State-level Unemployment Rates in the U.S. in September 2020

1. Introduction of the bad graph

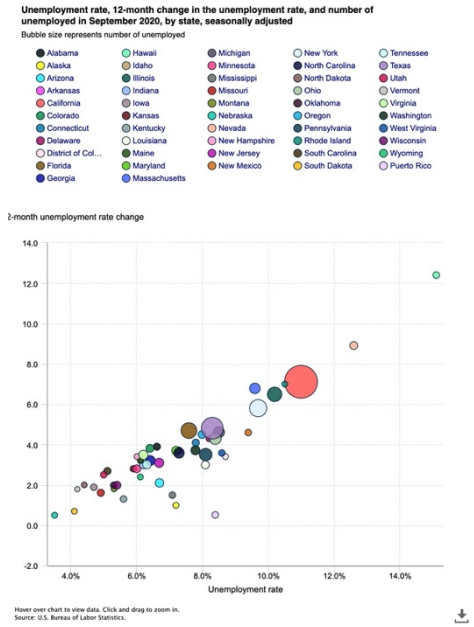

Figure 1: The Bad Graph (U.S. Bureau of Labor Statistics)

The original graph is displayed on the U.S Bureau of Labor Statistics website. The purpose of the graph is to represent information related to the unemployment situation in the U.S. in September 2020. The first variable is about the unemployment rate across states in September 2020, which is displayed on the horizontal axis. For the second variable, the value shown on the vertical axis represents the 12-month unemployment rate change from September 2019 to September 2020. The final variable is the number of unemployed persons in September 2020 and it is displayed by the size of bubbles. From the graph, it can be seen that the way to distinguish states is based on color of the circles.

2. The bad design flaws

The original graph does not provide an effective data visualization. Firstly, it does not facilitate the comparison about size of the bubbles. For example, the bubbles sizes of Nebraska and South Dakota are very similar. Even though they are juxtaposed with each other, it is still impossible for us to discover which bubble is bigger than the other. The only thing we can perceive is that California and Texas have the highest number of unemployed persons because the bubbles of these two states are much larger than the other states. For the remaining bubbles, we cannot make a comparison between them. The second weakness is that we cannot recognize states having the same base color. For instance, the base color of Tennessee and New York is light blue and the color of Tennessee is only a little darker than that of New York. For these cases, it is very challenging for readers to distinguish between them. Another disadvantage of the graph is related to overplotting problems. It can be seen that there are a lot of overlapping circles near the vertical grid line at 6.0% and some circles are completely covered by others. Finally, even

though the data set is associated with geographic locations, the graph does not illustrate any geospatial context which inhibits readers from exploring meaningful insights.

Based on these disadvantages, I would like to use the linked micromap design to improve the bad graph so that readers can easily recognize states, compare the values between states and extract a vital piece of geographic information from the dataset.

3. The first redesigned graph

+ Data preparation

The dataset of the original graph can be directly downloaded from the button "Show Table" below the plot on the website. The original dataset is comprised of four columns, including state, unemployment rate, 12-month rate change and number of unemployed persons. Each row represents the value of each state. For the redesigned graph, to visualize more clearly about the change in unemployment rate from September 2019 to September 2020, I add an additional column which is from a different dataset on the website (U.S. Bureau of Labor Statistics). It provides information about the jobless rate in September 2019, which is not provided in the original dataset. Therefore, I remove the column containing the 12-month rate change and instead will visualize the change by using two other columns, which are the unemployment rate in September 2019 and September 2020.

+ The redesigned improvements

The reason I choose the linked micromap is that we can divide all states into perceptual groups and use different colors to associate the states with their values. One further benefit from this design is that readers can pay more attention to the local panel and detect similarities and distinctions (Pickle, Pearson, & Carr, 2015). There are five columns shown on the redesigned graph.

For the first column, there are ten small maps indicating the regions, which helps us extract essential insights from geospatial data. It can be observed that the panels above median highlight the states which are mostly along the border from the West to the South and in the North East. For the panels below median, it highlights a cluster of states in the West and the Midwest. This is a very important pattern that the original graph could not display. Another improvement is that the second column performs state names explicitly, whereas we have to take time to find the color matching with the states on the top of the bad graph. Next, I use the dot plot to visualize the unemployment rate in September 2020. After sorting rows, we put similar rows close together, which simplifies appearance of the whole plot. From that, it also becomes more easier for us to know the states having the highest and the lowest rate. Additionally, the vertical reference line, which represents the national unemployment rate at 7.9% (U.S. Bureau of Labor Statistics, 2020), is added to compare the values not only between states but also between a specific state and the national rate. Combined with the maps on the first column, we can see that regions in the North East and along the border from the West to the South primarily have the jobless rate above the national rate. The next column is to display the 12-month unemployment rate change by an arrow. The starting value is the rate in September 2019 and the ending value is the rate in September 2020. By obtaining more data about the rate in September 2019, the redesigned graph can show the difference between two time points and the value of each month, while the bad graph only shows the change. This new design reveals that the five states in the first panel (from top to bottom), which have the highest unemployment rate, also have very high value for the 12-month unemployment rate change. The

rate in these five states in September 2019 was below 5% and approximate to the rate of the other states, however, it rose sharply for the past 12 months. During this period, we know that the drop in employment is as a result of COVID-19 pandemic's impacts. Therefore, the rate of 51 states increased rapidly compared to the rate of last year. The final column is to display the number of persons jobless in September 2020. I use the bar plot to differentiate the values between states, which are poorly visualized by the size of bubbles of the old graph. For states having similar values, we still can compare which value is higher based on vertical grid lines. From the plot, we can see that California and Texas have the highest number of people jobless, which is significantly higher than other states. The reason is that California and Texas are the most populous U.S. states.

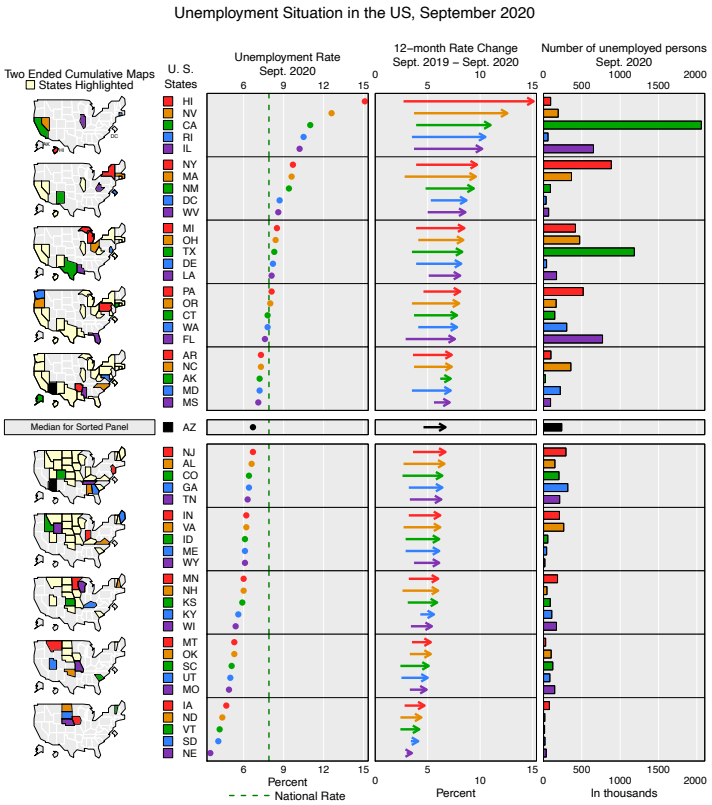

Figure 2: The first redesign graph

In summary, by obtaining more data and using the linked micromap design, the new graph completely overcomes the weaknesses of the bad graph, in terms of supporting the comparison, having the reference value and distinguishing the states in an easy way. Especially, it facilitates us to discover patterns related to geographic context. The states having the jobless rate above median are mostly along the border from the West to the South, while the states having the rate below median locate in the center of the West and the Midwest.

4. The second redesign graph

In addition to the linked micromap design, the choropleth map is a very good choice to execute an analysis on a geographic dataset. One of the most powerful strength of the choropleth map is that the color scheme can support readers to experience an easy visual perception. We can easily discover value differences and geographic patterns only by the color scheme.

Therefore, I decide to create a choropleth map to visualize the unemployment rate change during this period. Two different types of data visualization allow us to derive more useful ideas. This graph is created by using the data from the dataset of Figure 2 and adding more data from a different dataset on the website (U.S. Bureau of Labor Statistics). The additional data describes the unemployment rate in January 2020 and April 2020.

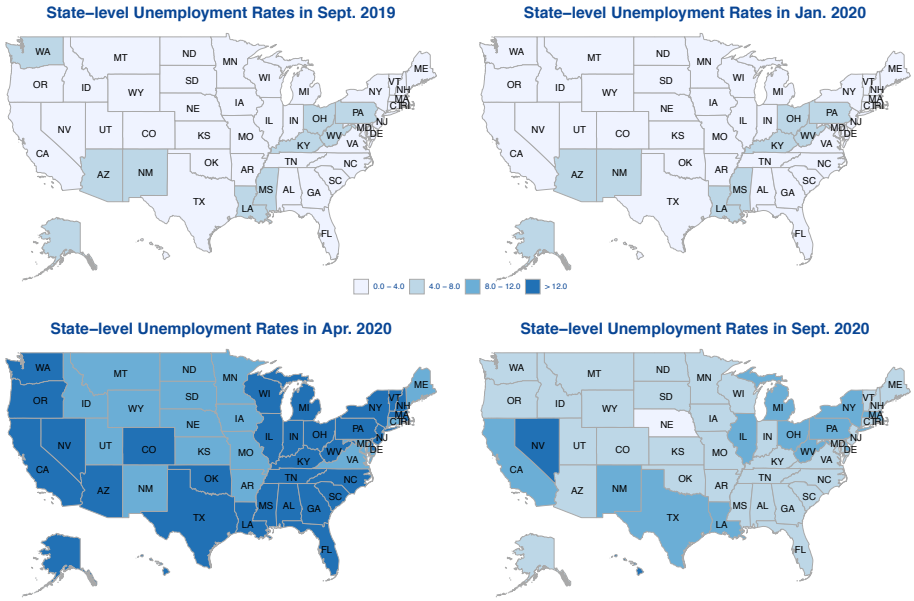

Figure 3: The second redesigned graph

To discover the change during the period, I use the 'ggarrange' function in the 'ggpubr' package to combine four individual maps onto the same page. Each map corresponds to the rate of a month. If each map is drawn separately with different colors and percentage intervals, there will be no connection and correlation between them. Therefore, the unemployment rate is converted to a factor variable for each small map, including four levels: 0.0-4.0%, 4.0-8.0%, 8.0-12.0% and over 12%. Also, there are four colors corresponding to each level of the rate. The lightest blue describes the states having the lowest rate and the darkest color is used for the states having the highest rate. From the plot, we can see that the unemployment rate from September 2019 to January 2020 is unchanged, however, the rate of all states in the U.S. rise dramatically in April 2020. In September 2020, the unemployment rate is mostly lower than the rate in April, but it is still much higher than the rate in September 2019.

In conclusion, the linked micromap and choropleth map are suitable ways for geographic data visualization. By obtaining more data, both redesigned maps can reveal the patterns related to regions and help us notice similarities and discrepancies between states. For the first redesigned graph, the linked micromap design can provide information about rate, rate change and the number of persons jobless together on one plot. Also, the national rate is added to allow us to make comparison between state-level rate and national rate. For the choropleth map, we can perceive how the rate is changing during the period without much effort. Even though it presents only one variable, it is still a good design to use for geospatial context thanks to its powerful color scheme. For the advantages mentioned above, the redesigned graphs are completely better than the original graph.

References

Pickle, L. W., Pearson, J. B., & Carr, D. B. (2015). micromapST: Exploring and Communicating Geospatial Patterns in US State Data. *Journal of Statistical Software*.

U.S. Bureau of Labor Statistics. (n.d.). Retrieved from https://www.bls.gov/: https://www.bls.gov/charts/state-employment-and-unemployment/unemployment-rate-12-month-change-in-the-unemployment-rate-and-number-of-unemployed-by-state.htm

U.S. Bureau of Labor Statistics. (n.d.). Retrieved from https://www.bls.gov/: https://www.bls.gov/charts/state-employment-and-unemployment/unemployment-rate-12-month-change-in-the-unemployment-rate-and-number-of-unemployed-by-state.htm

U.S. Bureau of Labor Statistics. (n.d.). Retrieved from https://www.bls.gov/: https://www.bls.gov/charts/state-employment-and-unemployment/state-unemployment-rates-animated.htm

U.S. Bureau of Labor Statistics. (n.d.). Retrieved from https://www.bls.gov/: https://www.bls.gov/charts/state-employment-and-unemployment/state-unemployment-rates-animated.htm

U.S. Bureau of Labor Statistics. (2020). *The Employmnet Situation - September 2020.*