

Assignment 5 - Interactive Visualization

Yifu Wu (916619585) *

12/16/2020

In this assignment, I will do some research on the interactive visualization based on the Covid-19 data on New York Times. I will first draw a time series plot for cases in each county of California, and then draw a map for the cases in each county of California.

1 Approach for Interactive Time Series Plot

The first step is to get the Covid data from New York Times using the code given by professor. One important thing is that we cannot use `getURLContent()` on Windows since it will report an error. So we choose the `GET()` function from the `httr` package to do the same thing.

Since we want to analysis the Covid cases in each county in California, we need to find the geoid for California and match the corresponding counties in the hierarchy section. We also need to remove the last one due to the unknown county name. From the list we can get useful information such as the number of daily cases, number of cumulative cases, pro-rated value of cases relative to the county population and so on. We put those information into a data frame.

Since we need to show information in the tooltip, we need to first use the code below to determine what will be shown in the tooltip.

```
text <- paste("Cumulative_Cases_Pro_rate:", CA_covid_data$Cumulative_Cases_PR, "%", "<br>",
             "Cumulative_Deaths:", CA_covid_data$Cumulative_Deaths, "<br>",
             "Cumulative_Deaths_Pro_rate:", CA_covid_data$Cumulative_Deaths_PR, "%", "<br>",
             "Daily_Cases:", CA_covid_data$Daily_Cases, "<br>",
             "Daily_Cases_Pro_rate:", CA_covid_data$Daily_Cases_PR, "%", "<br>",
             "Daily_Deaths:", CA_covid_data$Daily_Deaths, "<br>",
             "Daily_Deaths_Pro_rate:", CA_covid_data$Daily_Deaths_PR, "%")
```

We also need to set a highlight key to ensure we can highlight one or more counties while others still appear in the background.

```
covid <- highlight_key(CA_covid_data, ~County, "Highlight one or more county")
```

To change the static plot drawn by `ggplot()` into an interactive plot, we need to use `ggplotly()` function in the `plotly` package with the `highlight()` function to enable we highlight specific counties.

```
plty <- ggplotly(time_series_cases) %>%
  highlight(selectize = TRUE, persistent = TRUE, off = 'plotly_doubleclick')
```

The next step is to output a HTML file, we can do it with the `saveWidget()` function.

*Department of Statistics, UC Davis and ifuwu@ucdavis.edu

```
saveWidget(plt, "HW5.html", selfcontained = FALSE)
```

Apart from using the R, we also need to use JavaScript to hide all cases in the plot since we need to allow viewer to toggle the time series plot. The JavaScript code is stored in **HideLines.js**, and it is the rather similar with the JS code professor used in the **SacYolo2.html**.

We also add some text into the HTML file we just create including the title and some instructions of the plot, we also add a **Hide Cases** button to connect with what we do in the JavaScript.

```
<script src="HideLines.js"></script>
<body>
<h2>Time Series Plot of Covid-19 Cases for each County in California </h2>
Instructions: <br>
1. To highlight one or more county, you can select in the search box below.
You can use backspace to delete one or more counties have been selected.<br>
2. To select a few counties, you can double-click the county on the right
side to select the first county, then you can click other counties you want.
You can double-click unselected counties to remove all selections.<br>
3. To toggle all counties, you can click the "Hide Cases" button.
And if you want to add some counties to the plot, you can select in
the search box. You can click the button again to show all counties
when there is no county shown in the plot.
<p>
Click the button to toggle the cases.
<button onclick="hideLines()">Hide Cases</button>
```

The figure above shows the screenshot of the HTML file which contains the time series plot. You can see the instruction in the HTML page (Also from the HTML code above).

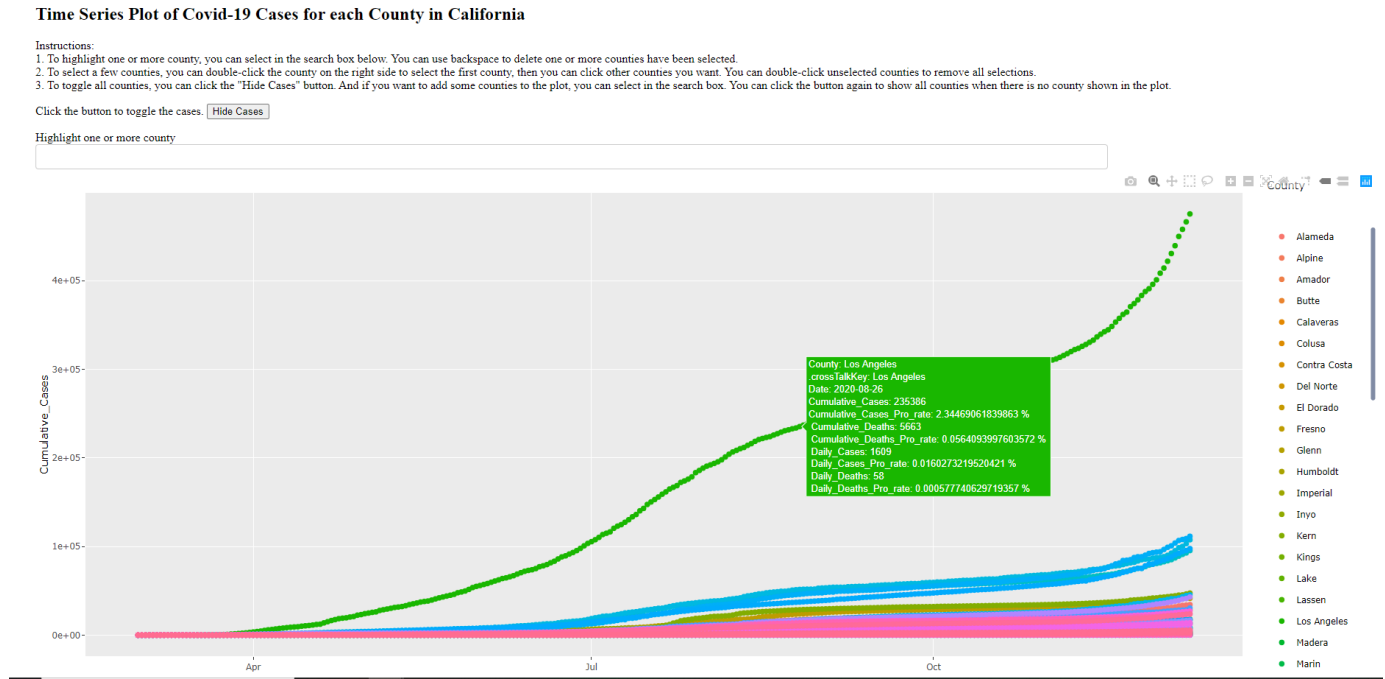


Figure 1: Time Series Plot of Covid-19 Cases for each County in California

2 Approach for the Covid Cases Map

For this part, we use **gganimate** package in R to generate an animated county map in California which contains the information of the pro-rated value for cumulative cases relative to county's population. The approach base on the tutorial from <https://www.5haw.com/posts/week-3-making-an-animated-map-using-maps-ggplot2-and-gganimate/>. However, it does not contain as much as information compared with the plot based on JavaScript.

The first step is to generate the county map for California. Here we use the **map_data()** function to get the geographic data of counties in California, and merge those into the data frame we created in the previous section. Then we can use **ggplot()** and **geom_polygon()** to draw a county map based on the new data frame. To generate a animated map, we should split the data into several states based on the Date using **transition_states()** function.

We then use the **animate()** and **anim_save()** functions to generate and save the map into a GIF file.

```
animated_map <- animate(ca_covid_map, nframes = 2*length(unique(covid_map$Date)))  
anim_save("covid-map.gif", animation = animated_map)
```

The following figure shows the pro-rated value in different counties in California on Dec.10. The full animated map can be seen in the **covid-map.gif**.

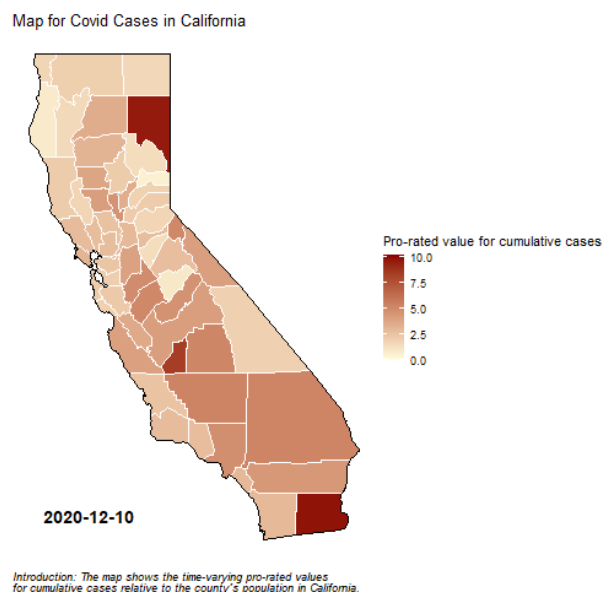


Figure 2: Map for Covid Cases in California