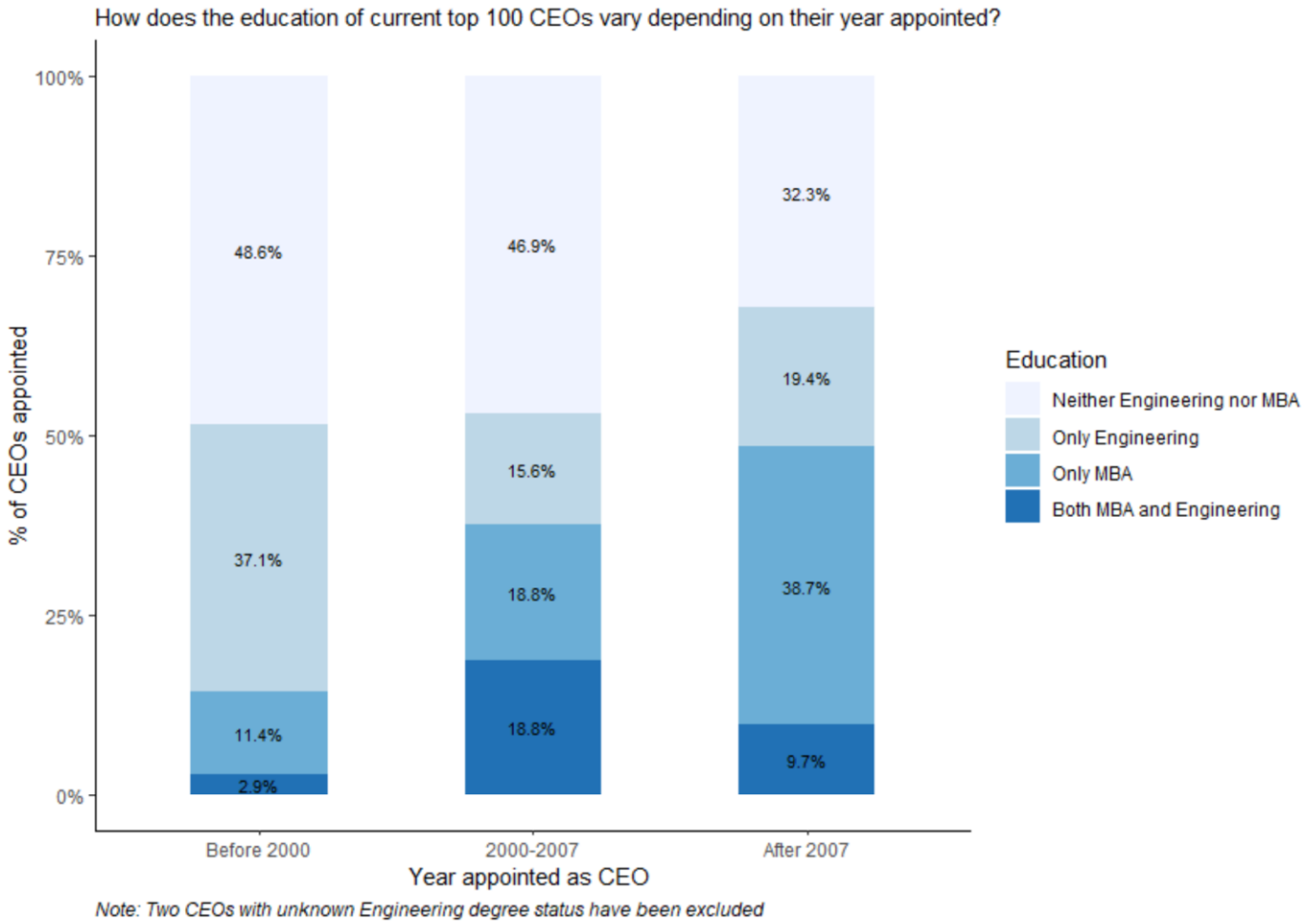
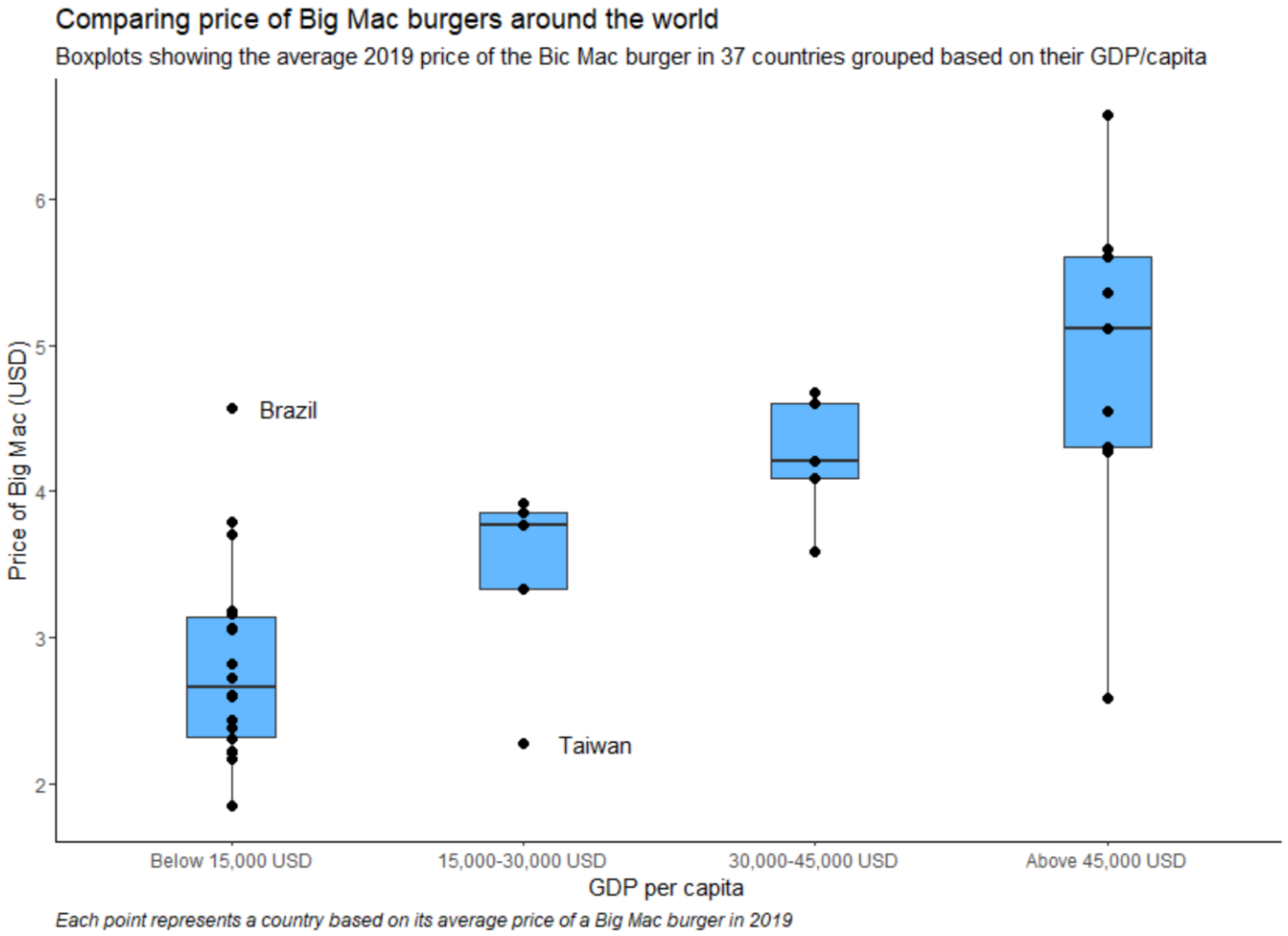
# Data Visualization- HW1




- The above stacked bar chart represents how the Education (degree) of the top 100 CEOs around the world has varied depending on the time (year) when they were appointed as the CEO of their respective companies. **We see that almost half (~48%) the CEOs appointed <u>after</u> 2007 do have at least an MBA degree as of 2018. This is a much higher proportion than top CEOs appointed before 2000 (~14%) and those appointed between 2000-2007 (~38%)**. Moreover, we see that this increase over time is attributed to the decrease in the proportion of CEOs with only an Engineering degree or those without either of those two degrees
- Choice of visualization: I chose a bar graph as I wanted to show the change in proportion over groups of discrete time intervals. Combining groups of time intervals would not be interpretable using a Line graph. Using a stacked graph allowed me to compare proportion variations over time by comparing the height of a group (shown using same intensity) over the x-axis. Further, I decided to use the proportions instead of raw counts as they are easier to interpret, especially when each bar represents similar counts. **By grouping the years in 3 buckets as I did above, I was able to ensure that each bar height represents ~1/3$^{rd}$ of the top CEOs (35 vs. 32 vs. 31), an n-size large enough for the proportions to make sense and see trends/variations over buckets of time. Using different intensities of the same color, I was able to show degree variations within the same category with higher intensity representing higher education (e.g. both MBA + Engineering has the highest intensity, while neither has the lowest).** I have shown labels (percentages) for each bucket to make interpretability of the proportions easier and also allowing for quick comparisons.
- In terms of data preparation, some data cleaning was necessary. I did not the Engineering degree status for two CEOs (shown as N/A in raw data), so I left them out of the visualization for a fair comparison and used a caption in italics to make it known to the viewer. The title of the graph would allow for the viewer to set expectations for

what's about to come: data only for the world's top 100 CEOs. I also had a make the variable "Education" using a combination of the two binary variables "Engineering" and "MBA". **The major insight indicates that more and more top CEOs tend to have an MBA degree in the recent years as compared to the past, and fewer have neither MBA nor Engineering backgrounds** (but does not necessarily mean one needs an MBA to be a top CEO)




- The above set of boxplots show the distribution of the price of the Big Mac burgers around the world. To allow for a fair comparison, I have grouped the set of countries with similar GDP/capita together, i.e., showing distribution of big mac prices for countries with low GDP/capita (below 15k USD) and likewise for higher GDP/capita groups. The above plot shows few major insights: **1)** The price of the Big Mac burger (in USD) tends to increase as the GDP/capita of a country increases. This can be seen in the increasing height and placement of the mean line of the blue box as we go from left to right .**2)** As always, there are a few exceptions. Brazil, for instance, has a very high price when compared to countries with similar GDP (it is a clear outlier, so much so that its price is almost same as the mean price of the countries with GDP/capita of 45k+). Taiwan, a country with relatively higher GDP/capita has a fairly inexpensive burger (in USD), much lower than other countries in its category. Hong Kong is another country with a low price in USD (the lowest black point in the 45k+ GDP/capita group) but is not pointed out with annotation (label) as it is not an outlier.
- Choice of Visualization: The visualization has two major aims: a) Showing the distribution of big mac prices, and b) Allowing for price comparison across multiple groups of countries based on GDP/capita. The two most obvious options were histograms and boxplots. I chose side-by side boxplots rather than multiple histograms because the data was distributed in a way which showed the increasing trend clearly and

allowed me to point out the outliers within each group individually as prominent points. The graph can be loosely interpreted as a fairly obvious price comparison between poor, moderate and rich countries I also chose to show each country as a point in the boxplot as there are some categories with limited n-sizes (for instance, the 2nd and 3rd categories have five countries each, while others have 10+). The GDP/capita scales are chosen to be uniformly distributed (<15k, 15-30k, 30-45k, 45k+) to allow for fair comparisons of categories with at least 5 countries each. The outliers have been labelled with the country names for interested viewers to explore more about why their prices seem unusual (read: Why brazil's big mac seems so expensive when converted to USD)

- The visualization required extensive data prep. For instance, I decided to show the 2019 price for each country by taking a mean of the January 2019 and July 2019 prices available in the dataset. Further, there were only 37 countries where data for both Big Mac Price and GDP/capita information was available. Hence, countries with limited information were excluded from the analysis by filtering. These facts have been pointed out in the visualization title/caption at the bottom.