Explore Boston's Neighborhood Inequity

Umarbek Nasimov
unasimov@mit.edu

Victor Chau
vchau@mit.edu

1 Introduction

Boston has a history of segregation due to discriminatory housing policies from the pre-civil war era. In fact, out of the country's 51 greater metropolitan areas with large Black populations, it ranks 15th for segregation. There also exists inequities and income segregation across Boston neighborhoods. A 2015 study conducted by the Federal Reserve Bank of Boston revealed that a white household's median net worth is $247,000 whereas in an African American's household, it is $8. In addition, different Boston neighborhoods are also affected by crimes in various ways. While some experience more overall crime, other neighborhoods are more susceptible to severe crime types such assault. As a result, we seek to understand the correlation between crime and demographics of Boston's neighborhoods through a scrolly-board, martini visualization. We hope that our visualization reveals how Boston's history continues to affect the lives of its residents today.

2 Related Work

There has been extensive discussion on the demographics of Boston's neighborhoods. In an article in the Boston magazine, Catherine Elton outlines how Boston's discriminatory housing and lending policies have left long-lasting effects of income and opportunity segregation [6]. She describes racial segregation evolved from individual decisions to not sell or rent to Black people in the late 1800s to formal, legal discriminatory housing laws in the 1900s. Importantly, she also highlights the experiences of specific individuals such as Sheena Collier, a graduate student at Harvard Graduate School of Education. After graduating from a historically black college in Atlanta, Georgia and arriving in Boston, she had the impression that there weren't a lot of people of color in Boston. When she moved to Roxbury and began managing after-school programs, she describes how her whole perception of the city shifted and that although Boston is diverse, "it's just really segregated".

There has also been a wide array of research into Boston's crimes. The U.S Department of Justice has written an extensive quantitative study on the nature of the extent of the most serious crimes occurring in Boston's police districts [5]. Crime reports, however, can also be manipulated in dangerous ways. A paper by Andrew J. Branauskas describes how the news media tends to disproportionately report more robberies and assaults in disadvantaged neighborhoods [4]. Though we don't directly discuss this in our visualization, we acknowledge in section 6 how future work needs to explore this issue.

A previous visualization called Crimes in Boston analyzes the same crime dataset we used, but neighborhood inequity isn't heavily discussed [1]. In our visualization, we aim to inform users of both of the issues.

3 Methods

In this section, we will describe the two datasets that we used as well as the pre-processing we implemented for each dataset. We also explain the way we generated maps.

3.1 Datasets

The first dataset we used is Boston crime data from Analyze Boston, which is Boston's open data hub [3]. It contains crime incident reports provided by the Boston Police Department where each report contains the crime type and location. The second dataset we used is Boston demographics data from Analyze Boston [2]. This dataset contains each Boston neighborhood's age, race, poverty rate, income, and education attainment distribution. Note that Analyze Boston, the source of both of these datasets, is aiming to be the default technology platform to support the publication of Boston's public information.

3.2 Boston Crime Data

In order to improve the focus of our crime visualizations, we filtered a subset of the available crime data. First, we filtered for crime incidents that occurred between 6/15/2015 and 7/15/2015 to remove the clutter on our point visualization, described in section 4.4. The dataset also contained all indicidents that officers reported, including helping sick people, so we further filtered the crimes on high severity and high relevance. We define severe crimes as ones that are directly life threatening and relevant crimes as ones that occur frequently such that trends are noticeable. We then aggregated crimes with similar offense codes, a feature of each crime incident, into more general crime types. For example offense group codes Larceny and Larceny From Motor Vehicle were combined into just Larceny. Using these filters, the final aggregated crime types we used are: drug violation, larceny (theft), vandalism, assault, firearms and explosives, burglary, disorderly conduct, robbery.

3.3 Boston Demographics Data

To join the two datasets together such that they can be used in a single visualization, we aggregated the demographics data of certain neighborhoods. In each crime incident, there existed a police district in which the crime incident occured while in each neighborhood demographics statistics, there existed a corresponding neighborhood name. Boston neighborhoods exist within certain police districts as shown in Figure 1. To visualize the correlation between crime and demographics, we aggregated the neighborhoods demographics data to create police district neighborhoods demographics. For the neighborhoods that do not fit inside of a single police district, we assigned them to the district that covers the majority of their area.

3.4 Generating Maps

For visualizing geospatial data of Boston neighborhoods and police districts, we applied the albers projection on their topojson representations. For the albers projection, we applied a dynamic scaling function based on the viewer's window size such that the map visualizations could be properly viewed.

4 Results

In this section, we will describe the visualizations that our system produces and our design justifications. We used a scrolly-board, martini visualization in order to add context to the complex issue of segregation and crime. In particular, the scrolly-board introduces the crime dataset and the demographics dataset separately, highlighting key statistics in each. Within these introductions, the scrolly-board also adds clarity and transparency to how we transformed our data.

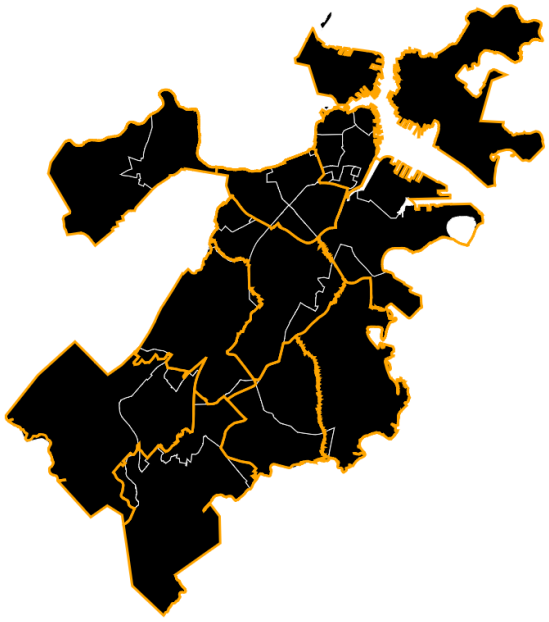

Figure 1: A visualization of Boston neighborhoods, shown in the white outlines, Boston police district, shown in the orange outlines

At the end of our scrolly-board, we combine both crime and demographics data in a single interactive visualization.

4.1 Police Districts and Neighborhoods

The first part of the scrolly-board describes Boston's police districts and neighborhoods, as well as their relationship. The accompanying map initially shows the map of Boston's neighborhoods, outlined in white with a black background. When the description of Boston's police descriptions is scrolled into view, the map updates to show the police districts, denoted by orange outlines. Afterwards, we instruct the user to hover over police district names to view the neighborhoods contained within them on the map. This interactivity describes police districts and neighborhoods, including their names and location. It also clarifies that neighborhoods exist within police districts. We make an explicit disclaimer of the fact that some neighborhoods cover more than 1 police district.

4.2 Crime Visualizations

Then, the scrolly-board introduces the crime dataset by highlighting key statistics. The first visualization is a choropleth map of the total count of crimes across police districts. This highlights the idea that certain districts are generally more affected by crime than other districts. Then, there is a pie chart of the overall distribution of crime types in Boston. This emphasizes that some crime types are more common than others as well as the fraction of total crimes that severe crime types make up. We chose to omit interactions on these visualizations to add focus to their respective messages. The final visualization is two choropleths of crime types in police districts. Within each choropleth, we allow the user to explore how a given crime type disproportionately affects some districts more than others. Note that we didn't display the overall distribution of a crime type across every district because that would omit the number of crimes that happen within a district. Instead, we display the portion of total crimes a given crime type within a district such that a valid comparison between districts can be made. We include two choropleths side-by-side to allow users to compare trends between

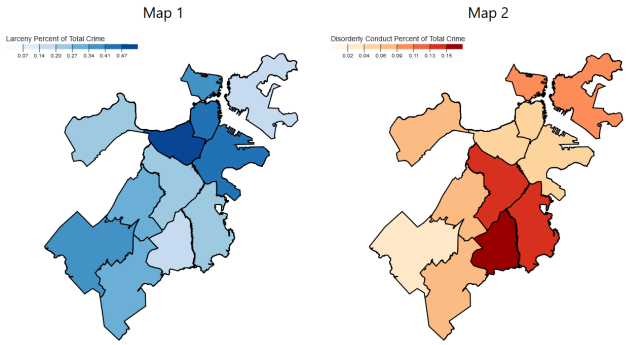

Figure 2: Default view of the crime type choropleths. Map 1 reveals how almost half of the crimes in the Fenway area are larceny, the most out of any district. Map 2 reveals how 15% of Mattapan's crimes are disorderly conduct, the most out of any district.

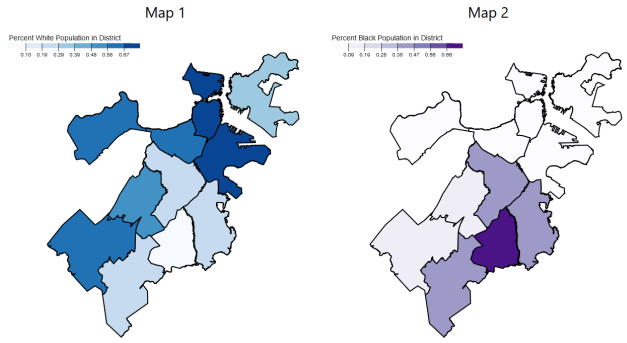

Figure 3: Default view of the race choropleths. Note how Mattapan's black population is 66% (Map 1) while South Boston's/Seaport's white population is 66% (Map 2)

crime types. The user can click on crime types under each map and the map as well as the legend will update accordingly. We omit to show tooltips on the map because information such as police district names are distracting to the main purpose. The default view of the crime type choropleths is shown in Figure 2.

4.3 Demographics Visualizations

The scrolly-board then introduces the demographics dataset by highlighting key race and wealth statistics. The first visualization is a pie chart of the overall race distribution of Boston's residents. This emphasizes the fact that there is a large minority population in Boston. Once again, interactions such as tooltips are omitted to focus on the message. The next visualization is two choropleths of residents' races in police districts. Note that these maps are similar to the crime types choropleths, making them easier to understand. Within each choropleth, we allow the user to explore how segregated Boston's police districts are. By transitioning from the pie chart to these choropleths, we also emphasize that the race distribution of each district is very different from Boston's average. Note that we didn't display the overall distribution of a race across every district because that would omit the number of residents within a district. Instead, we display the portion of total residents a given race within a district makes up such that a valid comparison between districts can be made. We include two choropleths side-by-side to allow users to compare trends between races. The user can click on a race under each map and the map as well as the legend will update

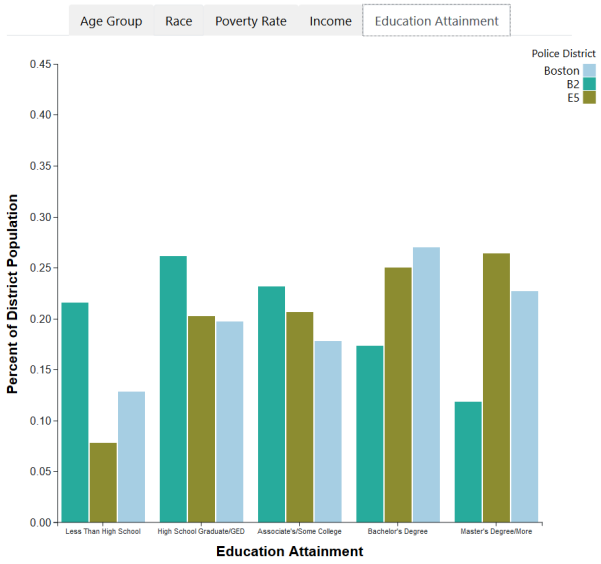

Figure 4: The overall education attainment for police districts E5 and B2. Note district E5's higher overall education attainment over both district B2's and Boston's education attainments.

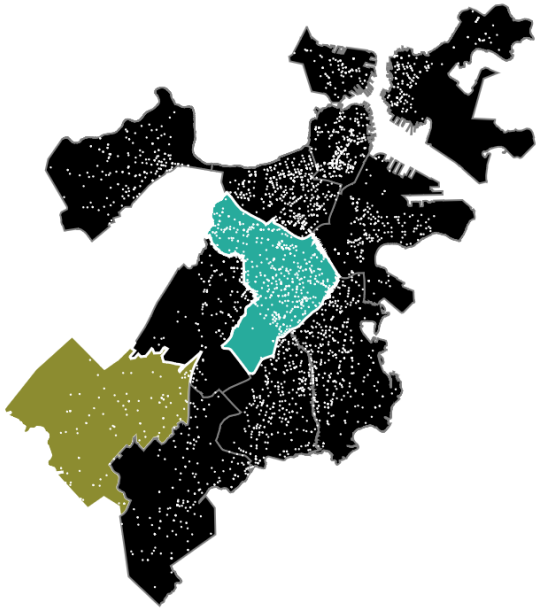

Figure 5: The accompanying point visualization to Figure 4

accordingly. The default view of the race choropleths is shown in Figure 3. The last visualization is a table of the 3 richest and poorest neighborhoods in Boston. This highlights that there is a massive imbalance in family income across neighborhoods, implying the resource inequity that exists within some of them. We included race daa directly below the incomes to reveal that minority families are disproportionately on the lower end of family incomes. Rather than encoding these numbers, we chose to use their raw form since their differences are easily interpretable.

4.4 Combining Crime and Demographics

The martini glass opens with our final visualization where we combine the crime and demographics data.

On the left side is a point visualization of crime incidents on a map. On the map, we plot an individual point for each crime incident. We didn't use a continuous encoding such as a heat map because each crime is discrete and distinct. We also didn't generalize the points into bins to emphasize how each crime incident impacts people. The user can hover over each police district on the map to view a tooltip with more information such as the neighborhoods included inside. On the top of the map, there are filters for different types of crimes. When a user hovers over a crime type filter, it highlights the crimes with that type on the map. If a user clicks on a crime type filter, only the crime points with that crime type are shown. Note that users can select multiple filters at once, and once they do so, they can still hover over each to highlight the specific crime types.

On the right side is a bar chart displaying the demographics data. The user can toggle between the different bar charts to explore a single type of demographics data at a time and view it alongside the map as a reference. The bar charts encode the magnitude of each group's percent population such as the percent population of the Hispanic group. We chose to use percent population rather than raw population to highlight distributions within a district. The default state of the bar chart includes only Boston's overall distribution, but the user can add more police districts. We used a group bar chart to allow for easier comparisons.

When a user clicks on a district on the map, that district's information will be added to the demographics charts. The user can select multiple districts at a time to compare the data between them. For example, Figure 4 shows the overall education attainment for police districts E5 and B2. Figure 5 shows the accompanying point visualization.

5 Discussion

From our work, the audience has learned the correlation between crime and demographics of each Boston neighborhood. The scrolly-board begins by revealing the relationship between Boston's neighborhoods and districts. Then, the scrolly-board effectively introduces each dataset, summarizing distributions such as overall race. For each dataset, the scrolly-board also instructs the user to interact with the data, allowing them to see how crime and demographics distributions within a district are distinct. Once the user is familiarized with each dataset, we reveal the main visualization where crime and demographics data is combined. With this opening of the martini glass, the user can explore certain correlation trends between our datasets. For example, neighborhoods that are low-income may be more susceptible to some crime types. We emphasize to the user that they are free to explore correlations, but causations between crime and demographics are not revealed due to its complexity. Throughout the scrolly-board, we also are transparent in describing our data transformations.

6 Future Work

We only focused on the trends within a specific time frame in our crime data, so future work can analyze if these trends are universal across seasons and/or years. We also used demographics data from a single year, so it would be interesting to visualize the temporal changes in demographics data. Finally, it would be important to research how certain issues such as over-policing may affect certain neighborhoods, and compare the extent of these issues with neighborhood demographics.

References

[1] *Crimes in Boston*.
[2] Analyze Boston. *Boston Demographics*.

[3] Analyze Boston. *Crime Incident Report*.
[4] A. J. Baranauskas. *Exploring the Social Construction of Crime by Neighborhood: News Coverage of Crime in Boston*. Sociological Focus, March 2020. doi: 10.1080/00380237.2020.1730280
[5] A. P. Cardarelli. *CRIME IN BOSTON - AN ANALYSIS OF SERIOUS CRIME PATTERNS WITHIN 81 NEIGHBORHOODS*. Department of Justice.
[6] C. Elton. *How Has Boston Gotten Away with Being Segregated for So Long?* Boston Magazine, December 2020.