GitHub Repo: https://github.com/milivij/SI206_Final

In addition to your API activity results, you will be creating a report for your overall project.

The report must include:

A. The goals for your project including what APIs/websites you planned to work with and what data you planned to gather (10 points)

Our goal in this project was to examine COVID cases and death rates and explore what factors could have been affecting them, including medication. Originally, we had planned to work with a COVID dataset that included case and death numbers per state, and an FDA dataset that had information on the medications being administered to COVID patients, including how many people were receiving each medication. We planned to work with the data to find how drug implementation varied across states and affected covid case numbers.

B. The goals that were achieved including what APIs/websites you actually worked with and what data you did gather (10 points)

We ended up finding it more interesting to explore what factors were linked with different amounts of COVID cases in the states. We collected the COVID case and death numbers from the COVID database using an API call, and we ended up using Census data to get more information about the states. Specifically, the education rates were very interesting to us, but we pulled a lot more data to explore, such as the poverty rate in every state, high school and bachelor's degree graduates, and median household income. The last data we gathered was from web scraping a Wikipedia page to collect data on which states voted Republican or Democratic in the 2020 election. We achieved the goal of exploring some factors that were related to the COVID case rates through creating graphs—we specifically focused on which party the state voted for in the 2020 election , bachelor's degree numbers and median household income .

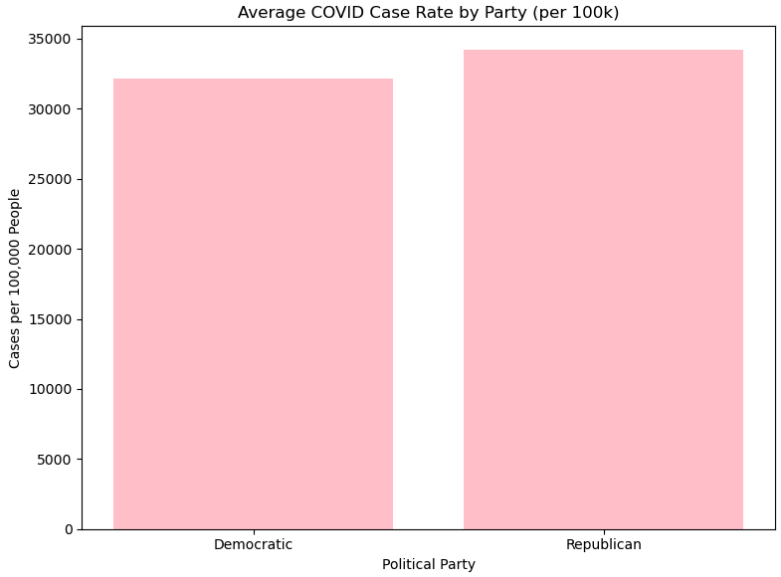
C. The problems that you faced (10 points)

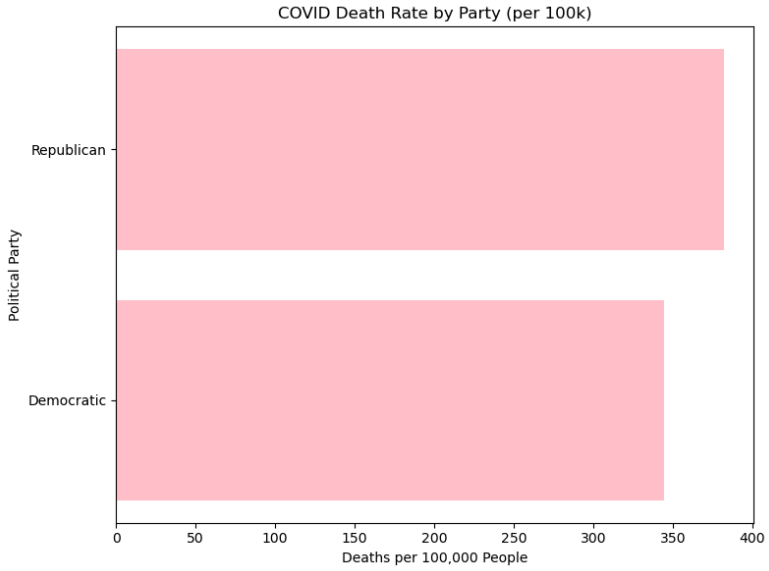
We faced multiple problems in this project. Firstly, figuring out how to code together, as we had to plan Git commits and make sure nothing was conflicting. We worked around that by increasing communication, using Live Share and meeting up to code together. Secondly, navigating the Census data compared to the other database—accessing the Census data was harder due to the way it is set up on their website. We accessed the wrong data multiple times before figuring it out. Thirdly, figuring out how to visualize what we wanted to convey was a difficult task—thinking of the calculations that we would have to implement to get the correct things to graph.

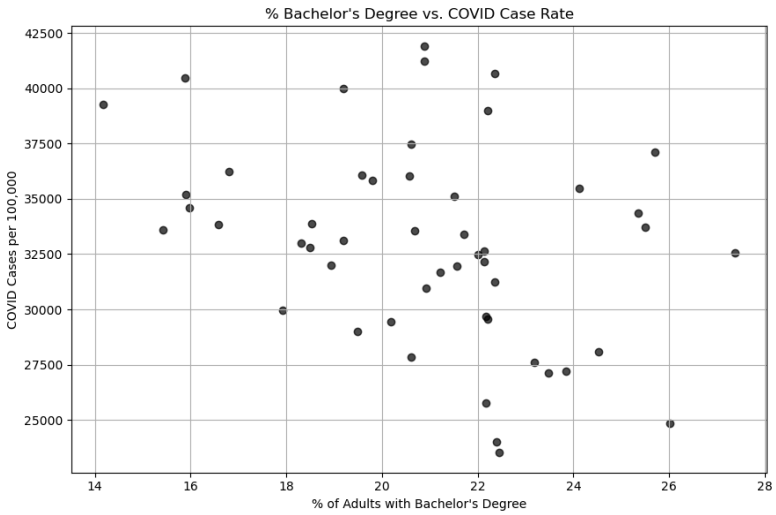
D. The calculations from the data in the database (i.e. a screenshot) (10 points)

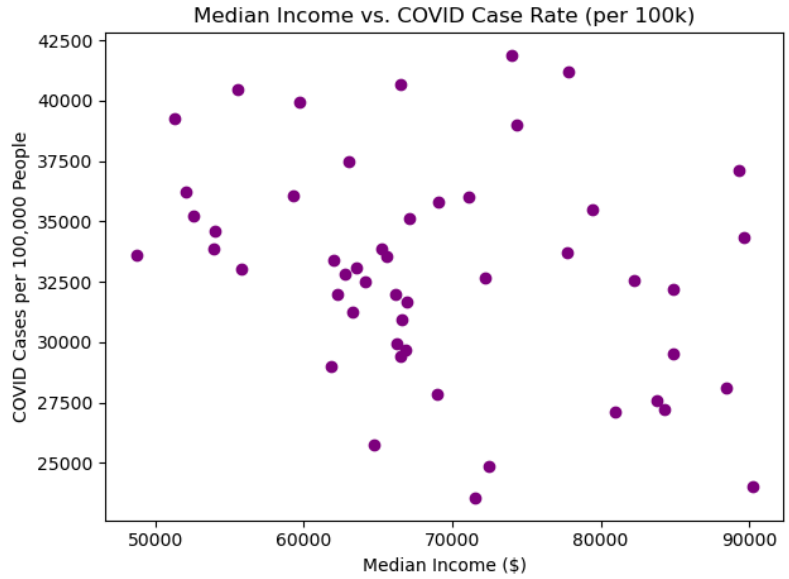
```
=== Raw State Data ===
Alabama: 1659936 cases, 21138 deaths, $53913 median income, Republican
Alaska: 301513 cases, 1485 deaths, $77845 median income, Republican
Arizona: 2607545 cases, 34402 deaths, $69056 median income, Democratic
Arkansas: 1062606 cases, 13246 deaths, $52528 median income, Republican
California: 12711918 cases, 112443 deaths, $84907 median income, Democratic
Colorado: 1874746 cases, 15552 deaths, $82254 median income, Democratic
Connecticut: 983652 cases, 12354 deaths, $83771 median income, Democratic
Delaware: 350706 cases, 3663 deaths, $71091 median income, Democratic
Florida: 8048191 cases, 95206 deaths, $63062 median income, Republican
Georgia: 3287483 cases, 44069 deaths, $66559 median income, Democratic
Hawaii: 418369 cases, 2167 deaths, $84857 median income, Democratic
Idaho: 525825 cases, 5482 deaths, $66474 median income, Republican
Illinois: 4136659 cases, 42005 deaths, $72205 median income, Democratic
Indiana: 2208419 cases, 28018 deaths, $62743 median income, Republican
Iowa: 1058274 cases, 10797 deaths, $65600 median income, Republican
Kansas: 946564 cases, 10229 deaths, $64124 median income, Republican
Kentucky: 1808735 cases, 19183 deaths, $55573 median income, Republican
Louisiana: 1684058 cases, 19270 deaths, $52087 median income, Republican
Maine: 346182 cases, 3394 deaths, $64767 median income, Democratic
Maryland: 1451442 cases, 17929 deaths, $90203 median income, Democratic
Massachusetts: 2367542 cases, 25586 deaths, $89645 median income, Democratic
Michigan: 3306221 cases, 44762 deaths, $63498 median income, Democratic
Minnesota: 1900794 cases, 16458 deaths, $77720 median income, Democratic
Mississippi: 1000415 cases, 13474 deaths, $48716 median income, Republican
Missouri: 1780715 cases, 22778 deaths, $61847 median income, Republican
Montana: 333758 cases, 3712 deaths, $63249 median income, Republican
Nebraska: 574399 cases, 5063 deaths, $66817 median income, Republican
Nevada: 923059 cases, 12485 deaths, $66274 median income, Democratic
New Hampshire: 382013 cases, 3329 deaths, $88465 median income, Democratic
New Jersey: 3295020 cases, 36873 deaths, $89296 median income, Democratic
New Mexico: 725653 cases, 9623 deaths, $53992 median income, Democratic
New York: 7587861 cases, 83374 deaths, $74314 median income, Democratic
North Carolina: 3501404 cases, 29059 deaths, $61972 median income, Republican
North Dakota: 309990 cases, 2513 deaths, $66519 median income, Republican
Ohio: 3741277 cases, 43896 deaths, $62262 median income, Republican
Oklahoma: 1306350 cases, 16157 deaths, $55826 median income, Republican
Oregon: 992925 cases, 9764 deaths, $71562 median income, Democratic
Pennsylvania: 3565499 cases, 51480 deaths, $68957 median income, Democratic
Rhode Island: 443803 cases, 4166 deaths, $74008 median income, Democratic
South Carolina: 1857853 cases, 20311 deaths, $59318 median income, Republican
South Dakota: 282895 cases, 3231 deaths, $66143 median income, Republican
Tennessee: 2729641 cases, 30729 deaths, $59695 median income, Republican
Texas: 9190299 cases, 104793 deaths, $66963 median income, Republican
Utah: 1137615 cases, 5719 deaths, $79449 median income, Republican
Vermont: 155080 cases, 1139 deaths, $72431 median income, Democratic
Virginia: 2315784 cases, 23748 deaths, $80963 median income, Democratic
Washington: 2070848 cases, 17339 deaths, $84247 median income, Democratic
West Virginia: 703856 cases, 8247 deaths, $51248 median income, Republican
Wisconsin: 2043838 cases, 16758 deaths, $67125 median income, Democratic
Wyoming: 196126 cases, 2126 deaths, $65204 median income, Republican

Average COVID cases across states: 2163907.12

=== Average Case Rate by Party (per 100k) ===
Democratic: 32173.19 cases per 100,000
Republican: 34181.49 cases per 100,000

=== Death Rate by Party (per 100k) ===
Democratic: 344.38 deaths per 100,000
Republican: 381.99 deaths per 100,000
```

E. The visualization that you created (i.e. screenshot or image file) (10 points)









F. Instructions for running your code (10 points)

Write “python final.py” in the terminal and select the enter button. This will run our code.

G. Documentation for each function that you wrote. This includes describing the input and output for each function (20 points)

get_covid_data() - Gets the COVID data from the COVID API and saves it. No inputs; output is the list of dictionaries or None if the file doesn't exist.

get_poverty_data() - Gets the data from the Census and saves it into its corresponding JSON file. No inputs; outputs the JSON data.

get_state_election_results() - Webscrapes Wikipedia to find the party each state leaned toward in the 2020 election. No inputs; outputs a dictionary with each state assigned "Democratic", "Republican", or "Neither".

convert_poverty_to_dict() - Converts the poverty JSON into a dictionary for easier use. No inputs; outputs the dictionary.

set_up_covid_database(db_name) - Initializes the database to hold all of our data. The input is the database name; the output is a tuple of the cursor and connection.

create_combined_table(cur, conn) - Creates a main table that holds relevant data for each state. The inputs are the cursor and connection. No Outputs.

create_parties_table(cur, conn) - Creates the party table which assigns a unique key to each political party. The inputs are the cursor and connection. No Outputs.

get_or_create_party_id(party_name, cur) - Retrieves or inserts the party into the database and returns its ID. The input is the party name; the output is the corresponding party ID.

insert_combined_data(covid_data, election_dict, cur, conn) - Joins and inserts the combined COVID data with political party information into the state_data table. Inputs: covid_data, election_dict, cur, conn. No Outputs.

create_split_tables(cur, conn) - Creates two separate tables, states and poverty_stats, linked by a foreign key to state codes. Inputs are the cursor and connection. No Outputs.

insert_split_poverty_data(cur, conn) - Inserts up to 25 rows of poverty data into the states and poverty_stats tables. Inputs are the cursor and connection.No Outputs.

get_joined_data(cur) - Retrieves joined data from multiple tables with all relevant information. Input is the cursor; output is a list of tuples: (state, cases, deaths, median_income, party_name).

write_results_to_file(joined_data) - Calculates the average COVID cases per state and writes results to a text file. Input is the joined data. No Outputs. (writes to file)

plot_death_rate_by_party(cur) - Calculates average death rate and plots a horizontal bar graph by party. Input is the cursor; output is a saved PNG graph.

plot_avg_case_rate_by_party(cur) - Plots a bar chart of average COVID case rate (per 100k) by party. Input is the cursor; output is a saved PNG graph.

plot_income_vs_case_rate(cur) - Creates a scatter plot of median income vs. COVID case rate. Input is the cursor; output is a saved PNG graph.

plot_education_vs_case_rate(cur) - Plots the percentage of bachelor's degree holders vs. COVID case rate. Input is the cursor; output is a saved PNG graph.

main() - Calls all the functions in the correct order to fetch, process, store, and visualize the data. No inputs; no outputs.

H. You must also clearly document all resources you used. The documentation should be of the following form (20 points)

Date	Issue Description	Location of Resource	Result (did it solve the issue?)
April 13th	Didn't know the regex	Used ChatGPT to help fix issues in regex while scraping data	Solved the issue perfectly!
April 15th	Didn't know how to put multiple apis into one table, initially attempted to do it with each api having its own individual table.	Asked GSI during discussion for help.	Solved the issue perfectly!