Cappy Funding

Group Members

Bryan Foo
Gongzi Chen
Yujie Jiang
Ziyang Chen

Table of Contents

Part 1: Project Requirements Compliance

- Two Data sources:
    - ✔ USA Spending API and USA Census
- Data analysis component
    - ✔ Five visualizations for the data analysis
- Specific components built by each member
    - ✔ Refer to the code responsibility part
- A visual or textual output
    - ✔ Refer to the following description of the three visualizations
- Each distinct component must be a subpackage
    - ✔ Refer to the relevant section
- Be able to run in a virtual environment
    - ✔ Refer to the code
- README.md
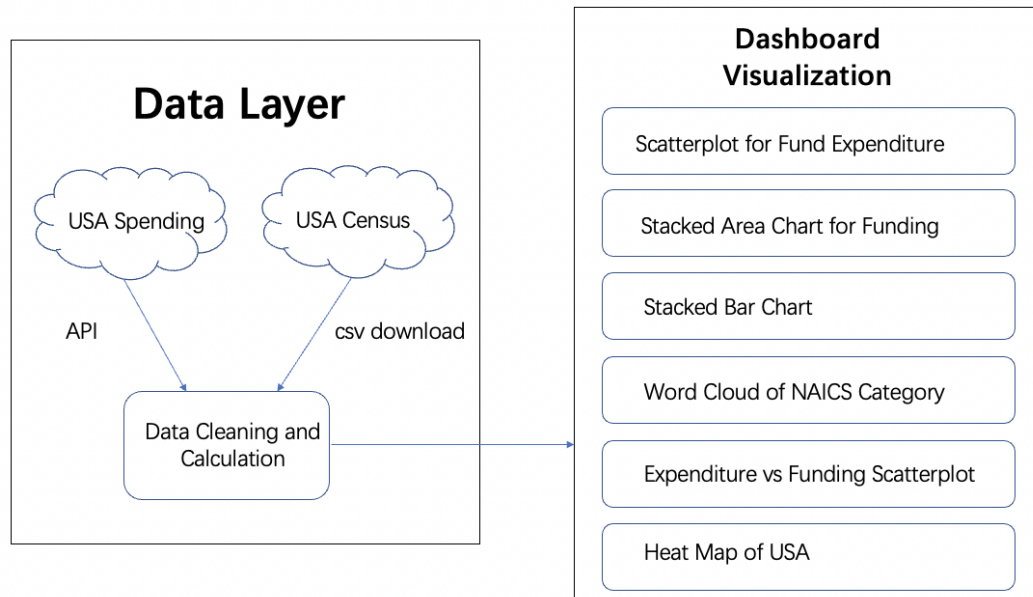   - ✔ Refer to the Git repository

Part 2: Project Abstract

The main purpose of this project, Cappy Funding, is to visualize the allocation of federal fundings in the US from 2016 to 2020, allowing the audience to have a better sense of (1) how funds at the federal level are spent on different industries and categories, and (2) how states manage expenditure in accordance to those categories. Our expected target audience include federal foundation managers, funding seekers, and others who are interested in the way federal funds are spent. The complete data set contains a breakdown of federal funding by industry (using the NAICS categorization) as well as by state.
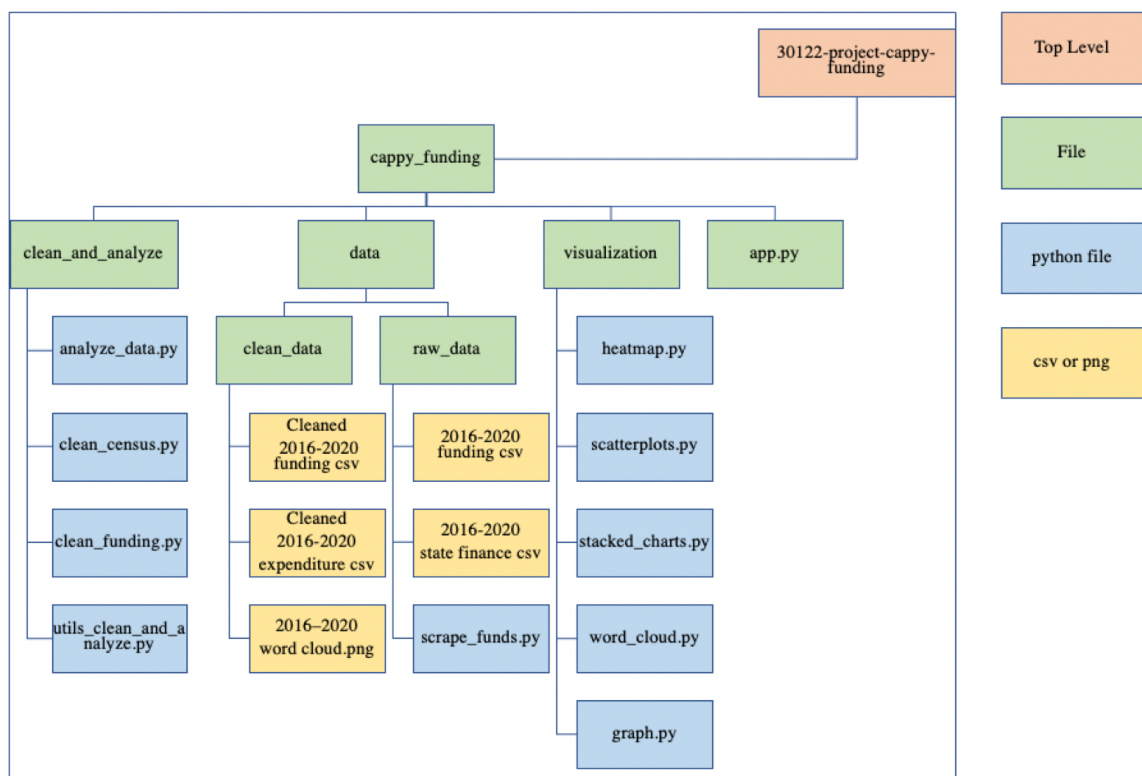
Our project performs data visualization and analysis for federal funding expenditure from 2016-2020 in the following five ways:
1. Interactive Funding Heat Map of the US: illustrates how fundings are distributed geographically to different states for each NAICS sector.
2. Interactive Scatter Plot on Expenditure per Capita of every state: illustrates how state expenditure per capita is correlated with population size.
3. Time-series Stacked Charts: illustrates how federal funding for each sector changes over time.
4. Interactive Scatter Plot for Expenditure per Capita and Funding per Capita: illustrates how state expenditure per capita is correlated with funding per capita in a given year and how the relationship changes over time.
5. Top 10 Categories Word Cloud: illustrates the top ten NAICS industries that have received the greatest amount of federal funds over the five years.

Part 3: Project Structure

Data Layer

USA Spending

USA Census

API

csv download

Data Cleaning and Calculation

Dashboard Visualization

Scatterplot for Fund Expenditure

Stacked Area Chart for Funding

Stacked Bar Chart

Word Cloud of NAICS Category

Expenditure vs Funding Scatterplot

Heat Map of USA

Top-level directory structure

30122-project-cappy-funding

cappy_funding

clean_and_analyze

analyze_data.py

clean_census.py

clean_funding.py

utils_clean_and_analyze.py

data

clean_data

Cleaned 2016-2020 funding csv

Cleaned 2016-2020 expenditure csv

2016–2020 word cloud.png

raw_data

2016-2020 funding csv

2016-2020 state finance csv

scrape_funds.py

visualization

heatmap.py

scatterplots.py

stacked_charts.py

word_cloud.py

graph.py

app.py

Top Level

File

python file

csv or png

Part 4: Code Responsibility

Modules	Tasks	Name
Data Scraping	Used API to scrape data.	Yujie Jiang
Data Cleaning and Environment construction	(1) Cleaned and manipulated data to make it usable for data visualization. (2) Made the application workable in virtual environment.	Bryan Foo
Data Visualization -1	(1) Interactive Scatter Graph (2) Time-series Stacked Charts (3) Interactive Time-series Graph (4) Word Cloud	Ziyang Chen
Data Visualization -2	Interactive Heat Map	Gongzi Chen

Part 5: Interacting with the Software

1. Interacting with the application

To interact with the application, clone the git repository and run the command line (explained in the README.md file). Our application allows the user to do one of the following four actions:

1. Open the visualization dashboard: Selecting this option starts the process of data visualization using the pre-stored datasets. In the process, two static data visualizations (time-series stacked chart and word cloud graphics) are saved into the visualization/ directory. Following which the user is provided with a URL to use to access the data visualization dashboard on a browser of choice.
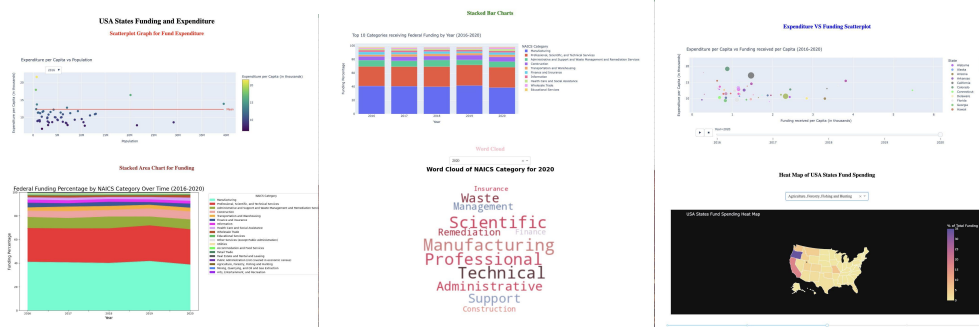
2. Run API and download files: Selecting this option starts the back-end process of API scraping from the USA Spending database. The user is provided with the option of selecting a range of years to download the raw files.

3. Data cleaning and analysis: Selecting this option starts the back-end process of using data already scraped from the API and cleans and processes the data sets. The user is provided with the option of selecting a range of years of raw datasets to clean. For the data visualization to work properly, the user should use the system default starting year (2016) and ending year (2020) by pressing

<u>the Enter button when prompted.</u> Five main datasets are cleaned in the process for each year selected: (a) federal funding by state, (b) federal funding by category by state (percentages), (c) federal funding by category by state (absolute values), (d) state expenditure per capita, (e) federal funding per capita. In addition, two more datasets from the USA census are cleaned: (a) 2020 population census, and (b) 2018-2020 poverty rate by state.

4. Exit application: Selecting this option allows the user to exit the poetry virtual environment.

2. Interacting with API

To interact with the API, run the command line '**poetry run python3 -m cappy_funding**' at the top-level directory and select option 2 for API interaction. It will ask for the input of a year range: a start year and an end year. This program will run a query on the USASpending API of the entered range. Note that the default start year is 2016 and the default end year is 2017.

The output will be stored in csv-files at "/cappy_funding/data/raw_data". For each year, there will be a csv file containing the fundings of every category in every state. The average run time for one year is around 10 minutes. As suggested, a shorter year range query is recommended for API testing.

Part 6: Goals and Accomplishments of the Project

The project successfully delivered visualization on the distribution of federal fundings in the states of every category of industries (defined by NAICS) and also visualized information on federal expenditure incorporated with state population data.

Insights:
<u>US Federal Funding Allocations Across NAICS Categories in (2016-2020) Analysis:</u>
Based on the percentage of federal funding allocated to different NAICS categories from 2016 to 2020 in the USA, it shown that the Manufacturing sector received the highest percentage of federal funding, ranging from 38.4% in 2016 to 38.7% in 2020. This is followed by Professional, Scientific, and Technical Services, which received around 29% of federal funding consistently over the years.The Construction and Administrative and Support and Waste Management and Remediation Services sectors also received a significant amount of federal funding, ranging from 5.6% to 6.9% and 7.5% to 10%, respectively.

On the other hand, the Agriculture, Forestry, Fishing, and Hunting sector received a relatively low percentage of federal funding, ranging from 0.09% to 0.16% over the years. Similarly, the Arts, Entertainment, and Recreation sector and the Public Administration (not covered in economic census) sector received the lowest percentage of federal funding.

It is interesting to note that the Utilities sector received a high percentage of federal funding in 2016 and 2017, but the percentage decreased in the following years. In contrast, the Finance and Insurance sector received a relatively low percentage of federal funding in 2017 and 2019, but the percentage increased in 2018 and 2020.

Overall, the data suggests that federal funding allocation in the USA is concentrated in a few sectors, with Manufacturing and Professional, Scientific, and Technical Services receiving the most funding.

Funding Received per Capita Analysis:
The funding received per capita for each state in the United States is calculated for the years 2016 to 2020. The average funding received per capita for the United States for the years 2016 to 2020 is $1.01 thousand. The state with the highest average funding received per capita is Maryland with $5.57 thousand, while the state with the lowest average funding received per capita is South Dakota with $0.50 thousand.

Expenditure per Capita Analysis:
The expenditure per capita for each state in the United States is calculated for the years 2016 to 2020. The average expenditure per capita for the United States for the years 2016 to 2020 is $10.90 thousand. The state with the highest average expenditure per capita is Alaska with $20.83 thousand, while the state with the lowest average expenditure per capita is Idaho with $8.67 thousand.

Expenditure per Capita vs. Funding Received per Capita Analysis:
The expenditure per capita and funding received per capita for each state in the United States are analyzed for the years 2016 to 2020. The correlation coefficient between the expenditure per capita and funding received per capita is 0.12, indicating a weak positive correlation.

Obstacles and Reflections:
We had to reassess objectives as some of the results achieved did not meet our initial expectations.
   (1) We initially wanted to provide insights for funding seekers by analyzing specific funding recipients in every state. However, we faced limitations in accessing the recipient information database, making it impossible to use API or web scraping methods, so we had to adjust the idea to focus on analyzing funding and expediture per state and categories.
   (2) One of our initial goal was to find a correlation between federal expenditure and federal funding, but the correlation scatter plot showed a weak relationship between the two variables. Similarly, we expected to observe changes in the top 10 categories that receive the most funding over time, but there were no significant changes from 2016 to 2020.