## Sample Data Visualizations

Group 5:

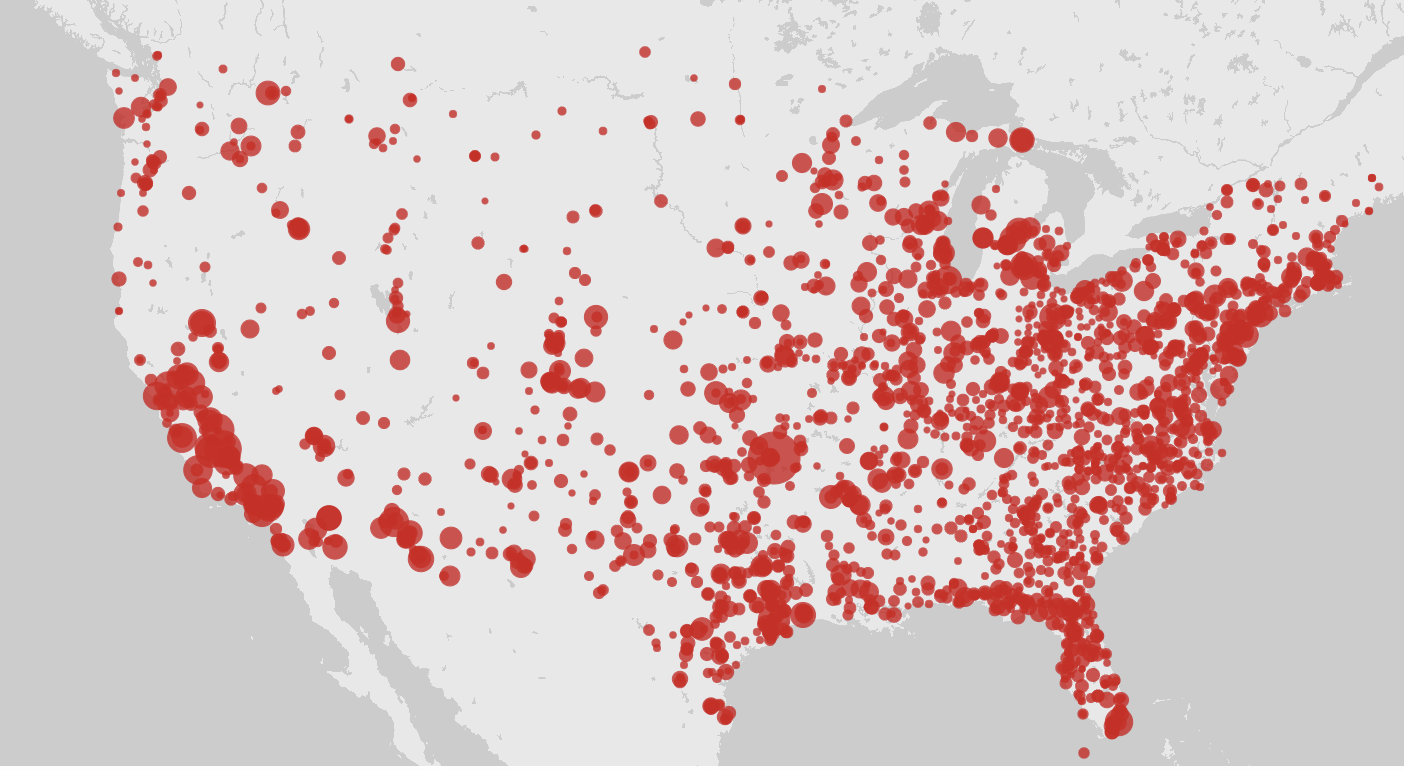
**Chinmaya V, Christy Y, Daniel J, Muthu P, Simon Z, Zack P**

**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.

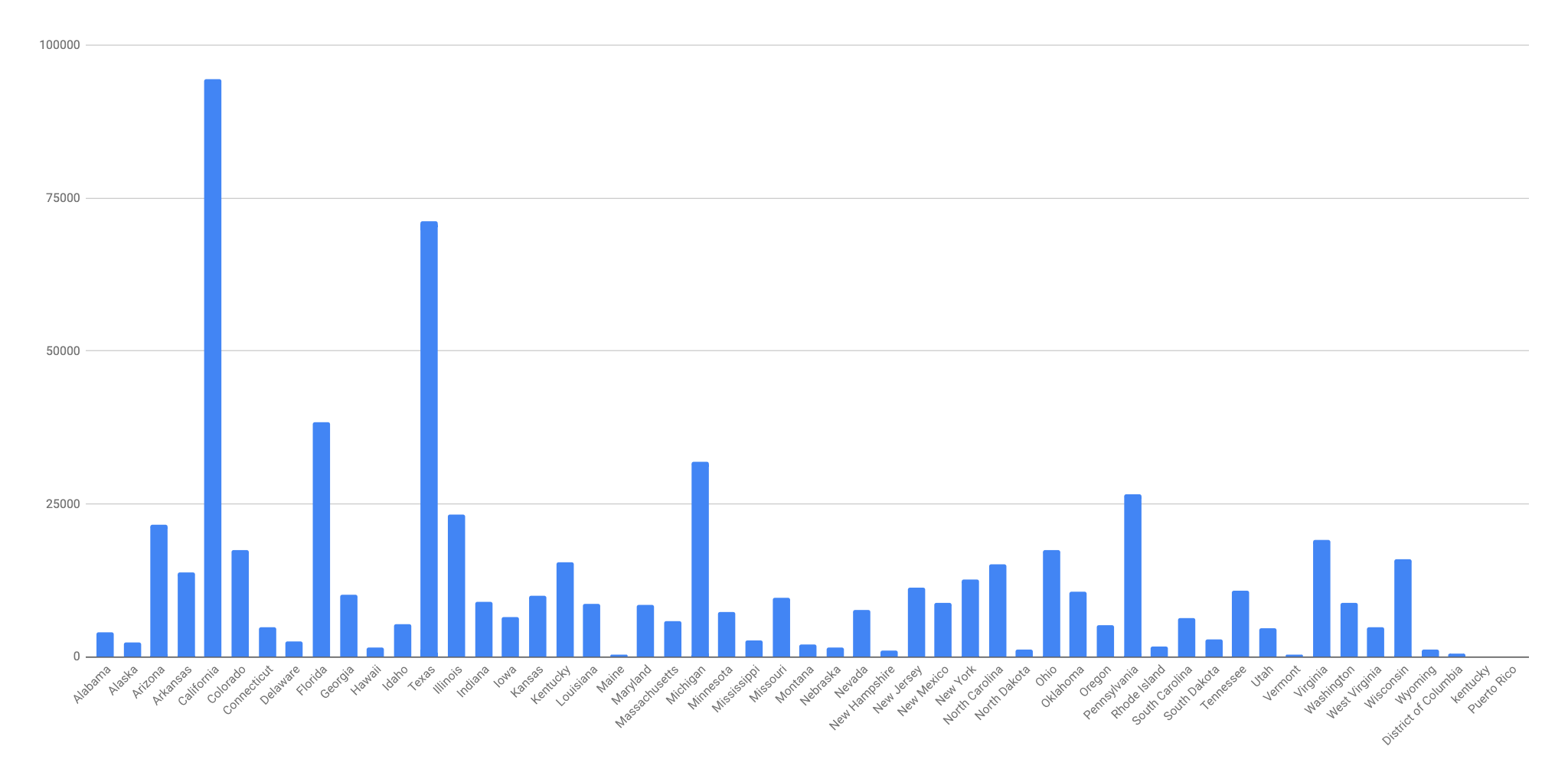
**Charts/Maps/Visualizations:**

Total Prison Cases (Inmate + Staff) in Continental US

Visualization process: data was cleaned to combine staff and prisoner cases. This visualization is the default provided through Palladio but we have to still have to make our own with the json.

Blurb:

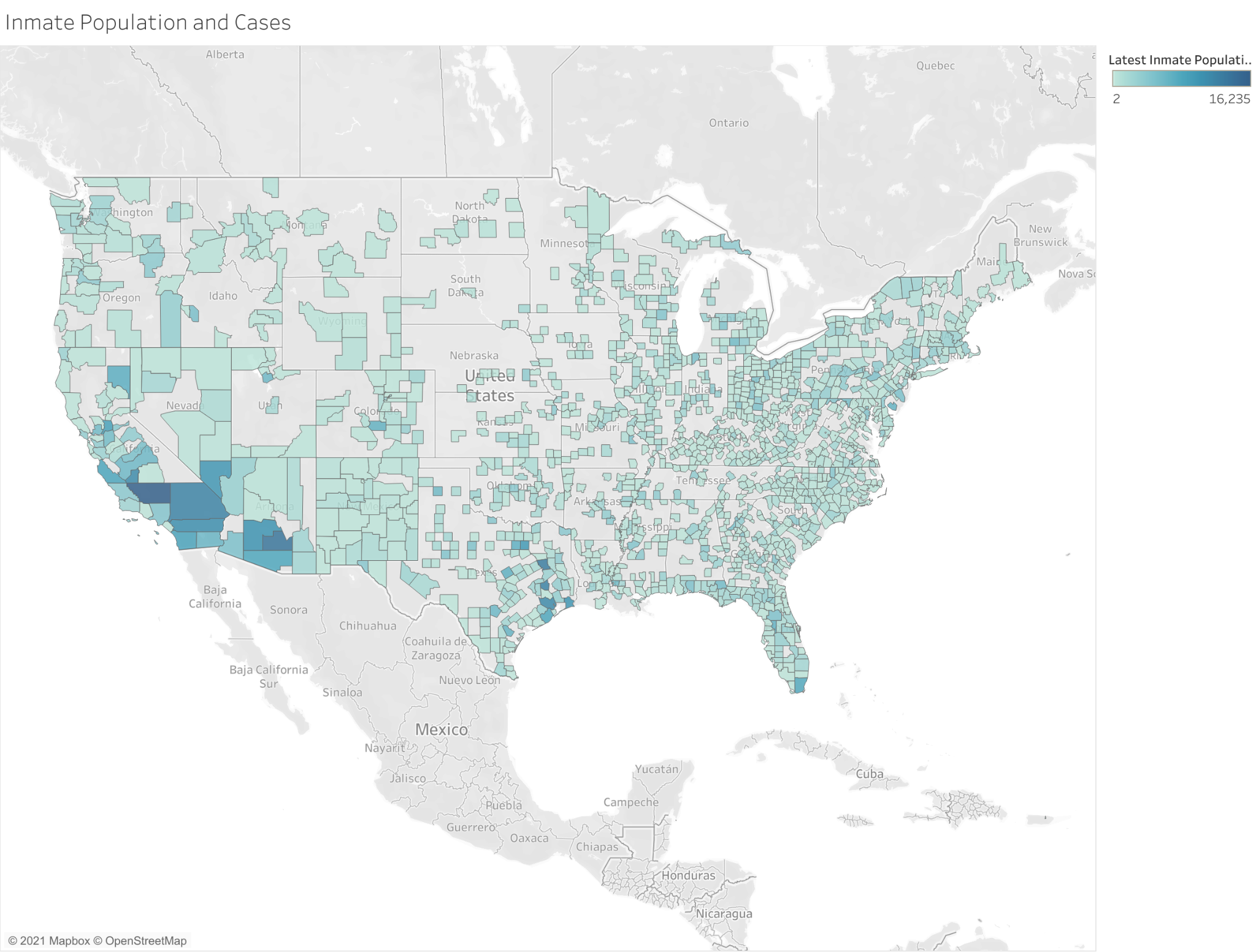
This data visualization shows the total number of prison cases in the continental US indexed by county through a bubble map. We believe this can detail for us what prisons in the US were hit hardest by COVID and gives us a very clear and low-bias (as it is more granular than the entire map area of a state, which can be misleading) impression of how COVID has impacted the prison system and whether particular parts of the country were hit harder than others as a whole. We have cleaned the data to combine staff and prisoner cases for each location, and the visualization above is a sample visualization from Palladio - we intend on creating our own version of this bubble map using a JSON.



Total Prison Cases by State

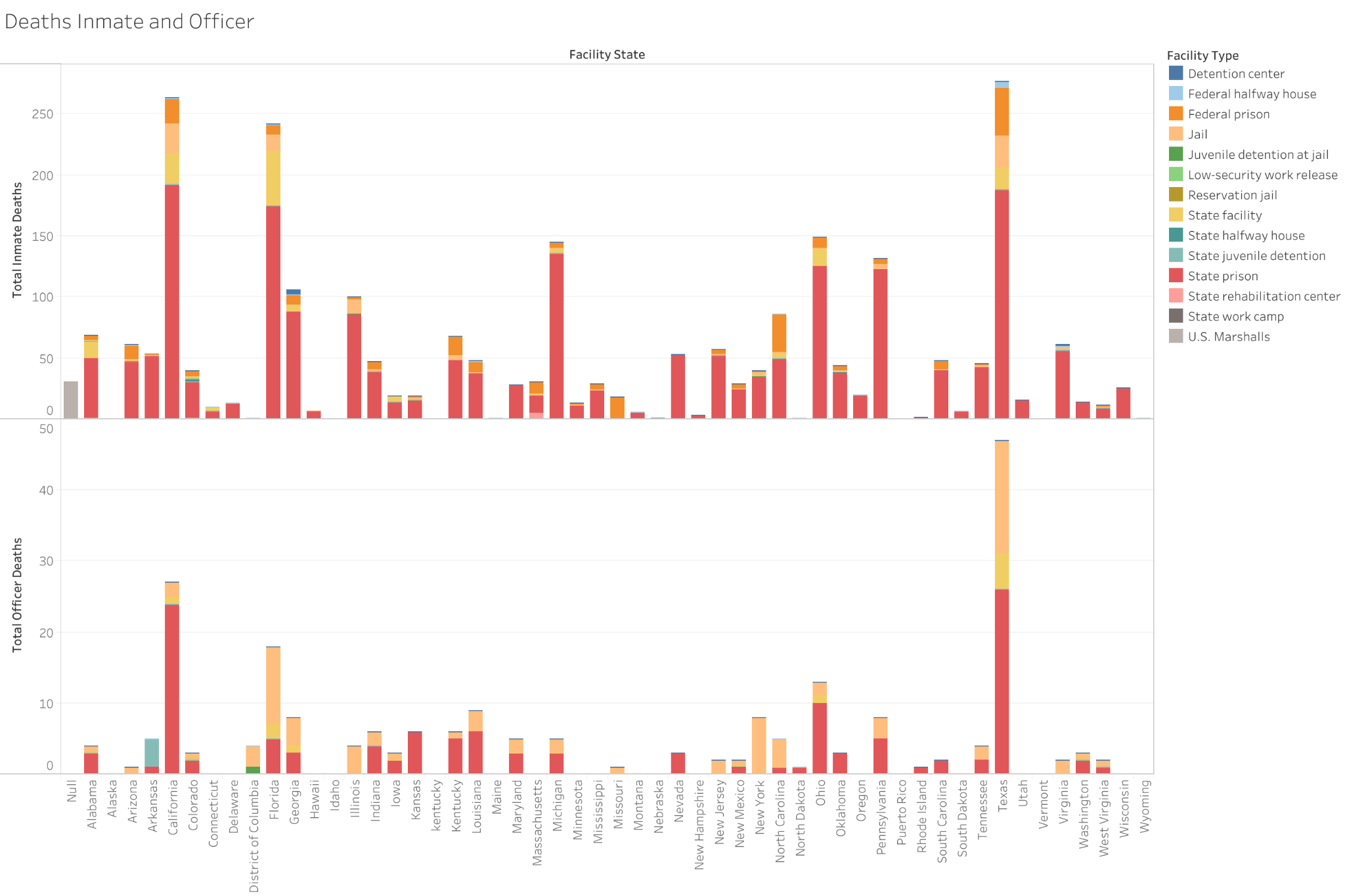
Visualization process: same as above. States including DC and Puerto Rico. Visualization is a bar chart through Google Sheets

This is a histogram showing the total number of prison cases in the US indexed by state, including DC and Puerto Rico. Through this, we want an easily comparable rule that can outline certain states with exceptionally many or few COVID cases, as this will help direct us to examine these states’ policies or conditions that may have contributed to their outcomes with COVID care in a high-risk setting. We visualized this as a bar chart using Google Sheets, but plan on developing our own using the JSON.



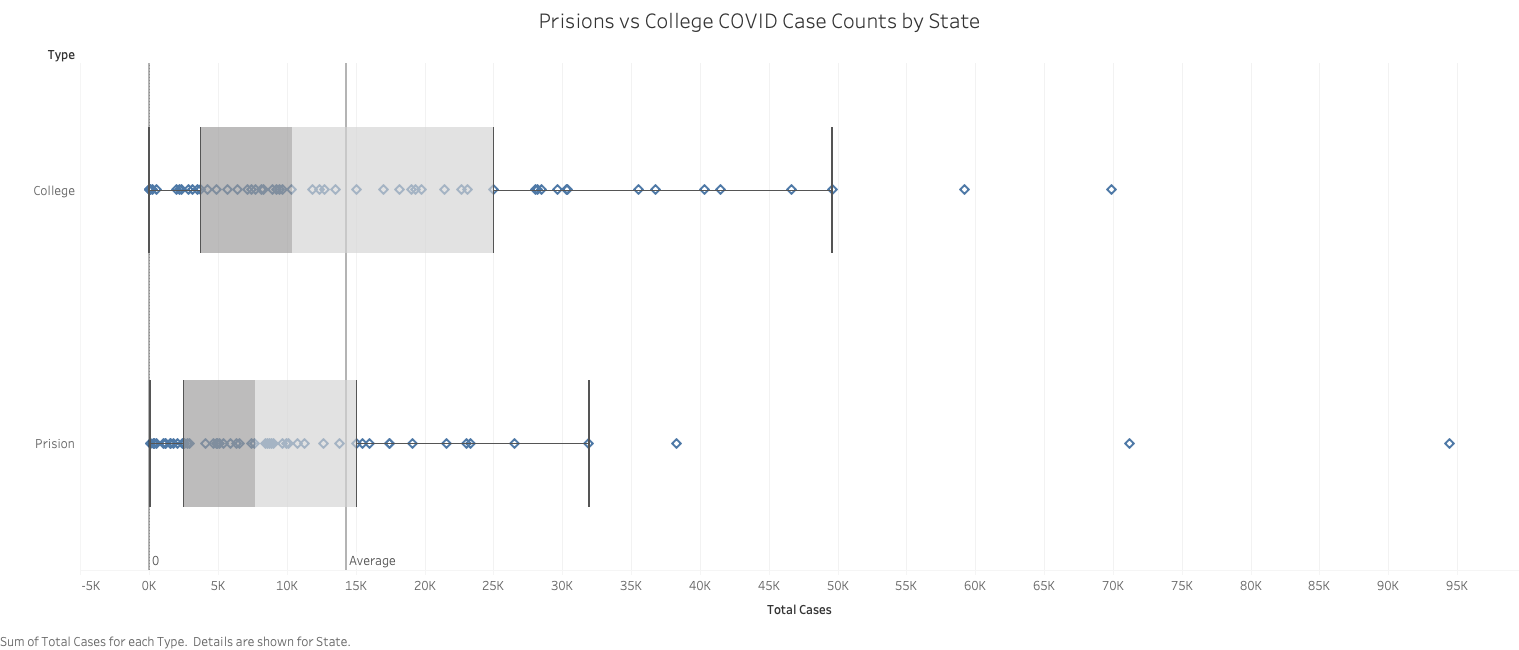
Inmate population and cases, created by Tableau. This may help us investigate whether a linear regression model exist between total inmate population and inmate cases.

This is a Tableau map portraying the inmate population per county as a regional map, and this can help show the density of incarcerated people in each county. We plan on deploying this to help us determine, through a linear regression model, whether a significant correlation exists between inmate population and inmate cases. How does this then connect with your research? There seem to be strong outliers in the southwest. Will you be focusing on those areas in a specific way within your narrative?



State-wise Total Deaths comparison between inmate and officer, in different prison systems. It may illustrate that different facilities have their own characteristic response to COVID-19. Created by Tableau.

This stacked bar chart was created in Tableau and compares the total COVID-related deaths for the different prison systems of each state. We believe this can illustrate how different facilities may have their own characteristic responses to COVID-19 and help us understand whether deficits in care may disproportionately exist in certain systems over others. This does not have any more stages of processing to go through. I might suggest a follow up to this visualization that removes the prisons and visualizes the other facilities and the prisons separately because there is such a drastic difference in scale. As you think about your narrative, consider how this chart might allow you to introduce your reader to these different facilities. Don’t assume they know how they differ. What’s important for your reader to understand about these numbers? For example, can you do some research around why New York and Florida have higher number of deaths in their jail facilities. New York shows that way more officers affiliated with jail facilities than inmates…that seems very strange to me. Why might that be, or perhaps there’s an error in the data?



**What the data visualization is meant to show**

* This compares COVID case count totals of colleges against prisons on a per state basis. By using a box and whisker plot, this graph helps to give a better understanding of how both systems performed in the country and helps to isolate outlier data points. Note that it needed to be shrunk to fit here but it will be much larger and easier to read on our website. I see the value in this graph, but it is not easily legible. Are the dots the states? If that is the case and they are easily exposed in the hover-over, that will be helpful for the outliers, but not as useful for those close together. I think perhaps that will be fixed when it is not shrunk to fit this space. In any case, you will want to make sure that you have a description that carefully walks the reader through what they are looking at and what you want them to get out of it.

**How you think it might fit in with your narrative**

* I believe we are beginning to find that while some prisons might have worse health outcomes, overall, these two very different institutions tended to have somewhat similar outcomes when it comes to the spread of the virus. We are missing a key data point though by not having access to death rates of college students and staff. If we had that as well, we could make more definitive comparisons regarding health outcomes between the two different institutions. Yes, I feel like this is a bit of a stretch without more data. Perhaps it can be reframed as future research that you would like to see the field explore?

**Where you are at within the visualization process (if there still data cleaning to do, anomalies to fix, or styling to add)**

* Our data has already been fully and thoroughly cleaned. With respect to this plot, it is in a near to final state. It could use perhaps a splash of color and small stylistic updates, but it clearly depicts what we desire without anomaly.

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