

Assignment 5 Report: Data Visualization

This assignment was completed with the help of RStudio, Atom and Firefox.

The final product is an interactive line plot with data on COVID cases and deaths in California since the beginning of the pandemic. The plot comes with the following indications on its function:

- Use the following button to hide all lines in the plot and click in it again to make them reappear.
- This will allow you to select the individual California counties you wish to see.
- Double click the name of the county from the legend to the right of the plot to isolate this county's line. You will be able to hover over this line to see only this county's information through time.
- Click once on other counties to add their lines to the display.
- Double click again on any name of the list to deactivate this function and show all lines.
- At any given time when more than one line is displayed, hovering over the lines will highlight the current line and dim the others as well as display a box of detailed information about that county on that date.

I began the assignment by loading the dataset containing the COVID case information by county in California provided by the professor to RStudio with `readRDS()`. Then I explored the data through actions like extracting the county names, viewing the first few lines of the data, checking that the number of observations for the counties match. Next, I proceeded to create the columns of information requested in the prompt from the data in the data frame as follows:

- `casesByPopulation`: contains the number of cases adjusted by population size rounded to 6 digits. I took the number of cases multiplied it by 100 and divided it by the population size for each observation.
- `deathsByPopulation`: contains the number of deaths adjusted by population size rounded to 6 digits. I took the number of deaths multiplied it by 100 and divided it by the population size for each observation.
- `casesToDate`: contains the cumulative sum of cases. I looped over all counties to obtain the cumulative sums with the help of `cumsum()`.
- `deathsToDate`: contains the cumulative sum of deaths. I looped over all counties to obtain the cumulative sums with the help of `cumsum()`.

After gathering all the data needed, I started building the plot with `ggplot`. First I used the most basic template and grouped the colors by name of the county and displayed it to check for errors. Then I continue to build the interactive plot with `ggplotly` and checked for errors once more. In the same manner I saved it with `htmlwidgets::saveWidget()` as “covidCA.html”. Once I knew this worked I started adding the labels and styling the annotation and axis’ ticks.

The first functionality I added was hiding the lines. This was done in atom by adding directly the script to the function `hideLines()` and the button to cause the action. Once in atom I started editing the text and headline because I thought I would try to complete the rest of the

functions directly in the html file. However when I was working on the next functionality (highlight one line over the others), I figured out how to do it in Rstudio. Thus, I went back to R to generate the plots again. This part was a little tricky because I could not understand how to use the functions `highlight()` and `highlight_key()`. I asked in piazza where I was advised to check one of the previous discussion sections. After seeing this it still took a couple of tries to get it to work but it worked. I set the highlight parameters to go on on hover and off on double click. Naturally I saved the changes to covidCA.html once more. What I did not realize was that it would undo the changes I did manually in atom.

In order to be able to proceed using this approach I created a copy of the html file called covidCA1.html after restoring the progress lost. This way I can continue to make and review changes to the r code without losing changes made in atom.

This became useful when I started editing the tooltips. I realize the county name appeared three times so I was going back and forth from r to Firefox to confirm the changes made to covidCA.html. I tried multiple ways of the argument tooltip in ggplotly but the name continued to have duplicates.

Once I was happy with the functionality of the r generated plot I copy and pasted the changes saved in covidCA1.html to covidCA.html and added some instructions in how to interact with the plot. Now covidCA.html contains the final result. In order to ensure that running the r code will not delete all my work I copied it again and named the copy covidCA_copy.html. This way if the r code is run to check if it works, my work is still saved in covidCA_copy.html.

Shortcomings: when one county is selected the name appears twice in the legend, and when we select on county after hovering over another line this new line appears dimmed. The first is inconsequential and the second can be easily fixed by hovering over this new line you want to see to highlight it. In addition another thing that I noticed is that if you hover over the plot after clicking the hide lines button it continues to highlight the single line you are hovering over. This can be seen as an unintended alternative version of highlighting and dimming the other counties, but it can be distracting to have both.