## Sample Data Visualizations

Group 5:

**Chinmaya V, Christy Y, Daniel J, Muthu P, Simon Z, Zack P**

**Our Approach:**

Based on our initial discussions, some ideas we had were:

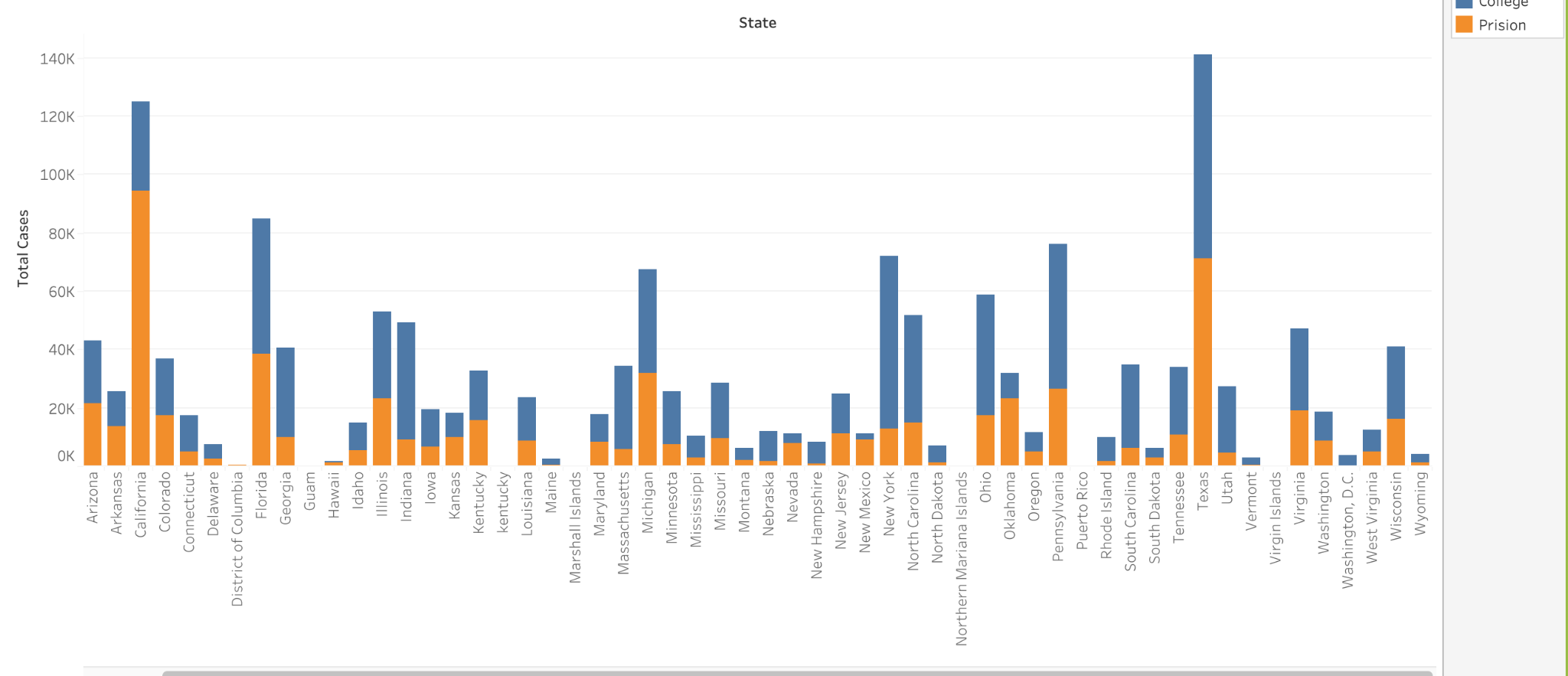
* Our first idea was a bubble map of the US to represent cases from colleges and cases from prison facilities in a county. This will be our main chart to visualize the comparison between the two institutions because the bubble map chart overlays data particularly well. By increasing the radius of the bubble based on case count and color-coding bubbles based on whether the cases originated from colleges or prisons, it would serve as a great first look into how the data sets compare from a zoomed out, first look type of way. Again, we decided to use a bubble map for this because this visualization overlays data in a way that makes comparison by location on a map very straightforward.
* Similar to our bubble map chart, we wanted to compare the number of COVID-19 cases that occurred at a college campus and at a prison facility within the same county. While the bubble map chart is important in helping readers visualize the data on a familiar geological view, we wanted to visualize that data as bar graphs as well because in Chapter 4 of *Data Points: Visualization That Means Something* by Nathan Yau in our assigned reading, the author made a point that bar graphs help highlight dramatic differences in numbers much more significantly than comparing relative areas. Thus, we don’t want readers’ comprehension of the difference in numbers to be sacrificed for a geographical understanding.
* Another visualization and comparison we wanted to make was between the number of covid cases in a particular college in 2020 versus the number in 2021. This comparison could show how the colleges combatted the pandemic and possibly saw improvement, speaking to the emphasis on healthcare in an educational institution. A gantt chart works great for this comparison because we are looking at two data points that match the institution but differ in time. Gantt charts are usually used to display how data varies over time as explained in the Hardin et al. reading.
* Since the reaction to the pandemic has also varied by state and state’s political ideologies, we wanted to make a comparison between the number of cases in prison facilities in each of the 50 states. To compare these data points, we decided a bar chart could work well because bar charts showcase trends and distributions well. Also since there are 50 states, we could make bars slightly thinner to make for a more readable chart. In the reading, it was stated that bar charts are great for comparing data across categories, and in our case, our category is state.
* We also wanted to compare the case rate and death rate between inmates and officers within the same facility to investigate to what extent their treatment differed despite interacting in the same environment. To do so, we wanted to use a pie chart to visualize the percentage of cases within a facility that were caught by officers compared to the percentage of cases within a facility that were caught by inmates. The pie chart also helps to visualize the percentage of limited resources that were allotted to inmates versus officers at the same facility. We could also have a similar visualization for the percentage of deaths within a facility by officers versus inmates.

**Charts/Maps/Visualizations:**

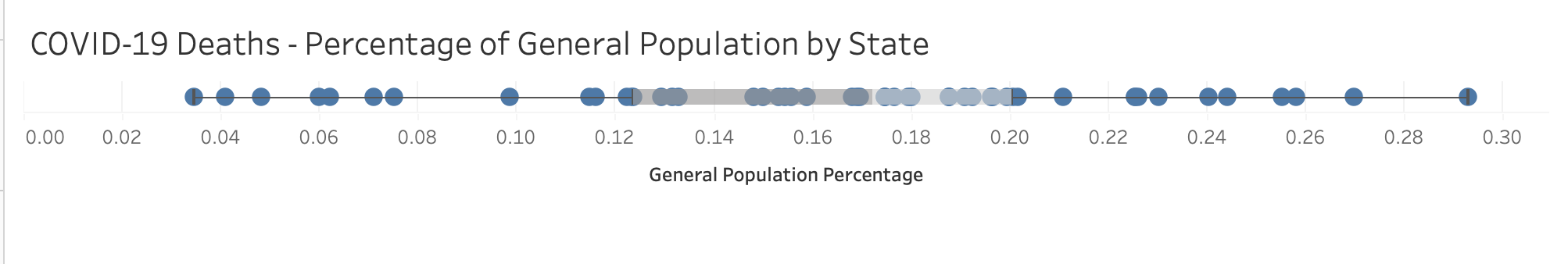
**Total Covid cases in colleges and prisons on map**

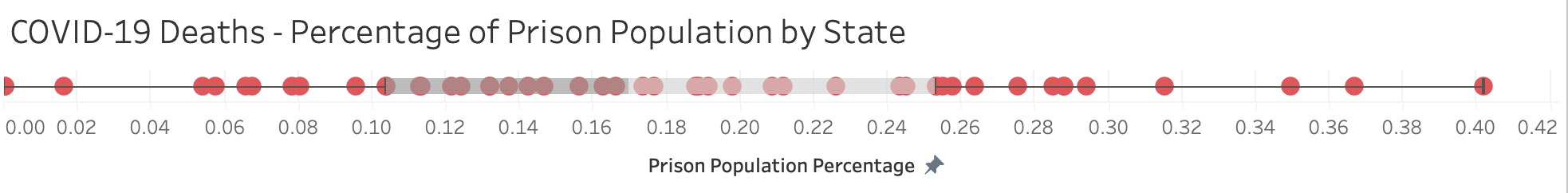
(Daniel)

**Total Covid Cases in Stacked Bar chart**

****

This Tableau-generated stacked bar chart depicts the raw Covid-19 case count in prisons and colleges, separated by state. From first glance, this graph shows that colleges have had significantly worse Covid positivity numbers than prisons. In reality, what will be shown in a later graphic is that, once cases are adjusted based on population, there was little difference between Covid-19 cases between the two communities. We feel that this graphic helps set the scene as it implies that there is a dichotomy between the treatment of the Coronavirus in colleges and prisons, but we will later show that this may not be the case.

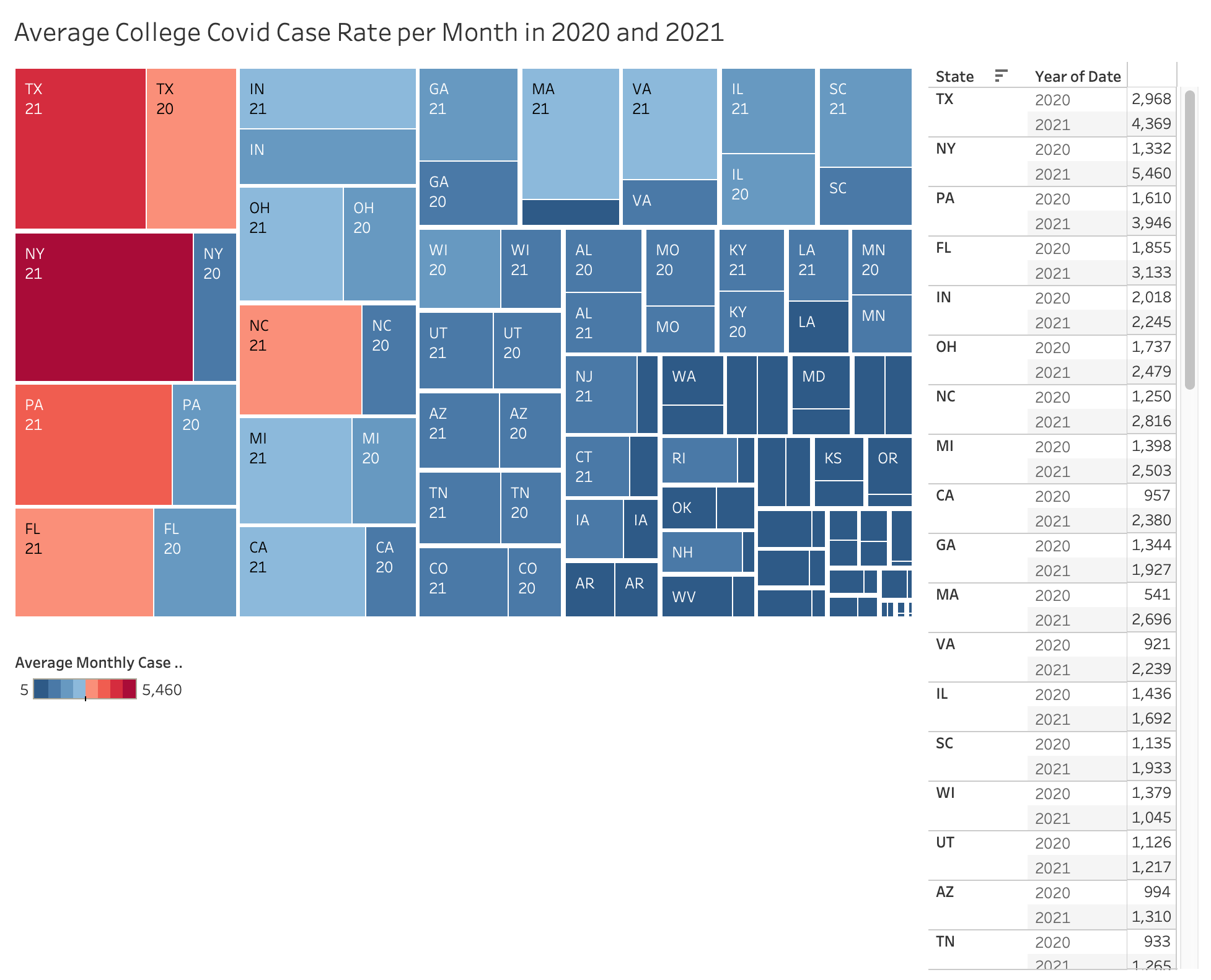
**Population Adjusted COVID-19 Deaths by State Versus Prisons**



The graphic above is a stacked box plot comparing the distribution of the percentage of COVID-19 deaths by state inside and outside of prisons. Based on the above graphs, we see that the median and IQR of death rates between general population and inmates is fairly similar.

**Adjusted Comparison of COVID-19 Cases in Prisons vs Colleges**

(Daniel)

**Comparison of 2020 to 2021 Covid Case Rates in Colleges **

Data cleaning: We divided the total number of Covid cases for 2020 by eleven months since February was the month the pandemic was officially announced as an emergency in the United States. Furthermore, we divided the total number of Covid cases for 2021 by five months since our data for colleges ends at the end of April.

Although this cleaning step falsely assumes that case numbers were uniformly distributed by month, we can use the graphic to demonstrate the average difference in college Covid cases from 2020 to 2021.

**One timeline**

Good work here. I think there is work that can be done to make your graphs more readable and also to focus in on the information that is important to your narrative. I’m a little concerned that only one of your visualizations covers the college data set. Is this intentional? If so, will the college portion play a more limited role within your narrative? Do you feel that it needs to be part of your project or would you prefer to focus solely on the prison data? With the prison visualization that you have, see if you can more clearly narrow in on what about these graphs and charts you’re planning on focusing on in your narrative. The connection to how they are helping you address a research question still seems a little underbaked. Begin to fill that out a bit more for your reader.

Preliminary grade: 87/100