CS573 Final Project Process Book
Visualizing the Massachusetts Affordable Housing Crisis
Samantha Crepeau

I. Overview and Motivation

According to the National Low Income Housing Coalition, the United States needs 6.8 million more affordable housing units for extremely low income (ELI) families. Affordable housing is housing that consumes less than 30% of a household's total income. There is no state or county where a full-time minimum wage employee can afford to rent a two-bedroom apartment [1]. For those who are faced with high housing costs relative to their income, other stresses such as food security, health care, retirement also cause increased financial strain [2]. These facts illustrate the severity of the affordable housing crisis in the US.

It is important that people are aware of the facts about the affordable housing crisis. Those who are not affected directly by it may benefit from understanding the situation that a significant number of their fellow citizens face, and those who are affected may benefit from knowing the facts about the situation they are burdened with. The aim of this project is to educate people about the affordable housing crisis through interactive data visualizations, with the goal for this knowledge to empower others through their newfound understanding of the situation.

We will develop the data visualization dashboard with a focus on certain aspects of the affordable housing crisis. Visualizations will be accompanied by text explaining their context and how to interact with them if applicable. The text and visualizations together will tell a story that summarizes important aspects of the affordable housing crisis, such as the ways in which it has worsened over time and who is most affected by it.

II. Related Work

A related data visualization dashboard that served as inspiration for the proposed design of our own dashboard is the Housing Data Dashboard by the Government Finance Officers Association (GFOA) [3]. The dashboard features several interactive data visualizations accompanied by information about the housing market and construction. This format presents information concisely and effectively and is organized well. Screenshots of the dashboard can be found in Figure 1.

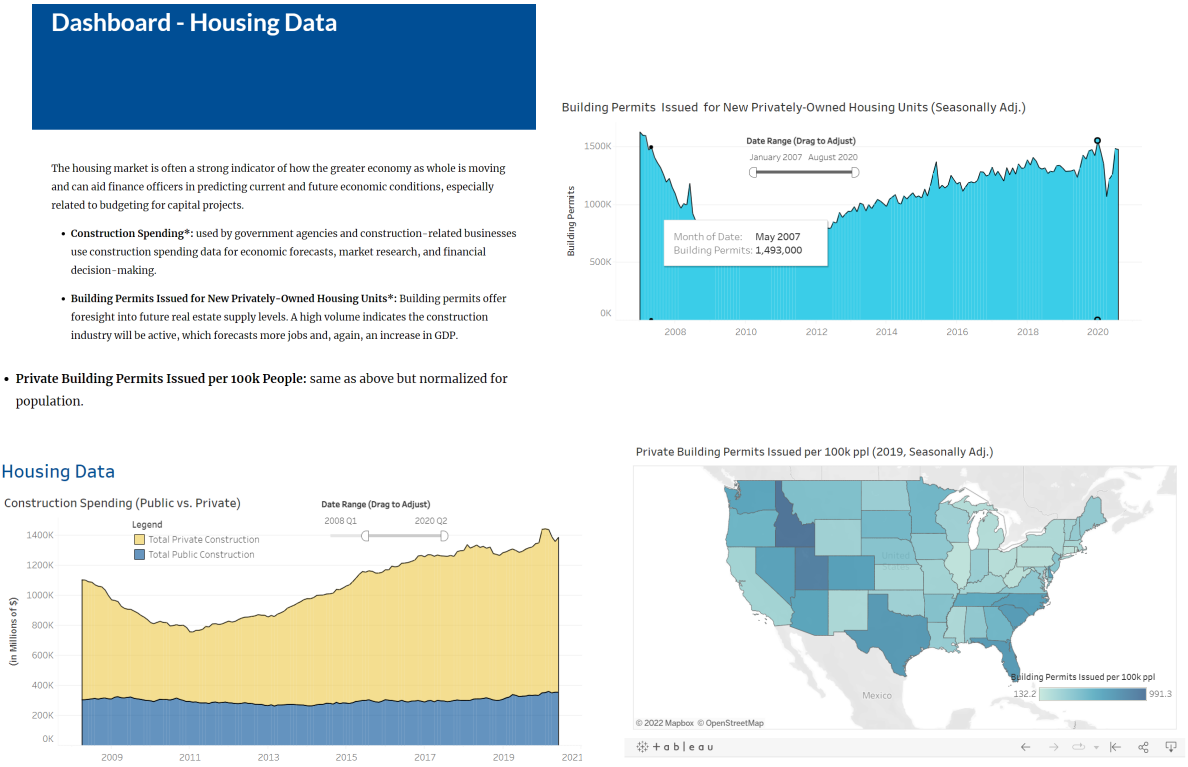

Figure 1: Screenshots of the GFOA Housing Data Dashboard. From https://www.gfoa.org/dashboard---housing-data

Another website that served as inspiration for our proposed dashboard is Massachusetts Housing Data by Towncharts [4]. The site features a series of visualizations related to different subtopics such as home supply and home cost. Each figure is accompanied by text explaining what the chart depicts and important points about the data. Figure 2 depicts a screenshot of one of these charts.

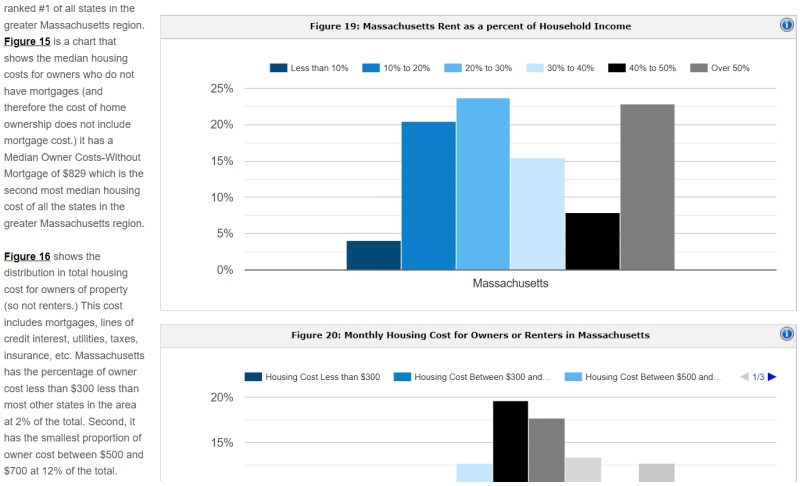

Figure 2: Example chart by Towncharts. From https://www.towncharts.com/Massachusetts/Massachusetts-state-Housing-data.html

The strongest source of inspiration for our dashboard was a Tableau visualization from the article "The Growing Shortage of Affordable Housing for the Extremely Low Income in Massachusetts" by Nicholas Chiumenti [5], which summarizes the issue of the affordable housing shortage in MA with a focus on the impact on extremely low income households. A couple paragraphs of text describe the data and its key points; they are followed by the visualization found in Figure 3 below. A drop down menu allows the user to select different measures from the data. By mousing over each town, the user is able to view the name of the town and the data specific to that town. The source also contains a downloadable file containing the data which we used to make a similar visualization.

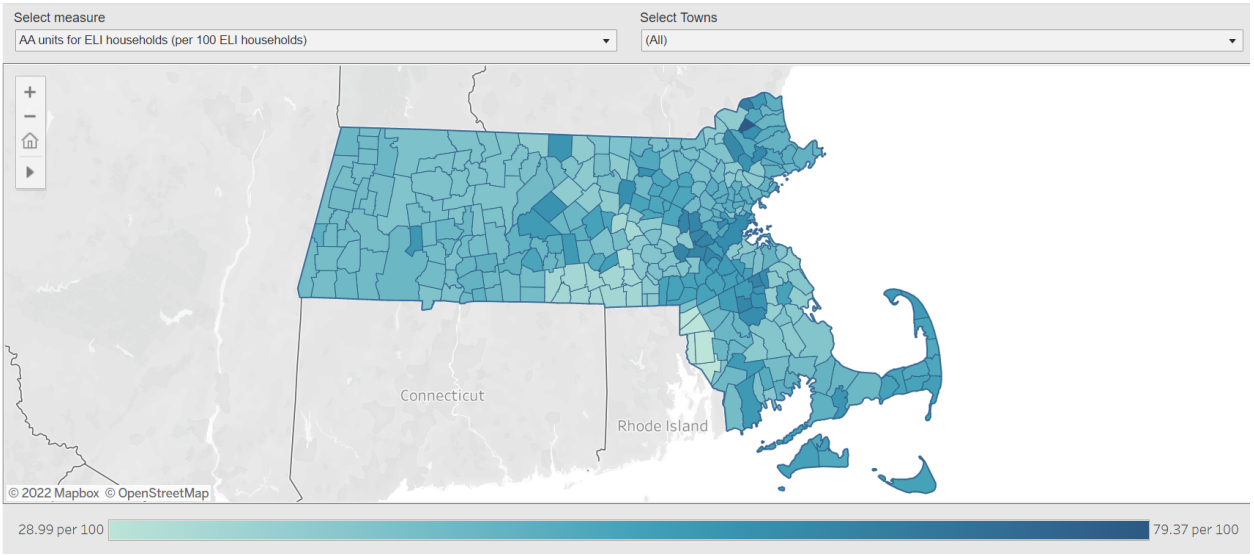

Figure 3: Interactive map of MA featuring affordable housing data. From https://www.bostonfed.org/publications/new-england-public-policy-center-policy-report/2019/growing-shortage-affordable-housing-extremely-low-income-massachusetts.aspx

III. Questions

The main questions we wanted to answer when starting this project were:

1. What are the most significant aspects of the US affordable housing crisis?
2. What data can be used to best characterize the US affordable housing crisis concisely and effectively?
3. How should this data be represented in visualizations?

These questions remained nearly unchanged over the course of the project, except that the scope was narrowed down to the state of Massachusetts, rather than the entire United States, in order to

keep the project focused and manageable. Eventually, we began to consider another important question:

4. How do we balance the amount of text explanation with the number of visualizations?

This question was prompted by the growing walls of text that began to take up more of our dashboard as it developed. At some point, too much text weakens the impactfulness of the visualizations. It was useful to consider the dashboard as more of an interactive infographic rather than a dashboard, highlighting key points that we want the audience to come away with and keeping text relatively minimal. Thus, the answer seemed to be that we should choose a few key points and definitions related to the data and allow the visualizations to do most of the explaining.

IV. Data

Our final three data sources for the three visualizations in our dashboard were:

1. https://fred.stlouisfed.org/graph/?id=MEDLISPRIMA,MEHOINUSMAA672N [6]
2. https://www.bostonfed.org/publications/new-england-public-policy-center-policy-report/2019/growing-shortage-affordable-housing-extremely-low-income-massachusetts.aspx [5]
3. https://www.americashealthrankings.org/explore/annual/measure/homeownership_disparity/state/MA [7]

The data did not require much in the way of cleaning. Each source allows the user to download the data as a CSV or ASPX file. The file that required any significant cleanup was Source 2, which was downloadable as an ASPX file. The file was converted into a CSV file, then manipulated using the Pandas data analysis library for Python. The file contained 20 columns, which was cut down to 5 columns featuring only the data of interest: number of ELI households that are rent burdened, number of ELI renter households, and number of renter households, in addition to the county name and the city/town name. These values were used to calculate the percent of renter households that are ELI and the percent of ELI households that are cost burdened.

Eventually, we decided to focus on the percent of renter households that are ELI to keep the project more manageable. This set of values, along with each corresponding town name, was merged with the CSV file containing the town names and data used for GeoJSON. This final set of data could then be used to create a map of all Massachusetts towns with each town colored based on the percent of renter households that are ELI. Screenshots summarizing the cleanup process for this source are shown below in Figure 4.

	COUNTY NAME	CITY/TOWN NAME	NO. OF ELI RENT BURDENED (30%)	NO. OF RENTER HOUSEHOLDS	NO. OF ELI HOUSEHOLDS
0	Plymouth County	Abington	228.8	2035.8	367.4
1	Middlesex County	Acton	304.6	1988.8	446.1
2	Bristol County	Acushnet	61.3	484.2	88.8
3	Berkshire County	Adams	336.1	1284.2	419.5
4	Hampden County	Agawam	842.2	3034.1	1036.5
5	Berkshire County	Alford/Egremont/Mount Washington	31.9	121.9	39.8
6	Essex County	Amesbury	328.3	2137.3	557.3
7	Hampshire County	Amherst	985.8	5032.5	1244.4
8	Essex County	Andover	352.6	2471.7	410.8
9	Dukes County	Aquinnah/Chilmark/Gosnold/West Tisbury	51.1	296.6	69.3

	COUNTY NAME	CITY/TOWN NAME	PERCENT ELI	PERCENT RENT BURDENED ELI
0	Plymouth County	Abington	0.180470	0.622754
1	Middlesex County	Acton	0.224306	0.682807
2	Bristol County	Acushnet	0.183395	0.690315
3	Berkshire County	Adams	0.326663	0.801192
4	Hampden County	Agawam	0.341617	0.812542
5	Berkshire County	Alford/Egremont/Mount Washington	0.326497	0.801508
6	Essex County	Amesbury	0.260750	0.589090
7	Hampshire County	Amherst	0.247273	0.792189
8	Essex County	Andover	0.166201	0.858325
9	Dukes County	Aquinnah/Chilmark/Gosnold/West Tisbury	0.233648	0.737374

	id	town	fips	percent
0	1	ABINGTON	25023	0.180470
1	2	ACTON	25017	0.224306
2	3	ACUSHNET	25005	0.183395
4	13	ASHFIELD	25011	0.247158
..	...	...	...	...
348	347	WOBURN	25017	0.238778
349	348	WORCESTER	25027	0.343722
350	349	WORTHINGTON	25015	0.247014
351	350	WRENTHAM	25021	0.232012
352	351	YARMOUTH	25001	0.233746

Figure 4: Summary of the cleanup process for ELI data, including the removal of extraneous data and merging of the desired data with geographic data.

V. Exploratory Data Analysis

One visualization that we thought would be important to create was one comparing income with a measure like cost of rent or housing prices over time. This visualization would illustrate how wages have relatively stagnated while the cost of living has increased, with a focus on housing. Initially, we looked at median income compared to median property value. As shown in Figure 5, this visualization was not particularly illuminating. We decided to shift the goal of the visualization to focus on what the housing market looks like for the average MA family household looking to become homeowners. Therefore, the comparison made in our visualization became median household income and median house listing price, the result of which is shown in Figure 6. This visualization better shows how houses are growing more unattainable for many. All exploratory data analysis was done in Python using Matplotlib.

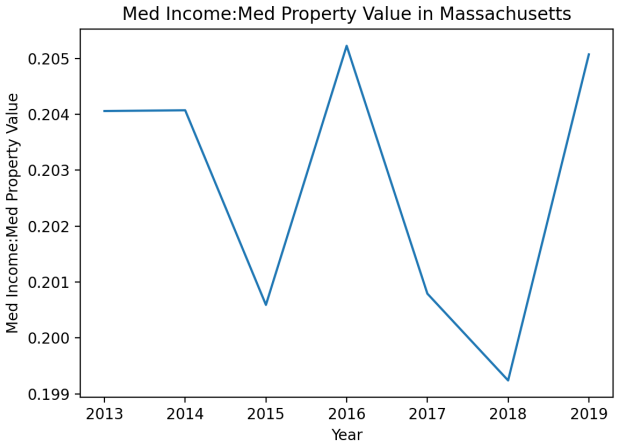

Figure 5: Ratio of median income and median property value. Data from https://datausa.io/profile/geo/massachusetts.

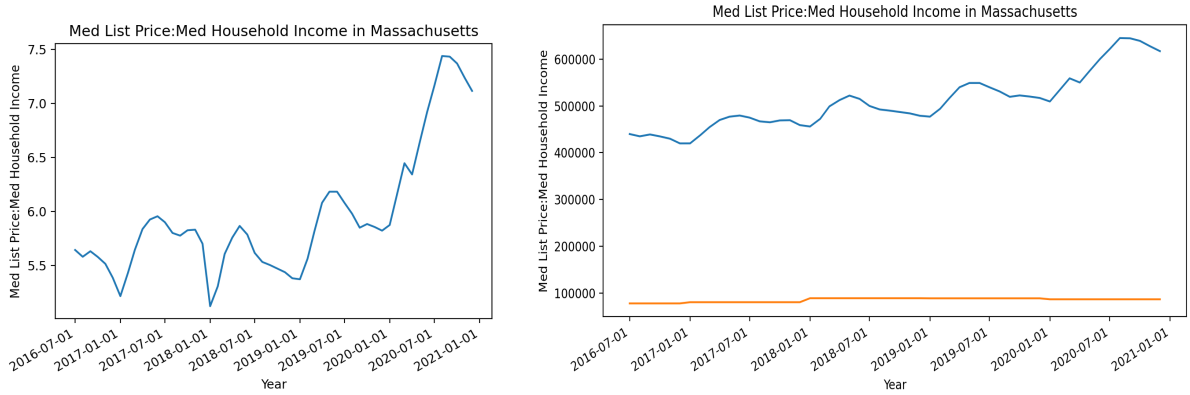

Figure 6: On the left, we have the ratio of median list price and median household income. On the right, we have median list price and median household income plotted separately. Data from https://fred.stlouisfed.org/graph/?id=MEDLISPRIMA,MEHOINUSMAA672N.

VI. Design Evolution

The first draft of the dashboard design is depicted below in Figure 7. Initially, there were four planned visualizations: average income vs. average house price/cost of rent over time, housing cost burden in MA, racial disparity in housing cost burden, and rates of homeownership over time. As previously described, the first visualization evolved to compare median household income and median house listing price instead. The other three proposed visualizations evolved into a map of extremely low income renter households and a chart depicting the racial disparity in rates of homeownership. This decision was made both for the sake of time and due to our inability to find data specifically on the racial disparity in housing cost burden. The proposed text was changed accordingly in the final product.

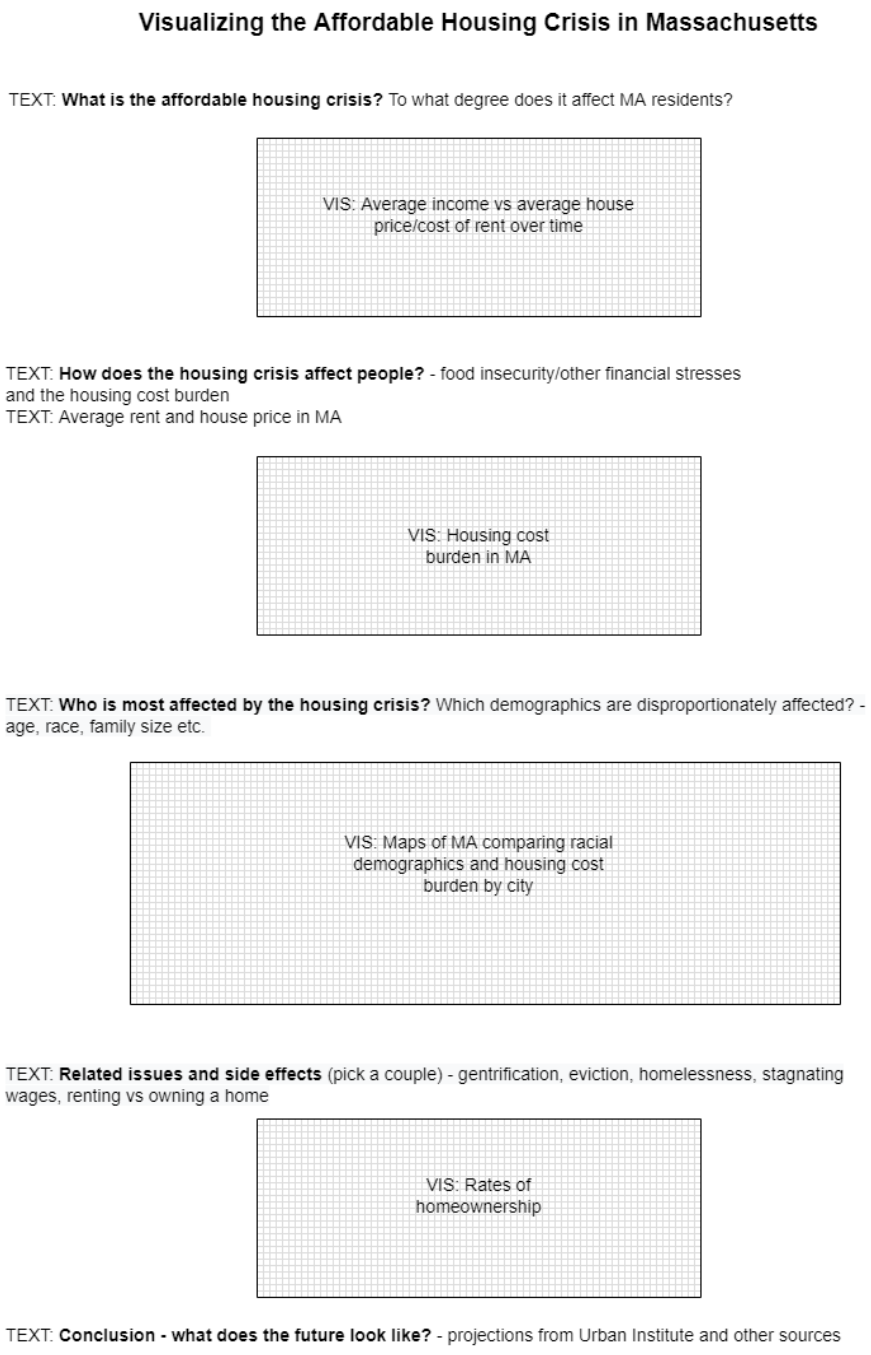

Figure 7: Original proposed layout of the dashboard.

Certain perceptual and design principles heavily influenced our decisions concerning the visualizations and the format of the dashboard itself. These include:

- Importance of interactivity
- Importance of color and other aesthetics
- Choosing the best visualization given a data source and a key point to emphasize

Interactivity and aesthetic were two important driving forces behind the choices made in designing the visualizations. Interactivity is an important element in order to improve the effectiveness and clarity of data visualizations. Therefore, we ensured that at least two visualizations would make use of interactive features. Aesthetics were also a significant part of the design process. We maintained a color palette consisting of mostly blue to give the website a professional look. We also took great care to manipulate margin sizes to prevent the page from becoming cluttered or hard to read. Poor aesthetics can distract the user from the content and reduce effectiveness of the visualizations.

To decide upon the visualization types, we examined each potential data source and considered the goal of each visualization. The goal for each visualization was to emphasize some key aspect of the affordable housing crisis, specifically what it is, how it affects people in Massachusetts, and who it affects the most. Given each of these goals, we considered which data from our data sources would best represent them. A simple line graph was useful for comparing trends over time, while a map is the obvious choice for data that varies with geography, and a bar chart was most effective to compare a value across groups.

VII. Implementation

The first visualization implemented compares median listing price with median household income over time in Massachusetts. The visualization is shown below in Figure 8.

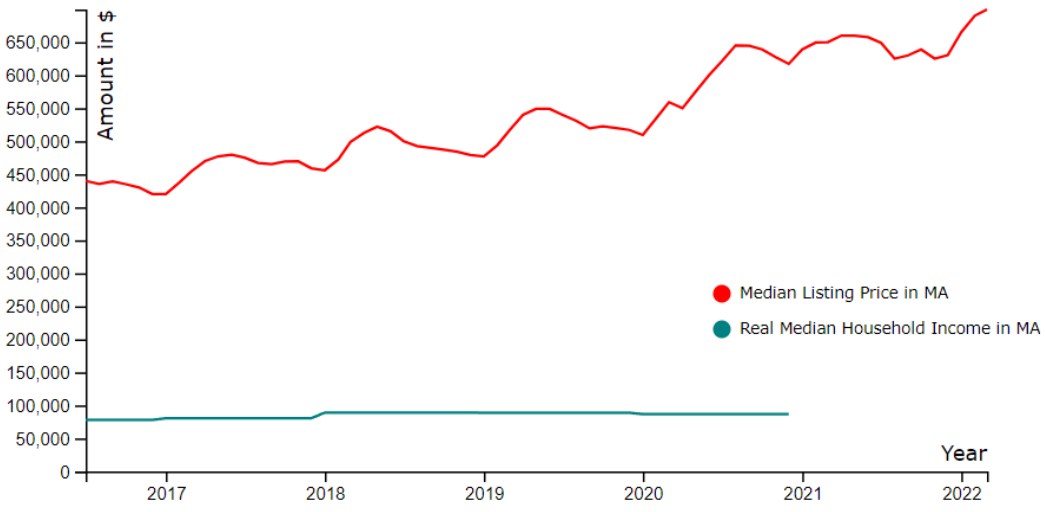

Figure 8: The first visualization on the website. A line graph comparing median listing price and median household income in MA over time. Data from https://fred.stlouisfed.org/graph/?id=MEDLISPRIMA,MEHOINUSMAA672N

The second visualization, shown in Figure 9, is an interactive map of Massachusetts. Each town or city is colored based on the percent of renter households that are extremely low income. By mousing over each town, the user can view the name of the town and the corresponding percentage of renters that are ELI.

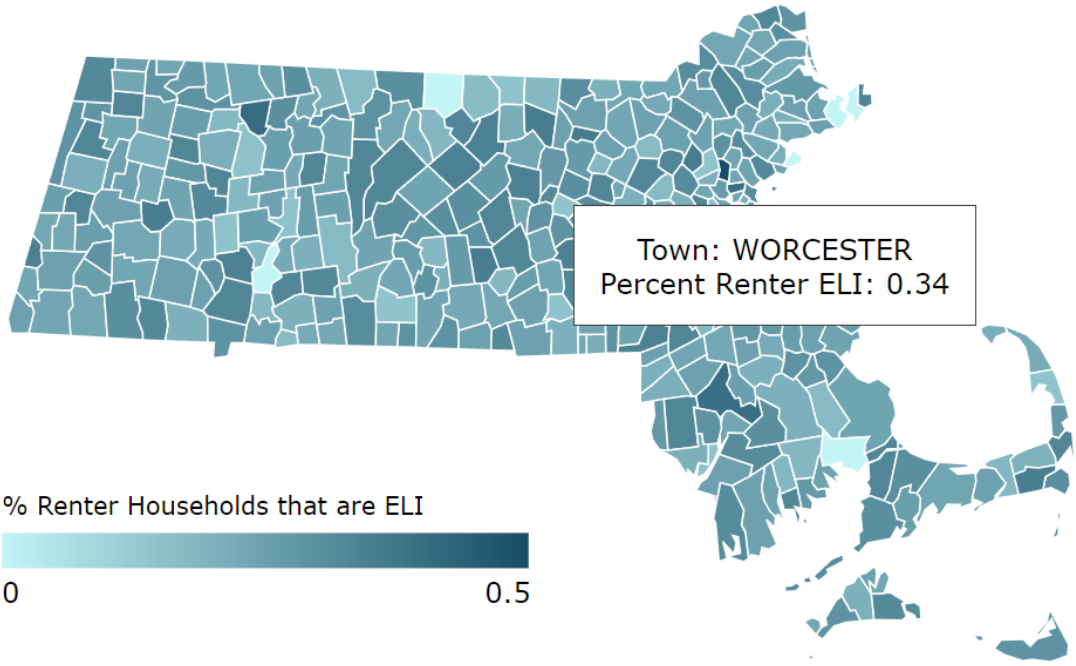

Figure 9: The second visualization on the website. A map colored based on percent renter ELI with a tooltip that presents each town name and corresponding percent renter ELI. Data from https://www.bostonfed.org/publications/new-england-public-policy-center-policy-report/2019/growing-shortage-affordable-housing-extremely-low-income-massachusetts.asp

The final visualization on the website, shown in Figure 10, is a bar chart representing the racial disparity in homeownership in Massachusetts as of 2019. By mousing over each bar, the user can see what percent of each group are homeowners.

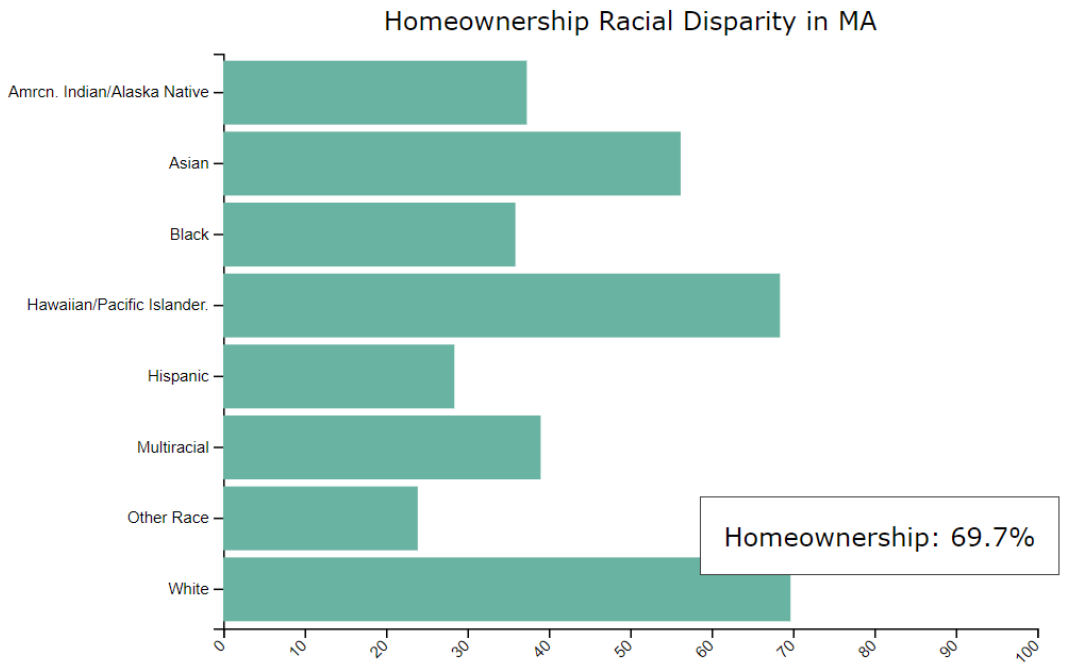

Figure 10: Third visualization on the website. A bar chart representing the racial disparity in homeownership, featuring a tooltip that shows the specific percentage. Data from https://www.americashealthrankings.org/explore/annual/measure/homeownership_disparity/state/MA

VIII. Evaluation

Throughout this project, we aimed to answer the following questions:

1. What are the most significant aspects of the affordable housing crisis in MA?
2. What data can be used to best characterize the affordable housing crisis in MA concisely and effectively?
3. How should this data be represented in visualizations?

We were able to answer all of these questions during the project's development. Through research, we determined that the most significant aspects of the affordable housing crisis included its effects on those looking to become homeowners, how its severity varies throughout the state, and how different groups are disproportionately affected. Through research, we were

also able to find relevant housing data that could be used to represent empirically these aspects of the housing crisis. Through data analysis and exploration, we were able to determine the best methods to represent this data effectively in our visualizations.

Only the map visualization does not function as it should. The tooltip shows the incorrect town and percentage when each town is moused over, indicating that there is an issue with how each GeoJSON town is assigned values from the CSV file containing the town name and percentage. As for other potential improvements to the visualizations, adding an interactive element to the first visualization would enhance it. It would be useful to be able to mouse over the graph and view the year, median household income, and median listing price as the user moves along the x-axis. Additionally, adding a dropdown menu to the map visualization that would allow the user to also view the percentage of renter households that are rent burdened would improve the visualization.

IX. References

[1] NLIHC. (n.d.). *The Problem*. National Low Income Housing Coalition. Retrieved May 1, 2022, from https://nlihc.org/explore-issues/why-we-care/problem

[2] Corradino, S. (2022, April 14). *There is no easy fix for the Affordable Housing Crisis*. Forbes. Retrieved May 1, 2022, from https://www.forbes.com/sites/forbesnonprofitcouncil/2021/09/07/there-is-no-easy-fix-for-the-affordable-housing-crisis/?sh=577b0bc773b7

[3] Government Finance Officers Association of the United States and Canada. (n.d.). Dashboard - Housing Data. Retrieved May 1, 2022, from https://www.gfoa.org/dashboard---housing-data

[4] Towncharts. (n.d.). *Massachusetts Housing Data*. Towncharts. Retrieved May 1, 2022, from https://www.towncharts.com/Massachusetts/Massachusetts-state-Housing-data.html

[5] Chiumenti, N. (2019, April 3). *The growing shortage of affordable housing for the extremely low income in Massachusetts*. Federal Reserve Bank of Boston. Retrieved May 1, 2022, from https://www.bostonfed.org/publications/new-england-public-policy-center-policy-report/2019/growing-shortage-affordable-housing-extremely-low-income-massachusetts.aspx

[6] Federal Reserve Bank of St. Louis. (n.d.). *FRED Graph*. FRED. Retrieved May 1, 2022, from https://fred.stlouisfed.org/graph/?id=MEDLISPRIMA%2CMEHOINUSMAA672N

[7] United Health Foundation. (n.d.). *Explore homeownership racial disparity in Massachusetts: 2021 annual report*. America's Health Rankings. Retrieved May 1, 2022, from https://www.americashealthrankings.org/explore/annual/measure/homeownership_disparity/state/MA