Chart 1

Referred to as Chart 1 or 'Boxplot' in the writeup on page 3.

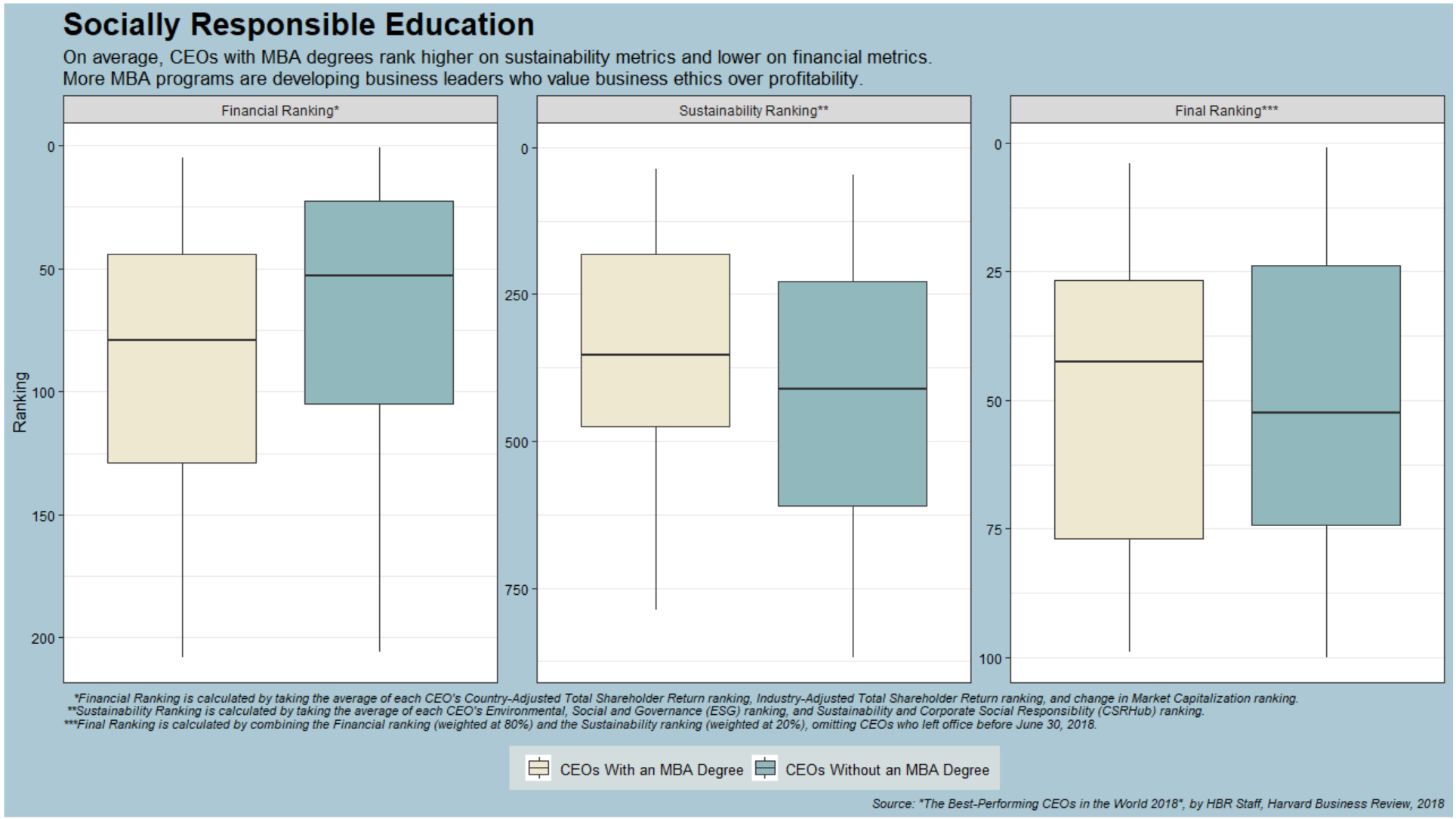
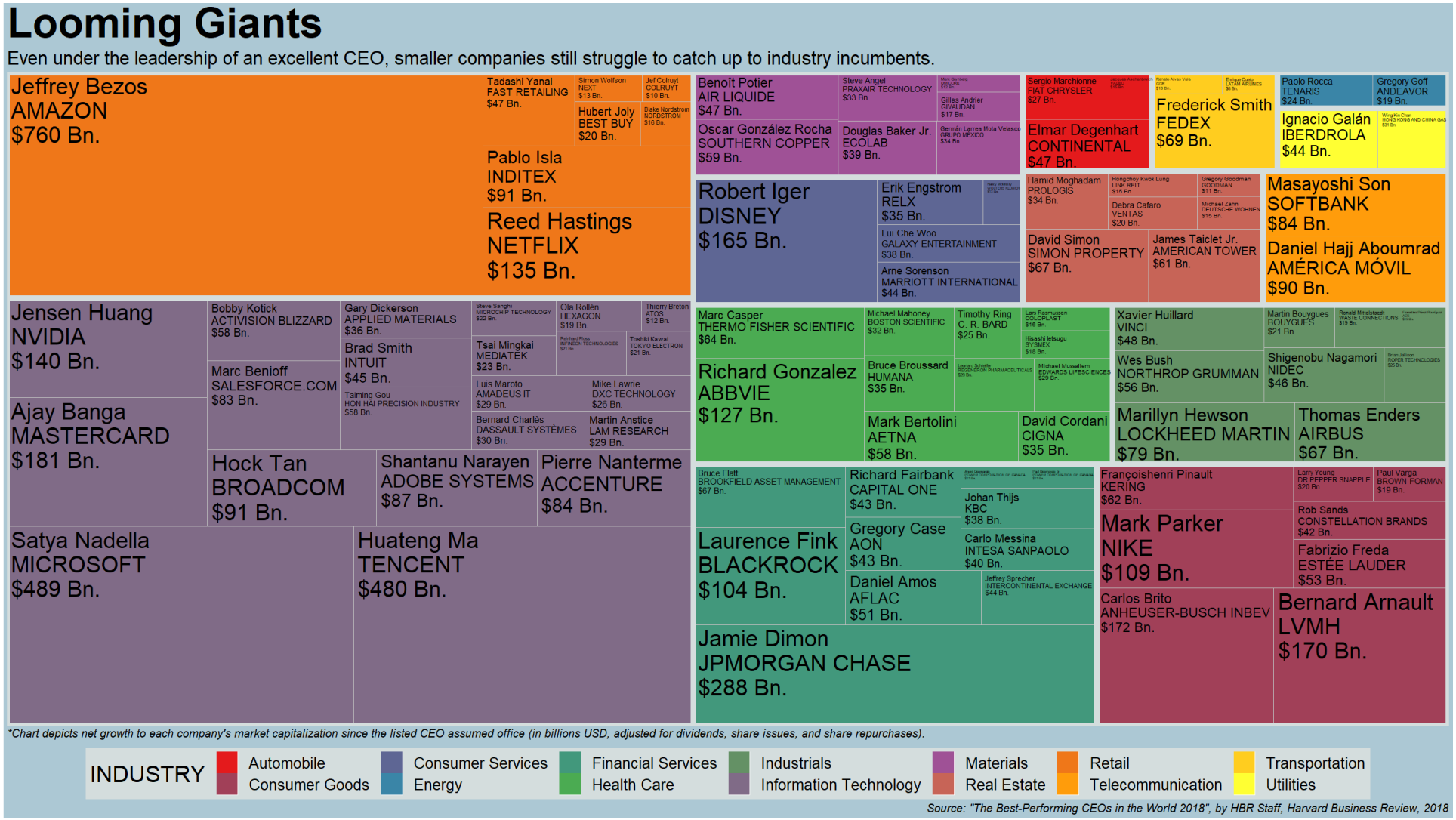


Chart 2

Referred to as Chart 2 or 'Treeplot' in the writeup on page 3.



Carl Xi, Data Visualization, Homework 1

Inspiration, Data & Design Discussion

Chart 1:

While reading the article, the first thing I noticed was that whether each CEO had an MBA degree or not was listed as a variable. I became intrigued and wanted to find out if there was a correlation between the various rankings and receiving formal masters-level business administration education. Under the methodology section of the article, I learned that the financial rank is calculated based on the average of the other four financial related ranks and that the final rank is calculated by combining the financial rank (weighted at 80%) and the two ESG ranking (weighted at 10% each). To keep my chart simple, I decided to only retain the final and financial rankings from the original dataset and combine the two ESG rankings into an overall "sustainability rank" by taking their average. This will enable the reader to easily cross-compare between the two components of the final ranking while still having the final ranking for reference. To best showcase the interquartile range and mean of CEO rankings with and without MBA degrees, I decided to use boxplots, as it delivers both the range and mean in a very simple and elegant manner. I made sure to vertically invert the boxplots as we learned in class that we naturally assume up is and down is bad. The final overall aesthetic choices are listed below.

Chart 2:

After getting the dataset doing some simple data cleaning, I conducted exploratory data analysis. Immediately, I noticed that almost all of the financial variables have really large ranges and standard deviations. Taking a close look, I noticed that even though we are looking at the top 100 CEOs around the world, the top performers financially represented most of their respective industries. I decided to create a chart that illustrates this overshadowing effect. Between the two total shareholder return percentages and change in market capitalization, I decided to go with the market capitalization variable as I can represent each company's gain in market capitalization as a part of the total wealth generated by these 100 CEOs. I decided to make a treemap with each box representing one CEO to help me demonstrate the overshadowing effect of the industry leaders. Instead of using color to convey size, I decided to use color to categorize the different industries, as adding industry labels into the treemap proved to be very messy. While the color palette of my second graph (TreeMap) is chosen by design as it is nominal and is very effective in helping the reader distinguish between the 14 different industries while still keeping the text legible. We learned from in class that for treemaps precision is not important and that treemaps are good for large data sets where the smallest categories are still relatively significant. This is perfect for our dataset as these are the top CEOs around the world we are illustrating. With this said, I decided to further convey this contrast in company size between small and large companies by **intentionally keeping** the labels on the smaller companies despite them being difficult to read. The final overall aesthetic choices are listed below.

General aesthetic/artistic choices:

The main goal of my charts is to convey spatial comparison neatly and elegantly. I have been reading The Economist magazine for a long time and like the design of their graphs. However, I decided to mimic the theme and create a personal modified version of it instead of directly applying theme_economist() in ggplot, as this gives me more flexibility in tailoring specific aspects of my graph to my liking and as needed. I was able to find The Economist color scheme from their website and used it for my graphs. The background is a soft light blue to be gentle on the reader's eyes, while all texts are fully black to ensure contrast, legibility, and ease of reading. Both graphs have a "hook" title to draw the reader in, while the subtitle underneath describes the message that the graph is trying to convey. Technical aspects of the graph (how the numbers are calculated, legend, and data source) are all listed beneath the graph for ease of reference. Lastly, I made sure to remove all "chart junk" (e.g., unnecessary tick marks, axis lines, titles, axis labels), and go through the class slides to amend my chart as needed (e.g., on principles of perception