Visualization of Football Data

Rasmus Bo Adeltoft
Sebastian Seneca Haulund Hansen
Steffen Berg Klenow
Christian Bjørn Moeslund
Andreas Staurby Olesen
Henrik Sejer Pedersen
Supervisor: Marco Chiarandini

18th May 2016

1 Abstract

2 Preface

Contents

3 Introduction

Data is collected at a rapidly increasing rate in all fields and it becomes necessary to present data in different ways in order for humans to make sense of it. One way to do this is through data visualization. Visualization can help human's understanding of large data sets, as the data can be summarized very effectively, and patterns can quickly be recognized by humans. When making visualizations it is important to understand how the human cognitive system works, such that visualizations can be designed to make it easier for humans to understand the data. In order to do this, we will apply principles from the field of visualization to present football data. We will use tools such as R to process data and plot static visualizations, and use D3 to make interactive and dynamic visualizations.

Specifically, we will do this both by making visualizations that can help 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.?
- How does a team's playing style (for example passes, possession and tackles) 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 considerations regarding the human cognitive system.

4 Theory

4.1 Something

4.1.1 Design Process

This section describes the typical work flow of a data scientist. We will focus on the following four faces: Preparation, Analysis, Reflection and Dissemination.

The process of getting the data, understanding the data and produce results is an iterative process. The process is seen on Figure 1.

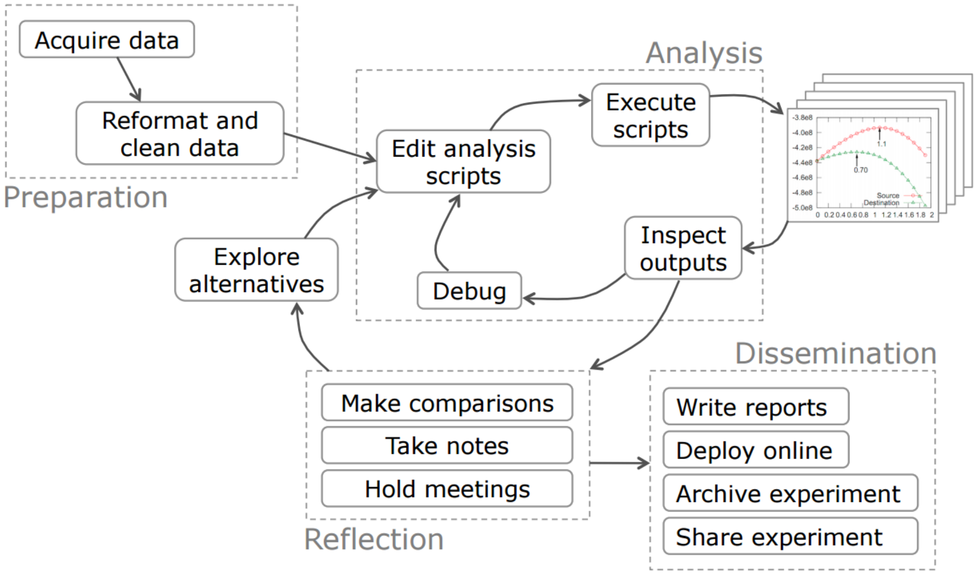

Figure 1: The model showing the iterative process[Guo12, Chapter 2]

- The first face is the preparation face. Here the you have to acquire the data, that could be from hard disks, servers, through an API ect. Where to store and how to organize the data files should be considered, so it is easy to replace the right files if the data gets updated. Then the data should be cleaned, meaning removing tuples with missing values, changing the formatting, sorting the data ect.

- The second face is the analysis face. Here the data is analysed to get more information about it. This is an iterative process, where the you create and run scripts, look at the output, maybe find some mistakes, debug these and run it again.

- The third face is the reflection face. Here the output results is discussed, for example by making comparisons between outputs, and exploring alternatives.

- The fourth and last face is the dissemination face. Here the the results are reported and maybe published in a report. [Guo12, Chapter 2]

4.1.2 Design

4.1.3 Data Types and Data Sets

4.1.4 Idioms

4.1.5 Analysis and Complexity

4.1.6 Facets and View Manipulation

When creating visualisations it is not always enough with one idiom to present the data in a understandable way. One way cram more information into the idiom, but still keeping the number of variables low is to either manipulate the view of the idiom or to facet into multiple idioms.

4.1.7 View Manipulation

By creating multiple views in an idiom, the idiom can contain more information, without clutter. The different views can include changes like switching between different idioms, changing the viewpoint, changing the order of the data, changing the number of items showed and so on. For example you can change the way the data is ordered by sorting the data by different variables. This is very powerful because of spatial position being the highest ranked visual channel. Many view manipulations is based on animation. Animating has a trade off, the cognitive load can be very high if too many elements change. This means that we get low cognitive load if either some elements are static and others moving, or some groups of elements are static and others moving. If few elements change by a graduate transition, the viewer can keep the context between the two views. [MM15, Chapter 11]

Selecting one or more elements is common in many interactive visualisations. The result of selecting some elements is then some change in the view. It has to be considered which elements the user can select, and how many the user can select. Choosing how to select items is also subject which has to be considered, clicking to select, hovering over some element with the cursor or something else. Changing the view by highlighting some element and creating pop out could be done by changing the channel, for example the color, the size, the outline or the shape of the element. This change should of course be so dramatic that the element clearly stands out from the rest. Look at Figure 2 for an example of highlighting. [MM15, Chapter 11]

Another option for an interactive idiom is the ability to navigate the view by changing the viewpoint. Here we think as if we have a camera pointed at the view. We can then change the view by zooming, panning or rotating the the camera around its own axis. There are two kinds of zooming,

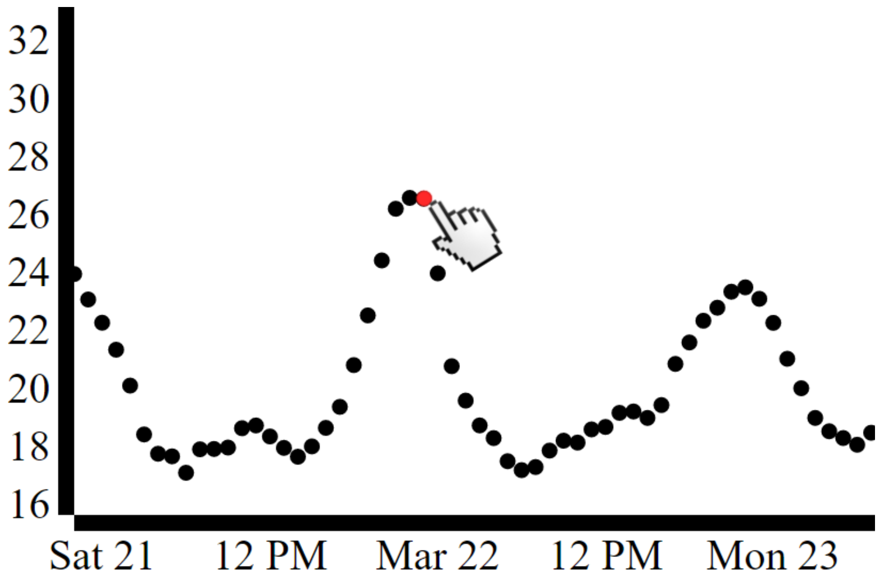

Figure 2: Highlighting the selected element

geometric zooming and semantic zooming. Geometric zooming is straight forward making some elements come closer to the camera. With semantic zooming the not only the size of the elements change, but the semantics too. Semantic zooming changes what is shown, and maybe the representation of it. For example zooming semantically could view more detail about some element showing new information about it. Navigation could also be changed through reduction of attributes, by slicing, cutting or projecting. These are all dimension reduction techniques. To slice, a specific value at a dimension is chosen, and only elements matching this value is shown. A cut is made by placing a plane in front of the camera, all elements in front of the plane is not shown, in this way it is possible to explorer elements behind other elements, or looking inside 3D objects. Projection is done by eliminating some dimension but still showing all the data. This is similar to what the human do when looking at a 3D object.

4.1.8 Facet

Faceting is splitting the view up into multiple views or into multiple layers. One of the main reasons to facet is to compare views. This is much easier than comparing two views in a changing view, because we do not have to remember the prior view, but continuously can compare them. Another reason to facet is to gain more information about the data through a multiform design, where data is shown using different encodings. By having multiple views

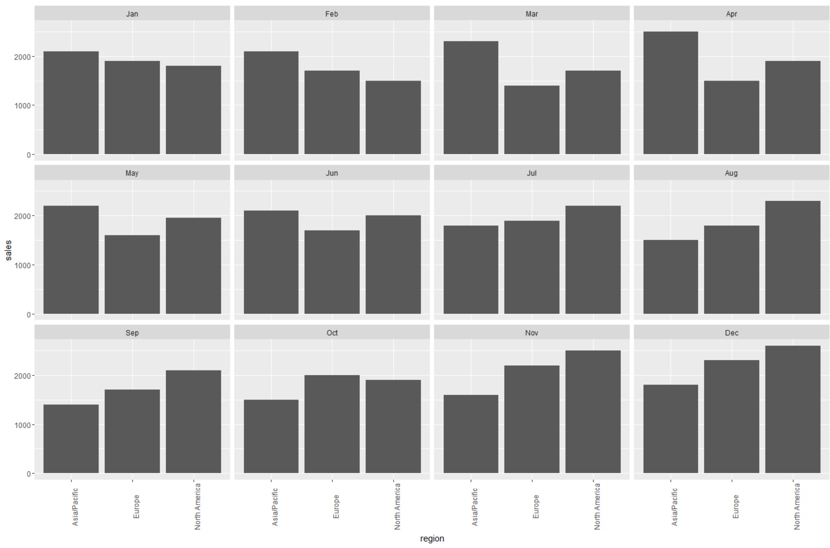

Figure 3: An example of small multiples

more attributes can also be shown. Of course when you have multiple views shown beside each other, each view has less space, which is one of the trade off's and why having multiple layers on top of each other, and switching between them sometimes is better. If having juxtaposed views it might be interesting to link the views. This could be done sharing the data, sharing the visual encoding, synchronizing the navigation or highlighting.

Each view could show different subsets of the data, or having different viewpoints like the classic overview in one view combined with detail in another view. Having multiple views sharing encoding but showing different parts of the data is called small multiples and is often structured in a matrix. This could be an alternative to animations, where we lay out all the frames. The cognitive load is smaller with small multiples, and it is easy to go one frame back or forth. An example of small multiples is seen at Figure 3. Instead of juxtaposing the views, it is a option to stack them into a single frame. The views should have the same horizontal and vertical extend and blend blend together as one frame, by being transparent where there are no marks. The problem with stacking is distinguishing between the layers. This is easy with only a few layers, especially if the layers use different visual channels. But distinguishing between more than three layers, can be a real challenge. [MM15, Chapter 12]

4.2 Exploratory Analysis

4.3 Tools and Technologies

4.3.1 R

4.3.2 D3.js

5 Results

6 Discussion

7 Conclusion

8 Usability

9 Bibliography

References

[Guo12] Philip Jia Guo. *Data Science Workflow: Overview and Challenges*. PhD thesis, Stanford University, 2012.

[MM15] Tamara Munzner and Eamonn Maguire. *Visualization analysis and design*. CRC Press, Boca Raton, FL, 2015.

10 Appendix

11 Process Evaluation