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

We have worked with visualization of football data. Visualizations can help humans understand large datasets, as the data can be summarized very effectively, and patterns quickly identified. We worked with football data provided by Prozone, a company specialized in collecting and analyzing sport data.

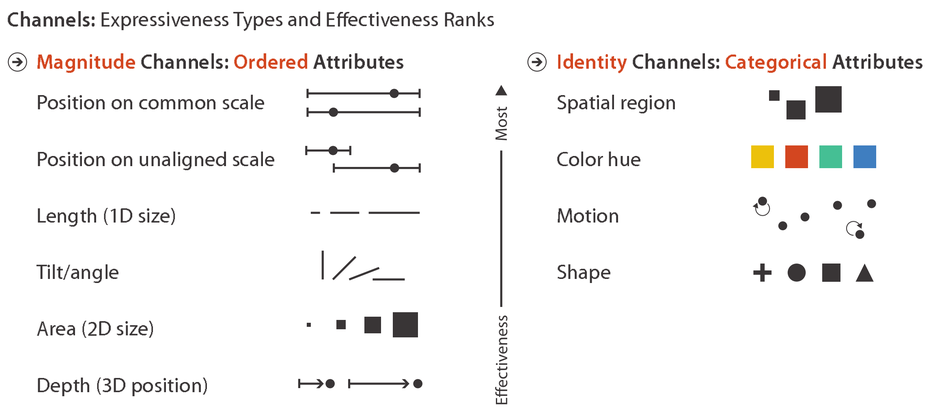
We set out to:

- Compare top-tier teams with low-tier teams.
- Examine how a team evolves throughout a match and a season.

Visualization Design

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?". *How* consists of how to design the visualization, as well as how to create it.

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.



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:

- *R*, a programming language for statistical computing.
- *D3*, a *JavaScript* library for making visualizations.

R is used to manipulate and clean data, preparing it for dynamic visualizations and making static ones. *D3* is used to make dynamic visualizations of the cleaned data, presented interactively in a web browser.

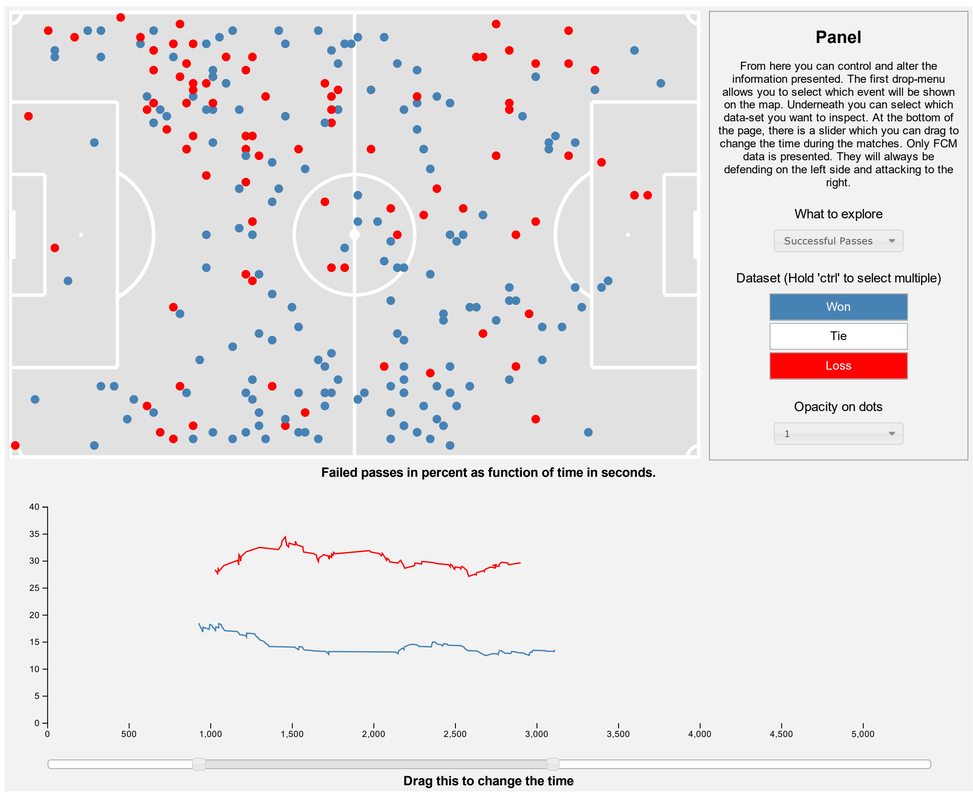
Our data has been acquired through Prozone's API. We implemented access to the API directly in *R*, making data acquisition and visualization generation automated.

Conclusion

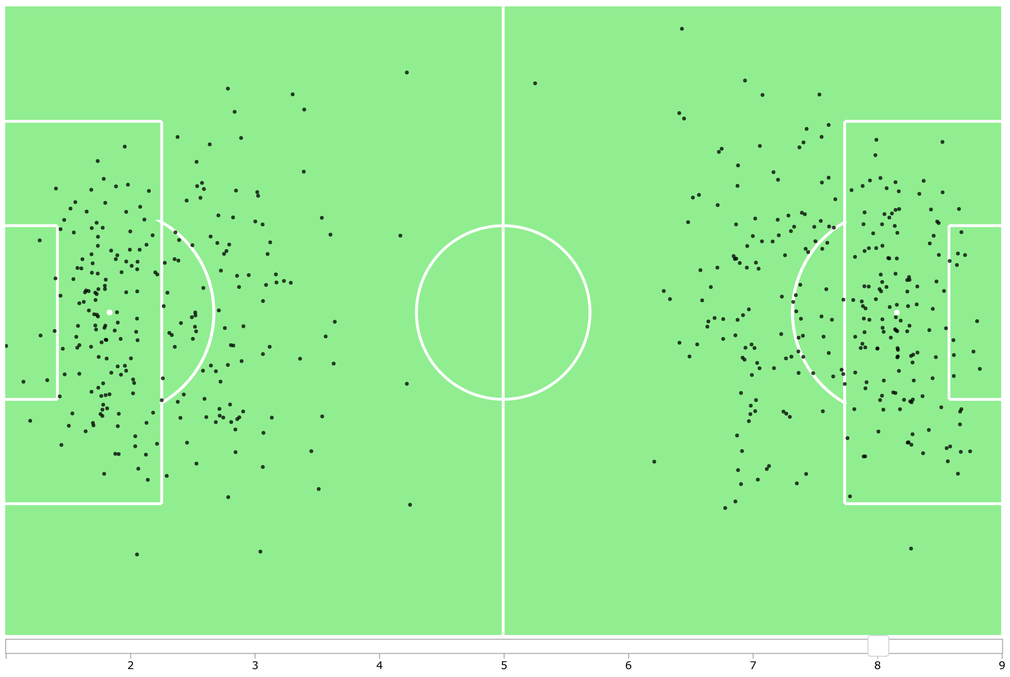
With the visualizations, a few tendencies became apparent:

- Teams make more goal attempts as a match progresses, and the attempts become more clustered.
- When a team loses to a lower ranked team, the higher ranked team fails more passes than usual, and plays more towards the middle of the field.

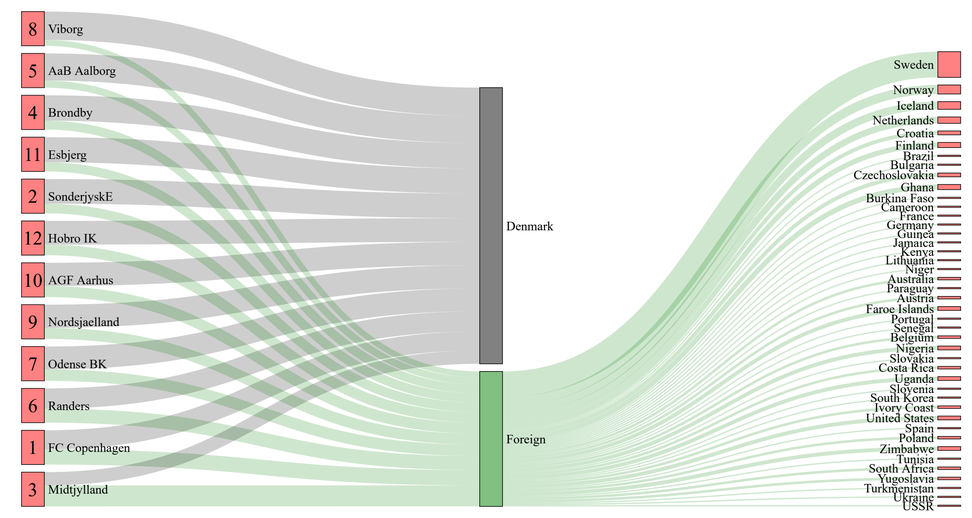
These tendencies were found on the field-based visualizations, which proved to be efficient idioms for our purpose.



Three matches played between FCM and Viborg, plotted from FCM's point of view. The color illustrates whether FCM lost (red) or won (blue). X-axis: time in seconds. Y-axis: failed passes in percent.



The map shows clustering of certain events in an entire season. In this snapshot, goal attempts between 70 and 80 minutes are shown. A dot represents the location of a goal attempt.



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