**CS 3300: Project 1 Writeup**

Danyal Motiwalla (djm453), Rachel Yunyann Liu (rl544), Paula Moya Nieto (pm445)

Datasets

Our main dataset was a collection of world leaders' social media data, obtained from CupoNation (cuponation.com/world-leaders-social-media/), recently released on Jan 22, 2018. The dataset included the names of the leaders, the regions they were from, their gender, age and total followers. There was an additional breakdown of the number of followers based on social media platforms, such as Facebook, Twitter, Instagram, YouTube and Linkedin. The dataset also included weblinks to their social media platforms, which we removed as we began to create our own compiled dataset.

After choosing to work with this dataset, we discussed how we could build in more data to make our visualization thoughtful and interesting. We decided to investigate trends related to population or approval ratings data. We found our population dataset from the World Bank (2016) World Development Indicators: (https://data.worldbank.org/indicator/SP.POP.TOTL) and UN Population Division, World Population Prospects (https://esa.un.org/unpd/wpp/DataQuery/) and our approval rating dataset from Gallup World Poll (2017) (https://analyticscampus.gallup.com). Additionally, we knew we wanted to incorporate a map into our visualization as it could be interesting to see which regions have leaders with a higher numbers of followers. We projected our map using a shape file (obtained from https://github.com/d3/d3-geo/blob/master/test/data/world-50m.json) outlining all the countries of the world. We cropped this within the svg to exclude Antarctica. To identify specific points on the map, we chose to use a dataset with the coordinates of all capital cities of the world, obtained from Techslides (2013) (http://techslides.com/list-of-countries-and-capitals).

Using the merge data feature of Excel, we started combining these datasets into a dataset that could be easily transformed into a json file, with information that we were highly likely to manipulate in our visualization. Since we were matching datasets based on country names, the hardest part was ensuring that those names were consistent across datasets. Additionally, we had to ensure that missing data would appear as null in the json file. Ultimately, we produced a json file which included an array of objects, with keys including country, region, world leader, status, gender, age, official title, total followers, twitter followers, facebook likes, instagram, followers, approval rank, approval ratings, population in 2016 and the log of that population, capital name, capital latitude and longitude, and country code.

The Map and Total Followers

The first story we wanted to tell was where all the leaders with social media accounts were located, so that viewers could more easily relate. Using a map projection, we plotted where all leaders with followers were located, based on their country's capital city. We noticed that while they were spread out amongst all the continents, there was a high number of leaders in Europe with followings, while in Africa the leaders with social media accounts tended to be along the coast lines.

Secondly, we wanted to see which parts of the world had leaders with the highest number of followers. After identifying the top 10 leaders, we enlarged those specific points on the map such that these leaders' total followers were proportional to the radii of the circles (using a square root scale). The leaders were

distributed fairly evenly across the globe, prompting us to analyze the data further to investigate possible trends.

Graph of population size and approval ratings

We noticed that the leaders with the highest total followers typically came from large, prominent countries in the world. Hence, we decided to investigate plotting a graph with a y-axis of population. We used a log scale (done earlier through excel) to measure population to spread the data more evenly. Since we thought it would still be interesting to differentiate the top 10 leaders from the rest of the world, and to allow a continuous shift in reading and understand of our visualization, we decided to choose longitude as the x-axis. That way, we could chart the circles identifying every leader from the map to the graph. The radius of the each circle would still be square root proportional to the total number of followers, for the top 10 leaders. We decided to incorporate one more variable - approval ratings - into the graph. We felt that it would be interesting to see where the leaders, especially the top 10, stood with regards to approval ratings, even though they had a high number of followers. This would be done through an opacity gradient of the points, where a higher opacity corresponds to a higher approval rating. For the countries we could not obtain approval rating data for, we gave the circles full transparency with a grey stroke.

We noticed a few interesting trends. Firstly, we noticed that the top 10 leaders typically came from a country with a higher population, except for Pope Francis, who was a clear outlier due to his global influence but small resident population of the Vatican City (Holy See). Amongst all the leaders, the spread was even amongst different population levels. There wasn't a clear pattern amongst approval ratings and population, but it was interesting to note that there was quite a range amongst the top 10 leaders. For example, Modhi (number 1) had a high approval rating, while Trump (number 2) had a low rating. Trudeau seemed to be the only leader in the Americas with a high approval rating.

Graph of age and social media platform

Since social media is usually assumed to be a "young person's" activity, we decided to investigate how the age of the leaders would relate to their number of followers. The y-axis represents a linear scale of age, while the x-axis continues to represent longitude, maintaining vertical visual continuity. Additionally, the color of each circle corresponds to the leader's most followed platform (among Facebook, Twitter, and Instagram). We used the official hues of color used by the companies, for easy recognition. Furthermore, the radius of each circle still was the number of followers using a square root scale. We found out that age was widely spread amongst all age groups. What was particularly interesting was that the leaders with the largest followers (Trump, Modhi, Pope Francis), tended to be older than the median age. Additionally, amongst the top leaders, Twitter seemed to attract more followers than Facebook.

Breakdown of Top 10 Leaders Followings

Lastly, to further delve into the platforms of social media world leaders use, we plotted bar charts with a specific focus on the top 10 leaders. We wanted to see how popular each of the leaders' social media platforms were. We once again used the typical color hues of the platforms, and broke them down in a graph where the y-axis was number of followers and the x-axis represented the leaders' names in the order of their capital's longitude. We noticed that their Twitter accounts had a larger following, but

Facebook was always a close second. Instagram had a much smaller percentage of total followers. Only President of Indonesia Joko Widodo had an even distribution amongst all 3 platforms.

Resources Used in the Code

Code obtained or inspired by external sources are referenced in comments within the code. In some cases, we encountered difficulty placing this code in an external file because it was a challenge to reference d3 in a .js file. For the sake of transparency, we are highlighting these code sources here:

The map was constructed with code from David Mimno's class on maps and projections (Friday, February 16, 2018).

The legend of opacity gradient for approval ratings was adapted from Darren Jaworski (2015) (http://bl.ocks.org/darrenjaworski/5397362).

Abhisek Jana's explanation was used as a guide for building the stacked bar graph (http://www.adeveloperdiary.com/d3-js/create-stacked-bar-chart-using-d3-js/).

Mike Bostock's wrap function was incorporated in-line (as it referenced d3) to wrap the bar graph's x-axis labels (https://bl.ocks.org/mbostock/7555321).

Susie Lu's legends library was used for the Age scatterplot and stacked bar graph (http://d3-legend.susielu.com/).