CS625 Project Milestone2

By Carol Zhou

Topic: Examining the Interplay Between Innovation, GDP, and Unemployment of U.S. States

The **final data sources** will be uploaded to GitHub, including:

- U.S. states' shape files for mapping
- Files containing GDP and population data for each state
- Files containing data on patents granted and R&D expenditure as a percentage of GDP
- Employment rate data file"

Refined Questions:

Question 1: Factors that affect the GDP of each state (Focus on Technology Factors).

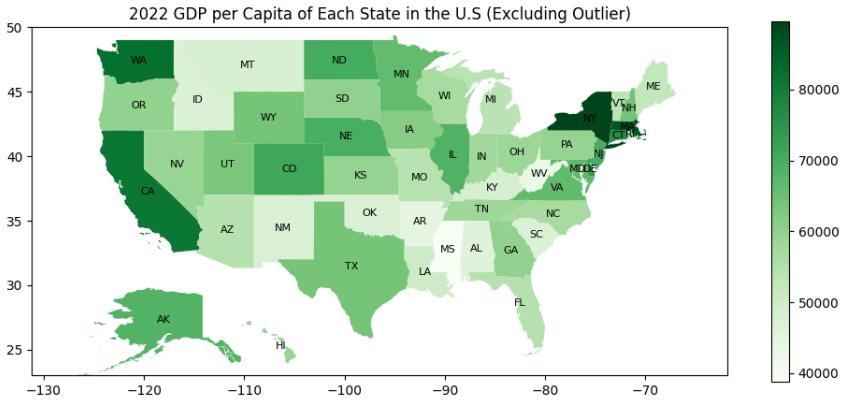
I will analyze two factors. One is the percentage of G&D (Research and Development) expenditure in the state GDP, which is akin to 'investment into technology'. The other is patents issued to each state, which is akin to 'output of the technology'.

Question 2: Does high GDP mean high employment rate?

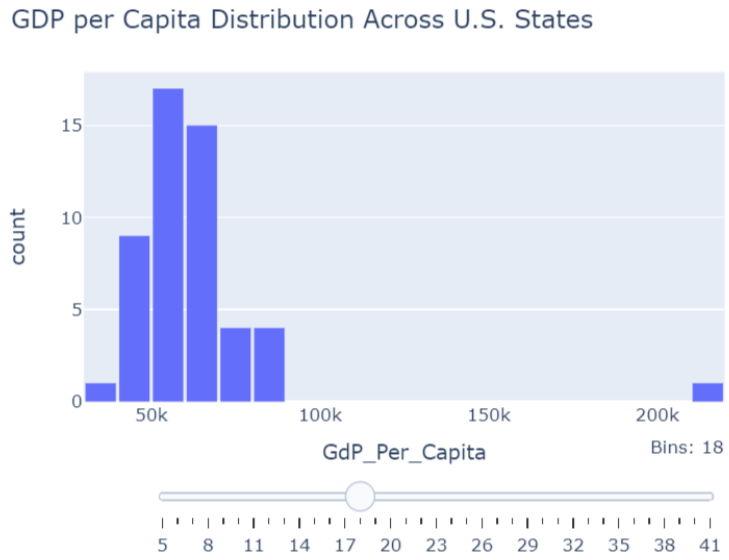
Hypothesis: High GDP means more opportunities, making it easier to find jobs.

Stech of Graphs:

Since my questions all pertain to state GDP per capita, I will begin by illustrating the overall picture of GDP per capita with a map.



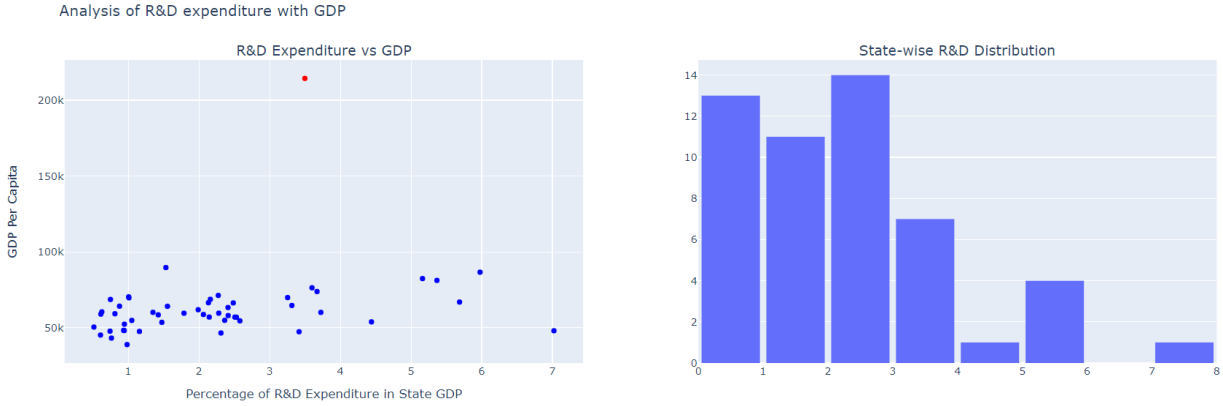
I will also plot the distribution of state personal GDP with a histogram. The plot will be interactive, featuring a slider bar that allows the user to adjust the number of bins displayed, as shown below:



Question1 Technology vs. GDP

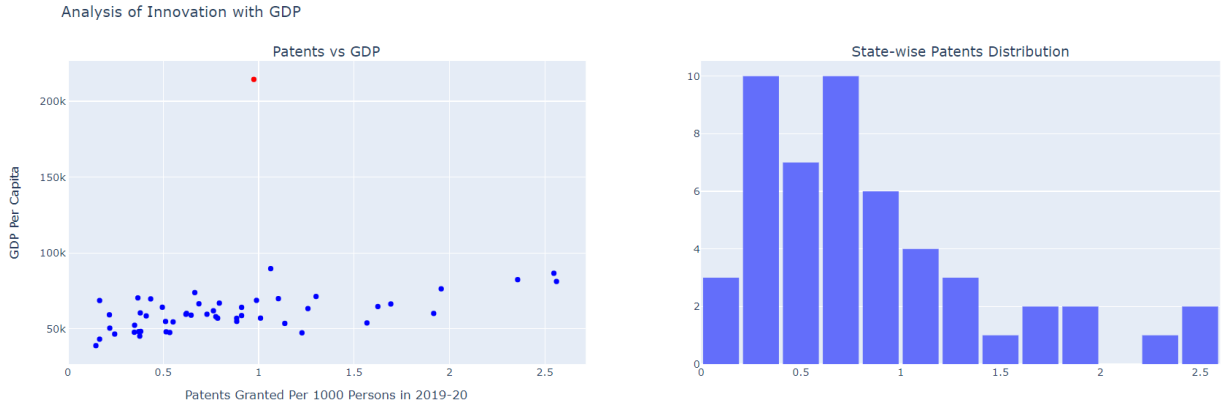
1) R&D expenditure vs. GDP

I plotted a scatterplot to examine the correlation between GDP and R&D expenditure, along with a histogram depicting the distribution of R&D expenditure state-wise. However, due to the presence of an outlier colored in red, it is difficult to discern the correlation. By hovering over the data points, it reveals that the outlier is Washington DC. Further research is following.



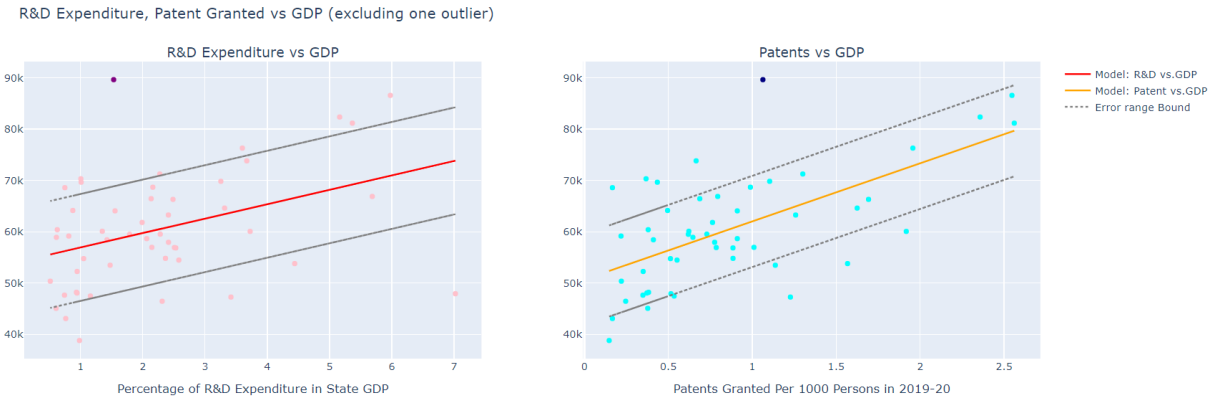
2) Patents Granted vs. GDP

As for Patents vs GDP, it looks similar to the previous analysis, with an outlier being Washington DC, which exhibits an unusually high GDP per capita. Further visualization is needed.



3) Exclude the outlier and redo the visualization..

After excluding DC, I created models for R&D vs. GDP and Patents vs. GDP by performing linear regression. Here, we also identified an outlier. By hovering over the data point, we determined it to be New York. I used a darker color to make it more obvious. The graphs are shown below.



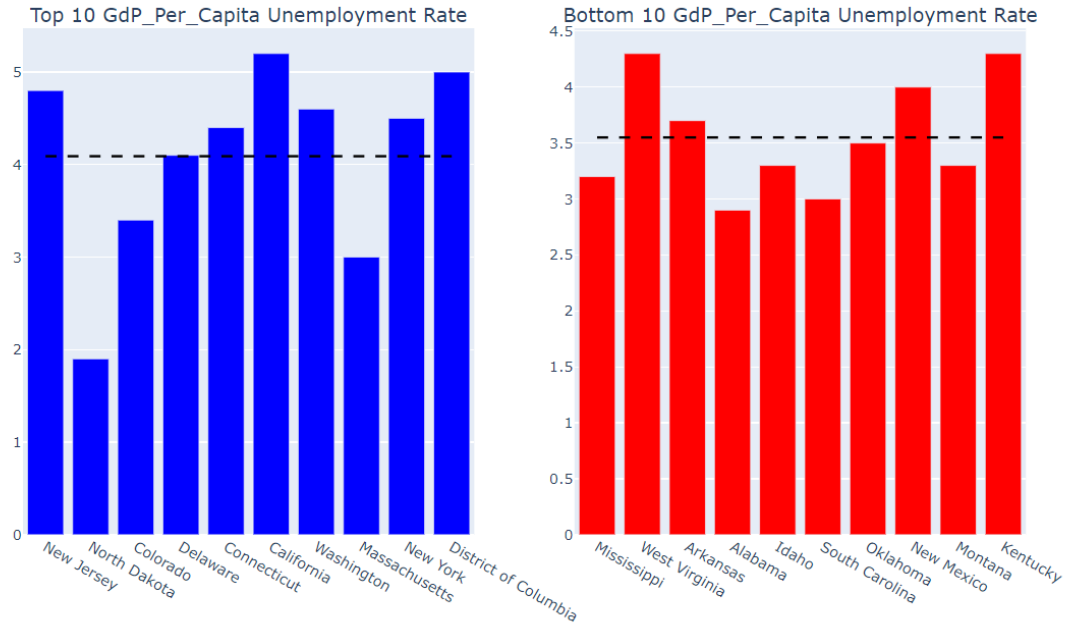
Conclusion from the visualization:
Washington, D.C. stands as the epicenter of political power and the largest taxing authority in the United States. In 2017, over 3 trillion dollars flowed through the institutions headquartered there, contributing to its exceptionally high GDP per capita. Similarly, New York State emerges as an outlier in statistical plots due to its status as the global economic hub, resulting in a notably elevated GDP per capita.

There's another exception: New Mexico. In contrast to New York and Washington, D.C., New Mexico exhibits an unusually high investment in research and development (R&D) compared to its GDP. Further investigation reveals that New Mexico receives significant state and federal funding for research, likely due to its largely arid desert landscape. This makes it conducive to infrastructure projects such as atomic research or clandestine studies on extraterrestrial life, which consume money yet do not reflect in its GDP. I drew inspiration from the film 'Oppenheimer,' which recently received many Oscars.

In general, the GDP per capita of other states tends to correlate with their investments in technology and innovation, as evidenced by the number of patents granted.

Question2: Does a high GDP per capita ensure a low unemployment rate?

I plotted bar charts for the top 10 states and bottom 10 states based on GDP and unemployment rate, adding a dotted line to represent the average. It suggests that the unemployment rate is slightly lower in states with high GDP per capita compared to states with low GDP per capita.



What's special:

1. In the map plot, I learned to move Alaska and Hawaii closer to the 48 contiguous states and resize Alaska to improve the map's appearance.
2. On the GDP histogram, I added a slider bar that can change the number of bins.
3. I used data retrieval to perform regression analysis.
4. All plots are interactive, allowing users to read the values by hovering.
5. Juxtaposition was employed in the presentation.