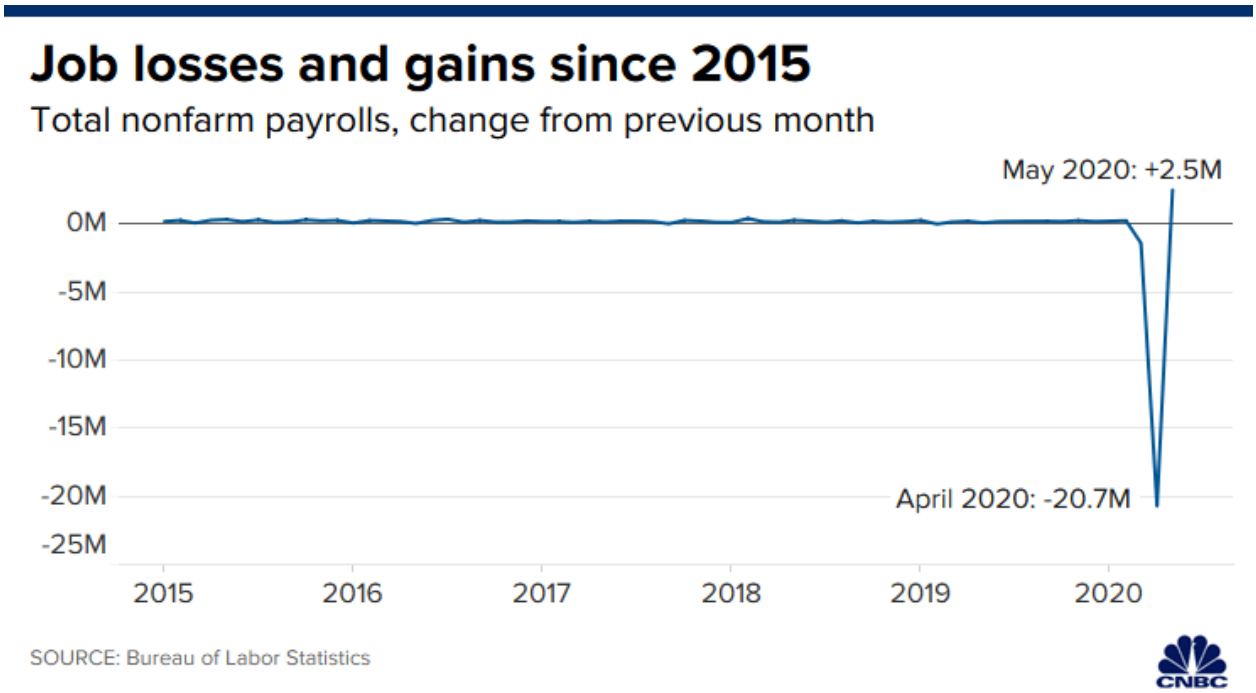
David Sant
U0454956

Data Visualization Homework 1

Figure 1



1. What data is represented in this visualization? Be specific.

   This graph shows the number of job losses (in millions) per month from January 2015 through May 2020. This is taken from the number of non-farm payrolls, so it should be a relatively accurate representation of numbers of job losses, with the exception of not taking into account deaths of working individuals during the COVID crisis.

2. What questions does the visualization answer?

   This visualization was meant to answer a question about job losses during COVID in comparison to the previous 5 years. It shows that there was a great decrease in the number of jobs in April 2020, but that the number of jobs is increasing in May 2020.

3. Describe one aspect of the visualization that is effective.

   This effectively shows that April 2020 had major job losses, much more than any month in the 5 previous years. This change is very dramatic. It also shows that the number of jobs is increasing in May 2020.

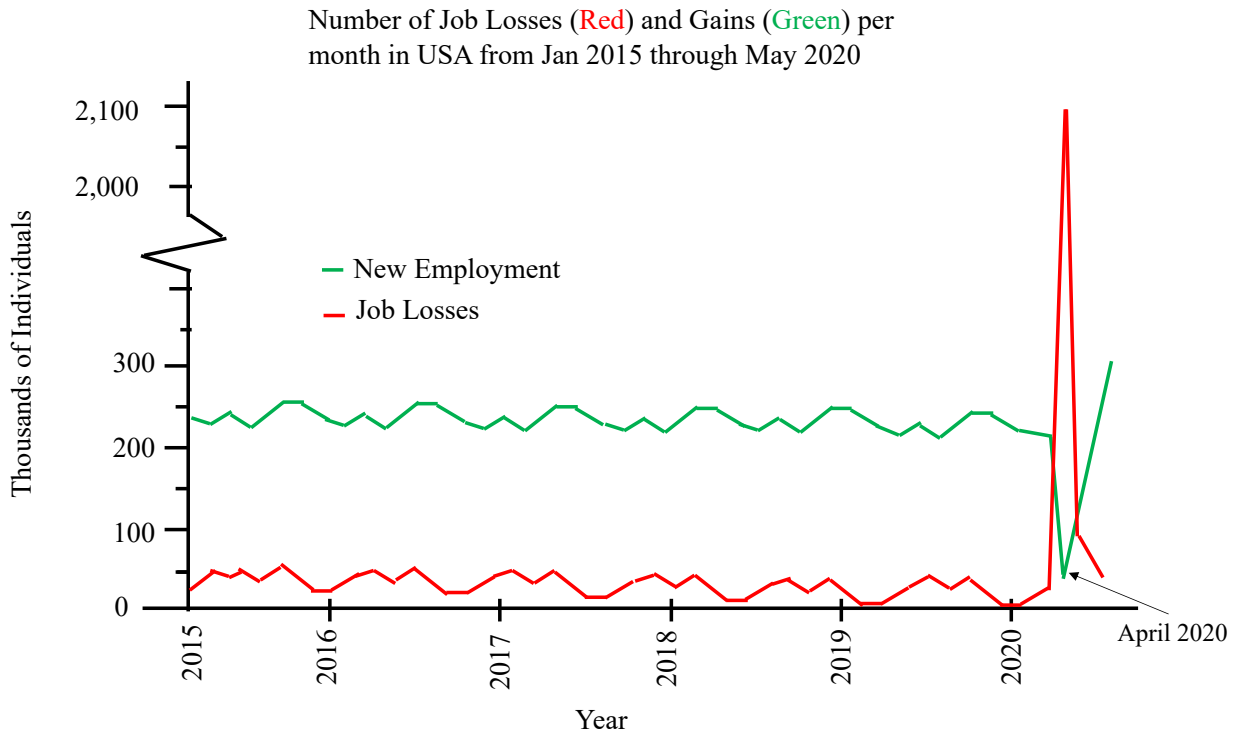
4. Describe one aspect of the visualization that is *not* effective.

   It was not effective at showing any change in number of jobs since 2015. It looks like it shows is that there were no changes in job stability for 5 years straight, then it suddenly shot down dramatically. In reality, there is constant change in the job market and April 2020 just happened to be very extreme. Additionally, the label says "Job losses and gains", so I expected it to show perhaps a line below the X-axis with the number of individuals that lost their jobs for reasons other than retirement or death and a line above the Y-axis for individuals that obtained new employment. Instead, it shows the change in the number of payroll employees from one month to the next. I am also not sure if the point for May 2020 shows that it is now better than baseline, or that it is better than April, but not as good as it was for the past 5 years.

5. Why do you like/dislike the visualization?

   I feel like the Y-axis should have a break so that you can see that change actually exists normally (not only in April 2020). I also feel like it should either show the raw number of positions filled or show separately the number of job losses and job gains. I especially don't like that it looks like in May 2020 there were so many jobs gained that there are now more people employed than ever in the previous 5 years.

Redesign 1

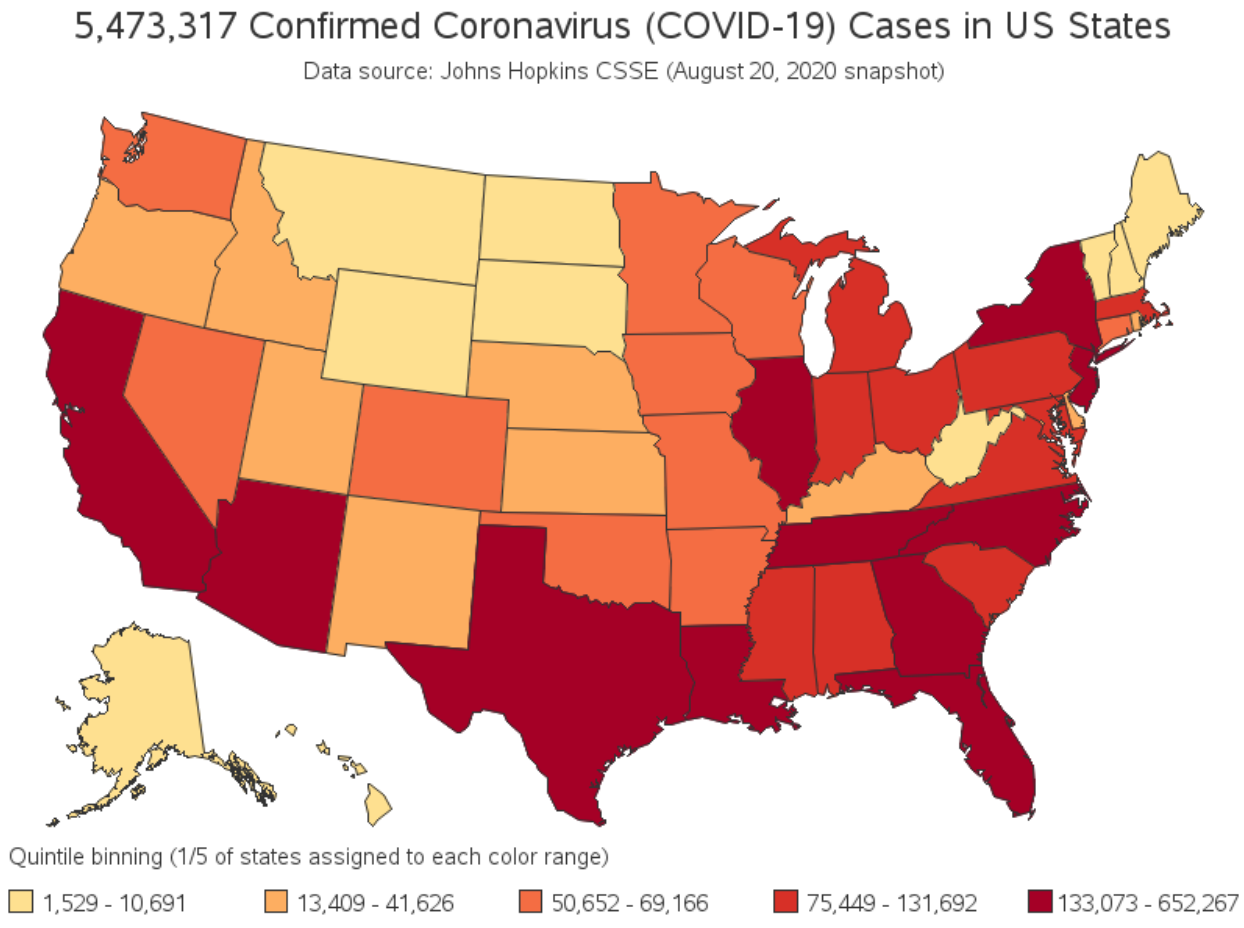


Above is my redrawing of the chart. It shows the number of job losses and the number of individuals with new employment per month from January 2015 through May 2020. The

separate colors make it easy to see which one is job losses and which one is job gains. The X-axis is split so that you can see that there is variance from month to month aside from only April and May 2020. This way it is easier to see that the number of job losses in April 2020 was far greater than the gains in May 2020. I considered using a blue line and a different Y-axis (with the label on the right) to show the overall number of employed people during each month, but I was afraid that adding an extra layer of data would make it too complex and not intuitive.

Figure 2

I chose a figure of a map of COVID-19 infection rates in the United States. This comes from Robert Allison's blog titled "Tracking the Coronavirus Disease 2019 (COVID-19) in the United States and can be found at https://blogs.sas.com/content/graphicallyspeaking/2020/03/13/tracking-the-coronavirus-disease-2019-covid-19-in-the-united-states/.



I chose this topic because it is obviously a big deal in the world currently. It affects our ability to go to social situations and has even been the cause of us moving our courses to entirely online for the majority of a year (possibly more than a year). He chose to use a map of the states and color each state according to the number of coronavirus cases. The image has links embedded to allow you to click on states, but I am critiquing only the image. Below is my critique.

1. What data is represented in this visualization? Be specific.

   This is a map of the United States. Each state has a color that represents a range of the number of people in that state that have tested positive for COVID-19. This does not distinguish between people that have died, recovered, or are still battling the illness. The map allows you to see where the cases are the highest and the lowest.

2. What questions does the visualization answer?

   This visualization answers the basic question of where in the country are the highest rates of COVID-19 and the lowest rates. Projecting the colors on the map also shows the location of the states relative to each other.

3. Describe one aspect of the visualization that is effective.

   This is effective at showing which regions of the country have the highest rates of COVID-19. For instance, it is easy to tell that states with a border against the ocean tend to have much higher levels of COVID-19. Those states east of the Mississippi river tend to have much higher rates of COVID-19 than inland states west of the Mississippi river. The color gradient has only 5 colors, so it is easy to tell the difference between rates in states even when they are not located nearby.

4. Describe one aspect of the visualization that is *not* effective.

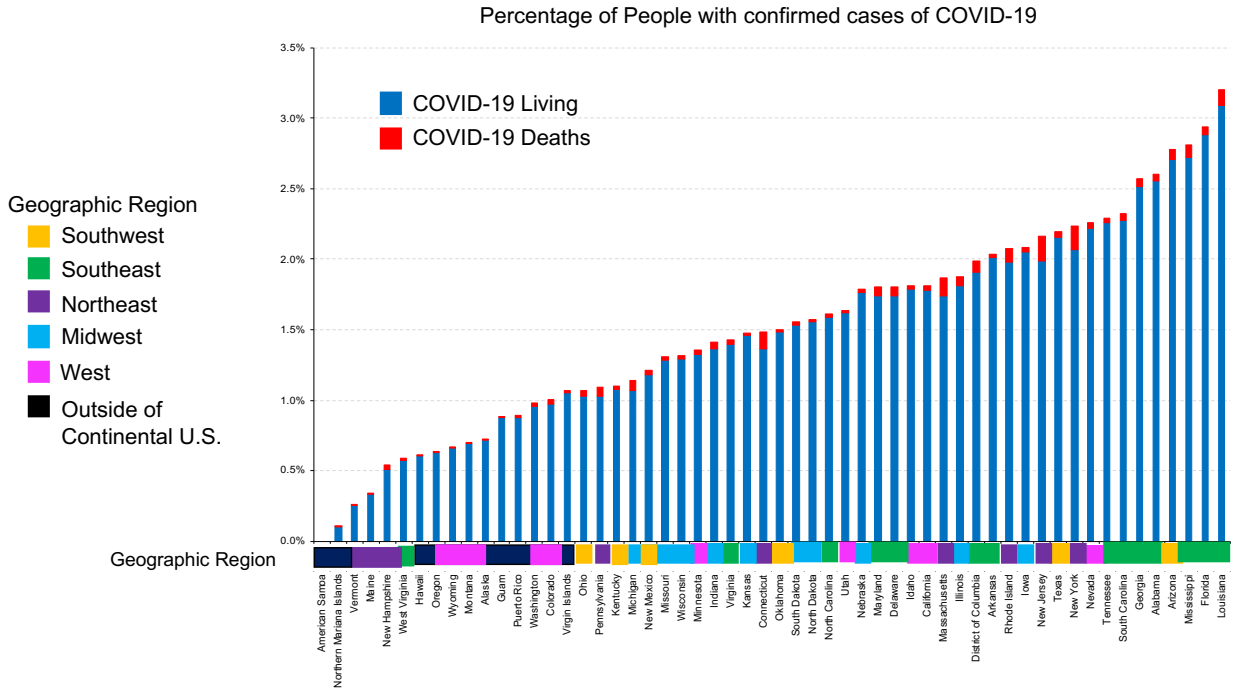
   This visualization does not take into account any information about the state populations, so you cannot tell which ones have high percentages of the population with COVID-19. I like the map, but it also does not show the COVID-19 rates per county which could be useful. For instance, upstate New York is extremely different from New York city and likely has much lower rates of COVID-19.

5. Why do you like/dislike the visualization?

   I really like the map because it is easy to perceive exactly where each state is. However, I do not like that the color gradient does not allow for any form of ranking. For instance, using only 5 colors means that you can't tell if Florida has more cases of COVID than New York. I would like to see these ranked. It also does not take into account death rates, so you can't tell which ones have higher death rates than others. I also don't like that it does not include information about the state populations. For instance, Georgia has a high

COVID-19 rate, but this map makes it look the same as California. The truth is that California has an extremely large population, and therefore likely has a much lower COVID-19 rate per capita than Georgia.

Redesign 2



This was made using Mocrosoft Excel and PowerPoint. The geographic region labels were taken from a map made by National Geographic. https://www.nationalgeographic.org/maps/united-states-regions/#:~:text=A%20common%20way%20of%20referring,West%2C%20Southeast%2C%20and%20Midwest. The numbers of COVID-19 cases, COVID-19 deaths and populations were taken from the GitHub page linked with the blog (https://github.com/CSSEGISandData/COVID-19).

As you can tell, I made this in PowerPoint and things aren't lined up quite right, so this is really a rough draft. However, you can get the idea for what I am going for. Looking at this makes it much easier to tell that the Southeast has a much higher prevalence of COVID-19 than the rest of the country in general. For instance, California and New York are much further down the list than you expect based on their raw numbers. You can also see that the West and territories outside of the continental U.S. have the lowest rates of COVID-19 per the population. You can also see what percentage of the COVID-19 cases have died in this graph. I think the map was more effective at showing the proximity of states, but this gives a rank so it is easy to see that Louisiana has more COVID-19 per the population than even Florida.