Grace Dwyer and Huda Shulaiba

APIs, SQL, and Visualizations: https://github.com/dwyerg/si206-final-project/

Original Goals

Initially, we had planned to use an API with criminal data from the FBI and the official Census website. Our intention was to collect general population data about race, ethnicity, sex, and age for each state by scraping the Census site and comparing it to the criminal data collected to analyze if the incarcerated demographics were proportional to the overall demographics of the area. With visualizations clearly mapping out the comparisons, we were hoping to be able to pinpoint inequalities in the system.

Goals Achieved

APIs/websites used: Wikipedia historical demographics, Wanted Criminals: FBI API

Both the original websites and APIs that we were planning to use caused us difficulty, and we had to switch to alternate resources to gather our data from. Rather than using the Census data, we scraped Wikipedia to gather demographic data by state for the years 2020 and 2019 as well as the corresponding demographics for people in poverty. Similarly, the initial API was not functioning properly and the data we were getting from it did not match what we needed to achieve the desired analysis. Instead, we used a different FBI API that provided data for wanted criminals in a format that was understandable and able to be transferred into the database. From there, we were able to completed the analysis that we had planned to and compared the race demographics for two states between the overall Wikipedia data and the wanted criminals in that state. The results of this analysis can be seen in the visualizations below.

Problems

Problem 1: I originally planned on using the census quick facts page to scrape the data from, but the way it was formatted made it so running the code took way too long, and it was not efficient at all. I also considered using kff.org, but their HTML was so many nested divs that it was ridiculously hard to scrape the information I needed.

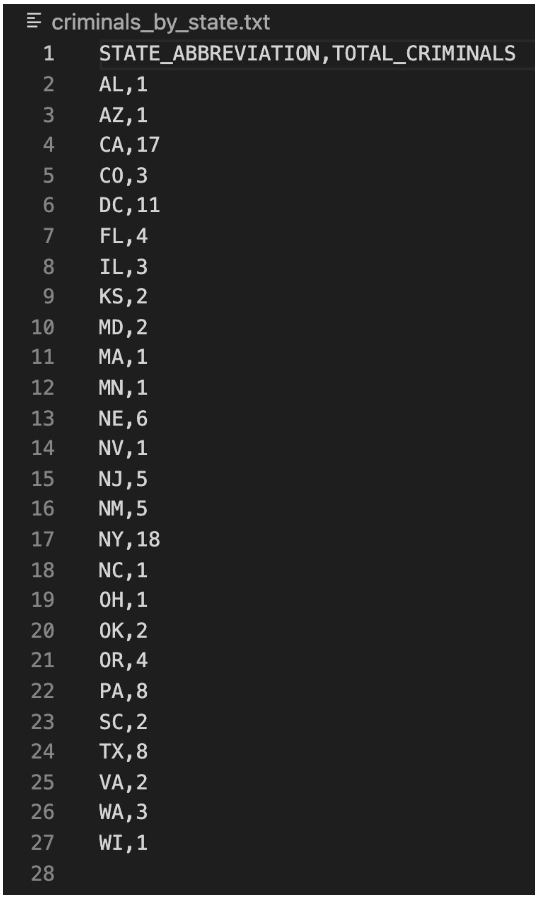
- Solution: I found the census data all on a Wikipedia page and scraped the information from there. For the poverty statistics, I just downloaded the CSV file from the website since we satisfied the project requirements with the Wikipedia page and read in the data from there.

Problem 2: The API that we originally planned to use was not functioning correctly. They required an API key in order to access the data and make API calls, which would've been fine except that the API key that I received originally did not work, causing me to receive a 403 (authentication) error. Not wanting to abandon our entire plan, I reached out to the customer service for the API and asked them with help formatting my queries and inquired about why the API key wasn't working. They were able to get my API key into the system, and I finally began making API calls, only to realize that the data I was receiving didn't fulfill the requirements for the analysis we wanted to perform.

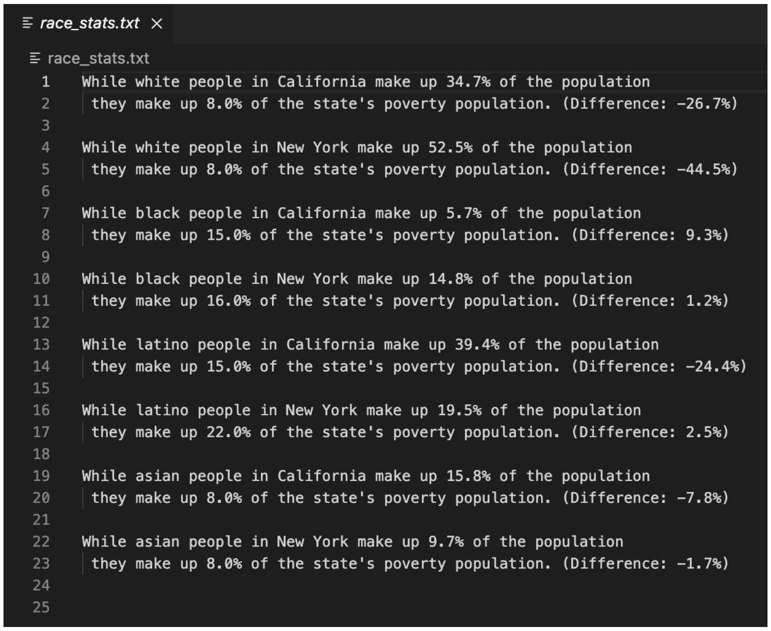
- Solution: At this point, instead of persevering and trying to change our plan to fit the API, I decided to do a bunch of research to find a different API that would be easier to use and fit our needs better. The wanted criminals API didn't even require an API key, and I was able to process the data received from the API calls much more efficiently in order to then preform the necessary calculations and analysis.

Grace Dwyer and Huda Shulaiba

Calculations



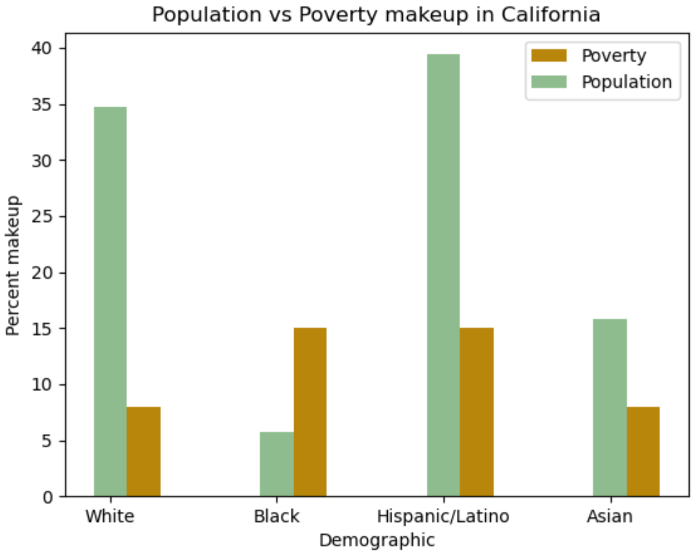
```
STATE_ABBREVIATION,TOTAL_CRIMINALS
AL,1
AZ,1
CA,17
CO,3
DC,11
FL,4
IL,3
KS,2
MD,2
MA,1
MN,1
NE,6
NV,1
NJ,5
NM,5
NY,18
NC,1
OH,1
OK,2
OR,4
PA,8
SC,2
TX,8
VA,2
WA,3
WI,1

```

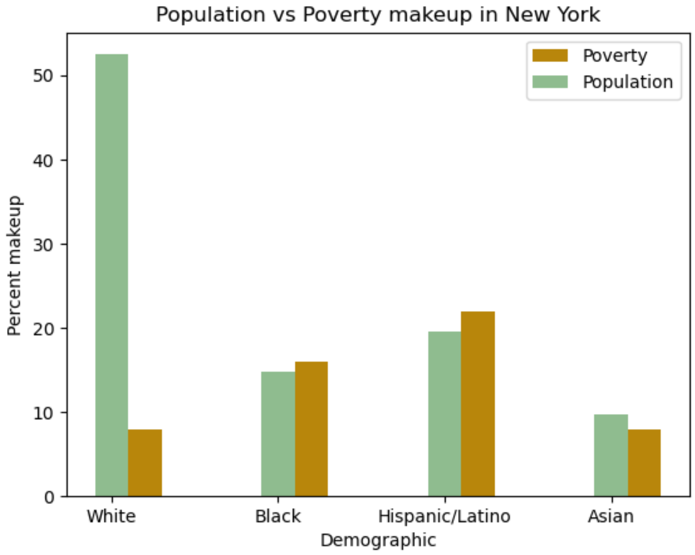


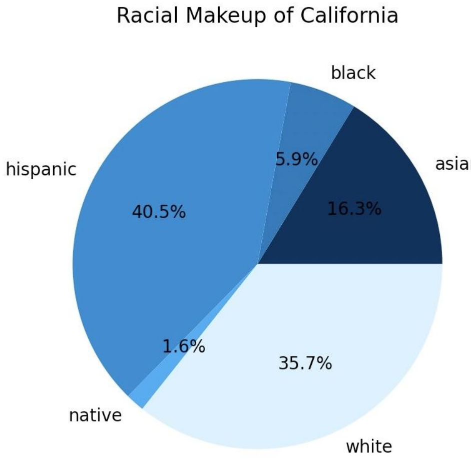
```
While white people in California make up 34.7% of the population
 they make up 8.0% of the state's poverty population. (Difference: -26.7%)

While white people in New York make up 52.5% of the population
 they make up 8.0% of the state's poverty population. (Difference: -44.5%)

While black people in California make up 5.7% of the population
 they make up 15.0% of the state's poverty population. (Difference: 9.3%)

While black people in New York make up 14.8% of the population
 they make up 16.0% of the state's poverty population. (Difference: 1.2%)

While latino people in California make up 39.4% of the population
 they make up 15.0% of the state's poverty population. (Difference: -24.4%)

While latino people in New York make up 19.5% of the population
 they make up 22.0% of the state's poverty population. (Difference: 2.5%)

While asian people in California make up 15.8% of the population
 they make up 8.0% of the state's poverty population. (Difference: -7.8%)

While asian people in New York make up 9.7% of the population
 they make up 8.0% of the state's poverty population. (Difference: -1.7%)

```

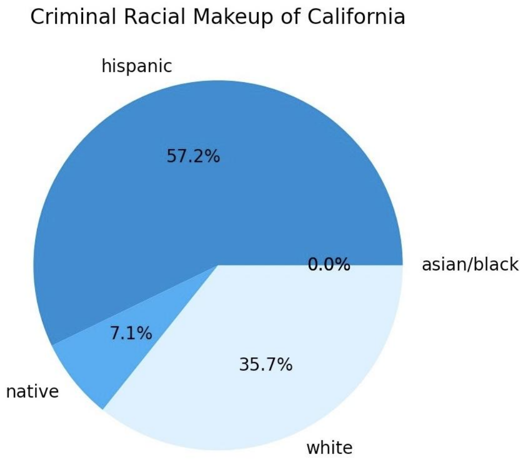
Grace Dwyer and Huda Shulaiba

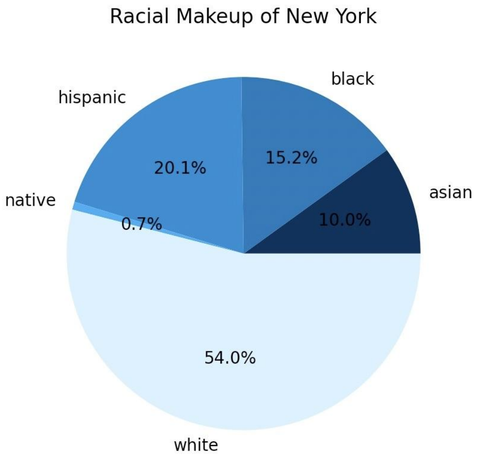
Visualizations

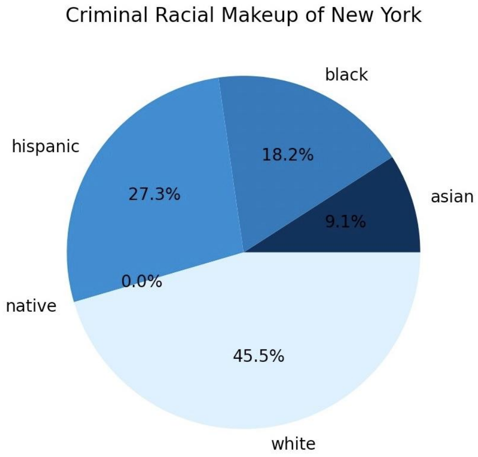












Grace Dwyer and Huda Shulaiba

Code Instructions

Run the census_data.py document first because the state table is needed in the FBI_data.py document.

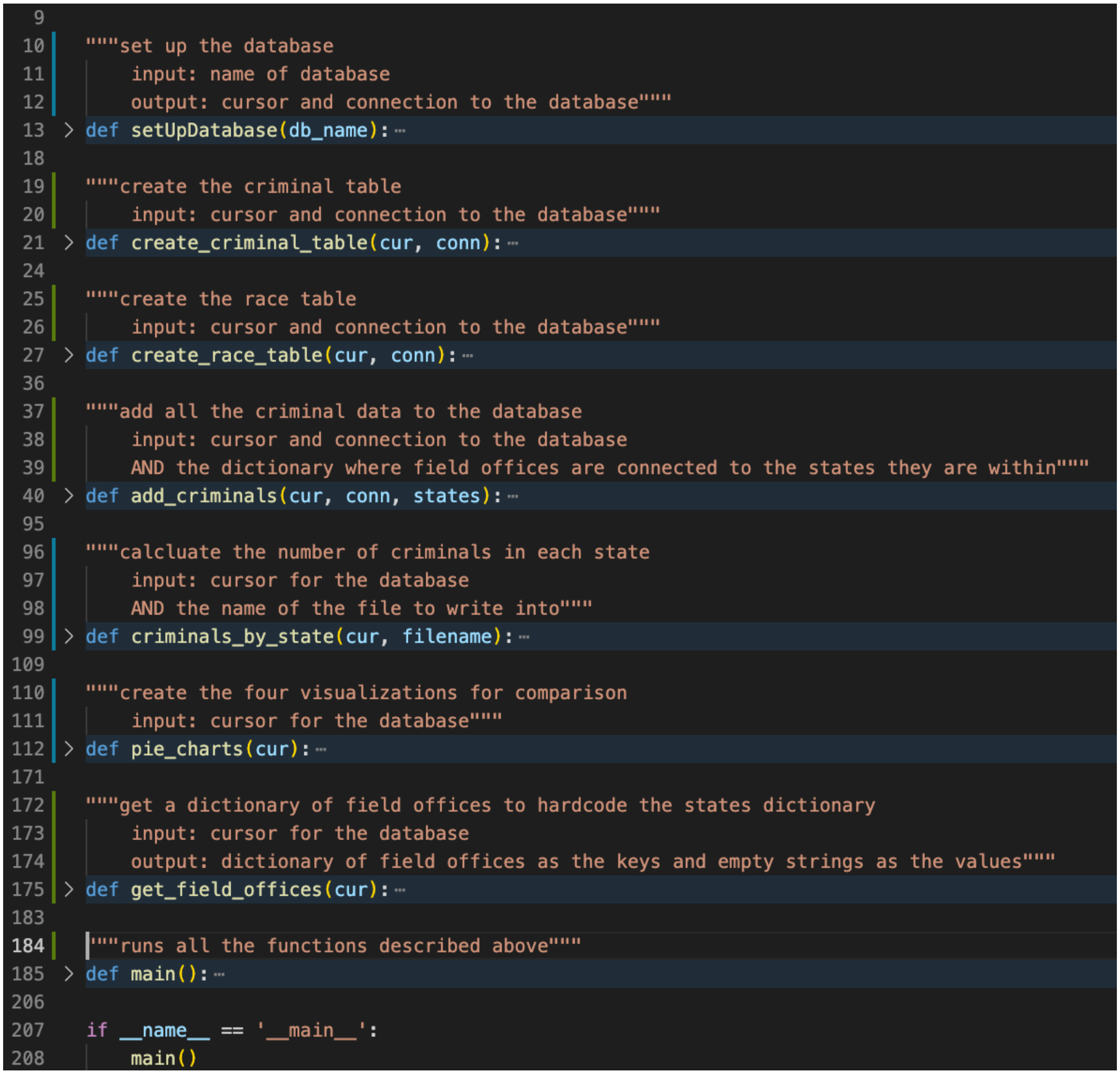
FBI_data.py
Find the main function at the bottom of the code doc. In order to set up the database, run the line of code under “`SET UP THE DATABASE`”. Then to create the tables within the database, run the lines underneath “`CREATE THE CRIMINAL AND RACE TABLES`” and populate the criminal data table with the states dictionary and the add_criminals function under “`ADD CRIMINAL DATA TO THE DATABASE`”. You would need to run the code 9 times to get all the data. You do not need to uncomment get_field_offices function. To run the calculations, run the line under “`CALCULATE NUMBER OF CRIMINALS BY STATE`”. Lastly, to create the visualizations, run the line under “`CREATE VISUALIZATIONS`”. Once you have done any number of these, simply run the code.

census_data.py
Scroll down to the main function at the bottom. To scrape the website set up the tables in the database, run the code (for this the lines under these comments are used: “`SET UP STATE IDS,`” “`SCRAPE DATA FROM WEBSITES INTO LISTS`,” and “`SET UP TABLES AND PUT SCRAPED DATA IN DATABASE`”). To make the calculations and write them to a txt file, use the code under “`CALCULATE DATA AND WRITE TO TXT FILE`”. Finally, to make the visualizations, uncomment the code under “`VISUALIZATIONS`” along with the code under “`CALCULATE DATA AND WRITE TO TXT FILE`”.

Grace Dwyer and Huda Shulaiba

Documentation

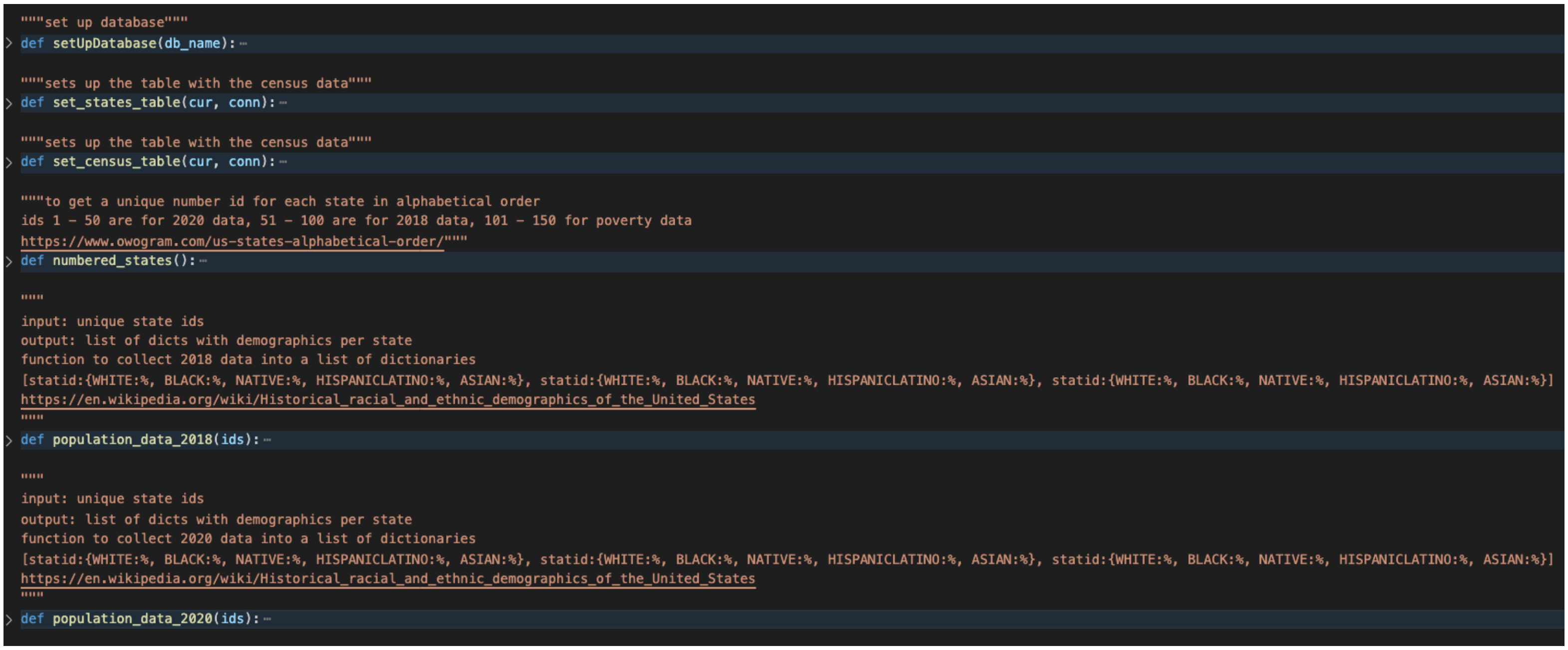
FBI_data.py



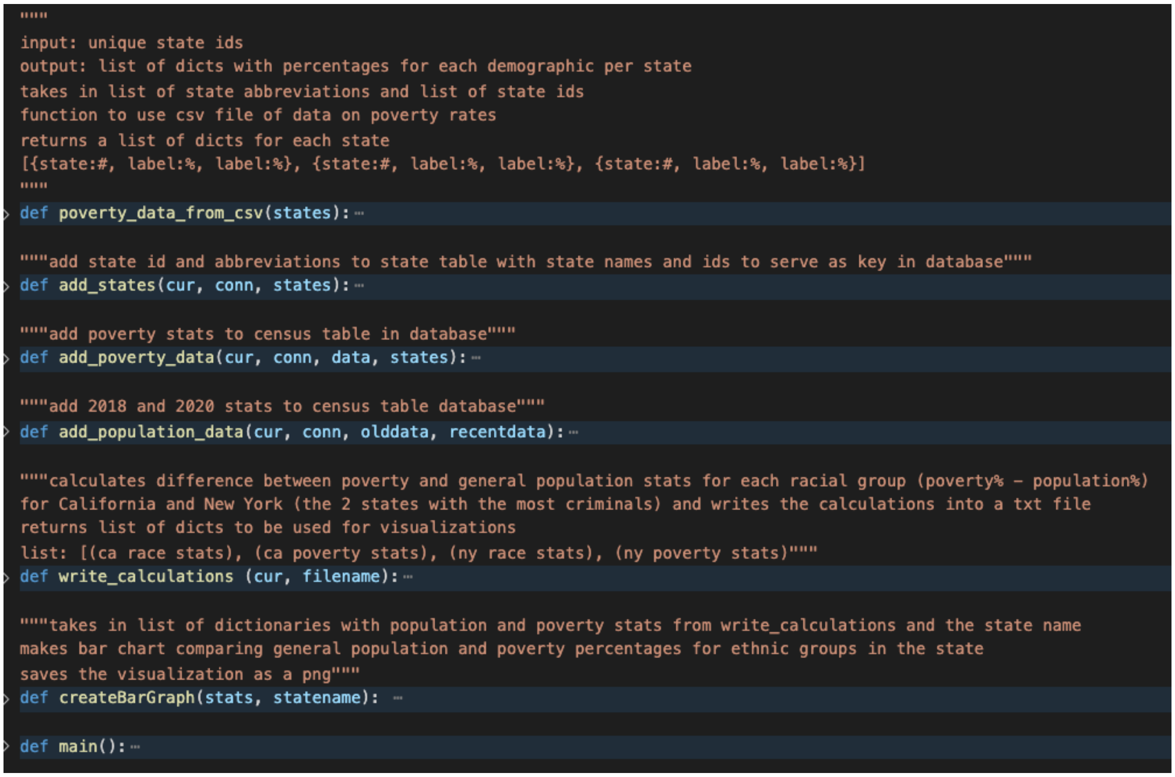
```python

"""set up the database
    input: name of database
    output: cursor and connection to the database"""
def setUpDatabase(db_name):⋯

"""create the criminal table
    input: cursor and connection to the database"""
def create_criminal_table(cur, conn):⋯

"""create the race table
    input: cursor and connection to the database"""
def create_race_table(cur, conn):⋯

"""add all the criminal data to the database
    input: cursor and connection to the database
    AND the dictionary where field offices are connected to the states they are within"""
def add_criminals(cur, conn, states):⋯

"""calcluate the number of criminals in each state
    input: cursor for the database
    AND the name of the file to write into"""
def criminals_by_state(cur, filename):⋯

"""create the four visualizations for comparison
    input: cursor for the database"""
def pie_charts(cur):⋯

"""get a dictionary of field offices to hardcode the states dictionary
    input: cursor for the database
    output: dictionary of field offices as the keys and empty strings as the values"""
def get_field_offices(cur):⋯

"""runs all the functions described above"""
def main():⋯

if __name__ == '__main__':
    main()
```

Grace Dwyer and Huda Shulaiba

Census_data.py



```
"""set up database"""
def setUpDatabase(db_name):⋯

"""sets up the table with the census data"""
def set_states_table(cur, conn):⋯

"""sets up the table with the census data"""
def set_census_table(cur, conn):⋯

"""to get a unique number id for each state in alphabetical order
ids 1 - 50 are for 2020 data, 51 - 100 are for 2018 data, 101 - 150 for poverty data
https://www.owogram.com/us-states-alphabetical-order/"""
def numbered_states():⋯

"""
input: unique state ids
output: list of dicts with demographics per state
function to collect 2018 data into a list of dictionaries
[statid:{WHITE:%, BLACK:%, NATIVE:%, HISPANICLATINO:%, ASIAN:%}, statid:{WHITE:%, BLACK:%, NATIVE:%, HISPANICLATINO:%, ASIAN:%}, statid:{WHITE:%, BLACK:%, NATIVE:%, HISPANICLATINO:%, ASIAN:%}]
https://en.wikipedia.org/wiki/Historical_racial_and_ethnic_demographics_of_the_United_States
"""
def population_data_2018(ids):⋯

"""
input: unique state ids
output: list of dicts with demographics per state
function to collect 2020 data into a list of dictionaries
[statid:{WHITE:%, BLACK:%, NATIVE:%, HISPANICLATINO:%, ASIAN:%}, statid:{WHITE:%, BLACK:%, NATIVE:%, HISPANICLATINO:%, ASIAN:%}, statid:{WHITE:%, BLACK:%, NATIVE:%, HISPANICLATINO:%, ASIAN:%}]
https://en.wikipedia.org/wiki/Historical_racial_and_ethnic_demographics_of_the_United_States
"""
def population_data_2020(ids):⋯
```

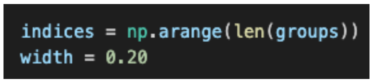


```
"""
input: unique state ids
output: list of dicts with percentages for each demographic per state
takes in list of state abbreviations and list of state ids
function to use csv file of data on poverty rates
returns a list of dicts for each state
[{state:#, label:%, label:%}, {state:#, label:%, label:%}, {state:#, label:%, label:%}]
"""
def poverty_data_from_csv(states):⋯

"""add state id and abbreviations to state table with state names and ids to serve as key in database"""
def add_states(cur, conn, states):⋯

"""add poverty stats to census table in database"""
def add_poverty_data(cur, conn, data, states):⋯

"""add 2018 and 2020 stats to census table database"""
def add_population_data(cur, conn, olddata, recentdata):⋯

"""calculates difference between poverty and general population stats for each racial group (poverty% - population%)
for California and New York (the 2 states with the most criminals) and writes the calculations into a txt file
returns list of dicts to be used for visualizations
list: [(ca race stats), (ca poverty stats), (ny race stats), (ny poverty stats)"""
def write_calculations (cur, filename):⋯

"""takes in list of dictionaries with population and poverty stats from write_calculations and the state name
makes bar chart comparing general population and poverty percentages for ethnic groups in the state
saves the visualization as a png"""
def createBarGraph(stats, statename): ⋯

def main():⋯
```

Grace Dwyer and Huda Shulaiba

Resources Used

Date	Issue Description	Location of Resource	Result

12/9/21	I couldn't figure out how to do a comparison bar chart with 2 bars per item.	https://codeboxsystems.com/tutorials/en/how-to-compare-values-bar-chart-python-matplotlib/	I was able to read their code and reverse engineer a way to apply the same concept to mine.
			
			`indices = np.arange(len(groups))`
			`width = 0.20`

12/9/21	Was struggling to make a legend for the bar chart.	https://www.geeksforgeeks.org/bar-plot-in-matplotlib/	I remembered how simple it was to make a legend. `plt.legend()`
12/1/21	Couldn't find adequate data concerning poverty by state demographics on the census site.	https://www.kff.org/other/state-indicator/poverty-rate-by-raceethnicity/?currentTimeframe=0&sortModel=%7B%22colId%22:%22Location%22,%22sort%22:%22asc%22%7D	I was able to use the chart on kff.org to get data to use in my calculations and comparisons.
11/24/21	Couldn't figure out how to run the API call with the necessary key.	https://aapi.io/api-directory/FederalBureauofInvestigationFBI_CrimeDataExplorer_CrimeDataAPI_v1	Explored their documentation and examples with help from GSIs and IAs. Eventually emailed them directly and found the formatting issue.
12/9/21	Wanted the colors to match across the bar charts.	https://medium.com/@kvnamipara/a-better-visualisation-of-pie-charts-by-matplotlib-935b7667d77f	Found a way to hard code the colors and make it specific to the slice of the pie for easier comparison.
12/12/21	Wanted to join three tables at once.	https://learnsql.com/blog/how-to-join-3-tables-or-more-in-sql/	I needed the races and the state abbreviations, so I wanted to see if I could join more than one table.
12/13/21	Wanted to delete a column in the database.	https://www.w3schools.com/sql/trysql.asp?filename=trysql_drop_column	Tried to do it the way they mentioned, instead ended up doing it directly to the database.