

Final Project Reflection — SARC 5400

Ronith Ranjan's COVID Counties Data Visualization

Background/Motivation

In our conversations about tolerance, the largest controversy that sticks out from this year of the pandemic is how differently people reacted to the social guidelines enforced by both government and private businesses. Responses ranged from public and loud anti-maskers to those all willingly to hunker down for a longer period of time to those with constantly evolving views. People's reaction likely differed based on what type of jobs they had, how many elderly people live with them, how hard their community was hit, and probably a dozen other factors.

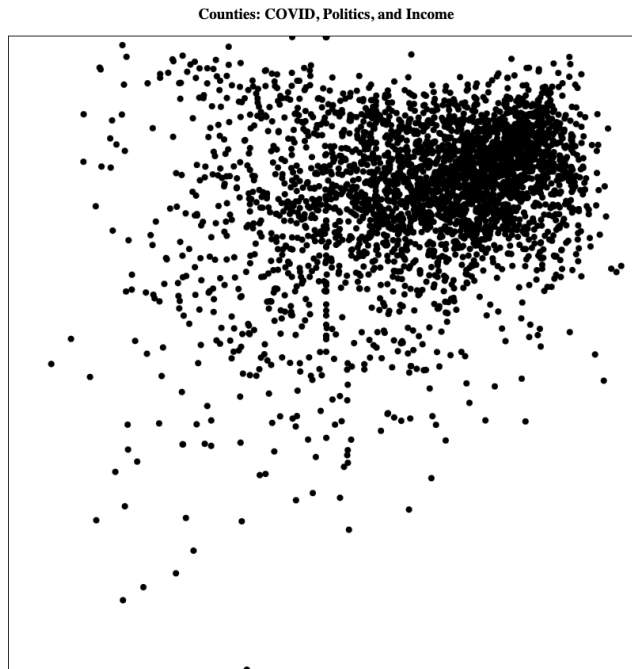
The question I was trying to get at was seeing the relationship of different factors that ultimately led to a difference in COVID cases. Some of the factors I was considering were vaccination data, education level, mobile data showing movement, and diversity. The three factors that I settled on were political views, median income, and population. I chose counties as the data points to explore because it provided a more specific lens than states, which would hide large differences in political and economic stratification, but large enough that the data would be consolidated and available.

Strategies and Data Exploration

Info	Source
COVID Cases	https://github.com/nytimes/covid-19-data
FIPS Codes	https://www.nrcs.usda.gov/wps/portal/nrcs/detail/national/home/?cid=nrcs143_013697
Median Income and Population	https://www.ers.usda.gov/data-products/county-level-data-sets/download-data/
2020 Election	https://github.com/tonmcg/US_County_Level_Election_Results_08-20
Mask Use by Survey	https://github.com/nytimes/covid-19-data/tree/master/mask-use

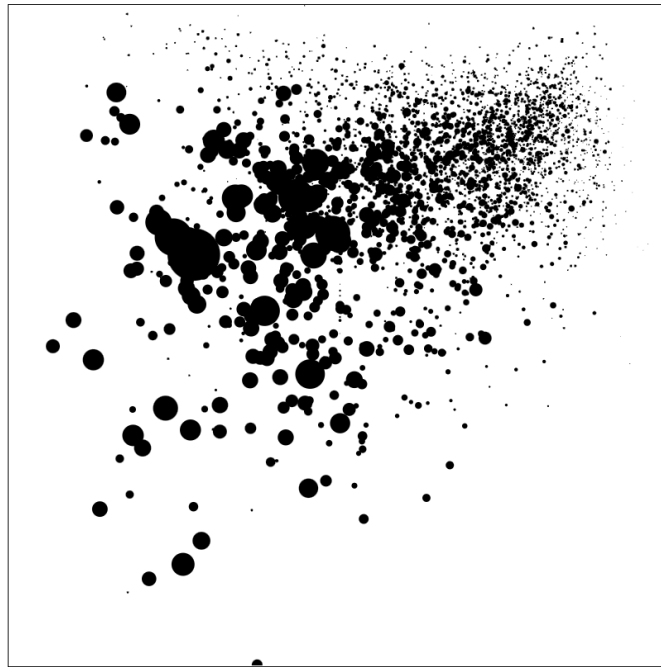
I spent a good portion of my time watching the panopto lectures and ensuring that I understood the basics of D3. Quickly thereafter, I was able to get all the counties plotted by the political differences (% GOP - % DEM) for the X axis and median income for the

Y axis. I manually used excel to figure out the range and only later found out I could have used extent to speed up the process. The values were scaled to fit the entire length of both axes.



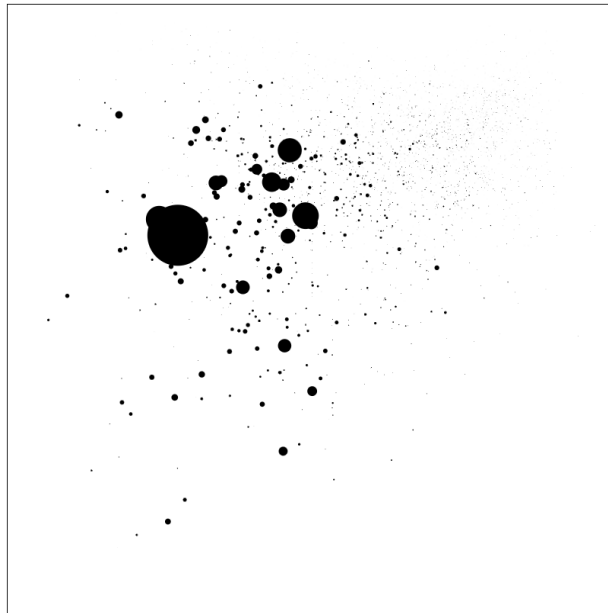
I then wanted to see the circle area based on the population so inputting $\sqrt{\text{population}/\pi}$ before scaling it linearly. Already we can see some pattern we were expecting, namely that the larger counties tend to lean to the left while the smaller ones are skewed to the right.

Counties: COVID, Politics, and Income



The next graph I created was solely based on the COVID cases but I forgot to do the same thing as I did for the population in terms of using the number as an area and then converting to radius. The radius here is just a function of the total COVID cases.

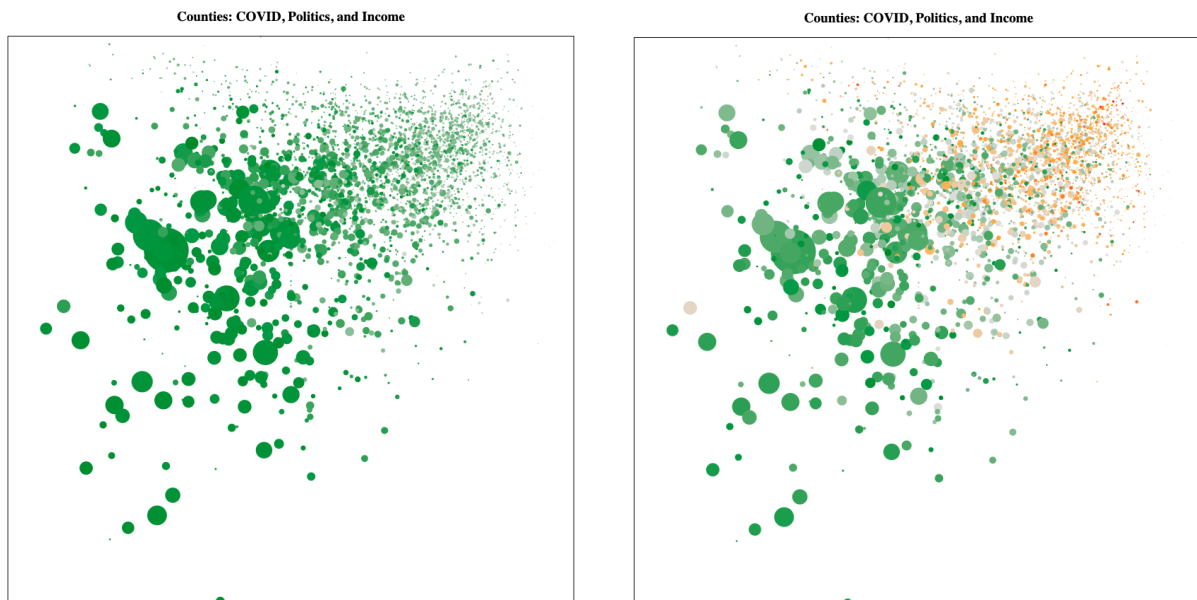
Counties: COVID, Politics, and Income



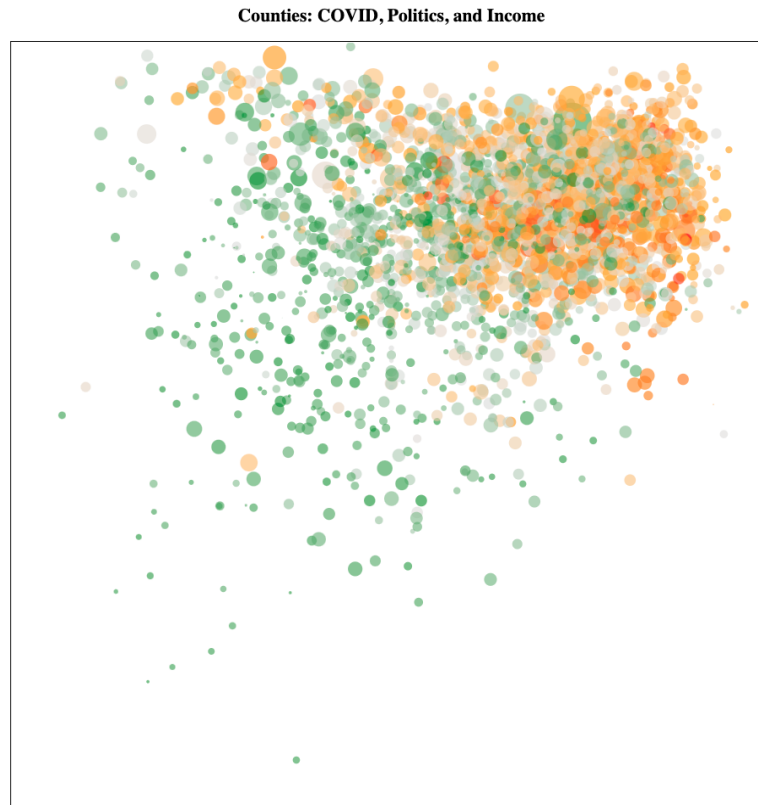
I changed the colors of the dots to be based on the COVID mask use data which uses the equation below and then is scaled linearly onto a color between green and orange, where green represents higher mask use.

```
let res = 4 * obj["ALWAYS"] + 2 * obj["FREQUENTLY"] - 2 * obj["RARELY"] - 4 *
obj["NEVER"];
```

The scalar values were meant to give more emphasis to the extremes (Always/Never) so that the differences could be accentuated. It's immediately clear that while the smaller dots to the right are a lighter shade of green than the ones on the left the differences are small. Furthermore, since the scaling was based on $[-4, 4]$ with -4 being orange, 4 being green, and 0 being grey, all the values end up being over 0. Only when I changed the "never" category scaling to be multiplied by 40 did some noticeable patterns emerge. This seems like some evidence that the whole "never masker" movement as a contributor to COVID masking is overhyped or at the minimum the data set this is based on (being self reported surveying) is not necessarily the most representative.



This next graph really shows something interesting. The top right corner is quite busy, which is a concentration of GOP strongholds and low income. The sizes are based on covid cases (the total as of May 7th) divided by the population of the county. The number are scaled from 0 - 0.7 to a radius size between 0 and 50.



Challenges

- Figuring whether to use scale log or scale linear
 - scale log wouldn't work for a while because I forgot you couldn't include 0 in the domain
- Deciding whether to use the heavily skewed -40 for the "never" masking category or to keep it the same scaling as the "always" category
 - Ultimately decided that because it was such a large part of the COVID debate, that it's worth showing which counties were likely to be portrayed as such
- The most challenging might have been the most simple. Adding the axis line and label proved to be more difficult than I expected.
 - Putting in the two axis lines were easy enough but I struggled to figure out how to put in the arrowheads. Later realized that I need to add a url to my SVG header but still not sure what that does
 - I also felt like I was hard coding the axis label positions and just guessing
 - I know the whole thing will get screwed if the SVG is displayed on a smaller screen and I wish I could learn how to adjust the vector based on media queries

- In a similar vein, I used a CSS grid to position the different elements. Not sure how I could fix it so that it could adjust for different devices like if someone wanted to see it on their phone
- The date function to display COVID cases based on a particular point in time in the pandemic year
 - Not necessarily difficult but I knew it would take another hour at least to figure out how to set up the function so I decided to implement it in the end
 - Ran out of the time and wasn't able to include that functionality; will need to return to this project to code that in

Reflection

Overall, this project has given me the confidence that I have the skills to build another data visual from ground up using the lessons I have learned from the class. There is a process from where you start with a question, gather the data, explore the data, develop a strategy, enact the strategy, revise the strategy, and ultimately create the end product that seemed so daunting at the start of the semester but now feels manageable. Part of the challenge is reigning in my frustration when I'm struggling with a particular part of the visualization. Often there is a feature that seems so easy and intuitive to put together in your head but translating that into code shows you just how much you underestimated some part of the implementation.

In no way do I think I have mastered Inkscape, Tableau, or D3. I would even argue that it's near impossible to master them because there seems to be some many cool different things you can do with them. But I feel excited and ready to apply these skills in projects I will do for other classes and beyond.