

All Cities Are Budgets:

1. Group members:

- Monica Nimmagadda | CNETID: monican
- Lisette Solis | CNETID: lisettesolis
- Josemaria Macedo | CNETID: josemaria
- JP Martinez | CNETID: jpmartinez

2. Project Abstract:

The project focuses on understanding the protest regarding the Blacks Lives Matter movement after the killing of George Floyd in May 2020 and their impact in relationship to media coverage and changes to municipal budgets using a sample from January 2017 to January 2023.

Using data on the number of protests from the Crowd Counting Consortium, news articles from The Guardian and The New York Times APIs, and police budgets collected for the top 8 cities with the highest police spending per capita,¹ we examined the relationship between number of protests, the amount of media coverage, and changes to budgets. We also performed some initial sentiment analysis and analysis of word correlations on the lead paragraphs and headlines of the news articles.

Our results at this time are inconclusive. We did not see strong relationships between these datasets, but this is likely due in part to poor data quality and the specific nature of our text. More discussion of this can be found in the *Data and Methods* section of the dashboard.

3. Overall structure of the software:

The main package of the project is “*project-protests*” which contains the following subpackages that all live in the same level underneath *project-protests*:

- *newspapers*: Collect news articles from The Guardian and The New York Times using the API’s; clean the files and compile them.
- *sentiment_analysis*: Perform sentiment analysis and pairwise correlation tasks on newspapers data.
- *html*: create dashboard application for final output.
- *police_budget*: Clean data of police budget.
- *visualizations*: Create visualizations to be used in dashboard.
- *protest*: Clean data obtained from the Crowd Counting Consortium.

The *project-protest* package calls the *html* package to run the application and if specified in the inputs, it calls the *newspaper* subpackage to collect news articles, clean the data and compile the information of The New York Times and The Guardian in a csv file.

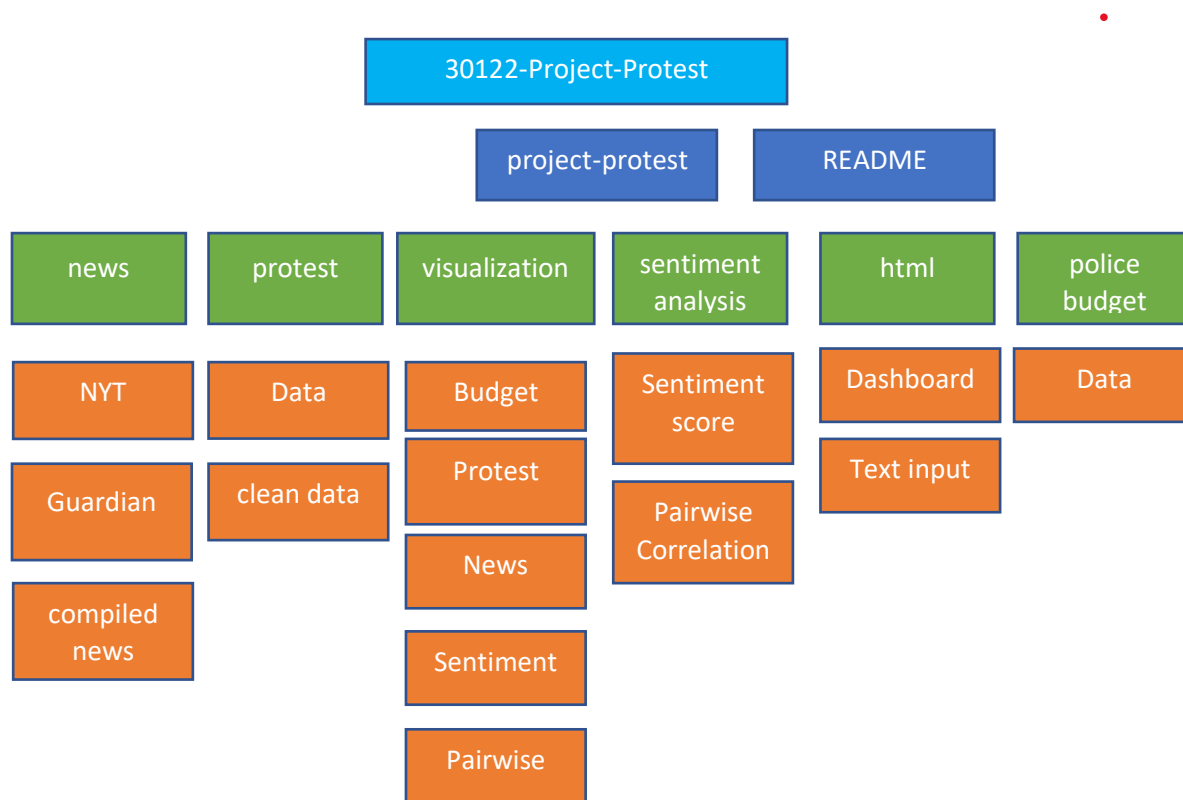
¹ <https://www.statista.com/chart/10593/how-much-do-us-cities-spend-on-policing/>

The *html* package calls the *visualizations* package to import the different graphs used in the application output.

The *visualizations* subpackage calls the *newspapers*, *sentiment_analysis*, *police_budget* and *protest* to collect the data used for the different visualizations.

The *sentiment_analysis* subpackage calls the *newspaper* subpackage to use the newspaper data to perform the analysis.

The *newspapers*, *police_budget* and *protests* subpackages are independent subpackages that don't utilize functions from other subpackages.



4. Description of code responsibilities:

Responsibilities by package:

- project_protest:
 - __main__: Josemaria Macedo
- newspaper:
 - The Guardian API requests and Data Cleaning: JP Martinez
 - The New York Times API requests and Data Cleaning: Josemaría Macedo
 - Compile news from newspapers API's: JP Martinez and Josemaría Macedo
- visualizations:
 - Pairwise correlation: Monica Nimmagadda

- Sentiment analysis visualization: Lisette and JP Martinez
- Other visualizations: Lisette Solis
- html:
 - Dashboard application: JP Martinez and Monica Nimmagadda
- police_budget and population per city
 - Process file: Monica Nimmagadda
 - Population projection for FY23: Monica Nimmagadda
 - Cleaning: Monica Nimmagadda and Lisette Solis
- protest
 - Process protest files: Monica Nimmagadda
 - Cleaning: Monica Nimmagadda
- sentiment_analysis:
 - Sentiment analysis: JP Martinez
 - Pairwise correlation: Monica Nimmagadda

Every file in the repository has a description in the header of the authors involved.

5. Short guide on how to interact with the application and what it produces:

The application is an html output that includes two tabs “Home” where the description of the project and the visualizations are included and “Data and methods” that described the data utilized, and methods used for collection and analysis.

Every visual can be filtered using the legend and includes a tooltip with the values of the point when you hover over it.

These visualizations included on the “Home” allow for interaction with the visual:

- i. Protest count: This graph has a dropdown where the user can select from a list of 8 cities and the nationwide option to visualize the number of protests in a given year. The default is to include all cities and the nationwide count.
- ii. Sentiment analysis: This graph has a dropdown where the user can select the section of the article in where the sentiment analysis was performed and can choose between “lead paragraph” (default value) and “headline”.
- iii. Budget: User can filter by the city.

6. What the project tried to accomplish and what it actually accomplished

The project tried to analyze the impact that the Black Lives Matter protests had in news coverage about the movement and municipal police budgets. We hypothesized that news coverage would follow the protest trend and that municipal police budgets might decrease as the major policy change called for was to defund the police.

Protest

The protests data comes from the Crowd Counting Consortium, an organization that started publishing protest data after the 2017 Women's march. This data is collected from newspaper, social media events, and self-reported data. However, once we started to analyze and visualize this data we realized that there was an unexpected upward trend in the data. Researching the organization more we learned that their staff had grown during this period, and therefore this upward trend was likely due more to internal improvements in data collection capacity rather than BLM protests. This made it difficult to analyze this dataset alongside our other datasets.

News Coverage

We tried to get news coverage from different news sources (liberal and conservative) so the bias of news coverage for the movement was balanced. However, most of the newspapers we wanted to use did not have available APIs. Therefore, we only use data from The New York Times (NYT) and The Guardian API. We focused our analysis on the first paragraph of the story because this paragraph usually includes the key information for news stories, and was readily available in both APIs.

Sentiment Analysis (Sentiment Score & Pairwise Correlation)

The sentiment analysis was limited by a couple of issues. First, we had to limit our API requests due to restrictions to collect only news articles pertaining to police brutality, Black Lives Matter, and other terms directly related to the movement. When doing the pairwise word correlation with "police", our terms were all related to the BLM movement and presented a skewed perception to how newspapers write about police in general. Looking forward, we should train the model with a more neutral dataset and from a variety of news sources ranging from conservative to liberal to understand the change in perception of police over these years as it relates to the BLM movement.

Second, sentiment analysis on articles related to police brutality, murder, injustice and so on is complicated. For example, describing the death of George Floyd can be described as death, murder, or killing. Describing the death as a murder is more aligned with the language BLM wants to use, but is calculated with a more negative score. Police brutality in general uses strong, negative language, which given the topic is necessary, but makes our analysis difficult to fit into a general sentiment analysis package. In the future, training our own data to determine which stories are associated with support and opposition to the BLM movement is necessary, but could not be included due to capacity and time constraints.

Police Budget

Municipal budgets were not as easy to find as we expected. Each city has their own process to find the data. Some were within PDFs and others were accessible through an open data portal. Because of the individual nature of how this data was stored, we had to manually search and enter the numbers into an Excel. This forced us to limit our budget analysis to eight cities with the highest police budget. To be able to compare the cities' budget, we manually found population data from the Census and projected the population change between 2020 and 2021 to 2022 and 2023. This allowed us to calculate budget per capita, but it's more of an estimate for the last two years. It could have been interesting to analyze a larger more diverse set of cities to see if there were still no meaningful changes to municipal budgets. Perhaps cities who do not already spend such a large portion of their budget on police forces were more willing to divert resources away from police departments.

