

CityView+: Enhanced Urban Data Visualization



by

Frank Campanelli

Zhung Keung Moo Feng

Tarek Safa

CityView+: Enhanced Urban Data Visualization

A Major Qualifying Project
submitted to the Faculty of
WORCESTER POLYTECHNIC INSTITUTE
in partial fulfilment of the requirements for the
degree of Bachelor of Science

by
Frank Campanelli
Zhung Keung Moo Feng
Tarek Safa

Date:
02 28 2019

Report Submitted to:
Professor Yanhua Li
Worcester Polytechnic Institute

This report represents work of WPI undergraduate students submitted to the faculty as evidence of a degree requirement. WPI routinely publishes these reports on its web site without editorial or peer review. For more information about the projects program at WPI, see <http://www.wpi.edu/Academics/Projects>.

ABSTRACT

The goal of this project is to deliver on visualizations that tell a story which is clear and understandable to the masses. To accomplish this goal, we will complete two main objectives. First, we will identify categories and subjects we want to tackle based on the datasets we can utilize. Afterwards, we will create a website which utilizes these up to date datasets and clear, understandable data visualizations which will each describe a phenomenon in the city of Boston.

TABLE OF CONTENTS

ABSTRACT.....	3
TABLE OF CONTENTS.....	4
TABLE OF FIGURES.....	5
Chapter 1: Introduction.....	7
Chapter 2: Background.....	8
2.1 Prior Work.....	8
2.1.1 Static Data Visualization.....	8
2.1.2 Time Series Visualization.....	8
2.1.3 Dynamic Visualization.....	9
2.2 Related Visualizations.....	9
2.3 Related Mobile Visualizations.....	12
2.4 Technologies.....	13
Chapter 3: Methodology.....	14
3.1 Goals.....	14
3.2 Design.....	14
3.3 System Overview.....	15
3.3.1 Client.....	15
3.3.2 System Infrastructure and Functionalities.....	15
3.3.3 GitHub.....	16
3.4 Categories.....	17
3.4.1 Demographics.....	17
Dataset.....	17
3.3.2 Events.....	19
Emergency Snow Parking.....	20

3.3.3 Livability.....	21
CityScore.....	21
3.5 Implementation	24
3.5.1 Leaflet	25
3.5.2 D3.js.....	25
CHAPTER 4: Results and Analysis.....	26
4.1 Usage.....	26
4.2 Results.....	26
4.2.1 Website	26
4.2.2 Livability.....	26
4.2.3 Demographics	26
4.2.4 Events.....	27
CHAPTER 5: Conclusion and Recommendations	28
5.1 Recommendations.....	28
Server Backend	28
Mobile App.....	28
Data Availability.....	28
REFERENCES	30

TABLE OF FIGURES

Figure 1: Climate Ready Boston Map Example	10
Figure 2: Minimal theming allows for patterns to be easily recognized.....	11
Figure 3: French team's unique visualization	12
Figure 4: An overview of the website's content.....	15
Figure 5: Overview of the structure of the website.....	16

Figure 6: Full-sized sunburst diagram	18
Figure 7: Screenshot showing overview-detail sunburst	19
Figure 8: Crime report marker clusters	20
Figure 9: Emergency snow parking	21
Figure 10: Free Wi-Fi access points	23
Figure 11: Choropleth map of tree density by neighborhood	24

Chapter 1: Introduction

Data visualizations are a relatively new medium of expressing data and has become a desirable means of communicating trends and phenomenon. When effectively implemented, they can communicate insightful information to the average person very quickly, which makes them very useful today. They help people grasp difficult concepts and/or identify patterns by converting a large amount of data into a graphical or pictorial representation.

The previous iteration of CityView was created to analyze large sets of data and possibly assist with urban development. What we wanted to accomplish with CityView+, was to create a system that presented data that was useful and interesting to anyone. For our data visualization to be useful, we needed to be clear in communicating what the data was, and how one might use the data visualization.

CityView+ is a web-based system designed to present urban data to a regular citizen in a simple and understandable way. The website provides a varies set of tools to understand different datasets pertaining to Boston via data visualizations. We utilized popular frameworks for visualizations, mainly d3 and Leaflet. With the high customizability of these frameworks, there is plenty of room for adaptation, and future updates.

Chapter 2: Background

2.1 Prior Work

Before CityView+, a previous MQP team developed CityView, henceforth referred to as CityView, an Urban Data Visualization application which visualized static, dynamic, and time series data for picturing an urban environment. The CityView application was a mainly front-ended application developed with HTML and JavaScript. The datasets they used to visualize their data were static and did not reflect recent data. Their website served as a crux for features we wanted in our application, mainly visualizing data on a map in a manner that would be understood by most people.

Most of the features the CityView team implemented were necessary in our application, however, we wanted to visualize data that did things differently: first, we wanted to visualize data that would tell a story that was more meaningful to the regular person; second, we wanted our data visualizations to reflect the most recent information possible; and finally, we wanted to focus on data from Boston and cater the app for people interested in Bostonian data differing from the CityView app which visualized data from New York and Boston.

2.1.1 Static Data Visualization

The static data visualization displayed the locations of streetlights and hubway bike share rack locations. In this data visualization, CityView used MapBox and Leaflet to visualize data that was static and fixed. The data was taken from the Analyze Boston website, a website we are currently utilizing for data as well. The purpose for this visualization as described by the team was to help bikers take paths that were safer for them during the night-time.

2.1.2 Time Series Visualization

In the Time Series Data Visualization CityView aimed to show change in an area over a period. The subject of their time series data was the recycling schedules in different areas in Boston. Again, CityView used Leaflet and MapBox to visualize these schedules over the period of a week. They again obtained their data from the Analyze Boston website. The purpose of the Time Series Visualization as described by CityView was to assist recycling companies in

determining routes for their recycling schedules, since they would be able to recognize patterns in the visualization that might give them insights they previously did not have.

2.1.3 Dynamic Visualization

In the Dynamic Data Visualization CityView aimed to show how objects might move and change over time. They did this with New York's green taxi pickup and drop-off schedules. This visualization was developed in Carto, and they obtained their data from the NYC taxi and limousine commission. The visualization demonstrated how taxi drivers might increase their travelling efficiencies

2.2 Related Visualizations

To gain some insight into how we wanted our final visualization to look, we did research into other data visualizations. One source that stood out to us was Imagine Boston's 2030 Metrics Dashboard. Their basic framework of having each dataset on the side in a hideable menu bar gives quick access to each data set, and a detailed description about each dataset is displayed front and center after a dataset is clicked. The visualizations are visible after scrolling down the page, which we decided to replicate. This layout ensures that the reader is exposed to the description of the visualization before seeing the data displayed, giving them a chance to learn more about the data so the visualization is less confusing from the start.

Another source that utilized a novel visualization concept is the Boston Climate Explorer. It utilizes multiple toggleable shaded overlays to show many data all at once.

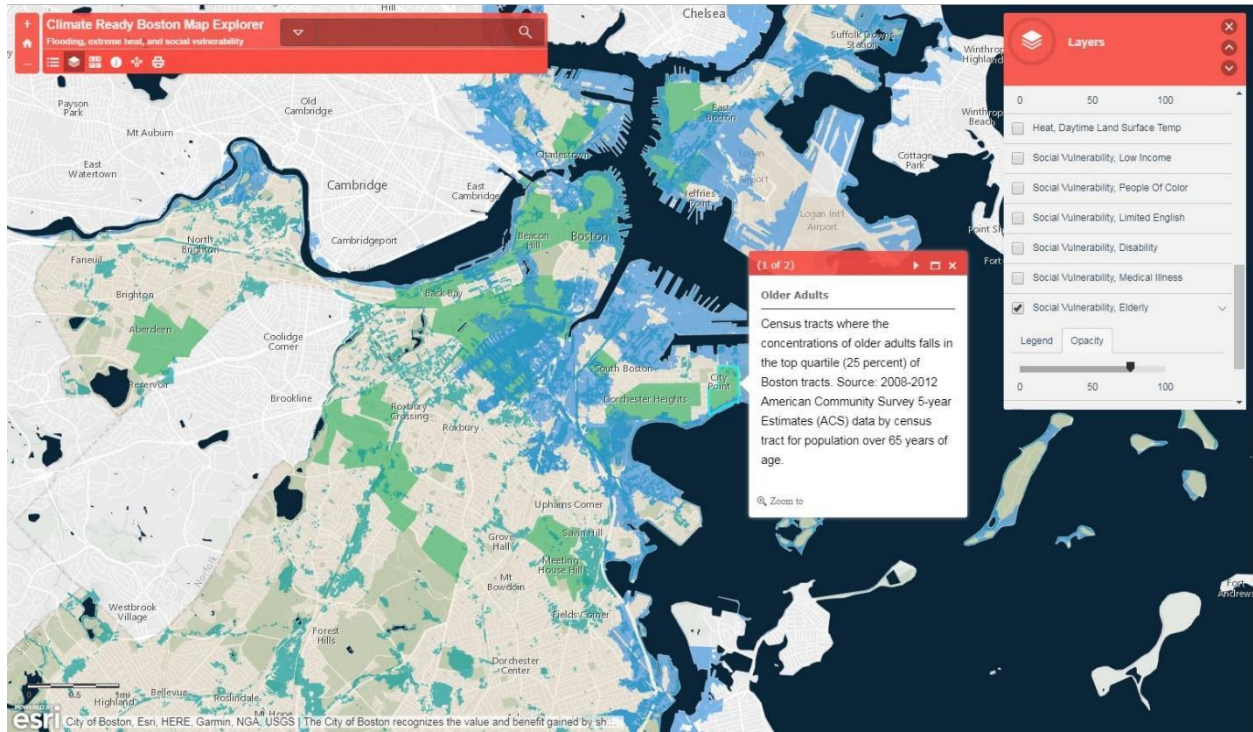


Figure 1: Climate Ready Boston Map Example

When the user clicks on a shaded section, it will highlight all related overlays and a small window appears giving more detailed information about the overlaid information. Giving users the ability to toggle the overlay visibility allows the user to consume information at their own pace without getting overwhelmed.

The visualization “Access to Taxicabs for Unbanked Households” demonstrates how simple themes and minimal extra information can be leveraged for conveying information effectively.

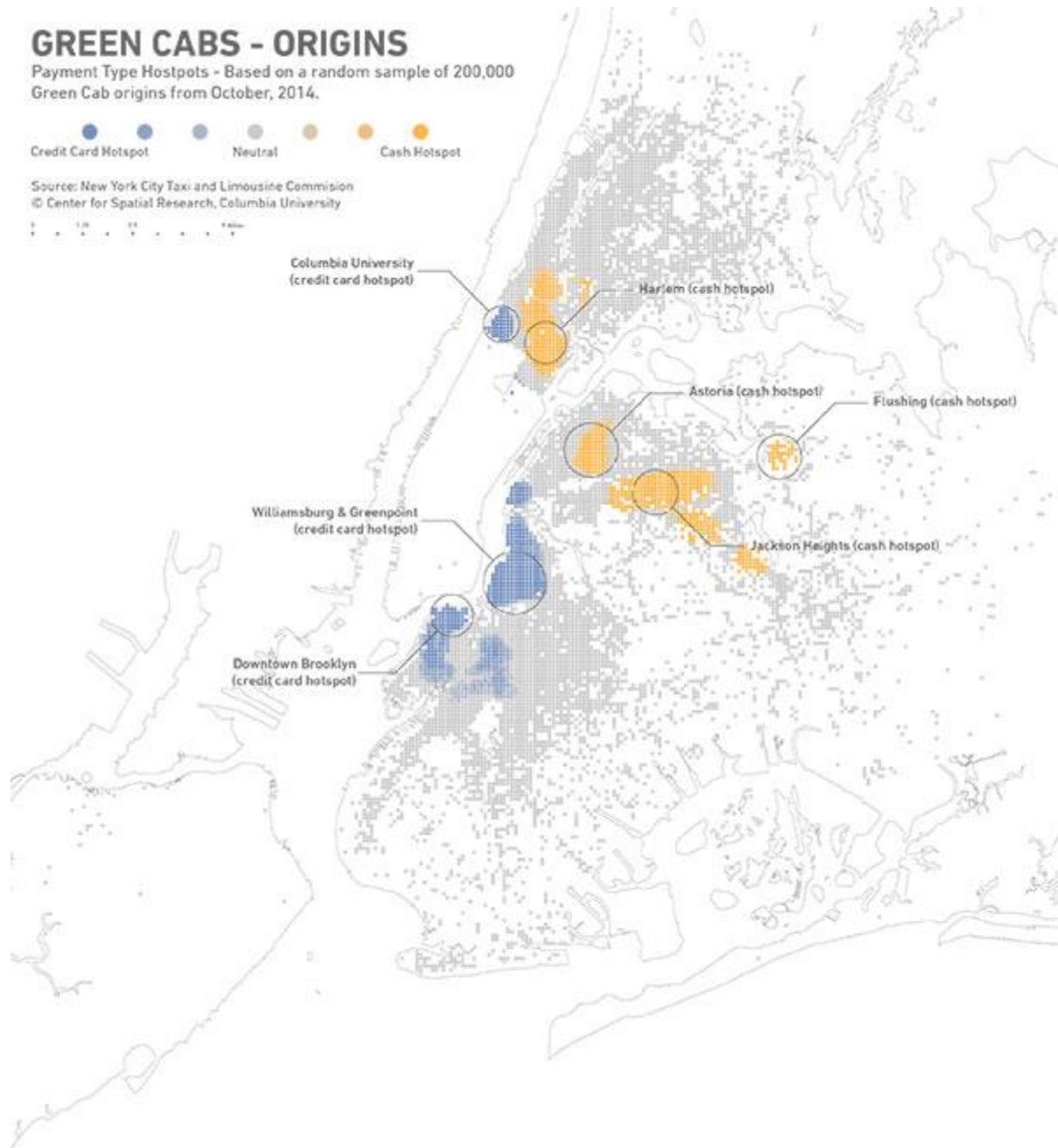


Figure 2: Minimal theming allows for patterns to be easily recognized

The use of a gray and white background for the map contrasts with the blue and yellow to focus the reader's attention on the origins of the taxicabs. This makes it easier to notice patterns in the data near points of interest, like the airport or prominent neighborhoods.

2.3 Related Mobile Visualizations

Some visualizations went the extra mile and included interactive input from users. A French project built an “educational simulator to help understand the issues around energy flexibility” (Dataveyes) and included input from the audience on their mobile devices to augment the presentation of the project. Along with the inclusion of mobile devices to aid in data visualization, the French team took a unique approach to visualizing the data.

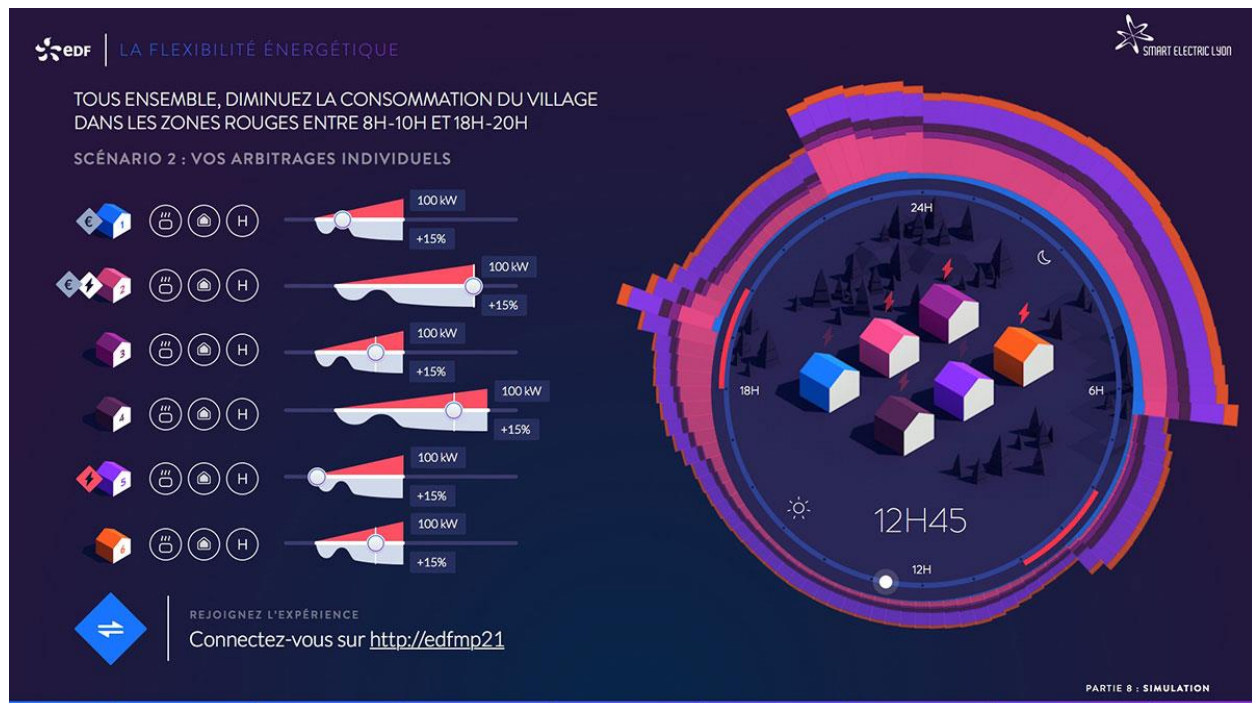


Figure 3: French team's unique visualization

The visualization would dynamically update the infographic as more data came in from the audience’s mobile devices after users submitted their data into a form on the team’s website. This kind of mobile integration assured better audience attention to the presentation, as the audience members’ own data was being visualized next to data from their peers. For our project, we are looking to make a mobile version of our visualizations, so they are more accessible.

2.4 Technologies

We have researched some technologies that will help us build the CityView+ website and visualize the datasets. When it comes to the website styling and design, we decided to use Bulma, a CSS framework based on Flexbox. We chose Bulma because it is responsive, modular, and open source. This makes it easy for us to develop the website with mobile users in mind. Additionally, modularity means that we can import whatever modules we need from Bulma. There are some drawbacks to using Bulma including the fact that it is purely a CSS framework. As opposed to Bootstrap that is an HTML, CSS, and JS framework, Bulma lacks when in terms of making a website intractable because it lacks jQuery plugins. Furthermore, Bootstrap has a larger community. We chose Bulma because of its low overhead and ease of use.

2.4.1 Leaflet

Leaflet is an open-source JavaScript library used to make mobile-friendly interactive maps. It can interface multiple map providers such as Mapbox to display a base map layer. Furthermore, Leaflet extends many plugins that improve upon the base interactions Leaflet provides. Like CityView, we will be using MarkerCluster, which adds animated marker clustering functionality to Leaflet.

2.4.1 D3.js

D3.js is a JavaScript library that allows us to make dynamic, interactive data visualizations in web browsers. Using D3.js, it is possible to create almost any type of visualizations imaginable. For the purposes of this project, we will be using it to visualize non-geospatial data.

Chapter 3: Methodology

3.1 Goals

Our main goal for CityView+ is to create a website with compelling information for users to draw meaningful conclusions about the data with. In doing so, we hope that the users of our website will come away with a better understanding of the urban environment in Boston, and how otherwise dull statistics might have meaning when combined with other data. We accomplished this goal by meeting the following requirements:

- The website must include data visualizations which are visually engaging and tell a story with the data displayed
- The website itself must be easily navigable and usable
- The website must use up to date datasets which and reflect the most current information possible

3.2 Design

The purpose of our design is to clearly display the data pertaining to the categories we defined, in a way that was engaging to the viewer. These categories were: demographics, livability, and events. Each of these categories had their own designated page on the website to ensure clarity when in any selected category. To construct the website, we employed the use of GitHub to keep our codebase organized and to deploy the website using GitHub pages. Each team member created their own fork of the main repository and we all made sure to keep up to date with the latest changes on that repository.

Our website was constructed using HTML, CSS and JavaScript. We implemented the Bulma framework to style our website. The data visualizations were implemented in JavaScript with the use of third-party APIs. We used Leaflet to implement easily navigable and interactive maps. We used D3 to implement easily understood and interactive data visualizations.

3.3 System Overview

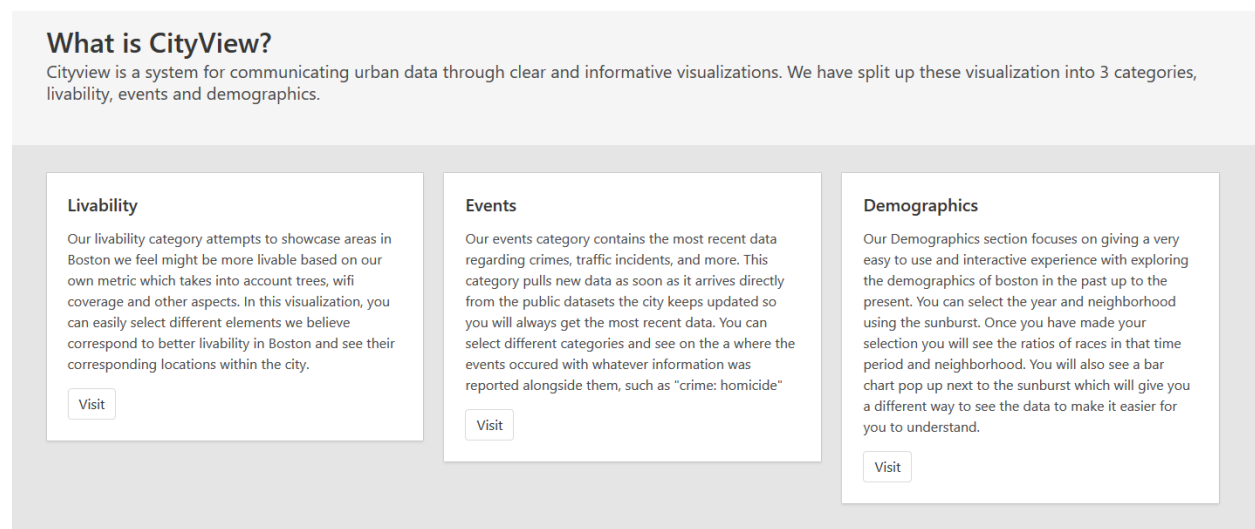
A simple yet effectively built system is an important feature to have in a project. We have built a functional site with the following characteristics:

3.3.1 Client

We built our website using JavaScript scripts, HTML pages, and CSS styling. The file system was organized to keep the scripts, HTML, and styles in their respective folders to simplify modifying and referencing. Some of the datasets we use in the website are not referenced directly from Analyze Boston because they are static. We have downloaded some of the static datasets and referenced them directly from our file system to reference in our visualizations.

3.3.2 System Infrastructure and Functionalities

Navigation and use of our system are simple. When first entering our website, the landing page contains an introduction that briefly describes the basic functionality of our website. From there, one can scroll down to find links to all our data visualization pages, all of which are accessible from the static navbar at the top of each page. The more specific details of our different visualizations will be described in the following section of this chapter.



The screenshot shows a website section titled "What is CityView?". Below the title is a paragraph: "Cityview is a system for communicating urban data through clear and informative visualizations. We have split up these visualization into 3 categories, livability, events and demographics." Below this are three white boxes, each representing a category. The first box is titled "Livability" and contains text about showcasing areas in Boston based on metrics like trees, wifi, and coverage, with a "Visit" button. The second box is titled "Events" and contains text about the most recent data regarding crimes and traffic incidents, with a "Visit" button. The third box is titled "Demographics" and contains text about exploring demographics of Boston from the past to the present, with a "Visit" button.

Figure 4: An overview of the website's content

On the navbar you will also find links to “About” and “Team”. If you click the link to “About” you will be led to a page with an in-depth description on each visualization with links to

each one respectively. If you click the link for “Team”, you will be greeted with a page describing giving some basic information on each team member who participated on this project.

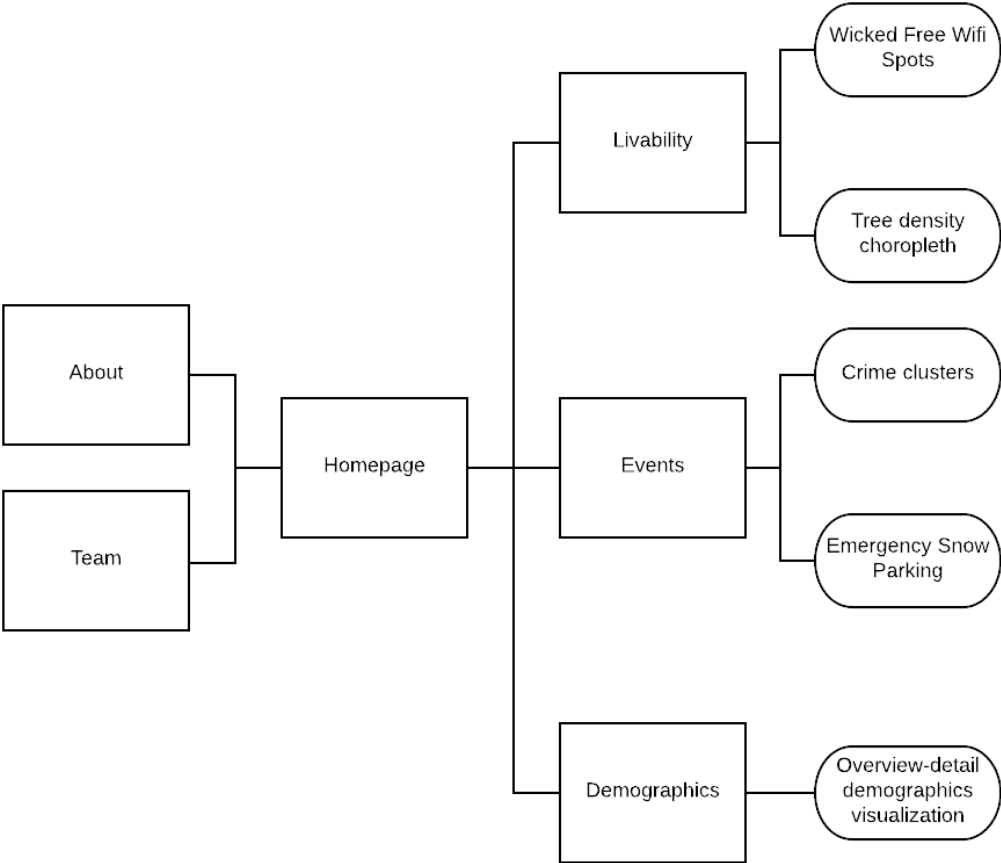


Figure 5: Overview of the structure of the website

Above is a diagram that represents the structure of our site. Visualizations and other technical content are shown on the right of the diagram while non-technical content is shown on the left.

3.3.3 GitHub

We used GitHub pages to host our site. It was an easy way to host our visualizations since all the code was implemented to work client side. Therefore, we did not need to set up a server to serve the pages.

3.4 Categories

3.4.1 Demographics

Our goal for the Demographics category was to produce interesting visualizations with data about the population in Boston and groups within it. Additionally, we want to use the data and visualizations in combination with our other categories to infer some useful information from it. For example, we could combine Demographics with Events to tell which population has more crime in the area. We will be using the Boston Neighborhood Demographics dataset from Analyze Boston as our main dataset for this category.

Dataset

The Boston Neighborhood Demographics dataset was created by the Boston Planning & Development Agency Research Division (BPDA) from the U.S. Census data to detail demographic changes in Boston's neighborhoods from 1950 through 2010 (Boston Neighborhood Demographics). The dataset consists of data per decade at a neighborhood level. Due to the characteristics of the Census data, some variables are not filled in for some decades. However, the BPDA indicates those cases in the data and tries to show consistent data. The Census Tract and the BPDA have different boundary definitions for Boston's neighborhoods. Most of the times the differences are minor. To minimize any major boundary differences, the BPDA has settled for the 23 color-coded group of neighborhoods defined in the 2010 Census Tract.

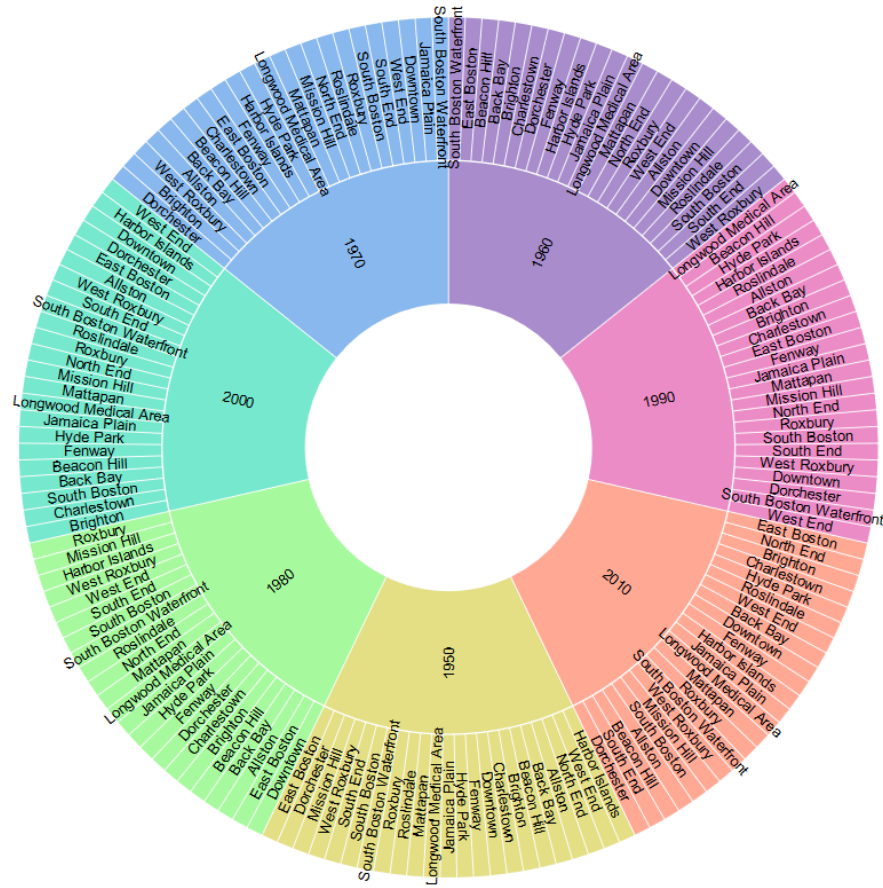


Figure 6: Full-sized sunburst diagram

We chose to represent the ethnic diversity of each neighborhood with a multicolored interactive sunburst diagram. Each slice is colored by decade, and each decade slice has an outer section with slices for each neighborhood.

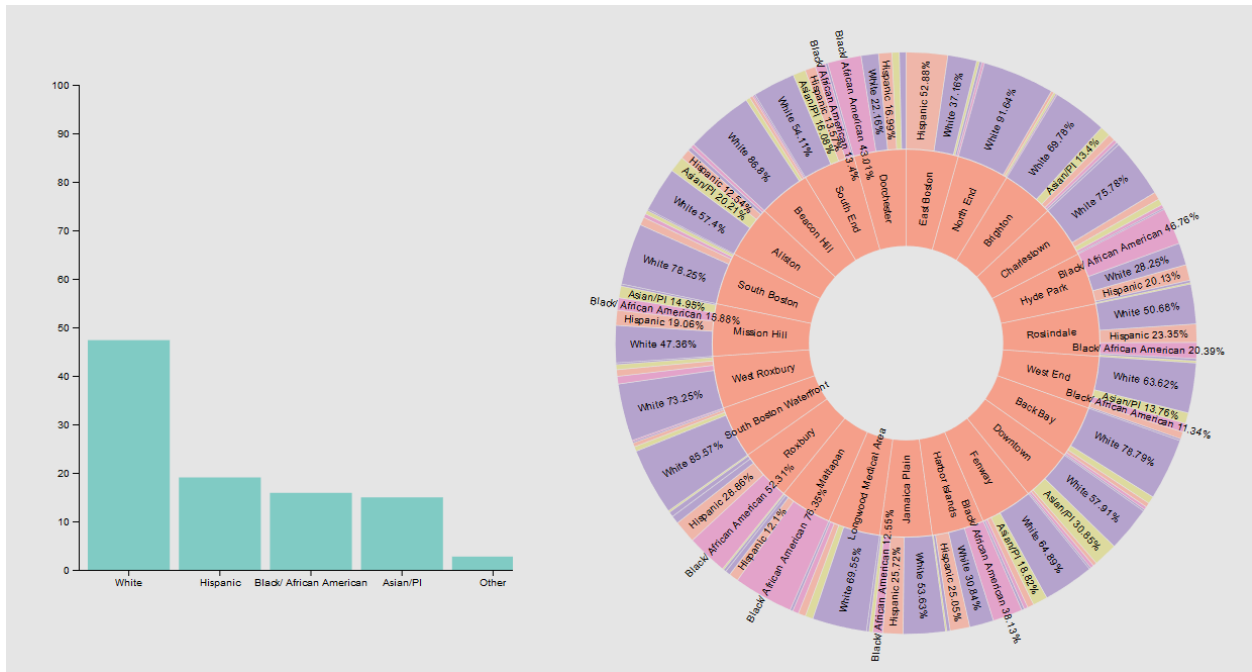


Figure 7: Screenshot showing overview-detail sunburst

Clicking on a decade slice will update the sunburst to only show neighborhood data for that decade. The outer slices now represent the ethnic diversity of each neighborhood and are sized according to the population percentage of each race in that neighborhood. Clicking once again on the outer slices, the sunburst will become a donut chart showing the population percentages more clearly. A bar chart will also appear next to the sunburst, which gives a different perspective of the population breakdown for the selected neighborhood. Clicking in the center of the sunburst will bring the visualization back up one level, which can be repeated until the sunburst is reset. The bar chart will remain until a new neighborhood is selected, causing it to update with new data.

3.3.2 Events

With the Events category, we aim to provide live updates with information useful to the public.

Crime Data

When a crime is committed, a record with the location and crime type is saved to a central database. We use this information to let consumers know when and where a crime was committed. Users can view the history of crimes in a certain area using a widget which allows

them to choose an area and time frame of crimes committed to view all the crimes in that area and time frame. They can use this information to infer whatever they want, such as, the most common type of crime, the frequencies of crime, and whatever else they might want information on.

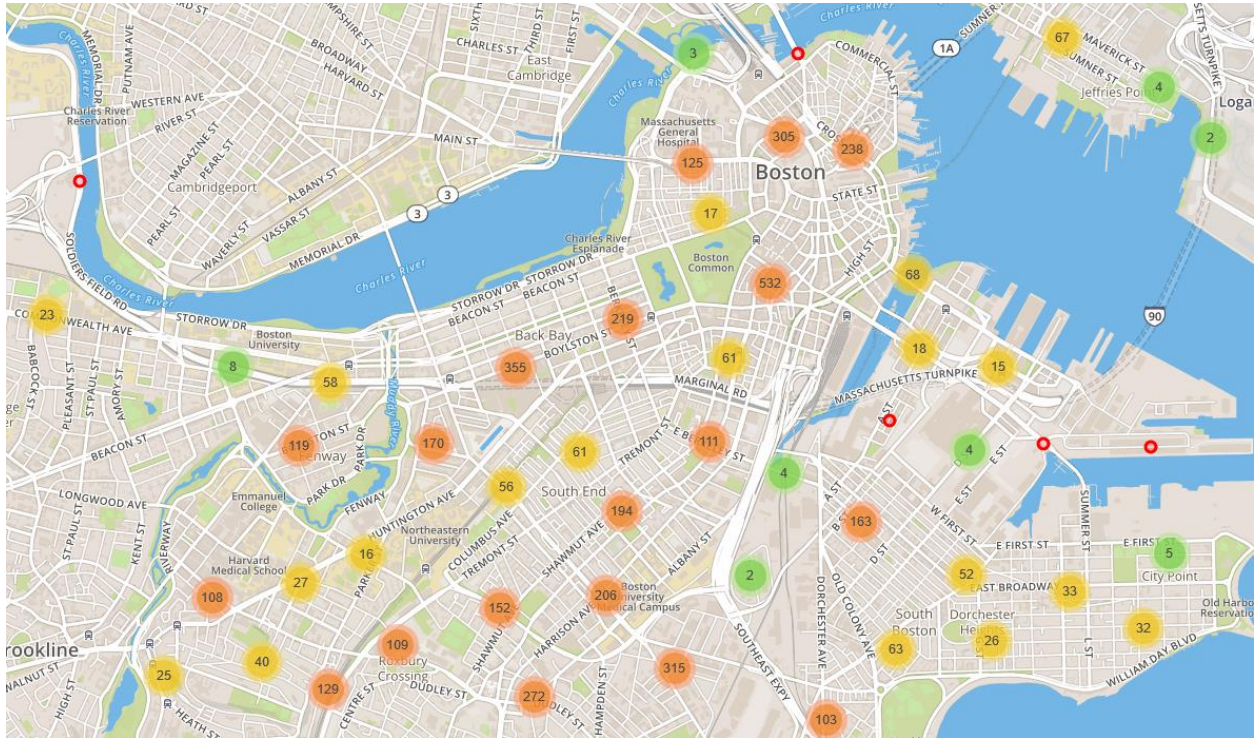


Figure 8: Crime report marker clusters

We visualized the crime data by first grouping them into clusters by geographical location, and then plotting them onto our map. The clusters are sized to be spaced out evenly across the city. The number inside each marker cluster represents the number of points that it is clustering. The color also is indicative of the amount of marks inside the cluster; colors range from green to yellow to orange, representing a single-, double-, or triple-digit number of markers in the cluster.

Emergency Snow Parking

When winter comes, there is important information a person might need access to. To ensure the safety of drivers in the winter, we decided to include emergency snow parking information in our Events category, which allowed drivers to locate the nearest place they could park and remain safe in the event of a rough snowstorm.



Figure 9: Emergency snow parking

You can locate emergency parking locations on the visualizations via the blue circles on the map. Each parking location can be clicked to show a popup with more information about each location. This way we can present the data in a meaningful and interesting way without losing detail about each location.

3.3.3 Livability

Our goal with the Livability category is to leverage several sources of data to create an index of how “livable” an area of a city is.

CityScore

Boston’s CityScore data platform is “an initiative designed to inform the Mayor and city managers about the overall health of the City at a moment’s notice by aggregating key performance metrics into one number” (CityScore). That one number is an index that ranges from zero to two. An index of exactly one means that the performance of that service is on target as defined by the mayor. Indexes above 1.0 indicate the service is above the target performance, and below one means the service is underperforming. Each metric is updated daily, and weekly

and monthly averages are tracked. The metrics span different areas of Boston's daily living conditions, such as crime rates, city maintenance on-time percentages, and library users to name a few. We augmented these metrics with our own interpretation of what makes a city "livable."

We included metrics that reflect the number of "green spaces" around the city, such as parks or areas with trees and grass, the density of free Wi-Fi connections, other crime rate statistics that may not be tracked by CityScore, and the proximity to areas of interest such as public buildings like libraries, schools, police stations, and other population hotspots.

The CityScore dataset contains fields that represent each service, the corresponding metric's index value, each metric's target value if specified, and weekly, monthly, and quarterly averages. The other datasets we will use in addition to CityScore include crime reports, Wi-Fi access point locations, and locations of trees in public spaces. The crime report dataset has fields that represent the type, date, location, and description of crimes that have been reported to the police department.

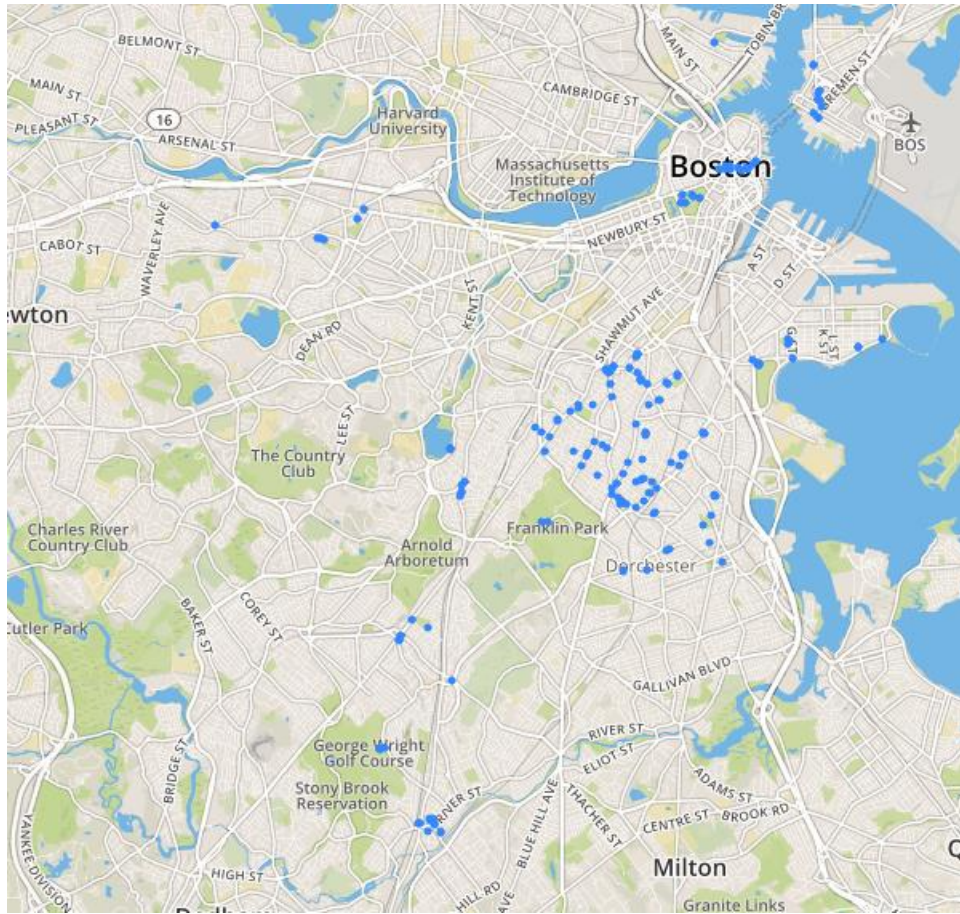


Figure 10: Free Wi-Fi access points

Wi-Fi

The Wi-Fi access point dataset lists locations and names for each access point around the city that is part of the city-wide “Wicked Free Wi-Fi” (How Wicked Free Wi-Fi Works) program. You can locate each Wi-Fi Location by the blue circles on the map. When you have selected a location, you will be able to view the address attributed to that Wi-Fi location.

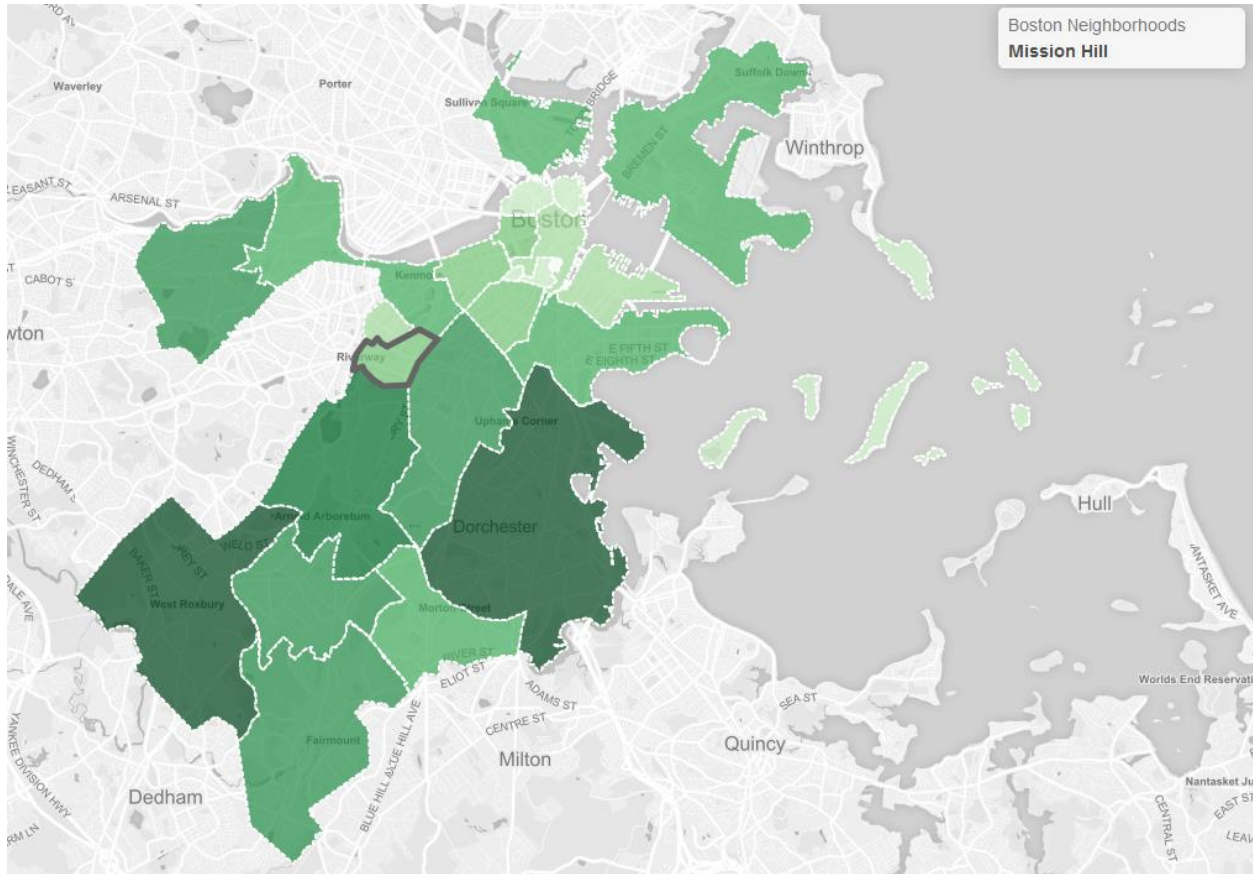


Figure 11: Choropleth map of tree density by neighborhood

Trees

The tree dataset contains locations and descriptions of trees catalogued by a planimetric project completed in 2011. We showcase the tree data by outlining certain Boston neighborhoods, and then shading each district by the number of trees in that area. When a neighborhood is selected, you can view the name of that respective neighborhood on the top right.

3.5 Implementation

We mainly use both Leaflet and D3.js to make interactive visualizations. To get the datasets, we query Analyze Boston for their datasets via SQL. If the datasets are not obtainable via SQL, we download the dataset and store it in a local directory. The queried datasets are preprocessed for later use in our visualizations.

3.5.1 Leaflet

We used Leaflet to create interactive maps for categories whose dataset contained geospatial data. Leaflet allows us to implement layer toggling between datasets. Toggling between datasets allows users to choose how they want to explore the data.

Like the previous MQP, CityView, we encountered performance issues when we try to render many markers from a large dataset. This was because each marker on the map was an individual DOM element. However, we solved this using the “preferCanvas” option when instantiating the map, in addition to using MarkerCluster. This way all markers (and marker clusters) are drawn in one HTML canvas element, as opposed to each being rendered as their own element.

MarkerCluster is a plugin for Leaflet that lets us add animated marker clustering functionality, clustering the markers together at different zoom levels of the map. The plugin offers a wide array of options to configure as well as customization of the clusters themselves. Depending on the zoom level, either individual markers are shown, or clusters labeled with the number of markers they represent. At max zoom level, individual markers are displayed; otherwise, clusters are shown. We configured at which zoom levels clusters are shown using the options provided by the MarkerCluster.

3.5.2 D3.js

We are using D3 for three things. Loading csv files, processing data, and visualizing demographics data using a zoomable sunburst chart using an implementation like the one shown here (Bostock).

The zoomable sunburst diagram shows multiple layers of data by leveraging the structure of nested JSON objects. After preprocessing the demographics data, we convert it to a nested JSON objects by using JSON manipulation functions provided by D3. The nested JSON objects are grouped in a hierarchical tree structure; much like using the GROUP BY operator in SQL, except that there can be multiple levels of grouping and the result is a tree rather than a table.

CHAPTER 4: Results and Analysis

4.1 Usage

CityView+ is currently fully functional and accessible from the following link: <https://cityviewplus.github.io/>. The interface is easy to use and communicates information clearly. The full functionality of CityView+ is outlined in the design section of our methodology. CityView+ gives the user tools to understand the landscape, amenities and events attributed to Boston through the most current data available.

4.2 Results

4.2.1 Website

The website has all the features we aimed to include during the initial planning phase of the project. What we have in place is a system for discovering different aspects to Boston including the livability, demographics and ongoing current events. This was accomplished using the most up to date datasets from “Analyze Boston” (City of Boston). For the visualizations that required it, we would query the most up to date information documented and distributed by the city of Boston.

4.2.2 Livability

The Livability section offers a simple visualization that allows users to view the conditions, amenities and features of different districts in Boston. Users can select what they would specifically like to view to get a better picture of the specifics of each district.

4.2.3 Demographics

The Demographics section allows a user to understand the demographics of each neighborhood in Boston, this is presented via a selectable sunburst diagram. We also allow the user to view what the same neighborhoods demographics were in the past few decades. We also show the data as a bar chart when interacting with the sunburst diagram, so the data might be more easily understood.

4.2.4 Events

The Events section offers users to view some of the most current occurrences in Boston. A user interacting with the events visualizations will be able to learn about the most recent crime statistics reported by the Boston police daily. We use the most recent data by querying the dataset every time the webpage is opened so the user is always getting the latest info. One can also get some basic info by zooming into a specific crime and selecting it to determine the nature of the crime, when it was committed, and any other documentation provided by the Boston police.

CHAPTER 5: Conclusion and Recommendations

When developing CityView+, our initial idea was to create an interface that was usable and that could provide valuable information relevant to the average citizen in Boston. Because of this we decided our goal should be to create a website with compelling information for users by telling a story with visualizations. We accomplished this by creating visually engaging data visualizations, making the website easily navigable and usable, and using up to date datasets which reflected the most current information possible.

5.1 Recommendations

We have a few recommendations for any teams looking to work on this project or any projects like CityView+:

Server Backend

Having a server backend to do the heavy lifting for creating maps and their corresponding layers would improve user experience. This is especially effective for storing maps and layers that do not require frequent updates from new data. Additionally, the server could allow more complex and computationally intensive visualizations to be shown client-side.

Mobile App

A mobile app would increase the usability of our website. Having a viewing experience optimized for smartphones means users expands our potential userbase to people without immediate access to a computer. While using a phone browser is an alternative to a computer browser, having a mobile app allows for more fine-tuned control over how the visualizations are presented to the user.

Data Availability

Something we did not anticipate when starting this project were issues finding datasets to satisfy our vision for each category. This made things difficult down the line since we only had a few datasets to work with to try and put forth as much valuable information as possible. We

would recommend future teams to consider what they would like to visualize, how they would like to visualize it, the feasibility of visualizing said data, and then move on from there.

REFERENCES

Bostock, Mike. “D3 Zoomable Sunburst.” *Beta.observablehq.com*, Observable, 30 Apr. 2018, beta.observablehq.com/@d3/d3-zoomable-sunburst.

“Boston Neighborhood Demographics.” *Analyze Boston*, Department of Innovation and Technology, data.boston.gov/dataset/boston-neighborhood-demographics.

City of Boston. “Analyze Boston.” *Analyze Boston*, City of Boston, 15 Apr. 2018, data.boston.gov/.

“CityScore.” *Boston.gov*, Department of Innovation and Technology, 19 July 2018, www.boston.gov/cityscore.

Dataveyes. “An Educational Simulator to Help Understand the Issues around Energy Flexibility.” *CityVis*, 2014, cityvis.io/detail.php?id=111.

Department of Innovation and Technology. “Demographic Data for Boston's Neighborhoods, ...” *Analyze Boston*, data.boston.gov/dataset/boston-neighborhood-demographics/resource/e6956d7e-92cb-49eb-8123-2a5a3452a2e1.

Walsh, Martin J. “How Wicked Free Wi-Fi Works.” *Boston.gov*, City of Boston, 2 Nov. 2018, www.boston.gov/departments/innovation-and-technology/wicked-free-wi-fi.