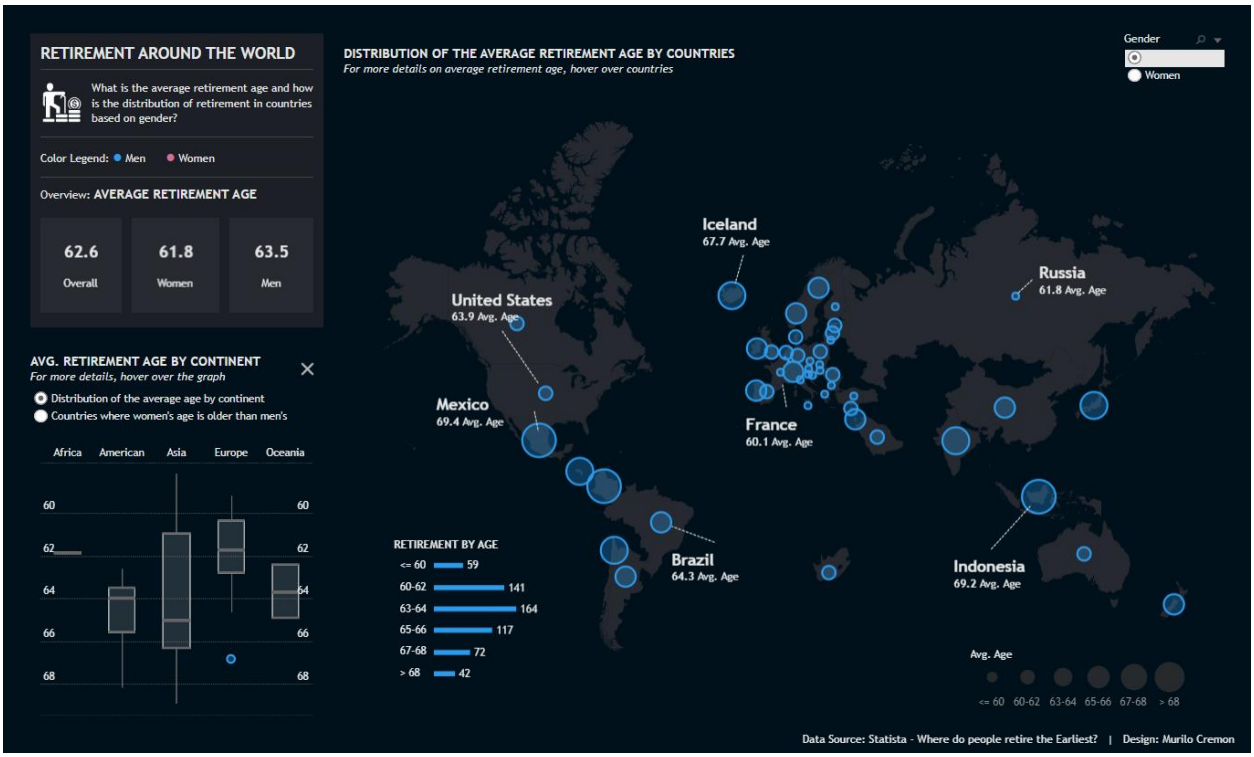
Eliezer Molina Mello
Assignment 2: Good and Bad Data Visualization

Good data visualization – Retirement Around the World

Link:
https://public.tableau.com/app/profile/murilo.cremon/viz/RetirementAroundtheWorld/Dashboard



There are several reasons why this graph is good and here are some of the most important reasons:

Firstly, based on the anatomy of a figure, one noticeable feature of a good visualization is visual cues such as title, legend, labels, and grid. The above graph, even having two visualizations, a box plot, and a map, can convey key information thanks to its well-positioned and formatted labels. Also, the plot appearance makes the graph very soothing to look at, the background and labels have very distinct colors that it is very easy to notice and understand each point or bar in the graphs. We can also see a very minimalistic noticeable grid line in the box plot, which is just enough for anybody to see the various age buckets.

Secondly, to further explain why this graph is good I will write a detailed explanation based on the qualities of the visualization discussed in class. First, the visualization has great aesthetics as it can be seen that the mix of colors, fonts, and positioning of the graphs are all very pleasing to look at. Second, the visualization is very substantive, meaning that it accurately and honestly presents data by correctly listing the data source at the bottom of the graph (Statista). Third, the visualization is very perceptual, meaning that we can clearly

understand the message the maker was attempting to convey, which is to show us a global representation of the various retirement ages for each country and by gender, male, and female.

Thirdly, I want to also describe why this visualization is good, based on its dependencies discussed in class. For context, given that retirement is a common topic, this visualization can be used in several contexts such as academic journals and posters, since its source data is valid. For the audience, even though retirement is only a topic that affects seniors, I believe people of all ages can be interested in knowing more about this topic. For data structure, the visualization does a pretty good job highlighting all the data it has and hasn't, which is the case of Africa, a continent that the visualization has no data of, hence it's not shown in the visualization.

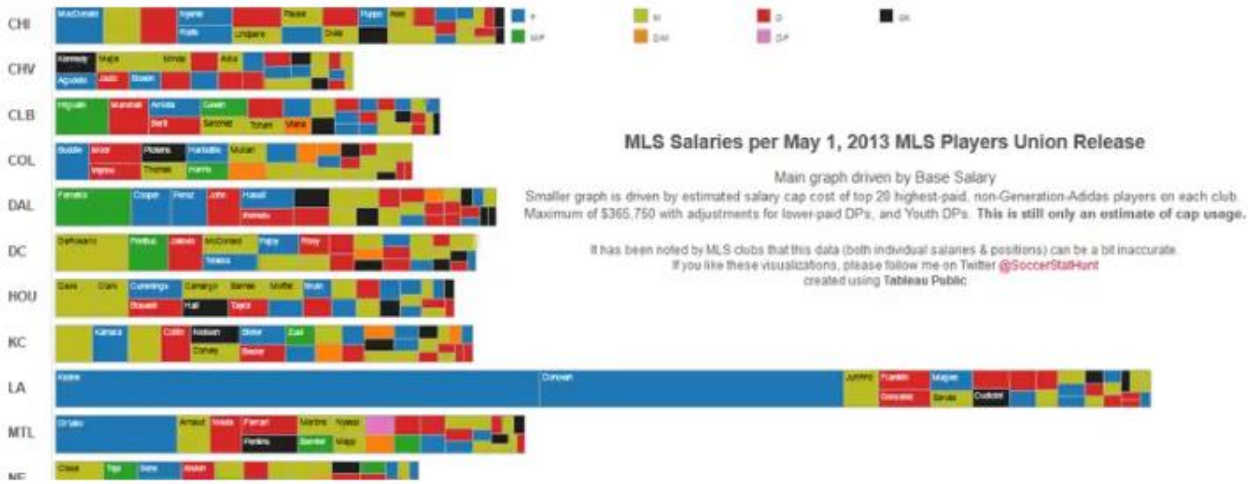
There are two improvements I can suggest to make this visualization even better than it already is:

One improvement would be to add a few more colors to distinguish the demographics better, as this was discussed in class on the topic of modifying data points. Even though the visualization is aesthetically good, it lacks some extra visual cues that can be very beneficial to understanding the various countries listed.

A second improvement I can suggest is on the topic of reproducibility discussed in class. Even though it's not required for the maker to do that it would be very beneficial to recreate this graph, to some degree, in an open-source software such as Python or R. That way, people can learn how to build a great graph and potentially improve it.

Bad data visualization – MLS Salaries

Link:
https://www.businessinsider.com/the-27-worst-charts-of-all-time-2013-6#i-never-thought-it-was-possible-but-i-actually-understand-soccer-less-after-looking-at-this-chart-3



There are several reasons why the above visualization is bad:

First, looking at the quality of data visualization discussed in class, the visualization is not aesthetic at all. There are way too many colors, shapes, and boxes of data which does not seem to tell us anything concrete. Then, the graph is not substantive, and you can even read that in the second paragraph "… data can be a bit inaccurate". Having unreliable data is a big red flag in any visualization. Other than knowing what the graph is about by reading the title, there is little else to understand about the various bars and legend.

Second, based on the anatomy of a figure discussed in class, this graph lacks several features. For starters, there is no x-label defined as the y-label is not intuitive, it does not contain proper labels. Also, not having a grid makes it almost impossible to compare bars as well as not having any major ticks to help us understand how much salary each bar tops. Then, the legend is poorly located and described which makes it even harder to understand each bar.

Third, looking at how bad this graph is based on its dependencies discussed in class, the graph lacks context. It is only understood that it is trying to show the various salaries, but it is hard to infer where and how the visualization will be used. Finally, the data structure is very poor. Emphasizing again that the data might be inaccurate makes us question whether the maker performed due diligence when collecting, parsing, and modeling the data.

There are three improvements I can suggest to make this visualization much better based on the "Data Visualization" best practices article from the University of Buffalo Libraries:

First, determine your purpose. The article mentions that understanding your purpose can help you decide what data to include in the visualization. This graph does not reflect what it's trying to convey and who the intended audience is. Also, the maker has to decide how to properly tell a story, perhaps by using more intuitive labels and legends to highlight trends.

Second, the maker needs to know their data better. Given the honesty of the maker, we can infer that the data might be incomplete, unclean, and retrieved from not-so-credible sources. To remedy this, the maker should either replace his data sources altogether or do further research to confirm the data sources are credible and complete. The main goal here is as it is laid out in the article: "You know where it came from and how it was created".

Third, choose the correct chart or visual. Having a messy bar chart does not accurately represent the data. As the article suggests, several sources can help the maker decide which chart type to choose. For instance, a better graph option can be to break down the data into various charts broken down by state or region.

Reference:

https://research.lib.buffalo.edu/dataviz