University of Zurich[UZH]

Department of Geography

Time-oriented Visual Analytics in Soccer

An implementation and assessment of timeline representations for depicting the temporal value progression of dominance indicators in a soccer game

GEO 620 Master's Thesis

Luke Alan Albright
10-726-875

Supervised by
Prof. Dr. Sara Fabrikant, GIVA, Department of Geography, University of Zurich
Dr. Halldor Janetzko, GIVA, Department of Geography, University of Zurich

Faculty representative
Prof. Dr. Sara Fabrikant, GIVA, Department of Geography, University of Zurich

28.09.2017
Department of Geography, University of Zurich

Contact Information:

Luke Albright

In den Reben 20, 8708 Männedorf, Switzerland

Prof. Dr. Sara Fabrikant

Head of Geographic Information Visualization and Analysis Unit

Department of Geography (GIVA)

University of Zurich

Winterthurerstrasse 190, 8057 Zürich, Switzerland

Dr. Halldor Janetzko

Senior Scientist

Department of Geography (GIVA)

University of Zurich

Winterthurerstrasse 190, 8057 Zürich, Switzerland

Abstract

Over the past few decades soccer has significantly grown in its magnitude and popularity across the entire globe. Hence, teams are becoming more and more professionalized and can be viewed as corporate entities selling a specific entertainment-based experience. Such an evolution has led to increased efforts of clubs to develop their product by means of economic optimization, marketing or marketing. A similar development can be observed in the tactical and athletic preparation of players in order to be a competitive force throughout the season. Soccer clubs employ significant coaching staff to ensure the best possible facilitation and monitoring of the players' abilities and development. An essential dimension of this development is the heightened importance of supporting analytical tools which aim to help coaches and analysts in better understanding relevant game patterns. This increased knowledge aims to enhance the capabilities of decision makers in adapting their team's style of play in order to boost the chance of being victorious at the end of the game. The main objective of this work is to make a contribution to the current state of visual analytics in soccer. Most existing representation-based applications visually convey specific aspects of the game which are relevant for detailed observation. However, a vast amount of them do not emphasize the temporal dimension sufficiently, implying a research void which still is to be filled by new formats of visual analytic tools. Within the scope of this work several different timeline representation formats are implemented, which depict the value progression of dominance indicators over the course of a game. The core aim is to understand how well these representations can be applied as means of support to enhance the understanding of the game by domain-specific decision makers, namely coaches and analysts. The initial groundwork underlying this work is mainly a combination of the absence of time-oriented approaches towards visual soccer analytics as well as the promising concept of feature-based analysis (e.g. Janetzko et al., 2016) which offers intriguing opportunities of being enhanced further. As a result, the output generated by the proposed approach aims to convey information regarding the temporal evolution of game-related indicators as an intriguing visual alternative which can be easily integrated into an analytical workflow. Conclusively, this work makes a contribution to the current state-of-the-art by demonstrating a workflow which dismantles and factorizes a broad soccer phenomena such as dominance and subsequently computes and visualizes the identified indicators.

Preface

Working on and completing this work over the course of the last 365 days has definitely been an exciting and instructive journey providing me with the opportunity to obtain new skills and strengthen my abilities to structure and conduct a research project of prolonged duration. On this note, I would like to extend my sincere thanks to all of the people involved in supporting me throughout my work progress. Especially, I would like to mention the following key individuals who have contributed greatly to the timely and successful completion of my Master Thesis:

- My supervisor, Prof. Dr. Sara Fabrikant, who gave me the opportunity to write my Master Thesis at her unit (*Geographic Information Visualization and Analysis*). Through her positive and encouraging style of communication, as well as her helpful and valuable advice, she has played a facilitating role regarding the accomplishment of this work.
- My co-supervisor, Dr. Halldor Janetzko, who has put a lot of time and effort into meeting with me on a regular basis and significantly supporting the implementation part of this work comprised of the computational and graphical programming procedures which often were very new to me.
- The domain experts, namely Markus Wanner, David Andreoli, Stefan Goll, Thomas Jent and Marinko Jurendic, who in the early phases agreed to provide me their insights and knowledge through expert interviews. An extended thank you goes to Markus Wanner, Thomas Jent and David Andreoli for additionally participating in the evaluation panel. The information derived from all of these candidates' responses has been an important and integral foundation for this work.
- My parents, Thomas and Brigitte Albright, for enabling me to follow my interests by pursuing a degree of my choice and further providing a highly supportive working environment over these past years, for which I am truly grateful. On this note, I would also like to thank my sister, Lynn Albright, for significantly contributing to creating this environment.
- Finally, I would like to thank my girlfriend, Danielle Lanser, for her unremitting emotional support throughout this past year. She often helped me through long and tough working days and kept me calm and composed whenever challenging phases appeared on the horizon. Truly, thank you.

Table of Contents

Abstract i

Preface ii

List of Figures v

List of Tables viii

1.) Introduction 1
1.1.) Intended contribution of this work 2
1.2.) Assessment of the need for dominance visualization in soccer 2
1.3.) Research Questions 5

2.) Related Work 9
2.1.) Information Visualization & Visual Analytics 9
2.2.) Visualization of time-oriented data 13
2.3.) Perception of timeline displays 17
2.4.) Visual Analytics in soccer – An overview of the state-of-the-art 18

3.) Technique / Implementation 26
3.1.) Dominance Factor Identification (Requirements Analysis) 27
3.1.1.) Definition of dominance 27
3.1.2.) Implementation of exploratory expert interviews 30
3.1.3.) Identification of suitable dominance determinants (phase of divergence) 32
3.1.4.) Assessment and selection of optimal dominance factors for further utilization (phase of convergence) 38
3.2.) Dominance Factor Computation 42
3.2.1.) Foundations of the computational environment 43
3.2.2.) Required supporting pre-calculations 43
3.2.3.) General computational workflow 46
3.2.4.) Computation of Dominance Factor I: Distance-based proximity 50
3.2.5.) Computation of Dominance Factor II: Zone-based proximity 56
3.2.6.) Computation of Dominance Factor III: Numeric predominance 60
3.3.) Dominance Factor Visualization 65
3.3.1.) Selection of timeline visualization formats 65
3.3.2.) Technical implementation of timeline visualizations 71
3.3.3.) Output 80

4.) Results 81
4.1.) Case study of visualization output 82
4.1.1.) Option panel visualization 82
4.1.2.) Timeline visualization outputs 84
4.2.) Evaluation of visualization results 92
4.2.1.) Structure of evaluation procedure 92
4.2.2.) Evaluation of results 93

5.) Discussion 108
5.1.) Discussion of research question outcomes 108

6.) Conclusion / Future Outlook 117

Bibliography .... A

Appendix .... F

Personal declaration .... AA

List of Figures

FIGURE 1: ILLUSTRATION SHOWING THE VALUE OF VISUAL ANALYTICS ROOTING IN THE SYNTHESIS BETWEEN HUMAN- AND MACHINE-BASED ANALYSIS APPROACHES (KEIM ET AL. 2008)....12

FIGURE 2: ILLUSTRATION DEPICTING THE WORKFLOW OF TRANSFORMING A LINE-CHART INTO A HORIZON GRAPH REPRESENTATION (HEER ET AL., 2009)....16

FIGURE 3: HORIZON GRAPH REPRESENTATION OF TIMELINES DEPICTING THE VALUE PROGRESSION OF THE REGARDED SPEED FEATURE. (SACHA ET AL., 2014)....19

FIGURE 4: ILLUSTRATION DEPICTING THE THREE KEY FACTORS SHAPING THE TRAJECTORY IDENTIFICATION PROCESS (LIN SHAO ET AL., 2016)....20

FIGURE 5: GLYPH-BASED VISUALIZATION DEMONSTRATING A TEAM'S SEASONAL EVOLVEMENT (LEGG ET AL., 2012)....21

FIGURE 6: INTERFACE VISUALIZING DISTINCTIVE GAME FEATURES (STEIN ET AL., 2016)....22

FIGURE 7: REPRESENTATION OF FACETED VIEWS DEPICTING SPECIFIC GAME PHENOMENA INDIVIDUALLY (PERIN ET AL., 2013)....23

FIGURE 8: COMPARATIVE TIMELINE REPRESENTATION SHOWING THE VALUE-EVOLVEMENT OF DEFENSIVE PRESSURE-INDICATIONS FOR TWO COMPETING TEAMS (ANDRIENKO ET AL., 2017)....24

FIGURE 9: MIND MAP PORTRAYING A SELECTION OF DIFFERENT POTENTIAL DOMINANCE DETERMINANTS DISCOVERED WITHIN THE DIVERGENCE PHASE....32

FIGURE 10: SCHEMATIC ILLUSTRATION OF A HORIZONTAL PASS OVERPLAYING SIX OPPONENTS, INDICATING A PACKING VALUE OF SIX. 33

FIGURE 11: SCHEMATIC ILLUSTRATION OF TEAMMATES LOCATED IN A PLAYABLE POSITION WITH REGARD TO THE BALL POSSESSOR....35

FIGURE 12: ILLUSTRATION SHOWING TWO CONTRASTING SCENARIOS IN WHICH THE LEFT ONE DEPICTS A HIGH DEGREE OF SPATIAL FREEDOM WHILE THE SCENARIO ON THE RIGHT SIGNIFIES ITS LOWER COUNTERPART....36

FIGURE 13: MATRIX DISPLAYING THE ASSESSMENT OF INDIVIDUAL DOMINANCE FACTORS BY DOMAIN EXPERTS....39

FIGURE 14: ILLUSTRATION OF THE SPATIAL PROPERTIES OF A RECTANGULAR AREAL EXTENT ON THE FIELD....45

FIGURE 15: CODE SNIPPET SHOWING THE ASSIGNMENT OF THE INITIAL ATTACKING DIRECTION TO A PLAYER....46

FIGURE 16: SCHEMATIC OVERVIEW OF THE GENERAL IMPLEMENTATION PROCESS STRUCTURED INTO 4 SEGMENTS....47

FIGURE 17: CODE SNIPPET SHOWING A FEATURE INSTANTIATION STATEMENT WRITTEN WITHIN THE FEATUREFACTORY.JAVA-CLASS..49

FIGURE 18: CODE SNIPPET SHOWING AN EXEMPLARY FEATURE REFERENCE TO THE ZONAL PROXIMITY VALUE FOR THE RESPECTIVE HOME OR AWAY TEAM....49

FIGURE 19: SCHEMATIC FLOWCHART PORTRAYING THE SINGLE-SCALE VALUE INTEGRATION DEPENDING ON THE CURRENT BALL POSSESSION....50

FIGURE 20: SCHEME SHOWING THE NINE LOCATIONAL RELATIONSHIP CASES BETWEEN AN OBJECT (CENTER-POINT) AND THE RECTANGULAR EXTENT BASED ON COORDINATE REFERENCES....52

FIGURE 21: CODE SNIPPET SHOWING THE CREATION OF NEW FEATURES REFERENCING THE RESPECTIVE SIDE-BASED VALUE-ARRAY FOR THE LEFT AND RIGHT PENALTY BOX....53

FIGURE 22: FLOWCHART ILLUSTRATION SHOWING THE ASSIGNMENT OF THE CORRECT SIDE-BASED VALUE BASED ON THE ATTACKING DIRECTION OF THE OBJECT (CENTER POINT OR SINGLE PLAYER)....54

FIGURE 23: CODE SNIPPET SHOWING THE CREATION OF RESULTING DISTANCE-BASED PROXIMITY FEATURES FOR BOTH TEAMS..........55

FIGURE 24: FLOWCHART ILLUSTRATION SHOWING THE CALCULATION OF ZONAL-PROXIMITY VALUES. 57

FIGURE 25: FLOWCHART ILLUSTRATION SHOWING THE CREATION OF TEAM-DEPENDENT FEATURES BY CONCATENATING THE SIDE-BASED VALUES DEPENDING ON THE ORIENTATION OF THE GAME FOR THE REGARDED TEAM. 59

FIGURE 26: SCHEMATIC FLOWCHART DISPLAYING THE CALCULATION OF PLAYER-COUNTS BASED ON THEIR POSITION IN RELATION TO THE DEFINED RANGE AROUND THE BALL POSSESSOR. 62

FIGURE 27: FLOWCHART ILLUSTRATION DEMONSTRATING THE CALCULATION AND WEIGHTING OF DOMINANCE VALUES ON THE BASIS OF THE PREVIOUSLY DERIVED PLAYER-COUNTS. 63

FIGURE 28: ILLUSTRATION SHOWING THE THREE DIMENSIONS OF A TIMELINE DESIGN SPACE (BREHMER ET AL., 2017).. 66

FIGURE 29: ILLUSTRATION OF A TWO-TONE PSEUDO COLORING SCHEME IN RELATION TO NORMAL DISCRETE OR CONTINUOUS FORMATS (SAITO ET AL., 2005).. 69

FIGURE 30: FLOWCHART SHOWING THE CONFIGURATION OF THE PROGRAMMING STRUCTURE FOR GRAPHICAL TIMELINE VISUALIZATIONS OF SOCCER DOMINANCE INDICATORS.. 71

FIGURE 31: CODE SNIPPET ASSESSING IF THE CHOSEN VISUALIZATION TYPE CORRESPONDS WITH THE PIXEL-CHART REPRESENTATION..73

FIGURE 32: CODE SNIPPET DISPLAYING THE ALGORITHM USED TO ASSOCIATE THE CORRECT VALUES WITH THE CORRESPONDING PIXELS.. 74

FIGURE 33: TWO-TONE PSEUDO COLORING SCALE DISPLAYING THE COLOR-RATIOS FOR THE CORRESPONDING DOMINANCE VALUES RANGING FROM ZERO TO ONE...75

FIGURE 34: ILLUSTRATION OF THE OPTION PANEL ELEMENT IN THE INTERFACE. 83

FIGURE 35: PIXEL-CHART TIMELINE VISUALIZATION FOR A SINGLE TEAM (VALUE-SMOOTHING PARAMETER = 0.0). 85

FIGURE 36: PIXEL-CHART TIMELINE VISUALIZATION FOR A SINGLE TEAM (VALUE-SMOOTHING PARAMETER = 0.5). 85

FIGURE 37: COMPARATIVE TWO-SCALE PIXEL-CHART TIMELINE VISUALIZATION FOR TWO TEAMS (VALUE-SMOOTHING PARAMETER = 0.5). 86

FIGURE 38: COMPARATIVE SINGLE-SCALE PIXEL-CHART TIMELINE VISUALIZATION FOR TWO TEAMS (VALUE-SMOOTHING PARAMETER = 0.0). 86

FIGURE 39: COMPARATIVE SINGLE-SCALE PIXEL-CHART TIMELINE VISUALIZATION FOR TWO TEAMS (VALUE-SMOOTHING PARAMETER = 0.5). 87

FIGURE 40: TWO-TONE PSEUDO COLORED TIMELINE VISUALIZATION FOR A SINGLE TEAM (VALUE-SMOOTHING PARAMETER = 0.0). ..88

FIGURE 41: TWO-TONE PSEUDO COLORED TIMELINE VISUALIZATION FOR A SINGLE TEAM (VALUE-SMOOTHING PARAMETER = 0.5). ..88

FIGURE 42: COMPARATIVE TWO-SCALE TWO-TONE PSEUDO COLORED TIMELINE VISUALIZATION FOR TWO TEAMS (VALUE-SMOOTHING PARAMETER = 0.5). 89

FIGURE 43: COMPARATIVE SINGLE-SCALE TWO-TONE PSEUDO COLORED TIMELINE VISUALIZATION FOR TWO TEAMS (VALUE-SMOOTHING PARAMETER = 0.0). 89

FIGURE 44: COMPARATIVE SINGLE-SCALE TWO-TONE PSEUDO COLORED TIMELINE VISUALIZATION FOR TWO TEAMS (VALUE-SMOOTHING PARAMETER = 0.5). 90

FIGURE 45: LINE-CHART TIMELINE VISUALIZATION FOR A SINGLE TEAM (VALUE-SMOOTHING PARAMETER = 0.0). 91

FIGURE 46: LINE-CHART TIMELINE VISUALIZATION FOR A SINGLE TEAM (VALUE-SMOOTHING PARAMETER = 0.5). 91

FIGURE 47: COMPARATIVE TWO-SCALE LINE-CHART TIMELINE VISUALIZATION FOR TWO TEAMS (VALUE-SMOOTHING PARAMETER = 1.0). 92

FIGURE 48: COMPARATIVE SINGLE-SCALE LINE-CHART TIMELINE VISUALIZATION FOR FOR TWO TEAMS (VALUE-SMOOTHING PARAMETER = 0.5). 92

FIGURE 49: TWO-TONE PSEUDO COLORED TIMELINE REPRESENTATION WITH NINE HIGH DOMINANCE PHASES (VALUES > 0.75). 96

FIGURE 50: COMPARATIVE TWO-SCALE PIXEL-CHART REPRESENTATION. 100

FIGURE 51: COMPARATIVE SINGLE-SCALE PIXEL-CHART REPRESENTATION. 100

FIGURE 52: COMPARATIVE TWO-SCALE TWO-TONE PSEUDO COLORED REPRESENTATION. 101

FIGURE 53: COMPARATIVE SINGLE-SCALE TWO-TONE PSEUDO COLORED REPRESENTATION. 102

FIGURE 54: COMPARATIVE TWO-SCALE LINE-CHART TIMELINE REPRESENTATION. 102

FIGURE 55: COMPARATIVE SINGLE-SCALE LINE-CHART TIMELINE REPRESENTATION. 103

List of Tables

TABLE 1: TABLE SHOWING THE USER-COUNTED EXTREME DOMINANCE PHASES (MINIMAL < 0.25; MAXIMAL > 0.75) IN COMPARISON TO THE ACTUAL NUMBER OF HIGH/LOW DOMINANCE OUTCROPS. 95

1.) Introduction

FIFA World Cup 2014 in Brazil – national teams from all over the planet gather to compete in the world's most viewed sporting event. Thousands of people take over the streets of Rio de Janeiro or Sao Paulo, loudly chanting and waving their country's flags in joyful anticipation of the upcoming games. Around the globe, billions of spectators try to extract a portion of the live excitement by following the matches over media resources such as television or radio. Despite its magnitude, the described scenario is only a single instance of soccer fascination, demonstrating its tight grip on a sizable portion of the global population. Soccer, in its essence, is omnipresent and its reach is almost immeasurable. As Stein et al. (2016) state, this enormous popularity has triggered a significant process of professionalization over the past couple of decades. Hence, soccer clubs can now be viewed as corporate entities aiming to conduct their operations as successfully as possible. A major cornerstone within this evolution is the amplified importance of game analysis. Given the rapid development of tactical and physical factors, teams increasingly resort to analytical means to enhance their understanding and knowledge of the game, thus aiming to heighten their chances of winning. Decision makers within the soccer realm, such as coaches or analysts, usually don't simply want to know "what happened" (e.g. team A wins the game) but much rather wish to understand the underlying "why" dimension. This situation becomes even more prevalent in so-called invasive sports like soccer, where game patterns are more complex, thereby opening up more potential for information extraction (Stein et al., 2017). Throughout the last couple of years, significant advances in sensory technology have boosted the collection of spatially-referenced movement data, enabling new formats of match data that can be evaluated. Subsequently, numerous strains of soccer analysis exist which aim to illuminate various specific facets of the game. A novel domain within these approaches, demonstrating enhanced potential for information retrieval and knowledge building, can be found in visual analysis. Visualization aims to enable an observer (e.g. coach/analyst) to view specific aspects of the game, aiding the derivation of important match information. Ultimately, games become more understandable as they can be viewed through different representational elements (Page and Moere, 2006). The most prominent and widespread visual technique, which is prevalent among coaches and analysts, is the use of manually processed video recordings which are edited for analysis and presentation (Stein et al., 2016). The field of visual analytics is still very much open to for new implementation ideas representationally-based tools which can assist decision makers in their workflow.

1.1.) Intended contribution of this work

Within the context of visual analytics in soccer, this work aims to complement the currently existing work with an alternative visualization approach for specific game-related phenomena. More concretely, the key objective of the implementation is to visualize the temporal value progression of dominance indicators in a soccer game based on timeline representations. The superordinate question underlying this work is therefore *"which team is more dominant?"*. Dominance is clearly a very broad and vaguely defined concept that needs to be dismantled into more formal, expressive factors which are ultimately displayed. A further focus area lies in assessing the utility potential of timeline displays in the professional soccer workflow, as these formats can possibly be used to visualize the temporal evolution of additional game factors or patterns.

1.2.) Assessment of the need for dominance visualization in soccer

In justifying the approach proposed in this work, there are two integral aspects that shape its legitimacy. On one hand, the essential question needs to be argued as to why the professional soccer realm is in need of a timeline visualization of dominance indicators. On the other hand, it is of further interest to understand at what point in the training-game cycle such an application could potentially be utilized. Within the context-oriented part of exploratory interviews with domain experts, they provided interesting indications regarding these questions which are presented in the following two sections. Ultimately, these answers underline the validity and need for the contribution proposed in this work.

A.) How desirable is a timeline visualization of soccer dominance in the professional soccer realm?

There are two core dimensions which need to be considered when trying to identify the need and desire to utilize a timeline visualization of dominance indicators within the professional soccer realm. First, one needs to understand the general attitude among domain experts (potential users) towards (visual) analytic tools. Secondly, it is vital to assess the perceived potential value of a timeline visualization of dominance to the users. The following two sections provide an overview of the answers given to these integral questions.

1.) How desirable is the general concept of an analytical tool in the professional soccer realm?
Reviewing the opinions provided by all of the domain experts, one can clearly identify that an inherent interest exists in using analytical tools to support the quality of play and practice. However, contrasting this initial attitude, they also strongly stated that their perception of such applications is accompanied by notions of skepticism. For instance, M. Jurendic (Coach SC Kriens) states that team sports such as soccer are always prone to a significant degree of "chaos and randomness". D. Andreoli (Coach SC Buochs), on this note, further highlights the dependency of performance on mental factors which are difficult to quantify. Despite existing limitations in encoding the entirety of a soccer game, domain experts believe that there are many complementing puzzle pieces that need to be understood and optimized individually to help generate a successful final outcome. In summary, visual analytics can help make single game aspects more comprehensible by offering tools which can quickly and understandably convey a specific information aspect.

2.) How desirable is a timeline visualization of dominance indicators in a game-analysis process?
All experts interviewed agree that a timeline visualization of soccer dominance can potentially be utilized to identify interesting game situations. In addition, throughout the exploratory interviews, various indications could be observed as to the beneficial functionality such an application could contribute to soccer analytics. On the basis of this foundation, this work assumes that temporal dominance visualization facilitates the following capabilities: 1.) Display of the opponent's performance, indicating time-frames of weakness or intensity loss, etc. 2.) Overview of one's own team's performance, potentially correlated to physical fitness, event impact on dominance, etc. 3.) Foundation for making tactical decisions based on displayed temporal game patterns, such as player substitutions 4.) Potential to aggregate numerous games in order to view big-scale trends. 5.) Utilization as a communication tool to visually convey information to other staff, players or even spectators and media outlets. As a result, it becomes apparent that a timeline visualization of dominance indicators in soccer

potentially caters to multiple audiences. Hence, a further pursuit of the proposed approach is clearly valid and justified, particularly considering the validating responses of domain experts.

B.) Can a timeline representation of soccer dominance be integrated into the professional soccer workflow?

The timeline representation of dominance indicators should be viewed as an application which is used by domain experts to support specific analytical tasks. Therefore, it is vital to discern at what point such a tool can be integrated into the workflow of potential users. This key aspect is subsequently addressed by the following questions and answers.

1.) Which domain expert individuals would potentially make use of a dominance visualization tool?

The first question of interest is concerned with identifying which roles are mainly involved in making use of a proposed analytical tool that visually displays soccer dominance. The answer is very straightforward: *the head coach.* In his/her role as the primary team decision maker, the ability to grasp and synthesize numerous elements of information simultaneously is essential and calls for supporting measures, for instance, through visual analytics. Analysts are mainly focused on identifying and annotating interesting game situations which aims to simplify the head coach's work by shortening the processing time. Tasks including the interpretation and communication of game patterns predominantly fall within the remit of the head coach. Most clubs that are not affiliated with prestigious leagues, thus having small budgets, do not employ a team analyst, as is the case in this work. As a result, there is a need for developing analytical tools that cater to a larger variety of individuals, such as coaches and other supporting staff. Given the rapid pace at which new information has to be absorbed and shared with other stakeholders, visualization applications need to be simple and quickly comprehensible.

2.) At what point in the training-game cycle should a dominance visualization tool ideally be integrated?

All domain experts interviewed have presented a similarly structured training-game cycle comprised of an initial regenerative practice, tactical/development trainings throughout the middle of the week, and a final, short, high-intensity practice before the game, which usually takes place on the weekend. Hereby, it was stressed repeatedly that analytical tools are ideally not used within games or to shape practice schemes on a short-term time horizon. Generally, coaches make a shift from the previous game to the next game before the first practice of the week. Therefore, any analysis tasks for both games are mainly done in this timeframe. It is essential to note that coaches usually work on multiple timelines, one concerning the short-term game preparation while the other one focuses on mid- and long-term development. As supporting analytical tools should aim to support both of these temporal dimensions, it is challenging to exactly define an ideal point in the workflow procedure where a visual analytic application is to be incorporated. However, all experts have underlined the legitimacy of attempting to make these means of support an integral part of the training-game cycle.

The gathered responses to these questions clearly indicate that a temporally-oriented visual analytic application can make a meaningful contribution to the current state-of-the-art, which is further elaborated in Chapter 2. Clearly, there remain numerous aspects of ambiguity and uncertainty, such as the limitation of analytical tools, in grasping and explaining certain game-related phenomena. As a result, this work should be seen as an initial approach towards placing more emphasis on timeline displays in the sport analysis context, and laying the groundwork for further research attention on more detailed aspects regarding this matter. Going forward, the general structure of this work is based on a set of research questions which are articulated in the succeeding subchapter.

1.3.) Research Questions

The research questions underlying this work are shaped based on the following three integral variables: *1.) What shall be represented? 2.) How shall it be represented, in a visualization sense? 3.) What utility level does the representation have?* It is vital to understand that this work aims to explore a new design space for a specific domain-based application. Subsequently, the research questions are formulated rather broadly in order to cover the most important aspects of a first-level assessment of the utility value of the implementation result.

<u>1.) Research question regarding the visualized phenomena (what-dimension):</u>

Research Question I: Is it possible to effectively express the broad concept of dominance in the soccer sport through a set of identified and subsequently formalized factors?

Among various existing approaches for visual analytics in soccer, one can find numerous representational elements aiming to depict specific game aspects including passing, speed, shooting opportunities, etc. However, it appears that the domain has not undertaken many attempts to visualize broad-level trends catering to generic but immensely common questions such as *"which is the better team?".* One of the most frequently mentioned psychological concepts associated with such a question is the notion of pressure. Andrienko et al. (2017) have introduced an initial approach to visual indications of pressure but solely from a defensive perspective, underlining the potential for representing such holistic game phenomena. This work aims to complement and further enhance this approach by focusing on offense-related dominance. Dominance has not explicitly been addressed by research and, therefore, represents an intriguing niche for further exploration. An exception is the work by Taki and Hasegawa (2000) which introduced a visualization of dominant spatial regions in team games. Whereas their approach aims to depict patterns of dominance on a locational dimension, this work sets its focus on the temporally-oriented representation, implying an alternate visualization format. Ultimately the first research question is concerned with assessing how well a broad concept such as dominance can be effectively expressed by a set of identified quantifiable factors.

Hypothesis: Dominance is a vague and ambiguous concept within the soccer realm. Each individual has a unique interpretation of the term, implying a heightened complexity in finding a commonly acceptable format for expressing its meaning effectively. The initial hypothesis assumes that, despite the difficulty of defining dominance, a satisfactory first solution can be found that encompasses a commonly agreed understanding of the term.

2.) Research questions regarding the visualization type (how-dimension):

Research Question II A: Are timeline displays comprehensible and interpretable by the domain-originating observer (soccer coach/analyst)?

Research Question II B: Are visually more complex, color-encoded timeline displays of higher information value to the domain-originating observer than simple time series representations such as line-charts?

When comprehensively evaluating existing visual analytic applications in the soccer sport realm it becomes evident that the approaches vary widely. In order to convey valuable game information, concepts such as parallel coordinates (Janetzko et al., 2016), glyphs (Cava and Dal Sasso Freitas, 2013; Legg et al., 2012) or spatial representations (Stein et al., 2016) are used. However, research-based approaches very often disregard the temporal dimension in their representations. Whilst a few instances of related work include time serial displays, such as Andrienko et al. (2017) or Sacha et al. (2014), they do not explicitly focus on this element, as it only represents a small proportion of the resulting interface. A distinct dedication to the application of timeline displays in the soccer domain is currently absent. This work mainly aims to fill this void by explicitly focusing on such temporal visualization formats and their information value for potential users. Therefore, the second research question is concerned with understanding and assessing the relationship between the timeline representation and the domain-originating observer. As these individuals do not have any specific information visualization background, they are usually not exposed to more complex, color-encoded representation formats. Subsequently, the scope of this work includes identifying if such enhanced visualization has a higher utility value than simple time-serial depictions, such as scatterplots, line or bar charts.

Hypothesis: Timeline displays can be generally comprehensible and interpretable by the domain-originating user. Further, the domain-originating observers/users prefer a visually more complex, color-encoded timeline representation as it is more interpretable, informative and aesthetically pleasing.

<u>*3.) Research questions regarding the utility level of the visualization (application-dimension):*</u>

Research Question III: Is a timeline visualization of the dominance phenomena in soccer utilizable for a domain-originating observer (soccer coach/analyst) and, thus, able to be integrated into their respective workflow?

As explained previously, information visualization in a soccer analytics context mainly aims to support decision makers within the domain, such as coaches and analysts, in gaining a better understanding of occurring game patterns. Therefore, it is vital to assess the actual level of utility such an application offers to the potential user. Most related work points out that the core functionality of a visual analytic tool for soccer is the identification and representation of interesting situations throughout the course of a game (Sacha et al., 2014; Stein et al., 2015; Stein et al., 2016). Hence, such an application should simplify and shorten the coach's/analyst's workflow of delineating and annotating such occurrences which are often relevant for further, more detailed observation (for instance, through video). Moving forward, it is of essential interest to understand the utility potential that the implemented timeline displays have when applied in a soccer analysis procedure. This mainly includes the following aspects: 1.) For what purpose can a dominance timeline visualization be used? 2.) Which tactical decisions can be made based on a dominance timeline visualization? 3.) Where in the training-game cycle can a dominance timeline visualization be utilized? 4.) Is a temporally-oriented visualization (timeline) preferred over other formats, such as depicting the phenomena directly on a soccer pitch graphic (spatially-oriented)? Based on indications regarding these fundamental questions, an assessment shall be made as to the degree of utility serving the needs of potential domain users.

Hypothesis: Timeline visualizations representing soccer dominance can be used for several different purposes including visual communication to players, analysis of one's own and opposing teams, and performance assessment. Hence, they are not confined simply to being a means of identifying interesting game situations. Coaches can use the application to shape their game-related tactics according to the displayed dominance patterns. Most prominently, this involves player substitutions and managing the style of play (offensive, defensive, level of pressure, etc.) on the basis of dominance indications for the opposing team.

Moving forward, this thesis is structured as follows: Following this introduction, *Chapter 2* provides an overview of past related work, covering fundamental time-oriented visualization concepts and the state-of-the-art for visual analytics in soccer. *Chapter 3* (*technique/implementation)* is concerned with demonstrating the implementation workflow, comprised of identifying, selecting and computing dominance indicators and, further, visualizing them in a set of timeline representations. Thereafter, *Chapter 4* displays the computed results which are further elaborated upon, including expert evaluations. Finally, a discussion and conclusion (*Chapters 5 and 6)* aim to embed the outcome of this work in the bigger research picture by critically positioning the work in relation to the state-of-the-art, providing a conclusive outlook.

2.) Related Work

Defining a clear-cut research context for this work is a rather challenging undertaking, as it is an exploratory attempt to assess a new design space for a specific visual application display within the sphere of the soccer sport. At this point it is vital to highlight that the main focus shall lie on the visualization aspect, distinguishing it from the vast majority of past analytical research which has focused on understanding sport-based game patterns in more depth. The following subchapter will provide a comprehensive overview of past related work, segmented into various categories of focus. In a first step, the core fundamentals of visualization and visual analytics are described to set the broad context within which this work is to be positioned. The subsequent part aims to provide insight into the various existing techniques dealing with the depiction of time-oriented data, such as timeline representations, and further covers benchmark works focused on their visual perception. Finally, the-state-of-the-art for visual analytic approaches in soccer is reviewed to provide the thematic context within which this work aims to make a contribution.

2.1.) Information Visualization & Visual Analytics

A logical entry point into structuring the visualization realm is the so-called domain of *information visualization*. Hearst (2003) summarized various notions of the term into a sound definition as follows: "Information visualization is the depiction of information using spatial or graphical representations, to facilitate comparison, pattern recognition, change detection, and other cognitive skills by making use of the visual system. Mazza (2004) underlines the novel character of the discipline concerned with creating visual artefacts amplifying cognition. Whilst Hearst (2003) states that the role of the computer is *"merely of a mean that facilitates visualization"* it remains clear that modern information systems are prominent throughout this research field and manifested in many definitions as core elements of the representation-building process, in general (for instance by the *User Interface Research Group at the Palo Alto Research Centre*). It is an indisputable reality that this work is linked closely to instances of computational procedures highlighting the interrelation between mind-based visual capabilities and informatics.

Historically, examples of pictorial representations date back to the 16th century and the first benchmark works of William Playfair and J.H. Lambert appeared before 1800, conveying economic, demographic and other data by the means of graphics (Kmetova, 2010; Mazza, 2004). A further important contributor was Jacques Bertin, who made a first attempt to build a legitimate framework in the sphere of information visualization in 1983. In the same year, Edward Tufte shared his theory on maximizing information density in graphics, which was a fundamental benchmark for the research focus in this area (Mazza, 2004). Since then the discipline has seen rapid development in the numbers of readings, conferences and workshops and has flourished tremendously becoming a vital research area with heavy ties to computer-based methodologies (Kmetova, 2010). Purchase et al. (2008), however, note that the absence of a rooting theory calls for the need to define such areas of theoretical foundation.

When shifting to declaring the degree of value for a domain such as information visualization it is conspicuous that there is a high level of ambiguity in the understanding of the curriculum. While human and natural sciences aim to understand specific phenomena and disciplines, such as mathematics and physics based on logic and truth, information visualization doesn't have the objective of comprehending a certain domain. Rather, it tries to gain generic insights from collected data (Fekete et al., 2014). Fekete et al. (2014) state two essential arguments in their attempt to convey the benefit of such a research area:

Cognitive Argument: Vast data structures displayed as tables, textual collections etc. usually contain an enormous amount of information which is shown to the viewer in its full detail. Grasping the desired knowledge from seemingly boundless data compilations is extremely challenging and usually overloads the brain in its endeavor to process the materials efficiently. If one uses a well-designed visual model to transmit the necessary data adequately, the process of answering specific inquiries (e.g. In what year was the unemployment rate highest in Spain?) is propelled and simplified significantly (Fekete et al., 2014).

Perceptual Argument: The perceptual argument builds upon the premise of information theory that our visual system is the sense with the highest bandwidth (100 Mb/s) and, therefore, is the most suitable conduit to transmit information to our brain. Ware (2004) distinguishes two theories explaining the utilization of perceiving features through vision: The preattentive processing theory states that our low-level visual system can recognize certain optical features very rapidly and accurately (e.g. noticing a red circle among many blue circles). The Gestalt theory explains distinct principles that visual systems follow when trying to understand an image and, therefore, should be taken into consideration by visualization designers. These principles are **proximity, similarity, continuity, symmetry, closure** and **relative size.** In summary, one can say that the careful regard for perception-loaded factors can tap the enormous potential that information visualization has as a communicator of data.

Riccardo Mazza (2004) demonstrates the usefulness of information visualization by identifying three core areas of application.

Graphics for presentation: Graphic representations are powerful communicators that can potentially display facts in an understandable and perceivable manner.

Graphics for exploratory analysis: Graphic representations can serve as a means of finding and identifying underlying structure and hypotheses about the data (Röber et al., 2000). Hidden patterns can be visually revealed.

Graphics for confirmative analysis: Graphic representations can be used to confirm or reject hypotheses.

It is unmistakable that, despite the existing challenges of defining information visualization explicitly, the domain has its legitimate place in research underlined by various quantifications of its usefulness as seen above. Moving forward, it is vital to underline that the aim of this work is to construct a visual output that provides the user with a certain amount of information to enhance his/her knowledge. Therefore, visualization needs to be understood beyond its elementary rationale of graphically representing specific phenomena in a visually pleasing manner. In many cases, we commonly associate the deepened attempts of individuals or groups to obtain more knowledge about distinct areas or circumstances of interest with the term *analysis*. Importantly, the domain of visual analytics shapes a vital research perspective by interpreting visualization as a state of application, implying its utility value as a tool to extract important insights and information.

The basic definition of *Visual Analytics* has been stated by Thomas and Cook (2005) as "the science of analytical reasoning facilitated by interactive visual interfaces". Keim et al. (2008) add that visual analytics explicitly combines automated analysis techniques with interactive visualizations in order to effectively make sense of complex data sets and use the derived knowledge as a basis for decision-making. The publication goes on to describe essential goals of creating such representation-based tools: 1.) To derive insight from massive, often conflicting, datasets and synthesizing the respective information. 2.) To detect the expected and discover the unexpected. 3.) To provide timely, defensible and understandable assessments 4.) To communicate assessment effectively for action.

As demonstrated by Keim et al. (2008) (see Figure 1 below), the visual analytics domain aims to synthesize the benefits of machine- and human-centered approaches into robust data analysis capability. The combination of these two spheres clearly boosts the power and utility value of representational displays conveying information to the user and making it an analytical tool with high potential.

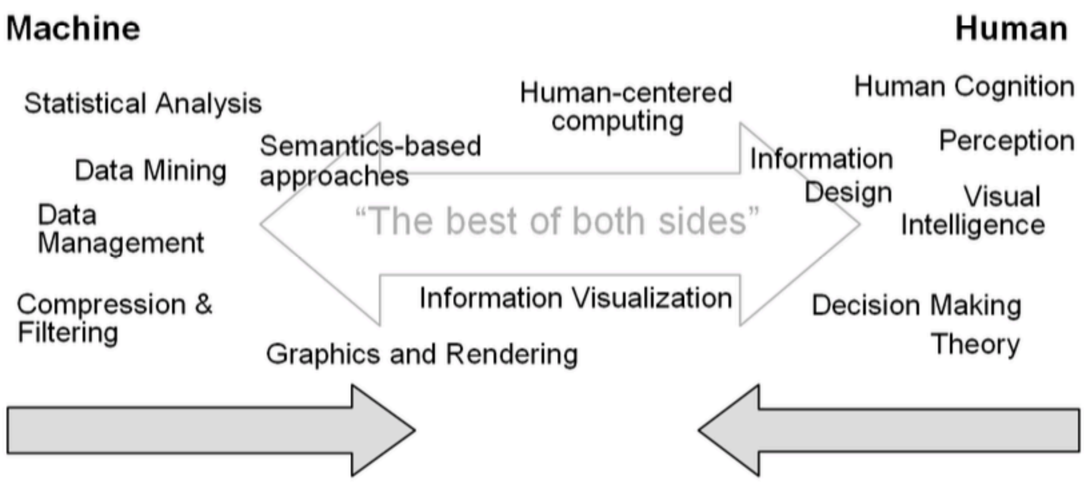


Figure 1: Illustration showing the value of visual analytics rooted in the synthesis between human- and machine-based analysis approaches such as cognition, perception, data mining or graphics and rendering (Keim et al. 2008).

Visual analytics appear to be a modern realization of visualization concepts, tightly linked to analysis and application. If we shift our attention to more current focal areas of the discipline, certain essential challenges become evident. Andrienko and Andrienko (2007; 2013), for instance, highlight the importance of visual analytics in the sphere of movement data.

Extending this notion, the omnipresence of spatio-temporal phenomena, similar to other visualization strains, are visible throughout published literature, such as Andrienko et al. (2013), Thomas & Kielman (2008) or von Landesberger et al. (2012)

2.2.) Visualization of time-oriented data

The researched timeline visualization of soccer dominance does not only imply the representation of information as such, but is further concerned with displaying its progression over the course of a game. As a result, the basic visualization concept is to be extended by the dimension of temporality. Aigner et al. (2011) make a clear distinction between the physical dimension of time and a time model in information systems. Modeling time is not aimed at perfectly imitating its physical dimension, but rather to provide a suitable reflection of the regarded phenomenon which can be utilized in a potential analysis task. On this note, Frank (1998 in Aigner) states that no single correct taxonomy of time exists, as it can be expressed in a vast array of different formats. There are numerous approaches to creating time-oriented visualizations which branch out into a variety of different design spaces. In their survey, Aigner et al. (2011) organize the various realizations of visualizing temporal data by defining key criteria categories as follows:

Frame of reference: The frame of reference dimension distinguishes between abstract and spatial data. Abstract data has no specific locational attribute, whereas its spatial counterpart inherently includes a *where* aspect.

Variables: This criterion categorizes the data used by the number of time-dependent variables. Univariate data associates one single value to each temporal primitive whereas multivariate data implies the representation of several variables at each time step.

Arrangement: Based on its characteristics, the temporal domain can be displayed in varying forms of arrangement. Naturally, time progresses linearly from the past to the future. However, certain time patterns are recurring (*e.g. weekly work schedule*) and, hence, are organized in a cyclic manner.

Time primitives: Primitives signify how data is related to time. One can differentiate between anchored and unanchored primitives. The first group can be pinpointed on an absolute scale, either as time instants (single point in time) or time intervals (specific portion of time). The latter group is shown in so-called time spans which are directed durations with no explicit scaling reference (e.g. 2-hour lecture).

Mapping: Time can be mapped based on the following two distinctions: Static representations, which show the data and the temporal dimension as a unified visual display, which does not change. Dynamic representations can change along the time continuum, ranging from a low discrete number of viewing frames for the respective points in time up to animation-like presentations which appear to be covering the extent of time more holistically.

Dimensionality: A presentation space can simply be classified as being either 2D or 3D. In two-dimensional representations temporality usually contains one axis that can vary in form and granularity while the data is usually signified along the opposing visual axis. Three-dimensional displays offer another dimension that can be exploited for enhanced 2D data representations.

These visual aphorisms build the basis for time-oriented graphics and are prevalent in many examples. Throughout the visualization sphere there exist numerous approaches to graphically displaying univariate, abstract data in relation to their temporal progression. Probably the simplest representation concept, which is prevalent among research and common media, is based on the depiction of data values in a two-dimensional Cartesian coordinate system. Hereby, the horizontal axis encodes time whereas the vertical dimension exemplifies the corresponding value-range. Common forms of two-dimensionally referenced visualizations include point and line plots such as bar and spike graphs (Harris, 1999). Over time these fundamental representation formats have been further enhanced. For instance, silhouette graphs attempt to fill the area below a plotted line-chart with a specific color shade and can additionally be stacked on top of each other creating a so-called layer area graph (Harris, 1999). Javed et al. (2010) aim to superimpose and braid such silhouette graphs, aiding the comparison of various temporal value progressions. Another basic attempt towards simply displaying time-series are *Sparklines* as proposed by Tufte (2006). Their key objective is to provide a broad overview of a variable developing over time through word-like graphics, which can be seamlessly integrated into text. Lee et al. (2010) have extended this idea by representing the evolution of textual key words by displaying them differently (font, size etc.) based on their number of occurrences (*SparkClouds*).

The previously described visualization formats can be categorized as fundamental approaches to displaying the temporal evolution of a specific data-value. Given the vast design space represented by the visualization domain, many further timeline displays exist which are all unique in their own way. The following segment's aim to provide a comprehensive overview of related work which has attempted to meaningfully depict time-oriented data.

Simple linear timeline representations: Tufte (1983) described the linear timeline concept as a very powerful visualization technique that dates back long before the first computers appeared. Hereby, time primitives are displayed as intervals implying that the visualization doesn't only express its location on the temporal dimension but also needs to be concerned with communicating its length (Aigner, 2011). Many other examples of linear timeline representations exist including *Gantt Charts* (Gantt, 1913), *Planning Lines* (Aigner et al., 2005) or *Time Annotation Graphs* (Kosara and Miksch (2001). These examples are predominantly useful in planning a multi-faceted undertaking of lengthy duration such as a construction or software project. In addition, Chittaro and Combi (2003) have founded their *Paint Strips* design on the core concept of timelines. Another interesting variation can be found in the *TimeNets* approach by Kim et al. (2010) where color bands indicate the time extent of specific phenomena while vertical drop lines are concerned with connecting them according to their semantic relationship.

Color-encoded linear representation of time-oriented data: Color-encoding is an essential part of many timeline representations. For instance, higher data-values are often displayed through darker, more saturated color shades whereas lower data-values evoke an inverse color association. Saito et al. (2005) propose a two-tone pseudo coloring technique, which depicts the value progression by a set of different color bands. Each data point is subsequently represented by a distinctive ratio of a lower and a higher color shading intensity. Extending this approach, Reijner (2008) developed the so-called horizon graphs to aid the visual comparison of a variety of different time-dependent variables. Figure 2 demonstrates the construction of a horizon graph whereby a set of operations is applied to the initial line-chart view to derive a color-encoded timeline visualization. An interesting alternative for linear, color-based timeline displays can be seen in the *Arc Diagrams* introduced by Wattenberg (2002).

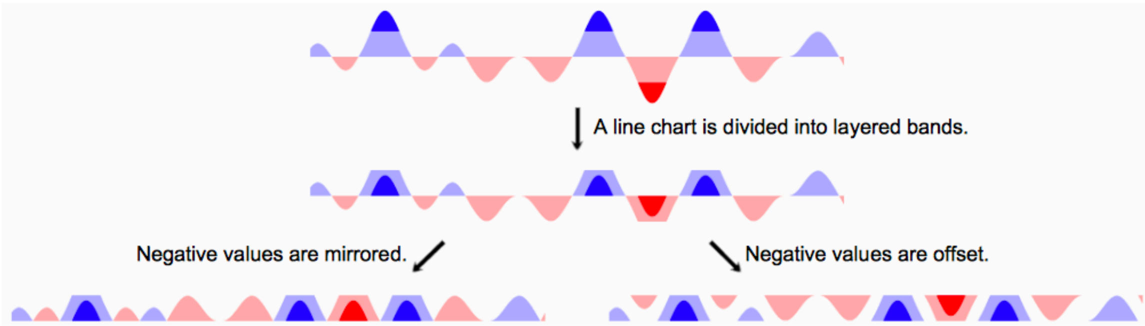


Figure 2: Illustration depicting the workflow of transforming a line-chart into a horizon graph representation. The value range of the initial line-chart is partitioned into a set of color-bands which increase in intensity towards the maximum and minimum. Each peak orientation is subsequently encoded with a different color hue (red or blue). The negative values can either be mirrored or offset to display a resulting horizon graph visualization (Heer et al., 2009).

Many linear timeline representations are based around the concept of pixel-based visualizations. Hereby, each color-based data value expression is assigned to a set of pixels (Keim, 2000). Pixel-charts have become popular formats of visual communication (Oelke et al, 2011; Keim et al. 2007; Keim, 2000). On this note Aigner et al. (2011) highlight that pixel-based visualizations are the most space-efficient way of displaying data. A prominent example of a suitable pixel-based visual format which can be utilized to depict large time-series is the *Recursive Pattern* approach by Keim et al. (1995). Shimabukuro et al. (2004) further introduced a *Multi Scale Temporal Behavior* technique which aims to create a visual matrix consisting of different temporal dimensions represented by varying color-encoding of the pixels. Another similar technique is propagated by Mintz et al. (1997), whereby pixel-based *Tile Maps* display a time-dependent variable in relation or proportion to different granularities of the temporal dimension (e.g. month and day).

Cyclical representations of time oriented data: Contrary to the linear timeline representations shown previously, cyclical arrangements of time-oriented data are usually visualized in a more circular manner. Weber et al. (2001) have developed a *Spiral Graph,* whereby the time axis is displayed as a spiral. Tominski and Schumann (2008) have attempted to integrate the two-tone pseudo coloring approach by Saito et al. (2005) into a cyclical visualization resulting in an *Enhanced Interactive Spiral* depicting each data point with a combination of two colors. Further spiral-based approaches include the *SpiraClock* (Dragicevic and Huot, 2002), the *SolarPlot* (Chuah, 1998), the *SpiralDisplay* (Carlis and Konstan, 1998) and *Ring Maps* (Huang et al., 2008; Zhao et al., 2008).

It becomes evident that there are numerous ways to visually represent the temporal progression of a data-variable through timeline displays. Alternatively, the *Small Multiples* approach by Tufte (1983) can be viewed as a contrasting concept aiming to visually convey the temporal evolution of a certain phenomenon. Contrary to the previously listed timeline visualizations, *Small Multiples* do not integrate the time-dimension in a single representation but rather show a new data frame for each temporal instance. Ultimately, the amount of time-oriented data visualization techniques is very copious. For further elaboration on these different approaches it is recommended to consult the survey made by Aigner et al. (2011).

2.3.) Perception of timeline displays

There are only few instances of published literature that set their focus on the visual perception of timeline displays. Heer et al. (2009) have concluded that line-chart representations are generally prone to higher estimation errors than color-encoded visualizations, such as horizon graphs, when conducting interpretational tasks. Further, within horizon graph timelines, estimation time and errors are significantly increased with additional color bands evoking more visual complexity. Finally, estimation errors decrease with increased chart height implying that charts of larger vertical extent are more conducive to analytical tasks. However, the estimation time remains shorter with decreasing chart height. Javed and Elmqvist (2010) have analyzed the graphical perception of multiple time series based on key criteria such as space management (shared or split), space per series, identity and visual clutter. They concluded that shared-space techniques were superior for tasks with a local visual span (specific area of the timeline) whereas split-space techniques were more conducive to tasks involving a dispersed visual span involving the entire graphic.

2.4.) Visual Analytics in soccer – An overview of the state-of-the-art

Visual analytic approaches in team sports have only recently started gaining more attention from a research perspective (Herdal et al., 2015) Hence, the young discipline is still at a very early stage and sets a fruitful breeding ground for novel applications which can potentially assist domain experts in gaining a better understanding of the game and improve their decision-making. Soccer, being the most popular sport in the world, is evidently no exception to this trend as more and more coaches and analysts are becoming interested in the usage of supporting visual tools. The following segments aim to provide an overview of the current state-of-the-art for visual analytic approaches within the soccer realm, which ultimately shall be complemented by the contribution of this work.

A.) Feature-driven visual soccer analytics

A significant contribution to the sphere of visual analytics in soccer is the paper *Feature-Driven Visual Analytics of Soccer Data* by Sacha et al. (2014). Its core aim is to support coaches and analysts in detecting interesting and relevant game situations which could be subject to further, more in-depth analysis. An interactive system is presented which shall facilitate the exploratory analysis of soccer data. The work is founded on a feature-based concept implying that the complex game of soccer can be dismantled into an array of different features such as speed, position, distance to ball or opponent, etc. These distinct features are seen as indicators expressing certain elements of the dynamic spatio-temporal team sport phenomenon, hence, factorizing it and making it more tangible. The authors highlight the adaptability and flexibility of the program regarding its applicability to varying data typology. More concretely, the designated system involves three essential levels of analysis. The *Single Player Analysis* is mainly concerned with detecting similarities in the feature-evolution of a player and partitioning the game accordingly, such that specific game phases can be discerned. Within the concept of *Multi Player Analysis,* the work proposes two specific ideas linked to visualization in order to observe the movement and performance of multiple players. By utilizing horizon graphs, displaying the feature values for each game participant along a time axis (see Figure 3), the temporal development of game phenomena plays a central analytical role. A second observation of interest displayed inside the *Multi Player Analysis* is the depiction of constellations and formations providing indications of collaborative movement patterns. The third strand of analysis is primarily concerned with the delineation of specific game-related events by means of observing the development of specific features. On one hand, each event is associated with a specific arrangement/configuration of feature values which can viewed in more depth to create event-specific knowledge. Complementing this concept, the *Similar Phase Analysis* seeks to encode specific situations by their feature value configurations.

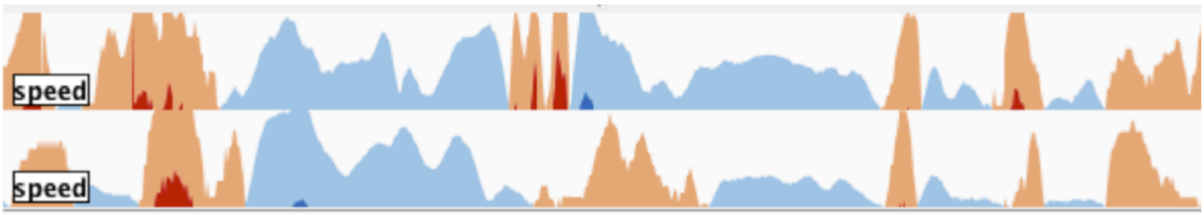
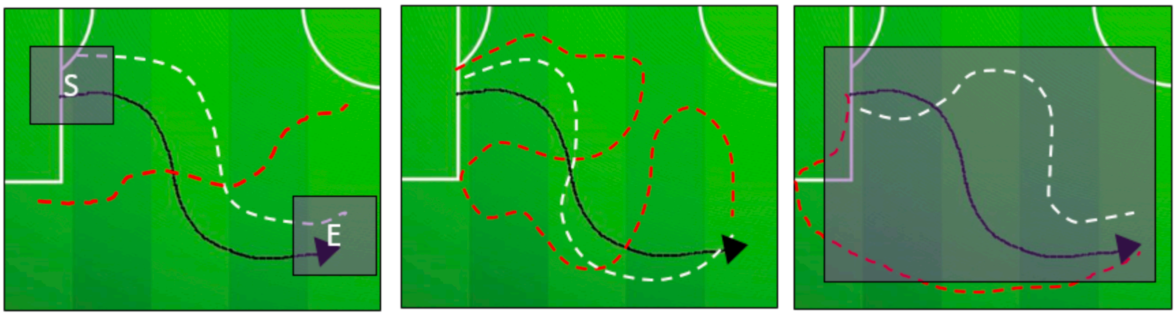


Figure 3: Horizon graph representation of timelines depicting the temporal value progression of the regarded speed feature (Sacha et al., 2014).

Finally, the feature-driven work by Sacha et al. (2014) can be seen as an integral approach towards building visualization interfaces that allow the exploratory analysis of soccer data on many different levels. The possibility of varying between different observation targets (single/multi player, events) and numerous features allows a holistic access to generating game-related knowledge on many different levels. The visual exposition is conducted by direct rendering on a schematic football pitch (trajectories etc.) or through time serial representations which, in its concatenation, gives an interesting combinatory portrayal. Given the ambitious attempt of dismantling the complex game of soccer into a multifaceted feature space, there still exists inherent potential to further create new feature factors that might display the game from another perspective. The work *Visual Soccer Analytics: Understanding the Characteristics of Collective Team Movement Based on Feature-Driven Analysis by Stein et al. (2015)* gives a complementary insight of this intriguing concept of feature-based soccer analysis.

B.) Visual-Interactive Trajectory Search

Similar to the feature-driven analysis, the paper *Visual-Interactive Search for Soccer Trajectories to Identify Interesting Game Situations* by Lin Shao et al. (2016*)* is concerned with identifying game situations of heightened interest. The provided interface is based on the concept of board sketches enabling the user to manually draw a desired trajectory inside the displayed field. The program aims to detect specific trajectories similar to the previously drawn one by assessing three specific factors of resemblance as shown in Figure 4. Candidate trajectories need to have a similar direction (start and end point), length and marginal restriction to be considered. As a result, the most akin trajectory will be displayed next to the initially drawn one with potential further attributes. This novel visual search system provides an interesting functionality to the analysis of soccer. The high degree of direct interaction aims to make the program simple to use and, therefore, delivers quick results to the user. It has proven to be an effective tool that aids the analytical process by retrieving specific events. However, the authors concluded that the retrieval of a single trajectory is insufficient and further movement or attributive data needs to be provided.



(a) Similar start and end position. (b) Similar trajectory length. (c) Similar marginal restriction.

Figure 4: Illustration depicting the three key factors shaping the trajectory identification process comprising of the start/end position, length and margin of restriction (Lin Shao et al., 2016).

C.) Visual Soccer Analytics through Parallel Coordinates

Enhancing Parallel Coordinates: Statistical Visualizations for Analyzing Soccer Data by Janetzko et al. (2016) provides a novel representational approach to feature-based visual analytics in soccer. The simultaneous observation of various game features results in a high-dimensional data space and triggers a desire for effective abstraction. Parallel coordinate plots vertically align the different features next to each other and use colored lines to connect the value points for a time instance. Ultimately, a certain combination of values (e.g. high distance to the ball, low speed, low distance to opposing goal) results in a specific game phase comparable to the feature-driven analysis. Similar value pairings resulting in a specific game phase are schematically encoded with a certain color. Founded on this partitioning scheme, the visualization aims to be used as a means of identifying interesting game events/phases. It poses a visually pleasing, intriguing alternative to standard feature-based event retrieval. It is further important to bear in mind the implied complexity of parallel coordinate plots for non-domain users which can impact the usability within the soccer analytics realm.

D.) Glyph-based soccer-related visualizations

An alternative visualization technique that has found some degree of application throughout soccer-related visualizations are the so-called *glyphs.* Legg et al. (2012) state the high value of this particular visual design form by making excellent use of human perception and cognition, hence, displaying high understandability. Glyphs are symbols that depict a certain set of data by incorporating different visual objects. The following two works are prominent instances of utilizing glyph-based representations in a soccer-context. Cava and Dal Sasso Freitas (2013) made use of a glyph-visualization in their work *Glyphs in Matrix Representation of Graphs for Displaying Soccer Game Results.* Within the linked SoccerMatches application a matrix configuration has been created that provides an overview of a whole league season. As can be observed in Figure 5, the glyph symbols encode various elements of information regarding games between team pairs in the matrix.

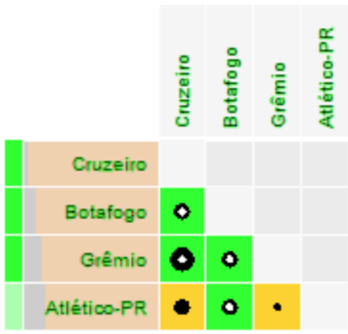


Figure 5: Glyph-based visualization demonstrating several teams' season. The color indicates the game outcome (win, loss, draw) for the home team listed horizontally on the left. Further the white and black circles display the amount of goals scored by the home and away side respectively (Legg et al., 2012).

The glyph-based visualization in this particular application is utilized to show high-level trends of the course of a whole season. Based mainly on scores, it does not allow in-depth analysis of the soccer game itself but potentially indicates long-term patterns that can be used to identify games or time frames where a more thorough insight could be valuable. Given the result-oriented data and visual appeal, this visualization form could be appealing in the context of sport media and journalism.

E.) Analysis and Annotation of Soccer Matches

A more direct attempt to establish a visual interface catering to visual analytics within the soccer realm is the work *Director's Cut: Analysis and Annotation of Soccer Matches* by Stein et al. (2016). The core objective of this specific approach is the appropriate transformation and visual representation of soccer motion data to assist analysts in assessing players and detecting interesting match patterns. Ultimately, the main contribution of this work is the implementation of effective automated annotation methods that dynamically analyze certain factors of the game. Combined with rule-defined annotation approaches, the process of manually analyzing and annotating interesting game situations is accelerated significantly.

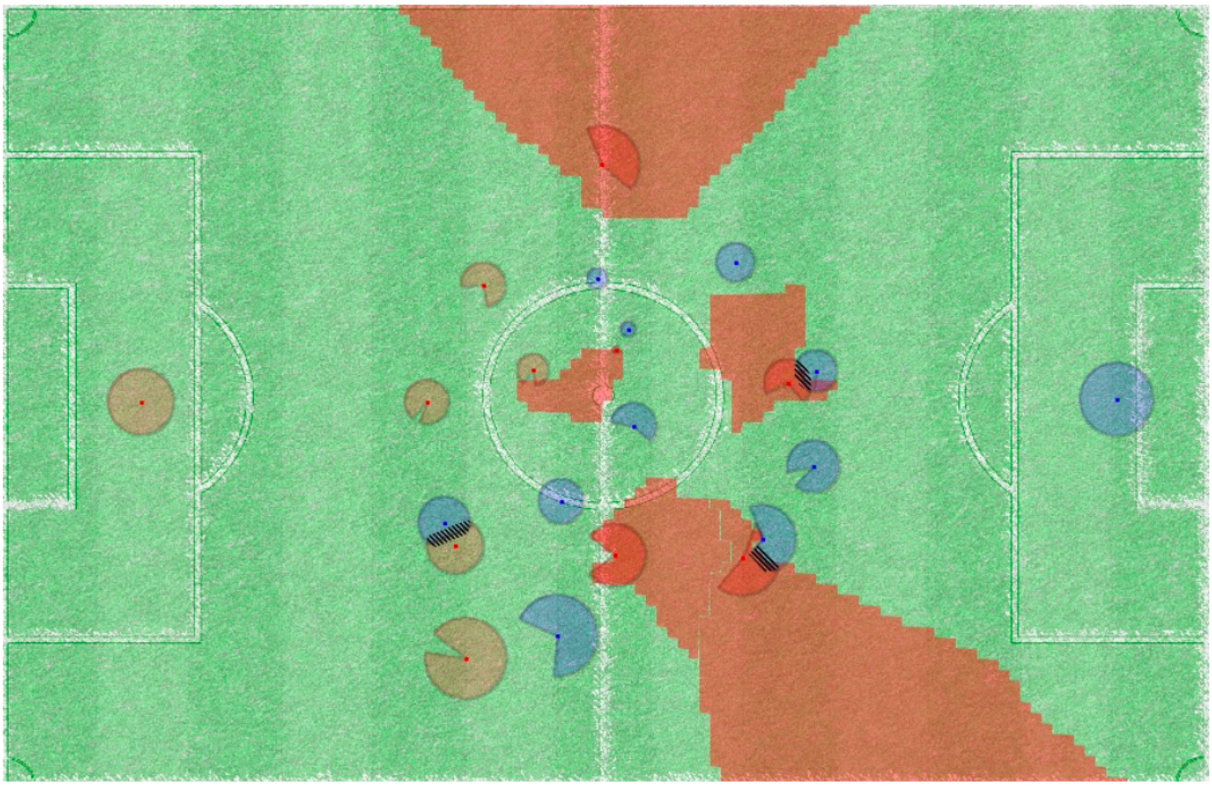


Figure 6: Interface visualizing distinctive game features including free spaces (red area), the area controlled by a specific player (blue/reddish circle) and the interaction (duel) space between two players (black overlapping area) (Stein et al., 2016).

Figure 6 above shows the resulting interface depicting a set of selected features accordingly. This type of visual representation displays game-related information directly on a soccer pitch rendering space. Subsequently, analysts or coaches can test their hypotheses derived from game observation directly on a field-based visualization. Such an approach provides a very effective and understandable view into relevant scenarios. Also, the notion of communicating information to other individuals is strongly supported since the application depicts the continuous movement on the pitch over the course of a game. In this method, there are inherent similarities to video sequences but, in addition, it provides informational and knowledge-building elements. The soccer-field view allows the display to be understood simultaneously by many people with soccer-related backgrounds without requiring an effort to learn about a new, specific form of information visualization.

F.) Multiple-View Analysis of Soccer Matches

Many visualizations cater to a certain aspect of the game that analysts seek to represent. In their work *SoccerStories: A Kick-off for Visual Soccer Analysis,* Perin et al. (2013) aim to create an application that integrates several different views in order to cover a whole array of different game scenarios with their respective representations. The basic idea underlying this approach is the partition of the game into a concatenation of specific phases. Each phase targets a coherent element of the game flow, such as a corner kick with a resulting goal scoring attempt or a fast-paced counter attack. To provide better insight into these semantically-loaded game segments, different essential game factors are individually represented through a fitting visualization as illustrated in Figure 7 below.

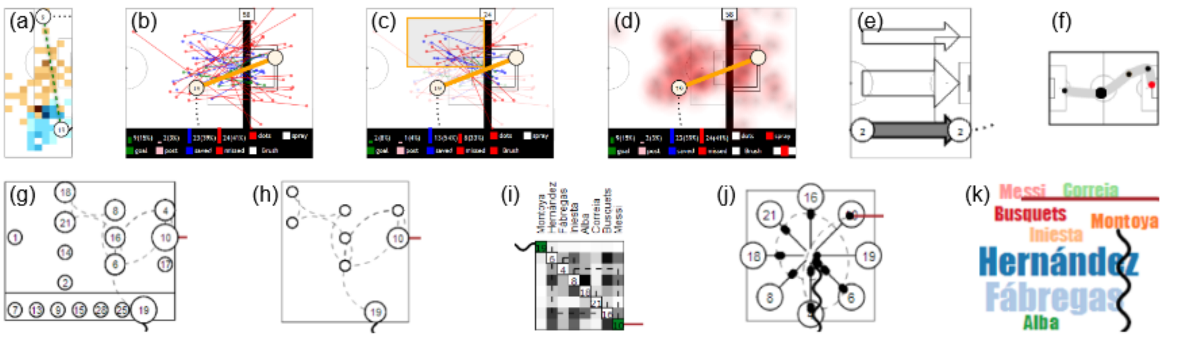


Figure 7: Representation of faceted views depicting specific game phenomena such as shot distributions, long runs, corners/crosses and pass clusters on individual visualization displays (Perin et al., 2013).

A timeline and a view of the global flow of phases incorporates all phases in a concatenated display providing the possibility to select them individually. For each phase, the relevant linked faceted views of the various game indicators are shown complemented by additional player attributes on the side. This allows in-depth exploration of many key factors related to a game phase, creating holistic information and knowledge about such a thematic and temporal instance of the game. With the *SoccerStories* application Perin et al. have created an intriguing approach to representing a high quantity of data simultaneously in order to increase the informational value for the observer. The work not only incorporates multiple factors into its computation, but shows them all at the same time in a single interface. Also, the novel approach of not simply detecting certain events, but partitioning the whole game into a sequence flow of annotated phases, provides an interesting variant to identifying desired game scenarios. Given the high information volume displayed through numerous simultaneous visualizations, the question remains as to how comprehensible and readable such an interface is to the viewer.

G.) Visual analysis of pressure in football

Perhaps the most novel work in the field of visual analytics in soccer is the paper *Visual analysis of pressure in football* by Andrienko et al. (2017). They propose the quantification of "pressure relationships" emerging within a game. Hereby, pressure is understood to be exerted by the defending team upon opposing players and the ball, respectively. Ultimately, a set of static and dynamic visualizations combined with interactive query tools aid the user in retrieving significant game information of interest. For instance, a two-dimensional time histogram demonstrates the calculated pressure values on the ball for the teams involved as is illustrated in Figure 8 below.

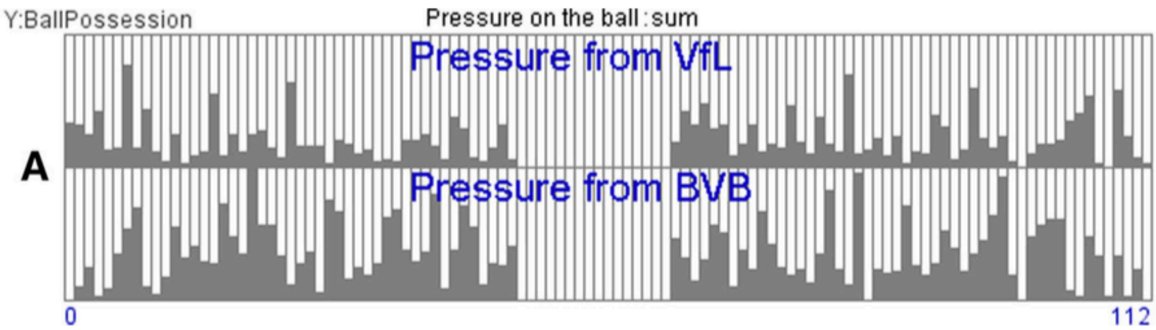


Figure 8: Comparative timeline representation showing the value evolution of defensive pressure for two competing teams (BVB and VFL) (Andrienko et al., 2017).

Further pressure visualization formats include density maps rendered on the soccer pitch, which signify the spatial distributions of players and high-pressure zones. An interactive time series display allows the user to identify and select game phases of interest based on query conditions. The approach outlined by Andrienko et al. (2017) is concerned with providing exact quantification of defensive pressure, thus setting the focus on a rather novel aspect of the game which has not yet been covered thoroughly. From a visual perspective, the synthesis of temporal (time series) and spatial (density maps) representations is an intriguing idea providing multiple dimensions of data depiction which can highlight varying aspects of game information.

3.) Technique / Implementation

As highlighted in the previous chapter, the prime research question of this work is concerned with identifying a suitable computable expression of dominant behavior within the soccer sport. The second core objective is to optimally visualize such a dominance concept in a timeline display which shall ultimately be utilized by coaches conducting their game analysis by means of visual analytics. This implies a multi-step process where, initially, the coarse understanding of soccer dominance is dismantled into a set of formalized factors which are subsequently computed. As a result, the dominance notion is expressed numerically allowing it to be utilized for successive steps of processing. The final product is a set of timeline visualizations which are to be evaluated and ranked by experts with regard to their utility potential. Hence, the implementation workflow consists of the following three segments:

1.) Dominance Factor Identification

The first partition of the workflow is signified by the dominance factor detection phase. With the help of domain experts, potential dominance indicators are identified and assessed on their contextual applicability and computational legitimacy. The result is a set of selected suitable factors that express dominant behavior in soccer as effectively as possible.

2.) Dominance Factor Computation

The second step of the workflow aims to compute the previously returned dominance factors by asserting them in a logic-based arithmetic scheme. Each factor generates a set of resulting numeric value-sequences, wherein each instance shows the indication of dominance for a specific time step in the game.

3.) Dominance Factor Visualization

In a concluding step, the numeric value-expressions conveying the notion of dominance are graphically displayed in a number of different timelines. These visualizations represent the final products of the implementation and are ultimately evaluated by potential users.

The following three subchapters provide detailed insight into the essential elements of the workflow.

3.1.) Dominance Factor Identification (Requirements Analysis)

One of the main parts of the implementation process in this work is focused on identifying suitable factors that express the notion of dominance in a soccer game. Initially, the coarse idea of dominance is defined and placed into a soccer context. Subsequently, with the help of expert interviews, this concept is further refined into factorial elements. Ultimately, in the concluding (convergence) phase, the determinants with the highest potential for further processing (computation/visualization) are delineated and selected.

3.1.1.) Definition of dominance

As previously stated, the overall aim of this thesis is to grasp and visualize the concept of dominance within a soccer game. But what does a generic term such as dominance actually mean in the context of a team sport? Before implementing methodological procedures, it is vital to examine the idea of dominance in further depth in order to understand its usage throughout the scope of this work. This includes addressing two integral questions: 1.) What does the term dominance signify on a very broad level and how does it manifest itself throughout other disciplines? 2.) How can the idea of dominance be understood and applied, specifically in the context of soccer?

1.) A broad definition of dominance and its existence in various domains

The Oxford Dictionary (“Dominance”, n.d., para. 1) defines the term dominance as the general notion of having power and influence over others. This implies a situation whereby the actors involved in the regarded scenario can be distinguished by their role of power, strength, and influence within the composition of individuals or groups. Therefore, the idea of dominance, in its essence, expresses the superiority of one subject over another.

The term dominance has been utilized and applied to a variety of disciplines. One can observe and differentiate two significant research poles. On one hand, the idea of dominance has played a vital role in Biology and Social Theory. Specifically, this has been the case in the focus area of genealogy, where the delineation of dominant strands can be observed within a gene pool (King et al., 2013). Similarly, *ecological dominance* demonstrates understanding as to which specie comprises the more numerous group in an ecological community (Clapham et al., 2006; Dong et al., 2010; Flinn et al., 2005). *Ethnological dominance* extends this observation by incorporating behavioral factors, implying a psychological element within the dominance discourse. On the other end of the spectrum, dominance is also strongly manifested in a more numeric fashion. In a mathematical sense, for instance, stochastic dominance is a form of ordering or ranking gambles according to their superiority (Bawa, 1982; Hadar and Russell, 1969; Linton et al., 2005). A significant extension of this numeric foundation is so-called *game theory,* concerned with the interaction of decision makers using certain strategies (Binmore, 2007; Osborne and Rubinstein, 1994). Thus, a strategy to an abstract decision problem is seen as dominant if it maximizes a player's pay-off, regardless of the opposition (Hargreaves-Heap and Varoufakis, 2004). Subsequently, one can say that dominance is a real-life phenomenon that appears in all sorts of scenarios of daily life. Operationalizing these instances numerically and declaring them through a game context shifts dominance towards being a desirable strategic trait that needs to be understood and, in some sense, calculated.

2.) How can the term dominance be applied in a soccer sport context?

As we shift towards understanding the application of the term dominance in a specific soccer sport context, this concretely implies that within this step we are rationalizing and legitimizing its usage. In common language, the broadly defined question as to "who is better than whom?" is omnipresent. The game environment of soccer is no exception - even more so, it carries such a notion strongly. The wide-reaching discharge of wordings such as *better* and *stronger* than the counterpart actor can often be directly linked to the score. If team X beats team Y by a score of two to zero, team X will be perceived as the stronger side. While the score is ultimately what counts, the game of soccer contains many more complex factors embedded in a constantly changing spatio-temporal environment. It is not only of interest which team eventually wins the game, but rather which party is perceived to be more dominant throughout the course of the game.

If we revisit dominance from a biological and sociological perspective, we can extract certain elements and concepts of its meaning that can be applicable in a sport setting. The notion of a team's performance being superior is, for instance, linked to factors concerning its physical abilities. Much more prominently, however, the concept of game-related dominance is tightly correlated with psychological and behavioral elements corresponding to ideas such as social order, etc. Individual players or whole teams will act a certain way, implying a style of play that has direct influence on the perceived dominance within the game. In that sense, team/player behaviors can also be affected by the dominance of others, indicating that the way a team/player acts is highly dependent on the perceived state of dominance. Subsequently, human conditions, be they physical or psychological, are a prominent element accompanying actions of dominance/superiority as seen throughout the sociological or biological realm. These circumstances further legitimize usage of the term *dominance* in the soccer sport from a human-centered perspective.

It is also evident that more numerical interpretations of dominance are very applicable to sport-related themes. Most prominently, this is displayed by game theory, which ultimately abstracts games into discrete rational elements and corresponding decision alternatives. This leads to the process of defining strategies which are useful to be dominant. The game of soccer is, in its dynamic character, significantly more complex and multifaceted than the often-simplified decision-making problems simulated in game theory. However, in order to strategize the soccer sport further, with respect to achieving a more dominant performance level, it is crucial to establish a more rational understanding of such a spatial-temporal phenomenon which again corresponds to the core objectives of this thesis. Gintis (2000), for instance, uses soccer as an applicable instance of the game theory approach while highlighting the fact that this particular concept can be used with numerous real-life examples. Therefore, one can confidently state that the notion of a *dominant strategy* in its meaning is transitive to the team sport environment to a certain degree.

In summary, the usage of the dominance term throughout this work comprehends all of the definitions provided above to some extent. Viewing soccer as social phenomenon, involving a set of actors and their physical, psychological and behavioral traits, taking place in a distinct extent of space, leads to the idea of dominance as a spatio-temporal occurrence. As a result, given the context of this work, dominance can be defined as the perceived superiority of one team over another based on the way they act throughout the game, being strongly influenced by physiological and mental capabilities, behavioral features and game-related skills. This view correlates heavily with the biological-social construction of the term. One of the central aims of this thesis is to dissect the broad idea of dominance and further attempt to express it through visible and operationalized factors, shifting the general task in a more numerically oriented direction. This procedure implies measures of simplification and mathematical/logical reasoning concerning the understanding of dominance, which again leans on the ideas of strategic and stochastic approaches. A key desired outcome of this work is to grasp dominance as a social/human-centered phenomenon and depict it in a numerical manner such that it can be processed and utilized further. It should, therefore, be conceived as a synthetic, integrative definition that incorporates the different forms of understanding hailing from various research domains.

3.1.2.) Implementation of exploratory expert interviews

In order to deepen the contextual understanding of the dominance concept in a professional soccer environment, exploratory expert interviews were conducted throughout the primary stage of this work. The main focus of this undertaking was to generate sufficient knowledge about how visual analytics can be incorporated into the workflow of soccer coaches and analysts, potentially based on their previous experience with such tools. Further, these dialogues aim to aid the detection of suitable dominance factors that can be utilized for further computational processing and, ultimately, effective visualization. The interviews were carried out with the following five domain experts (more in-depth profiles provided in the appendix):

1.) David Andreoli – Head Coach SC Buochs (1. Liga Classic – Switzerland)

2.) Stefan Goll – Head Coach YF Juventus (Promotion League – Switzerland)

3.) Thomas Jent – Head Coach FC Baden (1. Liga Classic – Switzerland)

4.) Marinko Jurendic – Head Coach SC Kriens (Promotion League – Switzerland); Specialist and Assistant to Technical Director at the Swiss Football Association.

5.) Markus Wanner – Head Coach FC Seuzach (1. Liga Classic – Switzerland)

The interview structure is generally founded on the work by Bogner, Littig and Menz – *Interviews mit Experten - Eine praxisorientierte Einführung (2014)*. It is flexibly structured, containing many elements that give the speaker a very narrative role, while the interviewer is more passively observant. The interview is ultimately partitioned into the following three elements which vary in their style and cater to different questions within the scope of the first two workflow components.

Part A: Context information – understanding the world of (semi)-professional soccer

The first segment of the interview is mainly concerned with discerning contextual information about the soccer realm. This encompasses the in-depth comprehension of workflows and processes throughout the training-game cycle. In order to maximize knowledge extraction, the utilized questions are of highly exploratory nature and aim to trigger rather extensive narratives, implying the derivation of procedural knowledge. The answers obtained in this partition should help to explain if and why soccer is in need of a visual representation of dominance and, subsequently, where such an analytical visualization can be usefully integrated into the coaching process (Chapter 1.2 in the introduction).

Part B: Dismantling dominance into factors – idea divergence

The second partition of the interview aims to identify as many valid game-related indicators as possible which signify dominance in soccer. Questions are asserted in a highly exploratory manner, triggering a very conversation-based narrative mainly focusing on the experts' elaborations. In a subsequent step, identified dominance factors are discussed in further depth. This approach should yield a varied assortment of potential game-related dominance determinants across a whole spectrum of different soccer ideologies and concepts.

Part C: Selecting the most suitable dominance factors – idea convergence

The final part of the interview attempts to step through the previously identified potential factors of dominance, ultimately assessing and rating these concepts considering their suitability and representational value. Questions are more structured and defined at this stage, though still allowing an elaborated narrative of the expert in order to seek out further contextual information. Therefore, in this phase, the highest-potential factors are identified and selected.

The following sections detail the outcomes of Part B and Part C of the interview process described above. The responses gathered in Part A are incorporated into the motivation of this work provided in the introduction (Chapter 1).

3.1.3.) Identification of suitable dominance determinants (phase of divergence)

The first step in the procedural detection of probable determinants of soccer dominance is signified by the phase of divergence. As the name suggests, the essential aim is to identify a broad variety of different game-related indicators that potentially express the notion of dominance in a certain way or form. In order to seek out viable candidates, a highly exploratory approach is chosen. Following the explanation provided in the previously explained *Part B* on page 43, the core gateway to relevant information on this matter is founded by conducting expert interviews. Common knowledge, derived mainly from popular media formats covering soccer games, and in a very limited amount from thematically associated research, is incorporated into the exploratory question structure as a conversational input. Based on this qualitative information retrieval procedure the following eleven dominance factors are identified as shown in Figure 9.

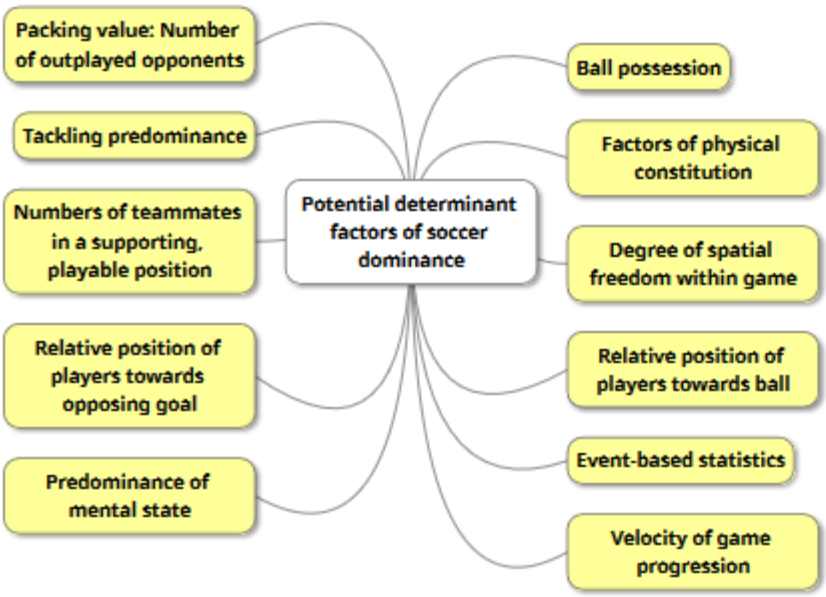


Figure 9: Mind map portraying a selection of potential dominance determinants discovered within the divergence phase.

The displayed variety of detected potential dominance determinants within the game of soccer are further elucidated in the following segments.

1.) Packing Value: Number of outplayed opponents

The so-called *packing value* is a concept coined by the former German soccer players Jens Hegeler and Stefan Reinartz (Sports Analytics start-up *Impect*) which gained a lot of attention during the FIFA World Cup in 2014. The focus lies on the means of advancing the ball efficiently, implying *packing value* as a novel and modern factor indicating dominance in the game of soccer. Many forms of analytical measurement related to the soccer sport observe passes in a singular sense, disregarding quality and effectiveness. On this premise, the packing value tries to enhance such a view by deducing the number of outplayed players by a pass. Figure 10 illustrates this concept schematically wherein the defender (left red player) attempts a direct forward progression of the ball to his offense midfielder (right red player). The resulting packing value is six, which is equivalent to the number of opposing players located between two involved attacking players.

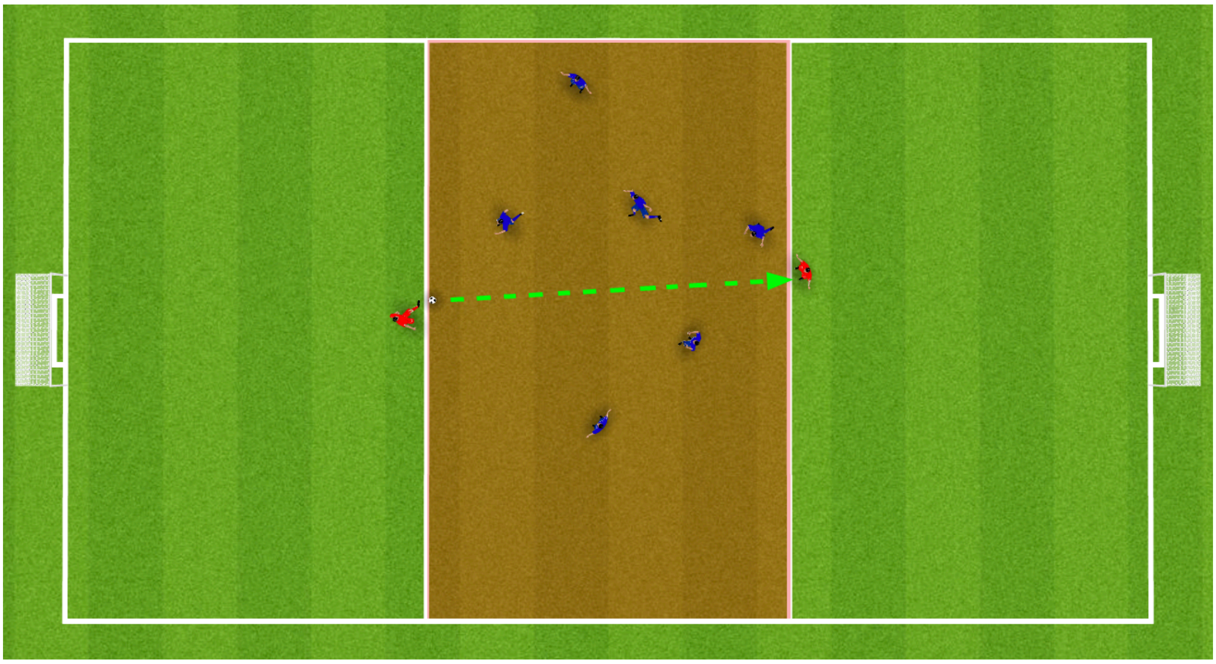


Figure 10: Schematic illustration of a horizontal pass overplaying six opponents, indicating a packing value of six.

2.) Ball possession

Ball possession is probably one of the most prominent measures used to convey the sport through statistical observations and is displayed on media outlets for a common audience. Stefan Goll and Markus Wanner both agree that, while holding the ball within one's own possession, teams are viewed as maintaining an active controlling role within the game. In contrast, the defending squad is perceived to be the more passive counterpart. This, consciously or subconsciously, implies a role allocation scheme that corresponds strongly with the notion of a superiority-inferiority relation. The simplicity of this determinant is shown in its universal applicability being reinforced in soccer discourses involving both experts and laymen.

3.) Tackling predominance

Similar to the ball possession statistic, tackling predominance has obtained quite a prominent status throughout the world of team sports, mostly on the account of media influence. Generally, this concept is expressed through the number of tackling duels a single player or an entire team has been able to decide for themselves. Being victorious in such a situation is often recognized differently depending on the individual observing the tackling scenario. However, in a simplified notion, the superior actor will prevail in keeping control of the ball. The notion of tackling predominance strongly corresponds with the awareness of physical and mental stability and further implies control of the ball which results in a perception of dominating the game.

4.) Factors of physical constitution

Measures encompassing the physical abilities of players have always been prevalent in various attempts to comprehend the soccer sport in more depth. Almost all experts have indicated the essential character of monitoring key physical virtues, such as the players condition, stamina, speed, and strength. Extending this notion, the realm of professional soccer has included an increasing number of medical observations, such as heart rates, physiological build-ups, and muscle analysis. The emphasis on factors of physical endurance has become very visible and, for many coaches, is a vital aspect of the game that can potentially influence the concept of dominance. Most of them agree that data regarding the state of fitness of the team can deliver valuable information to structure practice sessions.

5.) Number of teammates located in a supporting, playable position

Soccer is clearly a team sport. A vast majority of actions throughout the game require the involvement of more than one actor, most prominently teammates. Markus Wanner has stated that soccer can only be played effectively when the individual controlling the ball is sufficiently surrounded by valid targets that can potentially accept an incoming pass as illustrated in Figure 11. The opportunity of passing the ball effectively significantly increases the ability to progress the game in an offensive direction and create pressure and perceptual dominance. Thus, the idea of numerically overloading the opponent in proximal distance to the ball exemplifies a correspondingly dominant state for the attacking team.

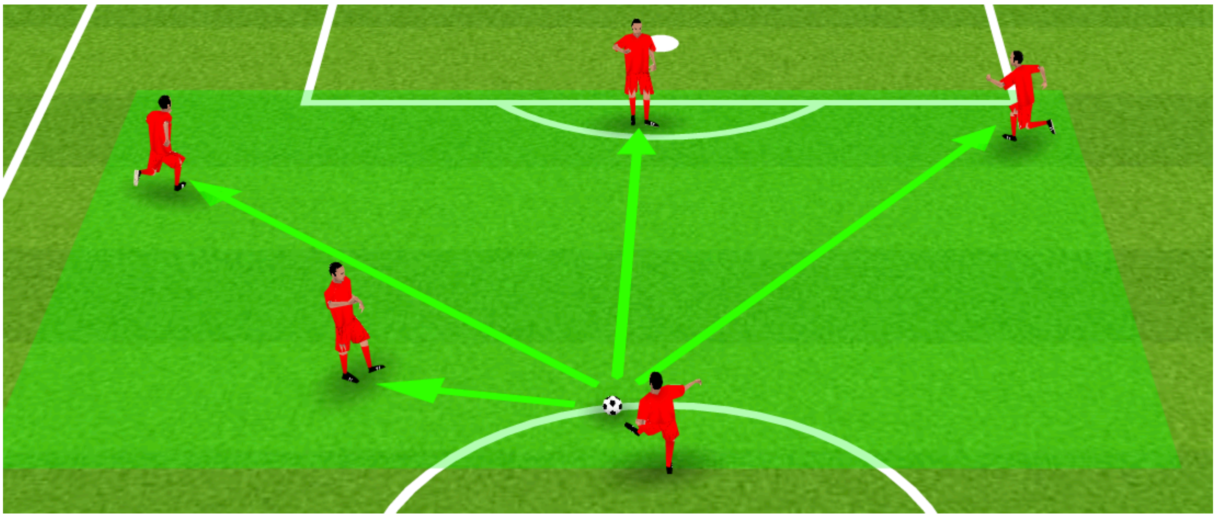


Figure 11: Schematic illustration of teammates located in a playable position in relation to the ball possessor.

6.) Degree of spatial freedom

Thomas Jent states that his understanding of game dominance is strongly correlated to how quickly control of the ball can be regained after handing its possession over to the other team. This concept asserts a desire of demonstrating dominant behavior on the pitch by spatially pressuring the opponent in order to obtain control over the ball. On this note, a key component throughout the dynamic flow of the game is the spatial freedom of the players. Figure 12 schematically illustrates two contrasting scenarios with varying degrees of freedom for the ball-possessing player. Hence, spatial "openness" can potentially be an instructive indicator as to how well one side is able to pressure the opponent. This dominance determinant can be of informative value from both an offensive and defensive perspective. It is essential to keep in mind that free spaces can vary significantly in their meaning depending on the field location and the game context. It is evident that defenders setting up the next attack from deep in their half are usually a lot more spatially unchallenged than an offensive midfielder trying to take action at center field.

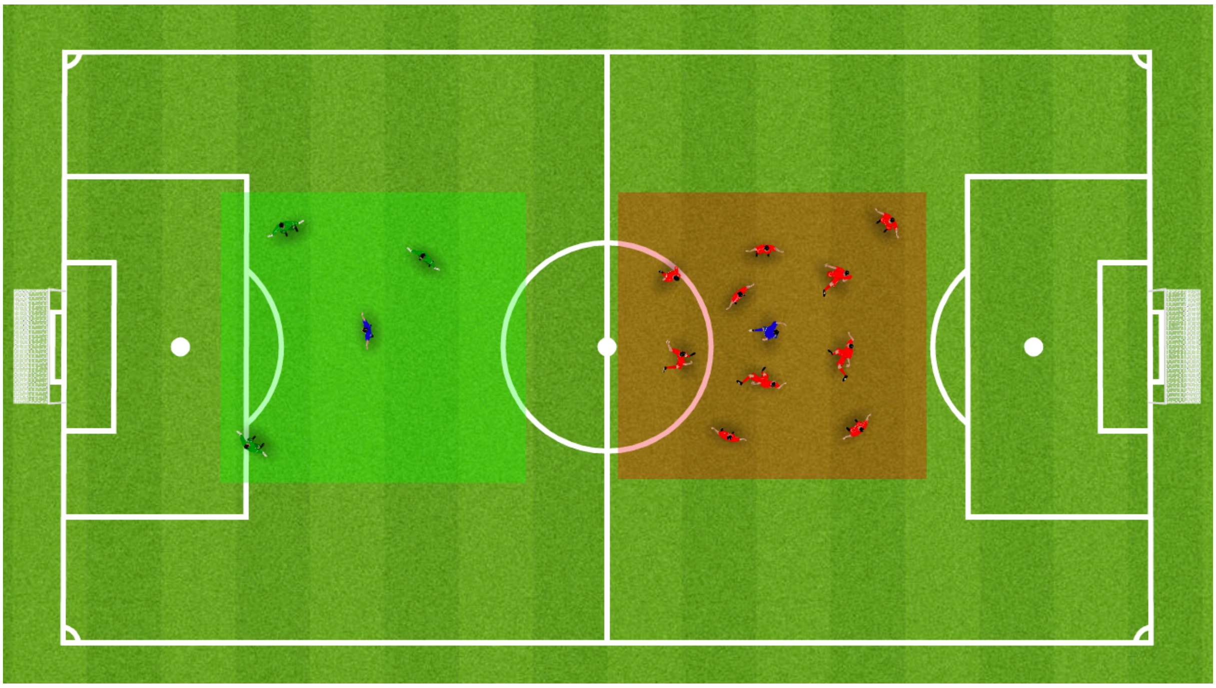


Figure 12: Illustration showing two contrasting scenarios in which the left one depicts a high degree of spatial freedom while the scenario on the right signifies its constrained counterpart.

7.) Position of players in relation to the opposing goal

When discussing offensive dominance, the notion of proximity to the opposing goal is vital. Throughout the expert interviews, Markus Wanner, Thomas Jent and Stefan Goll have all clearly highlighted the significance of this factor in the soccer dominance discourse. Despite the complexity surrounding a dynamic soccer game, the core objective remains to attack the opposing goal, indicating a clear target also in a spatial sense. The more attacking players that are positioned close to the opposing goal, the clearer one can conclude that they are obtaining a superior, dominant position within the game.

8.) Position of players in relation to the ball

If the game of soccer is observed on a very basic level, it is obvious that the ball is the key element lying in the center of every participant's focus. The core aim is to progress the ball into the opposing goal by controlling and shooting it. Many potential dominance factors revolve around ball possession (see factor 2 on page 45) or ideal positioning related to pass-oriented ball movement. Therefore, by aggregating and synthesizing these ideas, patterns of how players are spatially located and distributed around the ball is a highly interesting concept that can potentially indicate certain notions of dominance.

9.) Predominance of mental state

Dominance is often seen as a figure of perception and impression. Many of the factors described here are not only expressed in a spatial or physiological dimension, as they are also often strongly correlated with psychological implications in the minds of involved actors. On this note, Marinko Jurendic highlights the strong influence of the mental state on potential dominant behavior displayed by a team throughout the game. The sheer willpower to take control and dominate the opponent in particular categories such as tackling, speed, and pressure establishment can completely shift the momentum of the game. Despite an initially assumed distribution of strength and skill, the mindset of players can alter the course of a soccer match drastically (e.g. underdog beating the number one team in the league). Jurendic, therefore, emphasizes the importance of exploring mental characteristics as determinants for soccer dominance.

10.) Event-based statistics

Dominance factors such as ball possession, tackling dominance, or positioning patterns are of dynamic nature and, hence, are constantly changing throughout the course of a game. In addition, a soccer game is also strongly defined by specific events occurring along its temporal progression that potentially indicate notions of dominance. The game score is a very prominent example within this category of determinants. Almost all experts have further mentioned event-based statistics, such as the corner kick ratio, shots on target, and number of free kicks, as vital figures which characterize a soccer game in more depth.

11.) Velocity of game progression

Speed has become a vital aspect throughout many sports and soccer is no exception. As already stated, Thomas Jent has defined dominance as the time it takes to gain repossession of the ball after losing control of it. Offering a different view, David Andreoli has stated that certain teams orient their strategy towards “rushing soccer”, whereby the ball is progressed as quickly as possible towards the target goal by overrunning the opponent’s defense and subsequently creating scoring chances. Ultimately, the velocity of game progression can be seen as an intriguing determinant for the dominance phenomenon, as it is often conceptually linked to overpowering and pressuring the opponent.

3.1.4.) Assessment and selection of optimal dominance factors for further utilization (phase of convergence)

After having conducted the exploratory procedure of eliciting an array of potential factors signifying dominant game patterns, the current phase of convergence aims to select the most suitable factors for the subsequent computation and visualization process. Within the concluding part of the interview process, each dominance factor candidate was assessed with the expert regarding its ability to express soccer dominance. In a first step, these results are analyzed and further complemented with computational inputs. Areas with the highest development potential are delineated.

The results of individual factor assessments are aggregated in the chart below (Figure 13), which provides a summary illustration of expert opinion insights gathered through the interviews. When observing the chart, it is evident that some factors are generally not of interest, namely those with negative aggregate scores. Most prominently, the much discussed packing value did not receive a lot of acceptance among the experts. Similarly, the degree of freedom was viewed very critically, mainly due to the fact that it remained inherently complicated to seek out an understandable concept that would carry the meaning of this idea suitably. Factors of physical constitution were surprisingly assessed as having marginal relevance for an analytical visualization. Many coaches stressed that they see factors comprising of strength, stamina, and condition as mandatory prerequisites for the professional soccer sport in general. Hence, the teams already use extensive monitoring of such indicators to draw conclusions about the level of fitness within the squad. Ball possession and tackling predominance triggered mixed reactions among the domain experts, underlining that their relevance is highly dependent on the style of play intended. Further, these factors are very basic numerical observations and, therefore, have minimal potential to be further developed. Event-based statistics can be categorized equivalently given their low capacity for enhancement.

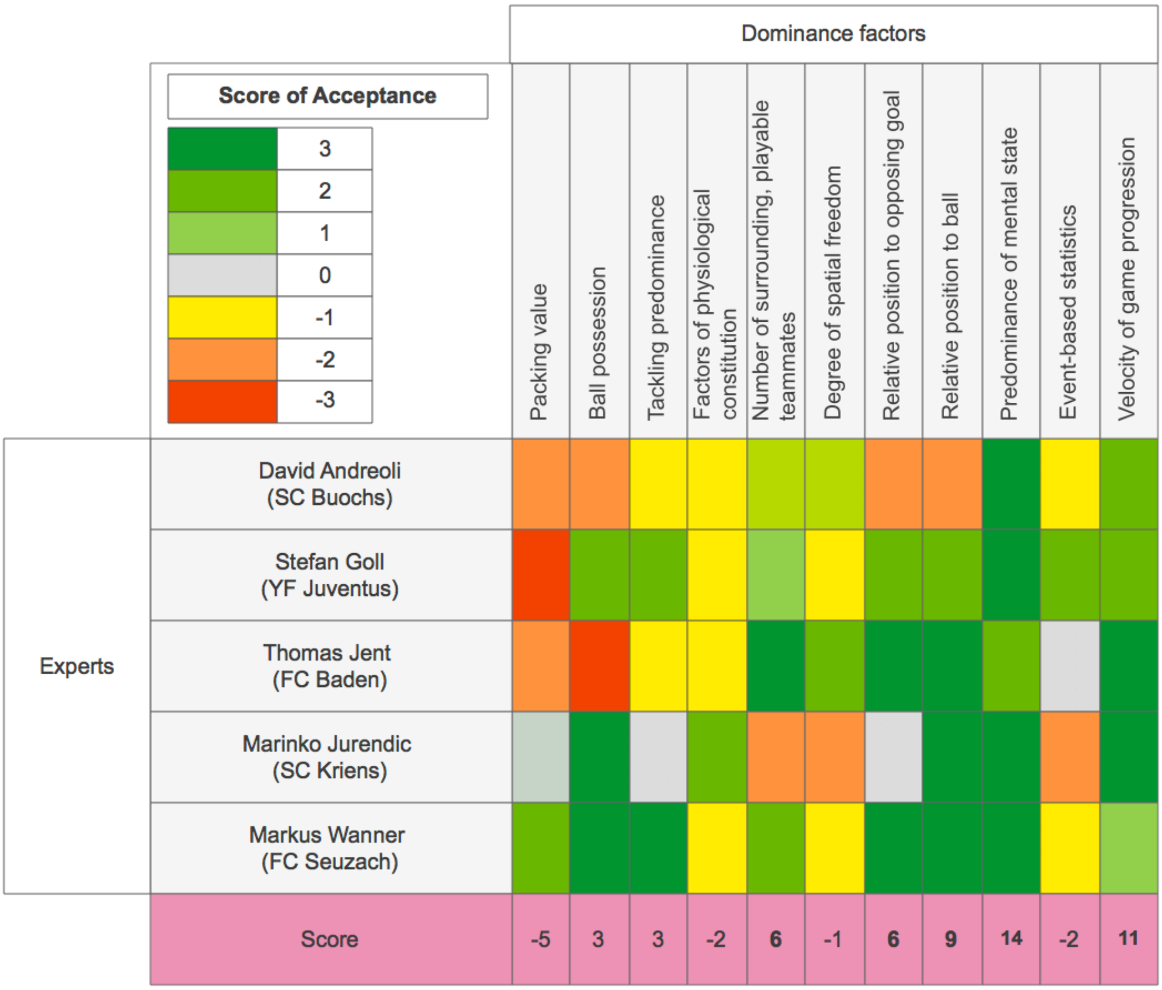


Figure 13: Matrix displaying the assessment of individual dominance factors. Each expert assigns a value to the respective determinant ranging from -3 (completely disagree) to 3 (completely agree), based on their opinion on how suitably it signifies soccer dominance. The aggregated scores are shown in the last row.

Based on their superior acceptance scores, as shown in Figure 13 above, the top five possible dominance determinants have been taken into further consideration. Experts overwhelmingly agreed that mental state is a game-deciding factor with immense influence on dominant performances of players and, ultimately, teams. It is apparent that an informative display of psychological traits throughout the game is a very intriguing idea. However, this determinant unfortunately cannot be considered as it falls under inherent limitations of knowledge and resources. The velocity of game progression has also been identified as a suitable dominance factor. Given the data's inability to of deal with speed on an extended basis without exponentially increasing the computational workload, this indicator significantly exceeds the scope of this work. A velocity-based dominance concept could, therefore, be explored in other undertakings. Ultimately, the three highest scoring factors are developed further as valid candidates for processing. The idea of dynamically computing and visually representing the positioning of players in relation to the opposing goal and the ball, has been rated positively among experts as a legitimate dominance indicator. Similarly, the observation of teammates located in a supporting, playable position has prompted a positive response among experts and the developer/designer. The following segments aim to form concrete, computable factors on the basis of these three selected concepts.

A.) Area of potential I: Number of teammates located in a supporting, playable position and position of players in relation to the ball

Throughout the expert interviews it became evident that the ball is viewed as the center of activity within a soccer game. Any action executed on the field, be it passing, shooting, running, tackling, etc. is, in one way or another, concerned with the location of the ball. The essential aim of scoring on the opponent's goal implies that the ball needs to be progressed forward efficiently. This can mainly be done by either dribbling the ball, randomly shooting the ball forward or passing the ball to a teammate. Coaches and common observers usually allocate their emphasis to the latter case. Hereby, the central question arises as to how (and how many?) players should position themselves in relation to the ball/ball-possessor in order to provide an optimal supporting role by remaining playable for as long as possible. Hence, the ball possessor potentially has more opportunities to progress the ball efficiently. As stated in the previous subchapter, M. Wanner mentions the concept of numerically overloading the opponent in proximity to the ball in order to be dominant within this highly significant area. This further provides the opportunity to attack the opposing goal by moving the ball towards areas of heightened scoring potential, which is a critical aspect of demonstrating dominant behavior. Continuously observing the ratio of one's own players to opposing players over the course of the game can potentially provide an informative indicator for game dominance on the basis of the relative positioning patterns of players around the ball. Further, this shall provide a first and simple quantitative figure as to how great the chances are for the ball possessor to find a suitable passing opportunity. Subsequently, this dominance phenomenon shall be expressed by the formalized factor, *Numerical Predominance*, explained below.

I.) Numeric Predominance

The dominance determinant *Numeric Predominance* aims to identify the ratio of one's own players to opposing players within a specific radius around the ball possessor. This is accomplished by first defining a spatial range within which players are assessed based on their team affiliation. The computation is conducted for the team in control of the ball, thereby taking on the attacking role. If observed players inside the predefined area belong to the same team as the ball possessor, they are marked as affiliated players whereas other players are assigned to the opposing team's count. Using this same technique, a ratio is also deduced signifying the number of offensive players located in close proximity to the ball in relation to opposing presence. Ultimately, this value provides insight as to how well the attacking side is able to numerically overload the opponent, as it implies the extended ability to progress the ball closer to the opposing goal. On that foundation, the *Numeric Predominance* factor encodes a varying perception of dominance over time by indicating the opportunity of reaching areas from which there are higher chances of scoring a goal. The defending team, not in possession, of the ball is not assigned a dominance value (dominance value = zero).

B.) Area of potential II: Position of players in relation to the opposing goal

Extending the observation of game patterns to a more holistic view, the overriding objective of the game remains to score more goals than the other team. The positioning of players in relation to their target is, therefore, vital to analyzing a soccer game. Simply formulated, the distance between players and their opposing goal provides an indication as to what degree their team is offensively dominant. The more a team can successfully access areas of heightened scoring opportunities, the stronger they are perceived as demonstrating dominant behavior by pressuring the target. On the other hand, a team is seen as inferior in a dominance-perception context when they retreat towards defending their own goal. This work suggests the two determinants described below to express dominance through the team's proximity to its target.

I.) Dominance expressed through Euclidean distance-based proximity

The simplest form of demonstrating spatial proximity in a geometric feature-space is to calculate the Euclidean distance between the location of player objects and specific delineated areas with high scoring potential (penalty box, goalkeeper space, high-danger zone extending goalkeeper space). This work aims to derive values of dominance by aggregating players into groups (e.g. whole team, certain positions) by expressing them through their center point. Subsequently, the distance between the center point and the designated zone around the opposing goal is identified as a dominance indicator. The distance is computed within a scaled unit-space ranging from zero to one and, hence, can be transferred directly to an equivalent dominance value spectrum.

II.) Dominance expressed through zone-based proximity

An alternative form of assessing the spatial proximity of players to the opposing goal is to count how many of them are located within a certain partition of the attacking half. Hereby, the zone-based proximity approach distinguishes the attack half, third, and quarter as potential areas of interest to the offensive team. These zones have been defined together with the domain experts who explained their cognitive partition of the soccer field. Ultimately, the number of attacking players positioned within the chosen area is identified at each time step to derive a representation of their proximity to the opposing goal. The resulting value is expressed as a fraction of one and can therefore be assigned directly to a zero-to-one dominance value scale.

3.2.) Dominance Factor Computation

The previous segment has thoroughly described the fundamental requirements analysis wherein the most suitable dominance determinants have been selected for further processing. At this point, the work shifts towards computing these factors such that they can be graphically visualized. The core aim of this part is to express these rather loosely defined indicators in a rationalized numerical logic such that a corresponding value scale can be computed. The succeeding subchapters offer an expanded overview of the computational workflow. In a first step, the basics of the programming environment are elucidated, that represent the foundation upon which the extended calculations are based. In a next step, a set of initial pre-calculations are carried out which provide a supporting function for upcoming procedures. Subsequently, the general structure of the workflow methodology is displayed and standard procedures, utilized for several factor-calculations, are executed. Ultimately, the three selected dominance determinants **numeric predominance, distance-based proximity** and **zone-based proximity** are computed, resulting in a variety of potential measures of dominance to evaluate and choose from.

3.2.1.) Foundations of the computational environment

The computational workflow is performed in an object-oriented Java environment wherein corresponding classes and methods are defined. It is comprised of several modular components which constitute a visual analytics tool for the soccer sport. This work aims to extend the existing framework by incorporating new elements within the scope of the dominance visualization concept.

To appropriately express the identified dominance indicators in a logical numeric format and, hence, conduct arithmetic operations throughout the program-methods, it is vital to define some key concepts which provide a foundation for the data and code. Since the operations carried out are based on spatial interrelations of areal extents and objects within the soccer field, the program consists of a specific spatial referencing system that calculate concrete positional coordinates. As a result, the soccer pitch is displayed as a two-dimensional space with an x- and a y-dimension forming a coordinate space. The initial zero point (x/y equal zero) is located at the top left corner. It is essential to note that calculations are strictly conducted in a single unit space, meaning that the x- (length) and y-scales (width) of the soccer field are expressed with numbers spanning from zero to one. Pre-existing scaling methods are called to transform the absolute figures into the unit space value range. Since the resulting dominance values are defined in an equivalent spectrum, the eventual derivation process is significantly simplified. The objects of focus are predominantly comprised of the players located on the pitch and, in some instances, may include the ball. Within the program realm they are defined as so-called position-time objects. Game progression is portrayed by a sequential series of observed time steps. At each of these temporal instances, the designated player or ball object has a specific coordinate position inside the previously defined spatial reference system.

3.2.2.) Required supporting pre-calculations

In order to appropriately conduct the computation of the selected dominance factors, a set of pre-calculations are necessary as supporting elements for the core methodological procedures throughout the program. There are 4 required pre-calculations which are elaborated upon in the following sections:

A.) Pre-calculation I: Creation of center points

Measures of Euclidean distance, as utilized in the distance-based proximity concept, are usually derived for each player individually. However, based on expert opinions, it is additionally intriguing to apply a such a proximity calculation to an aggregated group such as all players of a certain position or, ultimately, the entire team. In order to effectively express groups in a spatial sense, their central points need to be computed to provide an indicator of their mean location on the field. These point features are subsequently created in the *CenterPoints.java*-class which extends the *PositionTimeObject.java*-class in an equivalent manner as the *Player.java*-class. In this sense a center-point object is attributed, similarly asserting a temporal and a spatial reference. Within the associated class there exist two methods of calculating the average x- and y-position of an object array list which is passed in as a parameter. For each temporal instance, the method runs through all objects of the parameter list and assesses whether they are player-objects which are simultaneously located on the field. In the case of this being true, the current numeric position-value is added to the aggregate-variable and the player counter is incremented accordingly. When the loops have been executed completely, the average x- or y-position is deduced by dividing the aggregate by the player count. Both dimension references are called in the constructor of the class containing the same object-list parameter. Therefore, by creating a new class-object with a specific player-list, a position-time object representing the center point at every temporal point of observance is established. As a result, center points are created in the *Loader.java*-class for each team in aggregate, as well as for the respective playing position groups (defenders, midfielders, attackers).

B.) Pre-calculation II: Creation of areal extents

Proximity value calculations are consistently based on specific areal extents on the field, for example, a penalty box or a goalkeeper space. The zone-based approach, for instance, aims to count the number of players located within a certain attacking area. Another example is the Euclidean distance measure conducted between player-objects and high danger zones near the opposing goal. These designated areas are defined as rectangles in the

InitializationDefaults.java-class, whereby the top left corner remains the initial point with an X and Y-coordinate. By recognizing the height and width of the shape, it is further possible to derive the other corners, conveying the spatial properties for further processing as can be seen in Figure 14.

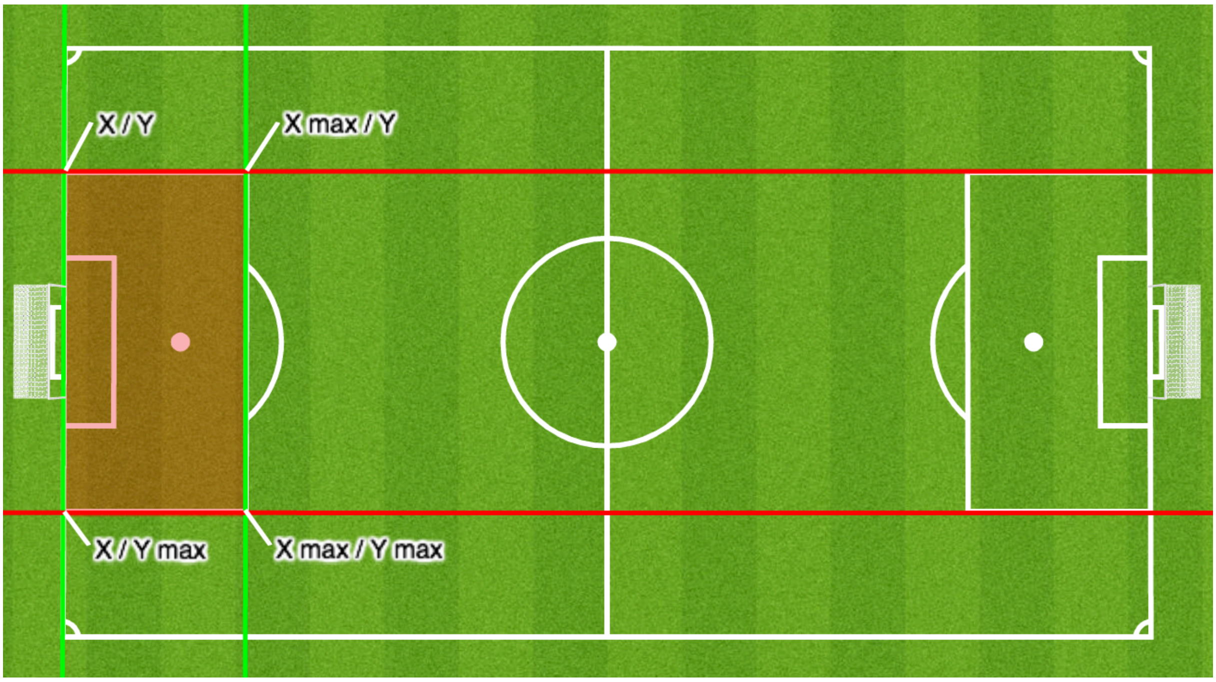


Figure 14: Illustration of the spatial properties of a rectangular areal extent on the field. The top left corner signifies the originating point. By adding the length and width of the shape, the maximum X and Y coordinates and, ultimately, the other corners can be derived.

The result is that three different attack zones (attack half, third, quarter) as well as three specific high danger zones (penalty box, goalkeeper space, extension of goalkeeper space) have been determined for potential further utilization.

C.) Pre-calculation III: Assignment of attack direction to players/teams

Game orientation is a vital characteristic which has a strong influence on factor calculation. In order to compute the value for each team, an attack direction needs to be identified for the objects such that instance proximity values can be derived for the correct side of the soccer field. This circumstance is further complicated by the teams switching sides at halftime. The code snippet below (Figure 15) shows the calculation of the starting attack direction in the *Player.java*-class which can be attributed to each object of focus. Depending on the temporal instance of observation this value can be inverted to provide the correct attacking orientation.

```
if (this.timeSteps[0] == 0 && this.posXScaled[0] < 0.5) {
    this.startingAttackDirection = ATTACK_DIRECTION.LEFT_TO_RIGHT;
}
else {
    this.startingAttackDirection = ATTACK_DIRECTION.RIGHT_TO_LEFT;
}
```

Figure 15: Code snippet attributing the initial attacking direction to a player. If the scaled x-coordinate value of the respective object lies below 0.5 it is located on the left side of the field implying a starting attack direction to the right. Otherwise the opposite scenario occurs.

D.) Pre-calculation IV: Identification of ball possessor

The information as to which team is currently attacking is essential in order to integrate data to a single scale representation where, at each time step, only the value of the ball-controlling team is represented. In addition, the numeric predominance computation process is in inevitable need for this distinction. Therefore, a method is written in the *Player.java*-class which, at each time stamp, loops through all players, identifying the ball possessor. Subsequently, a result-array is returned containing the player holding the ball at each time point of observation.

3.2.3.) General computational workflow

Having completed the pre-calculations described in the prior section, the core computational workflow can be pursued. Despite some distinctive differences between the calculation procedures of the individual dominance factors, the general workflow follows a common core structure, which can be seen in Figure 16 below. It is vital to understand that each determinant is expressed by a set of features that need to be created in the program. These features are usually derived for each team individually (two-scale representation) as well as on an integrated single value scale (single-scale representation). Since proximity-based indicators are strongly linked to the orientation of play, their computation is in need of antecedent pre-calculations which are conducted for each side respectively and are signified by so-called proxy-features which return only an interim result. The actual general workflow commences (Step 1, Figure 16) by instantiating all features and proxy-features in the *FeatureFactory.java*-class where they point to a new object-instance of the designated class, whereby the feature-calculation is executed. They are defined in the *FeatureType.java*-class.

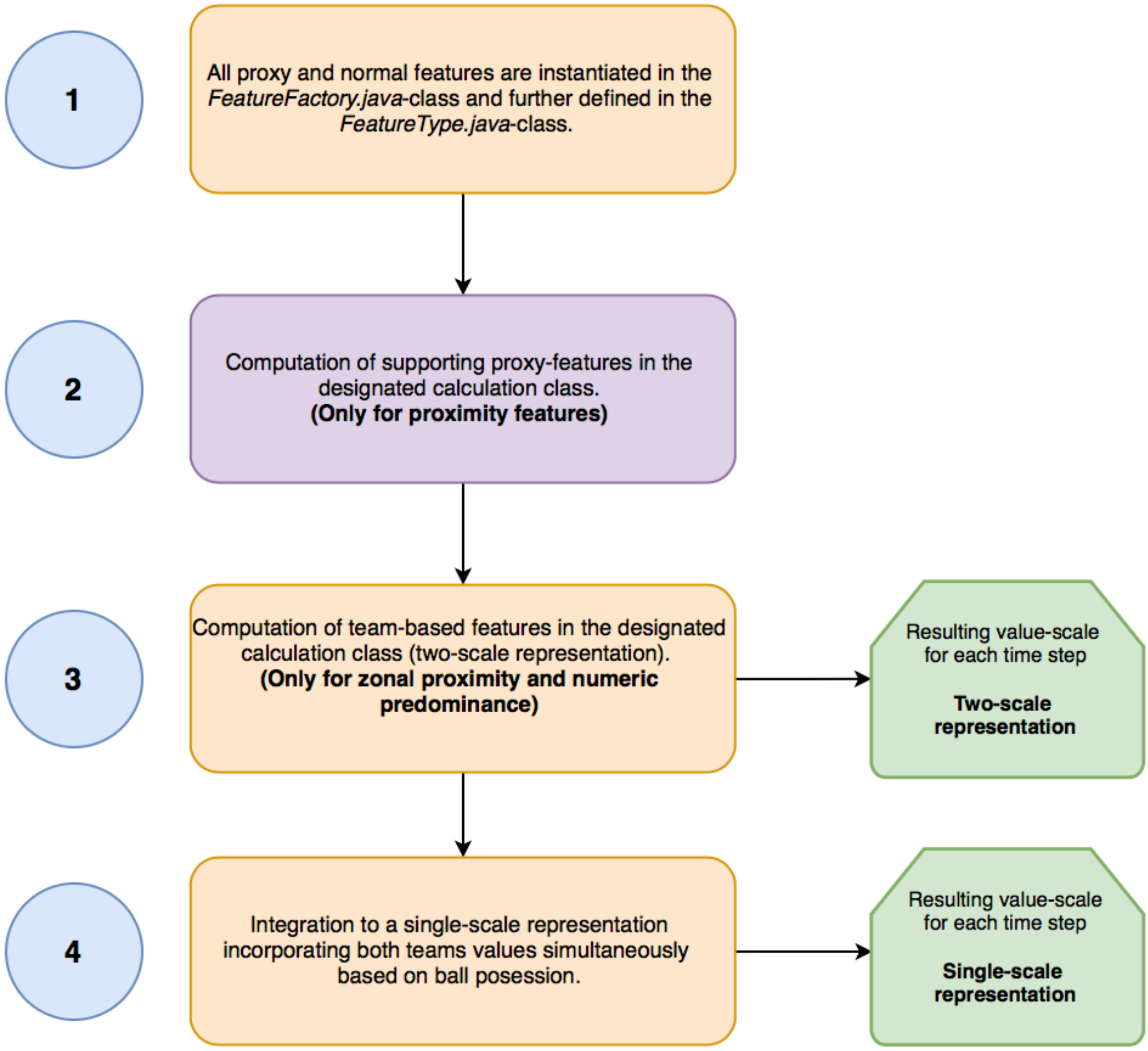


Figure 16: Schematic overview of the general implementation process which is mainly structured into 4 segments (beige = core procedural elements; purple = supporting procedural element; green = final output).

The second step revolves around deriving the supporting proxy-features for proximity-based determinants as previously described. Results returned by the associated calculations are not to be represented in a final feature as they are solely used for further computational steps. Moving forward, the third phase encompasses the computation of the team-based representation. This implies that a single dominance value range is asserted for the home and the away team respectively. A determinant is shown for each team individually resulting in two distinct dominance scales signifying the same dominance indicator. Subsequently, the results are referred to as *two-scale representations.* The concluding step (Step 4) integrates the previously derived individual team-based figures into a single value scale. This is mainly done by identifying the ball possessor for each time step. Hence, the respective team's feature value is assigned for the regarded temporal instance.

As stated before, the actual value computation is conducted within the instantiated feature-class. All of these classes implement a predefined, so-called feature-calculator, scheme implying that they contain respective data-calculation and result-return methods. By establishing a new instance of such a feature-class, as done in the *FeatureFactory.java*-class, the corresponding data-calculation procedure is prompted. Ultimately, a result-array is returned, containing one dominance indication value ranging from zero to one (unit space) for each defined feature.

The actual data calculation executed in steps two and three vary significantly between the different dominance factors and are, therefore, elaborated upon individually in subchapters *3.2.4.), 3.2.5.)* and *3.2.6.)*. Steps one (feature instantiation) and four (single-scale integration), however, are standard procedures conducted similarly for all features which are described in the two following segments.

A.) Feature instantiation statements (Step 1 of workflow)

As demonstrated in the previous workflow overview, the starting point for any factor computation is given by the creation of a new respective feature-statement. All dominance indicators are comprised of a set of proxy and normal features, with the latter actually asserting a resulting value-scale. In the following step, each feature and its associated data-calculation environment need to be established and instantiated. The feature's name is, therefore, defined in the *FeatureType.java*-class. On the basis of this naming rationale the corresponding feature-calculation is primed in the *FeatureFactory.java*-class. As can be seen in the code snippet shown below (Figure 17), this is done by creating a novel instance of the java-class within which the feature-calculation method is conducted. Each of these computation-classes implement the so-called feature-calculator scheme inferring the inclusion of a data-calculation method returning a value-array containing results for every time step of observation.

```java
case PLAYER_COUNT_FOR_HOME_TEAM_ATTACK_QUARTER:
    featureCalculator = new ZonalProximityForTeamsQuarter
    (ApplicationManager.getInstance().getHomeTeamName(), 0.0);
    break;
```

Figure 17: Code snippet showing a feature instantiation statement written within the FeatureFactory.java-class prompting an instance of the designated class, wherein the feature's value-calculation is conducted (i.e. class ZonalProximityForTeams- Quarter.java).

B.) Integrative procedure asserting a single-scale representation (Step 4 of workflow)

The integrative procedure for incorporating team-dependent values into a single-scale representation follows a standard procedure which can be seen in Figure 18

9 below. For each dominance factor, this final step is conducted in a separate feature-computation class. According to the common procedure described above, it contains a data-calculation method returning a designated result-array. Within this method, two feature references are initially created pointing to the respective team's value-scale (team-based two scale representation), as shown in the following code snippet for the case of the zonal proximity determinant.

```java
Feature zonalProximityHome = featureFactory.buildFeatureForObject
(null, FeatureType.PLAYER_COUNT_FOR_HOME_TEAM_ATTACK_HALF);
Feature zonalProximityAway = featureFactory.buildFeatureForObject
(null, FeatureType.PLAYER_COUNT_FOR_AWAY_TEAM_ATTACK_HALF);
```

Figure 18: Code snippet showing an exemplary feature reference to the zonal proximity value for the respective home or away team.

The method ultimately loops through each time step in order to identify the current player in control of the ball. It then continues to distinguish if the designated ball possessor is affiliated to the home or the away team and associates the respective value with the regarded temporal instance of observation. On the basis of the *zonal proximity* examples shown in Figures 17 and 18, this would imply that, in the case of a home team player holding the ball, the corresponding value of the feature *zonalProximityHome* would be associated. Clearly, in the converse scenario, the counterpart opposite feature-value would be assigned. Ultimately, a single result-array is returned, wherein values linked to the home team are scaled between 0 and 0.49999. Similarly, the away side's figures range from 0.50001 to 1, allowing a visible distinction between the two squads in a single timeline (single-scale) representation. The value 0.5 is generated at time steps where no ball possessor could be identified.

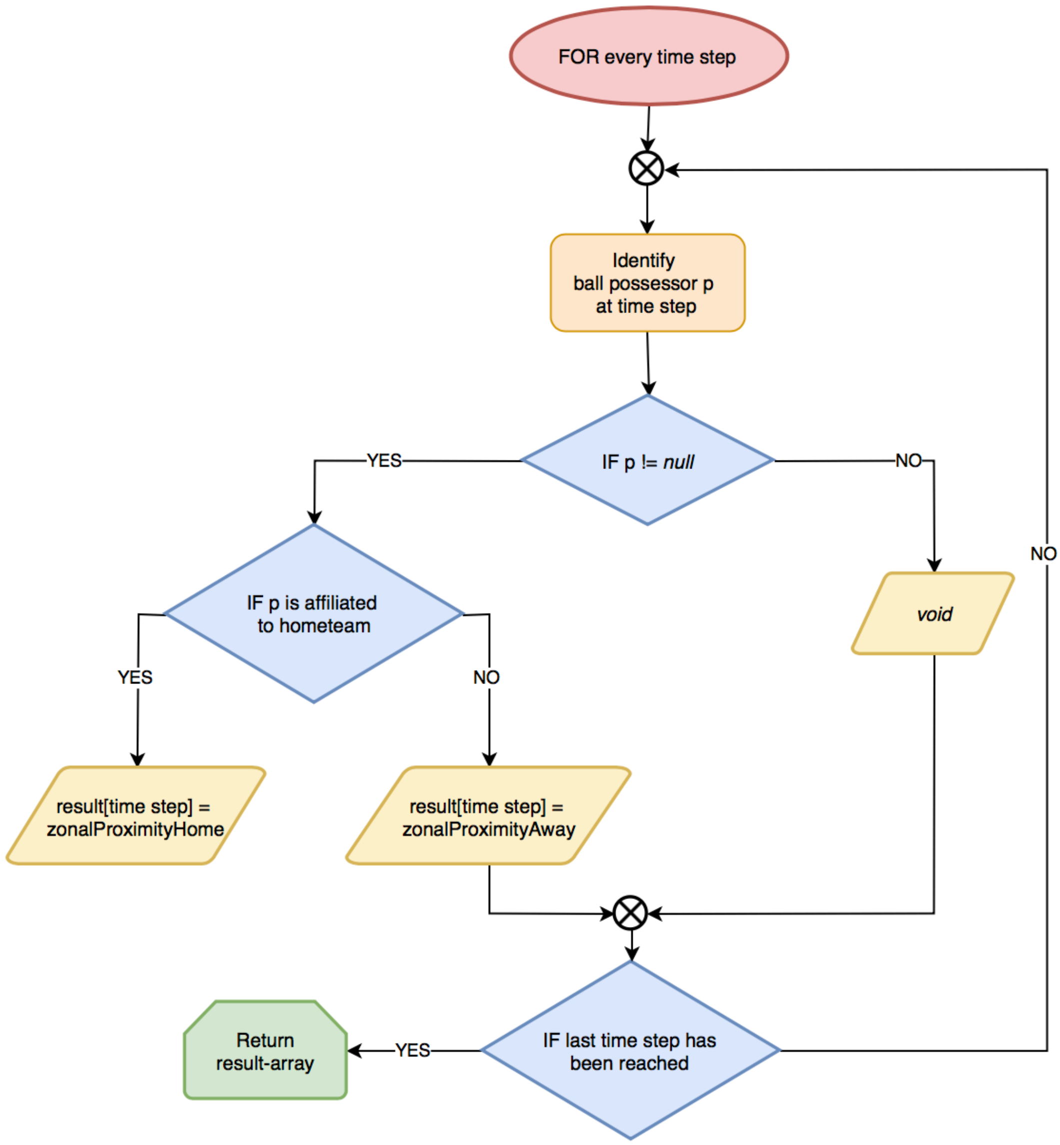


Figure 19: Schematic flowchart portraying the single-scale value integration depending on the current ball possession.

After having demonstrated the common computational procedures composed of the initial feature statements and the final single-scale value integration, it is now possible to elaborate on the particularities of the remaining core workflow elements (Step 2 and 3). The following subchapters aim to provide insight into the calculations executed for each dominance factor individually in order to reach the corresponding output value sequence.

3.2.4.) Computation of Dominance Factor I: Distance-based proximity

As explained in *Chapter 3.1.4.)* the distance-based proximity concept aims to continuously identify the current Euclidean distance between a team or a certain partition of players of a team and a predefined high-danger zone. In subsequent steps, a team or a player group is spatially defined by its center point which is pre-calculated on the basis of *Chapter 4.2.2.) – section A.)* and is further defined as the object. Similarly, high-danger zones are delineated in advance as shown in *Chapter 4.2.2.) – section B.).* According to the general workflow described previously, the initial step comprehends the initial feature statements referencing the various segments of the factor calculation. The computation of distance-based proximity is founded on three supporting proxy-features. The basic arithmetic distance calculation for each side of the field is carried out. Thereafter, these side-based figures are integrated into a single-proximity feature based the current attack direction of the object. Ultimately, actual result-feature values for each team are individually derived from the previously executed pre-calculations. Contrary to the other dominance determinants, these resulting features are not defined as feature statements in the *FeatureFactory.java*-class as is further explained later in section B.). The computational workflow shown here is executed for the case of having the high-danger zone being the penalty box.

A.) Computation of supporting proxy features

I: Side-based distance computation within the DistanceToSpecifiedRectangle-java-class

The prime proxy-features to be computed are the distance measures to each side's penalty box. This step is conducted within the data-calculation method of the *DistanceToSpecifiedRectangle.java*-class which iterates through each time step and assesses whether the spatio-temporal object is located on the field. In the case of this condition being fulfilled the Euclidean distance between the object and the penalty box is deduced by using the *computeDistanceToRectangle*-function in the *GeometricHelper.java*-class.

The geometric distance function is based on the concept of segmenting the positioning of the object (point) and the rectangle (area) in relation to each other depending on specific conditions. This approach is shown below in Figure 20. A point-feature can be split into nine different categories whereby the respective coordinates fulfill different conditions. As explained previously in *Chapter 3.2.1.),* all four corners of the areal extent can be derived by knowing the width and height. The spatial relation conditions are subsequently expressed by their orientation in relation to these four points. For instance, an object is located in the first areal partition if it's scaled x- and y-location references are both smaller than the top left corner of the rectangle. Once categorized, a corresponding distance value is asserted. In cases 1, 3, 7 and 9, the shortest distance is shown between the point and the respective closest corner of the rectangle. Cases 2, 4, 6 and 8, on the other hand, imply a minimal distance measure by calculating a straight line from the object to the nearest rectangle-side. Case 5 results in a distance value of zero since the point is located inside the areal extent of the rectangle.

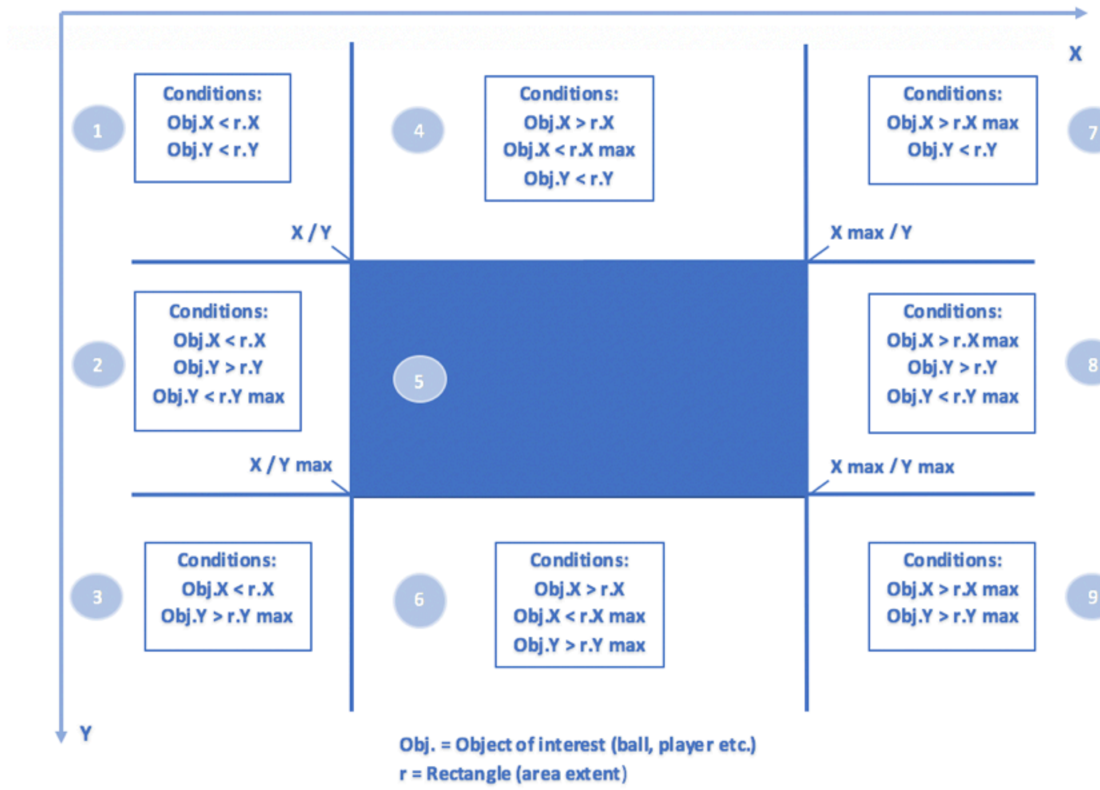


Figure 20: Scheme showing the nine locational relationship cases between the object (center point) and the rectangular area (indicated in dark blue) based on coordinate references.

Arching back to the data-calculation method of the *DistanceToSpecifiedRectangle.java*-class, the resulting distance value is subtracted from one, thereby calculating the actual proximity dominance figure. A high distance value implies low proximity and, hence, low dominance, whereas a low distance value indicates the opposite. In order to generate a suitable end result, a value inversion is necessary, such that low distance figures correctly correspond with indications of high dominance and vice versa. The method loops through all time steps and assigns the respective proximity value to the array.

II: General distance computation within the DistanceToPenaltyBox-java-class

Having computed the distance values to each side's penalty box, these two results are integrated based on the current attack direction of the object. New features are then stated, pointing to the distance calculation for the right and left penalty box as can be seen in the code excerpt below (Figure 21).

```
Feature distPenBoxRight = ff.buildFeatureForObject
(spatioTemporalObject, FeatureType.DISTANCE_TO_PENALTY_BOX_RIGHT);
Feature distPenBoxLeft = ff.buildFeatureForObject
(spatioTemporalObject, FeatureType.DISTANCE_TO_PENALTY_BOX_LEFT);
```

Figure 21: Code snippet showing the creation of new features referencing the respective side-based value-array for the left (distPenBoxLeft) and right (distPenBoxRight) penalty box.

Routinely, for each time step, the spatio-temporal object of concern is assessed as to whether it is located on the field. This being the case, the spatio-temporal object is identified as being either a player or a center point.

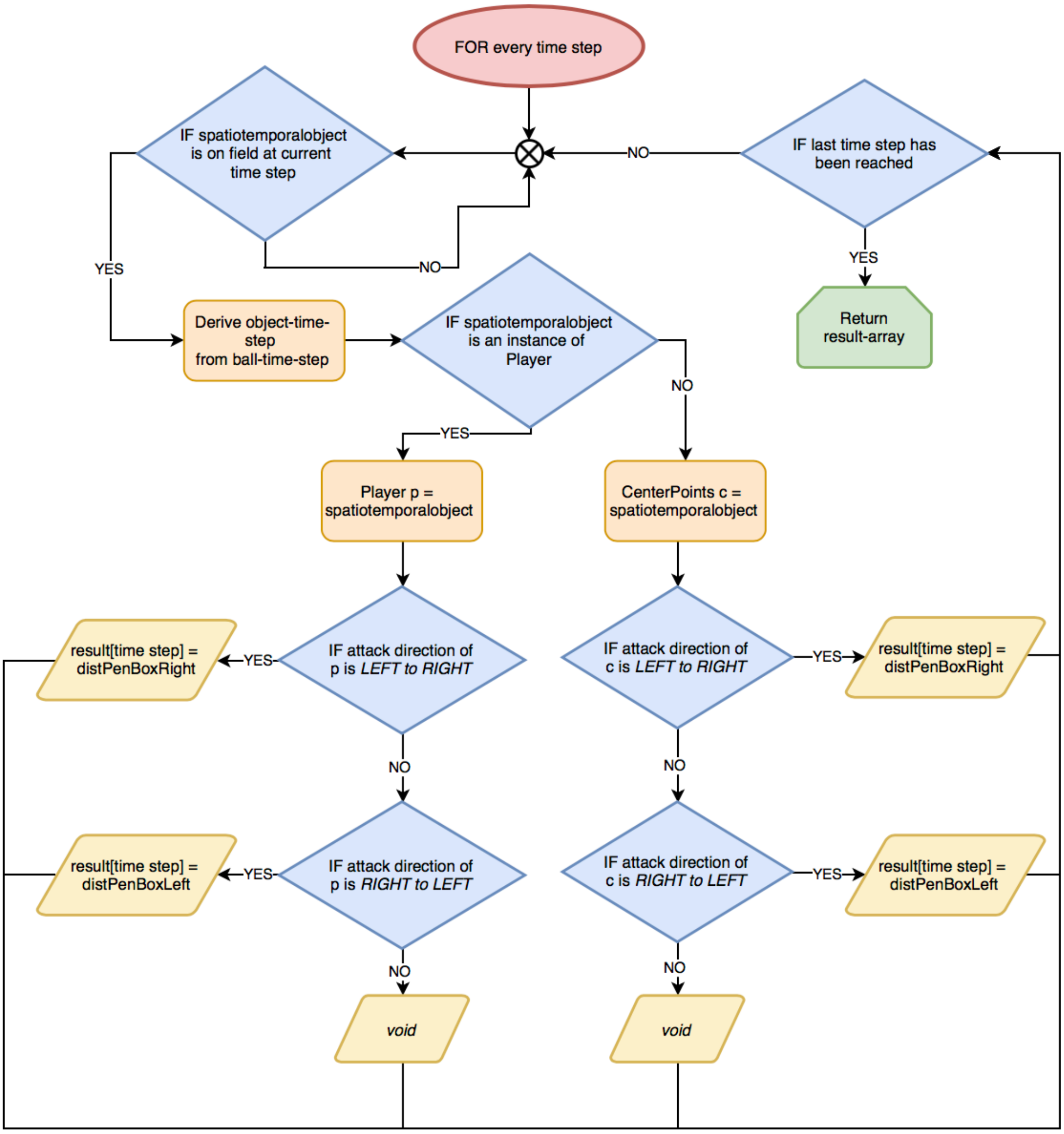


Figure 22: Flowchart illustration showing the assignment of the correct side-based value based on the attacking direction of the object (center point or single player).

For each scenario, a distinction is made between the current attack direction going from left to right, or vice versa. While in the first case the result at the current time step points to a distance measure to the right penalty box, the latter case results in the opposing feature value. Once again, the method loops through all time steps associating the correct number in the general distance-based proximity result array. Figure 22 provides a schematic flowchart overview of this procedure.

B.) Team-based feature-creation within the DominanceVisualization-java-class

The previous sections have demonstrated the creation of proxy features stating the distance to the penalty boxes on the soccer pitch. Contrary to the other dominance factor computations, the actual team-based features are not stated in the *FeatureFactory.java*-class but directly in the *DominanceVisualization.java*-class as switch-case statements shown in the code extract below (Figure 23).

```
case "DISTANCE_TO_PENALTY_BOX_FOR_HOME_TEAM":
    f = ff.buildFeatureForObject(ApplicationManager.getInstance().centerTeamAWithoutGK,
            FeatureType.DISTANCE_TO_PENALTY_BOX);
    break;
case "DISTANCE_TO_PENALTY_BOX_FOR_AWAY_TEAM":
    f = ff.buildFeatureForObject(ApplicationManager.getInstance().centerTeamBWithoutGK,
            FeatureType.DISTANCE_TO_PENALTY_BOX);
    break;
```

Figure 23: Code snippet showing the creation of resulting distance-based proximity features for both teams (expressed by their center points) in relation to the opposing penalty box.

As can be seen, the feature is created on the basis of parameters noting the desired center point of a team and potentially a position-segment as well as the proxy feature calculating the distance to the penalty box. For example, the first case computes a distance-based proximity value for the home team's center point in relation the opposing penalty box.

C.) Factors of variance

There exist two essential factors that can be varied within this dominance indicator computation. The core aim is to deduce Euclidean distance measures between center points symbolizing a certain group of players and a target area. For each team, the chosen center point can be altered between different subgroups comprising of defenders, midfielders, attackers or the whole team (disregarding the goalkeeper). The second factor of variance is the expression of the areal extent. As demonstrated in the workflow, a prime area of desire is the penalty box. However, the exact same proximity computations can also be made for the goalkeeper space or another predefined high-danger zone, in this case the extension of the goalkeeper space up to the penalty box boundary. These variations were defined in cooperation with the domain experts throughout the initial interviews.

D.) Results of distance-based proximity computation

Ultimately, a variety of results are returned that can be further used as founding data for subsequent visualization. For each of the three defined areal extents, a result-array is produced, showing the Euclidean differential to the center points of both teams' defenders, midfielders and attackers as well as all the field players. This implies a total of 24 resulting features that can be visualized. Each result-array consists of a proximity-based dominance value ranging from 0 (lowest) to 1 (highest) for each time step, running in inverted fashion to the actual distance measure. In any default or void-data case the value is set to 0.

3.2.5.) Computation of Dominance Factor II: Zone-based proximity

The zone-based proximity concept aims to convey the notion of dominance by counting the number of attacking players in a pre-defined offense zone close to the opposing goal. In order to derive the correct values, proxy-features for each side of the pitch are calculated. Building on that foundation, result-features for each team individually (two-scale representation) as well as an integrated single-scale display are computed. Following routine procedure, all five features are initially stated in the *FeatureFactory.java*-class. The following three segments demonstrate the associated feature-computation workflow in more detail. The example of a pre-defined offense zone is used in the subsequent process description, equivalent to the attack-third of the pitch (variations are discussed in *section C.)*).

A.) Computation of supporting proxy features

I: Side-based distance computation within the CountPlayersInRectangle.java-class

Equivalent to its distance-based counterpart, zonal proximity is an orientation-dependent indicator. Therefore, it needs to be computed as a side-based proxy-feature for the right and left ends of the pitch, respectively. The data-calculation method contains an IF-clause which identifies for which side the player count shall be calculated based on the Boolean parameter indicated in the feature statement. Hence, while looping through all players at each time step, the method assesses if the attacking orientation is pointing in the correct direction (see *chapter 4.2.2.) – section C.))* and if the player-object is located within the predefined attacking zone (see *chapter 4.2.2.) – section B.))*. In the case of these conditions being fulfilled, a player counter is incremented. This method-procedure is shown in the flowchart illustration below (Figure 24).

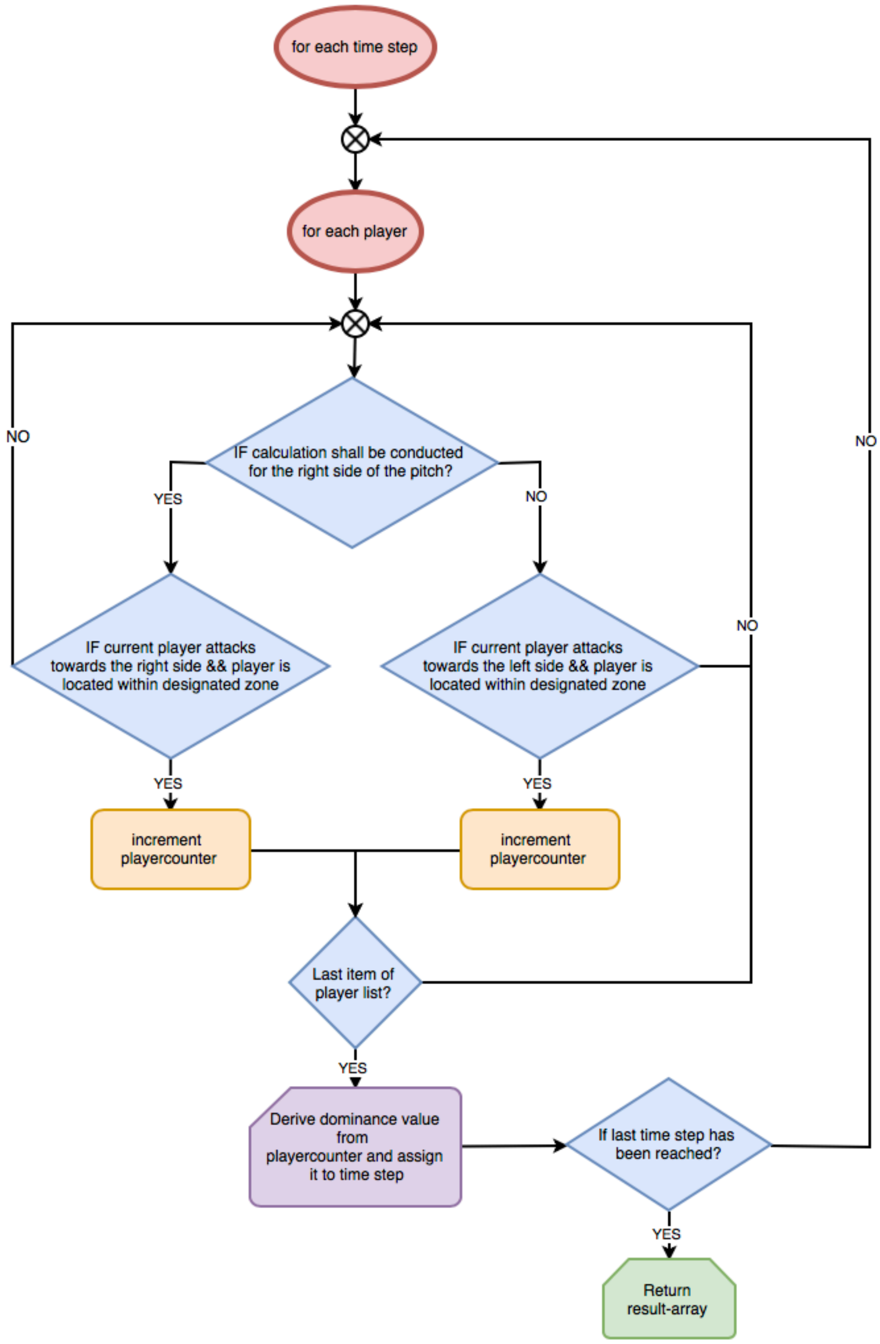


Figure 24: Flowchart illustration of the calculation of zonal-proximity values in relation to the two sides of the pitch by counting the number of attacking players located in the designated offense zone.

After having iterated through all players, the proximity-based dominance value is calculated by dividing one through all eleven players and then multiplying that number by the deduced player count. Ultimately, this figure indicates the partition of attacking players which are located in the defined offense zone in relation to the whole team.

B.) Computation of resulting final features

I: Team-based distance computation within the ZonalProximityForTeamsThird.java-class

Team-based zonal proximity features are routinely derived on the basis of side-dependent values, as calculated in the previous segment, prompting a value-scale for each team individually. Therefore, two new features are created at the methods initiating point which reference the proxy-features for each end of the field, respectively.

In a first step, the starting attack-direction for the designated team is identified. Given a left to right orientation, the variable *newstartingattackdirection* is set to one and, in the opposite case, a value of two. As is demonstrated in Figure 25 below, the starting attack-direction is assessed on the basis of this variable. If the designated team's initial game orientation runs from left to right, the corresponding values computed for the right side of the field are assigned throughout the first half (*countRightThird)*. In the second half, the designated team attacks in leftward direction, implying the association of values calculated for the left side of the pitch (*countLefttThird*). In the converse scenario, this value-assignment scheme is inverted due to the opposite game orientation. Ultimately, each time step is linked to one specific side-based dominance value inferring a result-array for the regarded squad. As the calculation is executed for the home and the away team, two individual features are returned, representing zone-based proximity dominance for each squad individually (two-scale representation).

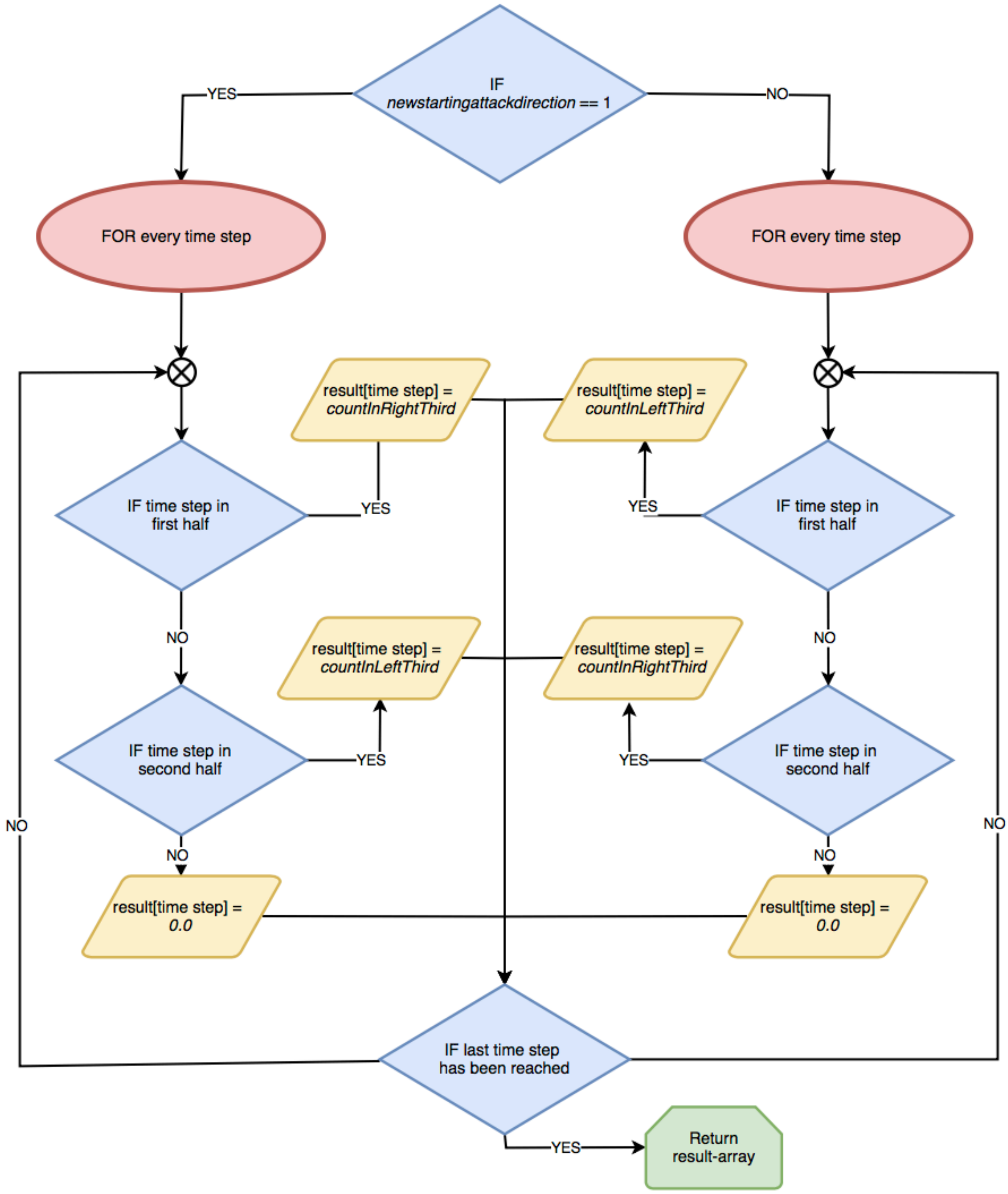


Figure 25: Flowchart illustration of the creation of team-dependent features by concatenating the side-based values depending on the orientation of the game for the regarded team (e.g. attacking from left to right at time step t1 asserts a zone-based proximity value for the right side (e.g. countInRightThird)).

II: General single-scale distance computation within the PlayerCountForRectangle.java-class

The previous segment has demonstrated the computation-process for team-specific dominance values. In a concluding step, these figures are integrated into a single feature representing both teams, simultaneously. Routinely, the incorporation method explained in *Chapter 4.2.2.) – section B* is utilized to execute this task. The result is a value-range returned, displaying at each time step the corresponding dominance figure of the ball-possessing team.

C.) Factors of variance

The sole factor of variance within this distance-based dominance concept is the definition of the designated offense zone within which the player count is to be conducted. Throughout the expert interviews different field partitioning schemes have been mentioned. Subsequently, three essential attacking areas can be distinguished: The offense half, third and quarter of the pitch extending to the opposing goal. All areal extents are represented as rectangles and are initially defined within the *InitializationDefaults.java*-class *(see Chapter 4.2.2.) – section B.).*

D.) Results of zone-based proximity computation

Three resulting features are returned conveying the notion of soccer dominance based on a zone-based proximity computation. For each team an individual set of data-values for each time step exist, ranging from zero (lowest dominance) to one (highest dominance). The single-scale feature at each temporal instance represents the dominance value of the ball-possessing team, subsequently integrating both teams into a single data-array. Hereby, the home team's value range extends from 0.50001 (lowest dominance) to 1.0 (highest dominance) whereas the away team's spectrum runs inversely from 0.49999 (lowest dominance) to 0 (highest dominance). All three features (home team, away team, integrated view) are ultimately computed for each zonal extent defined in the previous segment (*C.) Factors of variance*) implying a total of nine resulting data-value sets.

3.2.6.) Computation of Dominance Factor III: Numeric predominance

As explained previously, the numeric predominance concept evaluates the number of teammates located in a proximal position to the ball-possessing player. The associated calculation is mainly conducted by expressing their count in relation to the total number of players positioned within the predefined spatial range. Building on the idea of numerically overloading the opponent in close vicinity to the ball, the dominance indication clearly rises with higher numbers of surrounding teammates or lower numbers of surrounding opponents. Hence, a scenario with six playable supporting players and three opponents will infer a higher dominance value than a situation with only two affiliated individuals and eight opponents located inside the same designated radius around the ball possessor. It is apparent that the associated geometric operations aim to identify the inclusion of player objects (point-features) within a circumfixing areal extent (polygon). Due to reasons of consistency, calculations are again only conducted for the attacking team in control of the ball. Contrary to the proximity-based factors, the computational workflow for the numeric predominance indicator does not involve any supporting proxy-features. Subsequently the feature-instantiations conducted in the *FeatureFactory.java*-class are only stated for the home and the away team respectively (two-scale representation), as well as for the integrated single-scale view.

A.) Computation of resulting final features

I: Team-based predominance computation within the NumericPredominance.java-class

Numeric predominance features are computed for the two involved teams individually. For each time step, counters for the own and opposing players located within the defined radius are set to zero. The data-calculation method loops through all players and assesses if their distance to the ball possessor is greater or smaller than the range of interest. In the latter case, the corresponding player-count is incremented based on the team affiliation of the regarded player. Thus, if the player is affiliated with the individual in control of the ball, the own-player-counter would be raised by one, whereas in the contrary case the equivalent procedure would be executed with the opposing-player-counter. The flowchart below (Figure 26) demonstrates this first computational procedure in more detail.

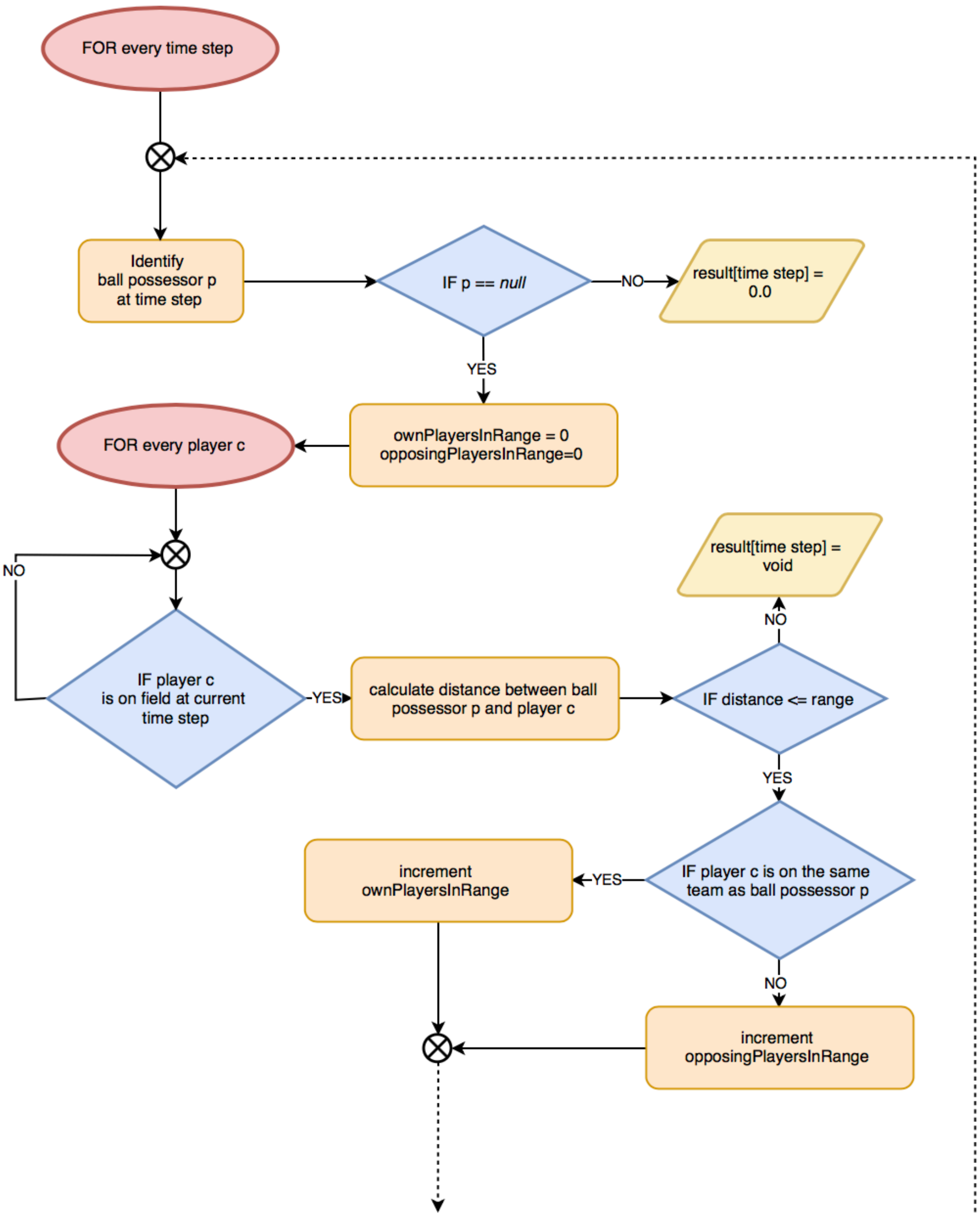


Figure 26: Schematic flowchart displaying the calculation of player-counts based on their position in relation to the defined range around the ball possessor.

Having derived the final player-counts for own and opposing players positioned in close proximity to the ball, a dominance value is deduced on the basis of these counts, as shown in the following part of the flowchart illustration (Figure 27).

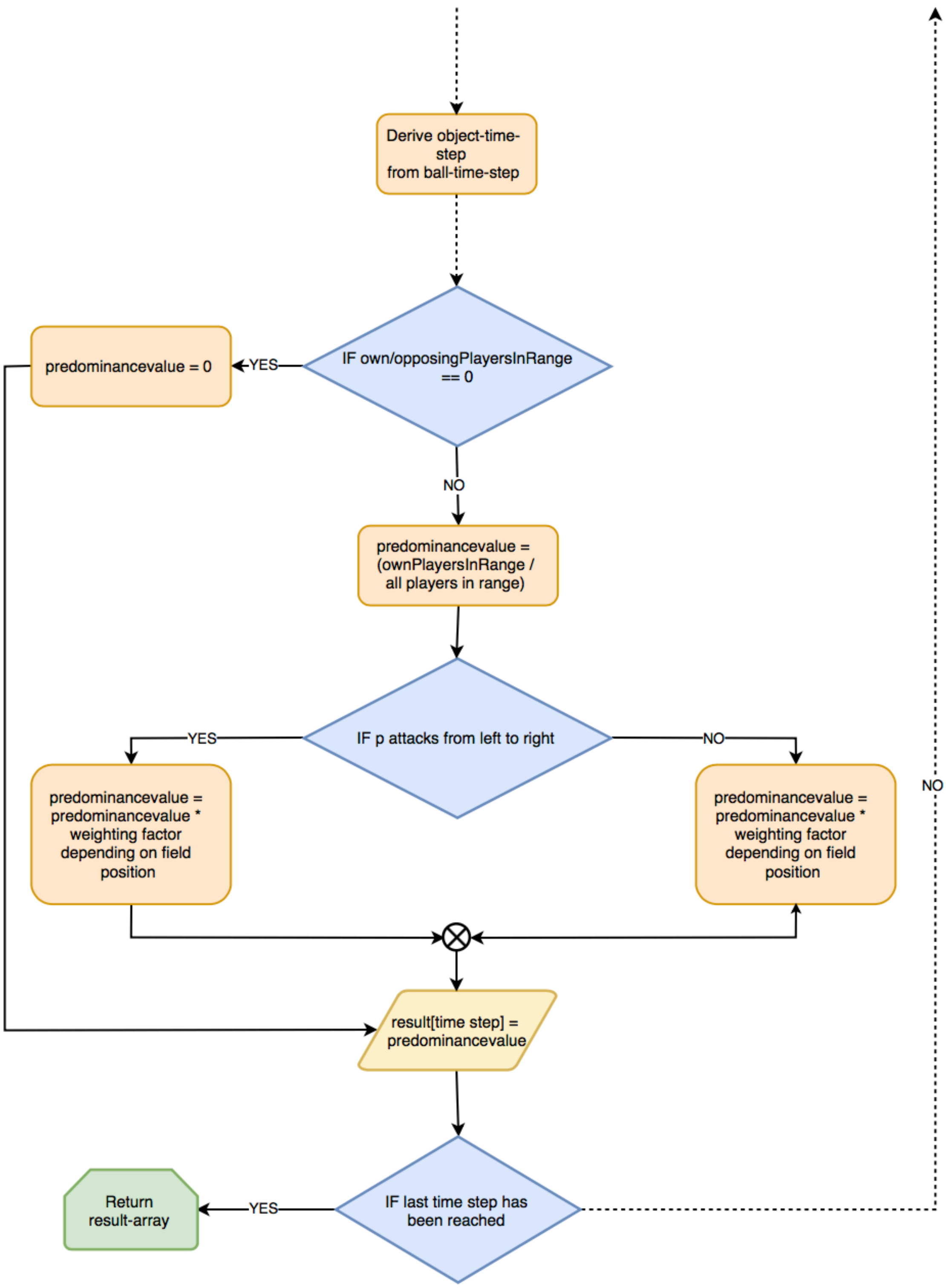


Figure 27: Flowchart illustration demonstrating the calculation and weighting of dominance values on the basis of the previously derived player-counts.

The resulting dominance values are computed by dividing the own player-count by the total number of players. It is vital to note that it is a lot more challenging to numerically overload the opponent the further the team progresses the ball into the opposing half. Therefore, the derived predominance values are weighted according to the field position of the ball. This is done by initially identifying the attack orientation of the ball-possessing team in order to apply the spatial weighting scheme for the correct direction. When partitioning the field into cognitive zones, all experts agree that the most forward quarter of the field can be defined as the highest-danger attacking area. Hence, if the ball possessor is located in this areal extent, a weight of 1.0 is applied, implying that the predominance value is expressed as strongly as possible. The middle zone is rather ambiguous in its delineation. As an intermediate solution, this area is shown by the quarter of the pitch ranging from the half-field line to the start of the high-danger attacking area and is subsequently associated with a weighting value of 0.75. The defending half of the pitch has overwhelmingly been defined as a non-attacking zone where the pressure on the ball possessing team is often significantly reduced. Therefore, the weighting factor is set much lower at 0.25 since numeric predominance is less important in this area from an offense perspective. The resulting predominance value scales for the individual teams, for every time step, contain a figure between zero (lowest predominance) and one (highest predominance).

II: General single-scale predominance computation within the TeamDependantPlayer-Predominance.java-class

The computational workflow aims to incorporate the team-specific predominance value-arrays into a single display. Hereby, for each time step, only the value of the currently attacking team shall be represented. Similar to the zone-based proximity approach, the routine procedure exemplified in *Chapter 4.2.2.) – section B* is utilized to create an integrated single-scale view depicting both teams simultaneously.

B.) Factors of variance

The factor of variance in the numeric predominance computation is signified by the focus range circumfixing the ball possessor, wherein the player ratio is to be identified. This value can be chosen freely. When asked for their opinion, experts have indicated that the ideal range of observation would be set at 12 meters. Therefore, this work has applied the provided measure as the default range extent.

C.) Results of numeric predominance computation

The numeric predominance concept is comprised of three resulting features aiming to convey the notion of soccer dominance. Firstly, for the home and away team, an individual value-sequence exists (two-scale representation), which displays a distinct predominance value for each time step, ranging from zero (lowest predominance) to one (highest predominance). Further, an integrated value scale (single-scale representation) is computed representing the numeric predominance phenomenon simultaneously for both teams. For each time step, only the predominance figure of the current ball possessing team is represented. Subsequently, a value between zero (highest predominance) and 0.49999 (lowest predominance) is asserted for the away team. The home team's value sequence ranges inversely from 0.50001 (lowest predominance) to one (highest predominance). The midpoint of the divergent scale is set at 0.5 and signifies no dominance indication for one team or the other (often because no ball possessor could be identified).

3.3.) Dominance Factor Visualization

Having calculated the identified dominance indicators in the previous phase, this work now shifts towards graphically rendering these factors in timeline representations. For each feature, a normalized value-sequence exists, containing an array of data points, each of which are associated with a specific time stamp. The current objective is to transfer such a numerical value progression to a designated timeline visualization by means of graphical computation. An appropriate variety of time-oriented visual representations are defined which will subsequently be used for the resulting timeline displays. Thereafter, the actual technical implementation, creating the resulting visualizations, is presented. The concluding subchapter illustrates the returned results of this workflow which are then evaluated by domain experts with respect to their utility potential.

3.3.1.) Selection of timeline visualization formats

The underlying data representing the dominance factors is organized as a set of value sequences represented by an array containing data instances for each individual time step. This work aims to display these temporal value progressions in visualization formats commonly referred to as timelines. From a soccer perspective, this implies a temporally-based representation scale which stretches from the first (minute zero) to the last (minute 90+)

observation point of the game. Hereby, the development of the respective dominance indication values shall be portrayed over such a defined timespan. In selecting an appropriate timeline representation format, it is essential to elaborate on the fundamentals underlying this visual approach. Aigner et al. (2011) defines a timeline as a depiction of interval event data (event sequence). This view is to be differentiated from continuous quantitative time-series data. At first glance, the computed value-scale appears to fall in the latter category when judged from a data-nature perspective. However, this work aims to display the initial ordering of discrete data-values as an event-sequence wherein the phases are encoded by their chronological position, duration (length) and intensity manifestation (color). Therefore, it is legitimate to express the utilized data visualization formats as timelines, given their objective of being an informative, event-based visual depiction. Based on their survey of significant work, Brehmer et al. (2017) propose a design space for timelines comprising of three dimensions: *Representation, Scale* and *Layout* as demonstrated in Figure 28 below.

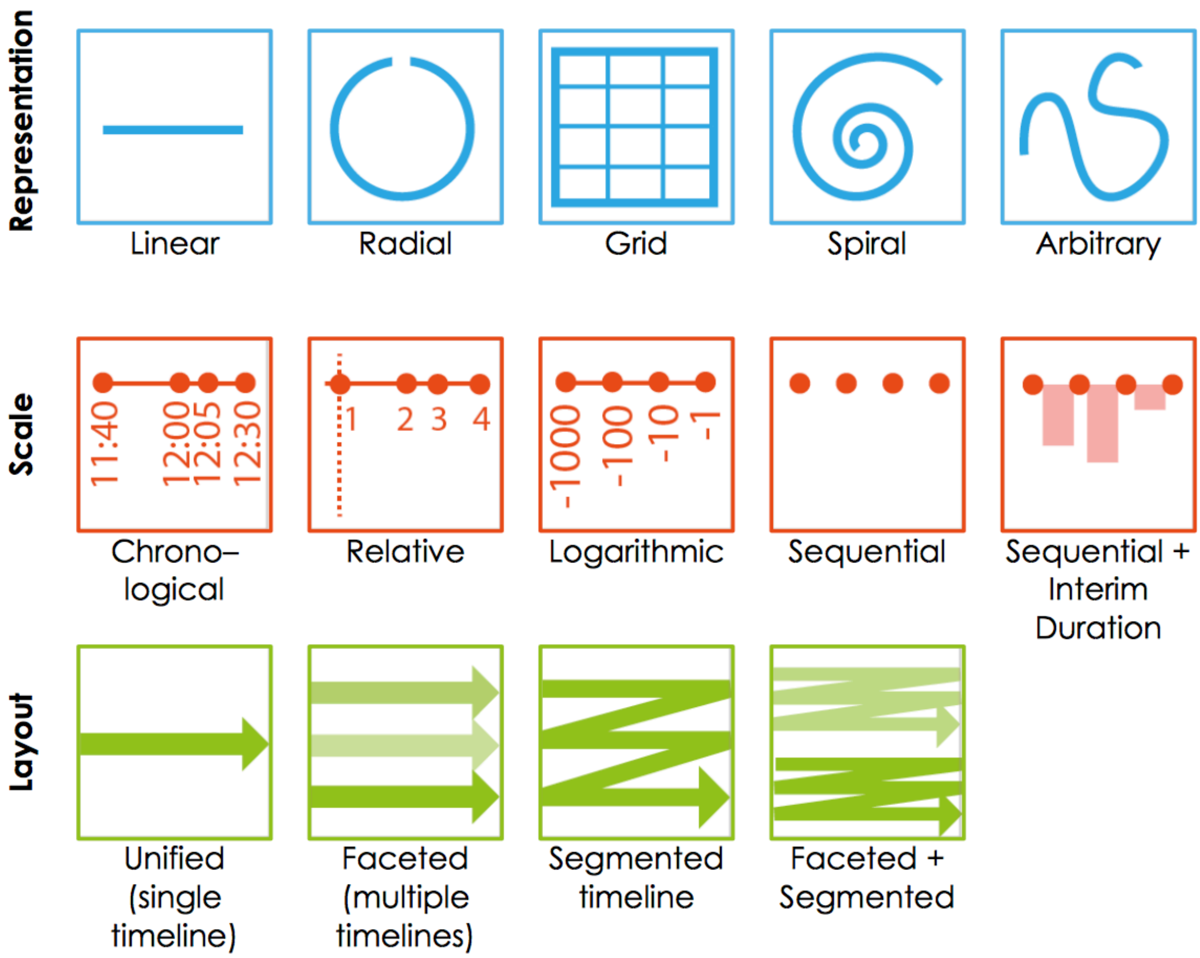


Figure 28: Illustration showing the three dimensions of a timeline design space as proposed by Brehmer et al. (2017).

This work follows these specific design principles to ensure a consistent and suitable timeline representation. The following sections describe how the time-oriented visualization shall be displayed with regard to the three design space dimensions defined by Brehmer et al. (2017).

Dimension I: Representation:

The representation of choice for the time visualizations created within the scope of this work is the linear display. Its presentation is a simple concatenation of sequential events in reading direction which makes it visually pleasing and comprehensible to the potential user. Radial displays are not ideal since they encode periodic data which does not correspond with the time-structure of a soccer game. Grid representations, to mention another alternative, are only suitable for encapsulated temporal granularities such as month-week-day. Finally, spiral and arbitrary displays are not of interest as they might be visually appealing but are portrayed in overly playful and ambiguous fashion, potentially implying more reading complexity and unnecessary distracting visual elements (Brehmer et al., 2017).

Dimension II: Scale:

Regarding scale, the time series classification is quite straightforward. The underlying soccer game data is ultimately an array of values which correspond to a sequence of evenly spaced time steps. Therefore, a sequential scale is ideal as it continuously depicts each temporal observation point. Chronological, relative and logarithmic scales imply a different spacing scheme for data points. As a result, they are not suitable for the representation of soccer dominance values (Brehmer et al., 2017).

Dimension III: Layout:

The final design space dimension of interest is the layout of the timeline. Based on the sequential linear data representation, a unified timeline appears to be the most suitable format to display such a visualization for a specific dominance feature. However, one of the main objectives of coaches when analyzing a soccer game is the comparison of both squads. A faceted timeline can be a useful representation view as it shows multiple temporal progressions of data points simultaneously and, hence, caters to conducting comparative analysis. Segmented timelines are disregarded in this work as they are arduous to correctly partition the course of a soccer game consistently (Brehmer et al., 2017).

As demonstrated in the *Related Work* chapter, there are numerous design approaches towards time-oriented data. Based on the previously discussed design space preconditions, three visualization types have been chosen which are adequate candidates to represent the calculated dominance factor value progressions in timeline representations. The objective is to select of a variety of potential timeline displays such that they can be assessed and compared by the domain experts. Together with an information visualization field expert the following candidates have been selected as suitable for further processing.

A.) Pixel-Chart Visualization

A prime format of visual representation examined in this work is the so-called pixel-chart which builds on the core concept of pixel-based visualizations. According to Keim et al. (2007), the mapping of data points to pixels provides an intuitive conversion of raw data to a graphical display. Pixel-based representations are advantageous in the sense that they are capable of depicting extensive amounts of data and, at the same time, have the ability to provide insight into specific details of interest (Oelke et al., 2011). This notion corresponds highly with the aim of this work to show temporal dominance patterns over the course of a whole game while simultaneously emphasizing the occurrence of particular patterns at a finer observation granularity. The basic idea of pixel-based visualizations is the process of mapping each data value to a pixel within the rendering space (Keim, 2000). In the context of a visual analytic timeline for soccer games, each time step infers a distinct dominance intensity value which is graphically displayed by coloring the pixel accordingly. The pixel-chart timeline is associated with a color-range which encodes the dominance value space. For the user viewing such a visualization format, two prominent variables of optical value distinction exist which shape the visual information retrieval process. Firstly, the color signifies the dominance intensity at each temporal instance of observation. Secondly, based on the number of temporally concatenated dominance intensity values of similar extent, specific phases can be observed in their absolute and relative length within the timeline. For instance, if 20 sequential time steps have a high dominance value above 0.9 this could potentially indicate a significant high-danger attacking phase due to the value-strength and extended duration. Given the existence of two visual indicators (length and color value) within the time line, the pixel-chart visualization format appears to be a legitimate candidate for time-oriented visual analytics in a soccer game. Oelke et al. (2011) further highlight that pixel-based visualizations are solely effective in the case of a non-sparse data set with a non-random distribution. As this condition is fulfilled by the soccer game data underlying this work, this representation type is also feasible from a data perspective.

B.) Two-Tone Pseudo Coloring Visualization

An intriguing alternative approach which also visually conveys extensive data sets in a comprehensible form is the so-called two-tone pseudo coloring approach by Saito et al. (2005). Similar to the previously explained pixel-based view, this representation concept focuses on simultaneously displaying a general overview as well as a more detailed view into the underlying data. Figure 29 below demonstrates the color-based representation scheme utilized within this visualization. Discrete and continuous coloring scales depict each data point by assigning a specific single color value from a defined color range. The two-tone approach, however, allocates a combined ratio of two discrete colors for each observed value in the data set. Concretely, this means that a dominance figure of 0.2 might be expressed by a colored vertical line consisting of 30% red below and 70% black on top, whereas a value of 0.4 might imply the reverse situation.

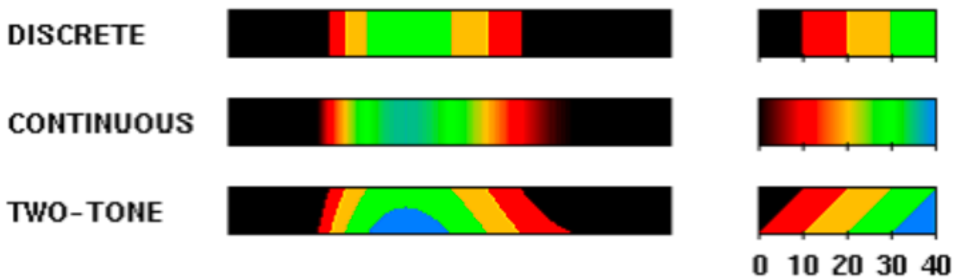


Figure 29: Illustration of a two-tone pseudo coloring scheme in relation to normal discrete or continuous formats (Saito et al., 2005).

Equivalent to pixel-based representations, two-tone pseudo coloring encodes the observed data by means of color value (dominance value intensity) and length (dominance phase duration). Saito et al. (2005) highlights further advantages which include the ability to precisely read out the absolute value at each data point. Also, the rate of increase/decrease can optically be derived from the slope of the color boundary. This implies the existence of visible value peaks signifying high or low dominance phases throughout the timeline. The vast number of different visual indicators included in the two-tone pseudo coloring visualization makes it a valid candidate for timeline-based visual analytics in soccer and, therefore, will be examined within the scope of this work.

C.) Line-Chart Visualization

The third visualization format to be implemented is the line-chart. It can be seen as the extension of a scatter plot and is often used to visualize a trend in data over a certain temporal interval (Khan and Khan, 2011). For each data point the value is expressed as a point positioned at a specific height on a numeric value-scale. If the value at time step 1 is 0.8, it is located higher than if it were to be 0.3 as the value expression is usually referenced to the vertical Y-scale. The line-chart subsequently connects these sequential data points with lines to depict the value-progression over time which is signified by the horizontal X-scale. Hence, such a line-chart can be utilized as a timeline display showing the development of dominance factors over the temporal extent of a game. Contrary to the visualizations discussed previously, the line-height is the only basic visual indicator exemplifying the dominance intensity and the line-slope encodes rates of value increase/decrease. The line-chart visualization is, therefore, an interesting alternative which, from a graphical perspective, is rather simplistic when compared to the other representation formats.

The previous discussion indicates that this work aims to graphically display the computed dominance factors by showing their temporal value progressions as timelines in three different visualization formats (pixel-chart, two-tone pseudo coloring, line-chart). The variance in optical complexity is a deliberate choice as it will be interesting to see which timeline concept is preferred by the domain experts in the scope of visual soccer analytics. Hereby, a significant emphasis lies on assessing the perceptional difference of interacting with more complex color-encoded representation types as opposed to commonly used displays such as a simple line-chart. An assessment of this kind can potentially return key opinions as to which form of visualization is in need of more attention when further developing similar end products in the future. The following chapter provides an overview of the graphical computation procedure wherein the three timeline visualizations are derived based on dominance factor values.

3.3.2.) Technical implementation of timeline visualizations

The technical implementation of timeline visualizations on a computational level involves two interrelated core elements as illustrated in the flowchart below (Figure 30). Firstly, the option panel is concerned with any interactive procedures within the dominance factor representation. Hereby, the user is enabled to make a selection regarding which calculated dominance features she/he would like to have displayed on the basis of which visualization format. The user may also select a data abstraction level on the integrated slider-tool prompting varying levels of visual granularity. Finally, a legend is added showing the absolute value-scale which assists the user in comprehending the visualization. The option panel can ultimately be seen as a graphical element wherein the selection is made, as to **what** exactly shall be represented, in **which** visual format.

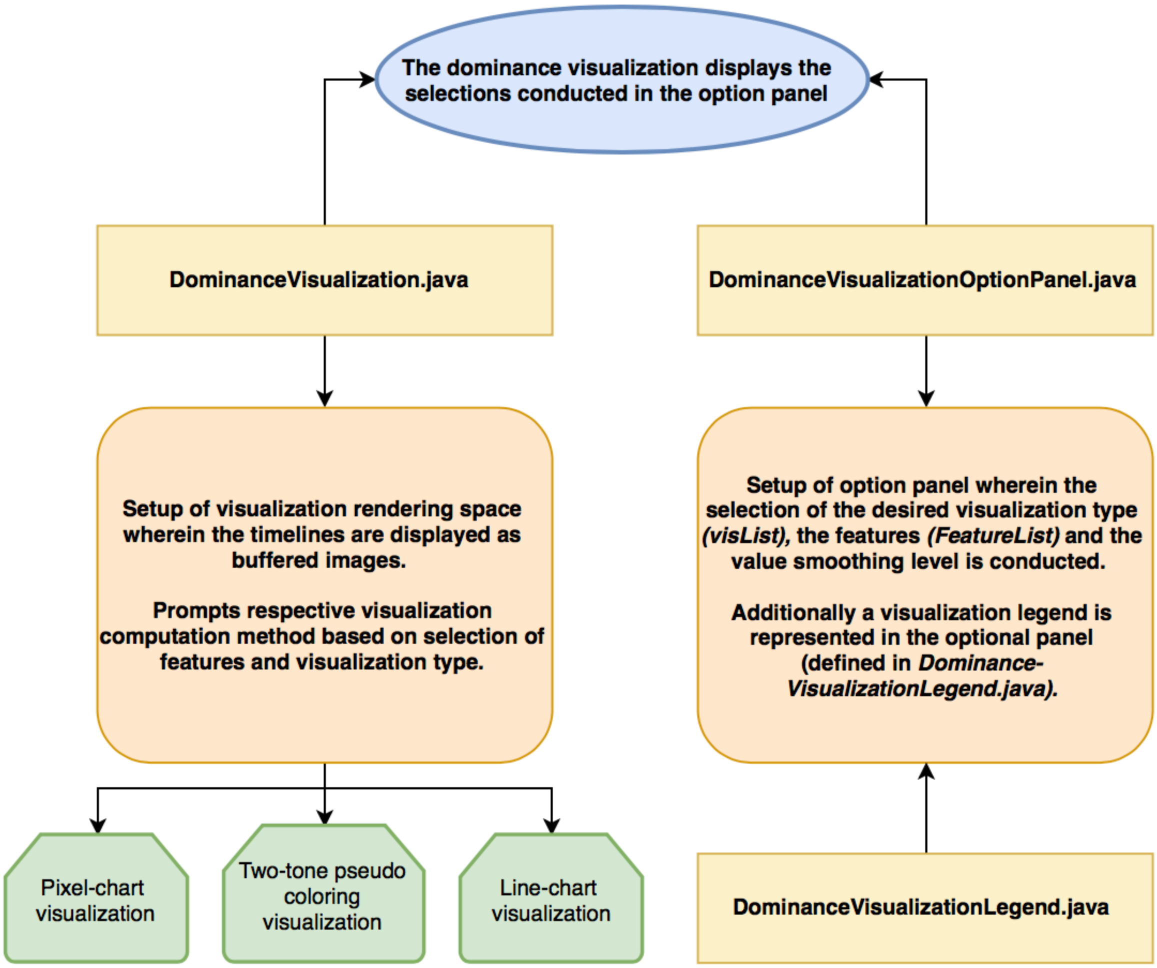


Figure 30: Flowchart showing the configuration of the programming structure for graphical timeline visualizations of soccer dominance indicators.

The second core element consists of the actual visualization types that are used to represent the panel selection. These are computed in the *DominanceVisualization.java*-class. The general program can be understood as an overarching algorithm which assesses the user-chosen parameters in the option panel and subsequently returns the appropriate timeline display in the predefined rendering space. Given the ability to adapt to any selection changes, a dynamic interaction experience for the user is ensured, which enhances the utility value of the tool. The following segments provide insight into the technical implementation of these computational elements.

1.) Implementation of option panel

The option panel is implemented in the *DominanceVisualizationOptionPanel.java*-class. The core element of this class is the constructor which defines the various partitions of the panel. Most prominently, list views are defined containing all dominance factor features (*featureList*) as well as the different types of visualizations (*visList*) which the user can select. Further, a slider tool is incorporated in the panel which allows a gradual visual abstraction of the value-display which is elaborated upon in the upcoming section 4. The class consists of three additional methods which repaint the visualization according to any changes made to either the feature or representation type selection as well as the degree of abstraction. In the case of an altered visualization format, the corresponding legend (computation outsourced to the *DominanceVisualizationLegend.java*-class) is updated and ultimately displayed within the option panel as well.

2.) Implementation of timeline representations

The computation of timeline representations is defined within the *DominanceVisualization.java*-class. Here the focal *getVisualization*-method is designed to create the appropriate visual representation according to the feature selection conducted in the option panel. Simply formulated, a buffered image is established which is geometrically expressed as a rectangle with an associated height and width figure. This area sets the range within which the timelines shall be rendered. In an initial IF-clause the method assesses how many features have been selected in the option-panel. The height of the rendering space is subsequently partitioned into the number of chosen features, inferring the possibility of portraying several dominance indicators simultaneously in vertical sequence. Ultimately, the method iterates through each of the selected feature names and derives the associated smoothed value-sequence calculated in the previous chapter. It then identifies the current visualization type of choice and prompts the corresponding graphical computation method for the designated timeline display as is exemplified in the code snippet below (Figure 31). This implies a specific algorithm for each of the three timeline formats which can potentially be invoked. The segments listed at the end of this section provide a more detailed insight into the respective timeline rendering methods.

```
if (visType.equals(VisualizationType.PIXEL_CHART.toString())) {
    visualizeSelectedFeaturePixelChart(graphics, borderLeft,
    heightPerLine * counter, width - borderLeft, heightPerLine,
    values, counter);
```

Figure 31: Code snippet assessing if the chosen visualization type corresponds with the pixel-chart representation. This being the case, the value-sequence of the regarded feature is rendered on the basis of the corresponding timeline-display.

A counter is incremented after each feature is visualized to re-partition the image scheme based on the number of features shown in the visualization. For instance, if two features are selected, the rectangular rendering space has to be divided into two areas. The method concludes by returning the resulting image containing the timeline(s) for the selected feature(s) based on the chosen visualization type.

The following three subchapters review the three visualization formats implemented in this work. Each of these formats are computed in separate methods which are defined in the *DominanceVisualization.java*-class. They share a common practice for the value to pixel assignment but differ in their graphical color association approach as is shown later. For reasons of comprehension it is vital to keep in mind that the rectangular rendering space initiates in the top left corner (startX/startY), from where its width dimension (X) extends to the right and its height dimension (Y) is oriented towards the bottom. This two-dimensional area can be viewed as a grid consisting of horizontal rows and vertical columns of pixels.

A.) Pixel-Chart Visualization

In an initial step a color-scale is defined which shall be used in the computation procedure. The actual choice of color is further discussed in the upcoming parameter subchapter. Secondly, the pixel-chart method aims to assign the correct dominance value to each vertical pixel column along the width of the rendering space as is illustrated in Figure 32 below. This procedure is routinely used for all visualization formats and, therefore, will only be explained in the current section.

```
for (int i = 0; i < width; i++) {

    double valuesum = 0.0;
    int counter = 0;
    double result1 = 0.0;

    for (int u = (int) Math.round((((1.0 * i) / width) * (values.length - 1)));
            u < (int) Math.round(((1.0 + i) / width) * (values.length - 1)); u++) {

        valuesum = valuesum + values[u];
        counter++;
    }

    result1 = valuesum / counter;
    pixelToValues[index][i] = result1;
```

Figure 32: Code snippet displaying the algorithm used to associate the correct values with the corresponding pixels.

The horizontal pixel alignment and the data value progression run on different scaling levels. Whilst the rendering space width, for instance, may be comprised of 400 pixel-columns, the number of data points that shall be represented is 4000. Subsequently, data values are aggregated before being assigned to such a vertical pixel pillar. The method initially iterates through all pixel-columns (*i*) along the X-axis of the representation rectangle. In an additional FOR-loop, the method then identifies which values (*u*) are proportionally allocated to the regarded pixel-column width (*e.g. pixel column 1 = data values 1-10; pixel column 2 = data values 11-20, etc.*) and calculates their average number. Ultimately, the concluding *pixelToValues*-operation assigns this mean value to the regarded column (*i)* before moving to the incremented width-division to the right (*i+1*). As can be seen in Figure 32, such a value-association is executed on the basis of two figures, one being the location of the pixel column on the horizontal axis (X) and the other referencing the vertical partition (*index*). Based on the previous explanation, several timelines can be represented in a vertical sequence simultaneously. For instance, if the user chooses to display three features, the rendering space is segmented into an equivalent number of rectangular representation bands. The *index*-figure thus indicates for which height-partition the values shall be assigned to the corresponding pixel-column, as the Y-dimension extent of such a band, in this case, only ranges over one third of the rendering space.

The color-allocation for the pixel-chart is conducted within the same iterative procedure of this method. For each pixel-column the assigned dominance value is multiplied by the predefined color scale, implying a specific color-expression being associated to the corresponding pixels. For example, a dominance-value of 0.4 will invoke a color intensity in the lower middle part of the range, whereas a 0.9 value-expression implies a color association from the highest range-partition (e.g. 0.4 = mild green; 0.9 = very dark green). As stated before, the color scale choice is further explained in the upcoming section 3.

B.) Two-Tone Pseudo-Coloring Visualization

In contrast to the pixel-chart representation, the two-tone pseudo coloring approach is founded on the concept of displaying each data point with two discrete colors aligned vertically as signified by the corresponding ratio. Saito et al. (2005) propose the usage of five colors for such a visualization. In an initial step, the predefined continuous color-scale is partitioned into five evenly spaced representative colors which shall be further used in the computation process. The successive assignment of data values to the corresponding pixel-column (*i*) in the rendering space is conducted in an identical manner to that described in the previous section.

Moving forward, the color allocation for each data-point is carried out by iterating through all pixel-columns (*i*) along the width dimension (X) of the rendering space. Each data value shall be graphically represented by a specific ratio of a top and a bottom "color-band". For instance, a dominance value of 0.6 would be encoded by the colors beige on the bottom and light blue at the top as can be seen in the color scale illustration (Figure 33) below.

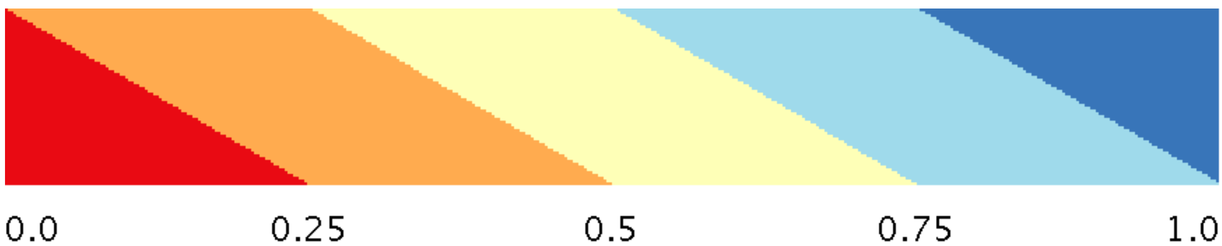


Figure 33: A two-tone pseudo coloring scale displaying the color ratios for corresponding dominance values ranging from zero to one.

In order to deduce an appropriate ratio value, the upper and lower color-band limits need to be identified. Staying with the example of a dominance value of 0.6, these boundaries are set at 0.5 (minimum) and 0.75 (maximum). Values not included within this range imply the usage of other color bands and are, therefore, to be disregarded. The proportion of the upper band color (light blue) subsequently increases gradually from 0% (at lower boundary) to 100% (at upper boundary). Hence, the ratio encodes the size of the partition of the upper band color based on the relative location of the data value to the boundaries (*ratio = data value – lower boundary / total range between boundaries*). A dominance value of 0.6 would imply the top 40% of the pixel column (*i*) being filled with the upper band color (*light blue*), whilst the remaining part remains beige (lower band color). Ultimately, within the iterative progression of the method, this graphical color-association procedure is conducted for each pixel-column (*i*) sequentially.

C.) Line-Chart Visualization

The line-chart visualization represents a line connecting the sequential data points. Hereby, the width dimension (X-scale) of the rendering space signifies the temporal game progression from minute zero to minute 90(+) whereas the height dimension indicates the dominance value scale ranging from zero (bottom boundary) to one (top boundary). The data points are evenly distributed in the horizontal direction from left to right. Their location, in relation to the vertical Y-scale, subsequently encodes the dominance factor intensity, resulting in higher values being located further up than lower ones. It is apparent that no color scale is chosen for this representation as it relies solely on a graph-based node-edge display. Routinely, the association of values to each pixel-column (*i)* is executed according to the procedure explained in the pixel-chart implementation segment above.

The execution of this graphical visualization is rather straight forward. As explained before, the method iteratively steps through the sequential pixel-columns (*i*) comprising the horizontal dimension of the rendering space and associates the respective dominance value accordingly. This procedure is repeated equivalently for the neighboring column to the right of the currently observed one (*i+1*). For each of these partitions the data point is displayed according to the derived dominance value expression on the zero-to-one height-scale. Hence, if a column consists of 50 vertically aligned pixels, a value of 0.6 would imply the data point being drawn around the 30th pixel (0.6 * 50) from the lower boundary. Ultimately, the *drawLine*-method connects these two node-features with a line. The procedure is repeated numerous times, linking the previously calculated line-segment with the next point to the right. The resulting representation is a line-chart, displayed in a two-dimensional space, signifying the temporal development of the regarded dominance indicator over the course of a game.

3.) Choice of visual variables

Any visualization has a set of attributes which shape its graphic depiction. In a cartography context, Bertin (1983) defines visual marks as visible objects which enable the representation by displaying relationships within data sets. Such a mark can be varied and, therefore, obtains a role of a so-called visual variable. Each visual variable is a specific representational property that can be altered accordingly. The timeline visualizations implemented in this work are framed by these variables which require additional explanation. The following segments provide an overview of how the timelines are displayed with regard to the different visual variable dimensions exemplified by Bertin. Carpendale (2003) categorizes marks on the basis of their geometric feature (point, line, area etc.). The timeline visualizations are expressed in a two-dimensional space containing a height and a width indication and, hence, are equivalent to an area. This form of mark can change in **position**, **color, value** and **texture** but not in **shape, orientation** or **size** without completely changing its meaning.

A.) Position

Due to the fact that the timeline is implemented in a pre-existing interface, it is confined to particular spatial arrangements. Naturally, the timeline is positioned at the bottom of the screen corresponding to the existing visualization of the time progression. This is ideal as it extends significantly from left to right and, therefore, doesn't interfere with the selection menus on the left side and the soccer pitch representation in the middle, top and right of the interface. The program setup allows the various comprising elements to be repositioned on the screen.

B.) Color and Value

When regarding the visual variable color, a distinction needs to be made between the line-chart view and the other visualization formats. Line-charts use a simple line representation and, thereby, do not require any attention from a color perspective. On the other hand, the two-tone pseudo coloring and pixel-chart approach are both based on color encoding. Both cases display quantitative data expressed by values lying in a range between 0 and 1. There are three dimensions which determine the perception of color: 1.) The *color hue* determines "the degree to which stimulus can be described as similar or different from stimuli that are described as red, green, blue and yellow" (Fairchild, 2011) 2.) The *color intensity* signifies the amount of pure hue in relation to neutral grey 3.) The *color value* is the perceived lightness and darkness of a color scale (Strode & Cross, 2013). These three components subsequently shape the resulting color schemes applied to represent the data underlying this work. For the pixel-chart timelines displaying a value range from 0 to 1, a sequential color scheme is chosen implying an order suited for data extending from low-to-high on a numerical scale (Harrower and Brewer, 2003). In order to derive an appropriate color solution, the application *ColorBrewer.org* has been chosen, which offers ideal color ranges for visual perception (Harrower and Brewer, 2003). This ensures that the underlying color representation corresponds with essential conditions of human cognition and comprehension. The ultimately selected scale consists of eight different color shades of green, encoding higher data figures with stronger (darker) color values. The number of classes comprising the color range was set to eight in order for the viewer to be able to distinguish finer differences in data values. Many color schemes do not go past nine classes as this would be visually overbearing and incomprehensible to the viewer. As stated previously, for the two-tone pseudo coloring approach, the number of classes in the color scheme is reduced to five as proposed by Saito et al. (2005). On the other hand, the single-scale views are based on value ranges initiating from the middle of the unit space and moving in opposite directions depending on the team (e.g. home team: 0.5 to 1.0; away team: 0.5 to 0.0). As a result, a diverging color scheme is chosen in the *ColorBrewer.org* application with nine classes for the pixel-chart (uneven number is chosen explicitly to have a "middle" class). The two-tone pseudo coloring approach remains at five classes.

C.) Texture

For computational reasons the texture variable has been disregarded in this work.

D.) Shape

Given the linear, sequential and two-dimensional representation of the timelines, the logical choice is to use a rectangular displaying shape. All incorporated timeline visualization approaches (pixel-chart, two-tone pseudo coloring, line-chart) suggest such a geometric property for the rendering space.

E.) Orientation

The orientation of line-charts has been discussed in previous literature. Cleveland (1993) and Heer and Agrawala (2006), for instance, propose a technique named *banking to 45 degrees* whereby the average absolute orientation of the line segments shall be equal to 45 degrees (Heer et al., 2009). However, most linear timelines which are shown in the survey by Aigner et al. (2011) are aligned horizontally from left to right as this ensures comprehensibility by considering the natural reading direction of people in most of the world. Therefore, this work remains conservative by following the general practice of implementing a level left-to-right orientation of the linear timeline display.

F.) Size

The high degree of interaction freedom for the user enables him/her to freely define the size of the timeline representation by a simple mouse drag. Subsequently, the visualization can be extended to the full screen extent. The timeline display is mainly thought to be a rectangular area whereby the length dimension is clearly larger than the height dimension. Therefore, any form of resizing the timeline can evoke visual distortions which may or may not be an issue for the observer. Heer et al. (2009) have concluded that estimation errors significantly drop with increasing chart height and levels out after the vertical extent reaches about 24 pixels. However, it is interesting to note that estimation time is prolonged with larger chart heights. As a result, no specific statement can be made as to what size a timeline should have. In the end, the functionality to freely resize the visualization according to personal preferences allows the user to choose his desired view, implying that size is not a limiting perception factor in the proposed visualization approach.

4.) Integrated data smoothing function

When displaying a series of sequential data points, the amplitude of value changes can sometimes be very rapid and random (O'Haver, 1997). On a visual dimension this circumstance can imply brisk variations of color displays resulting in visually unattractive, cluttered representations. In this case the occurring "noise" shall be reduced by a so-called smoothing process reducing or increasing adjacent high and low points respectively to receive a more leveled data sequence (O'Haver, 1997). Similarly, this work attempts to enable the user to choose an ideal visual abstraction level which can be set with a slider ranging from zero to full smoothing. On one hand, this procedure helps avoid visually unpleasing and unreadable timeline displays. On the other hand, it is of interest to observe to what extent the users wish to abstract and simplify the visualizations by smoothing the underlying data values, in order to reach the highest information level possible for them. Ultimately, an unweighted sliding-average algorithm is used which replaces each data point with the average value of *x* adjacent points (O'Haver, 1997). The *x* stands for the smoothing width or window which, in the case of this work, is set to 1. It is important to note that the identification of an ideal smoothing algorithm is not covered by the scope of this thesis. Rather, the provision of a fundamental data smoothing tool should be seen as a first step towards enabling the user to visually abstract the visualization to his/her preference.

3.3.3.) Output

The aforementioned computation process ultimately results in an interface consisting of an option-panel and a visualized timeline indicating the progression of soccer dominance factors. All results are presented in detail in the next chapter (*Chapter 4: Results*) where they are additionally evaluated by domain experts.

4.) Results

The following chapter provides a thorough overview of the computed results displayed in various forms of timeline visualizations. All of the visualizations depict the corresponding value progression for a specific dominance indicator over the course of a game. In a first step, a case study is shown wherein the various visualization outputs are displayed and interpreted. Thereafter, an evaluation is conducted with a panel of domain experts aiming to critically assess the final results of this work. A summary of insights into the evaluation outcomes, highlighting the key messages of the gathered responses, is also provided.

Disclaimer

The results demonstrated throughout this result section depict the relevant visualization elements computed in this work. It is vital to note that these are currently still at a premature “mock-up” stage and, therefore, do not yet represent a fully styled and usable interface to the user. The core objective of this work is mainly concerned with assessing the utility value of timeline visualizations and does not include an elaborated focus on the complete graphic implementation of an entire, coherently usable interface. First and foremost, the line-chart timelines in this work neither include a value range shown on the vertical axis nor supporting lines spanning across the whole length marking significant value levels (e.g. 0.25, 0.5, 0.75). As it was computationally not possible to display these elements in a suitable format, they were added into the timeline by hand for the evaluation in order to present a usable example of a line-chart. Secondly, the time series represented above the computed timelines depicting the temporal progression as well as meaningful events for both teams has been incorporated from previous computational work conducted in the scope of the interface environment. The timeline is mainly shown to provide an additional temporal scale complementing the timeline visualizations. On this note, it is essential to stress that this visual element has been implemented previously by an external person and evidently has not been computed within the extent of this work (similar to the general interface environment initially provided). The two teams taking part in the demonstrated game are Bayern Munich and Manchester City as can be seen in the time step display above the visualized timelines. For reasons of simplification, Bayern Munich will be referred to as the home team (or team A) and Manchester City as the away team (or team B). The timelines can be displayed for each dominance indicator separately as no consistent response could be retrieved from the evaluators as to which ones they wanted to be able to view and which were to be omitted. This circumstance also encompasses the fact that the timeline representations are shown for each parameter setting of the factors of variance. For example, the distance proximity indicator can be observed for the penalty box, the high-danger zone and the goalkeeper space as no dynamic tool exists yet to vary this setting within the interface. Hence, the user is free in independently choosing which indicators he wants to have displayed according to his preferences. Finally, the legend is missing an informative description as this couldn't be computed suitably and has subsequently also been incorporated by hand for the evaluation to ensure its usability.

4.1.) Case study of visualization output

The resulting interface is comprised of two main elements. On one hand, the option panel allows the user to interact with the program by selecting the level of visual abstraction and the desired dominance features that shall be represented. On the other hand, the actual value-sequence of the designated dominance indicators are visually displayed as timeline visualizations. In the following sections, the respective outputs for both elements are described in more depth.

4.1.1.) Option panel visualization

The option panel, which is displayed in Figure 34 below, enables the user to select specific desired parameters which drive the resulting timeline visualization.

The option panel is comprised of the following four elements:

1.) Dominance Feature Selection List:

Within the feature selection list the user can cull the dominance indicators that he/she wants to have displayed in the timeline visualization. These include all computed features from Chapter 4.2.) for the home and away teams as well as all integrated single-scale features. The selection can be executed by mouse-click, enabling the possibility of choosing several features simultaneously.

2.) Visualization Type Selection List:

The user can choose his/her desired visualization type among a choice of the three computed formats (pixel-chart, two-tone pseudo coloring, line-chart). It is only possible to select one variant at any given time.

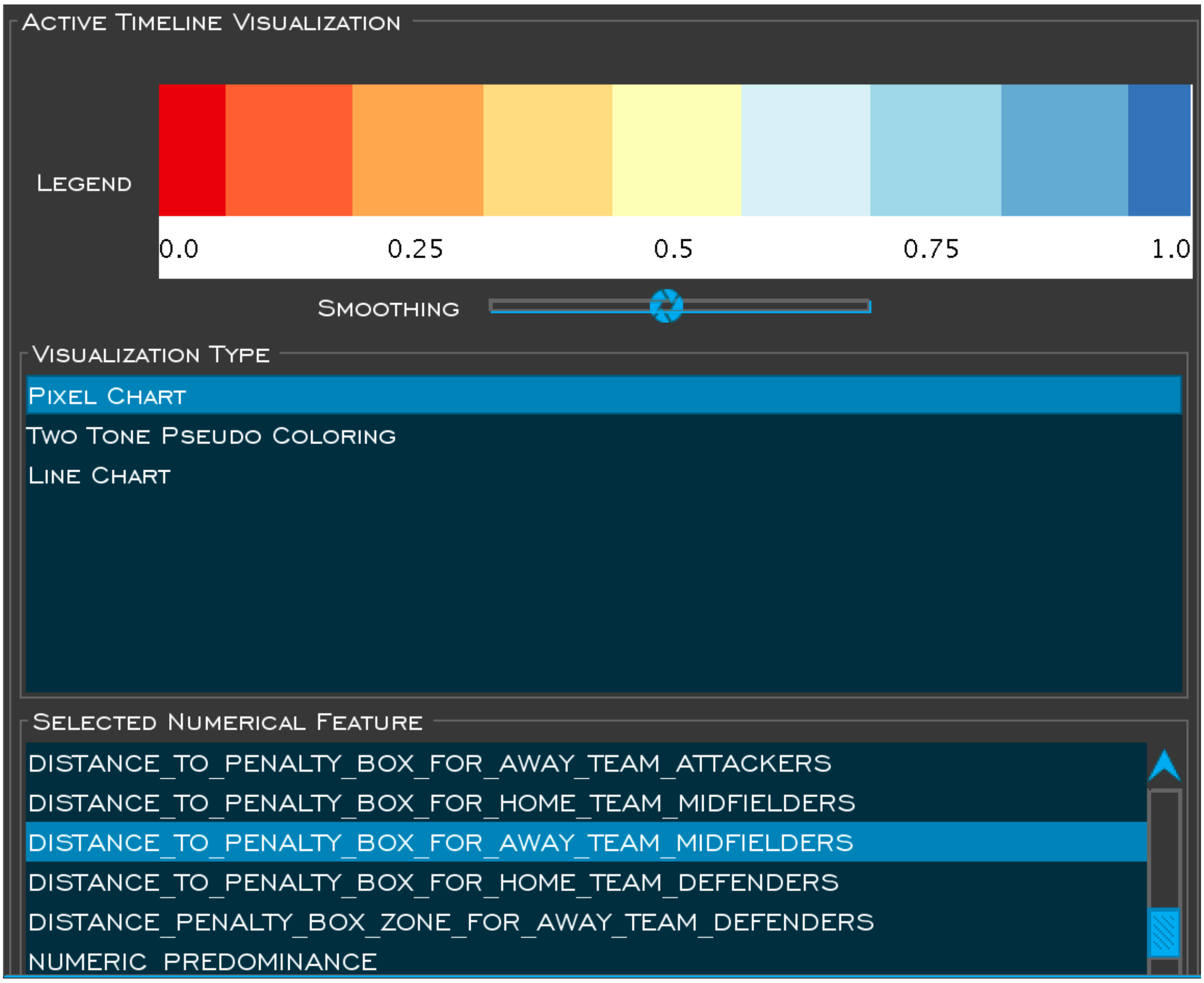


Figure 34: Illustration of the option panel element in the interface allowing the user to interact with the program by selecting the desired dominance features, visualization type and the level of data-smoothing (abstraction).

3.) Value-Smoothing Slider:

The position of the value-smoothing slider encodes a specific level of visual abstraction ranging from 0.0 (no abstraction) to 1.0 (full abstraction). It can be moved by the user at any time to invoke the desired abstraction level.

4.) Legend:

The legend aims to assist the user in understanding the timeline visualization by portraying the relation of the dominance value scale to the specified color range. When changing the representation type, the legend is altered accordingly.

4.1.2.) Timeline visualization outputs

The essential visualization features computed in this work are clearly the timeline representations of dominance indicators. Each timeline visualization is shown in the interface depicting the associated value-series for each selected factor. Three alternative formats have been created in this work for viewing the chosen features in the timeline display: 1.) A specific dominance feature can be viewed for one team individually, resulting in a sole timeline portraying the associated value progression over the course of a game. 2.) A dominance indicator can be viewed for both teams by selecting the feature instances linked to the respective teams and subsequently having the resulting timelines be displayed simultaneously in vertical alignment (two-scale representation). 3.) A dominance indicator can be observed in a single timeline representation integrating both teams' value-scales (single-scale representation). The following section provides an illustrative overview of the three timeline visualization approaches (pixel-chart, two-tone pseudo coloring, line-chart).

1.) Pixel-Chart Visualization

The first timeline visualization format covered in this chapter is the pixel-chart representation.

I: Dominance indicator visualization for a single team

The following two figures show a simple pixel-chart timeline visualization encoding a specific dominance indicator for a single team. The corresponding numeric value-scale ranging from zero (lowest dominance indication) to one (highest dominance indication) is depicted by a color sequence. Dark green color values represent phases of high dominance whereas the counterpart scenario invokes a lighter green intensity expression. The two distinctive visual marks embedded in this timeline display are the **color value** and the **duration (horizontal length)** of a specific dominance phase**.** As can be seen in Figure 35, the pixel-chart view already provides informative data at a non-abstracted representational level. Despite the high data resolution, specific low and high dominance outcrops can be identified and analyzed on a more detailed level based on the color intensities displayed in the timeline.

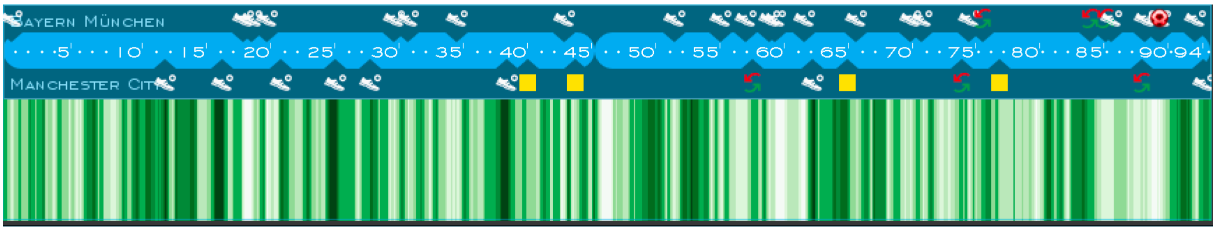


Figure 35: Pixel-chart timeline visualization for a single team (value-smoothing parameter = 0.0).

Figure 36 portrays the same pixel-chart data at a visually more abstracted level. The color differentiation is clearly not as distinct as in Figure 35 due to the smoothing of the data-values.

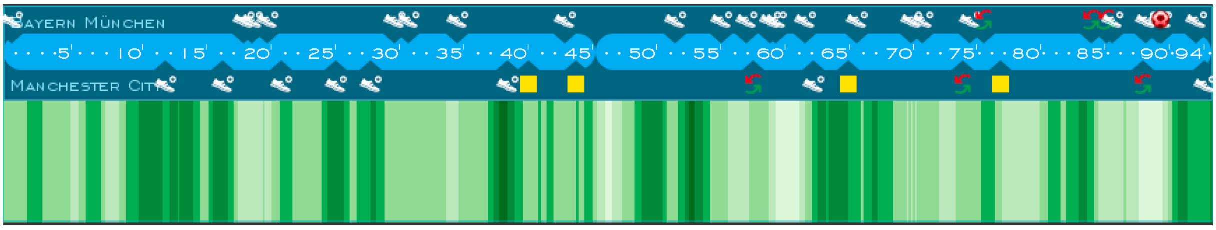


Figure 36: Pixel-chart timeline visualization for a single team (value-smoothing parameter = 0.5).

However, given the color-aggregation, dominance phases are generally easier to perceive on the basis of their length in relation to the time progression of the game. Ultimately, the pixel-chart timeline provides a simple and informative representation of the dominance value development.

II: Comparative dominance indicator visualization for two teams (two-scale representation)

The comparative two-scale representation aligns the value progression of a particular dominance indicator for both teams vertically above and below each other. Discrepancies in dominance values are mainly visible through the contrast in color and can further be observed on the basis of their duration indicated by the horizontal length of a phase (same/similar color values over a period of time). Figure 37 below exemplifies such a two-scale representation.

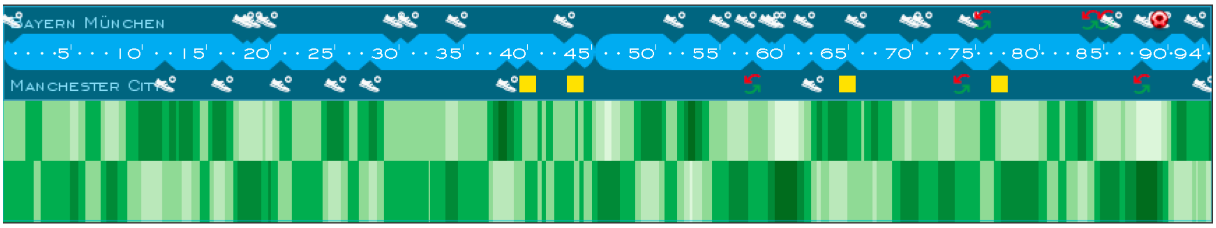


Figure 37: Comparative two-scale pixel-chart timeline visualization for two teams (value-smoothing parameter = 0.5).

III: Comparative dominance indicator visualization for two teams (single-scale representation)

The alternative approach to displaying a comparative view for a dominance factor is the single-scale representation integrating the value progressions of both teams into a unified timeline based on ball possession. As can be seen in Figure 38, dominance tendencies towards one team are mainly encoded by a diverging color scale (Orange/red = dominance team A; blue = dominance team B). The single-scale pixel-chart representation provides an insight as to which team is dominating the game at which time step and further quantifies the dominance intensity by the color value (e.g. dark red/blue = high dominance). From a viewer's perspective, the pixel-chart timeline appears rather abrupt in depicting phase-changes given that color is the main means of visual distinction.

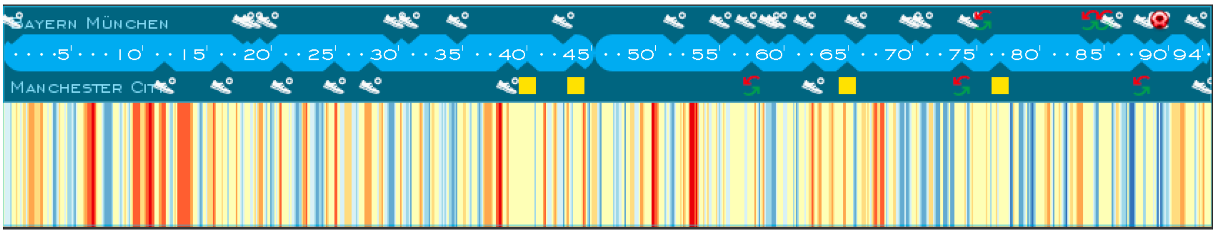


Figure 38: Comparative single-scale pixel-chart timeline visualization for two teams (value-smoothing parameter = 0.0).

Figure 39 shows a more abstracted view of the data-progression in the timeline, making it easier to define clear phases of dominance for one team. The length indication of dominance phases becomes more visible at this granularity level.

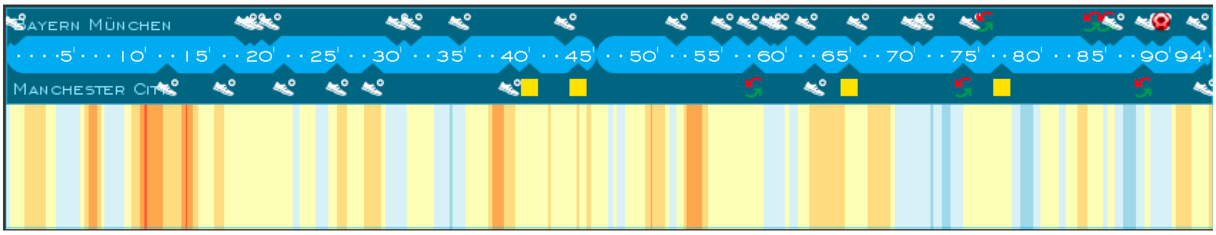


Figure 39: Comparative single-scale pixel-chart timeline visualization for two teams (value-smoothing parameter = 0.5).

B.) Two-Tone Pseudo Coloring Visualization

The second visualization implemented in this work is the two-tone pseudo coloring approach. Similar to the previous segment on the pixel-chart, this representation format is demonstrated for a single team as well as for both sides simultaneously in a comparative view.

I: Dominance indicator visualization for a single team

The two-tone pseudo coloring approach displays the designated dominance factor with a ratio-combination of two discrete colors. A heightened proportion of the upper-band color indicates a higher dominance value whereby, similar to the pixel-chart, the increasing intensity is encoded by a darker color (e.g. dark green). Two-tone pseudo colored timelines convey the dominance value progression mainly by three visual marks: **color** and **peak height** signifying the value markedness as well as the phase duration expressed by the **horizontal length**. This particular arrangement of visual variables adds an additional dimension in comparison to the pixel-chart view, boasting ancillary information potential. Figure 40 shows an unsmoothed two-tone pseudo coloring value depiction. One can identify short phases of strong dominance around the 14th to 18th and 49th to 55th minute. Lighter shaded indications of low dominance appear slightly more prevalent in the second half. It becomes apparent that such a representation seems to be linked to a high degree of visual complexity as the color progression is compressed together. Specific high and low dominance phases of interest are visible but value changes of finer granularity are more difficult to see due to the heightened number of colors involved as well as the compressed representation space. Therefore, it remains burdensome to recognize general dominance trends occurring throughout the timeline.

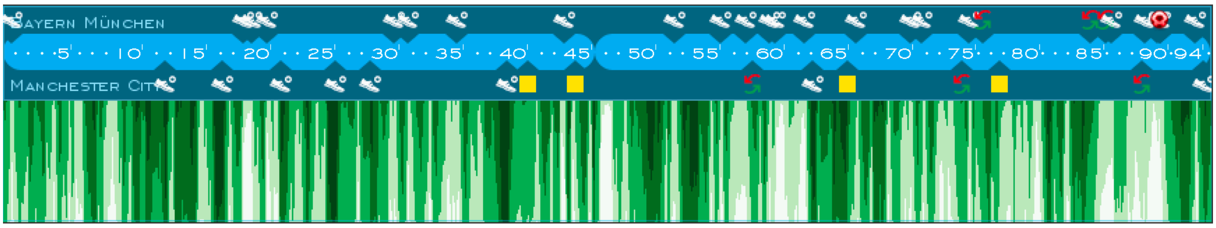


Figure 40: Two-tone pseudo colored timeline visualization for a single team (value-smoothing parameter = 0.0).

Figure 41 below displays the two-tone pseudo colored timeline on a higher abstraction level. Variations in peak size, form and color value are represented more clearly. This allows a more detailed observation of the dominance indicator's value development over the course of the game including the rate of value increase/decrease signified by the slope of the peaks. For instance, throughout the first half of the game the medium green shade is omnipresent and only disappears once at the 20-minute mark. On the other hand, the second half shows more interruptions of lighter color intensity indicating a higher dominance tendency. Specific dominance phases, such as the one between the 10th and the 20th minute, can be observed in more detail as they are encoded by color, length and a particular peak-shape. The two-tone pseudo coloring approach subsequently shows supplementary informative aspects through added visual variables but appears dependent on the abstraction-level.

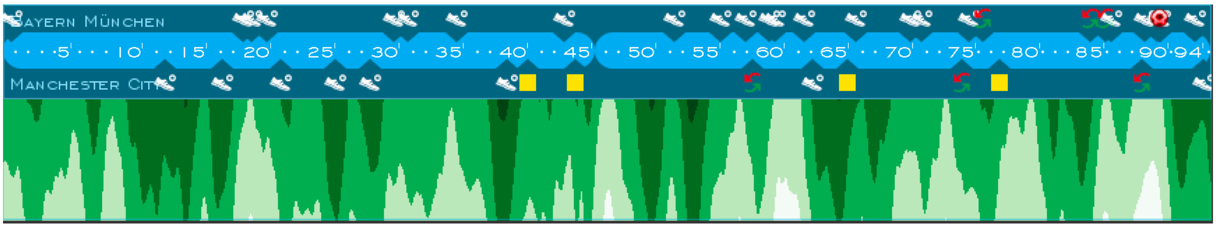


Figure 41: Two-tone pseudo colored timeline visualization for a single team (value-smoothing parameter = 0.5).

II: Comparative dominance indicator visualization for two teams (two-scale representation)

Figure 42 shows a comparative two-scale representation for two-tone pseudo colored timelines. Symmetries are observable throughout different visual variables such as color value, peak shape/height and phase length. This provides a detailed insight into the dominance discrepancy between the both teams. For instance, a segment of the timeline indicating team A being dominant over team B can further be analyzed viewing the exact phase development based on the nature and number of peaks, the color variation and the duration. When considering the timeline contrast between the 67th and 91st minute, one can discern numerous sizable value discrepancies on the basis of peak shape and color which appear to occur over a prolonged period of time as is implied by the vast horizontal extent (length).

Figure 42: Comparative two-scale two-tone pseudo coloring timeline visualization for two teams (value-smoothing parameter = 0.5).

III: Comparative dominance indicator visualization for two teams (single-scale representation)

The comparative single-scale representation of the two-tone pseudo coloring format provides an intriguing alternative approach to displaying the dominance distribution between both teams over the course of a game. Hereby, as displayed in Figures 43 and 44, a reciprocal pattern is shown consisting of upper-band peaks oriented from top to bottom and lower-band peaks going in the inverse direction. These pinnacles are associated with the respective teams. Specifically, indications of dominance for team A are encoded by the red/orange peaks coming from the lower boundary of the rendering space. The reverse scenario, implying team B dominating team A, is manifested through the blue peaks initiating from the opposing top boundary.

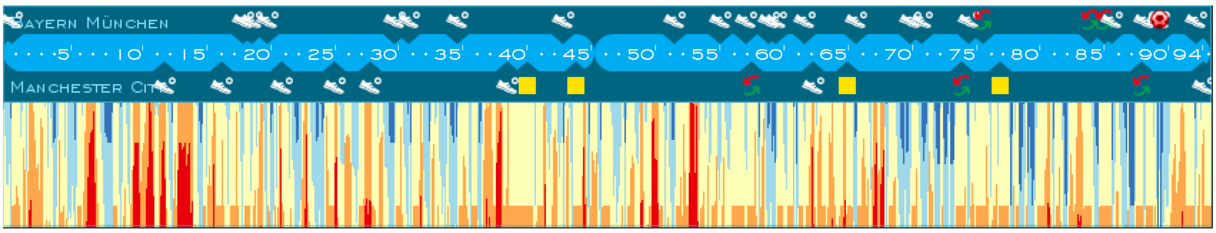


Figure 43: Comparative single-scale two-tone pseudo colored timeline visualization for two teams (value-smoothing parameter = 0.0).

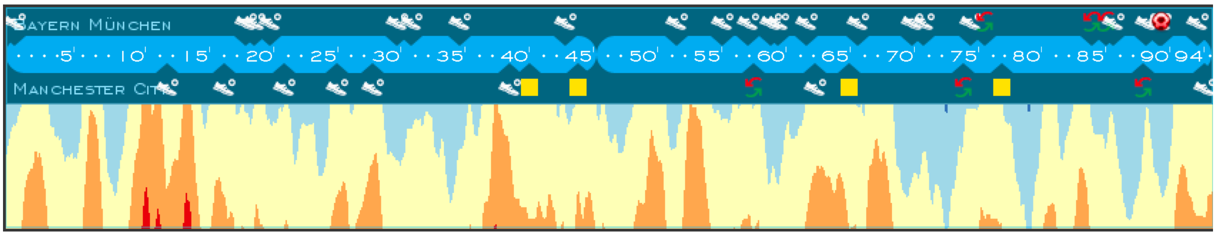


Figure 44: Comparative single-scale two-tone pseudo colored timeline visualization for two teams (value-smoothing parameter = 0.5).

Ultimately, a color-coded proportion scheme of the dominance value progression for both teams is shown in a single timeline display enabling a comparative analysis. Figures 43 and 44 demonstrate the need to increase the abstraction level to achieve better readability and comprehension. As can be seen in Figure 44, the respective phases of dominance are countable and more quantifiable than at a low abstraction level. Figure 43 displays an unsmoothed timeline which poses significant challenges in identifying coherent dominance phases, as they can solely be discerned according to their color encoding. For example, the dark red peaks between minutes 8 and 15 stand for a strong dominance tendency for team A. Nonetheless the local development of the dominance value sequence is barely visible as peak attributes such as height and slope cannot be clearly seen.

C.) Line-Chart

I: Dominance indicator visualization for a single team

The line-chart representations shown in Figures 45 and 46 are, in their visual essence, highly contrasting to the previously described timeline formats. In this approach, it is apparent that data values are strictly encoded by line height disregarding any additional characteristics such as color. The vertical axis of the rendering space uses a value scale ranging from zero to one. At each time step along the continuous value progression, the corresponding dominance intensity level can be deduced by observing the height position of the associated line segment. Figure 45 demonstrates a line-chart with very low value smoothing, thus consisting of a large number of outcrop peaks oriented towards the top and the bottom. On this representation scale visual clutter is inevitable, resulting from numerous rapid value increases or decreases.

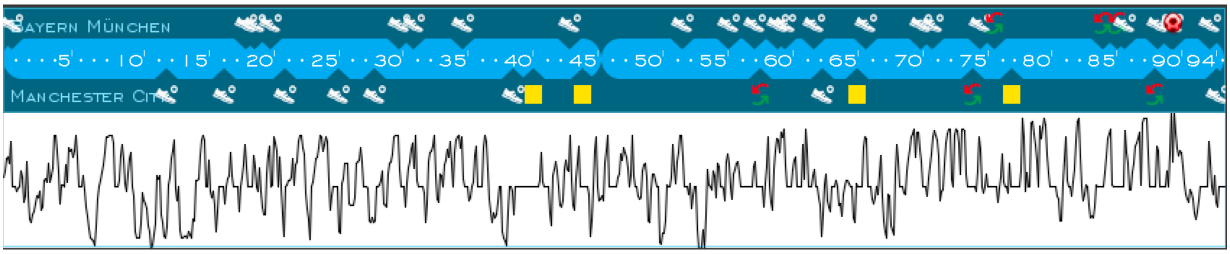


Figure 45: Line-chart timeline visualization for a single team (value-smoothing parameter = 0.0).

Figure 46, on the other hand, displays a high level of abstraction applied to the line-chart representation. The temporal value progression indicated by the line continuum is displayed a lot more smoothly and extreme outcrops are only of marginal significance. At each time step the viewer can identify if the dominance value is low or high, however, it remains challenging to observe general trends shown in the data.

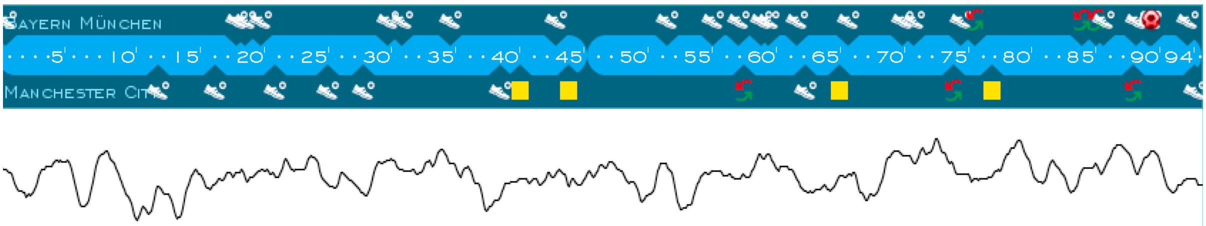


Figure 46: Line-chart timeline visualization for a single team (value-smoothing parameter = 0.5).

II: Comparative dominance indicator visualization for two teams (two-scale representation)

The two-scale representation vertically aligns the line-charts indicating the dominance value progression for both teams. Comparative observations are mainly conducted on the basis of symmetric features. For instance, as can be seen in Figure 47, the respective lines get closer to each other between the 10th and the 20th minute, implying a tendency of dominance towards the team represented below (away team). In contrast, if the line spacing becomes bigger, the inverse scenario is the case.

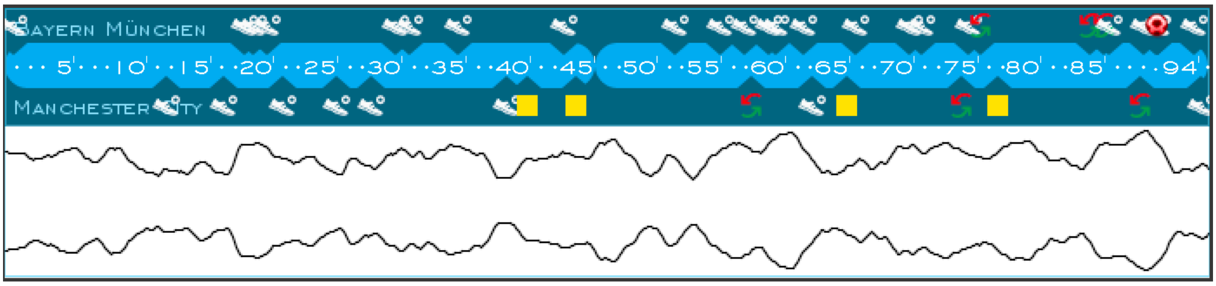


Figure 47: Comparative two-scale line-chart timeline visualization for two teams (value-smoothing parameter = 1.0).

III: Comparative dominance indicator visualization for two teams (single-scale representation)

The single-scale representation depicts the dominance value progression for both teams in a single line-chart. Any outcrops lying above the middle mark of the vertical dimension (0.5) indicate that the home team was dominant at that particular instance of time. Figure 48 displays such distinctive value outliers around the 5th, 10th, 80th and 87th minute. Peaks oriented towards the bottom, for instance around the 12th and 39th minute, display dominance tendencies of the away team.

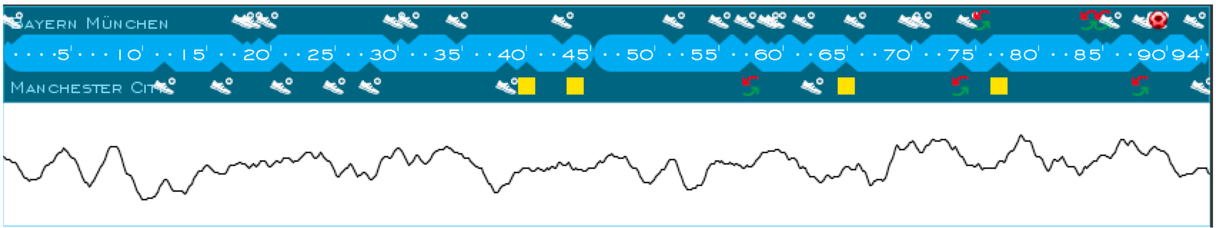


Figure 48: Comparative single-scale line-chart timeline visualization for two teams (value-smoothing parameter = 0.5).

4.2.) Evaluation of visualization results

A thorough review of the computed results is vital for critically assessing the quality and utility level of the timelines. The following subchapters provide a comprehensive overview of the chosen evaluation procedure including specific reference to the responses provided by the domain experts.

4.2.1.) Structure of evaluation procedure

There are numerous approaches discussed in literature regarding the evaluation of user interfaces. It is vital to note that the end product of this work does not include a robust program structure with many different windows and interaction fields, but is mainly confined to two core elements comprised of an option panel and, more prominently, a timeline display.

Based on the research questions defined at the beginning of this work, the main objective of an evaluation is to assess how understandable and useable a timeline visualization is for potential users. Further, it is of high interest to identify how well such a representation format depicting specific game-related phenomena can be integrated into the workflow of soccer coaches and analysts. To address these key questions, the evaluation process was partitioned into to three specific segments with varying approaches. First, a cognitive walkthrough was conducted on the basis of a single timeline view. According to Nielsen (1994) a cognitive walkthrough is defined as an "explicitly detailed procedure to simulate a user's problem solving process through dialogue, checking if the simulated user's goals and memory content can be assumed to lead to the next correct action". In the case of this evaluation procedure, the evaluator is asked to solve two very simple tasks concerned with correctly interpreting the timeline representations of soccer dominance. Through exact observation and open dialogue, the cognitive walkthrough aims to derive knowledge as to how comfortable the users are in understanding the timeline displays and using them for analytical tasks. Hereby, strengths and shortcomings were interpreted to gain an overview of the utility value of the dominance visualizations. The second phase of the evaluation followed the same scheme, asking the users simple questions concerning the comparison of the two respective team's dominance representations. This observation provided further insights regarding the utility level of these temporally-oriented visual support tools. Finally, a qualitative interview was conducted with the domain experts as proposed by Kopp (2011). Kopp (2011) indicates that an expert review should ideally include 4-5 reviewers as this implies a maximal return on investment. Given the challenging circumstances of finding a sufficient number of domain experts, this work resorts to three reviewers, as the informational value is proximal to the optimal scenario. The evaluation panel is comprised of David Andreoli (Headcoach SC Buochs), Thomas Jent (Headcoach FC Baden) and Markus Wanner (Headcoach FC Seuzach).

4.2.2.) Evaluation of results

The following three segments provide an overview of the evaluation results for the timeline visualizations of soccer dominance.

A.) Cognitive Walkthrough Evaluation I: General comprehension of the timeline view

The first part of the cognitive walkthrough was concerned with observing how the user generally interacts with the timeline visualization. For each representation type, two very simple tasks were to be executed on three different visual abstraction levels (smoothing parameter set to 0.33 (low abstraction), 0.66 (medium abstraction) and 1.0 (full abstraction). On one hand, the user was asked to count the number of phases displaying minimal (dominance value below 0.25) and maximal (dominance value above 0.75) outcrops of the visualized dominance indicator. In addition, the phase indicating the highest dominance for the designated team over the course of the whole game was to be identified by the evaluator. Whilst conducting these tasks, the user provided insights into his/her thought process and cognitive perception. Ultimately, the walkthrough comprehended how comfortable the user was with utilizing the timeline visualizations. Further, the results of the tasks signify how well the user was able to seek out specific visual encodings of data information. The following sections summarize the key conclusions derived for each visualization type.

I: Pixel-chart visualization:

All of the evaluation participants understood the visual concept underlying the pixel-chart representation fairly quickly. The few questions asked by the users were mainly concerned with local color value variations. It became apparent that the color variable was the sole element of value distinction given the absence of additional alternative visual marks. This circumstance resulted in the users tediously looking through the timelines to identify any potential dominance phases meeting the criteria, thus making them look rather strained. One evaluator distinctively pointed out that certain areas almost appear to be an optical illusion making it very difficult to assess whether a specific value threshold is undershot or surpassed. Subsequently, the users resorted to the legend numerous times to identify exact color distinctions. Users pointed out that an increase of abstraction value aided them in visually detecting dominance phases. As can be seen in the table below (Table 1) the count of high dominance phases were consistently too high, implying difficulties in distinguishing darker high-intensity values (e.g. dark green color range). On the other hand, low dominance phases were identified correctly concerning their number of occurrences. This is further underlined by the shorter duration of counting such phases in comparison to their higher counterparts. When identifying the phase of maximal dominance, most evaluators did not recognize the highest visually represented value. They commonly stated that the length variable encoding the phase duration was taken into much stronger consideration than the actual color value.

Despite the heightened importance of the length indication, the extent of the maximal dominance phase, as marked by the user, did not increase significantly with increasing

abstraction. Generally, the pixel-chart view appears to be an understandable timeline visualization concept which is predominantly interpreted correctly by the user but potentially overstates the occurrence of high dominance phases over the course of a game especially at lower abstraction levels.

Visualization type	Level of abstraction	Actual number of high dom. phases	Actual number of low dom. phases	Average user count of high dom. phases	Average user count of low dom. phases
Pixel-chart	**0.33**	23	11	28	9.67
	0.66	9	7	16	7
	1.0	4	2	9.33	2.67
Two-tone pseudo col.	**0.33**	15	5	19.33	14.67
	0.66	9	0	10	9
	1.0	3	0	8.67	4
Line-chart	**0.33**	11	6	17	7.67
	0.66	7	0	9.33	3.67
	1.0	6	0	5.33	2.67

Table 1: Table showing the user-counted extreme dominance phases (minimal < 0.25; maximal > 0.75) in comparison to the actual number of high/low dominance outcrops.

II: Two-tone pseudo coloring visualization:

The two-tone pseudo coloring approach was commonly referred to by the experts as a rather novel sight. Given the inclusion of more visual variables such as the peak size/shape and the encoding of data values through two colors, the evaluators needed extra time to consult the legend and observe the timelines more thoroughly. However, the counting process for the phases of minimal and maximal dominance were completed rather rapidly. The evaluators have pointed out that such value outcrops can be easily identified according to their peaks' origin. Light colors depicting low dominance are oriented from bottom to top in vertical direction whereas their darker high dominance counterparts aligned inversely. This circumstance was easily observable in the low abstracted timeline with a high variety of data values. In comparison to the pixel-chart visualization, the users conducted their phase count much more quickly and in a more decisive fashion. One evaluator commented on the fact that he didn't regard each peak as a separate indication of an extreme dominance situation, but rather he grouped the pinnacles together on the basis of a common color value. This statement implies that the two-tone pseudo coloring representation can potentially be visually deceptive to a novel observer as it overemphasizes the appearances of peaks throughout the timeline. For instance, maximal dominance phases (value(s) > 0.75) only occur when the darkest shade of green is visible, initiating from the top boundary of the rendering space, as can be seen in the following illustration (Figure 49).

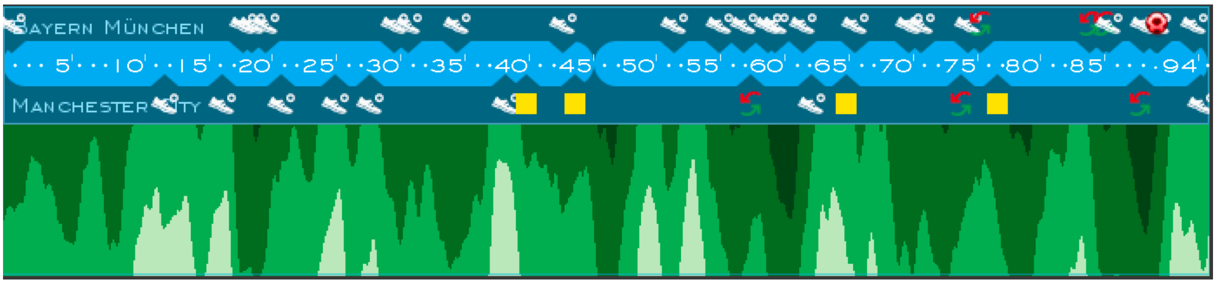


Figure 49: Two-tone pseudo colored timeline representation with nine high dominance phases (values > 0.75) which are encoded by peaks oriented from top to bottom colored in the darkest shade of green.

Many evaluators have pointed out that they are including all sizable peak groups colored in darker green. This implies a lack of understanding of the two-color value encoding which further becomes evident as the user counts of extreme dominance phases are consistently too high as can be seen in Table 1. The identification of the maximal dominance phase was overwhelmingly based on color value and, more prominently, phase length. Most evaluators included the global value summit in their phase delineation which eventually covered more length than in the pixel-chart representation. Ultimately, the two-tone pseudo coloring approach is commonly interpreted correctly on a macro-level. This means that the evaluators are quickly able to identify phases of high and low dominance for the respective teams.

However, observing the timeline on a more detailed level, it remains a challenge for the users to understand the concept of a bottom and a top color band encoding a specific data value.

III: Line-chart visualization:

The line-chart was presented to the evaluators as a simple alternative concept to the color-based representations. All evaluators commented that identification of specific dominance-phases appeared more difficult than in the previous visualization types. Outcrops of maximal/minimal dominance were mainly selected by the height of the line, disregarding the duration length. It was apparent that evaluators struggled significantly to execute the task quickly as they were trying to delve deeply into the timeline. The evaluators almost unanimously agreed that they were unsure if their count was correct implying that they had to guess on numerous occasions. Table 1 underlines the fact that the user-counts were generally positioned too high. Similarly, the identification of the maximal dominance phase was rather ambiguous as the duration-extent varied heavily for different evaluators and abstraction levels. This again indicates an inconsistent perception of the visualization among users implying the existence of visual complexity.

Evaluation of visual abstraction level (data smoothing value)

As shown previously, the tasks were conducted by the user on different levels of abstraction. This parameter is a vital element of the visualization and is taken into further consideration within this evaluation. The users were asked to specify the range of abstraction for each representation type which appeared acceptable to them as well as their preferred abstraction level for visual data analysis. Throughout the cognitive walkthrough the users additionally stated their impressions regarding the different value abstraction levels. The following sections summarize the perceived implications of data-smoothing for the respective timeline formats:

I: Pixel-chart visualization:

The degree of abstraction shown in the pixel-chart has evoked a set of varying responses by the evaluators. Many evaluators found the less abstracted representations to be more informative as patterns within the data can be distinguished and analyzed at a higher granularity level. It was stated numerous times that a high degree of abstraction leads to an exaggeration of dominance tendencies towards one team. For instance, the comparative two-scale view with a smoothing parameter of 1.0 (highest possible abstraction), indicates that the home team was dominant throughout almost the whole game. On this note evaluators have mentioned that lower abstraction levels allow them to observe specific high dominance phases of the inferior team which are hidden in views with heightened data aggregation. In contrast, for some observers, comparative single-scale timelines are more visually pleasing when they are simplified through value-smoothing. This allows them to more easily identify general dominance trends throughout the entire game. When asked to define their suitable range of abstraction, evaluators on average identified a rather broad portion of the smoothing-scale (0.18 – 0.7). The pixel-chart appears to be utilizable at a variety of different abstraction levels depending on the desired aspect of analysis. It was further noticeable that the ideal point of visual simplification at 0.48 was higher than in the other two representations, indicating that certain informative elements are more visible at such an abstraction level. This corresponds with the impression that minimal data smoothing in a pixel-chart visualization leads to a vast number of color-value changes which can be hard to detect for the user. A certain amount of abstraction needs to be available for optimal utility, even more so given the fact that color is the main visual mark of distinction.

II: Two-tone pseudo coloring visualization:

Within the two-tone pseudo coloring representation the tendency towards low abstraction among evaluators is more prevalent than in its pixel-based counterpart. Low levels of data smoothing were generally associated with a higher degree of information. Again, as has been mentioned, highly abstracted timelines appear to hide information which could be interesting when conducting game analysis. These general conclusions are underlined by the fact that users, on average, pinpointed the ideal abstraction-level at 0.27, which is significantly lower than for the pixel-chart. However, similar to the previous timeline representation format, evaluators expressed difficulties with defining holistic dominance trends in a comparative single-scale view with low levels of abstraction. On this note, it was commonly expressed that, in such visualizations, color remains the only form of value distinction whereas higher degrees of data smoothing additionally invoked visual marks such as peak shape/size and length.

III: Line-chart:

The line-chart view was generally perceived as visually challenging to detect specific game patterns. It has become evident that any sort of abstraction appears to amplify the difficulties of counting or delineating phases of maximal/minimal dominance. This circumstance is further highlighted by the fact that evaluators, on average, pinpointed the desired abstraction level between 0.2 and 0.4, both of which are significantly low values.

B.) Cognitive Walkthrough Evaluation II: Assessing the utility value of comparative timeline views

The second part of the cognitive walkthrough assessed the user's capability to compare the dominance patterns of two opposing teams. Hence, they were asked to name the more dominant team and further identify the phase of maximal dominance value discrepancy based on the two-scale and the integrated single-scale timeline representations. The following segments provide an overview of the corresponding evaluation results.

Pixel-chart visualization:

I: Comparative two-scale representation

Referencing Figure 50 below, all evaluators have identified the home team (encoded by the upper timeline) to be more dominant due to the symmetry of color values. The color contrast allows a very rapid differentiation of the dominance levels of both teams as proven by the short response time of the experts. Two evaluators have clearly stated that the lighter color values (low dominance) are more salient and recognizable than the darker ones. With increasing abstraction, the dominance tendency of one team becomes more evident. This implies that slighter dominance discrepancies are more hidden, potentially reducing the information value. This situation is also shown in the delineation of the phases indicating the highest dominance discrepancy, which are clearly indicated over a longer duration at higher abstraction levels. The comparison of the respective team's dominance value progression appeared to be very straight forward and clear to all participants.

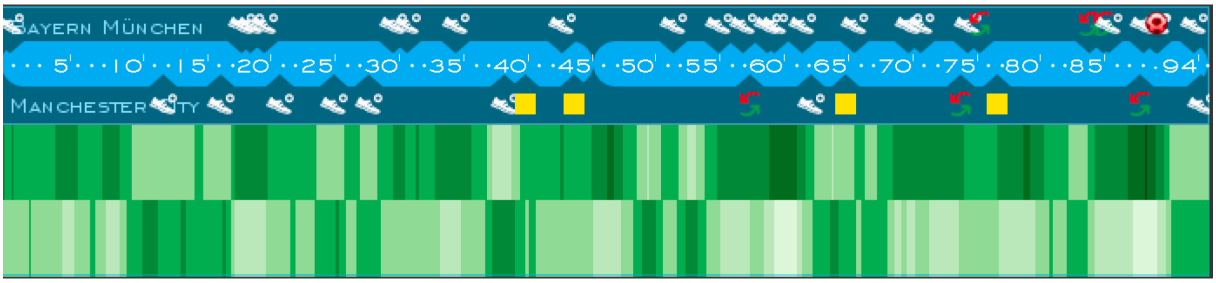


Figure 50: Comparative two-scale pixel-chart representation.

II: Comparative single-scale representation

Equivalent to the two-scale representation, all evaluators have identified the home team (blue color values) to be more dominant, based on the timeline shown in Figure 51. The single-scale representation was understood very quickly resulting in a fast completion of the tasks. It was noticeable that the phase length was of marginal consideration, as the evaluators commonly counted the respective orange-red and blue segments. One user underlined the circumstance that middle values indicating an equal dominance distribution are represented in beige, thus visually leaning more towards the orange color values as can be seen in Figure 51 below. This issue potentially also applies to alternative diverging color scales not used in this work. When identifying the maximum dominance discrepancy, the evaluators chose varying approaches. For some the key indicator was the duration (length) of a dominance period without being interrupted by a contrasting phase, disregarding the color values. Another evaluator placed more emphasis on the color intensity resulting in very short length marking. Ultimately, the single-scale pixel-chart appears a bit more diffuse but is generally perceived positively by the users with similar results as in the two-scale representation.

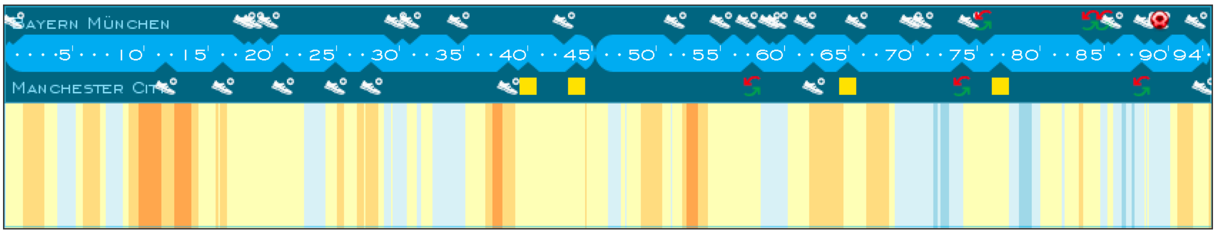


Figure 51: Comparative single-scale pixel-chart representation.

Two-Tone pseudo coloring visualization:

I: Comparative two-scale representation

When utilizing the comparative two-scale representation displayed in two-tone pseudo coloring format, evaluators unanimously identified the home team (upper band) to be dominant. Similar to the pixel-chart representation, observers pointed out the symmetrical visual properties which, in this case, are not only indicated by color but also by peak height/shape/size, making differences even more distinguishable. Most evaluators stated that dominance discrepancies can be viewed on a much more detailed level due to the multi-variable encoding. This allows more exact insights into the dominance distribution throughout phases which appear more balanced but strong tendencies can't be pinpointed as easily. Figure 52 shows contrasts between two value scales based on numerous indicators such as color, peak shape and duration simultaneously.

Figure 52: Comparative two-scale two-tone pseudo coloring representation.

II: Comparative single-scale representation:

The task of identifying a more dominant team within the comparative one-scale representation indicated a change over the range of abstraction levels as to which visual marks were observed by the viewer. At the lower end of data-value aggregation the evaluators oriented themselves to the displayed colors whereas more abstracted timelines resulted in a stronger focus on peak-counts and length (peak-width). Most evaluators stated that they counted the peaks originating from the top and bottom boundary of the rendering space and ultimately calculated a ratio to make a valid assumption. Further, they pointed out that the task became easier when done on a highly abstracted timeline. One observer mentioned that peaks oriented from bottom to top could potentially be more striking and detectable, indicating a potential visual bias. For instance, in Figure 53 below, the orange-red pinnacles could possibly be perceived as more eye-catching to the viewer.

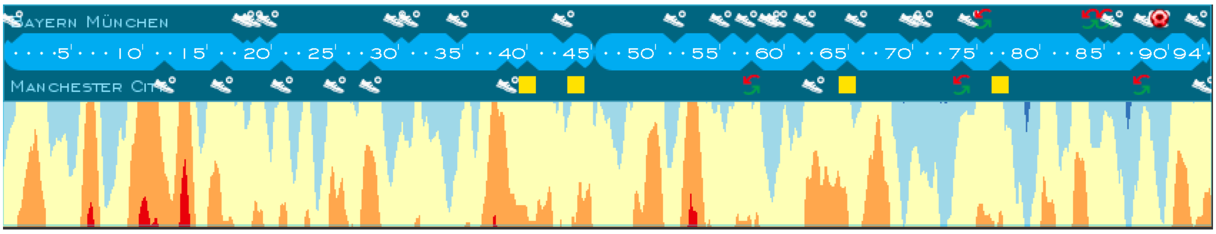


Figure 53: Comparative single-scale two-tone pseudo coloring representation.

Line-chart visualization

I: Comparative two-scale representation:

When using the two-scale line-chart representation to compare the value progression of the respective team's dominance patterns, all evaluators aimed to identify the discrepancies by observing the spacing between the lines. The illustration below (Figure 54) provides such a display. One can observe that the closer the lines are together, the more dominant the team below (away team). It took the evaluators a significant amount of time to determine the more dominant team. They mostly stated that it was close to impossible to make such a distinction simply based on a line-chart figure. Even when adding supporting lines into the graphic, which most line-charts have (not shown in the illustration below), it appeared challenging to obtain a decisive answer. Subsequently, the evaluators gave different answers when executing the task on the timelines with low abstraction. Timelines which were fully smoothed commonly gave no clear indication as to which team was more dominant throughout the game and resulted in responses stating the visual impossibility of conducting the task. When asked to delineate the phase of maximal dominance for one team, the evaluators mainly proceeded on the basis of observing the temporal extent (horizontal length). On all abstraction levels, they provided a wide range of different answers underlining that the visual complexity results in inconsistent interpretation patterns of the timeline.

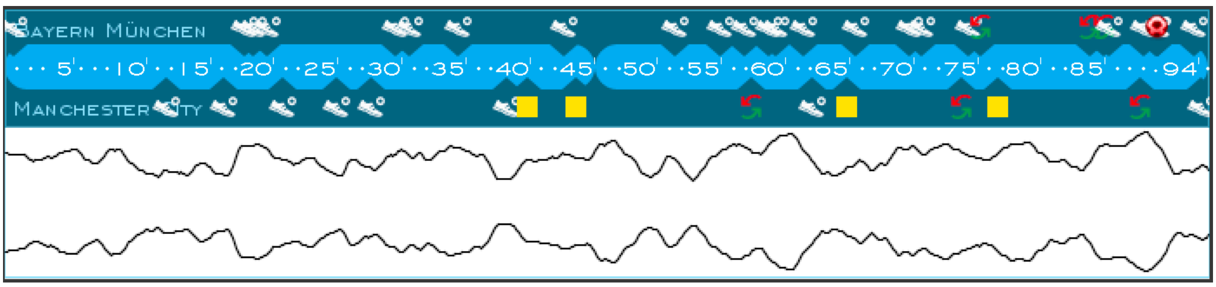


Figure 54: Comparative two-scale line-chart timeline representation.

II: Comparative one-scale representation:

The comparative one-scale line-chart representation was deemed of no utility value to conduct analytical tasks by the evaluators. They unanimously stated that decisively identifying a more dominant team or specific phases of maximal/minimal dominance is not achievable on the basis of such a line-chart view as is displayed in Figure 55. Stronger levels of abstraction further increase the imposed difficulties. While seeking out the phase with the highest discrepancy of dominance values, the evaluators aimed to find the highest/lowest outcrop-peak. However, the results varied heavily among the different observers implying an inconsistent view of the line-chart display. It is vital to note that the displays were complemented with supporting lines, which are not shown in the illustrations within this chapter, to make them more readable for the viewer.

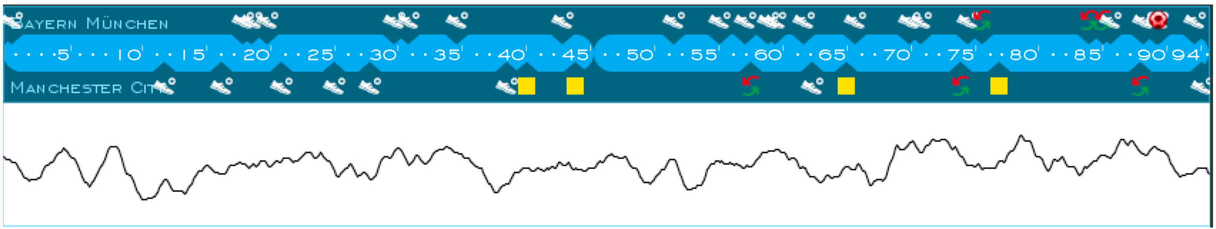


Figure 55: Comparative single-scale line-chart timeline representation.

Additional questions

In the final step of the task-based cognitive walkthrough evaluation the domain experts were asked the following additional questions regarding the utility value of comparative timelines:

I: Which visualization type is most conducive to performing visual analysis tasks?

All evaluators agreed that the best visualization format for conducting game analysis is the two-tone pseudo coloring approach. They stated that it was clearly the most visually pleasing alternative of the three. Given the encoding of data values through numerous visual marks such as color, peak size, shape, slope and duration (length), a more detailed observation is achievable. Dominance phases can be viewed on a finer granularity providing more insight into their local development. The pixel-chart representation was also seen in a positive light as it was consistently interpreted correctly and, thus, carries utility value. Some evaluators commented that the color distinction was challenging at times, mostly at low abstraction levels. Some difficulties can arise when using such a visualization format but it is, nonetheless, perceived as an understandable and useful timeline visualization. The line-chart was unanimously rated as the least favorite alternative, given the severe challenges underlying its correct interpretation. Evaluators have commented that analysis tasks can be arduous on such a visual basis often resulting in imprecise assumptions and incorrect conclusions.

II: Which scale-view is preferred for comparative analysis?

When asked for their preference among the comparative representations none of the evaluators provided a differentiating answer and implied that they thought that both the single-scale and two-scale were sufficient. However, after some further observation, they expressed mildly favorable tendencies towards the integrated single-scale view.

C.) Qualitative Evaluation: Assessing the utility potential of dominance visualizations in the professional soccer workflow

In order to suitably assess the potential of integrating visualizations of dominance in the workflow of soccer coaches and analysts, a qualitative interview has been conducted with each domain expert. The following segments provide a summary of their responses to key questions regarding the utility of the computed timeline representations.

I: How useful/meaningful is a timeline visualization of dominance?

Timeline visualizations depicting the temporal value progressions of dominance indicators were generally perceived as useful tools that could potentially assist analytical tasks concerning the game of soccer. Domain experts have mainly highlighted its supporting character to convey general trends that can be observed throughout the game. Further, the idea of aggregating several single-game timelines appears to be highly intriguing to coaches as it would allow them to obtain long-term tendencies in a team's performance. On the other hand, the simple macro-level depiction of the timelines could be of interest for media outlets since it offers a simple form of visually communicating game data. In addition, such insight could be of interest for general spectators who are more interested in the broad pattern occurring throughout the game, similar to ball possession statistics, etc. Concluding, one can say that a timeline visualization of dominance appears to evoke positive resonance as experts have identified several potential paths of exploitation and development.

II: At which stage of the practice-game cycle during the season would the application of a dominance visualization make the most sense?

Most evaluators view the dominance visualizations as useful tools to be applied retrospectively to an individual game, whereby the previous game is analyzed thoroughly by the coach. Even more significantly, they highlighted the potential value of aggregating several games together to derive larger-scale trends. This could, for instance, be of interest to view a team's development over the course of an extended time period of interest, such as the preparation camp, the preliminary round or the last ten games. Being able to visually represent such information allows the monitoring of progress or the detection of specific patterns in the opponent's and own team's style of play. As a result, domain experts appear to have an interest in periodically using a dominance timeline visualization to gain insight into the long-term development of a team's approach to soccer.

III: For what purpose, could a timeline visualization of dominance be useful?

A first identified area of application for dominance visualization is to monitor the performance of a team over the course of a game or a prolonged phase comprised of a certain number of games, aggregated accordingly. Evaluators have stated that often the impressions at a live game are contrary to more rationale observations conducted, for instance, over video after the game. It is, therefore, intriguing to have access to a visual tool which allows the verification or rectification of subjective game impressions. Long-term developments can potentially be captured providing feedback as to how well teams are progressing in becoming more dominant.

Clearly, single game analysis is an additional key area where the application could make a distinct contribution. On the one hand, aggregated timeline representations of dominance indicators can convey certain temporal patterns as to where in the game the opponent usually acts more or less dominant (e.g. consistent clear drop in dominance during the last 15 minutes of the game). The opponent potentially loses the element of surprise making them more predictable. Evaluators also see an opportunity in assessing their own team's performance, providing a foundation for delineating areas of further improvement.

Dominance visualization could also be used to communicate information to the players in a rapid and visually pleasing manner. Domain experts commonly agree that verbal interaction does not always reach the player effectively. Timeline representations can also be beneficial for transmitting a graphic insight into an opponent's temporal game patterns. This could potentially trigger thought processes as to how the game plan shall be strategically constructed (e.g. at what point shall the opponent be pressured; at what point do we retain our energy and act more defensively, etc.). Some evaluators have mentioned that dominance visualizations should only be used on focused situations to underline very specific areas of high importance. If players are exposed too strongly to such infographics they might be overwhelmed with information and will tend to disregard the provided game insights.

Finally, all evaluators agreed that timeline visualizations can definitely be used as an encoding scheme for significant situations of interest throughout the game which require more attention. For example, a phase between the 70th and 80th minute implying a very heavy dominance tendency towards the opponent could prompt a coach's or analyst's curiosity and he/she might want to review this scene in more depth on video.

IV: Does it make sense to base tactical decisions within a game, such as substitutions, on information provided by a timeline visualization of soccer dominance?

The domain experts unanimously stated that decisions made within a single game should not be based on such broad trend overviews. They prefer to utilize visual tools for long-term monitoring as opposed to dramatically influencing their game-related tactics on the basis of a single set of indications. Consequently, a real-time application seems to be of marginal interest for this purpose factors were commonly perceived as reasonable indicators covering some of the most vital aspects of the dominance concept. It is essential to note that many evaluators stated that it is almost impossible to find a common definition of dominance since every individual has a different mindset as to how he/she aims to approach the soccer sport. Domain experts would likely prefer to be able to establish and customize particular features according to their personal taste. This would imply a significant amount of additional work and one-on-one interaction between the developer and the user, potentially facing severe limitations of data availability and computational challenges. Ultimately, the experts have underlined the importance of proximity to the opposing goal as an inevitable factor in notions of pressure and dominance. In conclusion, the evaluators agree that the computed features are a solid foundation to formally express such a broad phenomenon.

V: Are temporally-oriented visualization formats generally perceived as appealing and useful?
Temporally-oriented visualization formats have evoked a generally positive response by evaluators. It has been noted that timeline representations can be informative tools for coaches and analysts, whereas spatial visualization depicting the soccer pitch are more appealing for players. The optimal solution appears to involve a combination of time-based visualizations exposing large-scale trends and spatially-oriented representations showing specific situations.

VI: Are feature-based visualization formats generally perceived as appealing and useful?
The approach toward dismantling a specific soccer game phenomenon into different features appears to make sense to most domain experts. One evaluator stated that single indicators are more informative, despite the trend to aggregate data into holistic views. However, most experts were open to the idea of aggregating certain factors into a single metric.

5.) Discussion

The approach provided in this work can be summarized as a procedural chain, wherein a broad game phenomenon is dismantled and formally defined before being computed and visualized in a timeline representation. Given the very recent research evolution of the visual analytic domain in the soccer sport, there is still a lot of potential for novel developments extending this focus area. This work's main contribution is a critical assessment of timeline representations depicting soccer dominance indicators according to their visualization type and utility value. Timeline displays have only marginally been explored in a team sport context through existing literature. The results of this work have yielded some interesting findings which provide an initial insight into the compatibility of such time-oriented, visually-based applications and highlight further areas in need of research attention. In this section the work is discussed in the context of the initially stated research questions. This should ultimately provide a perspective as to how the work is to be interpreted and valued in relation to existing literature in this field.

5.1.) Discussion of research question outcomes

The following segments aim to provide a comprehensive overview of the outcome of this work with regard to the three research questions defined in the Introduction.

1.) Research question regarding the visualized phenomena (what-dimension):

Research Question I: Is it possible to effectively express the broad concept of dominance in the soccer sport through a set of identified and subsequently formalized factors?

Hypothesis: Dominance is a vague and ambiguous concept within the soccer realm. Each individual has a unique interpretation of the term, implying a heightened complexity in finding a commonly acceptable format for expressing its meaning effectively. The initial hypothesis assumes that, despite the difficulty of defining dominance, a satisfactory first solution can be found that encompasses a commonly agreed understanding of the term.

Given the loose definition of dominance in a soccer context, this work has attempted to dismantle its broad meaning into a set of more concrete factors. A feature-driven approach has been applied, mainly inspired by the works of Sacha et al. (2014) and Stein et al. (2015). Within the evaluation it has become apparent that most users were open to representing game-related phenomena by a collection of different indicators. It is vital to recognize that the aforementioned influential papers were mainly focused on segmenting an entire soccer game into various factors generally concerned with passing, speed, free spaces or number of involved players. In contrast, this work concentrates on a single phenomenon operating on a more detailed level of game analysis. The positive reaction of potential users to the multi-feature expression of dominance underlines the legitimacy of further pursuing more sophisticated and advanced approaches to conveying specific indicators more precisely. As stated, Sacha et al. (2014) and Stein et al. (2016) have proposed very intriguing concepts that holistically aim to dissect a whole game into a set of features which can even be sorted to elicit and annotate specific game phases. Based on this work, it appears desirable to revisit these factors on a finer granularity level in order to understand and analyze the game in more detail. This conclusion is supported by some domain experts stating that, despite the trend to aggregate and simplify information accordingly, the core interest of coaches often lies in specific, individual aspects of the game. Therefore, the feature-driven approach, as applied in this work, can be seen as an intriguing and valid procedure towards making a soccer game more comprehensible on many different levels of viewing granularity.

Dominance is a very broad phenomenon which triggers a wide variety of responses from domain experts. Whereas Andrienko et al. (2017) note the absence of visual analytics on the topic of pressure, this work similarly mentions the equivalent gap with respect to offense-related dominance. Taki and Hasegawa (2000) have attempted to visualize dominant regions in a spatial sense, based solely on the condition as to which player can reach a specific field location more quickly than any other players. This work can be seen as an initial attempt within the realm of visual analytics in soccer to grasp and formalize such an ambiguous concept in a more holistic measurement approach. Throughout the exploratory interviews, the domain experts emphasized varying ideas of dominance linked to observations of physical and mental factors as well as situational game aspects (e.g. ball control). Despite these diverging opinions, this work has successfully identified the general desire to express dominance through more complex and multi-faceted measures linked to specific game patterns. Simple common factors, such as ball possession and tackling predominance, were of marginal interest. In contrast to Taki and Hasegawa’s (2000) work, a more multi-dimensional feature-driven approach as conducted in this work appears more realistic and suitable for informative game analysis. Ultimately, this work has formalized perhaps the two most powerful interrelationships driving dominance on the soccer pitch – *players in relation to the opposing goal* and *players in relation to other players.* All evaluators agreed that the chosen indicators are clearly among the most significant when aiming to convey the notion of offense-related dominance. Subsequently, the domain experts were generally satisfied with the formalized definition of dominance. However, it has become evident that each potential observer desires the ability to personally customize and redefine a specific game aspect for their own use. Within this work, dominance is only expressed by a set of indicators, purposely disregarding their aggregation to a single figure portraying the concept as a whole. There appears to be interest in computing such an integrated view, however at this point no reasonable suggestions can be made as to how the additive calculation shall be conducted.

Arching back to the initial research question and the associated hypothesis, one can say that this work has been able to delineate a more or less commonly agreeable expression of the dominance idea in soccer. Mainly, the feature-based approach has proven to be a legitimate attempt towards dismantling and formalizing such a broad and ambiguous aspect of the game. Subsequently, this work supports feature-driven analysis attempts proposed in Janetzko et al. (2016), Sacha et al. (2014) and Stein et al. (2016) as this concept was deemed favorable by the domain experts. It remains vital to stress that the elicited and computed factors do not provide a conclusive definition of dominance. Rather, the resulting dominance expression should be viewed as an initial approximation derived from a set of three specific indicators. Based on the adaptive character of the feature-driven analysis approach, the current state of formalization can be extended in collaboration with other experts in order to more precisely convey dominance. In conclusion, the conducted formalization of the dominance concept in this work appears valid and legitimate to prompt informative exposure of a soccer game. For future work, it is vital to constantly reiterate and explore new indicators which can be further incorporated in order to more robustly approximate such an ambiguous phenomenon.

2.) Research questions regarding the visualization type (how-dimension):

Research Question II A: Are timeline displays comprehensible and interpretable by the domain-originating observer (soccer coach/analyst)?

Research Question II B: Are visually rich, color-encoded timeline displays of higher information value to the domain-originating observer than simple time series representations such as line-charts?

Hypothesis: Timeline displays are generally comprehensible and interpretable by the domain-originating user. Further, the domain-originating observers/users prefer a visually rich, color-encoded timeline representation as it is more interpretable, informative and aesthetically pleasing.

Throughout the evaluation, timeline visualizations have proven to be a comprehensible analytical tool evoking positive responses from domain experts. The common concept of representing the progression of data values along a linear temporal extent was generally understood by the users without much further explanation. Hence, the findings of this work confirm that time serial displays in soccer game analysis are a legitimate form of visually conveying valuable match information and deserve more research attention. Domain experts have stated that temporally-based views could be increasingly relevant for their own observations whereas spatial depictions of the soccer pitch are more applicable to players.

Examining the various timeline formats which have been implemented within the scope of this work, several interesting aspects have arisen as to how well the potential user is able to interpret and derive valuable information from such representations. Pixel-charts solely discriminate values with color variances which was very noticeable to the evaluators. It was stated numerous times that lighter colors appeared more salient which corresponds with the *simultaneous contrast* phenomena. Hereby, a unit displayed in higher lightness surrounded by darker instances appears lighter to the viewer (Brewer, 1997). On the other hand, the distinction of higher data values represented by darker shades has proven to be challenging implying the need for an extended discourse on the ideal number of color classes for timeline displays. This work has followed the guidelines provided by Harrower and Brewer (2003) to ensure a consistent and valid selection of used colors. However, it is vital to note that no elaborated focus was placed on observing the influence of different coloring schemes as this would have exceeded the defined scope. The two-tone pseudo coloring approach graphically encodes data with an extended number of visual marks in comparison to the pixel-chart display. Given the heightened complexity of integrating these indicators in a single visualization format, the evaluators clearly needed more learning time. Thereafter, the interpretation of the timeline appeared much more seamless and decisive as they were able to orient themselves to a variety of different indicators including color, peak shape (slope) and peak size. As a result, the two-tone pseudo coloring approach was deemed most favorable from an aesthetic and information-value perspective among all candidates. This finding supports the general attempt in research to maximize data density in a spatially confined area as is noted in Heer et al. (2009). However, throughout the observation of user interactions with the two-tone pseudo colored timeline, it has become apparent that the different visual marks don't always accumulate to a higher information value. In certain scenarios, users will almost solely focus their attention on one visual indicator while marginalizing or even disregarding other factors shaping the display. For instance, peaks are almost always associated with very high dominance values without further looking at the underlying color ratio. This conclusion calls for more attention from a research perspective regarding the interrelationships between visual variables occurring simultaneously in a representation of high data density. On the other hand, line-charts have evoked rather negative feedback based on the fact that the value interpretation was inherently challenging. This work concludes that timeline visualization formats encoding data with a set of different visual attributes are more usable and interpretable than simple line-charts, which corresponds with the result explanation in Sacha et al. (2014). Similarly, Heer et al. (2009) have found line-charts to induce a higher error rate when conducting interpretation tasks. The unfavorable tendency towards line-based representations is also present in comparative views which are more prone to visual clutter. On this note, Javed and Elmqvist (2010) have highlighted the importance of *identity* for distinguishing between multiple time series. Hereby, "area" techniques such as the pixel-chart or two-tone pseudo coloring approach can display value progressions more uniquely based on color or fill pattern whereas line-charts are usually graphically limited to a single thin line. Another important aspect which has become apparent within the results of this work is the influence of visual abstraction of data on the user's perception of the timeline. Evaluators have generally preferred lower levels of abstraction, implying more detailed and insightful information. Heavily smoothed timelines appear to hide particular patterns underlying the game and further overemphasize the dominance of one team over another in a comparison view. All visualization formats have in common that, despite the trend towards low abstraction, it is desirable that a lower boundary (about 0.2 on the smoothing-scale) is not undershot due to the strong presence of visual clutter. This work does not provide a conclusive answer as to which visual abstraction levels are ideal, but much rather aims to highlight that this aspect appears to be highly ambiguous and, therefore, requires more research attention.

In conclusion, the initial hypothesis assuming that timeline displays are generally comprehendible and interpretable for decision makers in the soccer domain can be confirmed. Also, it has become evident that more complex timeline visualizations with heightened data density are clearly favorable in comparison to commonly used line-chart representations. However, the findings of this work also indicate that there are numerous factors which need to be addressed in more detail, including the level of visual abstraction or the interrelation of visual variables in the timelines. Hence, this work only provides a first attempt at quantifying and understanding the role of linear timeline representations as a means of support for soccer game analysis. As noted in the Related Work chapter (Chapter 2), there exist an extensive number of timeline formats that could equivalently be implemented. This underlines the fact that this work can only cover a small partition of an enormous design space which is still significantly growing through an abundance of new visual implementations (Aigner et al. 2011). Nonetheless, the chosen timeline formats are very simple and basic concepts which have a gateway function towards prompting further research focus on understanding the interplay between time-oriented visual analytics and invasive team sports. In summary, this work has made an initial contribution in deeming visually rich, color-encoded timeline displays to be compatible, even favorable, information visualization tools which can easily be interpreted by soccer domain experts. However, many influencing factors, such as the chosen visual variables or the level of abstraction, have only been regarded in an initial attempt. They evidently need further elaborated research focus, mostly in the thematic context of applicability in a team sport environment.

3.) Research questions regarding the utility level of the visualization (application-dimension):

Research Question III: Is a timeline visualization of the dominance phenomena in soccer utilizable for a domain-originating observer (soccer coach/analyst) and can it be integrated as a useful tool into their respective workflow?

Hypothesis: Timeline visualizations representing soccer dominance can be used for several different purposes including visual communication to players, analysis of one's own and opposing teams, and performance assessment. Hence, they are not confined simply to being a means of identifying interesting game situations. Coaches can use the application to shape their game-related tactics according to the displayed dominance patterns. Most prominently, this involves player substitutions and managing the style of play (offensive, defensive, level of pressure, etc.) on the basis of dominance indications for the opposing team.

In a very general sense the timeline visualization concept for dominance indicators has prompted overwhelmingly positive responses by domain experts. The application was deemed an intriguing approach towards visually depicting game patterns over the entire temporal extent. Users were able to quickly grasp the idea of timeline displays, hence indicating a fundamental sense of utility.

Reflecting on the purpose of such a supporting tool, a number of different aspects have been highlighted. The existing literature on visual analytics in soccer has overwhelmingly stated that the core objective is to simplify the identification process of interesting game situations (e.g. Legg et al., 2012; Sacha et al., 2014; Shao et al., 2016). This notion has been confirmed unanimously by the evaluators, indicating that a timeline provides an insightful overview encoding specific trends which, in a second step, can be re-observed in more detail. However, the results have clearly shown that a visual analytic application can also be used for various other purposes. For instance, the utilization of such representation-based means to communicate game information to players appears to be a legitimate usage. This work has made a positive step towards discerning visually pleasing timeline formats for coaches and analysts. Due to the simple and quickly comprehensible concept of timeline visualizations, an extension of the potential user group to include players, is likely valuable. The responses gathered by the domain experts have indicated further possible fields of application, such as monitoring the team's performance, general game analysis or validation of subjective impressions which are often thought to be skewed versus reality. It is vital to highlight that this work cannot exactly quantify for which particular purpose visual analytic tools in soccer are best suited, as this is most probably highly dependent on individual preferences. However, the findings clearly show that the discourse on the functional reasoning of such applications needs to be dismantled and focused on different potential scenarios of usage. The first part of the research hypothesis can be confirmed, based on the indication that visual analytic tools can be utilized for many different purposes, all of which are in need of deeper research. An essential factor that has been mentioned numerous times while evaluating the timeline display's utility value, is temporal scaling. Most domain experts have stressed that the feature observations should be aggregated over several games or even an entire season to derive big-scale trends, similar to the conclusion by Legg et al. (2012). Extending this notion, coaches/analysts appear more interested in long-term development than in single game indications. The initial hypothesis assumed that the presence of visualization tools, such as timeline representations, could support and structure tactical decisions. However, this presumption has proven to be incorrect based on the deduced evaluation results. Domain experts were almost unanimously opposed to overemphasizing the significance of single game data and aligning their tactical inputs according to such information. Only aggregated trends are believed to offer some informative elements which could be valuable for game preparation. Yet, it remains important to highlight that any derived knowledge is often seen as a sole factor among many and does not have the power to shape entire game plan considerations, such as player substitutions. On this note, experts also stressed that it is challenging to exactly define when and how often to use a timeline depiction of dominance indicators as a means of communicating game-related information to the players. In summary, the findings of this work have identified that the temporal scaling factor (single game, several games, entire season) is clearly a relevant cornerstone when discussing the utility level of such a time-oriented visual analytic tool. Existing literature has mainly been concerned with developing visually-based solutions for identifying and annotating interesting game situations which are based on single-game views (e.g. Sacha et al., 2014; Legg et al., 2012; Shao et al., 2016). However, the temporal framing appears not as clear-cut for tasks encompassing visual communication, performance monitoring/assessment or game-trend analysis. As a result, this aspect needs to be taken into further consideration when developing new time-oriented visualization approaches for team sport analysis.

Ultimately, the applicability of timeline representations, as proposed in this work, is surely legitimate and has received positive feedback from domain users. It is essential to understand that the implemented approach only covers a very specific aspect of the soccer phenomena, both on a representational and thematic dimension. Therefore, when placing this work in the broader context of its possible utilization value, various factors of ambiguity need to be taken into consideration. For instance, the diffuse results in this thesis regarding application purposes for visual analytics in soccer call for a stronger focus on eliciting which tasks should be supported. As a result, the outcome of this work does in no way provide a conclusive quantification of how applicable a timeline visualization is to analyzing soccer. However, the evaluators have provided clear indications deeming timelines as valid representations that can provide insightful information on the temporal progressions of specific game factors. It is also important to stress that many domain experts are still hesitant about directly influencing game tactics based on information derived from supporting visual tools. As a result, this work advises a stronger focus on depicting long-term trends as this appears to be a desired area of interest among coaches and analysts. Based on the fact that potential application purposes exceed the singular task of delineating interesting game situations, it is further challenging to specifically pinpoint an ideal time of usage within the training-game cycle as is done in Stein et al. (2016). Coaches have stated that analytical tasks are usually conducted during the time between games, predominantly at the beginning of the week. However, the additionally identified interest of observing lengthy trends adds further complexity to finding a decisive answer as to where in the analysis procedure such a visual application can be integrated.

6.) Conclusion / Future Outlook

When summarizing the contribution of this work there are three core elements of relevance. First, an approach has been undertaken towards grasping the underlying structure of a broad and vaguely defined phenomena referred to as soccer dominance. Second, specific metrics were defined in order that the dominance concept could be transferred into tangible, numerically-expressed values and computed accordingly. The third element was concerned with visualizing the temporal value progression of these indicators in timeline displays. All three of these elements have engaged the critical input of domain experts, validating each phase of the work.

Based on the above, this work contributes an intriguing alternative approach to the research domain by introducing a new focus on timeline visualizations. This has led to the ability to highlight new aspects of the game, specifically the notion of dominant play. The proposed representation concepts have triggered positive feedback among domain experts which legitimizes the groundwork laid by this thesis and provides a very encouraging environment for further enhancing the generated results. Also, the feature-driven approach towards visual analytics in soccer appears to be a very unique and promising dimension which can potentially facilitate advances in future research projects.

It vital to note that there are various factors of limitation underlying this work. The executed implementation represents an initial approach to formally defining, computing and temporally visualizing a broad soccer phenomena, namely dominance. The acquired results cannot be seen as final and conclusive, rather, they provide a set of solid, primary indications regarding the interaction between users and timeline visualizations in a team sport context. The first limitation of significance is the subjective nature of the dominance concept as developed through exploratory expert interviews. While these talks have provided informative responses by domain experts, it has become apparent that each individual has his own definition of soccer dominance, evoking a wide collection of opinions. As a result, it is challenging to suitably and definitively enforce a single, valid formalization scheme leading to a unique logical and numerical expression. Further, the complexity of the dynamic game measurements implies that this work, in its scope, has necessarily resorted to computing a small number of very rudimentary dominance indicators. Additional factors have, therefore, been disregarded within this approach. The scope of this work has also only allowed the implementation of a small, selected set of timeline visualizations. As a result, conclusions can only be drawn regarding these specific representations. Alternative graphic displays could also not be discussed within the scope this work. The visual implementation has been conducted according to research-based frameworks to ensure a consistent and appropriate representation to the viewer. However, no additional emphasis was placed on assessing the perception of different variations of the visual variables underlying the timelines. Finally, it has to be taken into consideration that the evaluations were executed on a qualitative basis and, therefore, provide insightful information on the observer's perception of the timeline, but clearly no empirical results.

Shifting the attention to a future outlook, several focus areas can be identified on the basis of this work which represent potential further research areas. Given the recentness of this research domain, the number of existing proposed approaches is still rather small. Arching back to the previous limitation paragraph, the formalization state of the dominance concept is still at a rather embryonic stage and suggests large potential for expansion and refinement. Similarly, in an opposite direction, a big question that remains open is how to suitably weight and integrate single indicators into an aggregated figure, and explore whether this would actually provide value. Another potential area of work would be to develop and examine other display alternatives, for example, different sizes, resolution-levels, positions or color schemes. Perhaps the most essential consideration of this work for the future, as emphasized by the domain experts, is a comprehensive approach to aggregating numerous single game measurements to a cumulative picture showing trends occurring over multiple matches. In summary, the findings of this work suggest numerous fields of potential to further develop this young and growing research niche.

Bibliography

Aigner, W., Miksch, S., Thurnher, B., & Biffl, S. (2005, July). PlanningLines: novel glyphs for representing temporal uncertainties and their evaluation. In *Information Visualisation, 2005. Proceedings. Ninth International Conference on* (pp. 457-463). IEEE.

Aigner, W., Miksch, S., Schumann, H., & Tominski, C. (2011). *Visualization of time-oriented data*. Springer Science & Business Media.

Andrienko, G., Andrienko, N., & Wrobel, S. (2007). Visual analytics tools for analysis of movement data. *ACM SIGKDD Explorations Newsletter*, *9*(2), 38-46.

Andrienko, N., & Andrienko, G. (2013). Visual analytics of movement: An overview of methods, tools and procedures. *Information Visualization*, *12*(1), 3-24.

Andrienko, G., Andrienko, N., Budziak, G., Dykes, J., Fucks, G., Von Landesberger, T. & Weber, H. (2017). Visual Analysis of Pressure in Football. *Data Mining and Knowledge Discovery.*

Bawa, V. S. (1982). Research Bibliography—Stochastic Dominance: A Research Bibliography. *Management Science*, *28*(6), 698-712.

Bertin, J. (1983). Semiology of graphics: diagrams, networks, maps.

Binmore, K. (2007). *Game theory: a very short introduction* (Vol. 173). Oxford University Press.

Brehmer, M., Lee, B., Bach, B., Riche, N. H., & Munzner, T. (2017). Timelines revisited: A design space and considerations for expressive storytelling. *IEEE transactions on visualization and computer graphics*, *23*(9), 2151-2164.

Brewer, C. A. (1997). Evaluation of a model for predicting simultaneous contrast on color maps. *The Professional Geographer*, *49*(3), 280-294.

Carlis, J. V., & Konstan, J. A. (1998, November). Interactive visualization of serial periodic data. In *Proceedings of the 11th annual ACM symposium on User interface software and technology* (pp. 29-38). ACM.

Carpendale, M. S. T. (2003). Considering visual variables as a basis for information visualisation.

Cava, R., & Freitas, C. D. S. (2013). Glyphs in matrix representation of graphs for displaying soccer games results. In *SportVIS-Workshop on Sports Data Visualization. Atlanta, Georgia, USA: IEEE VIS*.

Clapham, M. E., Bottjer, D. J., Powers, C. M., Bonuso, N., Fraiser, M. L., Marenco, P. J., ... & Pruss, S. B. (2006). Assessing the ecological dominance of Phanerozoic marine invertebrates. *Palaios*, *21*(5), 431-441.

Cleveland, W. S. (1993). *Visualizing data*. Hobart Press.

Chittaro, L., & Combi, C. (2003). Visualizing queries on databases of temporal histories: new metaphors and their evaluation. *Data & Knowledge Engineering*, *44*(2), 239-264.

Chuah, M. C. (1998, October). Dynamic aggregation with circular visual designs. In *Information Visualization, 1998. Proceedings. IEEE Symposium on* (pp. 35-43). IEEE.

Cook, K. A., & Thomas, J. J. (2005). Illuminating the path: The research and development agenda for visual analytics.

Dominance. (n.d.). In *Oxford Living dictionaries (online)*. Retrieved from https://en.oxforddictionaries.com/definition/dominance

Dong, Z., Liu, D., & Keesing, J. K. (2010). Jellyfish blooms in China: dominant species, causes and consequences. *Marine Pollution Bulletin*, *60*(7), 954-963.

Dragicevic, P., & Huot, S. (2002, April). SpiraClock: a continuous and non-intrusive display for upcoming events. In *CHI'02 extended abstracts on Human factors in computing systems* (pp. 604-605). ACM.

Fairchild, M. (2011). Color-Appearence Models: CIECAM02 and Beyond [PDF]. Retrieved from http://rit-mcsl.org/fairchild/PDFs/AppearanceLec.pdf (26.09.2017).

Fekete, J. D., Van Wijk, J. J., Stasko, J. T., & North, C. (2008). The value of information visualization. In *Information visualization* (pp. 1-18). Springer Berlin Heidelberg.

Flinn, M. V., Geary, D. C., & Ward, C. V. (2005). Ecological dominance, social competition, and coalitionary arms races: Why humans evolved extraordinary intelligence. *Evolution and Human Behavior*, *26*(1), 10-46.

Frank, A. U. (1998). Different types of "times" in GIS. *Spatial and temporal reasoning in geographic information systems*, 40-62.

Gantt, H. L. (1913). *Work, wages, and profits*. Engineering Magazine Co..

Gintis, H. (2000). *Game theory evolving: A problem-centered introduction to modeling strategic behavior*. Princeton university press.

Hadar, J., & Russell, W. R. (1969). Rules for ordering uncertain prospects. *The American Economic Review*, *59*(1), 25-34.

Hargreaves-Heap, S., & Varoufakis, Y. (2004). *Game theory: a critical introduction*. Routledge.

Harris, R. L. (1999). Information graphics. *Atlanta, Georgia: Management Graphics*.

Harrower, M., & Brewer, C. A. (2003). ColorBrewer. org: an online tool for selecting colour schemes for maps. *The Cartographic Journal*, *40*(1), 27-37.

Hawley, P. H. (1999). The ontogenesis of social dominance: A strategy-based evolutionary perspective. *Developmental Review*, *19*(1), 97-132.

Hearst, M. (2003). Information visualization: Principles, promise, and pragmatics. In *Handouts of the tutorial at CHI 2003 Conference on Human Factors in Computing Systems*.

Heer, J., & Agrawala, M. (2006). Multi-scale banking to 45 degrees. *IEEE Transactions on Visualization and Computer Graphics*, *12*(5), 701-708.

Heer, J., Kong, N., & Agrawala, M. (2009, April). Sizing the horizon: the effects of chart size and layering on the graphical perception of time series visualizations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1303-1312). ACM.

Herdal, T., Pedersen, J. G., & Knudsen, S. Designing Personal Visualizations for Different People: Lessons from a Study with Elite Soccer Teens.

Huang, G., Govoni, S., Choi, J., Hartley, D. M., & Wilson, J. M. (2008). Geovisualizing data with ring maps. *ArcUser*, *11*(1), 54-55.

Janetzko, H., Stein, M., Sacha, D., & Schreck, T. (2016). Enhancing Parallel Coordinates: Statistical Visualizations for Analyzing Soccer Data. *Electronic Imaging*, *2016*(1), 1-8.

Javed, W., McDonnel, B., & Elmqvist, N. (2010). Graphical perception of multiple time series. *IEEE transactions on visualization and computer graphics*, *16*(6), 927-934.

Keim, D. A., Ankerst, M., & Kriegel, H. P. (1995, October). Recursive pattern: A technique for visualizing very large amounts of data. In *Proceedings of the 6th Conference on Visualization'95* (p. 279). IEEE Computer Society.

Keim, D. A. (2000). Designing pixel-oriented visualization techniques: Theory and applications. *IEEE Transactions on visualization and computer graphics*, *6*(1), 59-78.

Keim, D. A., Schneidewind, J., & Sips, M. (2007). Scalable pixel based visual data exploration. In *Pixelization Paradigm* (pp. 12-24). Springer, Berlin, Heidelberg.

Keim, D. A., Mansmann, F., Schneidewind, J., Thomas, J., & Ziegler, H. (2008). Visual analytics: Scope and challenges. In *Visual data mining* (pp. 76-90). Springer Berlin Heidelberg.

Khan, M., & Khan, S. S. (2011). Data and information visualization methods, and interactive mechanisms: A survey. *International Journal of Computer Applications*, *34*(1), 1-14.

Kielman, J., Thomas, J., & May, R. (2009). Foundations and frontiers in visual analytics. *Information Visualization*, *8*(4), 239-246.

Kim, N. W., Card, S. K., & Heer, J. (2010, May). Tracing genealogical data with timenets. In *Proceedings of the International Conference on Advanced Visual Interfaces* (pp. 241-248). ACM.

King, R. C., Mulligan, P., & Stansfield, W. (2013). *A dictionary of genetics*. Oxford University Press.

Kmeťová, M. (2010). On history of information visualization.

Kopp (2011). Human-Computer Interaction – Session 7: User Interface Evaluation [PDF]. Retrieved from https://www.techfak.uni-bielefeld.de/~skopp/Lehre/MMI_WS1112/MMI_7.pdf (26.09.2017)

Kosara, R., & Miksch, S. (2001). Metaphors of movement: a visualization and user interface for time-oriented, skeletal plans. *Artificial intelligence in medicine*, *22*(2), 111-131.

Lee, B., Riche, N. H., Karlson, A. K., & Carpendale, S. (2010). Sparkclouds: Visualizing trends in tag clouds. *IEEE transactions on visualization and computer graphics*, *16*(6), 1182-1189.

Legg, P. A., Chung, D. H., Parry, M. L., Jones, M. W., Long, R., Griffiths, I. W., & Chen, M. (2012, June). MatchPad: Interactive Glyph-Based Visualization for Real-Time Sports Performance Analysis. In *Computer Graphics Forum* (Vol. 31, No. 3pt4, pp. 1255-1264). Blackwell Publishing Ltd.

Linton, O., Maasoumi, E., & Whang, Y. J. (2005). Consistent testing for stochastic dominance under general sampling schemes. *The Review of Economic Studies*, *72*(3), 735-765.

Littig, B., & Menz, W. (2005). *Das Experteninterview*. A. Bogner (Ed.). VS, Verlag für Sozialwiss..

Mazza, R. (2004). Introduction to information visualisation. *Faculty of Communication Sciences, University of Lugano*.

Mintz, D., Fitz-Simons, T., & Wayland, M. (1997). Tracking air quality trends with SAS/GRAPH. In *Proceedings of the 22nd Annual SAS User Group International Conference (SUGI'97)* (pp. 807-812).

Nielsen, J. (1994, April). Usability inspection methods. In *Conference companion on Human factors in computing systems* (pp. 413-414). ACM.

Oelke, D., Janetzko, H., Simon, S., Neuhaus, K., & Keim, D. A. (2011, June). Visual Boosting in Pixel-based Visualizations. In *Computer Graphics Forum* (Vol. 30, No. 3, pp. 871-880). Blackwell Publishing Ltd.

O'Haver, T. (1997). A Pragmatic introduction to signal processing.

Osborne, M. J., & Rubinstein, A. (1994). *A course in game theory*. MIT press.

M. Page and A. V. Moere. Towards classifying visualization in team sports. In *Proceedings of the International Conference on Computer Graphics, Imaging and Visualisation*, CGIV '06, pages 24–29, 2006.

Perin, C., Vuillemot, R., & Fekete, J. D. (2013). SoccerStories: A kick-off for visual soccer analysis. *IEEE transactions on visualization and computer graphics*, *19*(12), 2506-2515.

Purchase, H. C., Andrienko, N., Jankun-Kelly, T. J., & Ward, M. (2008). Theoretical foundations of information visualization. In *Information Visualization* (pp. 46-64). Springer Berlin Heidelberg.

Reijner, H. (2008). The development of the horizon graph

Röber, N., Möller, T., Celler, A. M., & Strothotte, T. (2000, May). Multidimensional analysis and visualization software for dynamic SPECT. In *JOURNAL OF NUCLEAR MEDICINE* (Vol. 41, No. 5, pp. 191P-191P). 1850 SAMUEL MORSE DR, RESTON, VA 20190-5316 USA: SOC NUCLEAR MEDICINE INC.

Sacha, D., Stein, M., Schreck, T., Keim, D. A., & Deussen, O. (2014, October). Feature-driven visual analytics of soccer data. In *Visual Analytics Science and Technology (VAST), 2014 IEEE Conference on* (pp. 13-22). IEEE.

Saito, T., Miyamura, H. N., Yamamoto, M., Saito, H., Hoshiya, Y., & Kaseda, T. (2005, October). Two-tone pseudo coloring: Compact visualization for one-dimensional data. In *Information Visualization, 2005. INFOVIS 2005. IEEE Symposium on* (pp. 173-180). IEEE.

Shao, L., Sacha, D., Neldner, B., Stein, M., & Schreck, T. (2016). Visual-interactive search for soccer trajectories to identify interesting game situations. *Electronic Imaging*, *2016*(1), 1-10.

Shimabukuro, M. H., Flores, E. F., de Oliveira, M. C. F., & Levkowitz, H. (2004, July). Coordinated views to assist exploration of spatio-temporal data: A case study. In *Coordinated and Multiple Views in Exploratory Visualization, 2004. Proceedings. Second International Conference on* (pp. 107-117). IEEE.

Simmons, B. (2015, January 9). The tale of two Flaccos. Retrieved from http://grantland.com/the-triangle/the-tale-of-two-flaccos/

Stein, M., Janetzko, H., Breitkreutz, T., Seebacher, D., Schreck, T., Grossniklaus, M., ... & Keim, D. A. (2016). Director's Cut: Analysis and Annotation of Soccer Matches. *IEEE Computer Graphics and Applications*, *36*(5), 50-60.

Stein, M., Häußler, J., Jäckle, D., Janetzko, H., Schreck, T., & Keim, D. A. (2015). Visual soccer analytics: Understanding the characteristics of collective team movement based on feature-driven analysis and abstraction. *ISPRS International Journal of Geo-Information*, *4*(4), 2159-2184.

Stein, M., Janetzko, H., Seebacher, D., Jäger, A., Nagel, M., Hölsch, J., ... & Grossniklaus, M. (2017). How to Make Sense of Team Sport Data: From Acquisition to Data Modeling and Research Aspects. *Data*, *2*(1), 2.

Strode, G., Cross, L. (2013, January 13th). Using Color in Maps. Retriebed from https://visual.ly/community/infographic/geography/using-color-maps

Taki, T., & Hasegawa, J. I. (2000). Visualization of dominant region in team games and its application to teamwork analysis. In *Computer Graphics International, 2000. Proceedings* (pp. 227-235). IEEE.

Tominski, C., & Schumann, H. (2008, November). Enhanced interactive spiral display. In *SIGRAD 2008. The Annual SIGRAD Conference Special Theme: Interaction; November 27-28; 2008 Stockholm; Sweden* (No. 034, pp. 53-56). Linköping University Electronic Press.

Tufte, E. R. (1983). The Visual Display of. *Quantitative Information*.

Tufte, E. R. (2006). Beautiful evidence. *New York*.

Von Landesberger, T., Bremm, S., Andrienko, N., Andrienko, G., & Tekusova, M. (2012, October). Visual analytics methods for categoric spatio-temporal data. In *Visual Analytics Science and Technology (VAST), 2012 IEEE Conference on* (pp. 183-192). IEEE.

Ware, C. (2012). *Information visualization: perception for design*. Elsevier.

Wattenberg, M. (2002). Arc diagrams: Visualizing structure in strings. In *Information Visualization, 2002. INFOVIS 2002. IEEE Symposium on* (pp. 110-116). IEEE.

Weber, M., Alexa, M., & Müller, W. (2001, October). Visualizing time-series on spirals. In *Infovis* (Vol. 1, pp. 7-14).

Zhao, J., Forer, P., & Harvey, A. S. (2008). Activities, ringmaps and geovisualization of large human movement fields. *Information visualization*, *7*(3-4), 198-209.

Appendix

A.) Guideline for exploratory interview

Experteninterview – Fragen

Masterarbeit – Luke Albright – GIUZ – Universität Zürich – 01.02.2017

Einführung

Begrüssung

Unterschreiben der Einverständnis-Erklärung

- Nachfrage ob Audio-Aufzeichnung in Ordnung ist.

Vorstellung der Masterarbeit, des Kontextes, des Interviewleiters

- Orientierung der zu befragenden Person bezüglich der Thematik sowie der Relevanz des Gesprächs.

Vorstellung durch die befragte Person

- Formalitäten: Name, Rolle, Berufstitel, Trainer-Diplome / Ausbildungen.
- Jetzige Rolle im Verein, Erfahrungen im Fussball / im Vereinsleben etc.

Themenblock I: Arbeitsprozess als Fussballtrainer

A.) Leitfrage 1: Wie sieht eine typische Arbeitswoche für Sie als Trainer aus?

A1.) Wie gestalten Sie den Trainings- und Spielzyklus generell?

A2.) Mit welchen Personen / Rollen im Verein interagieren Sie?

- Sportchef, Co-Trainer, Vereinsfunktionäre, Vereinspräsidenten
- Inhalte dieser Interaktionen?
- In welchen Zusammensetzungen werden taktische oder spielrelevante Informationen ausgetauscht, besprochen und analysiert?

B.) Leitfrage 2: Wie gestalten Sie Ihr Training grundsätzlich?

B1.) Durchlaufen Sie ein typisches Training

B2.) Wie ist die Zusammensetzung zwischen spielerischen und physischen Elementen in der Vorbereitung?

B3.) Was für taktische Situationen trainieren Sie regelmässig?
- Spezifische Konstellationen / Laufwege?
- Gewisse Drucksituationen (Offensiv oder defensives Tackling?)

B4.) Wie zeigt sich bei Ihnen der Gegensatz zwischen individuellem (einzelne Spieler) und gesamtmannschaftlichem (gesamtes Team oder grössere Spielergruppen)?

B5.) Wie observieren Sie die Leistung im Training generell?
- Individuelle Spielerleistungen
- Team-Leistung

B6.) Wie funktioniert die Integration von Spiel-spezifischem Wissen ins Training?
- Wird wissen über den designierten Gegner eingebaut?
- Wie integrieren Sie gesammeltes Wissen aus dem vorherigen Spiel ins Training.

B7.) Handeln Sie generell agierend oder reagierend?

C.) Leitfrage 3: Wie sieht Ihre Rolle in einem Spiel aus?

C1.) Wie nehmen Sie Einfluss auf das Spiel?
- Impulsive Zwischenrufe vs. Langfristige Entwicklung des Spiels?
- Wie benutzen Sie die Halbzeitpause?
- Individuelle Spieler vs. Gesamtes Mannschaftsauftreten?

- Stellen Sie sich vor Ihr Team lässt leistungsmässig merklich nach in einem wichtigen Spiel – Wie reagieren Sie?

C2.) Inwiefern observieren Sie die Umsetzung von Dingen die zuvor im Training angeschaut wurden?

- Folgen von taktischen Richtlinien?
- Kommunikation/Körpersprache auf dem Platz?
- Einstudierte Spielzüge / Standards?

C3.) Wie viel Einfluss kann man in einem Spiel überhaupt nehmen?

- Wo sind die Grenzen der Beeinflussbarkeit?

Themenblock II: Spielanalyse durch Hilfstools

A.) Leitfrage 1: Wie ist Ihre generelle Einstellung zu Analysetools?

A1.) Finden Sie solche Tools sinnvoll?

- Für welche Sachverhalte können solche Tools brauchbar sein?
- Wo sehen Sie die Grenzen der Analysierbarkeit im Fussball?

A2.) Haben Sie selbst schon Erfahrungen mit solchen Analysetools gemacht?

- Selbst benutzt? Davon gehört? Im Fernsehen gesehen?
- Benutzen Sie Video als Analyse-Medium?
- Spezifische Software oder andere Programme die Sie kennen oder benutzen?

A3.) Wie analysieren Sie ein Spiel generell?

- Analyse-Workflow?
- Benutzen Sie spezifische Hilfsmittel hierzu?
- Wenden Sie gewisse Observations-Techniken an, um informative Erkenntnisse zu generieren.

A4.) Wie (und wann) kommunizieren Sie diese Erkenntnisse?

- Mit wem? Auf welche Art?

A5.) Auf einer Skala von 1-10 für wie relevant halten Sie die Spielanalyse mittels Hilfstools im Fussball?

B.) Leitfrage 2: Welche Spielsachverhalte empfinden Sie als zentral für eine Spielanalyse?

B1.) Was ist für Sie der Grundzweck von Analyse-Tools?
- Zur Exploration von neuen, bisher unentdeckten Informationen?
- Zur Leistungsüberwachung (Halten an Taktik / Körperlich)?
- Zur Kommunikation von Informationen an die Spieler etc.?

B2.) Wo würde Ihre Präferenz liegen: Analyse von individuellen Spieler-Leistungen oder Gesamtspiel-Muster?

B3.) Stellen Sie sich vor es ist Sonntag nach dem Spiel und Sie erhalten die aufbereiteten Informationen des gestrigen Spiels für die Analyse und Trainingsvorbereitung – Welche Informationen hätten Sie gerne aufgezeigt?
- Eigenes Team? Gegner?

B3.1) Stellen Sie sich vor Sie bereiten sich auf das nächste Spiel gegen Gegner XY vor – was würden Sie gerne über diesen Gegner wissen?

B4.) Würden Sie Ihre Taktik auf gewisse Analyse-Inputs anpassen?

C.) Leitfrage 3: Was für eine Einstellung haben Sie gegenüber geläufigen Analysekonzepten?

C1.) Bei meiner Recherche habe ich die folgenden Analytik-Kategorien erörtert – Kommentare? Sinnhaftigkeit?

a.) Heatmaps – Bewegungsverteilung

b.) Generelle Bewegungstrends / Bewegungsströme
   a. Trajektorien-Analysen

c.) Pass-Netzwerke

d.) Befolgung der taktischen Aufstellung?

e.) Aufzeigen von Freiräumen / Passräumen f.) Statistische Auswertungen (Zweikampfstärke/Ballbesitze/Laufwerte etc.)

g.) Event-Retrieval in Video-Aufnahmen

Themenblock III: Was ist Überlegenheit / Dominanz im Fussball

A.) Leitfrage 1: Was ist Dominanz im Fussball für Sie?

A1.) Was fällt Ihnen generell zu diesem Begriff ein?

A2.) Ist ein valides Ziel ein Spiel zu dominieren oder zu kontrollieren oder nicht so entscheidend?

A3.) Kategorisieren Sie den Fussball-Raum in spezifische Raumpartitionen, welche unterschiedliche Wichtigkeit haben?

- Rolle von Spielfeldmitte vs. Aussenbahnen.
- Abschluss-Positionen ähnlich zum Slot (Eishockey) Sweet Spots (Basketball)
- Rolle des Strafraums

A4.) Ist Dominanz nur ein offensives oder auch ein defensives Phänomen?

- Angriff vs. Pressing etc.?

A5.) Bei einem Angriff möchte man ja eine numerische Überlegenheit in Ball-Nähe erzeugen aber gleichzeitig den Raum breit machen - Ist es möglich einen Grenzwert zu finden, womit man bestimmen kann wie viele Spieler nahe am Ball und wie viele weiter Weg sein müssen?

- Erklärungsversuch
- Unterteilung aktive vs. Periphere Spieler möglich?

B.) Leitfrage 2: Die folgenden Faktoren / Ansätze werden oftmals mit fussballerischen Überlegenheit in Verbindung gebracht. Können Sie diese in Ihrer Eigenart, Relevanz und Sinnhaftigkeit kommentieren? Bewertung der Relevanz für Dominanz auf einer Skala von 1-10.

a.) Ballbesitz

b.) Packing-Werte / Überspielung des Gegners

c.) Zweikampf-Stärke

d.) Physische Komponenten (Laufwerte etc.

e.) Freiheit im Spiel (Quantifizierung von Freiräumen/beherrschter Radius der eigenen Spieler)
   a. Voronoi-Analysen

f.) Anzahl anspielbarer Mitspieler

g.) Relative Positionen der Spieler zum Ball (weit/eng)
   a. Kompaktheit / Convex Hull

h.) Relative Positionen der Spieler zum Tor

i.) Überzahl-Regionen
   a. Dominant Regions

B2.) Abstrakte Faktoren / Ansätze

a.) Mittelpunkt der 10 Spieler

b.) Mittelpunkt gewisser Spielergruppen (Zusammensetzungen)

c.) Ausdehnung zwischen hinterstem und vorderstem Spieler

d.) Mittelpunkt einer spezifischen Position über den zeitlichen Verlauf hinaus.

e.) Aktionsradius eines gewissen Spielers (Personal Heatmap)

C.) Leitfrage 3: Wie attraktiv wäre das Konzept einer Dominanz-Übersicht?

C1.) Was für Erkenntnisse könnte man daraus ableiten?
- Attraktiv um gewisse Bruchpunkte im Spiel zu erkennen?

- Auswechslungszeitpunkte bestimmen?
- Trends über mehrere Spiele beobachten?

C2.) Könnten Sie sich vorstellen eine Dominanz-Übersicht als Kommunikationsmittel zu benutzen?
- Zu welchem Zweck?
- Zu welchem Zeitpunkt im Coaching-Prozess

Themenblock IV: Inputs für die eigene Dominanz-Visualisierung

A.) Leitfrage 1: Wie sehen Sie den Faktor Zeit im Fussball?

A1.) Wie würden Sie ein Fussballspiel partitionieren?
- Spezifische Zeitintervalle / Halbzeiten / Gesamtspieltrends
- Erkennung von spezifischen Einzelevents

A2.) Was ist für Sie wichtiger bei einer analytischen Visualisierung im Fussball? Raum oder Zeit?
- Zeitauflösung / Raumauflösung

A3.) Ist eine Zeitstrahlansicht ein attraktives Visualisierungstool?
- Alternativen?

A4.) Wünschen Sie die Daten in Real-Time oder Statisch nach dem Spiel?

B.) Leitfrage 2: Visualisierung von Dominanz: Sinnvoll?

B1.) Wie sinnvoll wäre eine Dominanzvisualisierung?
- Allgemein kommentiert sowie auf einer Skala von 1-10

B2.) Wie kann man verschiedene Dominanzgrössen am besten aufzeigen?
- Farbliche Variation
- Variation der Grösse

B3.) Welcher Detailgrad wäre bei einer Dominanz-Visualisierung sinnvoll?
- Nur Zeitstrahl-Gesamtübersicht über das Spiel
- Zeitstrahl-Übersicht + zusätzliche detailierte Informationen
    - An spezifischen Eventstellen oder Überall?
- Interaktionsgrad mit Zeitstrahl

C.) Leitfrage 3: Spezifische graphische Wünsche an eine solche Dominanz-Zeitstrahl-Übersicht?

C1.) Wie soll der Zeitstrahl farblich gestaltet werden?

C2.) Auf welchen Geräten einsehbar?

C3.) Zusätzliche spontane Wünsche

Abschluss

Frage bezüglich weiterer Zusammenarbeit

- Iterative Begleitung der Entwicklung
- Qualitative Analyse

Beendung des Gesprächs sowie der Audio-Aufnahme

Bedankung und Verabschiedung

B.) Guideline for evaluation

Evaluationsinterview / Leitfaden

A.) Cognitive Walkthrough

Fragen zum allgemeinen Verständnis eines Dominanz-Zeitstrahls

Für jeden Visualisierungstyp (Pixel-Chart, Two-Tone Pseudo Coloring, Line-Chart) werden Ihnen Dominanz-Zeitstrahle auf drei verschiedenen visuellen Abstraktionsstufen (Werteglättungsparameter = 0.33, 0.66, 1.0) gezeigt. Beantworten Sie die folgenden Fragen basierende auf den Ihnen gezeigten Dominanz-Zeitstrahlen für eine einzelne Mannschaft. Erläutern Sie bitte simultan Ihre Eindrücke sowie Ihr Vorgehen:

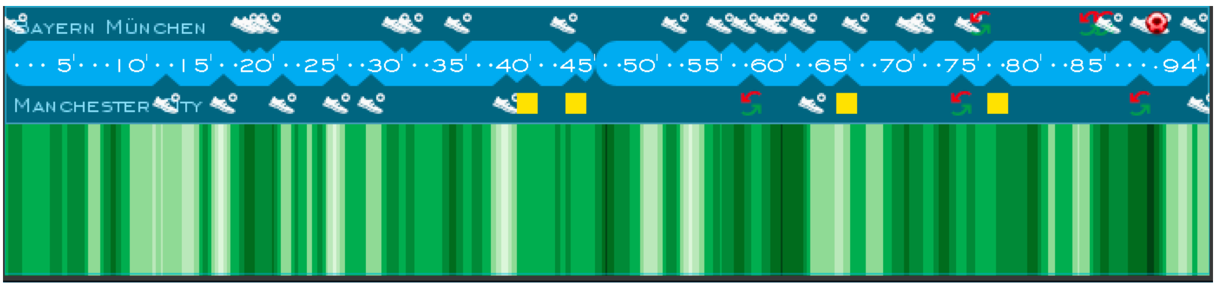
1.) Wie viele Phasen mit tiefen Dominanz-Werten (unter 0.2) sowie mit hohen Dominanz-Werten (über 0.8) können Sie in der Dominanz-Visualisierung erkennen? Wie gehen Sie vor? Wieso? Gibt es Schwierigkeiten? Wenn ja, welche? Gibt es Unterschiede zwischen der wenig und der stark abstrahierten Visualisierung? Wenn ja, welche?

2.) Markieren Sie die dominanteste Phase des Spiels für die designierte Mannschaft? Wie erkennen Sie dies? Wieso? Gibt es Schwierigkeiten? Wenn ja, welche? Gibt es Unterschiede zwischen der wenig und der stark abstrahierten Visualisierung? Wenn ja, welche? (*Markieren mit Rosa Marker unterhalb des Zeitstrahls*).

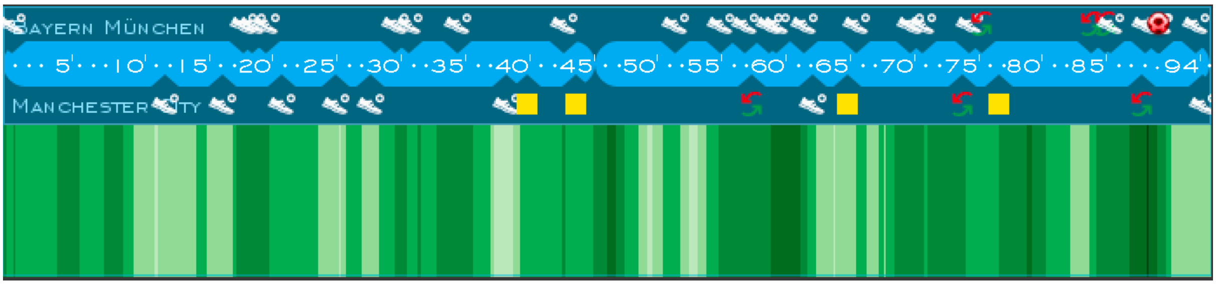
3.) Welche Halbzeit zeigt die höhere Dominanz für die designierte Mannschaft in der Visualisierung? Wie erkennen Sie dies? Wieso? Gibt es Schwierigkeiten? Wenn ja, welche? Gibt es Unterschiede zwischen der wenig und der stark abstrahierten Visualisierung? Wenn ja, welche?

A.) Pixel-basierte Visualisierung (Pixel-Chart)

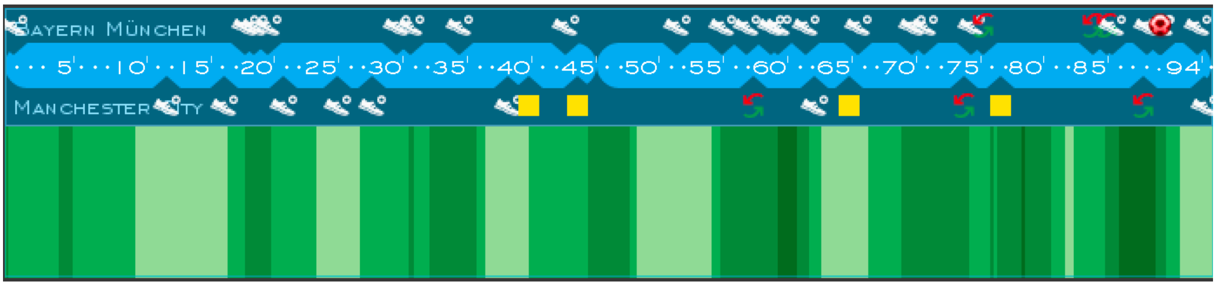
Abstraktionsstufe auf Werteglättungsregler = 0.33



Abstraktionsstufe auf Werteglättungsregler = 0.66

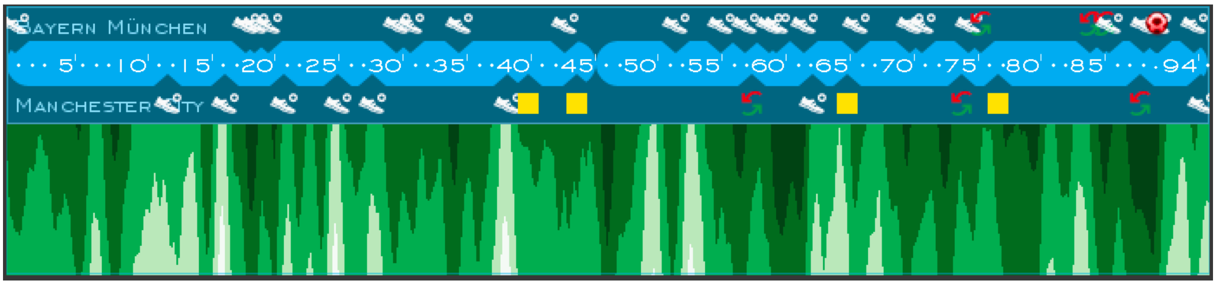


Abstraktionsstufe auf Werteglättungsregler = 1.0

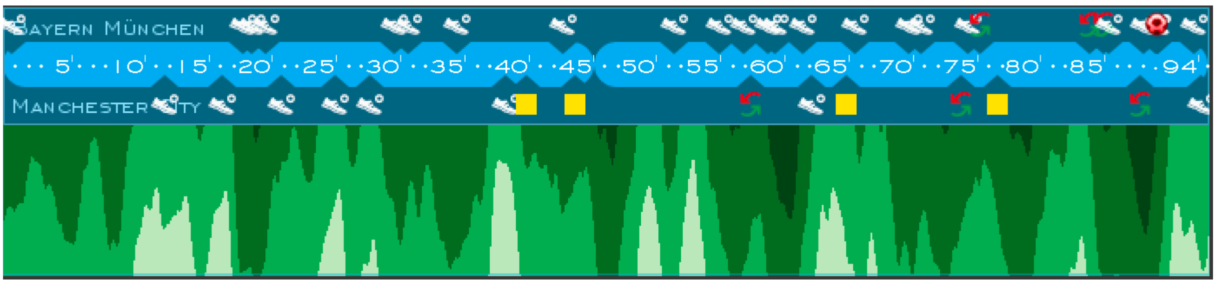


<u>B.) Zwei-Ton Pseudoeinfärbung (Two-Tone Pseudo Coloring)</u>

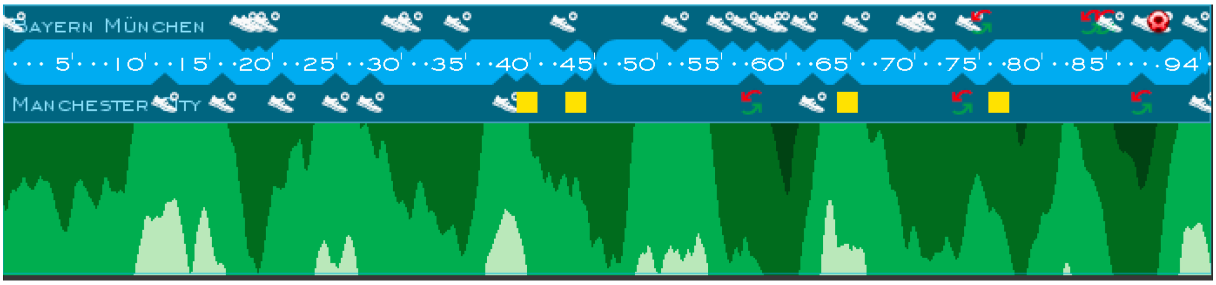
Abstraktionsstufe auf Werteglättungsregler = 0.33



Abstraktionsstufe auf Werteglättungsregler = 0.66

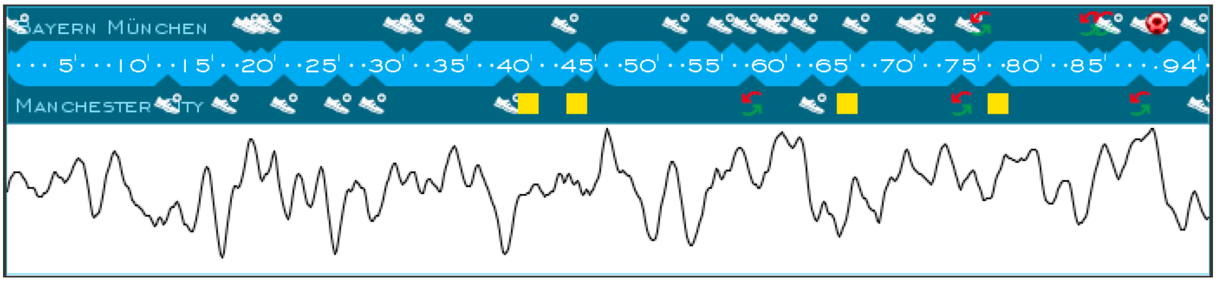


Abstraktionsstufe auf Wertеglättungsregler = 1.0

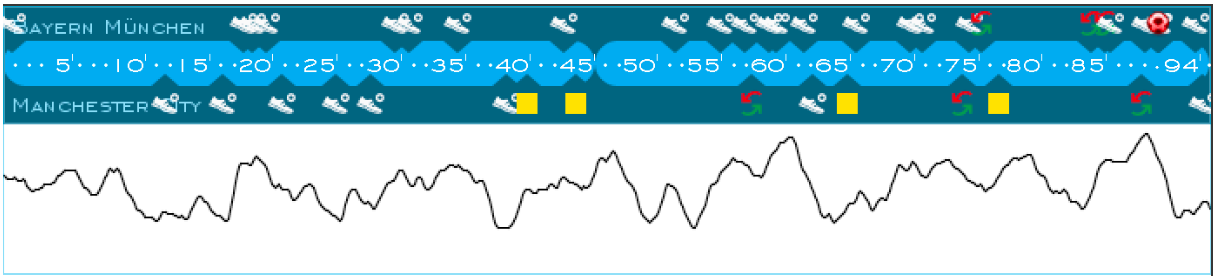


C.) Linien-basierte Visualisierung (Line-Chart)

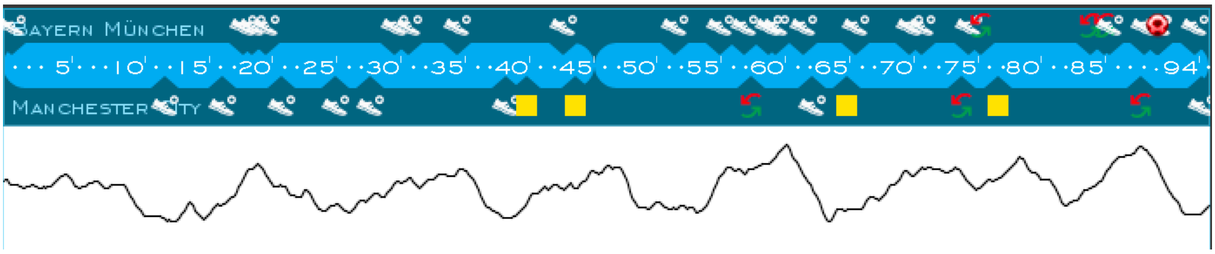
Abstraktionsstufe auf Werteglättungsregler = 0.33



Abstraktionsstufe auf Werteglättungsregler = 0.66



Abstraktionsstufe auf Werteglättungsregler = 1.0



Fragen zur Vergleichsansicht mit individuellen Dominanz-Zeitstrahlen für beide Teams (Two-Scale Representation)

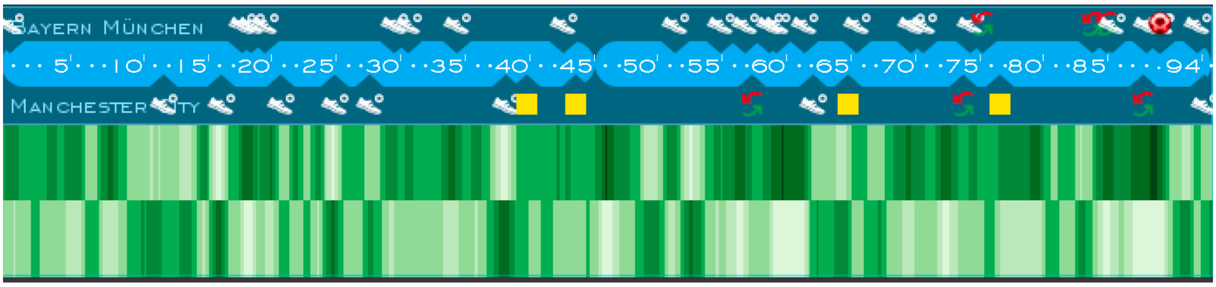
Für jeden Visualisierungstyp (Pixel-Chart, Two-Tone Pseudo Coloring, Line-Chart) werden Ihnen zwei Dominanz-Zeitstrahle auf drei verschiedenen visuellen Abstraktionsstufen (Werteglättungsparameter = 0.33, 0.66, 1.0) gezeigt. Beantworten Sie die folgenden Fragen basierend auf den Ihnen gezeigten Dominanz-Zeitstrahlen für zwei Mannschaften simultan. Erläutern Sie bitte simultan Ihre Eindrücke sowie Ihr Vorgehen:

1.) Welche Mannschaft (Heim oder Auswärts) war über den Verlauf des Spiels dominanter laut der Visualisierungen? Wie erkennen Sie dies/Wie gehen Sie vor? Gibt es Schwierigkeiten? Wenn ja, welche? Gibt es Unterschiede zwischen der wenig und der stark abstrahierten Visualisierung? Wenn ja, welche?

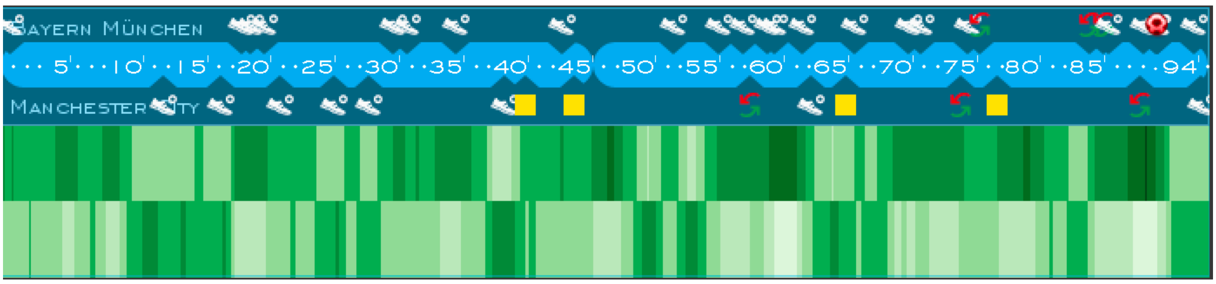
2.) In welcher Phase ist die Diskrepanz der Dominanz-Werte zwischen den Teams am grössten? Wie erkennen Sie dies/Wie gehen Sie vor? Gibt es Schwierigkeiten? Wenn ja, welche? Gibt es Unterschiede zwischen der wenig und der stark abstrahierten Visualisierung? Wenn ja, welche? ***(Markieren mit Rosa Marker)***

A.) Pixel-basierte Visualisierung (Pixel-Chart)

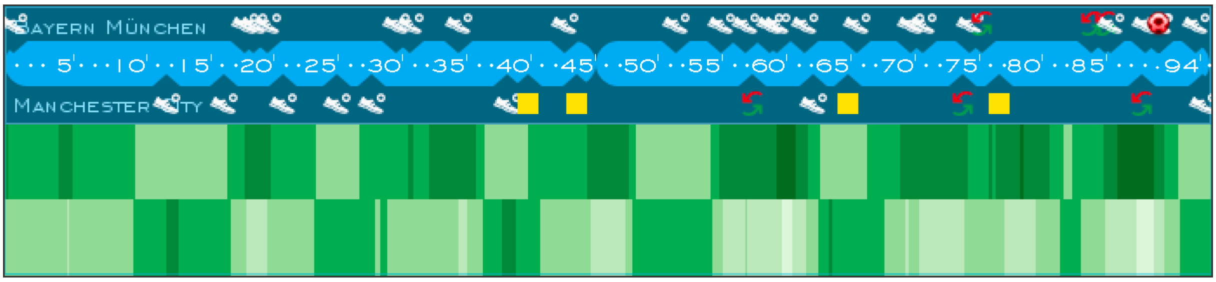
Abstraktionsstufe auf Werteglättungsregler = 0.33



Abstraktionsstufe auf Werteglättungsregler = 0.66

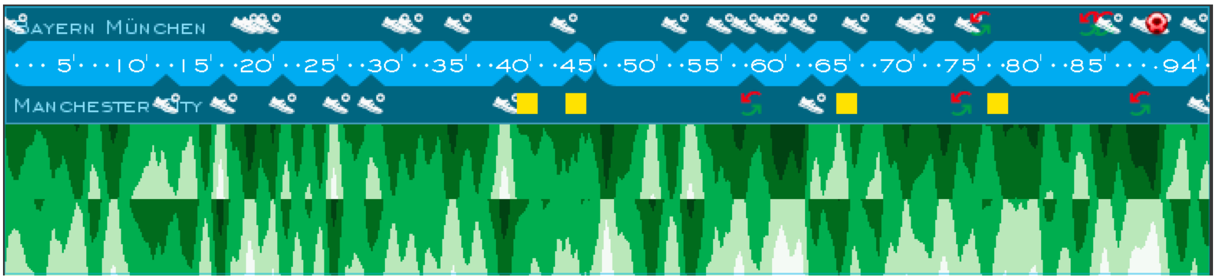


Abstraktionsstufe auf Werteglättungsregler = 1.0



B.) Zwei-Ton Pseudoeinfärbung (Two-Tone Pseudo Coloring)

Abstraktionsstufe auf Werteglättungsregler = 0.33

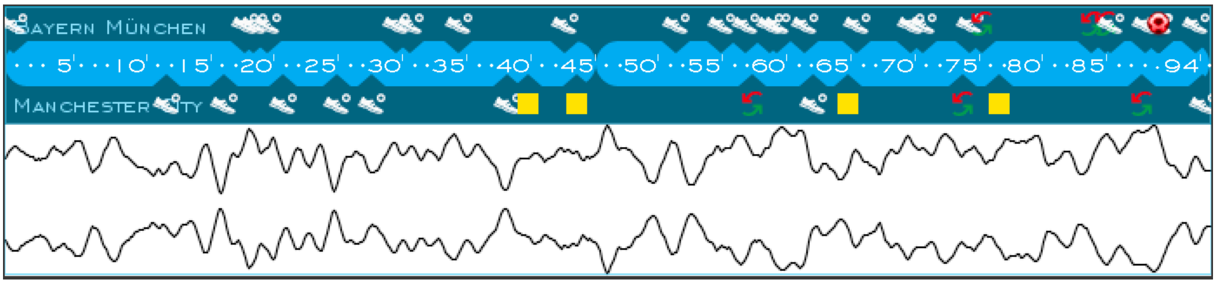


Abstraktionsstufe auf Werteglättungsregler = 0.66

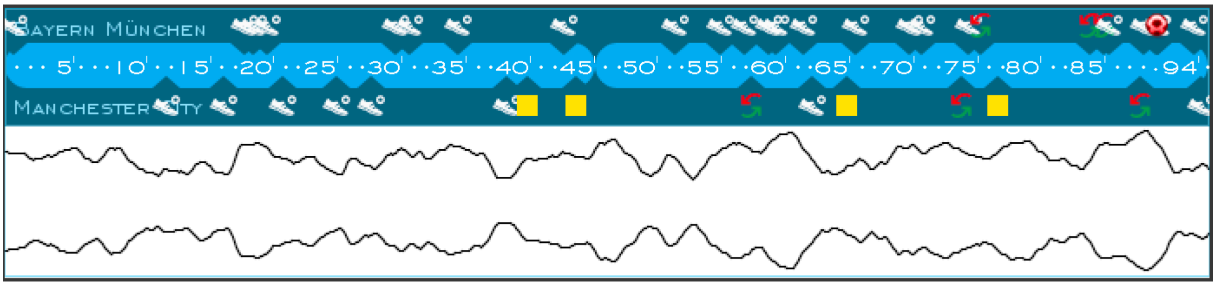
Abstraktionsstufe auf Werteglättungsregler = 1.0

C.) Linien-basierte Visualisierung (Line-Chart)

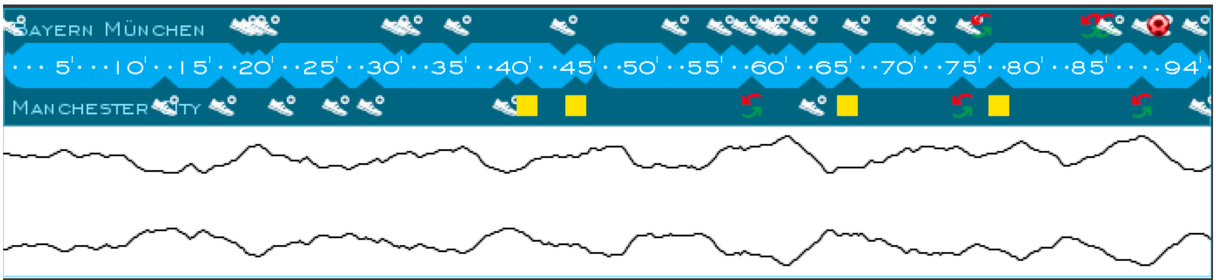
Abstraktionsstufe auf Werteglättungsregler = 0.33



Abstraktionsstufe auf Werteglättungsregler = 0.66



Abstraktionsstufe auf Werteglättungsregler = 1.0



Fragen zur Vergleichsansicht mit vereinheitlichtem Dominanz-Zeitstrahl für beide Teams (Single-Scale Representation)

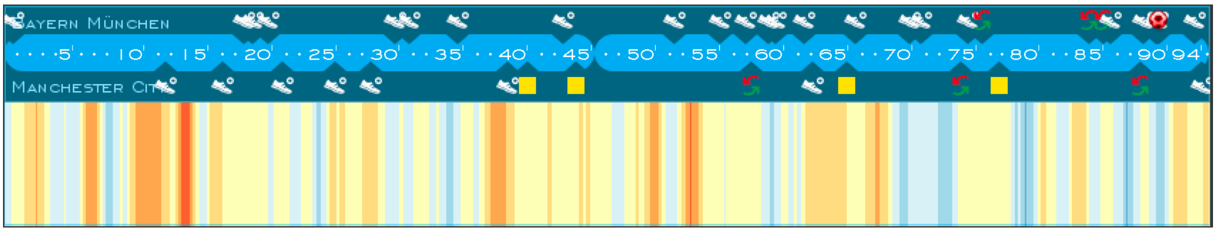
Für jeden Visualisierungstyp (Pixel-Chart, Two-Tone Pseudo Coloring, Line-Chart) werden Ihnen ein Dominanz-Zeitstrahl auf drei verschiedenen visuellen Abstraktionsstufen (Werteglättungsparameter = 0.33, 0.66, 1.0) gezeigt. Beantworten Sie die folgenden Fragen basierend auf den Ihnen gezeigten Dominanz-Zeitstrahlen für zwei Mannschaften simultan. Erläutern Sie bitte simultan Ihre Eindrücke sowie Ihr Vorgehen:

1.) Welche Mannschaft (Heim oder Auswärts) war über den Verlauf des Spiels dominanter laut der Visualisierungen? Wie erkennen Sie dies/Wie gehen Sie vor? Gibt es Schwierigkeiten? Wenn ja, welche? Gibt es Unterschiede zwischen der wenig und der stark abstrahierten Visualisierung? Wenn ja, welche?

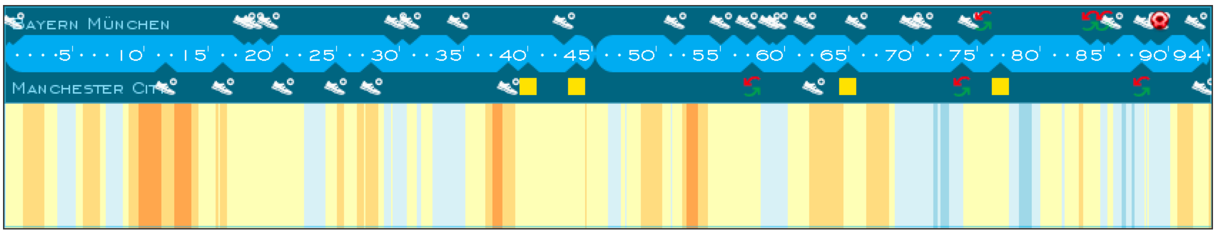
2.) In welcher Phase ist die Diskrepanz der Dominanz-Werte zwischen den Teams am grössten? Wie erkennen Sie dies/Wie gehen Sie vor? Gibt es Schwierigkeiten? Wenn ja, welche? Gibt es Unterschiede zwischen der wenig und der stark abstrahierten Visualisierung? Wenn ja, welche? *(Markieren mit Rosa Marker)*

A.) Pixel-basierte Visualisierung (Pixel-Chart)

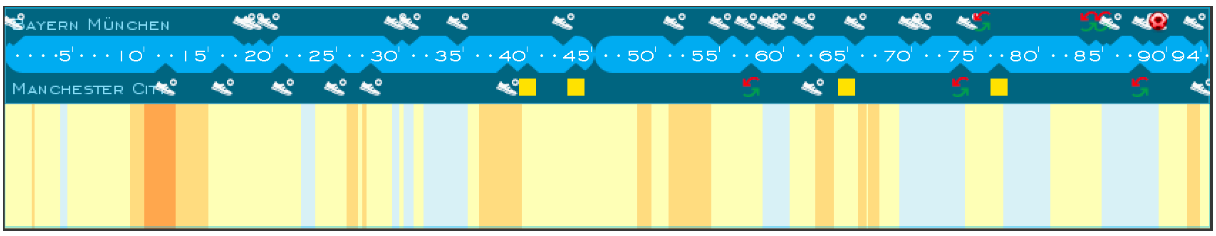
Abstraktionsstufe auf Werteglättungsregler = 0.33



Abstraktionsstufe auf Werteglättungsregler = 0.66



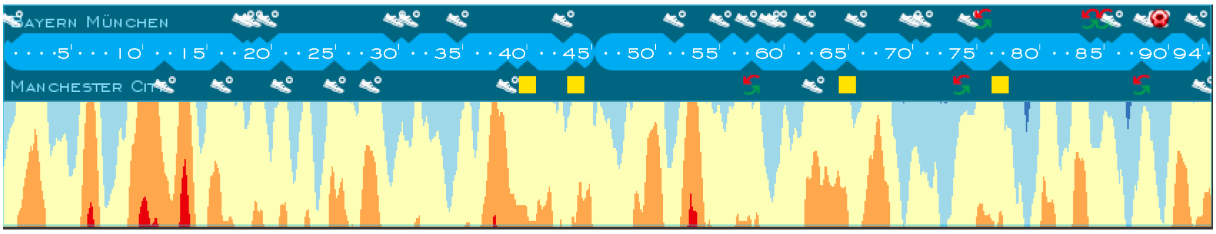
Abstraktionsstufe auf Werteglättungsregler = 1.0



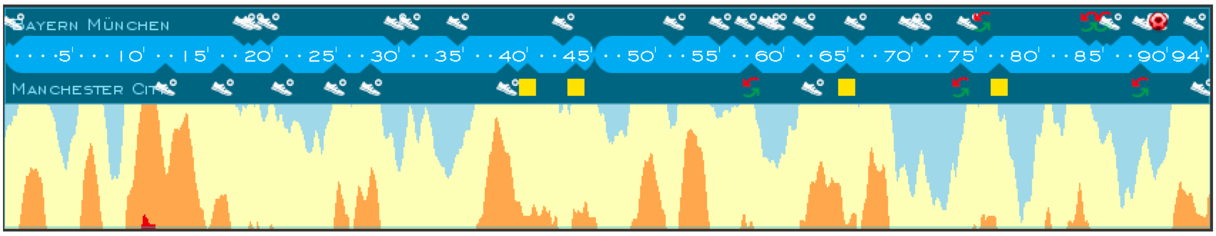
T

B.) Zwei-Ton Pseudoeinfärbung (Two-Tone Pseudo Coloring)

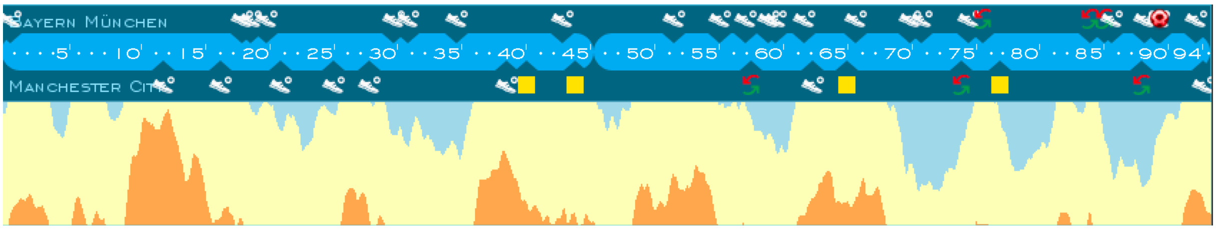
Abstraktionsstufe auf Werteglättungsregler = 0.33



Abstraktionsstufe auf Werteglättungsregler = 0.66

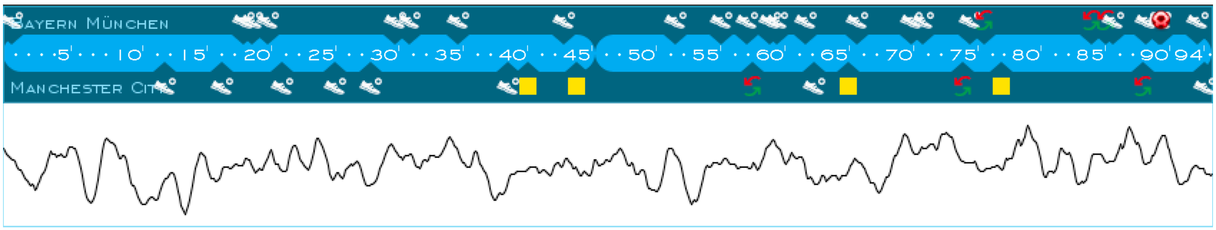


Abstraktionsstufe auf Werteglättungsregler = 1.0

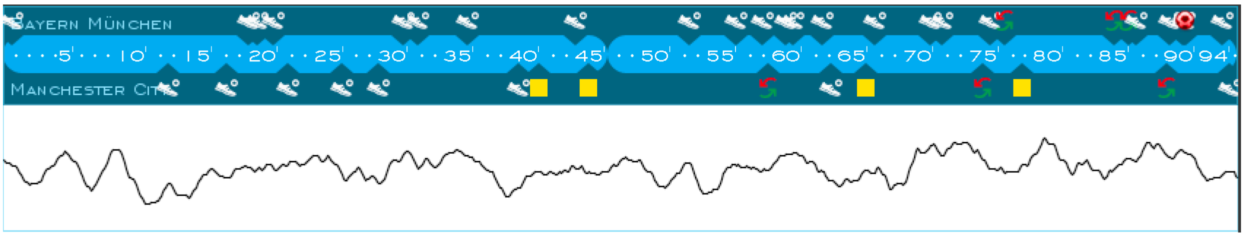


C.) Linien-basierte Visualisierung (Line-Chart)

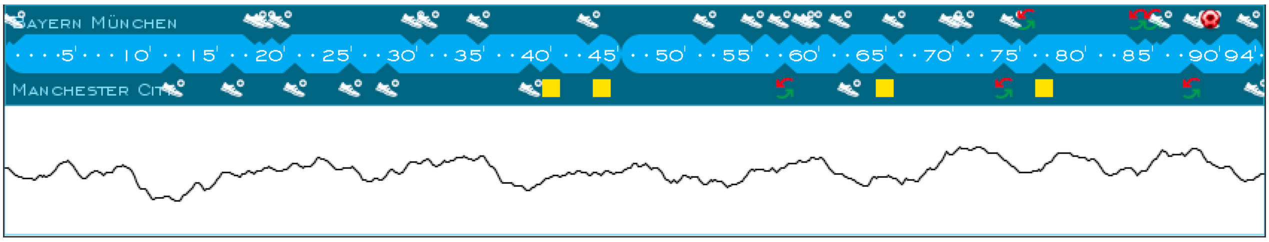
Abstraktionsstufe auf Werteglättungsregler = 0.33



Abstraktionsstufe auf Werteglättungsregler = 0.66



Abstraktionsstufe auf Werteglättungsregler = 1.0



Allgemeine Nachbetrachtungsfragen zu Vergleichsansichten

Beantworten Sie die nachfolgenden Fragen bezüglich der beiden Vergleichsansichten der Dominanz-Zeitstrahle:

1.) Welche Visualisierungsform ist für den Vergleich am sinnvollsten/nützlichsten? Weshalb?

a. **Pixel-basierte Visualisierung**
b. **Zwei-Ton Pseudoeinfärbung**
c. **Linien-basierte Visualisierung**

2.) Welche Visualisierungsansicht ist für den Vergleich der Dominanzwerte der beiden Teams am sinnvollsten/nützlichsten? Weshalb?

a. **Darstellung mit zwei übereinander angeordneten Zeitreihen für die beiden Teams (eigene Zeitreihe für jede Mannschaft).**
b. **Darstellung mit einer einzelnen Zeitreihe für beide Teams zusammen.**

Aufgabe zum Werteglättungsregler (steuert Abstraktionslevel)

1.) Der Werte-Glättungsschieberegler abstrahiert Daten-Werte visuell und vereinfacht Ihre Darstellung. Definieren Sie die Ober- und Untergrenze des für Sie sinnvollen Abstraktionsbereichs sowie die ideale Abstraktionsstufe für: (Wieso entscheiden Sie sich so?)

- **a.) Pixel-basierte Visualisierung (Pixel Chart)**
- **b.) Zwei-Ton Pseudoeinfärbung (Two-Tone Pseudocoloring)**
- **c.) Linien-basierte Visualisierung Line-Chart**

2.) Qualitative Bewertung der Dominanz-Visualisierungen

Beantworten Sie die folgenden Fragen zur potenziellen Verwendung der zuvor aufgezeigten Dominanz-Zeitstrahl-Visualisierungen:

1.) Wie gut/sinnvoll finden Sie die Pixel-Chart Visualisierung?

- Lesbarkeit/Verständlichkeit/Komplexität (Positives/Negatives)
- Farbwahl
- Einsatzfähigkeit

2.) Wie gut/sinnvoll finden Sie die Two-Tone Pseudocoloring Visualisierung?

- Lesbarkeit/Verständlichkeit/Komplexität (Positives/Negatives)
- Farbwahl
- Einsatzfähigkeit

3.) Wie gut/sinnvoll finden Sie die Line-Chart Visualisierung?

- Lesbarkeit/Verständlichkeit/Komplexität (Positives/Negatives)
- Farbwahl
- Einsatzfähigkeit

4.) Welche dieser Visualisierung finden Sie am besten, welche am schlechtesten? Weshalb entscheiden Sie sich so?

- Einsatzfähigkeit/Verständlichkeit etc.

5.) Wie wichtig ist die Legende zum Verständnis der Visualisierungen?

- Lesbarkeit/Verständlichkeit/Komplexität (Positives/Negatives)
- Welcher informative Wert besitzt die Legende?
- Wie verwenden Sie die Legende im Zusammenhang mit den Visualisierungen.

6.) Wie ist Ihre Meinung zur Werte-Glättungsfunktion?

- Ist es verständlich was der Werteglättungs-Regler macht?
- Abstraktion der Datenvisualisierung wichtig/nützlich/verständlich?

3.) Generelle qualitative Nutzerbefragung – Verwendung der Dominanz-Visualisierung im Fussball-Kontext

1.) Wie wirksam/sinnvoll finden Sie eine Dominanz-Visualisierung allgemein?

- Allgemeine Kommentare
- Positives/Negatives
- Sinnhaftigkeit
- Potenzial/Limitationen

2.) An welcher Stelle im Training-Spiel-Zyklus würde der Einsatz einer Dominanz-Visualisierung am meisten Sinn machen?

- Während der Spielvorbereitung? (Im Trainingskontext mit Spielern oder als Input für die Spielanalyse des Coachs/Analysten ohne Spieler?)
- Während dem Spiel selbst? (Input während Halbzeit/auf der Bank?)

3a.) Wie sinnvoll ist die Verwendung einer Dominanz-Visualisierung als visuelles Kommunikationsmittel?

- Ist ein Informationstransfer von Coach an Spieler möglich?
- Ist ein interner Informationstransfer zwischen Coaches/Analysten etc. möglich?

3b.) Wie sinnvoll ist eine Dominanz-Visualisierung als Indikator für spezifische Spiel-Phasen, welche eine vertiefte Betrachtung benötigen?

- Sind ausserordentlich hohe bzw. tiefe Dominanzphasen als interessante Anschauungsphasen zu verstehen?
- Können z.B. Hoch-Dominanzphasen als Indikatoren für angriffsbasierte Szenen mit hoher Torgefahr gesehen werden, welche für den Trainer von erweiterter Bedeutung sind, sodass er diese Phasen nochmals genauer betrachten möchte?

3c.) Wie sinnvoll ist eine Dominanz-Visualisierung für die Leistungsüberprüfung einer Mannschaft?

- Könnte man eine Dominanz-Visualisierung dafür verwenden, um die Folgen von taktischen Spielanweisungen zu observieren?
- Z.B. wenn der Coach ab der 65 Minute neue Spieler einwechselt und mehr Druck auf das gegnerische Tor fordert – Kann die Umsetzung dieser Anweisung mittels der Dominanz-Visualisierung erkannt werden?

3d.) Wie sinnvoll ist eine Dominanz-Visualisierung für die Spielanalyse der eigenen bzw. gegnerischen Mannschaft?

- Können Dominanz-Muster im eigenen/gegnerischen Spiel erkannt werden, welche einen gewissen informativen Wert in der Spielanalyse haben? (z.B. Gegner in den ersten 20 Minuten drückend überlegen; Heimteam mit vielen intensiven Angriffen in der ersten Halbzeit etc.)
- Kann die Dominanz-Visualisierung benutzt werden, um beide Teams mit einander zu vergleichen? (Inwiefern ist dies möglich?).
- Dominanz-Muster für ein Spiel vs. Dominanz-Muster für eine gesamte Saison (bzw. längerer Betrachtungszeitrahmen).

4.) Würden Sie aufgrund der Dominanz-Visualisierung taktische Spielanweisungen geben? Falls ja: Wie, wann und weshalb?

- Taktische Spielanweisungen einfliessen lassen (z.B. offensiver stehen, mehr Pressing, Rückzug in defensive, abwartende Formation etc.)
- Aufstellung ändern / Auswechslung vornehmen

5a.) Bewerten Sie den verwendeten Dominanz-Faktor *Numerische Überzahl (Numeric Predominance).*

- Sinnhaftigkeit
- Positives/Negatives
- Potenzial zur Weiterentwicklung

5b.) Bewerten Sie den verwendeten Dominanz-Faktor *Zonen-basierte Proximität zum gegnerischen Torraum (Zonal Proximity).*

- Sinnhaftigkeit
- Positives/Negatives
- Potenzial zur Weiterentwicklung

5c.) Bewerten Sie den verwendeten Dominanz-Faktor *Distanz-basierte Proximität zum gegnerischen Torraum (Distance Proximity).*

- Sinnhaftigkeit
- Positives/Negatives
- Potenzial zur Weiterentwicklung

6.) Wie ist Ihre Meinung zu Zeit-orientierten Visualisierungsansichten?

- Zeitstrahl bzw. zeit-basierte Darstellungen als informative Visualisierungen? (Positives/Negatives/Potenzial)
- Wo würde Ihre Präferenz liegen – zeit- oder raum-basierte Visualisierungen? (z.B. Zeitstrahl vs. Ansicht des Fussballfeldes)

7.) Wie ist Ihre Meinung zu Faktor-basierten Analyse-Tools im Fussball?

- Kann man ein Spiel-Phänomen bzw. Spiel-Muster mit verschiedenen Faktoren korrekt abbilden (z.B. Dominanz aufbrechen in verschiedene Faktoren unterteilt).
- Wäre es interessant solche Faktoren selbst definieren zu können?
- Aggregation mehrerer Faktoren zu einer Gesamt-Dominanz-Übersicht?

8.) Die Werteglättungsfunktion erlaubt eine Glättung der Werte und dient somit einer erhöhten visuellen Abstraktion – Ist dies sinnvoll/verständlich?

- Wie abstrahiert/detailliert sollen Datenwerte visuell ersichtlich sein?
- Ist visuelle Abstraktion nützlich/unbrauchbar?
- Z.B. jeder Angriff ersichtlich vs. 4-5 grobe Dominanz-Phasen?

Personal declaration

I hereby declare that the submitted thesis is the result of my own, independent work. All external sources are explicitly acknowledged in the thesis.

Männedorf, 28.09.2017

Luke Albright