Sen. Ed Markey: Budget Earmarks
Team B

MEET *the* TEAM

Labeeb Alam
Team Lead
Junior
lalam@bu.edu

Duarte Albuquerque
Team Member
Junior
duarte21@bu.edu

Prathmesh Sonawane
Team Member
Sophmore
prsona@bu.edu

Ethan Singer
Team Member
Senior
singerep@bu.edu

Jerry Ying
Team Member
Sophmore
jhying@bu.edu

Introduction

Our project analyzes congressionally directed funding (CDF), otherwise known as earmarks, within the framework of Senator Edward J. Markey's office. This lens was inspired by a commitment to fostering fair allocation of resources, a commitment that we share with Senator Markey, ensuring that marginalized communities are not left behind in the federal budgeting process.

Congressional earmarking offers lawmakers the ability to direct discretionary funds to specific projects within their constituencies. With the reinstatement of earmarks in Fiscal Year 2022 after a decade-long hiatus, it became imperative to scrutinize this process to ensure equitable distribution and outcomes.

Under the guidance of Senator Edward Markey and his office, our project aims to dissect the journey of earmark requests from filing to fruition, examining who benefits, who is excluded, and the underlying dynamics of the allocation process.

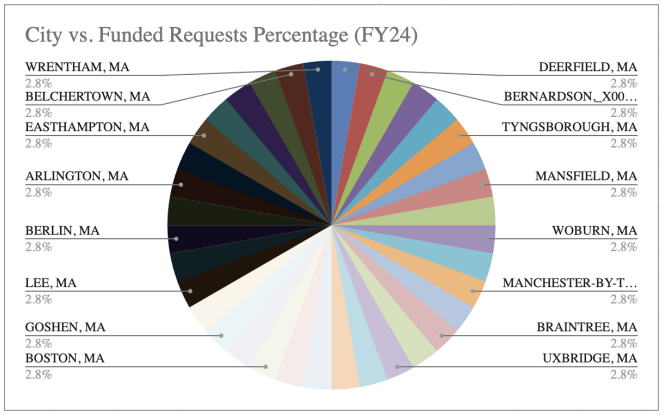
Through a retrospective analysis of CDF allocations, our project seeks to develop a repeatable (scripted) process documented in a comprehensive report, enabling application to future budget cycles. We present our key findings and various visualizations in this document, and in an accompanying presentation, to reveal the geographic distribution and demographic composition of funded projects.

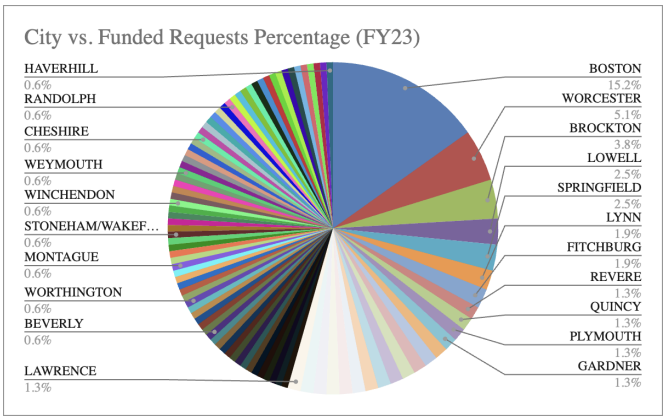
We also built an interactive platform allowing users to explore the data filtered by various other variables of interest, including geography, income level, education, and race/ethnicity. By analyzing trends in applications, sponsored requests, and allowed grants, we aim to uncover patterns, disparities, and potential areas for intervention.

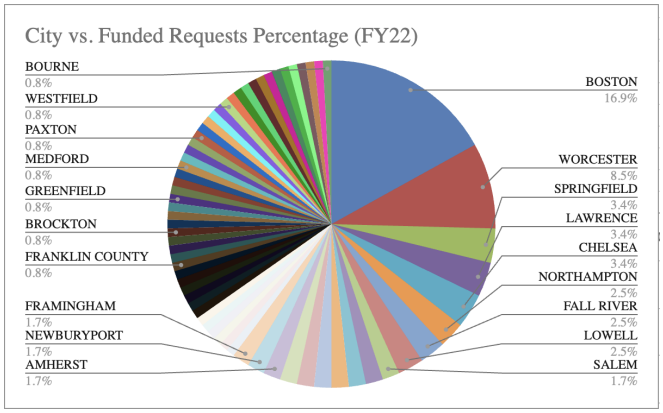
Base Analysis

When starting the project, we had two initial questions to analyze in regards to the earmark data we were given. Firstly, what demographics were positively impacted by the funded projects? Secondly, what demographics were underrepresented by the earmark funding? When starting our project, all of our earmark data was initially provided in PDF format, so our initial task was to convert 2022 - 2024 earmark data into a tabular format. Using tabula, we extracted the information from the pdfs to start initial data analysis. During our first look at the data, we tried to figure out a way to geocode projects to particular census tracts. Later on, this would allow us to pull census variables for each project location to actually determine demographics most and least impacted by them. However, we noticed that the information from the Project Recipient/Project Name columns did not always have clear locations to identify census tracts. We considered two solutions: We could either manually go through the data and map each project to census tracts using Google Earth or automate the process with a python script, searching for location information (like city names) with projectName/Recipient data. In addition, during our first meeting with Ed Markey's office, we were told that for the data to be useful for the stakeholders, we needed to focus on demographics at a city/municipality level. Thus, our group decided to use python scripts to scrape city names from each record’s data, mapping each project to a given city.

Our first analysis involved looking at the allocation of funding for each city encountered in our earmark dataset. The analysis for 2022-2024 is shown below.



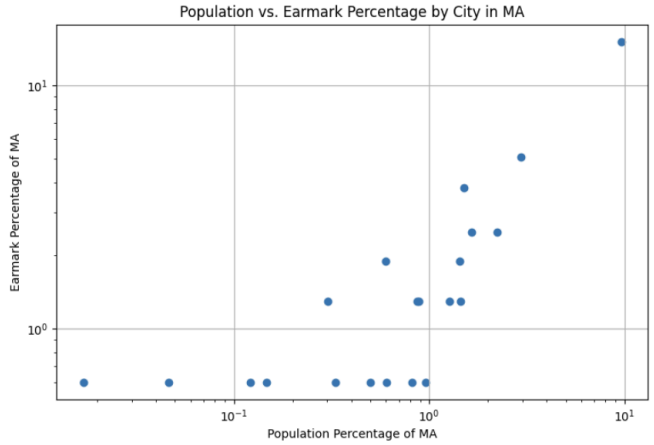




2022 - 2024 Breakdown of Funding Allocations per Municipality

These three graphs show us municipalities in Massachusetts and the percentage of funding they received in our datasets that we were given. For FY22, and FY23 the largest value of allocated funding was for Boston, which makes sense since it is the most populated area in the state. It is seven times larger (in terms of population) compared to any other city, and as such, it would make sense for that area to receive a higher amount of funding.

To better understand the relationship between city size and funding allocation, we hoped to compare a city's population with the funding they received. In doing so, our hope was to reveal any biases related to cities receiving more/less funding than expected. Below, using 2024 data and the most populous cities in MA, we graphed their population percentage and earmark allocations.



General Trend:

Higher Population Percentage

↓

Higher Earmark Earning Percentage

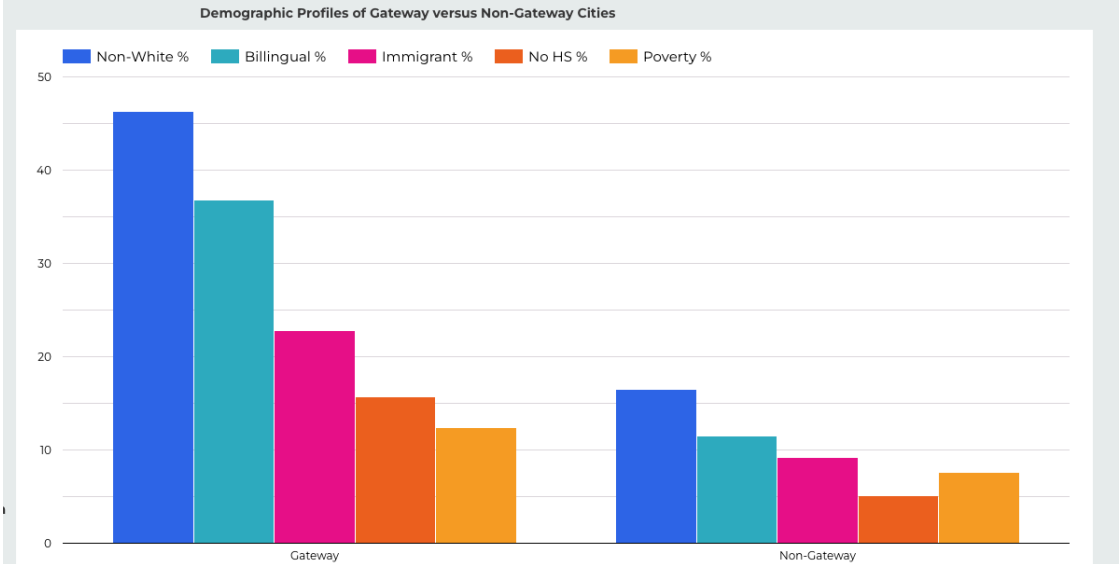
The graph above highlights a positive correlation between city population and earmark funding, which we would have expected. In this case, in terms of population size, more funding is going towards cities with a higher population, meaning no large inequitable bias is apparent when strictly looking at population size.

After the initial insights presentation (Deliverable #1), the team decided to focus on what demographic data we could look into to create a thoughtful analysis. At this time, Ed Markey's office had clarified that the final report was very open-ended and could be determined by the teams. To decide on variables to choose in our research, we looked into the different variables

provided by the Census to get a better understanding on what demographics we wanted to look into. We decided to focus on race, age, urban/rural, income, poverty rate, $/person, and $/person/squareArea. Additionally, we wanted to identify what projects were environment or education related since Ed Markey had a big focus on those topics. If we had extra time, we would look into bilingual (spanish v english only) or foreign born data as well. By comparing the funding of these different groups, we hoped to discover communities that were being over/under represented in our dataset. Using this information, our hope is to create new funding allocations to communities that need them, resulting in more equity.

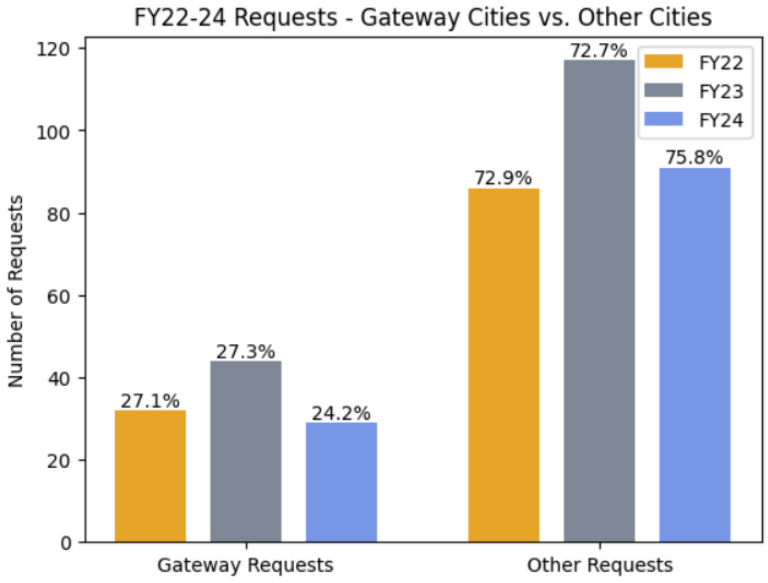
Deeper/Extension Analysis

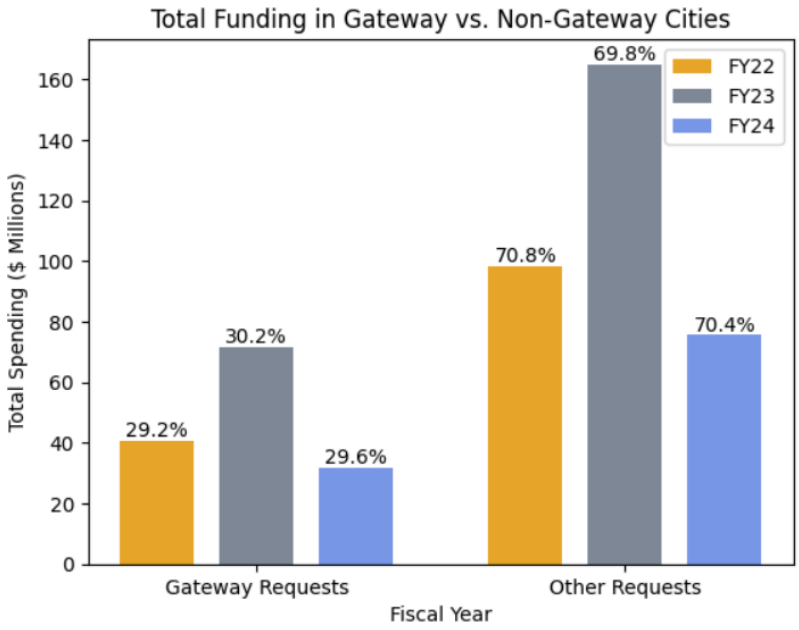
For our deeper analysis of federal earmark funding, our team not only looked into census variables but also spending on gateway cities vs other Massachusetts municipality-approved projects. According to the Massachusetts state government, a gateway municipality is defined as a city with a population greater than 35,000 and less than 250,000, where the median household income and rate of educational attainment of higher education is below the state average. These cities are important to Massachusetts residents that may face social and economic challenges because they offer residents opportunities to work in a variety of industries, providing them a *gateway* toward the “American Dream”. These cities are also demographically different from non-gateway cities.

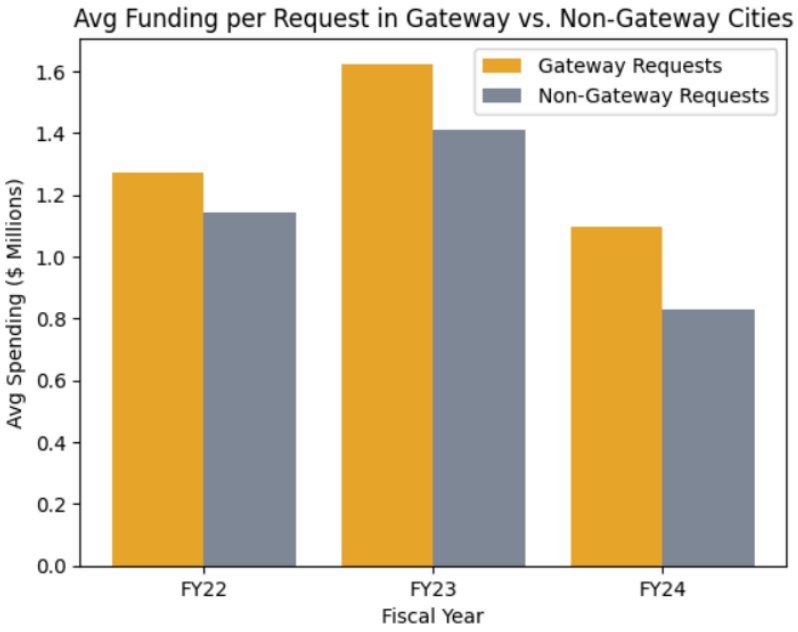


In general, gateway cities are characterized by a higher share of non-white residents, bilingual residents, immigrants, non-high school graduates, and residents below the poverty line.

We hypothesized that the current and future funding in gateway cities should actively promote the cities' ability to promote economic growth for its citizens For our initial comparison with Gateway cities vs other municipalities, we looked into the # of projects, funding, and trends for each, shown by the figures below.







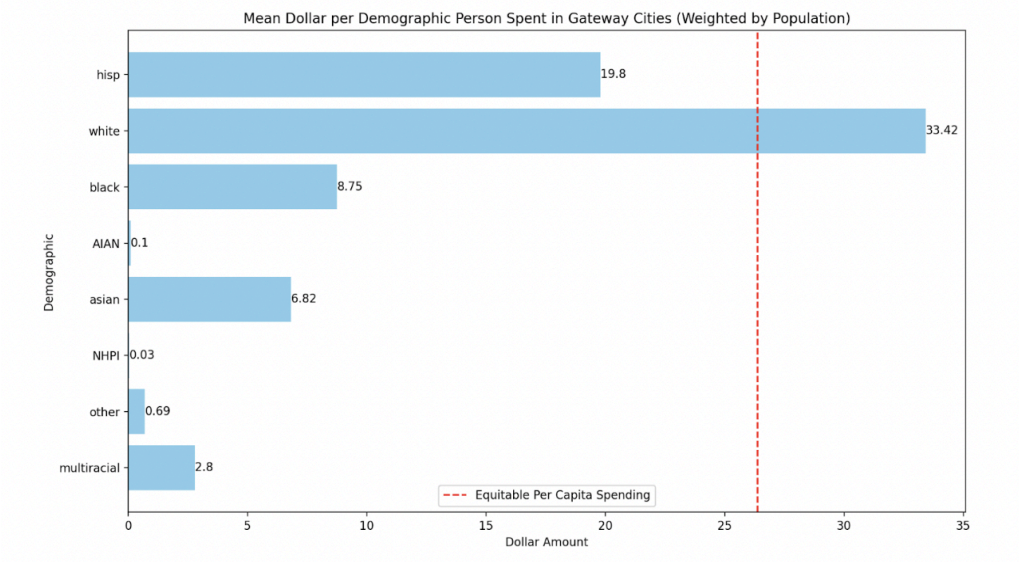
After conducting this analysis, we found that funding requests generally heavily favored gateway cities as expected and that FY23 especially saw a rise in overall funding requests. Additionally, we found that the average funding per request for gateway cities has become proportionally larger than for non-gateway cities over the three years that earmarks have been reintroduced for. This all points to the fact that gateway cities are heavily prioritized when it

comes to congressionally directed spending requests, which makes sense in the context of what a gateway city represents. Gateway cities are generally places of huge unrealized opportunities, so it only makes sense that efforts should be made to fund and support the population of those cities.

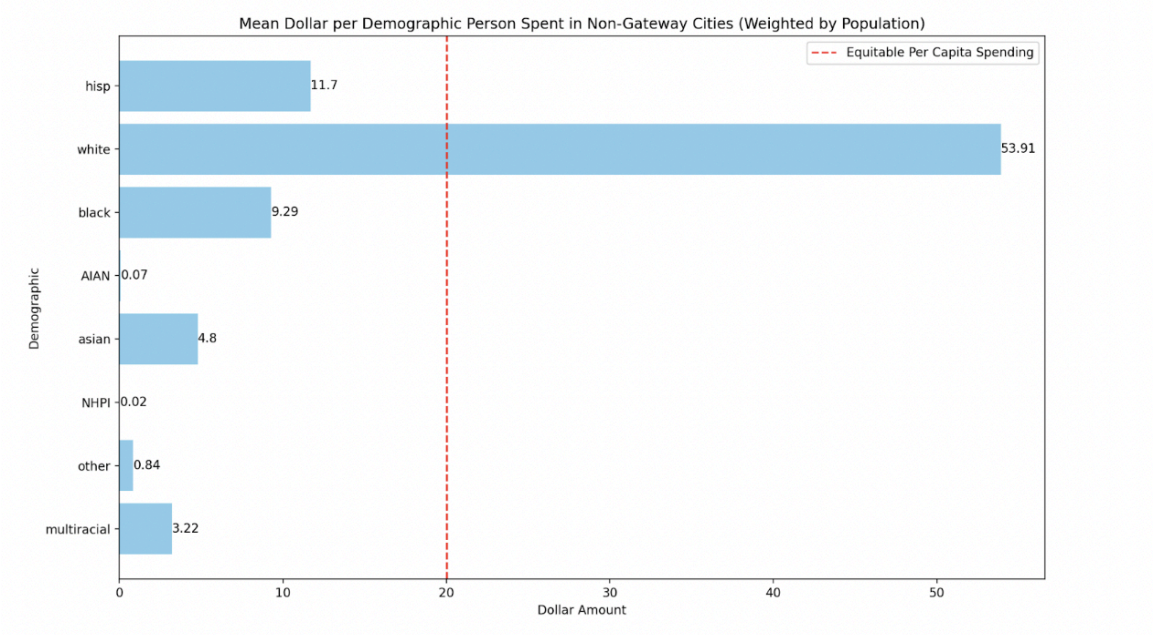
For the next steps of our project, we decided to utilize census variables/data to look at the demographics of both gateway/non-gateway cities to see if the overall funding supported certain communities or neglected others.

Firstly, we thought it would be interesting to look at how each demographic was being impacted by the earmarks spending. Thus, we decided to focus on the mean amount spent per capita per demographic. In layman's terms, this analysis allowed us to understand how much the average person would receive from the earmarks spent based on their race and where they live. To calculate such numbers, we first determine the total spending for each demographic group in each group of cities (gateway vs non-gateway cities) by multiplying the total spending in each group of cities by the percentage of that demographic group within it. For instance, if City X has 50% Hispanic people and City Y has 20%, with respective spendings of $100 and $200, the total spending for Hispanics would be $50 + $40, resulting in $90. Then, we divide this total spending for each demographic group by the corresponding total population of that demographic in the group of cities being analyzed to obtain the per capita spending per demographic. Finally, to determine equitable per capita spending, we divide the total spending across all demographic groups in each group of cities and divide by the total population of each group of cities, providing a baseline for fair allocation.

Across all gateway cities, we found the metrics in the plot below.

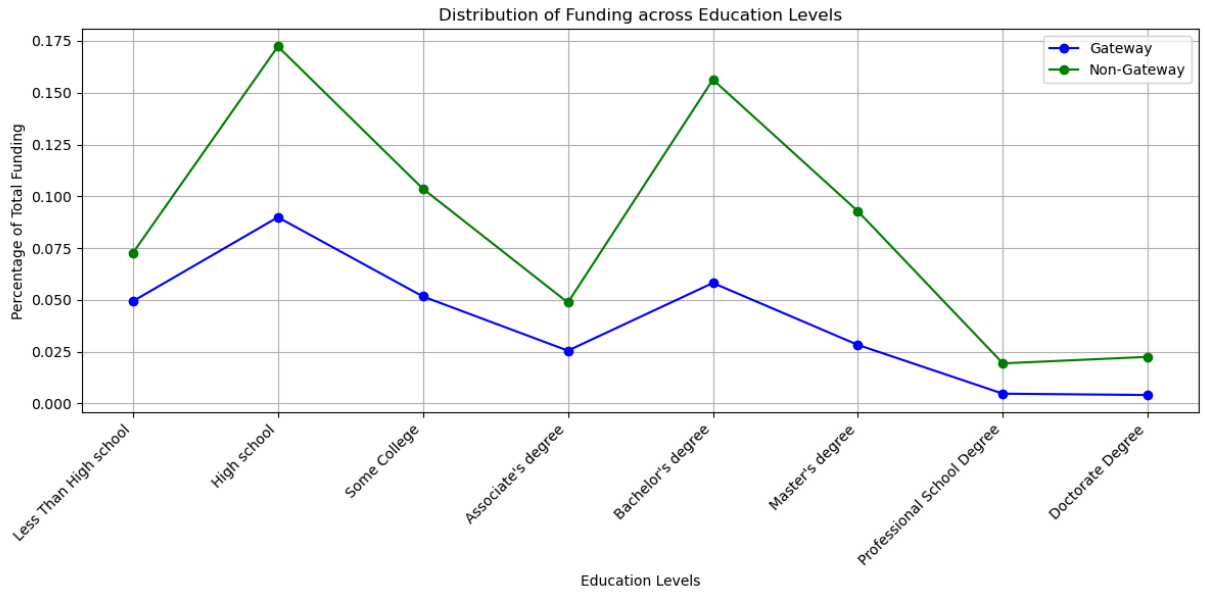


As we can see, we found that most of the funds were being allocated towards the White population ($33.42 per person), versus all other demographics with the second-largest receptor being the average Hispanic person ($19.80 per person). The equitable per capita spending also shows that all demographics, besides the White population, are significantly underrepresented in earmarks. Across all non-gateway cities, we found the following metric shown in the plot below.



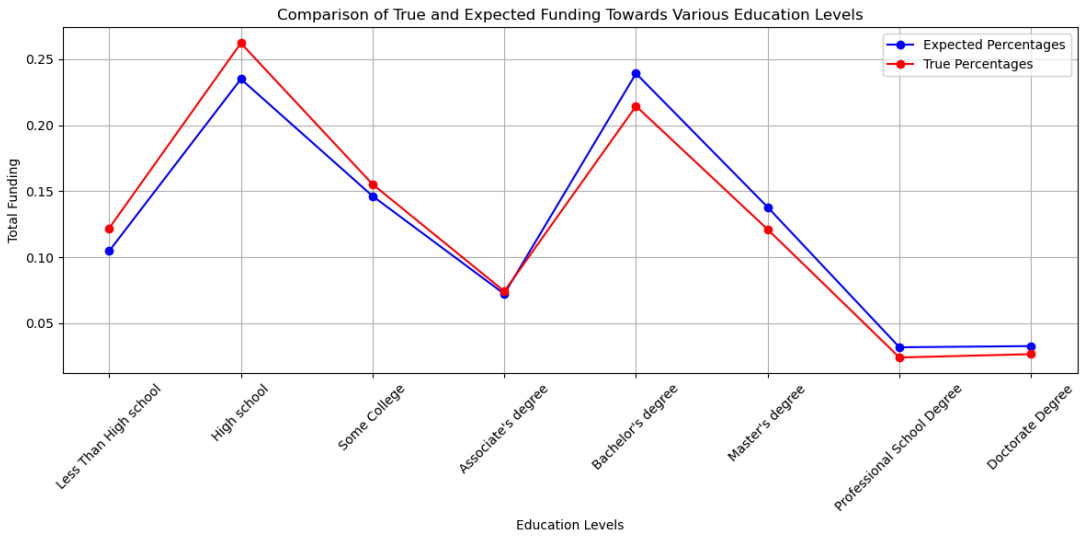
As can be seen in the above plot, we found an even larger disparity in the spending for non-gateway cities. The largest receptors once again are the White population ($53.91 per person), while the second largest receptors once again the Hispanic population ($11.70 per person). As can be seen, by the equitable per capita spending, the non-White demographics are once again significantly underrepresented, although perhaps not as much as in the gateway cities.

Next, we decided to look at the education distribution of both gateway and non-gateway cities to find biases in funding allocations if they existed. Utilizing Censaurus, for each project, we were able to utilize its location, often given by a city name, to find the entire education distribution for that city. In total, for any given funding allocation, our Python census package returned data that could be reduced to a data frame with columns "Less Than High school", "High school", "Some College", "Associate's degree", "Bachelor's degree", "Master's degree", "Professional School degree", "Doctorate Degree", and “Total”. Each column specified the number of people in each city that had achieved and stopped at a certain education, with the “Total” column providing a count for all individuals included in the dataset. Using the “Total” column, our team found the population percentage in each education category. Then, for any given record which contained an “Amount” column, specifying the amount spent in a funding allocation, the amount value was multiplied by each education category to find the amount of money allocated to each educational category for each record. By then dividing our records between the gateway and non-gateway cities, summing up the total funding for each educational category in each group of records, and dividing it by the sum of all funding, we found the below graph.



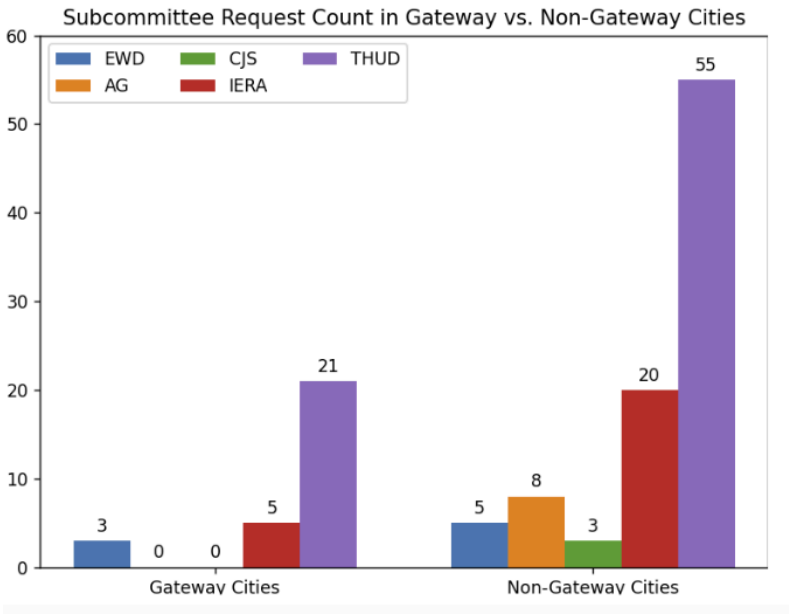
In this figure, we found that across all educational levels, non-gateway cities received higher amounts of funding compared to gateway cities. In addition, for both groups, we found that a higher percentage of funding was allocated to those with just a high school or bachelor's degree.

To get a better perspective on whether this was expected or not, we calculated the percentage of funding expected for each educational category by finding out the education distribution of residents in all the cities in Boston by a similar process as specified above. In addition, we summed the values for each educational category in the above graph to find how much spending was allocated to each educational category no matter the type of city. To compare the expected and true values, we created the below graph.



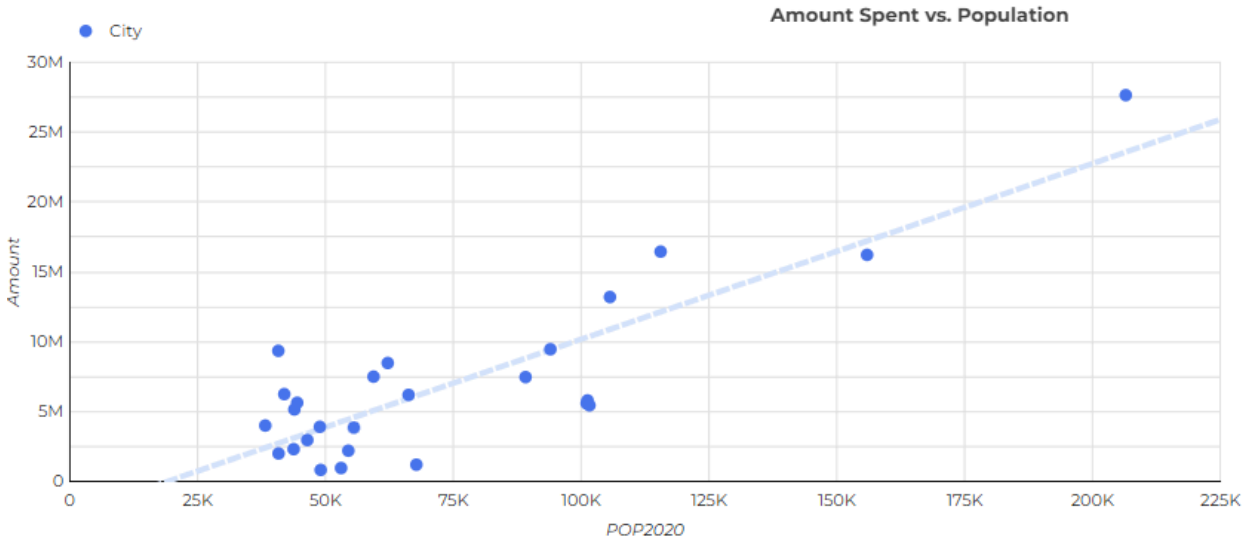
From this graph, we find that the true amount of money allocated to each educational category is very near to the expected amount. Thus, funding allocations are not heavily biased and do not favor/disfavor residents in certain educational categories.

We then continued by separating the requests by each subcommittee to see if specific subcommittees might have a certain bias or trend in their approved requests. By separately graphing each subcommittee's requests, we hoped to find patterns in how each subcommittee receives and approves their requests. We ended up with the following graph for FY24, as well as graphs for FY22/FY23 that I will not show for brevity's sake:



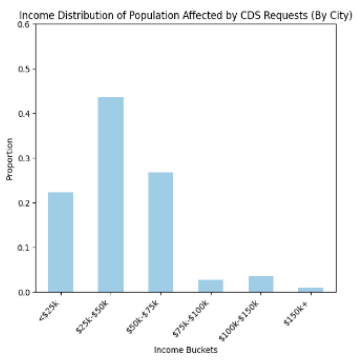
As one can see, most of the subcommittees' requests did not have enough data to be statistically significant, so there was not much analysis we could realistically do with the data. However, what we were able to find within the yearly trends was a steady increase of IERA (Interior, Environment, and Related Agencies) requests which is consistent with Ed Markey's goals of environmental justice. More specifically, we saw a 47% increase in overall IERA requests from FY22 in both FY23 and FY24. Additionally, we found that, especially in gateway cities, there was a rise in THUD (Transportation, Housing and Urban Development) requests which shows that work is being done to properly develop gateway cities.

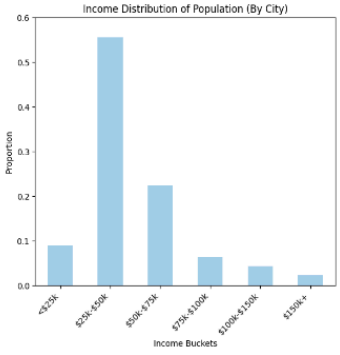
Following this, we wanted to examine the relationship between the population size of a city and the amount of funding it received. By retrieving census data on the population size of each city/town in Massachusetts, we were able to attach a population size variable to each funding request so that we could visualize the trend between spending amount and population size. We created a scatter plot of every gateway city's population size and total funding to find any potential outliers:

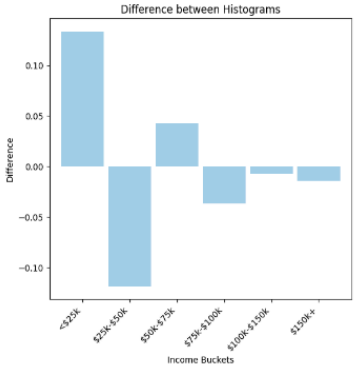


By creating this graph and fitting a linear regression over the points, we found that there were no significant outliers in gateway cities when comparing their population size and total funding amount. The only potential outlier was Chelsea, which received more funding than expected given its population size. However, given that the number of funding requests for each city is well below the threshold desired to be statistically significant, we were content with each individual point as long as it did not wildly differ from the expected value. We additionally conducted this analysis for non-gateway cities and found a very similar linear trend between population and funding amount.

Finally, we wanted to analyze the relationship between income range and spending requests and thus placed each city within a certain income bracket. We hypothesized that lower-income cities should be more heavily targeted, while higher-income cities would likely have relatively lower amounts of funding given that they likely already had the infrastructure needed to support themselves. We ended up coming up with the following figures by analyzing the income distribution by CDS requests and also the income distribution by city/towns:



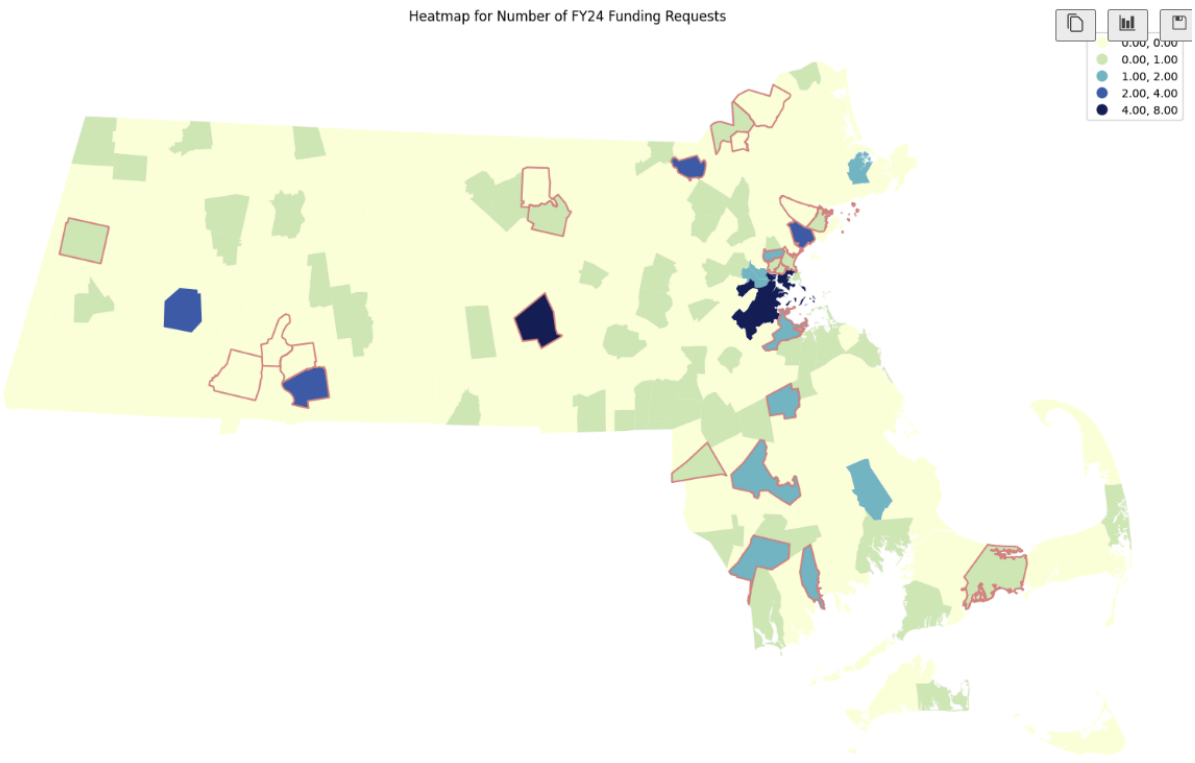


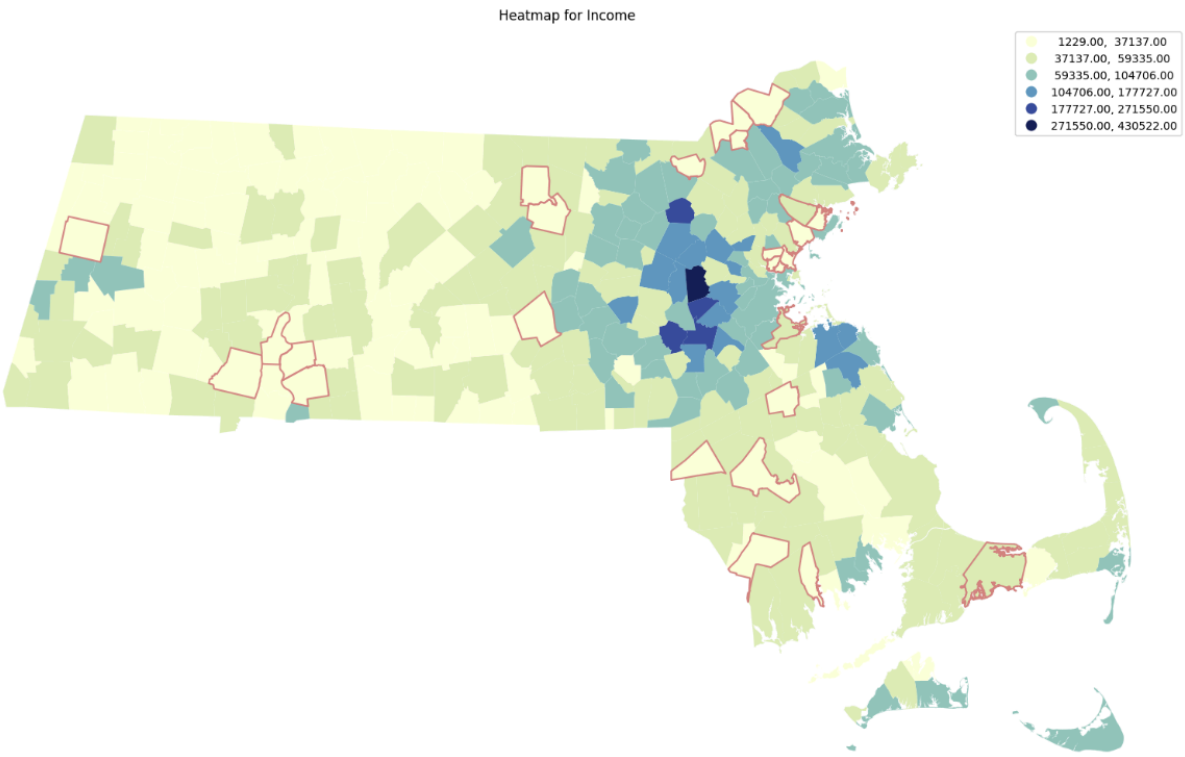


As you can see, our hypothesis held true that lower-income cities were specifically targeted by CDS requests, with a proportionally larger amount of requests going towards low-income cities. However, one thing we did find a bit troubling was that it seemed like the lower-middle class was being underrepresented in spending requests. When calculating the

difference between the two histograms, cities with a median annual income below $25k had a noticeably positive difference while cities with a median annual income between $25k and $50k had a noticeably negative difference. We thought up a few reasons why this may be the case and eventually concluded that the most likely explanation for this difference is the abundance of small towns/cities that are in the $25k-$50k income range. Because smaller towns and cities do not have the support structures that a larger city might have, their access to funding requests is lower. This inevitably leads to many small lower-middle-class cities receiving little to no funding while larger cities with better support structures make and receive more requests. If the bias against the lower-middle class does indeed exist, this means that addressing such a bias equitably will be a challenging issue to deal with.

To visualize these trends more effectively, we wanted to create geographic heatmaps, or choropleths, to better understand the trends that exist in spending requests. By combining our data with census-provided geospatial data, we were able to create the following choropleths:





These confirmed our suspicions that a large reason for the theorized inequality was the abundance of small towns with lower-middle-class income ranges, particularly in the rural parts of Western Massachusetts. Note that because of our chosen specialization in gateway cities, we opted to highlight those in red.

Future Scope

While our project has conducted an impactful analysis of current federal earmark spending, there are several avenues for future research that can further enhance our understanding of the data and inform better decision-making. By pursuing these subsequent ideas, this project will help Senator Edward Markey and his office make better decisions backed by better data, and help underrepresented groups and communities get the funding they need and deserve.

Historically, federal earmarks were only reinstated in 2021, thus, limiting the scope of our data and analysis. As more time goes on, however, it would be interesting to conduct a longitudinal study to track changes in earmark allocations, and demographic profiles of benefiting communities. This could reveal evolving trends and identify areas where equity has improved or deteriorated. This longitudinal analysis could inform policy decisions and interventions aimed at addressing persistent disparities.

Furthermore, our project was limited to an analysis by demographic group, income, education, and type of city (gateway vs non-gateway city). However, in the future, one can go beyond these variables and conduct an intersectional analysis by considering factors such as age, gender, and disability status which could provide a more comprehensive understanding of equity in federal budget earmarking processes. This approach would acknowledge the complex interplay of multiple identities and their impact on resource allocation. Moreover, by supplementing quantitative analysis with qualitative research, such as interviews or focus groups with community members and stakeholders, we could garner contextual insights into the lived experiences of those affected by earmark funding. Qualitative data could offer nuanced perspectives on equity challenges, potential solutions not captured by quantitative data alone, and a broader understanding of the needs of each group.

Finally, the project could benefit from engagement with other cities. Through a partnership with similar cities, in size and demographic landscape (such as San Francisco), we could gain valuable insights into disparities and equity in federal budget earmarking processes across similar regions. This comparative analysis could highlight differences in funding allocation, demographic profiles, and community needs, leading to recommendations for a more equitable distribution of resources.

Individual Contributions

Duarte:

- Early Insights: Duarte was present at all meetings, scripted and converted the PDF data to tabular format, and analyzed the data from the FY22/FY23/FY24 sheets to create plots demonstrating the percentage of funded requests coming from each city in MA. He also helped work on the slides.
- Mid-Semester: Duarte helped brainstorm ideas for the second deliverable, and conducted an analysis to understand the distribution of financial support across demographic groups in gateway and non-gateway cities. He preprocessed data, mapped projects to cities, calculated dollar per demographic person spent, and visualized the results through graphs. Additionally, he scripted this process and created and edited slides.
- Final: Duarte attended all meetings and scripted/generated new data frames to analyze the distribution of financial support among demographic groups in both gateway and non-gateway cities, utilizing the latest decennial data. Leveraging these refined data frames, he created visuals in Looker for both the final presentation, dashboard, and demo. In addition, Duarte created the title page, wrote the introduction and the future scope, and helped write the deeper/extension analysis for the final report. Wrapping up his contributions, Duarte presented during the demo session.

Labeeb:

- Early Insights: Labeeb attended all meetings. Worked on Project Kick-off tasks. Explored methods for more precise location data. Updated trello sprint boards. Prepped some slides for client presentation.

- Mid-Semester: Labeeb attended all meetings. Presented first deliverable to stakeholders. Communicated with project leads on deliverable requirements. Compiled geocoding scripts with other teams to produce a single geocoding file for other teams to use. Discussed research tracts and brainstormed ideas for second deliverable.
- Final: Labeeb attended all meetings. Presented mid-semester deliverable to stakeholders answered any questions and suggestions. Communicated with project leads on weekly tasks. Formatted looker graphs into one dashboard. Wrote findings and created a new layout for Demo-day posters. Setup final report and wrote base analysis for team to complete. Labeeb will present for demo day.

Prathmesh:

- Early Insights: Prathmesh helped complete preliminary questions and familiarized himself with the data at hand. After finding some potential problems, like insufficient location data, Prathmesh amended the Team Deliverable to make such concerns clear. In addition, they have explored utilizing project name/recipient data via Google Earth to find their relevant census tracts.
- Mid-Semester: Prathmesh brainstormed ideas for Deliverable #2, later presenting the group's findings during the client presentation. For the Mid-Semester project, Prathmesh manually geo-coded project titles with null locations, added Lat/Long coordinates to project data, and created education related graphs for the project data.
- Final: Prathmesh was a significant contributor in creating our Looker dashboard and final written report (Base Analysis + Deeper Analysis), and they entirely re-made the GitHub readMe for the team’s project as per the requirements. In addition, Prathmesh developed

and tested more graphs related to funding allocations and their educational distributions to better discover any educational-related biases present. Finally, Prathmesh helped demo the project during Demo-Day.

Ethan:

- Early Insights: Ethan requested and cleaned relevant City- and County Subdivision-level Census data, and began exploration of matching Project and Recipient names to latitude/longitude coordinates using Geocoding APIs.
- Mid-Semester: Ethan wrote scripts to download, process, and clean US Census demographic data for cities and counties in Massachusetts. Data included race, age, education, immigrant status, language, and more. Ethan also worked on graphics and created slides to characterize gateway cities.
- Final: Ethan finalized datasets for multiple Census surveys, including the American Community Survey and Decennial Census. Ethan also assisted in initial set up of the interactive Looker dashboard, and contributed looker graphics based on Census data and gateway status.

Jerry:

- Early Insights: Jerry helped answer preliminary questions and explored methods to categorize earmark funding requests by recipients' income category.
- Mid-Semester: Jerry worked on creating visuals for deliverable #2, matching requests to city and analyzing FY22/23/24 funding trends for gateway/non-gateway cities. For the mid semester report, he analyzed different subcommittees' funding request trends and

began exploring the creation/use of choropleths to help visualize trends. Also worked on slides.

- Final: Jerry attended all the meetings and conducted further analysis of the effects of income and population on funding requests, updating the Looker dashboard with the new graphs and insights. He also identified anomalies in the data and addressed them accordingly. Finally, he helped present the project during Demo-day.