Boston University

Equity and Philanthropic Giving in Massachusetts: Team C

Sharon Zou 2024 szou00@bu.edu
Fiona Mei 2024 fionamei@bu.edu
Esther Choi 2025 eschoi@bu.edu
Tyler Nguyen 2024 t198@bu.edu
Justin Lee 2024 juslee@bu.edu

CS506: Data Science Tools and Applications
Professor Lance Galleti
May 1st 2024

Introduction

This project's client is the New Commonwealth Fund, a philanthropic fund created to reform harmful power structures that have created generational economic and social inequities for Black, Latino, and indigenous people. The goal of the project, in particular, was to utilize the datasets provided to us to determine how much philanthropic funding is going to nonprofits headed by Black and LatinX leaders/racial justice programs in Massachusetts, along with identifying any visible trends in the data throughout that process.

By diving into these questions, this project will seek to confirm ongoing beliefs that past promises made to support the Black/LatinX community may not have fully materialized and/or that much of the funding given to the Black/LatinX community is not led by the community itself.

To answer these questions, the client is providing us with data collected from CANDID, an organization that transforms tax filings from philanthropic organizations and nonprofits into searchable data sets. More specifically, they provided us with 3 datasets:

1. the Base Bundle, which contains information about all nonprofit organizations - all their basic information, revenue, target population, etc.
2. the Grants Bundle, which contains information about all the grants made in recent years, the organization funding them, the organization receiving them, the grant amount, and more.
3. the Demographics Bundle, which contains information about the demographical data for each leader in the organizations.

We began our projects by sifting through these datasets and obtaining a basic understanding of it (what columns there are, which ones are more important than others, what we might be missing, etc). Then, through our base analysis and our extended analysis, we were able to make more solid findings to support the ongoing beliefs that our clients shared with us, and the key questions that they had asked us to answer. These questions are:

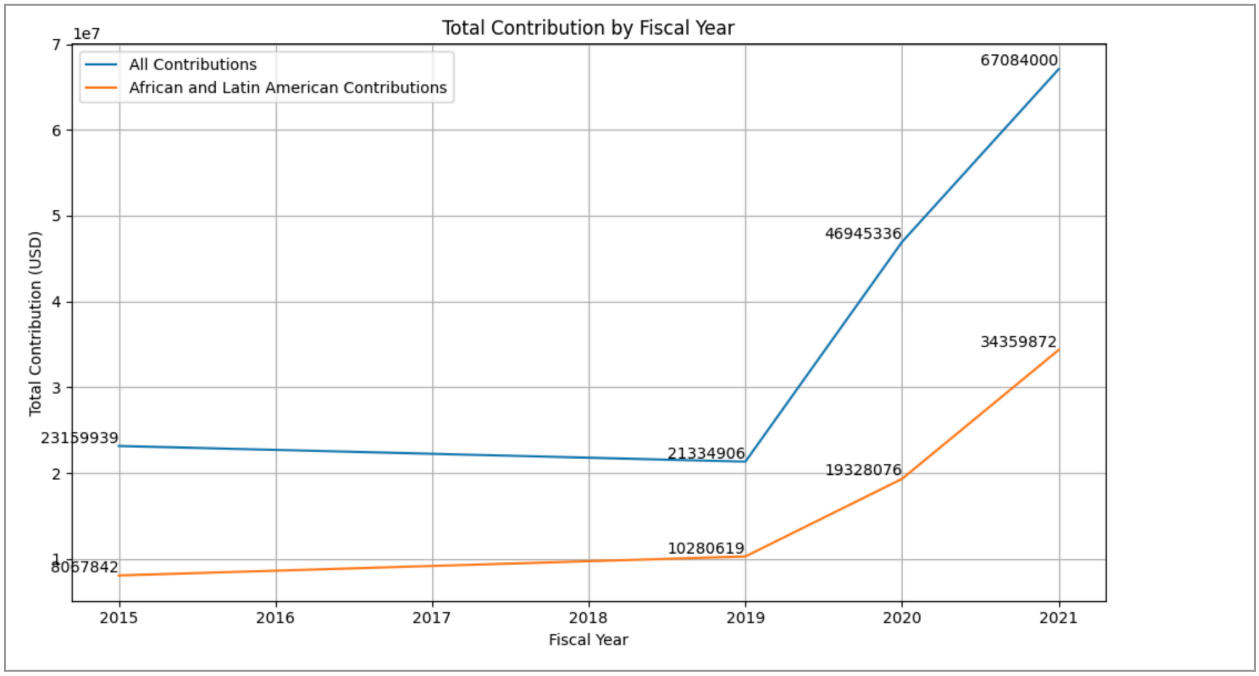
- How much funding is going toward racial justice programs (based on donations to nonprofits that focus on these issues) in Massachusetts and how has this shifted over time?
- How much funding is going toward nonprofits led by Black or LatinX individuals (president, executive director, board chair, etc.) in Massachusetts?

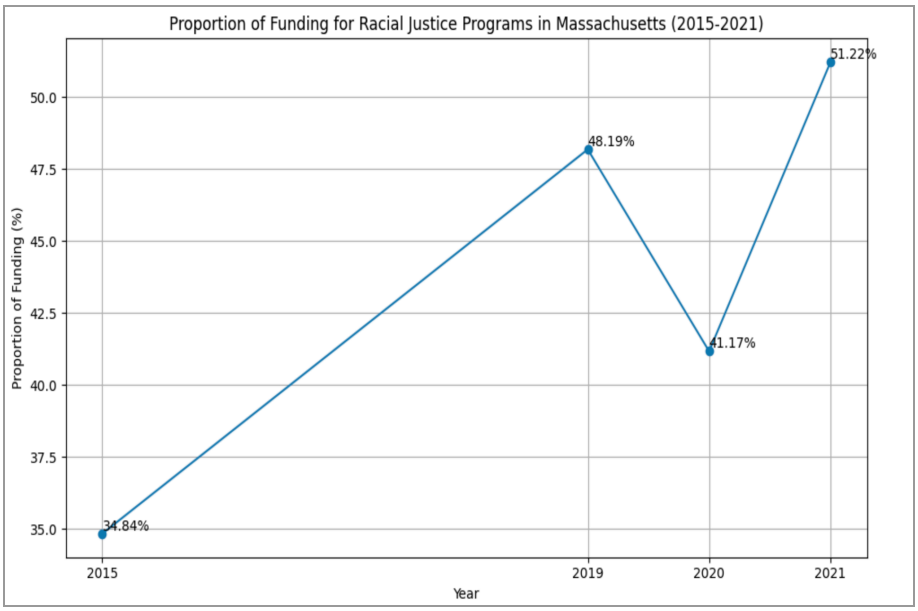
- What are other trends related to the type of organizations and grants that are being funded serving the Black and LatinX communities in Massachusetts e.g. missions of the nonprofits, issue focus of the grants, changes in amounts donated over time, etc?
- Who are the top funders of racial justice programs in Massachusetts?

Executive Summary

Throughout our semester of exploring the datasets and answering the key questions, we were able to create a variety of visualizations that helped us better understand trends within these non-profits and the grants that are being given out. In this executive summary, we will cover a general overview of the most important analyses that we've developed. *To understand them more in detail and/or see their methodology, please utilize the rest of our report and find a deeper explanation in the 'Base Analysis' and 'Extension Analysis' sections.*

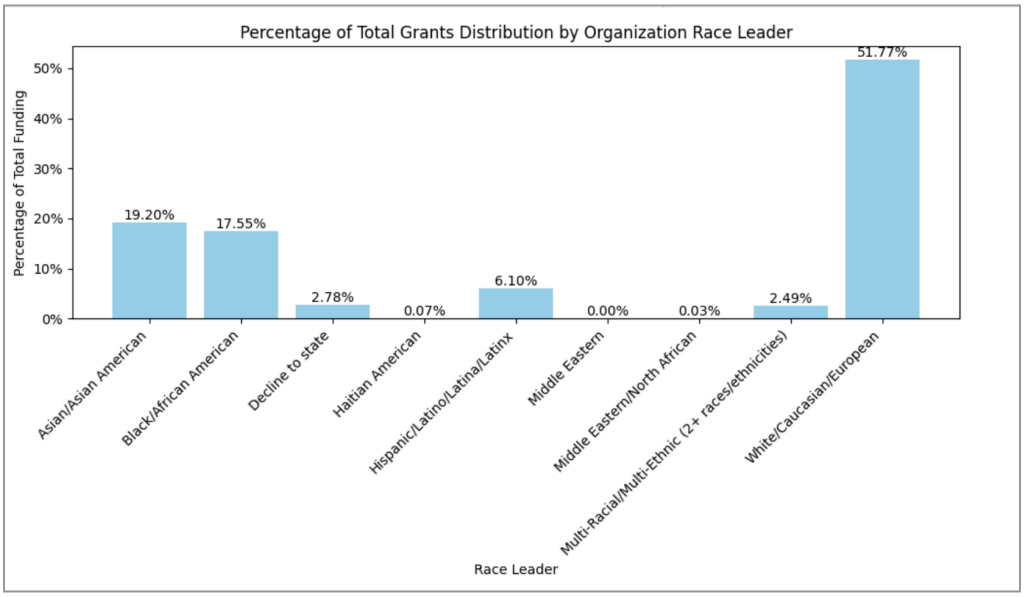
To start off, we were able to quickly understand that funding for racial justice programs has increased alongside general funding for all programs. For context, we defined 'racial justice programs' as any non-profit organization whose target population is the Black/LatinX community. We found this through a quick visualization that we created for Key Question 1:





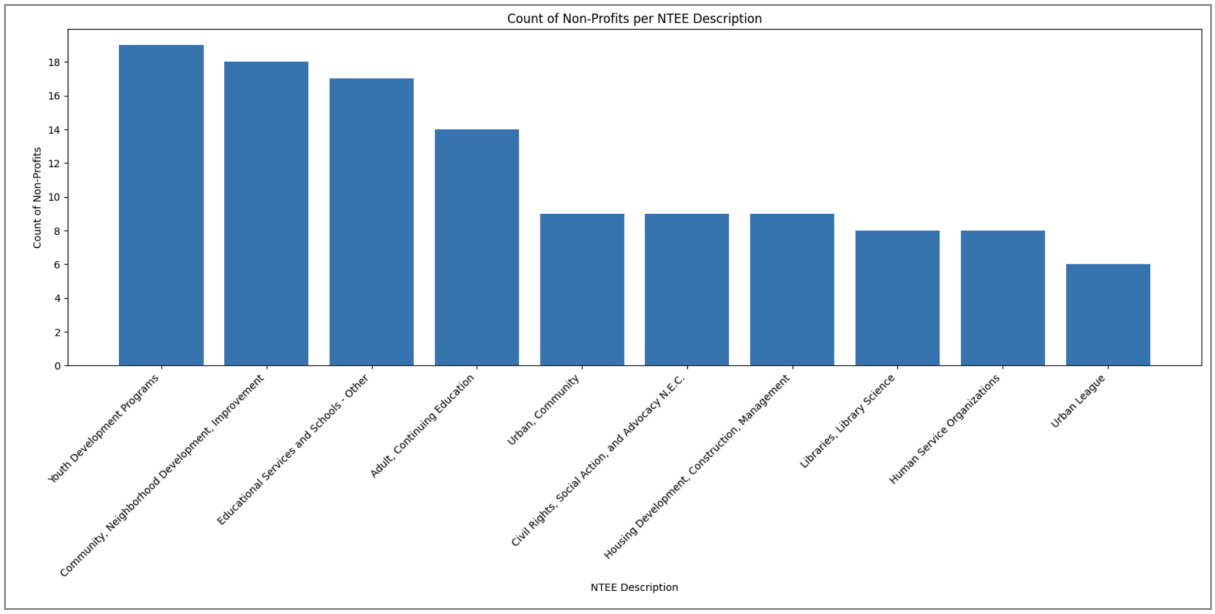
However, the proportion of the total funding that was provided for the Black/LatinX community had a slight dip during the year 2019. As a group, we weren't able to figure out why, but it could be due to a variety of factors, such as current world events at that time. After 2020, it was back up again.

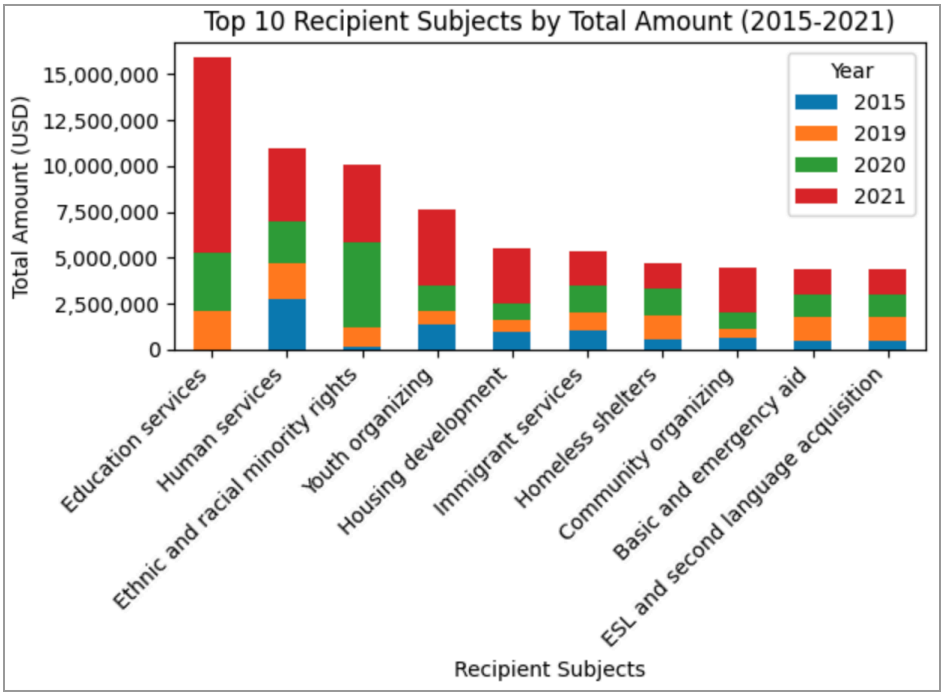
We also confirmed that most grants are going to organizations that are *indeed* not held by the same community that they seek to serve by combining the grants dataset alongside the demographic dataset. The majority of the leaders of these organizations are White, then Asian, and then Black.

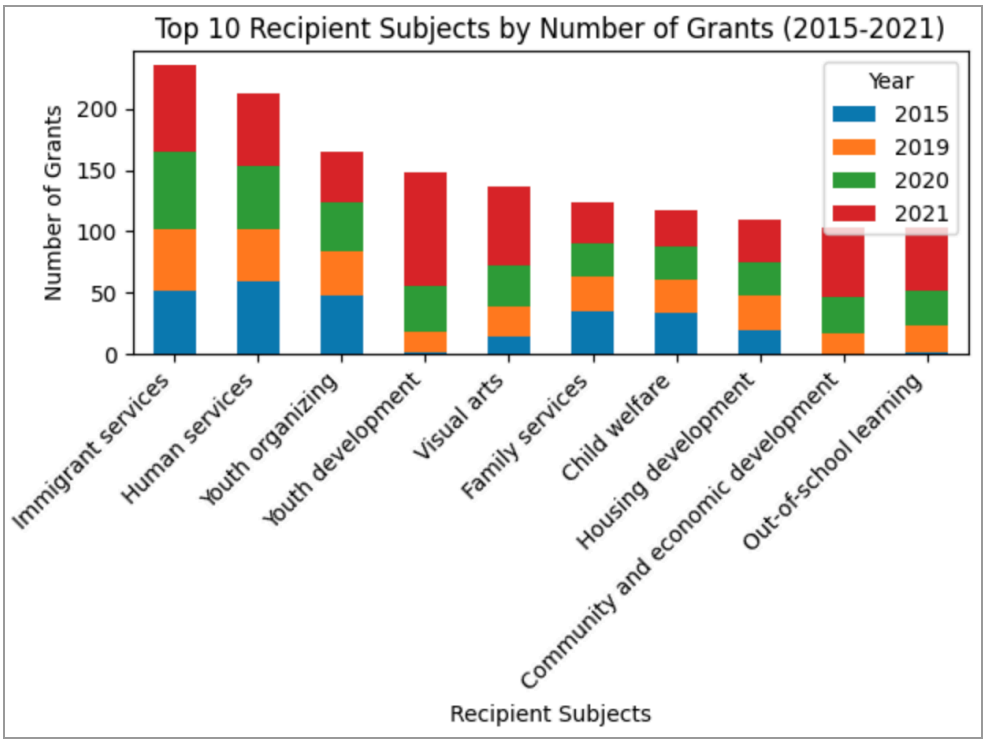


When it came to finding other trends that exist within these non-profits, we visualized the missions of the non-profits and the grants. It appeared that 'Youth Development Programs' and 'Education' seemed to be among the most common themes that reappeared throughout.

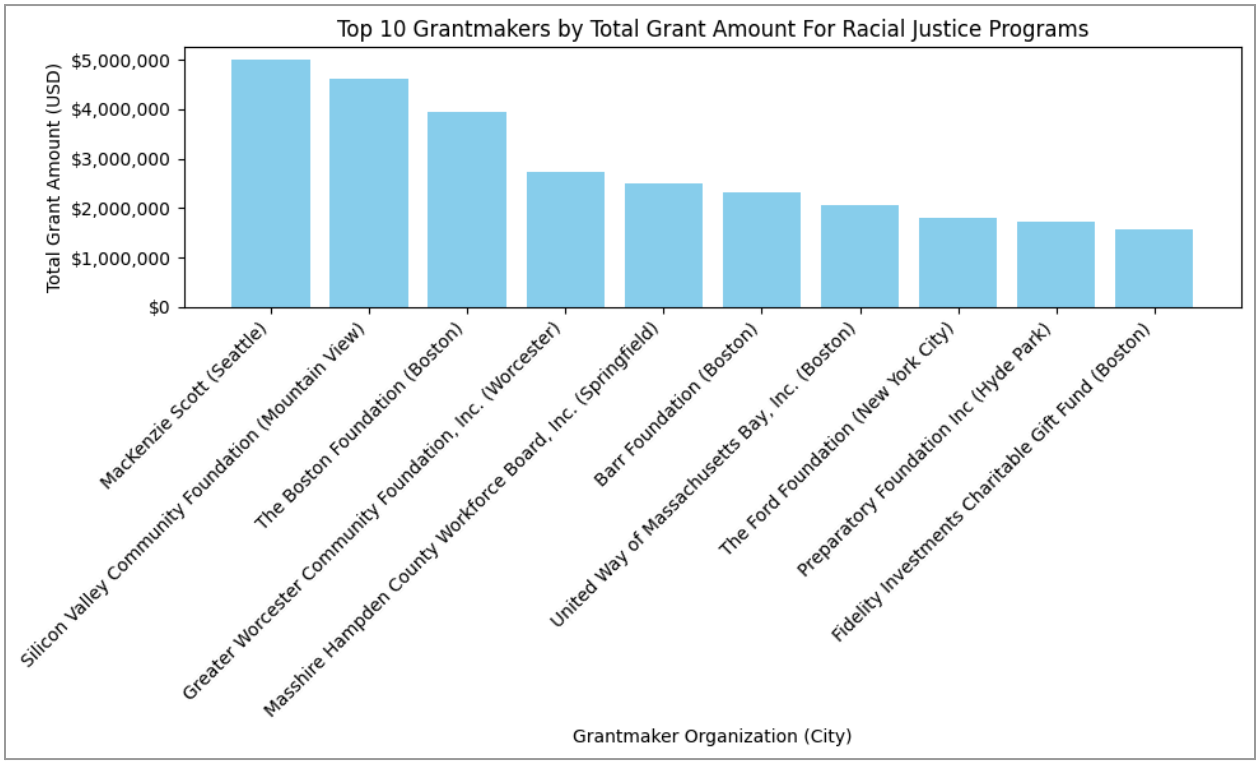
Missions of overall nonprofits who serve the Black and LatinX community







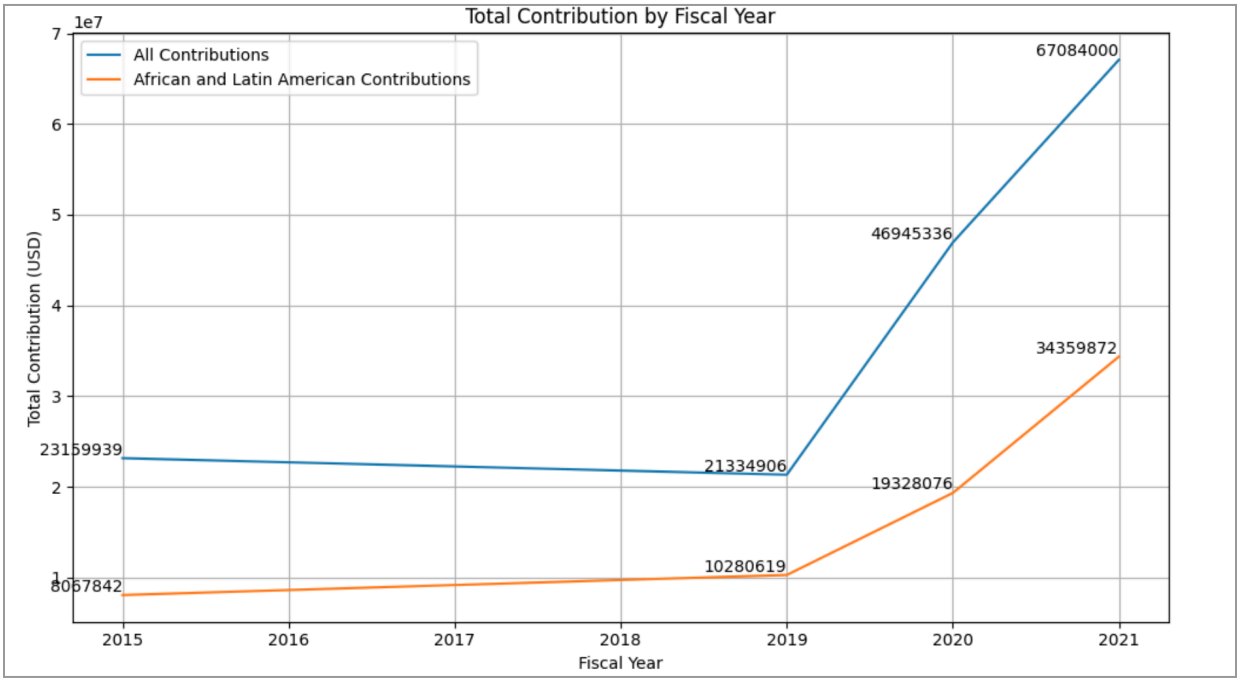
When it came to top funders, it was clear that these were some big, notable names in the philanthropic sector. Some of them were from Boston, while others were across the country.



These were our most notable findings when it came to figuring out answers to the key questions, and as expected from the beginning, they do confirm the notions that: there is increasing funding going towards non-profits supporting the Black/LatinX community, but many of those organizations aren't being led by the community that they support.

Base Analysis

Key Question 1: How much funding is going toward racial justice programs (based on donations to nonprofits that focus on these issues) in Massachusetts and how has this shifted over time?



Source: 'Grants' Bundle; we grouped contributions based on 'fiscal_year' on the entire dataset and a subset of the dataset that only contains grants with corresponding 'recip_population_code' which corresponds to the populations served by the organization.

The general trend of contributions to Black and Latin American organizations seems to be rising with the trend of total contributions to all organizations. However, the exact proportion of funding for racial justice programs in Massachusetts has gone up and down. It rose in 2019, went down slightly in 2020, and rose even higher in 2021 to where over 50% of funding went towards racial justice organizations.

Key Question 2: How much funding is going toward nonprofits led by Black or LatinX individuals (president, executive director, board chair, etc.) in Massachusetts?
$ 11,058,130

The df_grants and df_organizations are merged on ein (unique employee ID number). There are 63,668 unique IDs in the organizations bundle. There are 412 unique recipient IDs in the grants bundle. After the merge, only 119 IDs are retained. This is because some recipient IDs in the grants dataset are not in the organizations dataset. Specifically, 293 eins or 32.74% of total IDs from the grants dataset are not in the organization dataset.

Thus, only 28.88% of the total eins in the grants dataset are found in the organization dataset. When merging on organization name previously, 82.68% of data was *lost* because of naming inconsistencies. Thus, merging on employee ID numbers yields more accurate results.

Furthermore, a marginal amount of data is lost after grouping the merged data by race leader. Our primary analysis uses race leader as a categorical variable, meaning that organizations with no race leader specified (NaN values) are omitted. Specifically, 5.04% of the data is lost, which is only marginally better than 5.06% when we merged on the organization name.

Nevertheless, we get a total of $11,058,130 of funding towards organizations led by African or Latinx communities, which is 23.68% of total funding. Again, the total funding is based on only matched IDs and only ones where the race leader column is specified.

Using ein (unique employee ID number):

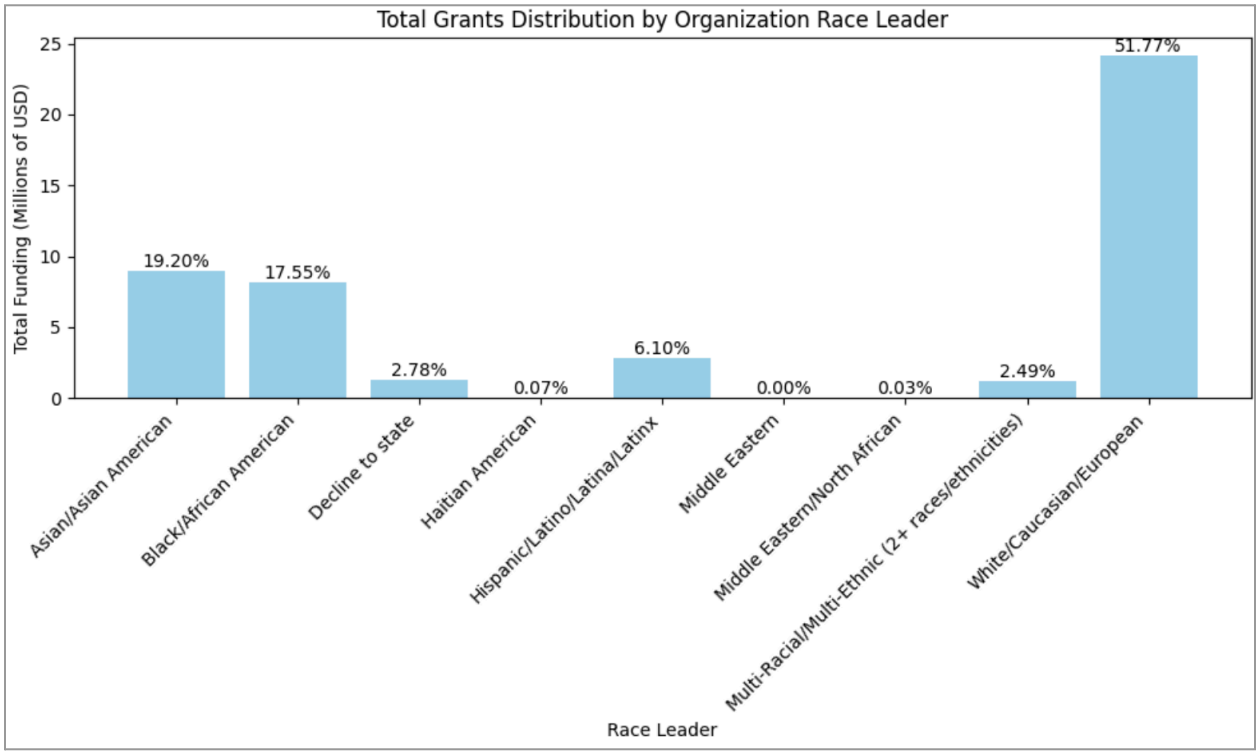
```
10.26 % of organizations have no race_leader data
73.90% of organizations that received grants were lost during the merge
5.04% of additional data was lost during grouping because of missing race leader values:
```

Using org_name (name of organization):

```
10.26 % of organizations have no race_leader data
82.68% of organizations that received grants were lost during the merge
5.06% of additional data was lost during grouping because of missing race leader values:
```

Source:'Grants' Bundle in conjunction with 'Organizations' Bundle; we merged the two datasets based on unique employee ID numbers, then grouped by the 'race_leader'.

The graph below displays the percentage of funding given to each organization grouped by the race of the leader out of the total percentage of funding. While there are limitations in the data, the overall trend is very telling. Specifically, the majority (51.77%) of funding is going towards organizations led by Caucaisan people. Additionally, 17.55% is going towards organizations led by African Americans, 6.10% towards organizations led by Latinx individuals, and 0.03% going towards organizatiosn led by Middle Eastern or North African people.

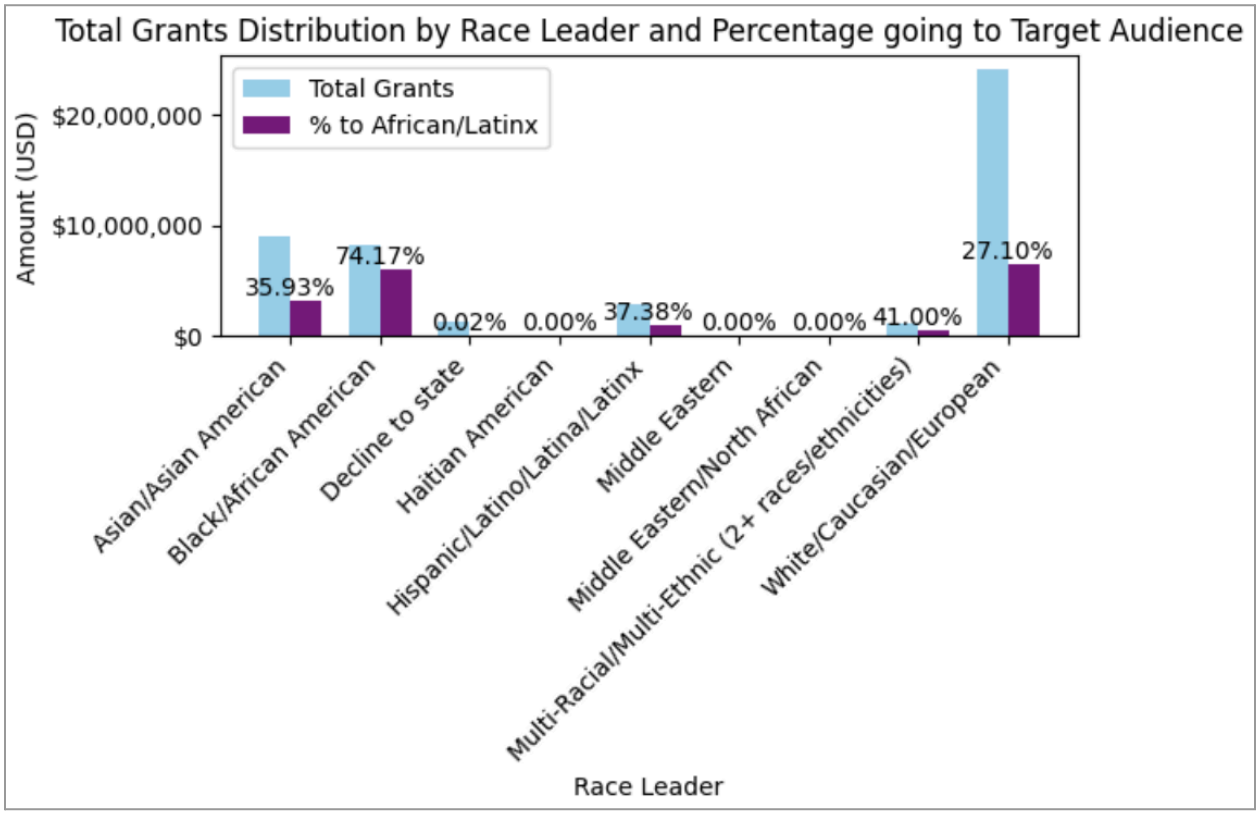


Source:'Grants' Bundle in conjunction with 'Organizations' Bundle; we merged the two datasets based on unique employee ID numbers, then utilized columns 'race_leader' and 'amount_usd'.

Out of the 127 organizations that target the African or Latinx community, 38 total organizations have also been given grants or 29.92% of all organizations. Again, this number is higher than the previous 24 total organizations due to naming inconsistencies from merging on the organization name.

We know that all of these organizations listed as having received grants are also in the merged data frame above; they are a subset of all organizations that have received grants. This is because these organizations that target the African/Latinx community are grouped by race leader, which wouldn't be omitted from all of the organizations that have received grants and are grouped by race leader. Because of this, we can easily divide the percentage of amount_targeting_african_latinx_by_race_leader / total_amount_by_race_leader to find the percentage of organizations that target the African/Latinx communities out of the larger set of organizations who are given grants and are led by the African/Latinx community.

The graph below shows the total grant amount received by organizations led by African American or Latinx individuals, and what percentage of those organizations are also aimed at targeting minority communities.

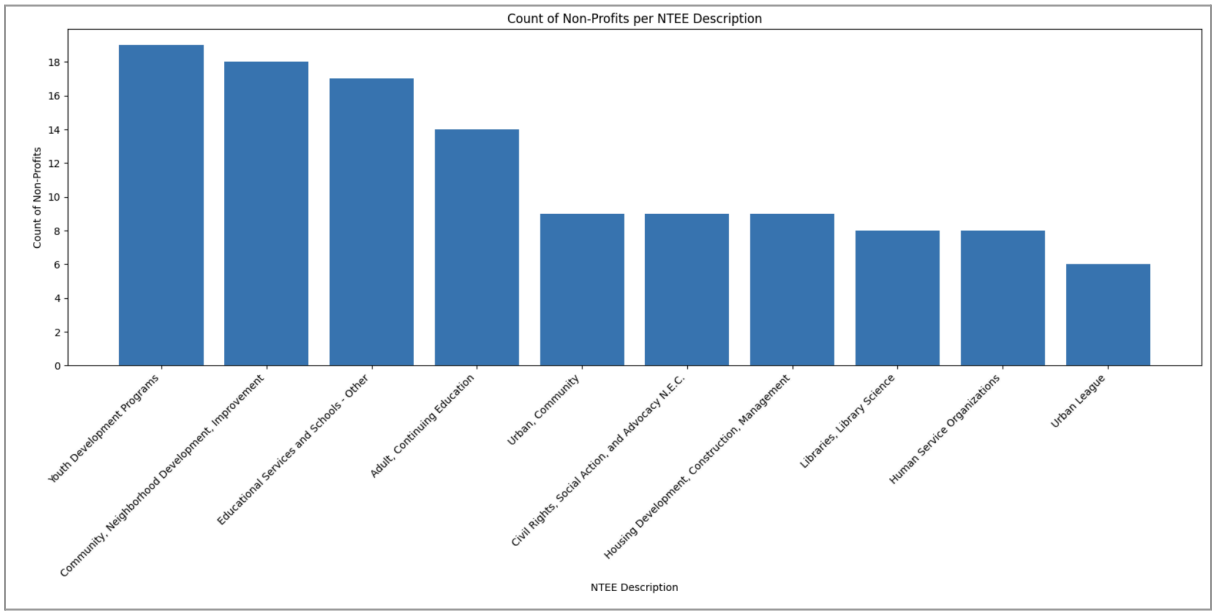


Source:'Grants' Bundle merged with 'Organizations' Bundle and grouped by race leader similar to the graph above.

Key Question 3: What are other trends related to the type of organizations and grants that are being funded serving the Black and LatinX communities in Massachusetts e.g. missions of the nonprofits, issue focus of the grants, changes in amounts donated over time, etc…

* NTEE, also known as National Taxonomy of Exempt Entities (NTEE) Code, specifies the exempt organization's purpose, so this tells us more about the mission of each of these organizations.

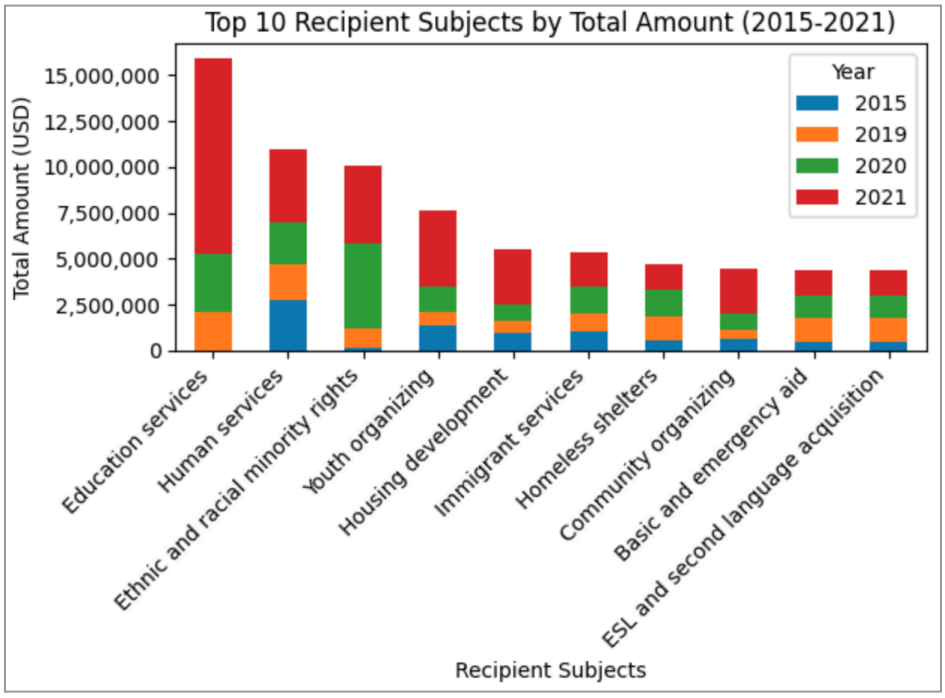
This visualization provides a overview of the number of non-profit organizations with the same NTEE descriptions. A quick glance shows that, after grouping organizations with the same NTEE code together, the mode of this data is 3 – meaning most commonly, there are 3 organizations with the same NTEE codes. This number jumps for the most popular NTEE codes. We analyze the specific objectives and programs later in more detail within the extension analysis.



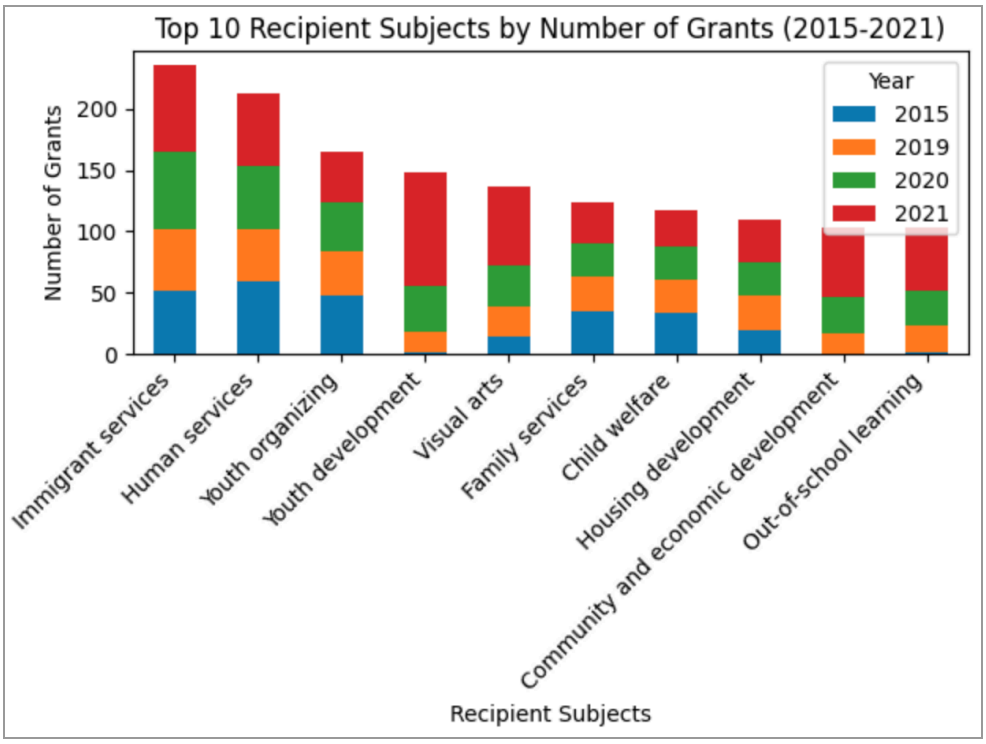
Source: 'Base Bundles' dataset; grouped by 'NTEE' counts.

The three most common NTEE* descriptions that non-profits share are 'Education Services and Schools', 'Community, Neighborhood Development, Improvement', and 'Youth Development Program'.

Additionally, the graphs below specify what subjects the organizations (from African/Latinx population codes) that received the grants are focused on. As you can see, the varied colors indicate the proportion of money granted towards the recipients by fiscal year. Clearly, Education services received the highest number of grants, with a total of over $15 million in funding from the years 2015-2021. There is a large difference between Education services and the next highest-funded subject, which is Human services, with a difference of around $4 million. This is roughly equivalent to the amount of funding going towards the least-funded subject, ESL and second language acquisition.



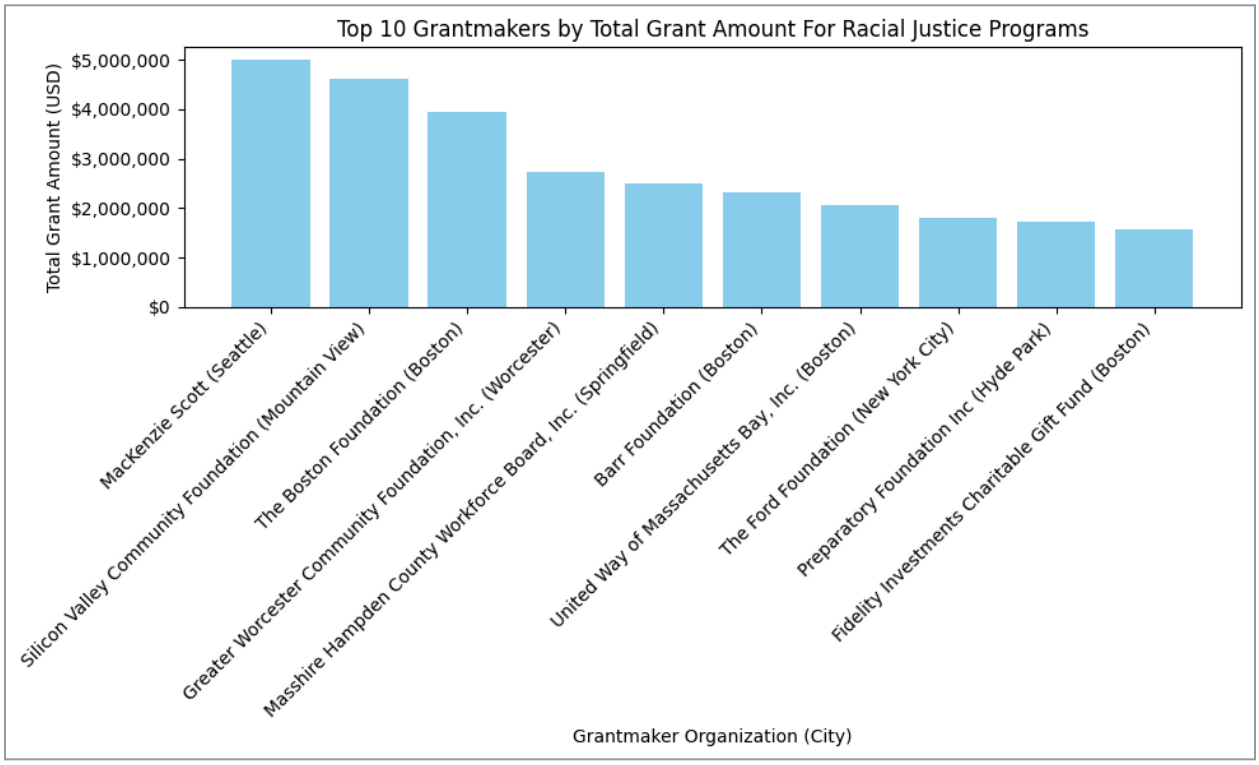
Source: 'Grants' Bundle'; grouped by 'recip_subject' which represent the core activities and services of the organization.



Source: 'Grants' Bundle; grouped by 'grant_subject_tran' which are the subject categories of grants.

Looking at the top 10 total amount of grants given per category versus the top 10 number of grants given per category, there is some overlap in the categories (which is to be expected since it makes sense if more grants would lead to a higher total amount), but not all the categories overlap with each other. This can mean that even though some categories have more grants being made, their value is actually not that high. And looking at the years of contributions, it seems that 2021 generally had more contributions for these top categories, which also lines up with the trend that more funding was given during 2021 compared to other years for African/Latinx communities.

Key Question 4: Who are the top funders of racial justice programs in Massachusetts?



Source: 'Grants' bundle is filtered to contain only the African or Latin American population code. It is grouped by the 'grantmaker_id' column and aggregates the data using gm_city and amount.

The graph above shows the top 10 grantmakers for racial justice programs based on grant amount. Interestingly the top two organizations, Mackenzie Scott and the Silicon Valley Community Foundation are organizations not based in Massachusetts at all as you might expect. The Mackenzie Scott foundation awards grants to various organizations of different causes. While the Silicon Valley Foundation specifically aims to serve the Silicon Valley community which makes it very interesting to see them here.

Extension Analysis

Key Extension Prompts / Questions

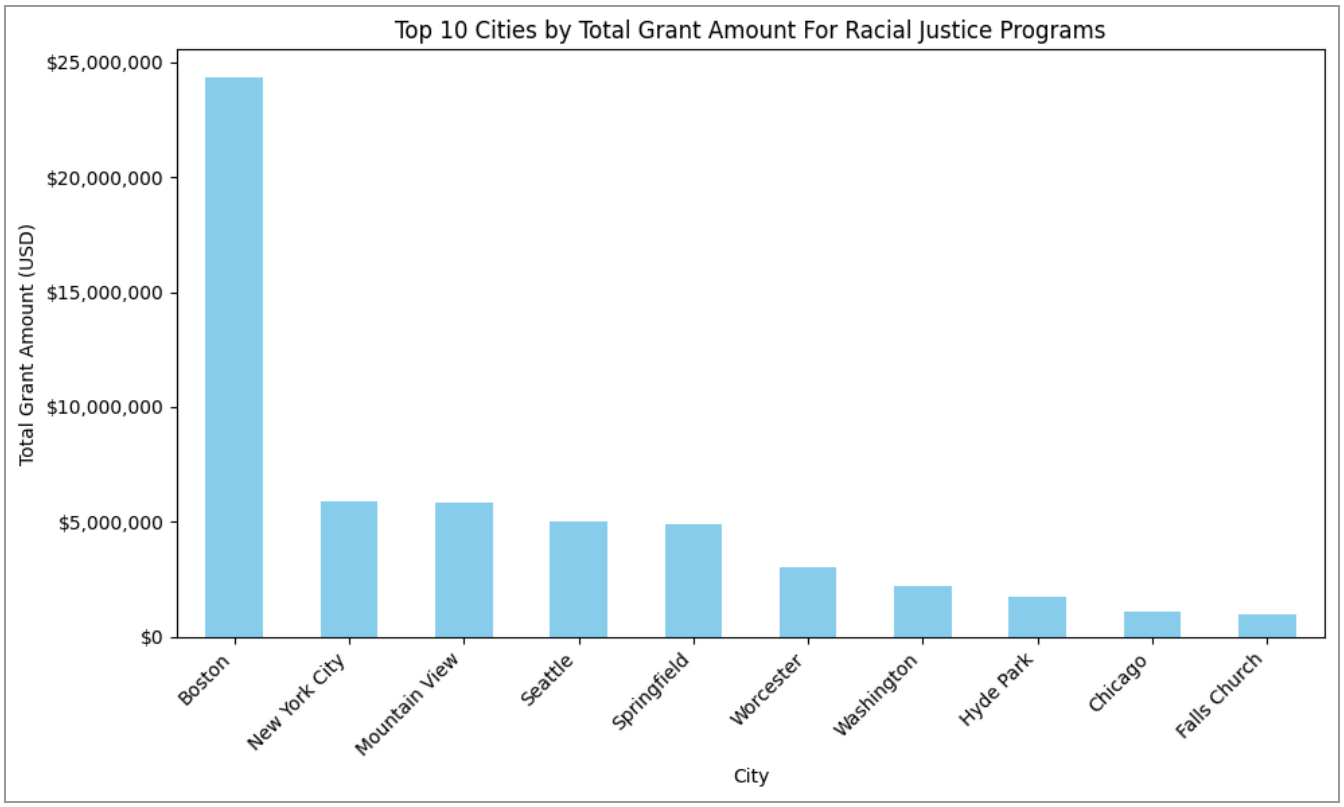
1. Refining analysis of contributions with additional parameters:
    a. Filter out hospitals, churches, and higher education institutions
    b. Filter out grantmakers
2. Propose a methodology for using and filtering by NTEE codes to extract more meaningful insights from organization mission. Why did you make the decisions that you did?
3. What is the proportion of contributions to Black/brown communities compared to all grants?
    a. How has this share of contributions changed between 2015 and 2020?
    b. Where are the recipient organizations located? How does this compare to state demographic data?

Extension #1: Refining analysis of contributions

For the base analysis of our project, we explored much into the non-profit organizations that were receiving the grants. This included how much funding they were receiving, the demographic and their missions. However other than identifying some of the top grant makers funding these non profits for racial justice, we did not look into much else about them. We decided to pursue this in our extension analysis.

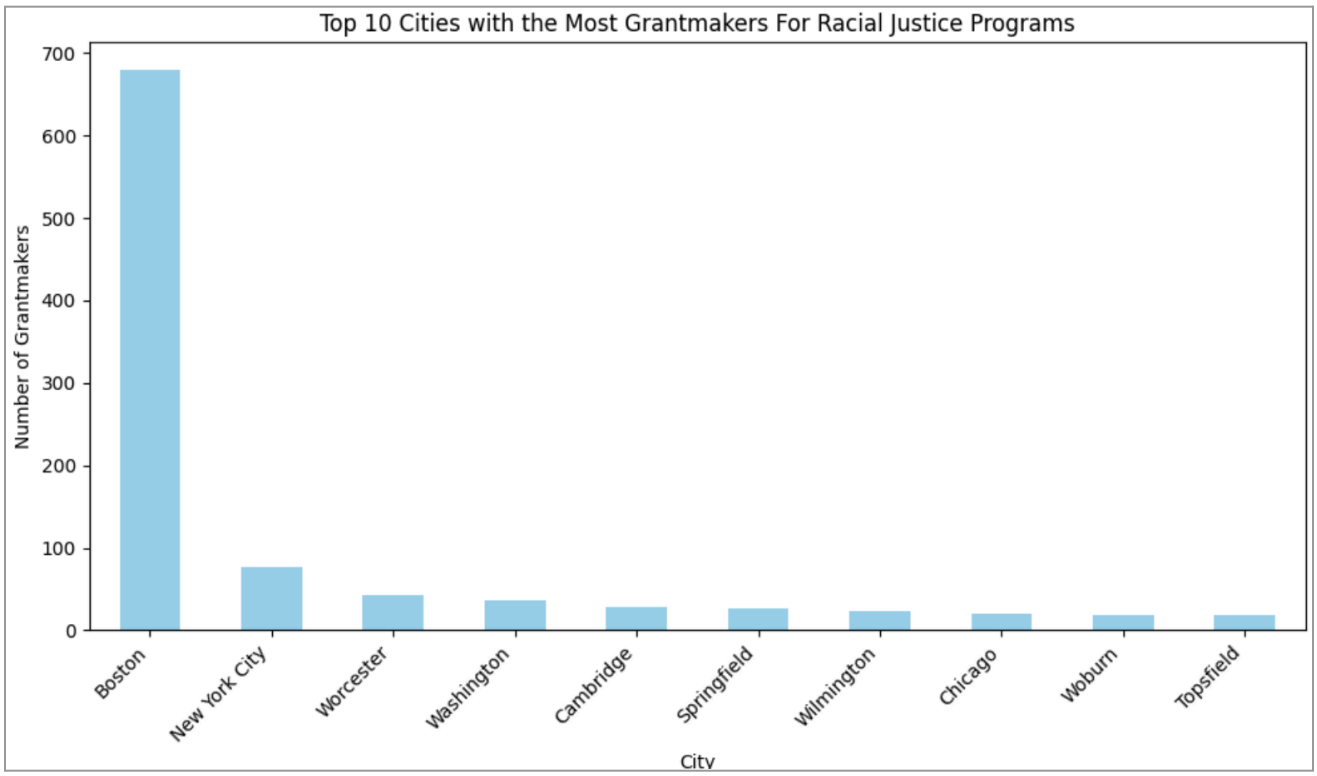
Looking into grant maker information helps nonprofits in many ways. It shows what funders care about so organizations can ask for money that matches those interests. It also helps find new funding sources, make friends in the charity world, see where money is going, ask for fair funding, check if their work is making a difference, and use that data to improve. This helps organizations navigate the funding landscape effectively and achieve greater social impact.

This led us to dive deeper to explore the primary contributors by city. As you might expect Boston emerges as the frontrunner, significantly outpacing other cities by allocating nearly $25 million, which is nearly five times greater than the funding from the city with the next highest contribution.



Source: 'Grants' Bundle; we grouped grants together by 'gm_city' and aggregated based on 'amount_usd'

Seeing this prompted us to do further analysis on where the most grant makers are. If Boston had the highest amount of grants, it would make sense that most of the grant makers are donating from Boston, since they would be giving back to their own communities, so we looked into that next.



Source: 'Grants' Bundle; we grouped grants together by 'gm_city' and aggregated based on unique values of 'gm_name'

This graph shows the top cities by total number of grantmakers. Boston again, by far the city with the greatest number of grantmakers as we suspected.

Since Boston came up on top for the most grant amount and grant makers, it would be good to do a qualitative analysis in the future for why this is the case. Understanding what makes the city special for nonprofits could be very important. We'd talk to people who give grants, run nonprofits, and know about the community to figure out why Boston stands out. This could show us what Boston does well and what challenges it faces, helping us make better decisions to support nonprofits and make the city even better, and expand that to beyond Boston potentially to provide support in other places.

Additionally, we know that certain types of organizations may skew the data in a certain way (ex: since education receives a lot of money, then perhaps it would be interesting to see how the data looks like without educational institutions), so we looked into how it would be possible to filter out certain types of organizations. We found that there's two methodologies that we can use.

Our first methodology would be filtering out organizations based on key words in their names. Our hypothesis is that, for instance, if a non-profit organization is a hospital, it would likely contain the word 'hospital' in its name. However, this methodology didn't provide much accuracy, because we ended up with these results:

```
There are 98 organizations whose name contains `hospital`
There are 0 organizations whose target population is the Black/LatinX community and whose name contains `hospital`
```

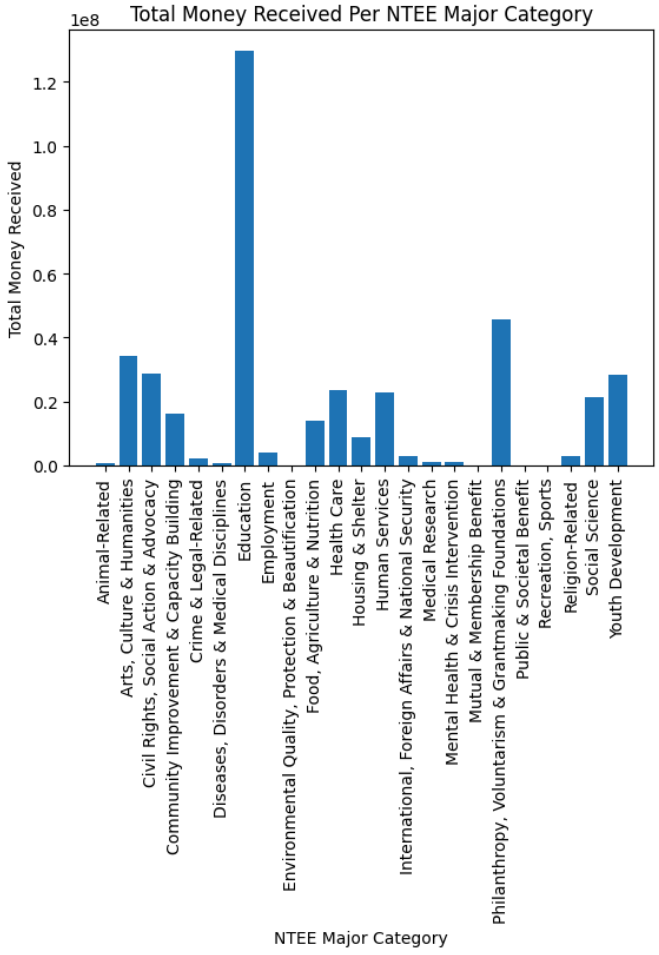
```
There are 86 organizations whose name contains `church`
There are 0 organizations whose target population is the Black/LatinX community and whose name contains `church`
```

```
There are 64 organizations whose name contains `university`
There are 0 organizations whose target population is the Black/LatinX community and whose name contains `university`
```

Source: 'Organizations' Bundle
Columns used: 'organization_name'; checking if it contains corresponding word

Our second methodology was looking into the NTEE codes and seeing if each of these filters pertained to a different category. Turns out, they do. For instance, all non-profit organizations that pertain to health care start with 'E', and all non-profit organizations

that pertain to Education start with 'B'. Therefore, we were able to re-categorize all non-profits to fit into a broader NTEE category so that we can visualize a larger trend.



Source: 'Grants' and 'Base_Bundle'; merged on 'ein_name' and 'ein'; added a new feature for major NTEE category and plotted based on that.

However, there was a challenge when it came to merging Grants data and organization data without aggregating the Grants data in a certain type of way - it's hard to preserve all grant information when it comes to merging the two datasets, because when there's duplicate values in either table in the column you are trying to merge it, Pandas creates duplicate values. Therefore, the resulting merge created issues in accuracy, and this table is bit off when it comes to the numbers. For our previous questions, we just immediately aggregated the total amount an organization receives, but it would be more useful if we were able to merge the information without having to immediately aggregate data. We weren't able to figure out how to make this merge without losing data in the meantime, but this would definitely be a good next step for the project and/or an area of an exploration for the DS working on this project.

Extension #2: Propose a methodology for using and filtering by NTEE codes to extract more meaningful insights from organization missions. Why did you make the decisions that you did?

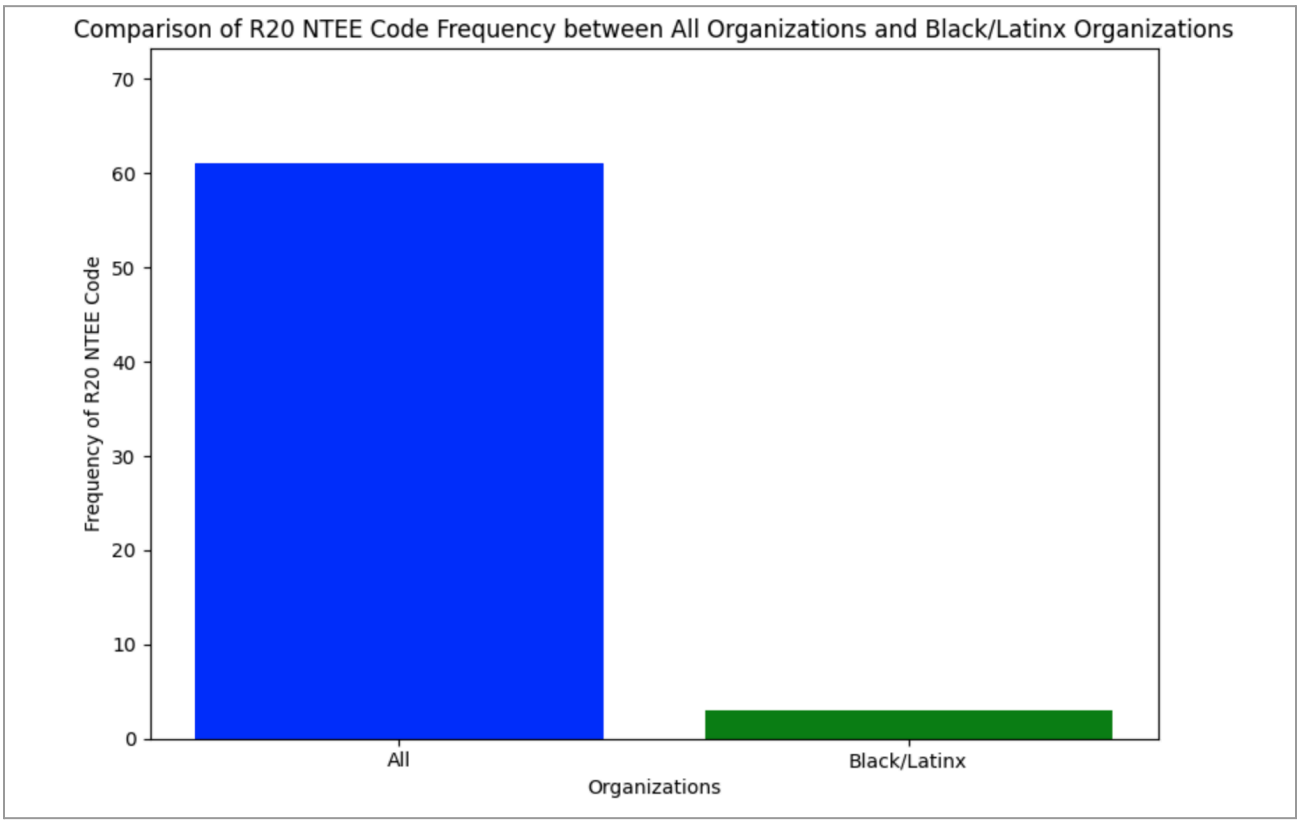
For most of our project, we did analysis using PCS (Philanthropic Cause Classification System) codes to determine funding allocations and trends related to racial justice programs and nonprofits led by Black and LatinX leaders. This involved examining how financial resources were distributed across various philanthropic causes, including racial

justice initiatives, and assessing the extent to which nonprofits led by Black and LatinX individuals received funding support. We analyzed trends in grantmaking, funding sources, and the alignment of financial allocations with community needs and priorities that align with black/brown PCS population codes.

Given this foundation, we chose to look at NTEE (National Taxonomy of Exempt Entities) codes for a different perspective on organizational missions, activities, and legal classifications within the nonprofit sector. This shift allows us to gain insights into the broader landscape of nonprofit organizations, beyond just funding considerations. We are now exploring the overarching purposes, objectives, and programs that nonprofits undertake to achieve their social or charitable goals, as well as understanding the legal and administrative context in which these organizations operate.

So to start, we can utilize filters based on NTEE codes starting with R20, as these codes are associated with Civil Rights, Social Action, and Advocacy specifically aimed at the Black community. While R20 is tailored for Black causes, R40 is designated for Latinx causes, so ideally looking into both would be what we want to include both black and Latinx communities. However, within our dataset, only R20 and R99 codes are present, with R99 representing Civil Rights, Social Action, and Advocacy activities that are not classified under more specific categories (Not Elsewhere Classified). Therefore, narrowing down to only the Black NTEE social justice codes indicates a broader involvement in social advocacy initiatives supporting the Black community, beyond organizations explicitly focused solely on Black causes.

Looking at the graph, there are a lot more organizations indicating they support civil rights and racial justice issues affecting the Black community than just the ones with PCS codes for Black/Latinx organizations. Some useful analysis we could do in the future is compare the organizations with PCS supporting black/brown and NTEE codes supporting those causes, looking at how the funding trends differ for the different groups.



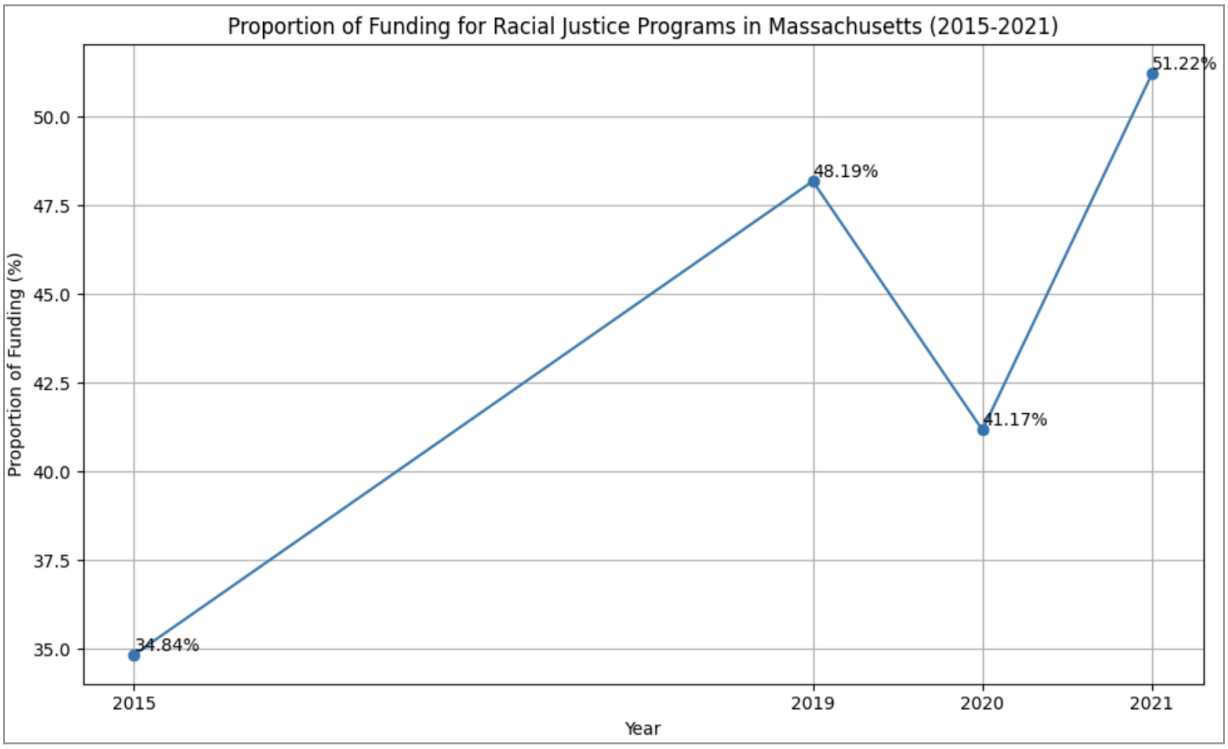
Source: 'Base_Bundle'; filtered by 'NTEE' codes

Extension #3: What is the proportion of contributions to Black/brown communities compared to all grants? How has this share of contributions changed between 2015 and 2020?

This question is important because it looks at whether Black and Brown communities get a fair share of funding compared to others. We want to know how much money goes to these communities compared to everyone else, and if this has changed from 2015 to 2020. This helps us see if there's been progress in making funding more equal for everyone.

Since we only have data for 2015, 2019, 2020, and 2021, we could not do adequate analysis on funding trends between 2015-2019 to determine if funding trends usually fluctuate this much or if it was a result of external factors happening around the world. However, the proportions of funding for racial justice programs went up from 2015 to 2019 from 35% to where almost 50% of all funding in our data went to racial justice programs. There was a slight dip to 41% in 2020, and the proportion went back up to

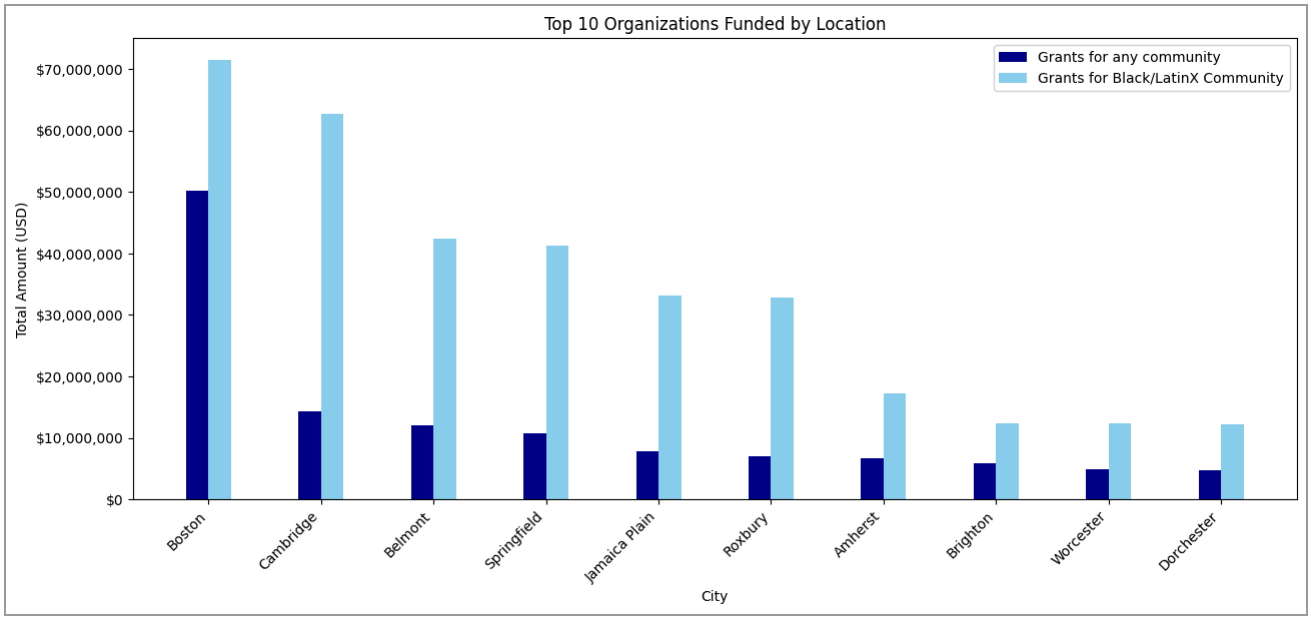
51% in 2021. We lack enough data to explain why this happened or identify if any of these numbers are unusual.



Source: 'Grants' Bundle; grouped by 'fiscal_year' on the entire dataset, filtered by 'recip_population_code' and compared that to the unfiltered dataset

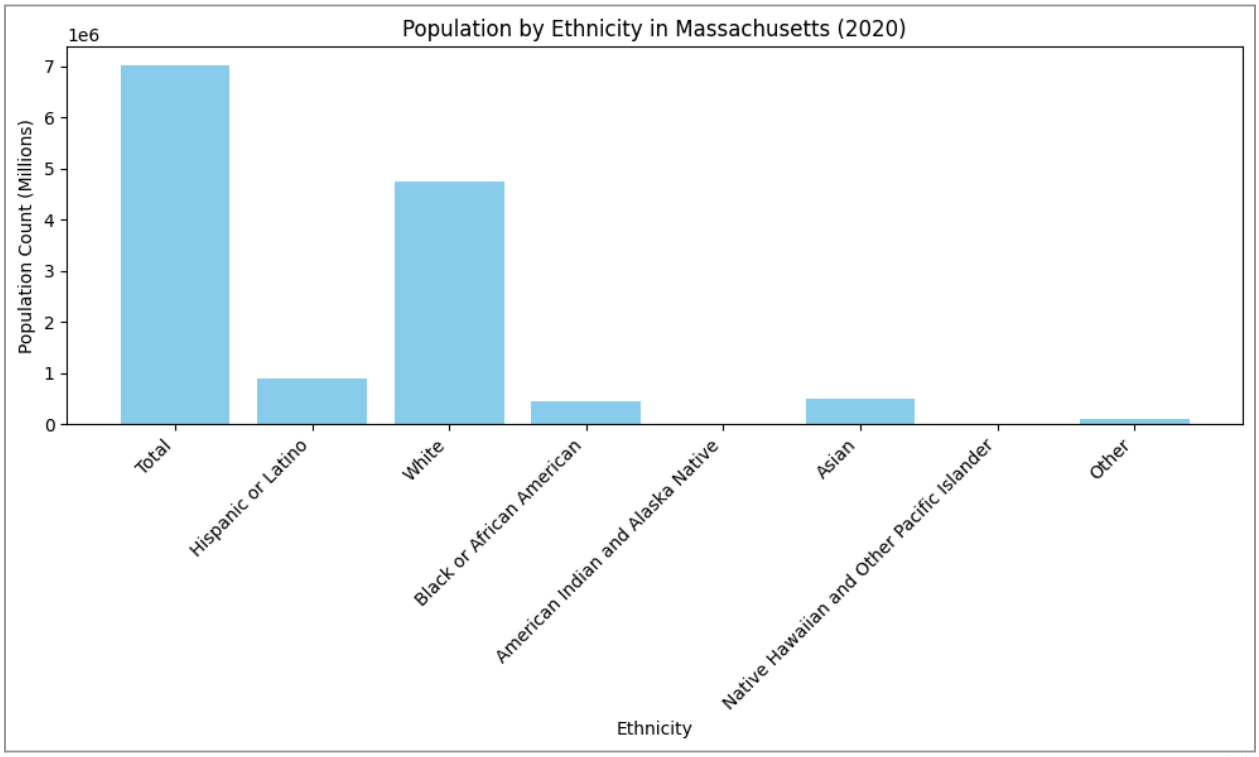
Where are the recipient organizations located?

In addition to comparing the amounts of donations for Black/LatinX organizations vs all donations, we also wanted to see if trends for where the donations were given stayed the same when we looked at all donations (rather than just Black and LatinX donations).



Source: 'Grants' Bundle; grouped by 'recip_city' and aggregated by 'amount_usd'

This graph shows the comparison between total grants to Black/Latinx communities versus any community per city. We see that in Massachusetts at least, the majority of funding goes towards these racial justice programs, but in other cities, there is a split between funding going more in other places. This further supports the trend we've noticed in Boston, where there is a consistent emphasis on donations towards racial justice initiatives.



Source: U.S. Census

When analyzing funding allocations for Black and Latinx communities, we thought it was important to put it into the context of demographic data. As shown above, the black and latinx community makes up an actually small portion of the population overall. So it is significant to note that these communities receive a considerable portion of funding, indicating recognition of their needs. However, further investigation is needed to ensure funding aligns with population realities and addresses any disparities. Understanding systemic barriers and engaging stakeholders can help promote equitable resource distribution, ensuring support for the diverse needs of these communities.

Future Scope

1. Integration of Qualitative Data

Incorporative qualitative research methods, such as interviews and case studies, to complement the quantitative data analysis. This could potentially provide deeper insights into the effectiveness of funded programs, the challenges faced by nonprofits, and the real-world impact of philanthropic funding on communities.

2. Research Donor Motivations and Strategies

Investigate the motivations, strategies, and decision-making processes of the major donors and grant-making organizations that we found from our research. Understanding why donors choose specific organizations or focus areas, and how they measure the impact of their contributions, could lead to better-targeted and more effective philanthropic efforts.

3. Long term Analysis

Explore the temporal scope of the data beyond just the current dataset. A study over the next couple of years could provide insights into the sustainability and long-term trends of funding towards racial justice. This could also allow for the assessment of the impact of global and national events on funding trends and how they might align with commitments made by philanthropic organizations. Identifying certain milestones and policy changes that correlate with donation patterns could pinpoint the impact of external factors on grant donation trends.

4. Geospatial Analysis

Analyzing the geographical space could provide insights into which parts of Massachusetts receive the least and most donations, and that along with analysis for the demographic data around that area and distance to donors could provide more information about donation trends.

5. Impact Analysis

Figuring out the impact of the grants on the organization could provide better marketing in the future to encourage donors to donate by showing them what their contributions are bringing for these organizations.

Individual Contributions

Sharon: As the team lead, I acted as a liaison between our team and Spark! - I presented at client meetings and PM meetings on behalf of the team. Additionally, I worked on finding mission focus for organizations for Key Question #2 and filtering out fields for Extension Question #1.

Fiona: I worked on using the data to discover the proportion of funds from African/LatinX compared to all donations, and looked into how the amount/count of grants for different missions of organizations/goals of grants changed throughout the years.

Esther: I worked on using the data to answer key question #2, how much funding is going toward nonprofits led by Black or LatinX individuals (president, executive director, board chair, etc.) in Massachusetts. I was able to generate graphs for the funding based on race of the organizations' leader and also what percentage of those organizations are also aimed at targeting minority communities.

Justin: I worked with the data to create graphs for some key questions. The first key question I helped create a graph for was about the proportion of contributions to Black/brown communities compared to all grants. In particular I made a graph answering this question based on the recipient organization's location or city.

Tyler: In summary I worked with the data to find and create visuals to help answer some of the key questions. Key questions I worked on included figuring out more information about the grant makers. This means diving deeper into what organizations they were and where they were located as well as the grant receiving side to better understand context.