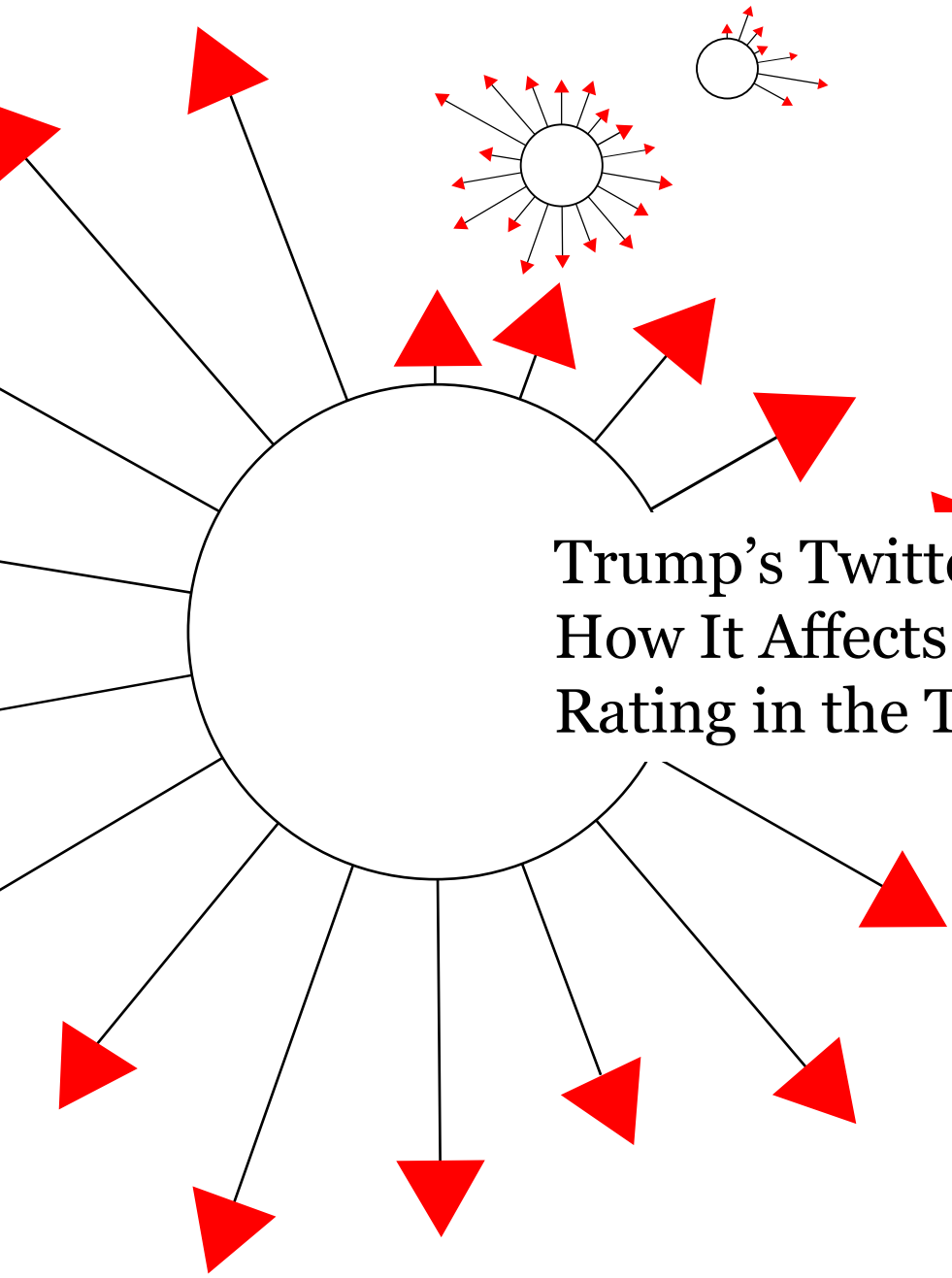
FALL
2020

FINAL
REPORT



Trump's Twitter Presence and How It Affects His Approval Rating in the Time of COVID-19

by Lauren Li

DGM 6108
Northeastern University

Overview

I harvested almost a years worth of data to try to see if President Donald Trump's Twitter activity (number of daily tweets and retweets) has any correlation with his approval rating.

I also compared his approval rating with economic factors such as major stock market indices, submitted unemployment applications, and national gas prices. Since we've been living in a pandemic for most of the year, I also recorded daily new and confirmed COVID-19 cases in the U.S. My goal for collecting this supplementary data was to see which of these factors might have the strongest correlation with Trump's approval rating if I failed to find one with his Twitter activity.

About the Data

All of my data is available for download here:
http://digitalmedia.neu.edu/li6108/final/finaldata.xlsx

In order to be as unbiased as possible, I collected approval rating data from RealClearPolitics.com. Real Clear Politics reports an aggregate of Trump's approval ratings everyday. Their aggregate is calculated based on polling data from websites such as Rasmussen, NPR/PBS, The Hill, etc ("President Trump Job Approval").

Trump's Twitter activity proved to be a challenge to tally on Twitter.com, but I was fortunate to find a website called TrumpTwitterArchive.com, which later changed their domain to TheTrumpArchive.com. Both websites check and record Trump's tweets every 60 seconds, and even keep Trump's deleted tweets as long as they've been archived (Brown, "FAQs"). As you'll see in my data, I separated regular tweets from retweets by distinguishing retweets with records starting with "RT" and regular tweets as those that did not.

For economic data, I collected average weekly national gas prices (dollars per gallon) for regular and diesel gas from the U.S. Energy Information Administration's (EIA) website. These weekly prices were repeated for each day of the same week to show the appearance of daily figures.

Since the stock market is regularly referred to as a good metric for calculating the wellbeing of the U.S. economy, I also chose to record three of the most widely followed U.S. stock indices: NASDAQ Composite (IXIC), Dow Jones Industrial Average (DJI), and Standard & Poor's 500 (S&P 500).

In addition to these three stock indices and gas prices, I also wondered if increased unemployment (UI) claims had an affect on President Trump's approval rating. I decided to record total weekly submitted unemployment claims available on the U.S. Department of Labor's website. Claims are record in the thousands, which I converted to daily averages by dividing the provided figures by 7 days of the week.

My final piece of recorded data is daily new and confirmed COVID-19 cases at a national level. For good measure, I also included total confirmed and probable cases. Both figures were recorded from The COVID Tracking Project website, which seems to update their historic and current data regularly.

Recording Data

I originally recorded all of my data in an Excel spreadsheet, starting with August 1st, 2020. After recording about two months worth of data, I decided I wanted to see the full effects of COVID-19 on President Trump's approval rating and began collecting data starting from January 1st, 2020.

One thing you may notice if you look closely at my recorded figures and crosscheck my sources is that my UI and COVID-19 figures may not line up correctly with data found on these websites. While collecting data, I learned that these two sources update their figures regularly. My data was last updated and recorded on December 7th, 2020.

As you can see in the last Excel sheet labeled "All Data," I combined all of my data into one sheet. This means I converted all weekly data to daily data.

When it came to converting my recorded data to JSON, I originally entered the first few months of data manually. Eventually, I learned about CSV files and CSV to JSON converters. In order to save time and reduce input errors, I saved my "All Data" sheet as a CSV file and used www.convertcsv.com to convert my data to JSON format whenever I needed to use an updated JSON file.

Exploring My Data

I wasn't entirely convinced I'd find much of a correlation between Trump's approval rating and his Twitter activity, but I did expect to see at least some sort of correlation between one of my properties and his approval rating.

When I explored my data through scatter/bubble plots in Homework #7, two of my visualizations proved to be helpful for comparing my data. Scatterplot #1

(Figure 1) and Scatterplot #2 (Figure 2) below, share properties that I ultimately used for my final visualization.

Scatterplot #1

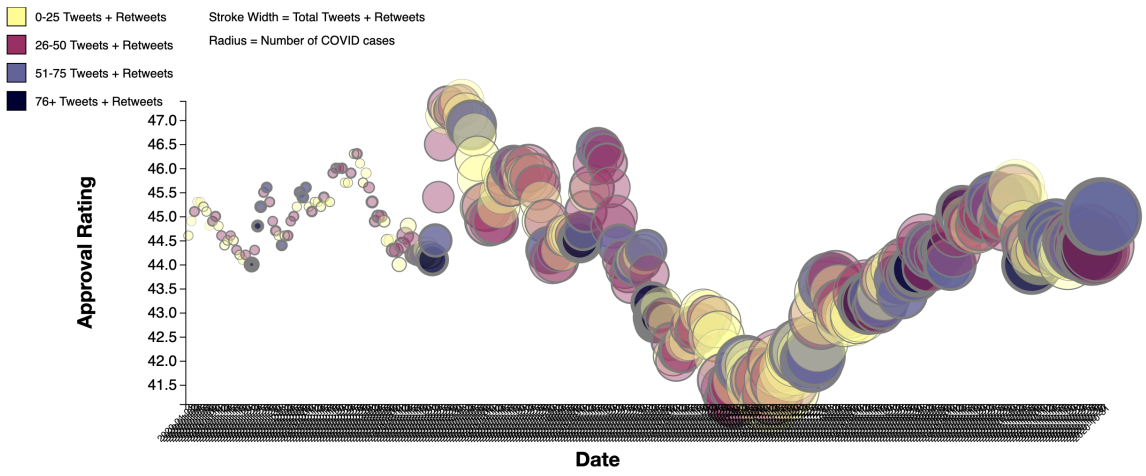


FIGURE 1

Scatterplot #2

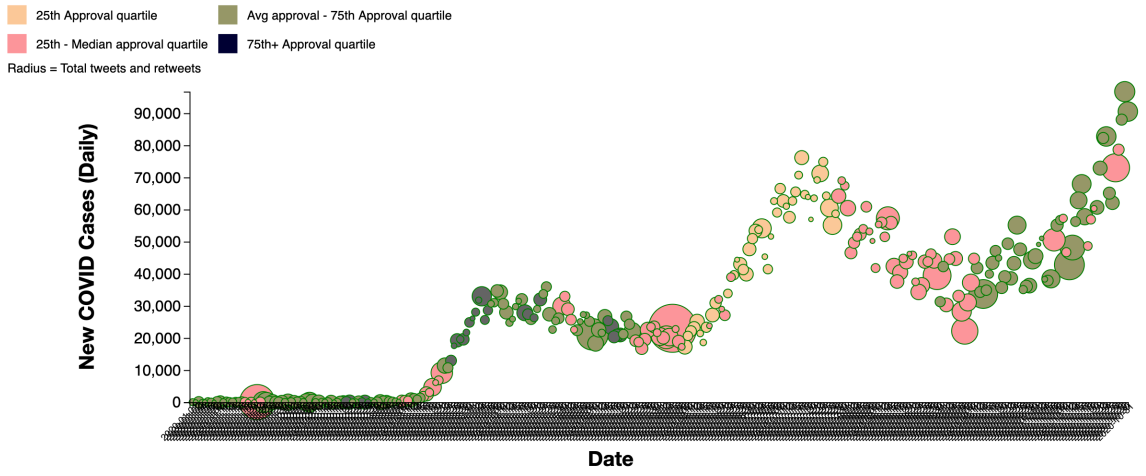


FIGURE 2

Sketches

Prior to coming up with my final visualization, I continued to explore other visualization ideas. I came up with a proposal in Homework #8, where users could change the x-axis property to whichever piece of data they wanted to compare with Trump's approval rating (Figure 3). The benefit of this feature is that radio buttons allow users to explore all of my data. This would allow them to confirm my conclusions for themselves. Separating my properties with radio

buttons also decluttered my scatterplot and solved the issue of each property needing different domains of values.

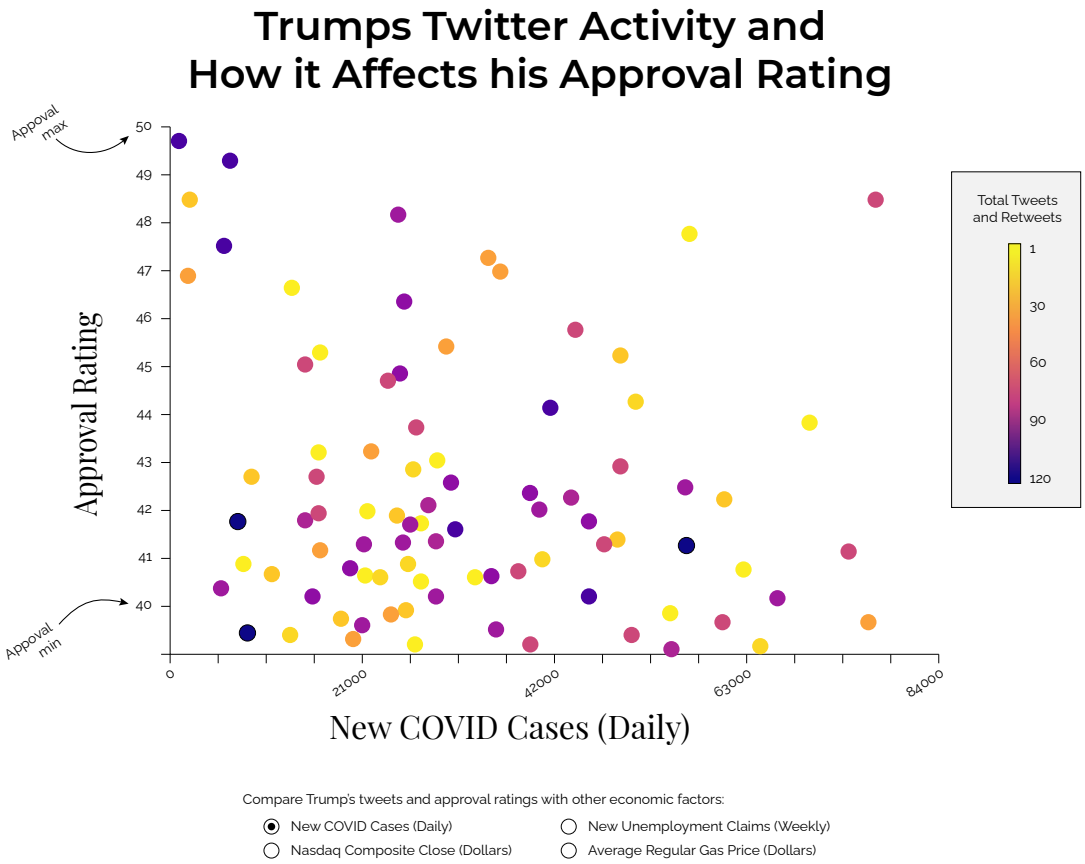


FIGURE 3

In addition to my radio button idea, I came up with another sketch to display my data in the shape of a COVID molecule (see Figure 4 below). Although it appears as if the x and y-axes have been removed, they instead have taken on a different shape. The body of the COVID molecule (large gray circle) is now the x-axis, and the length of the lines represent the y-axis. Triangles and yellow circles are the equivalent of points, or bubbles on a scatterplot.

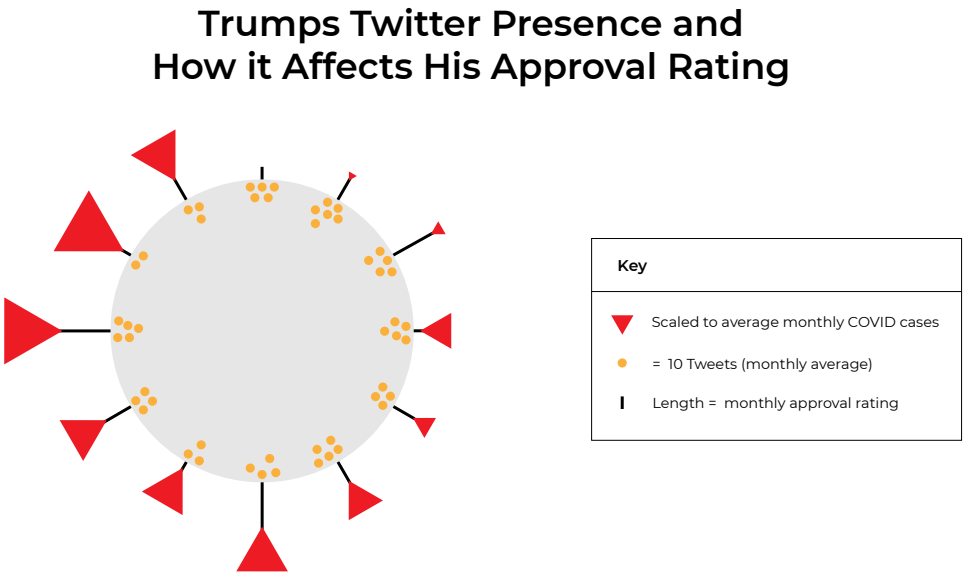


FIGURE 4

The benefit to this COVID molecule sketch is that my data, once with 300+ data points, now has only 12 main data points or COVID "spikes" (line and triangle combinations). This visualization simplifies my data substantially and remains visually relevant to my project.

While both of these visualizations have their own benefits, my radio button scatterplot lacks what I found to be an important variable discovered in Scatterplot #1 and #2 of Homework #7: the date. It also included too much information, which may give users enough knowledge to make conclusions of their own. The downside of this is that their conclusions wouldn't necessarily support my conclusions, possibly weakening my argument.

When it came to my COVID molecule sketch, that visualization almost simplified my data too much. As a result, it made comparing line length with triangle and circle scales/sizes too difficult of a feat.

In the end, I combined these two sketches with my lesson learned in Homework #7. I condensed my data and designed a bubble plot of COVID molecules, making sure to include the date on the x-axis.

Final Visualization and Process

I chose to visualize my data in the form of a creative bubble plot after narrowing down my properties. I decided to include three quantitative/numerical continuous

variables in my visualization (approval rating, total daily tweets and retweets, and daily new and confirmed COVID cases), while saving my x-axis to display months. As recommended by Claus Wilke in Chapter 2 of Fundamentals of Data Visualization, bubble plots are an effective means of visualizing data with a minimum of three quantitative variables.

In my case, these bubbles, or circles were also the shape I needed in order to draw COVID molecules. The fact that they could be scaled to different sizes and combined with various line lengths and triangles scales made them effective for representing my continuous data (Wilke).

I had a final visualization in mind, but before writing any drawing code, I had to manipulate my JSON data to fit my visualization and avoid over plotting. I started coding with a smaller set of data and drew visualization ideas from my COVID molecule sketch. I had to separate my data by month and then by category before calculating their monthly averages. Each of these steps required me to create new array data structures, which in return allowed me to customize my COVID molecules. Molecule bodies would be drawn relative to a monthly average, and each COVID “spike” would represent a day of one category of data. In the end, I also drew my COVID triangles to scale with another monthly average.

This entire process was much more challenging than I had expected. When I solved one coding challenge, I was faced with another. When I finally drew and presented my intended visualization (Figure 5), my data wasn’t clear.

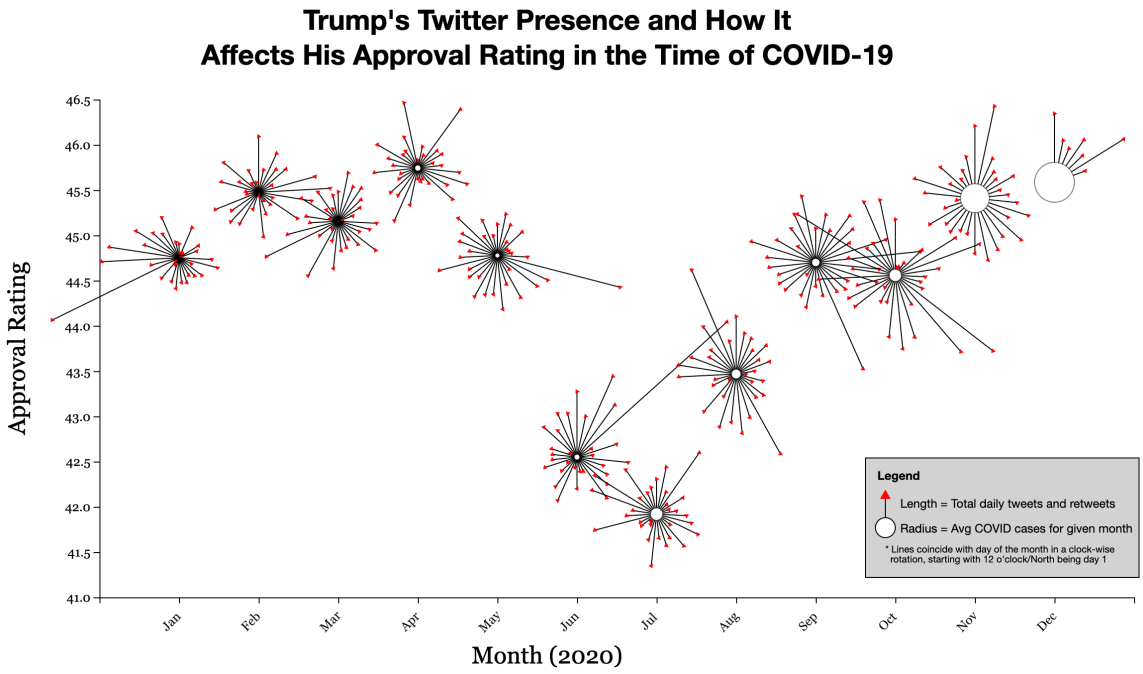


FIGURE 5

Although pleasing to the eye, it was too difficult to measure how often Trump tweeted, or which month he tweeted more often. The circle sizes, or average monthly COVID cases, are easier to read but don't necessarily help with my initial question: Is there a correlation between Trump's approval rating and his Twitter activity?

I went back to the coding board and modified my visualization one last time. The result is my final visualization below (Figure 6).

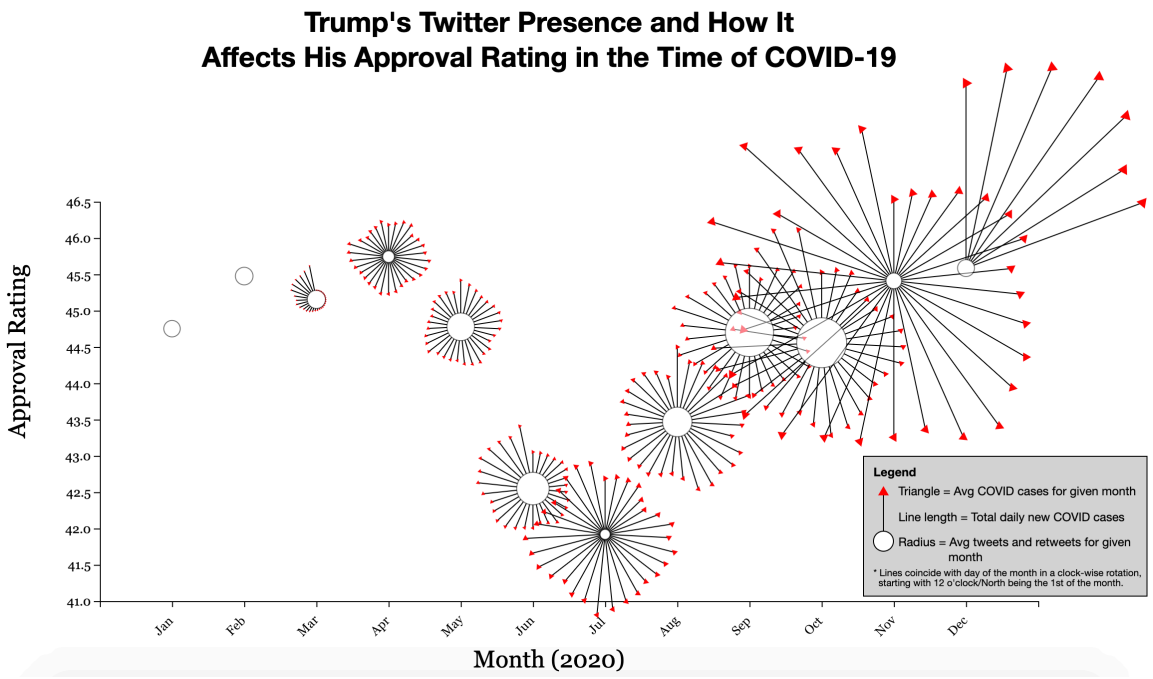


FIGURE 6

While my final visualization is more crowded between September—December, it displays the principal parts of my data (approval rating and tweets) more clearly and supports my conclusions.

Analysis

I originally predicted that Trump's Twitter activity would affect his approval rating more than my recorded economic factors and our current COVID-19 pandemic, but as you can see in my final visualization above (Figure 6), my hypothesis is unsupported. Although there are times when increased Twitter activity appears to have a correlation with increased approval rating (July—October) this observation is inconsistent.

I chose to include my COVID figures as a third variable because Homework #7, Scatterplot #2 (Figure 2), had me convinced that this piece of data had the strongest correlation with Trump's approval rating. August—September seemed to support my alternate hypothesis of Trump's approval rating being most affected by the pandemic, but as October came and went, so did that conclusion.

In more recent months, Trump's approval rating has been on the rise even with low Twitter activity and a significant increase in daily new COVID cases. I decided to emphasize my surprise and confusion about this realization by

intentionally leaving the crowded molecules of September—December as is. I wanted that part of the visualization to feel chaotic and relative to the increase in COVID cases. Somehow, while COVID-19 plagues our nation, Trump has less to tweet and receives greater approval ratings.

Reflection

Unfortunately, I couldn't find a correlation between my other economic factors and Trump's approval rating. If I had time to collect additional data for my project, I would look for sources that broke down Trump's approval rating by political affiliation. What I really expected to see was increased approval ratings from Republicans as Trump tweeted more, and the opposite from Democrats.

Prior to this assignment, I didn't track any data about the world or much about myself. Now that I've learned the foundations of JavaScript and created a visualization of my own, I feel empowered to continue collecting and visualizing data. Since I am much more interested in collecting data about the world, I have been thinking about pursuing my original final project idea about possible relationships between Boston area COVID cases, weather, and MBTA ridership. I've also thought about creating visualizations about the cost of higher education, and Boston real estate prices over time. For a smaller project, I may attempt to learn how to work with API's by exploring personal rock climbing data tracked on mountainproject.com.

Needless to say, I thoroughly enjoyed this project. It was immensely challenging, extremely rewarding, and set me on a path to continue learning about JavaScript, data collection, and data visualization.

Works Cited

Brown, Brendan. “FAQs.” *Trump Twitter Archive*, www.trumptwitterarchive.com/about.

---. “FAQs.” *Trump Twitter Archive V2*, www.thetrumparchive.com/faq.

Wilke, Claus O. “Fundamentals of Data Visualization.” O'Reilly, O'Reilly Media, Inc., 27 Mar. 2020, learning.oreilly.com/library/view/fundamentals-of-data/9781492031079/ch02.html#aesthetic-mapping.

“President Trump Job Approval.” *Real Clear Politics*, 2020, www.realclearpolitics.com/epolls/other/president_trump_job_approval-6179.html.