



Data Visualization Portfolio

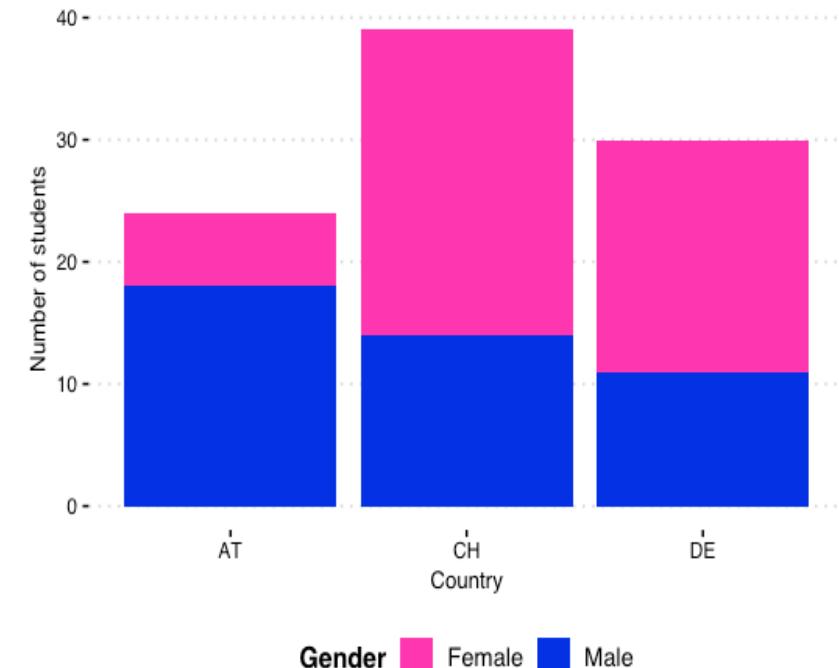
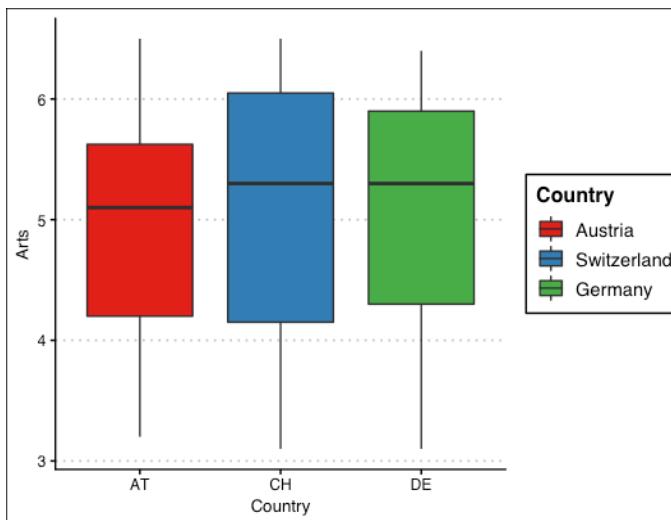
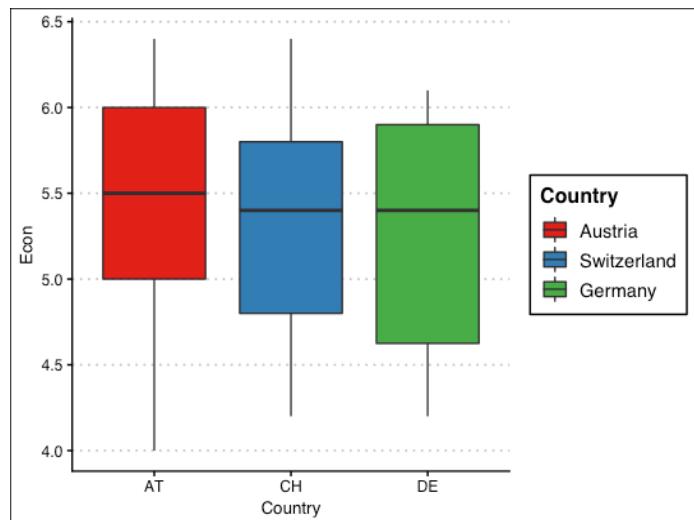
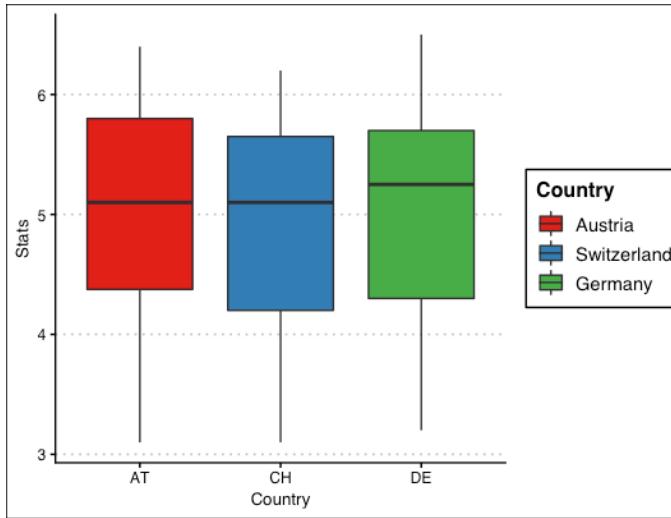
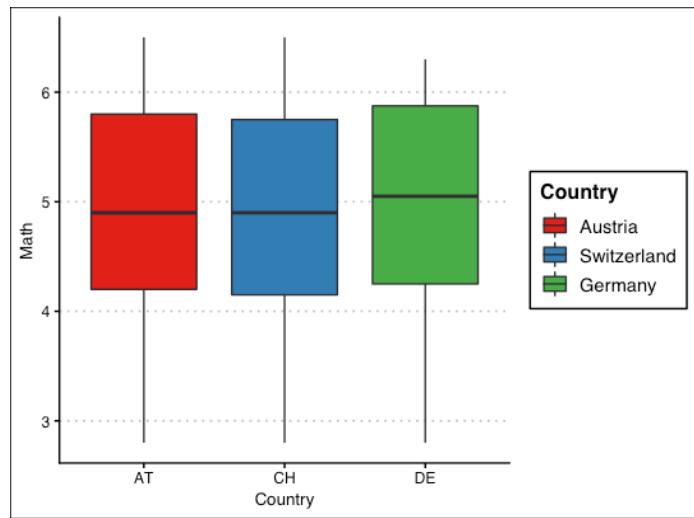
Arbian Halilaj

16-609-828

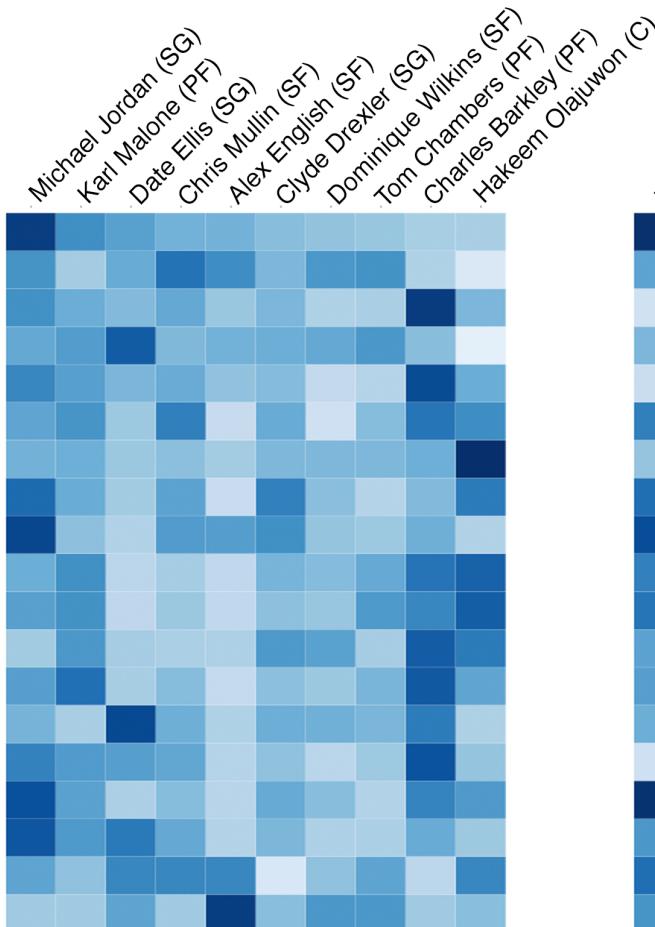
Content

Student Data	3
Color as important aesthetic	4
Black & White	5
Bad Visualization	6
Improved Version	7
Good Visualization	8
Visualization about myself	9
World Bank Data	10
Swiss Data	11
Creation Process	12
Data Map with Tableau	14
Visualization with many aesthetics	15
Favorite Tools	16

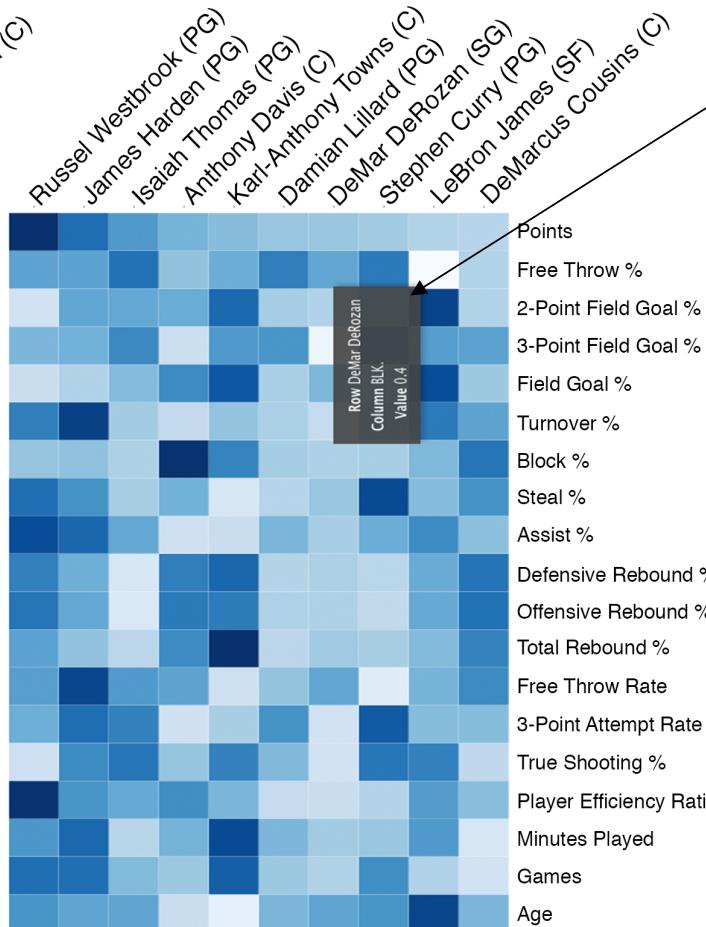
Student Data



Color as important aesthetic



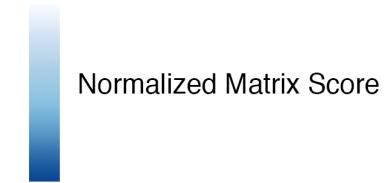
NBA Top-10 Scorer Statistics 1989 (season)



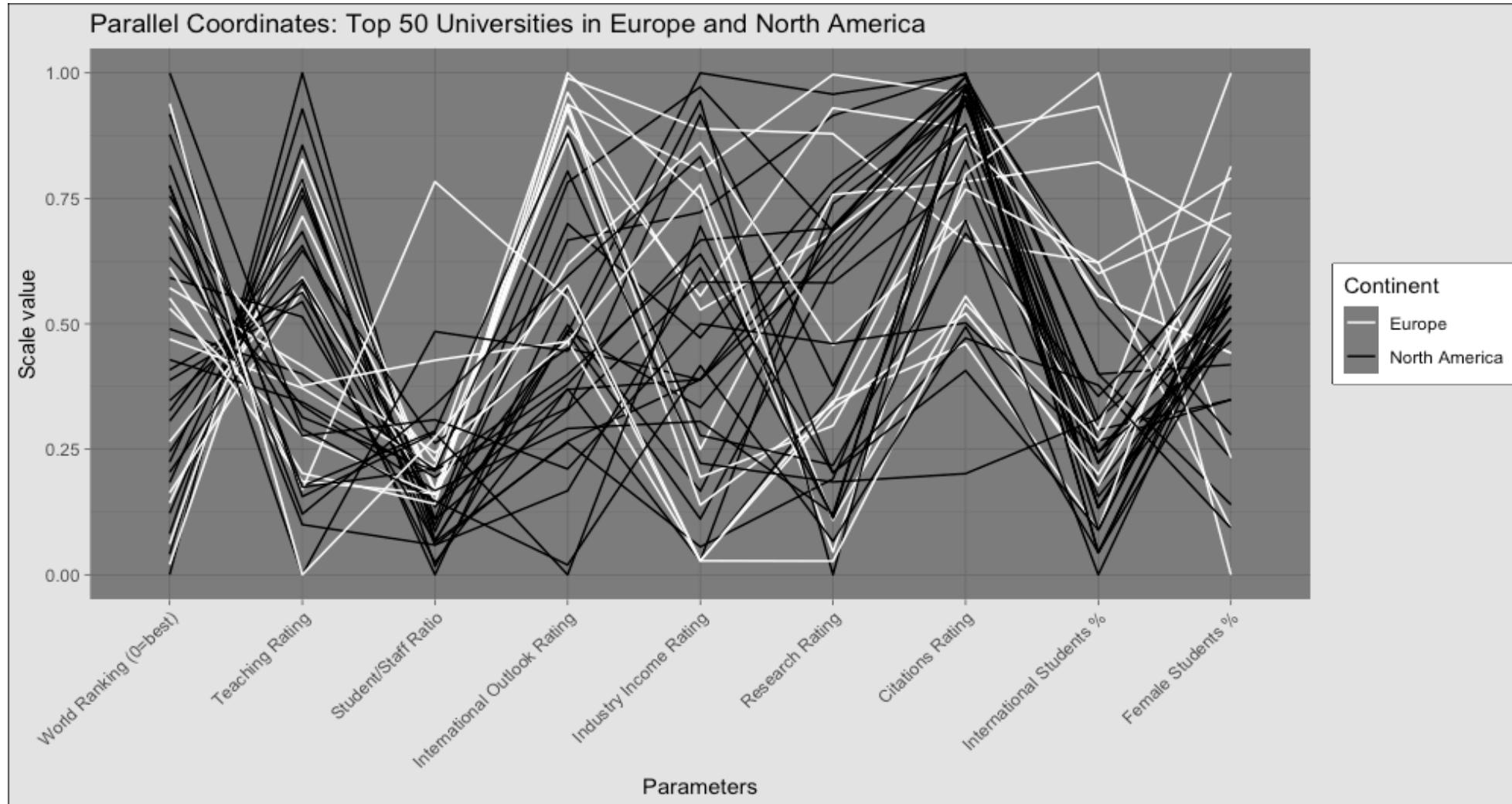
NBA Top-10 Scorer Statistics 2017 (season)

Interactive HTML Widget:

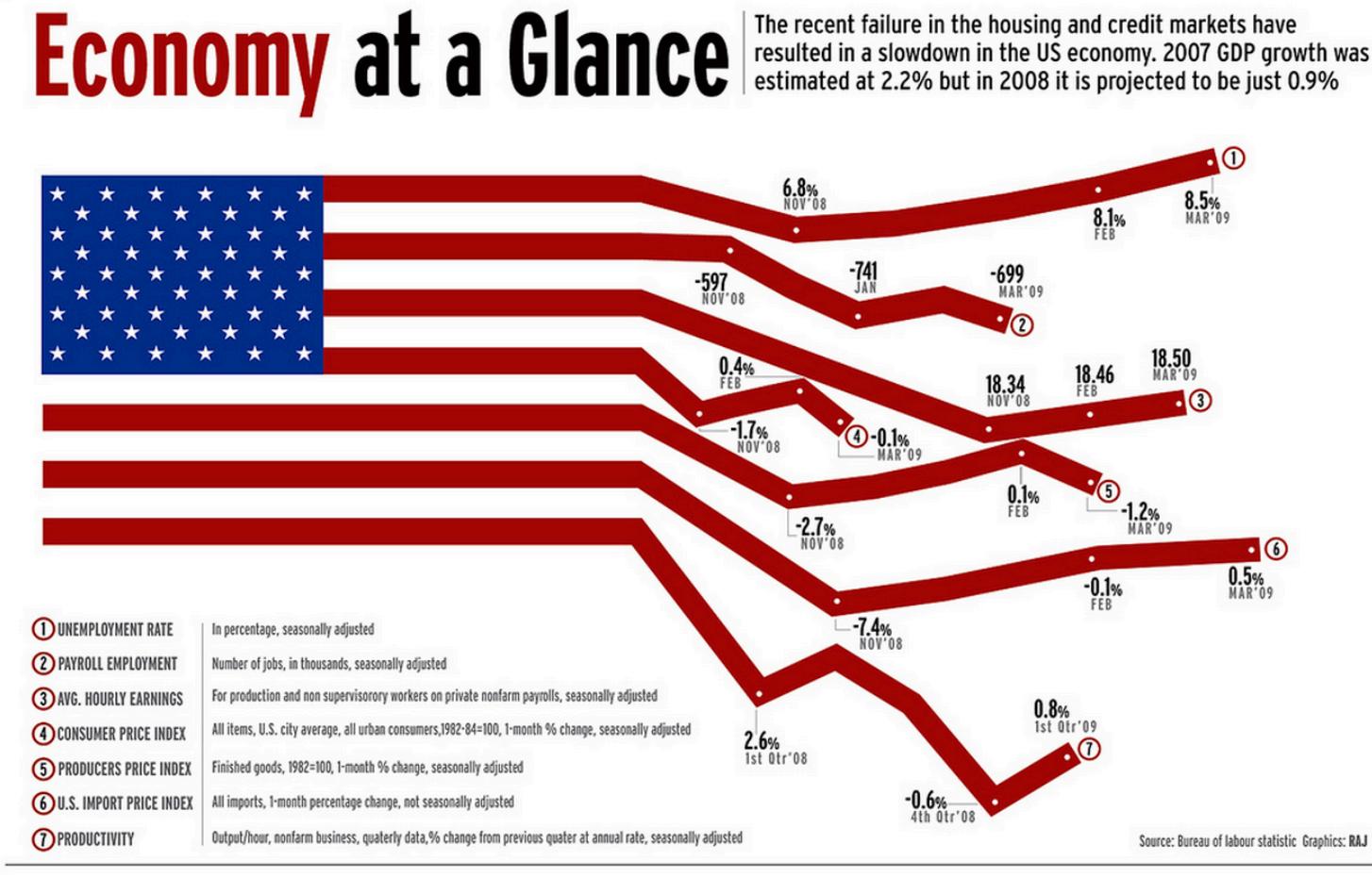
<https://drive.google.com/open?id=1Bqy5arfSz0cB70SvlokRzUFpEQTZBNrQ>



Black & White



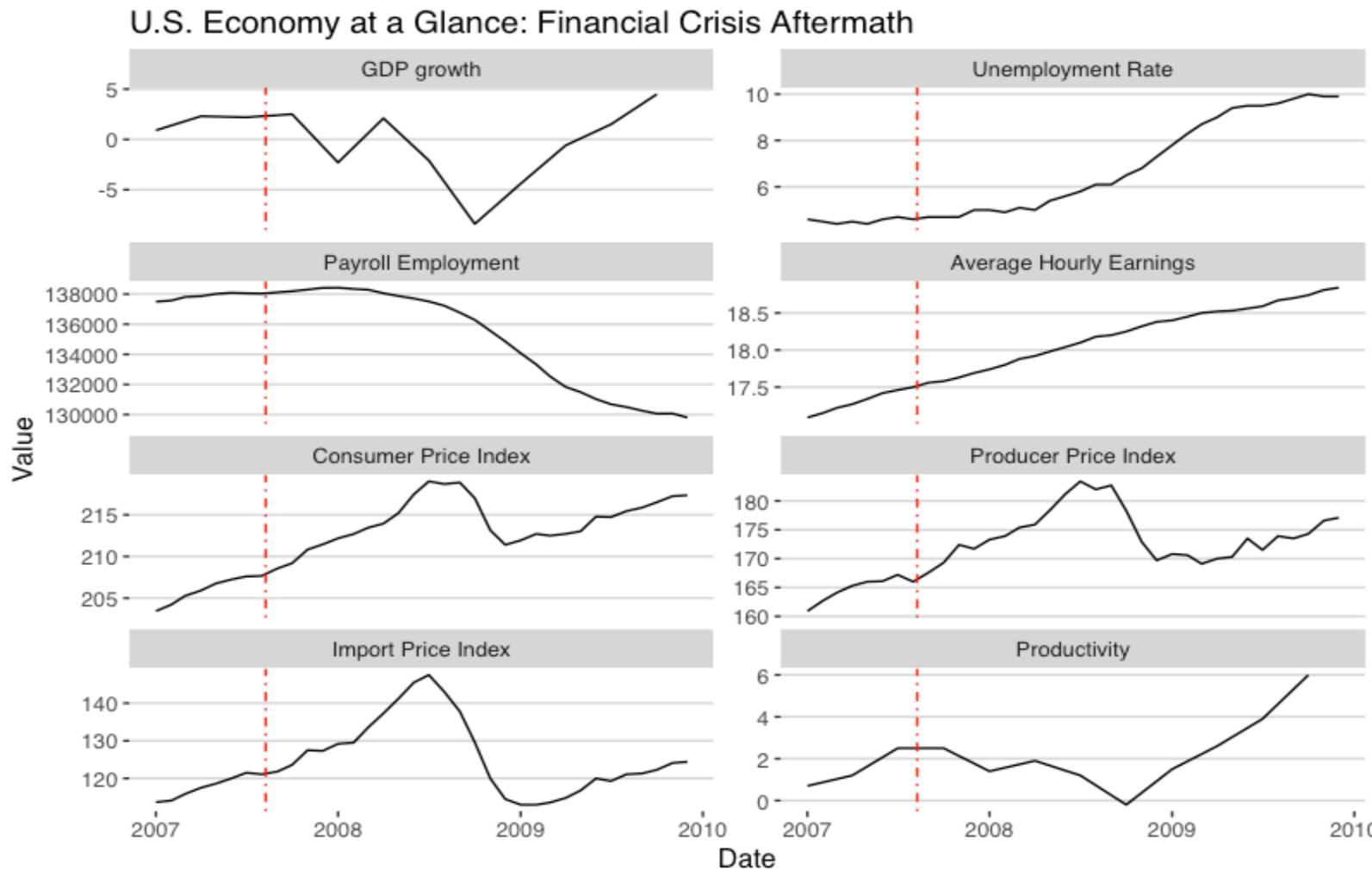
Bad Visualization



The shown graphic is a perfect example of a bad visualization. Before I explain why, here is an interesting fact: I found this viz on a blog and I couldn't trace back the source (Bureau of labor statistics in the U.S), which means it must have been removed.

- 1) The scale of the red stripes is really bad. The up and down movements do not show the true scale (e.g. Nr. 2; -741 & -699). Furthermore the core question is what happened after the financial crisis. I do not see any year label of 2007. Also the year (x) axis changes between variables. To sum up it is really difficult to interpret the values of this viz.
- 2) It is kind of funny that they used exactly seven variables, which is equal to the seven red stripes of the U.S flag. The question arising is why they used such a specific payroll (Nr. 3) and why they didn't include exports and GDP growth into the viz.

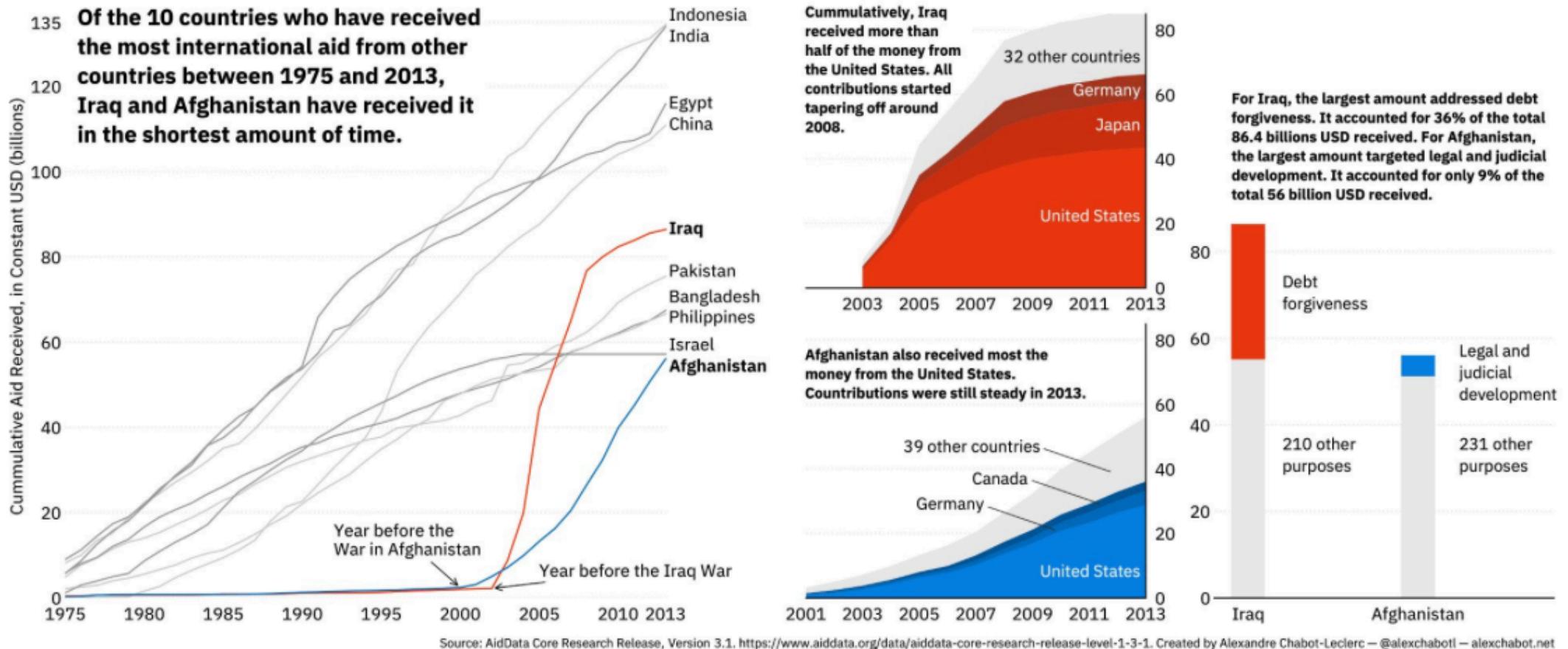
Improved Version



For the improved version I used data from the Federal Reserve data base. I made for each variable of interest a line plot. In addition to the bad version I included GDP growth and a red dotted line showing the beginning of the financial crisis (09/08/2007).

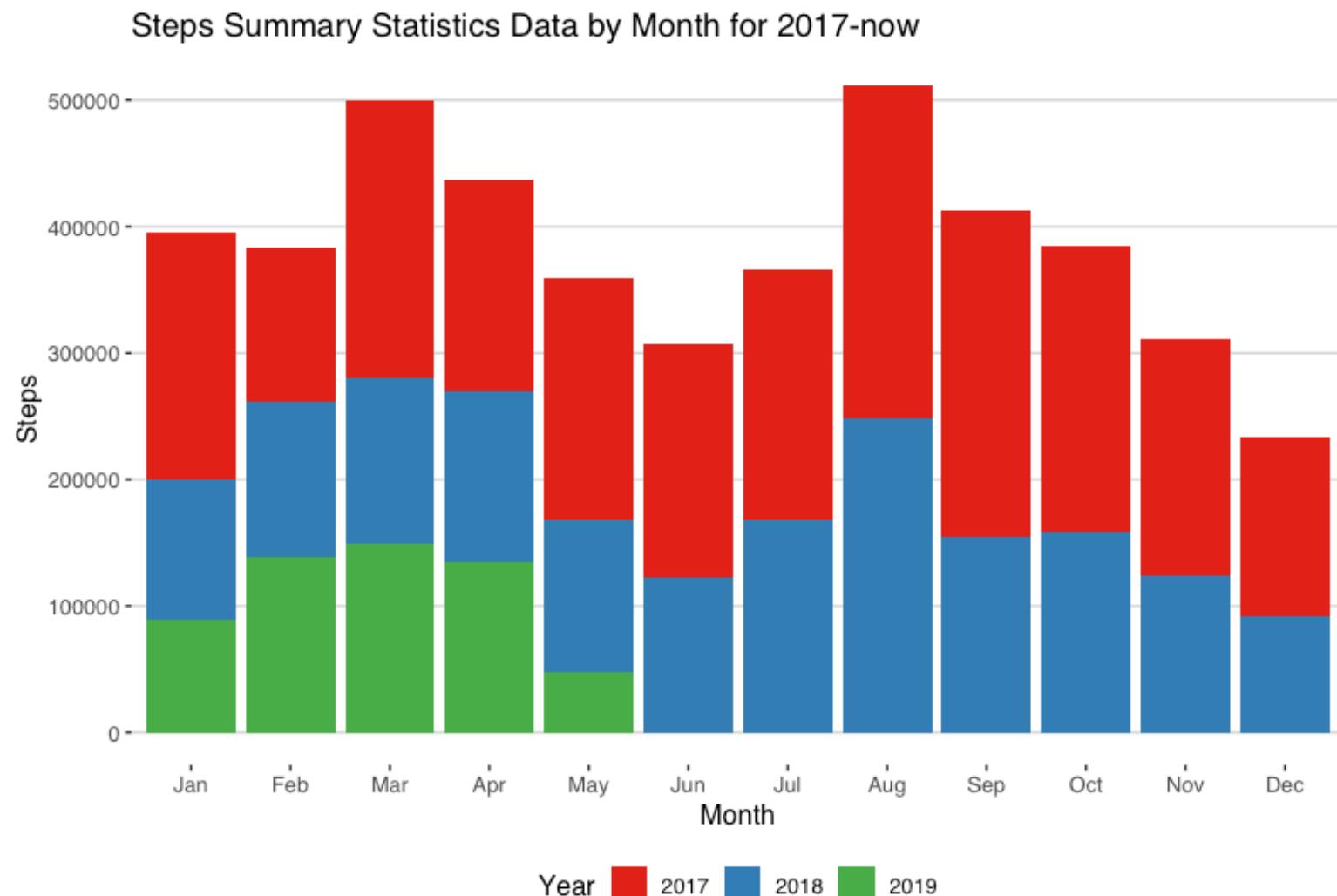
Comparing the lines of my improved version with the bad version is mindblowing. The lines of the latter one are completely misleading and also the data seems to be different.

Good Visualization



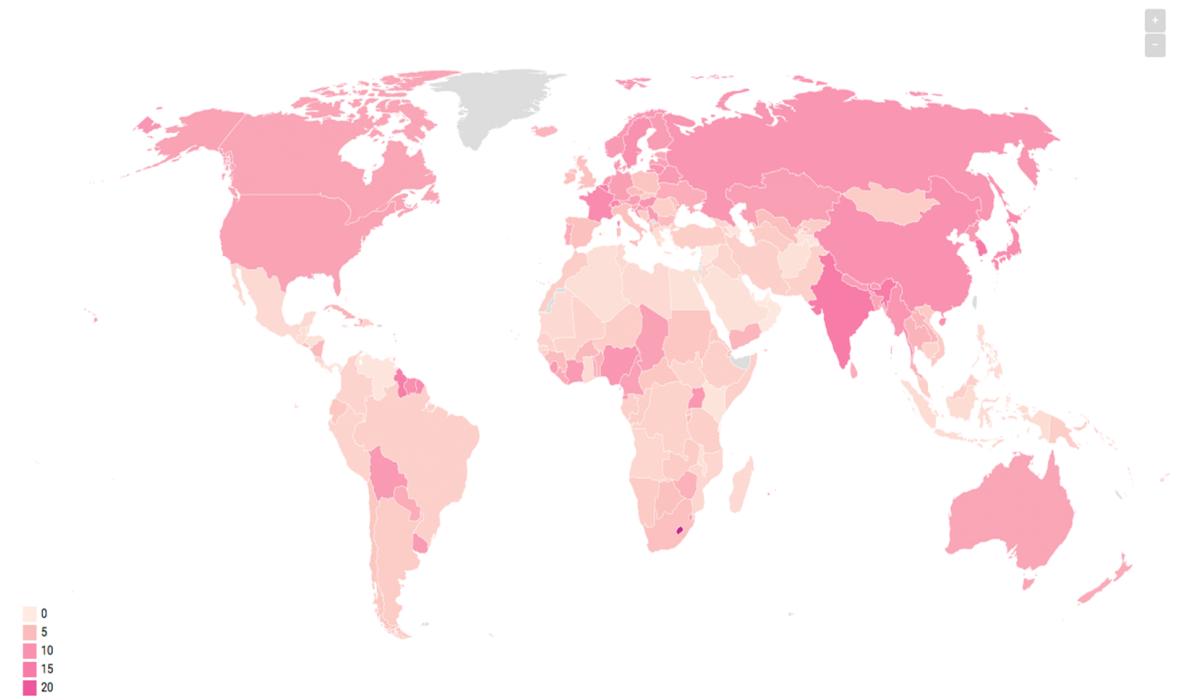
Very interesting story shown by three types of plots.

Visualization about myself

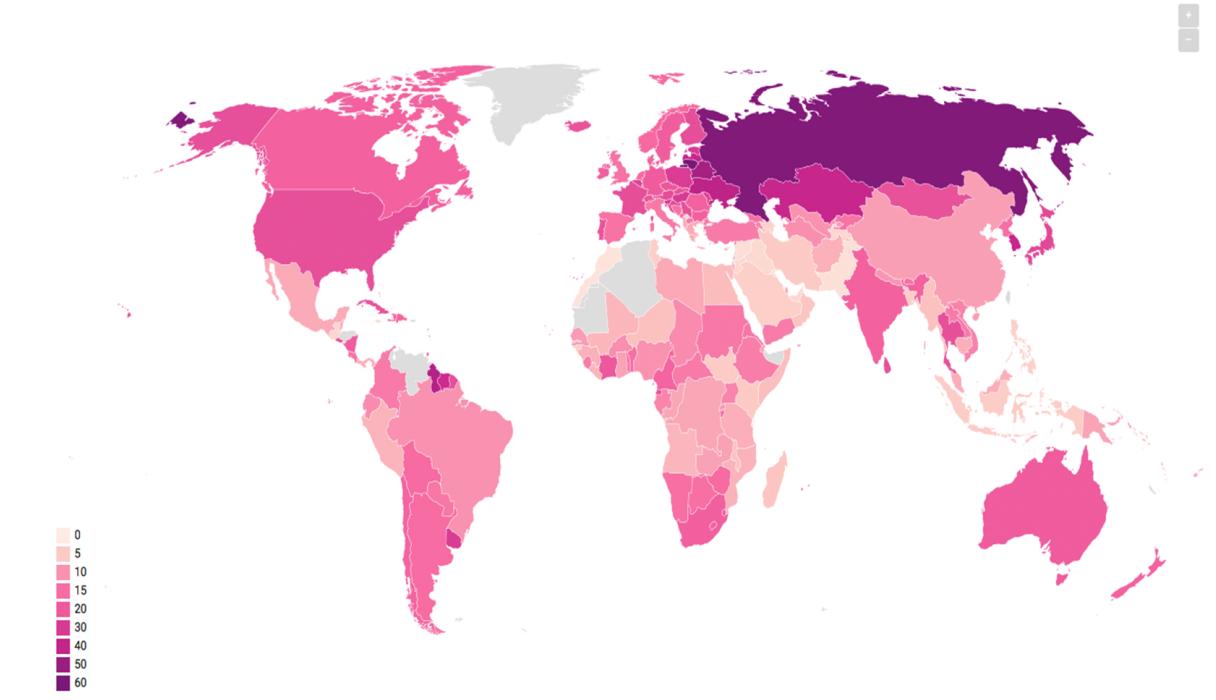


World Bank Data

Suicide mortality rate (per 100,000 female population)



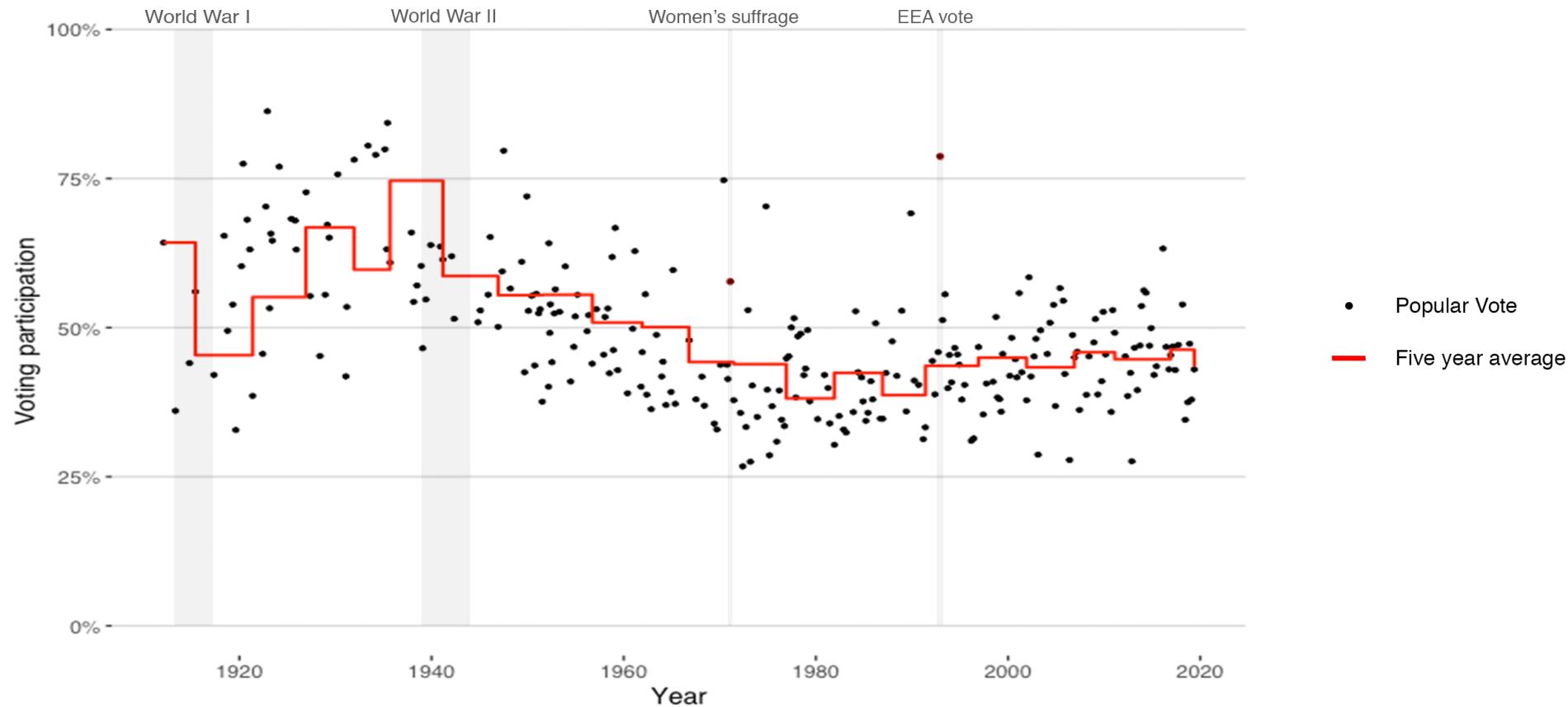
Suicide mortality rate (per 100,000 male population)



Source: World Bank Data • Get the data • Created with Datawrapper

Swiss Data

Participation in popular voting in Switzerland

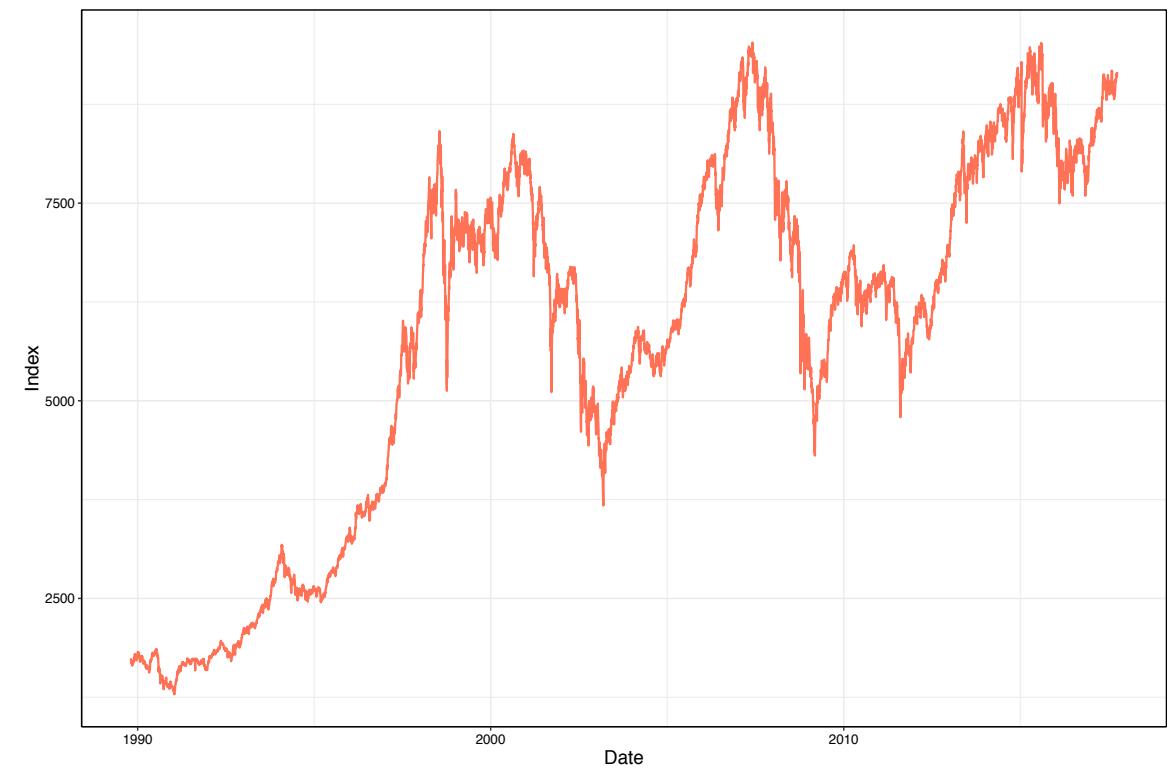


Creation Process

After downloading the data, we started the cleaning process. Eventually we obtained a clean data set of the largest Swiss companies' share prices as well as the Swiss Market Index valuation. We decided to start plotting the UBS share price over the years.



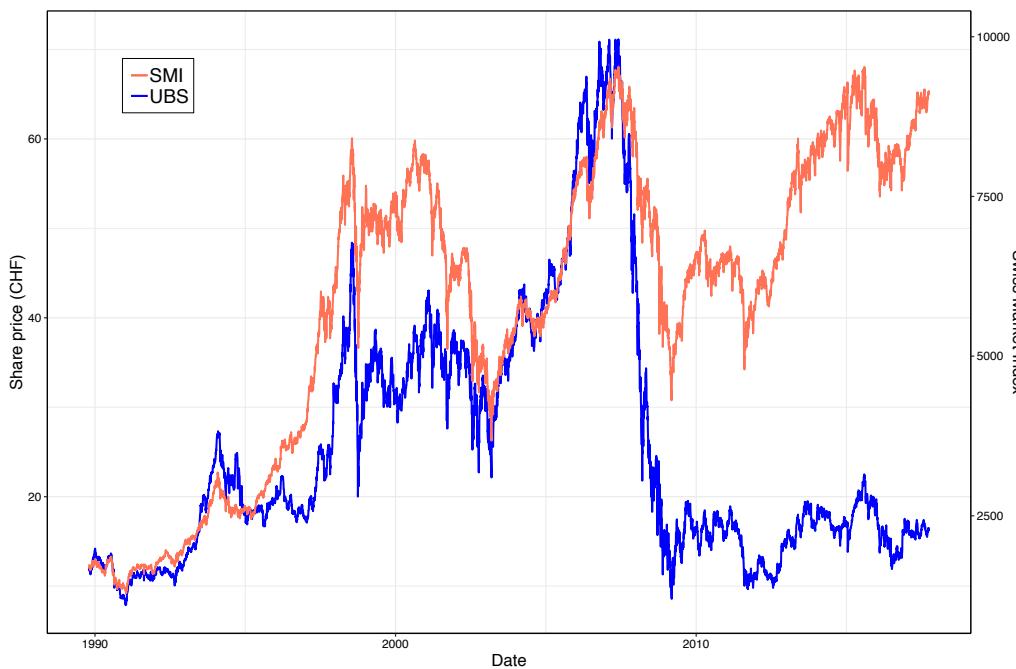
Subsequently, we adopted the same approach to illustrate the development of the Swiss Market Index for the same time frame.



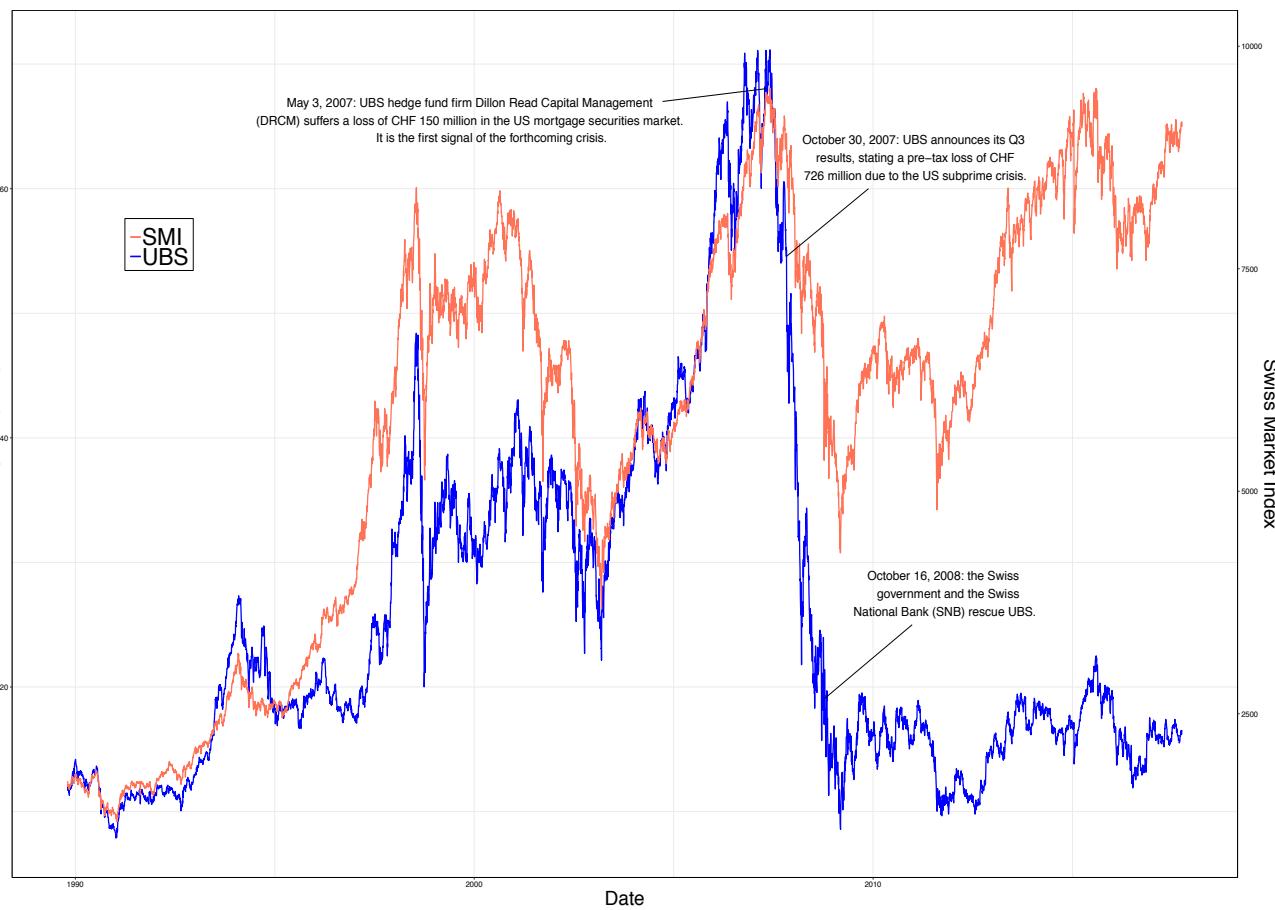
Creation Process

At this point, we had to figure out how to visualize the two plots together. The solution was to create a secondary axis, dividing the SMI data by a factor of 140, therefore making the two time series comparable. The code is the following:

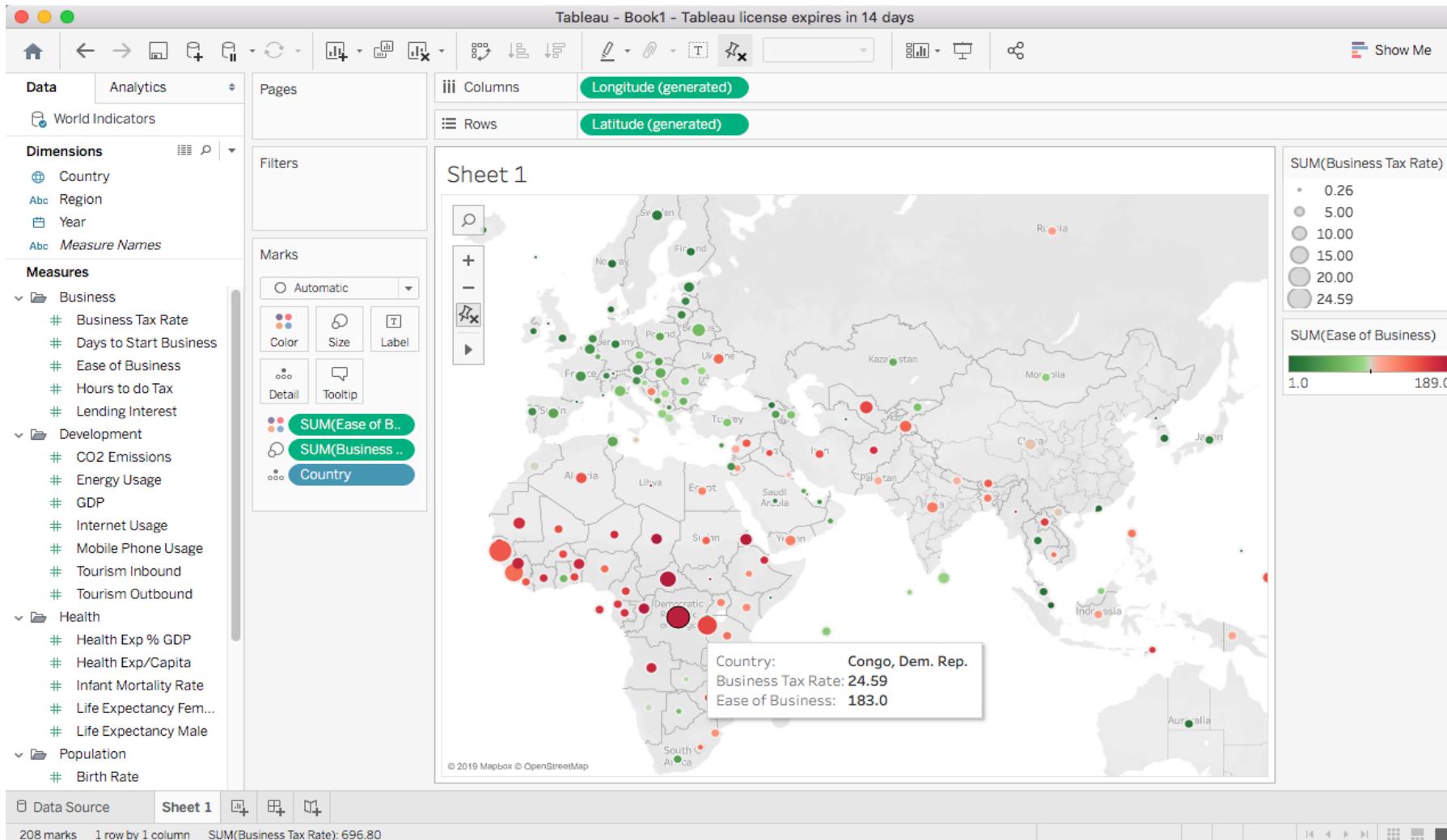
```
ggplot(data3, aes(x = Index)) + theme_bw() +  
  geom_line(aes(y = UBS, colour="UBS"),size=0.8) +  
  geom_line(aes(y = SMI, colour="SMI"),size=0.8) +  
  scale_y_continuous(sec.axis = sec_axis(~.*140, name = "Swiss Market Index")) +  
  labs(y = "Share price (CHF)",  
       x = "Date", caption = "Data Source: Swiss Market Index"  
) + scale_colour_manual(values=c(UBS="blue", SMI="coral1")) +
```



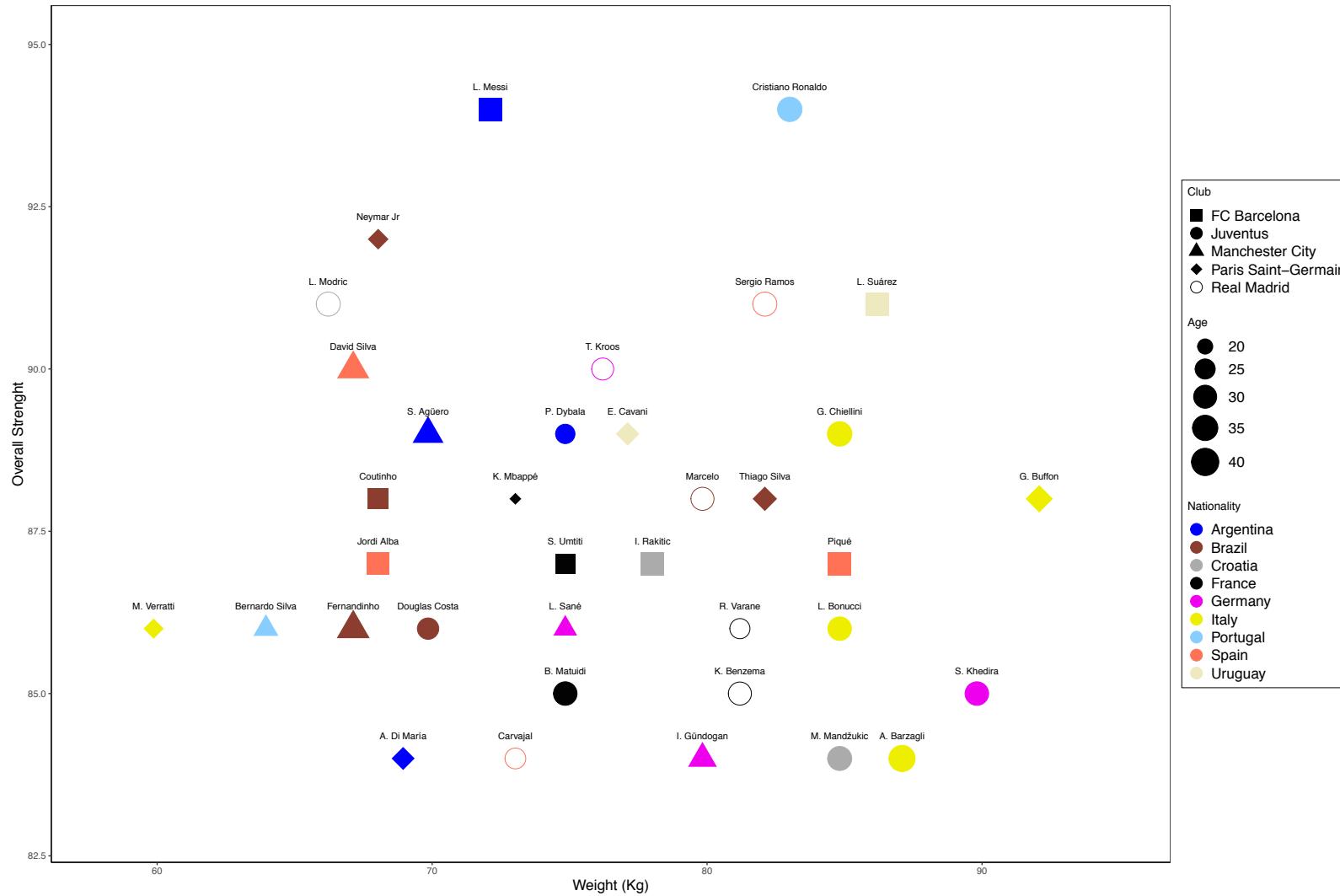
The last step was to add few important indications of crucial moments in the history of the 2007-2009 crisis concerning UBS. Furthermore, some details were adjusted.



Data Map with Tableau



Visualization with many aesthetics



Favorite Tools

To find an appropriate visualization technique and some inspiration I used the R Graph Gallery.

The majority of visualizations in this portfolio were made with R. I have played around with several packages to get the results I wanted. The variety of packages is amazing, although each has pro's and con's.

Other programs used for my visualizations were datawrapper.de and Tableau.

Sometimes it was really difficult to get the results I wanted so I had to use Photoshop or Inkscape to fix the issues.

Among the various blogs I enjoyed storytellingwithdata.com the most. They are doing monthly a data viz challenge.



Sources

Student Data	Student.RData	ggplot2, ggthemes
Color as important aesthetic	https://www.kaggle.com/drgilermo/nba-players-stats	d3heatmap, Photoshop
Black & White	https://data.world/hhaveliw/world-university-ranking-2016	GGally
Bad Visualization	http://bankinginsuranceworld.blogspot.com/2016/10/economic-overview-of-united-state.html	
Improved Version	FRED Database	ggplot2, reshape2, ggthemes
Good Visualization	http://www.storytellingwithdata.com/blog/2019/3/25/how-you-visualized-this	
Visualization about myself	Health App & QS Access App (get csv file)	ggplot2, ggthemes
World Bank Data	https://data.worldbank.org/	datawrapper
Swiss Data	https://www.bk.admin.ch/ch/d/pore/va/vab_2_2_4_1.html	ggplot2, ggthemes, scales
Creation Process	https://www.bloomberg.com/professional/solution/data-and-content/	ggplot2, xts, zoo
Data Map with Tableau	World Indicators (Data set inside Tableau)	Tableau
Visualization with many aesthetics	https://www.kaggle.com/karangadiya/fifa19	ggplot2