Problem Set 3

The third problem set focuses on advanced visualilzation. Your solution should be composed of a well-structured R script which should provide the required analyses. Besides the functions the code should be directly runnable or at least sufficiently well documented (working directory, path settings) to be executed.

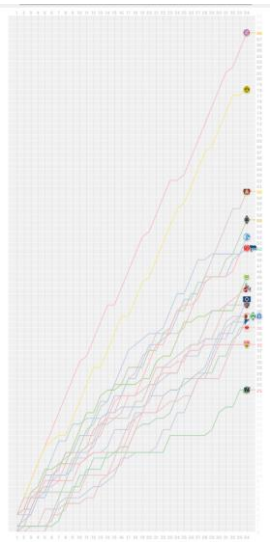
Furthermore, provide a documentation (in text format or powerpoint) where you document and illustrate your general approach, document the qualitative tasks and provide instantiations of your graphs. Provide some intuition of what your plots show and what you hypothesize.

Grading will reflect your performance on both the coding as well as the documentation and interpretation tasks – however, there is no fixed grading scheme between the two categories.

This problem set is **due on May 31st by 12.00** through the wuecampus group functionality.

Cooperation with other groups is not permitted and will lead to severe credit deductions.

1. Another fairly boring Bundesliga season is finished and the gigantic lead of Bayern and Dortmund prompted the newspaper Zeit to create a "better table:" http://www.zeit.de/sport/2016-04/bundesliga-tabelle-bessere-abstaende



Your task in this assignment is to replicate the chart as closely as possible using ggplot. The dataset D1 (curtesy of http://www.football-data.co.uk/germanym.php) provides all the necessary information to recreate this point chart – see http://www.football-data.co.uk/notes.txt for an exact documentation. For people not interested in sports, a win yields 3 points, a draw 1 point.

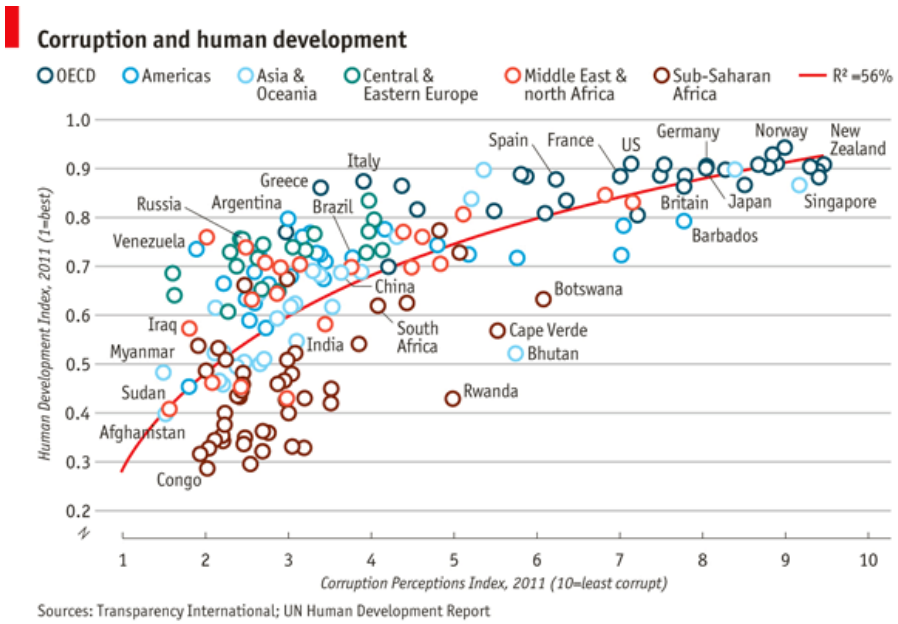
Some remarks:

- A good start is to melt the dataset with measure.vars = c("HomeTeam","AwayTeam")
- You do not need to implement the team logos, however nice labelling of the teams (possible abbreviated) next to the final tally would be nice start
- Aligning team colors with the line color would be nice as well
- The total points on the right are also nice

For those who want to go the extra mile

- http://stackoverflow.com/a/36172385/3124909 may provide insights on how one could include the team logos (if fame and bonus points motivate you)
- Another bonus point avenue would be to use ggvis and implement a dynamic visualization – e.g., allowing to choose which table to be displayed (e.g., home / away / total or points / places)

2. The Economist published the following graphic which relates a society's human development index to its corruption perception index.



The file *EconomistData.csv* provides you with the necessary data for replicating this graph. Again use ggplot and try to accomplish the following aspects (in decreasing order of importance):

a. The basic plot structure
b. Color-coded regions
c. Textual labeling of selected countries (Afghanistan, Greece, China, India, Rwanda, Spain, France, United States, Japan, Norway, Singapore)
d. Axis labels and title
e. Fit line
f. General look and feel (hint: shape=21 offers a circle with fill and line for geom_point)

3. A famous visualization by Hans Rosling and gapminder illustrates countries developments towards a better level of living: http://www.gapminder.org/videos/200-years-that-changed-the-world-bbc/
    a. The package "gapminder" includes the dataset gapminder which fuels these graphs. Replicate the graph for a (choosable) year.
    b. Now approximate the animation effect using faceting (essentially we create a "Daumenkino" version). Play around with selective timeslots and additional faceting for regions.
    c. We are now interested in trajectories: Use geom_path to plot the country over time. Ensure that the graph remains understandable by means of appropriate filtering, labelling, transparency or color/shape coding.
    d. Discuss the following countries: Kuwait, Singapore, China, India
    e. Finally, develop an own alternative visualization which captures the idea behind the original gapminder visualization. Explain what your visualization does better and where it is less powerful.

4. World population is an important topic for economic, societal and political reasons. The Economist recently reviewed the current situation and put it into perspective. http://www.economist.com/blogs/graphicdetail/2015/08/daily-chart-growth-areas

   You can find the underlying data here: http://goo.gl/EbUf4a with the dataset description (http://goo.gl/qhHNWu).

   Replicate the top two graphs as closely as possible (ignore colors). Subsequently, create a similar visualization of the most populous countries chart. It is ok to create three charts separately, however using faceting you may be successful as well. It is not necessary to use barplots if you can capture the required information.