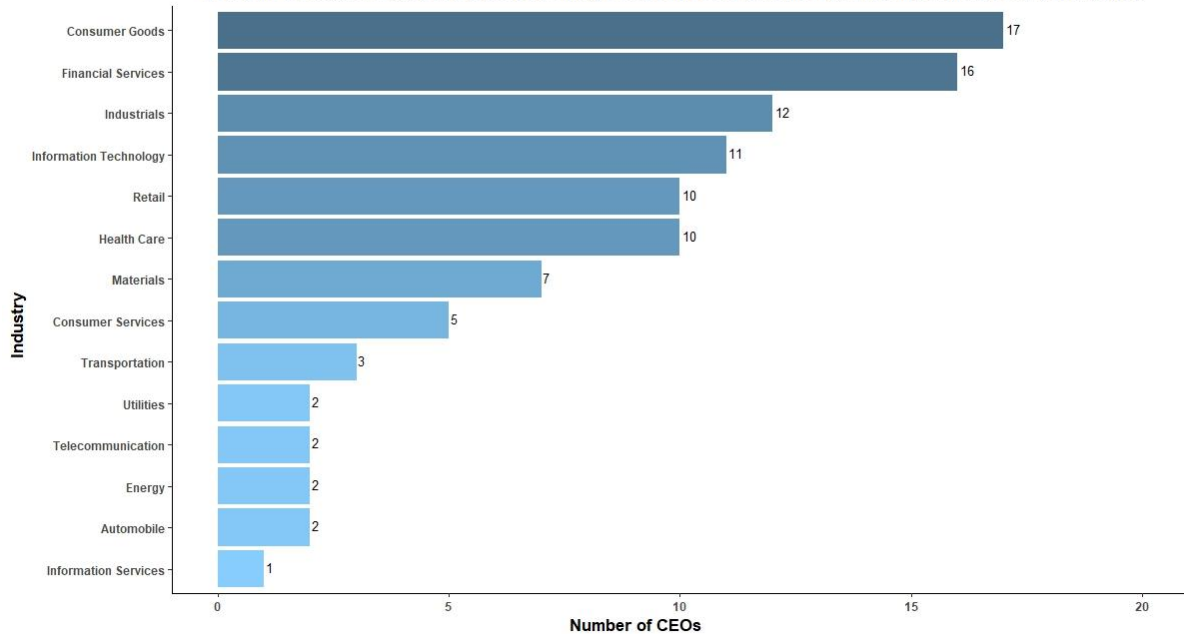


ISOM 675 Data Visualization – Assignment 1 (Agnes Liu)

Top 100 Best-Performing CEOs by industry, 2016

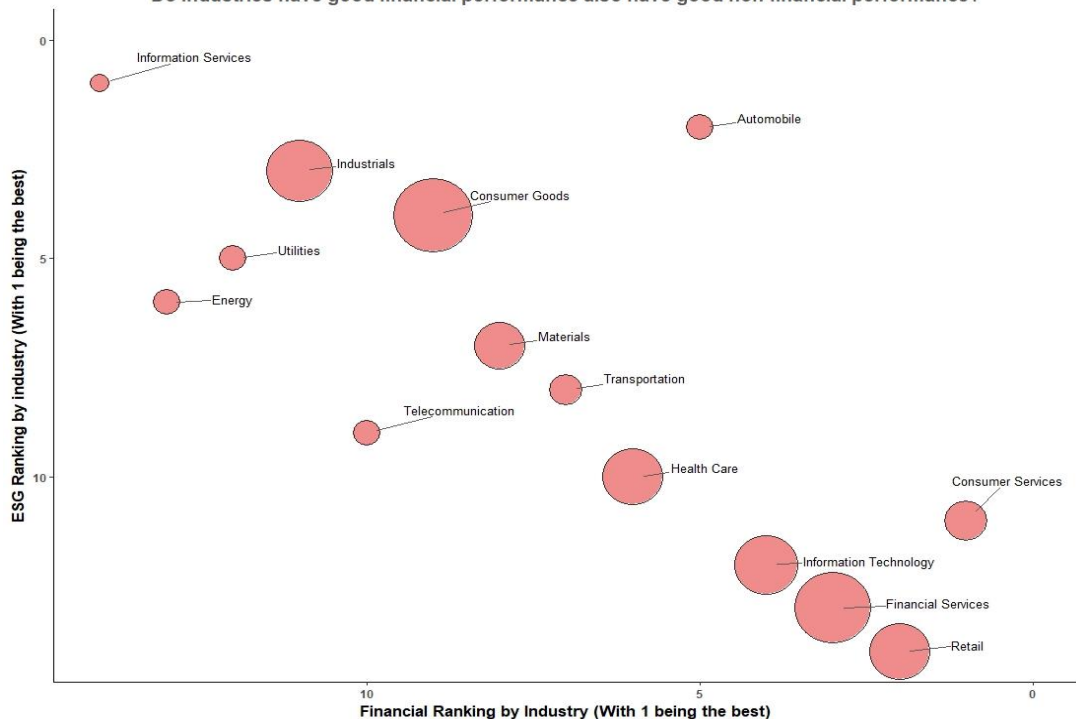
Harvard Business Review compiled a list of world's top 100 performing CEOs from 14 industries in 2016. Based on the data, one-third of top 100 performing CEOs come from Consumer goods and Financial services industries.



Data Source: Harvard Business Review

Financial Performance vs. Non-Financial Performance by industry, 2016

Do industries have good financial performance also have good non-financial performance?



Data Source: Harvard Business Review

About graphs:

Harvard Business Review compiled a list of the best-performing CEOs in the World in 2016. To compile the list, HBR obtain data from 895 CEOs based in 32 countries. Here, I created the above two graphs based on the data of the Top 100 best-performing CEOs. My main objective is to compare the performances of 14 industries and examine the relationship between financial performance and non-financial performance based on the industries.

The horizontal bar chart gives information about the number of Top 100 best-performing CEOs by 14 different industries in 2016. From the graph, we can tell that both Consumer Goods and Financial services industries have the most best-performing CEOs while information service. **The bubble chart** illustrates the relationship between financial performance and non-financial performance which based on ESG (Environmental, Social and Governance criteria) by industries. According to the chart, there is a negative linear relationship between financial non-financial measures of performance by industries. It was interesting to see that industry that did well financially might not have good environmental and social performance.

About designs:

For the first graph, I choose bar chart since I have one continuous variable and one categorical variable. I decided to use horizontal bar chart because I have longer label names and I want to show the order more clearly. I use darker blue for more values and lighter blue for less value so it's easier to see the order and the difference in numbers. In additional, I marked the exactly number of CEOs by the bar so readers can compare different industries by numbers. Bigger size for title highlights the main topic of the chart. I used dark grey and median font size for subtitle since I have longer sentences and I didn't not want it to take the attention for title. Additionally, I set small size of caption for data source because it is relative unimportant compare to other information on the graph. In R, I used the original dataset and cleaned NAs in the dataset. Then, I created another dataset using `count()` from library `plyr` to get the number of CEOs by industries. The graphs and designs are created using library `ggplot2`.

For the second graph, I choose bubble chart (3 variables) since I want to show the relationship between two different performance but also visualize the industries by the number of CEOs (readers can relate to the first graph for more information). The bubble size also indicates that information on industries that have small data might not be so accurate. For bubble color, I used coral instead of blue, green or purple color that's similar to the color of the first graph so it's easier for readers to separate the messages from two different graphs. For readability, I flipped the range of axis (with 1 being the best) and used segments to connect bubbles and industry names. For label and segment color, I used dark grey to weaken the effects since they are used as reference. The design of titles, subtitles and caption is similar to that of the first graph. The dataset I used for this graph was created by Excel. In Excel, I combined measures of Sustainably and CSRHUB for non-financial performance since they are both used as ESG. In additional, I also ranked average financial measures and average non-financial measures based on industries so the difference in range would not interfere the relationship we are trying we see from graph. In R, I utilized library `ggplot2` and `ggrepel` for graph and design. Specifically, I used `geom_text_repel` function in `ggrepel` to repel overlapping text labels away from each other and away from the data points that they label.