Visualization of Football Data

**Rasmus Bo Adeltoft, Sebastian Seneca Haulund Hansen,
Steffen Berg Klenow, Christian Bjørn Moeslund,
Andreas Staurby Olesen, Henrik Sejer Pedersen**

Department of Mathematics and Computer Science, University of Southern Denmark

Introduction

In this project, we have worked with visualization of football data. In general, visualizations can help humans' understanding of large datasets, as the data can be summarized very effectively, and patterns can quickly be recognized. The data at our disposal was provided by Prozone. Prozone is a company that specializes in collecting and visualizing football data. Our visualizations were designed to compare top-tier teams with low-tier teams, and to explore how a team evolves throughout a season as well as a match.

Visualization Design

Marks and Channels Marks are the individual elements in a graph, while channels define the appearance of these. Marks could be points in a scatterplot, and the channels could be shape or positions on common scales.

What-why-how Visualizations are made by following the what-why-how principle. *What* is comprised of the semantics of the data as well as the type. *Why* includes questions such as "Why make the visualization?" and "Should it consume or produce data?". *How* consists of how to design the visualization, as well as how to create it.

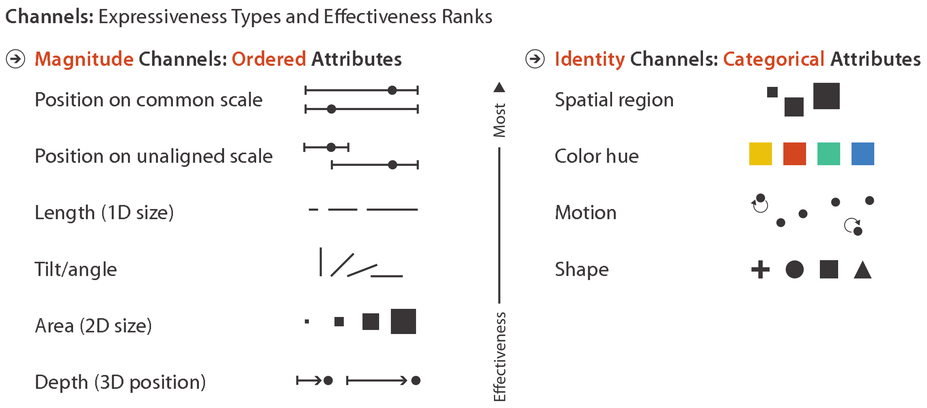

Figure 1: *Effectiveness of different channels. Source: Munzner, T. and Maguire, E. (2015) "Visualization Analysis and Design", p.115*

Technology and Data

To make the visualizations, we have used two kinds of technology. We have used *R*, which is a programming language for statistical computing, and we have used *D3*, which is a *JavaScript* library for making visualizations.

R is used to manipulate and clean the football data to make it ready for dynamic visualizations, but it is also used to make static ones. *D3* is used to make dynamic visualizations based on the data we have cleaned in *R*. Using *D3* allows us to make the visualizations interactive, and it allows users to view and interact with the visualizations in a web browser.

The data we have used in our visualizations have been acquired through an API provided by Prozone. We have implemented access to the API directly in *R*, making data acquisition and visualization generation automated.

Conclusion

With the visualizations, a few things became apparent. One of the more telling finds was that there is a tendency for teams to make more goal attempts as a match progresses. Furthermore, when a team loses to a lower ranked team, the higher ranked team fails more passes than usual, and tends to play more towards the middle of the field. These tendencies were found on the field-based visualizations, which proved to be efficient idioms for our purpose.

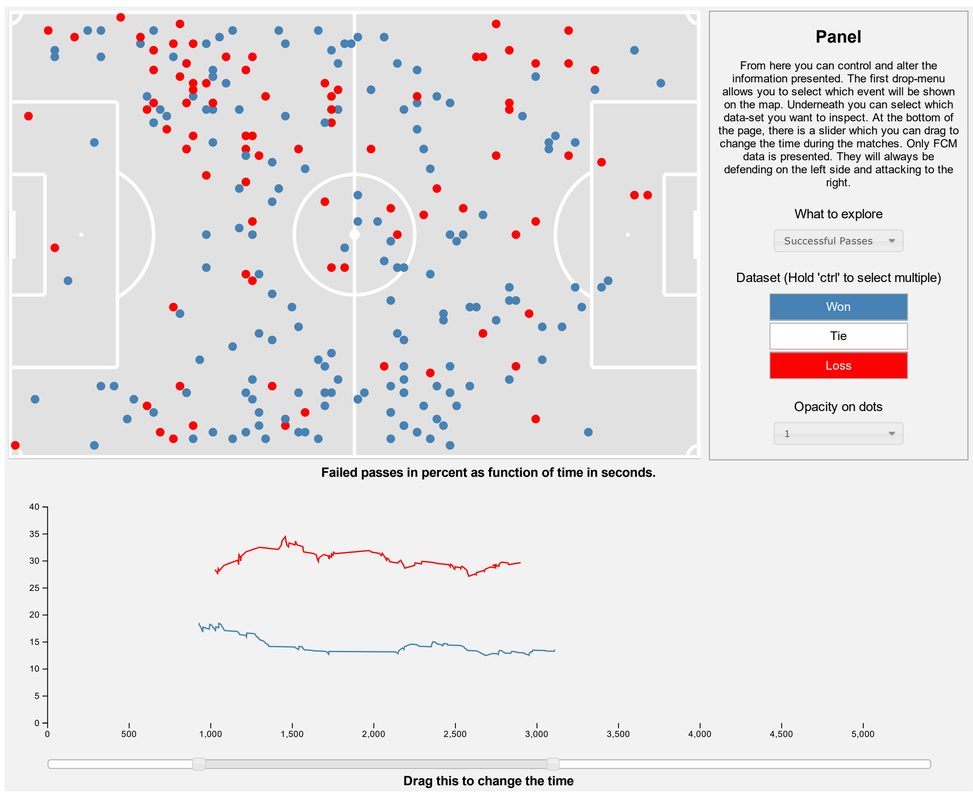

Figure 2: *Three matches played between FCM and Viborg, plotted from FCM's point of view. The color illustrates whether FCM lost (red) or won (blue).*

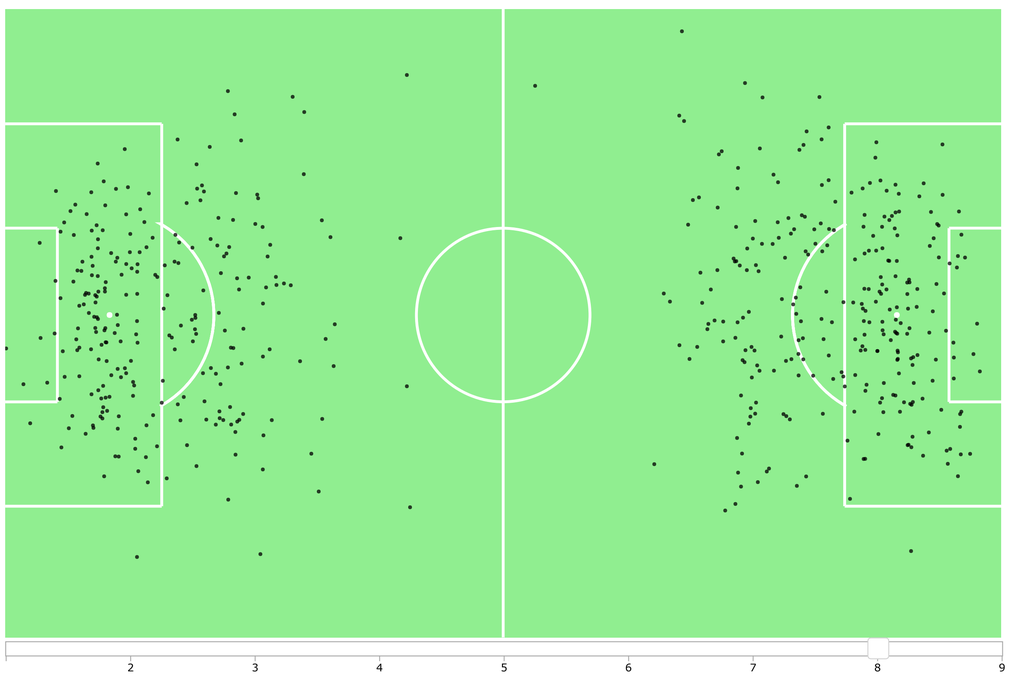

Figure 3: *The map shows clustering of certain events in an entire season. In this snapshot, goal attempts between 70 and 80 minutes are shown.*

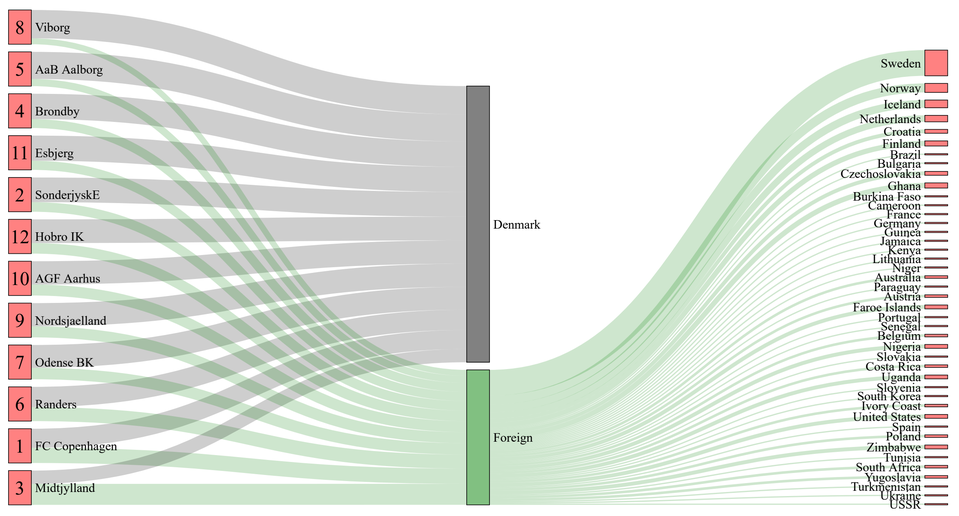

Figure 4: *A Sankey-Chart showing the distribution of players' nationality in the Danish Super League.*