Abstract

As data is being collected at a continuously increasing rate it has become a necessity to represent it in alternative manners for humans to make sense of it. One way of doing so is through visualization. Visualization is a powerful tool which augment human capabilities making it easier for us to interpreted data. However, there are many pitfalls and bad design decisions which may cause the visualizations to not perform as intended. For this paper, we have been working with football data and tried to create new forms of visualization whilst also answering some common questions regarding how you can statistically tell the difference between a good and a bad team, whilst also looking at what their play-style is like, both in single matches and seasonal. To aid in creating visualization we have been using technologies such as R and $D3.js$ for preprocessing and visualizing, respectively. Throughout the the course of the project, we have also been using a variety of statistical methods to help finding patterns but also to prove some of the tendencies which were found along the way. We ended up with some visualizations, both new and common ones, which aided in finding patterns in the datasets. Some of the things which we found was that more shots are actually taken towards the end of a game. The number og shots will already from the beginning of the game progressively increase all the way until the end. Some other tendencies the visualizations showed was that teams usually makes more mistaken passes at the beginning of each half relative to the match, which may imply that some psychological factors affects the players and how they play. Another visualization which also undermined the idea of psychological affects on players was also made. The visualization showed us that when a good team faces a worse one and loses, one can actually see a difference in how many failed passes they do relative to when they win. It also indicated that more play would enroll around the middle of the field, which may indicate a rather frustrated play-style. Though these tendencies seemed clear in the visualization, the quantitative foundation was not there to confirm the hypothesis and thus allows for further investigation to either confirm or disprove the idea. All said, it turned out to be rather difficult to predict anything concrete about the differences between top- and bottom-tier teams, as the error margins were simply too high.