

In modern day society, data is collected at a rapidly increasing rate in all fields, and it has become a necessity to present data in different ways in order for humans to make sense of it. One way of doing so is through data visualization. Visualization of data can help humans' understanding of large data sets, as the data can be summarized very effectively, and patterns can quickly be recognized. When making visualizations, there are several considerations the designer must make, including: What are the semantics of the data being visualized, what is the target of the visualization and how the visualization is going to be created. It is also important to consider the human cognitive system when designing visualizations, such that the visualizations are designed to make it easier for humans to understand the data. Examples of this include limiting the amount of variables in the visualizations, as well as using contrasting colours. Given that visualizations are created to simplify large and complex data sets, sacrifices must be made to allow efficient interpretation. Too complex visualizations will not accomplish the task of presenting the data in an easily understandable manner, as humans will not be able to process all the information.

In this project we have been working with visualization of football data. The data at our disposal for the project was provided by Prozone ¹. Prozone is a company that specializes in collecting and visualizing football data. They provided comprehensive data from the Danish Superliga, such as team stats like points, shots, goals and individual player stats.

The visualization techniques and idioms we will present in this report can be applied in many different settings. Some of the visualizations are very specific to football, for example the ones with a football field, but these can easily be adapted to visualize data from other branches of sports. The more generic ones, such as the football parameter bubbles and the radar plots can easily be used to present other kinds of data. This could be financial data, biological data etc.

The visualizations we have made can specifically be used by a team to analyze how other teams play and adopt their strategy when they meet those teams, or to analyze in general how teams play throughout a match, and then see if they can learn something which can improve their own strategy. The visualizations can also be used on websites where normal football fans can view them and see if they can produce any interesting hypotheses, which they can try to test the next time they watch a game on TV. The visualizations can be used by betting companies such that they can give their users tools such that the users can find out how they want to place their bets. The visualizations can also be used during live transmissions of football games, for example by showing statistics about the match during the intermission. In the academic setting, some of the visualizations can be used by psychologists, to learn something about the mentality of players during a game, by using the field representation of shots, to see where a player is likely to shoot from under certain conditions.

In order to learn to design good visualizations, we will apply principles from the field of data science to present football data. We will use tools such as the programming language R to process data and plot static visualizations, and use the JavaScript library D3.js to make interactive and dynamic visualizations. Specifically, we will do this both by making visualizations that can help to explore the questions that we present below, and by doing exploratory analysis such that new patterns can be discovered. The specific questions that we will be investigating are:

- How does a team evolve throughout a season in terms of goals, points, etc.?

¹<http://prozonesports.stats.com/>

- How does a team's playing style change throughout a match?
- How does a winning team differ from a losing team?

During the visualization process we will consider different visualization techniques and choose a suitable one based on principles and analysis tools given by Tamara Munzner in "Visualization Analysis and Design" to make sure that the data is presented in an accurate and easily understandable manner. This includes understanding the type of data that is to be visualized, considering how to filter the data to get the relevant data for the individual visualizations and how to create the visualization. We will discuss the results of the visualizations and specifically what we can and can't learn from them.

In this report we will start by introducing relevant theory, then present the data we have from Prozone, followed by analysis and visualizations of the data. The design of each visualization will be explained as well as the findings. The findings and the visualizations will then be discussed, and followed by a conclusion.