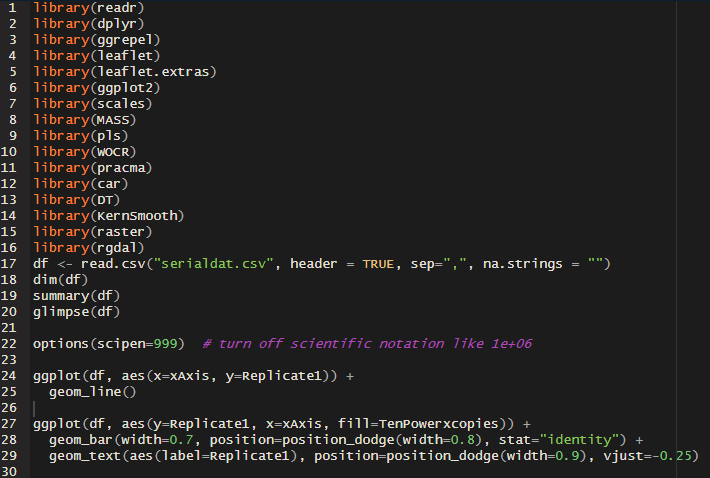
***HW 5: Use the provided data at the website (***[***https://ourworldindata.org/coronavirus/country/united-states?country=~USA***](https://ourworldindata.org/coronavirus/country/united-states?country=~USA)***)*** ***or All of Us data to complete the following visualization of distributions. Choose a country or region of US (if using All of Us data) and display the distribution of a continuous variable of your choice. Additionally, choose a categorical factor (gender, age groups, ethnicity/race, etc…) and compare the distributions in the way you find most useful. Clearly state the kind of message you intend to convey in the data visualization and explain how the visualization aligns with that message.***

I tried using the COVID19 dataset from the website but I couldn’t find a good variable to group by. Hence, I used the new dataset you posted called serialdat.csv.

First, I wanted to plot SUMOvar on the x-axis and Replicate1 in the y-axis but I wasn’t able to do it due to SUMOvar being a categorical variable. I fixed that issue by creating a new numerical column called xAxis. Using that new column I created the first line chart plot shown below.

Now, for the grouping part, I used TenPowerxcopies column to see all values according to the 6 different values of that variable (TenPowerxcopies). I created a bar chart plot and added the values at the top of each bar. The colors are default settings. I believe that the dataset was very simple and didn’t give me a lot of space to do a more complex visualization but I am curious to see what my classmates have done.

My Code



My charts

