Bar Hopping Visualizer

Zoheb Aziz, Ning Yu Chen, Marco Tortolani, Jason Zhang
Northeastern University

Abstract

In this project, our group will create a tool that filters and displays the location and attributes of different bars throughout boston. This interactive visualization tool should allow users to rank Boston bars by different attributes that are mostly appreciated by them. The attributes consist of reviews, hours of operation, type of establishment, location, pricing, and capacity pulled from Yelp's business.json dataset and other external sources that might be explored in the future. This tool aims to help Boston residents navigate the late night drinking scene.

GitHub Repo:
https://github.com/DS4200-S22/final-project-bar-hopper

1 Introduction

While Boston is a city rich with diverse bar options for all types of people, it can sometimes be difficult to locate or decide on the particular establishment that most fits one's needs. One's desired location involves an abundance of factors: location, price, style, etc, and researching into all these factors is not an easy process. It is common for people to spend lots of time searching for their target location, which they could have instead spent at the bar having a great time. We want to minimize the decision making time by simplifying the process.

2 Related Work

We start by looking for existing visualizations for our specific use case. From Pinterest (a non peer reviewed source), we have the "Boozetown Massachusetts" map posted by Philippe Guelton [1]. This visualization is a static map of Boston with heat maps overlaid representing the amount of liquor licenses in a given area. This is an ideal for the "generalist" approach to bar hopping, where you want to know which direction to go and figure out the details later. We will strongly consider using a heatmap such as this one as it is so effective in pinpointing active bar hotspots. In developing this heat map we will look to "Mapping Color to Meaning in Colormap Data Visualizations" which draws conclusions about how color map palette choices have effects on readers [2]. Particularly it will help us take advantage of how the human brain automatically perceives opaque and darker colors as more intense in a color mapping. Another relevant paper in mapping is "cpmViz: A Web-Based Visualization Tool for Uncertain Spatiotemporal Data" [3]. This is based on the 2019 VAST visualization challenge, and we can use their mapping of points to locate earthquake hotspots to either create static maps of bar hotspots or to track other app users and their bar activity in an active map.

For recommendations on bars, we can look to create an active visualization with tools as described in "MCORE: a context-sensitive recommendation system for the mobile Web" [4]. By using implicit data such as location of access or time of access, we can create a visualization that presents the best bars for you without having you figure out what's close or what you'd like. Additionally, work from "Designing Guidable Bus Transit Database for Chinese Cities" can be used as a framework to how the bars should interact as objects on the map, which will aid in adding features to the visualization in the long run [5].

3 Use Case

With this project we have two aims: to provide a concise summary of the bar scene in Boston (something that may be difficult simply searching 'bars' in yelp), and to create a tool that can filter to individual needs in this use case. A high rated bar does not do a student well if it's brutally out of their price range, while the same establishment might be the perfect fit for a middle-aged patron. Our tool will be able to aid each user in finding their perfect fit.

4 Data

The Yelp dataset offers insight into the thousands of restaurants and bars with profiles on the site. It contains attributes such as: location, opening hours, categories, and in depth information on reviews and ratings of each business. This dataset can be used to extract useful information such as ratings and filtering categories to bars, clubs, and restaurants in the area to find the closest and most highly rated establishments to visit around the area.

5 Task Analysis

Task ID #	Domain Task	Analyze Task (high-level)	Search Task (mid-level)	Analytic Task (low-level, "query")
College Student	I want to quickly locate cheap and highly rated bars for my friends and I to go to.	Consume -> Discover	Browse	Identify
Someone planning a date	I want to use a visualization to get an in-depth understanding of various locations which fit my criteria, so I can guarantee that I choose the best spot.	Consume -> Discover	Browse	Identify
A couple planning a vacation	I want to determine if there are suitable locations which will fit our busy vacation schedule.	Consume -> Discover	Browse	Identify

Normal Users	I want to quickly rank all the bars according to some attributes.	Consume -> Discover	Browse	Summarize
Normal Users	I want to just look up a known bar.	Consume -> Discover	Lookup	Identify
Normal Users	I want to filter out bars with certain attributes.	Consume -> Discover	Browse	Identify

The primary consumer of our visualization will be college students who are planning a night out. They will benefit from the simple interface and clear presentation of relevant information that they need to make their decisions. Our visualization will be primarily developed for the purpose of Discovery. This is because our end users will be working to discover and identify the most optimal and suitable locations which they would like to visit. While our visualization can also be leveraged in the contexts of Presentation and Enjoyment, its primary focus is to aid users in discovering the most suitable locations based on their individual preferences, thus allowing them to make more informed choices, given the new discoveries.

6 Implementation Plan & Preliminary Work

The base of our visualization would be a map of Boston, with an interactive node representing each location, which can be accessed for further information about the selected establishment. After the user clicks on the node, we will display organized information such as rating, hours, etc. We will utilize the following technologies to build our visualization: D3.js, HTML, CSS, Javascript, Python.

Graphs:
- A main interactive Massachusetts map that displays all the bars
- A search table to display all the bars and their attributes
- A customized line chart that displays the hours of operation
- A radial graph that displays the distance and direction of the bars from the user
- A scatter plot that compares the price tag, review count, and review rating of the bars
- A paginated view window for displaying the written reviews

Interaction/Interactive Components:
- The main map can be zoomed
- Users can filter out bars by a list of attributes(price tag, review count, categories….)
- Brushing on the main map shows the brushed bars on the table, radial graph, and scatter plot.
- Users can search up for bars by keywords
- Selecting a bar on the map, table, radial graph, and scatter plot show the bar's hours of operation and all its written reviews with pragenation

Annotations:
- The main map is a **must** because all other graphs have dependencies on the map
- The **bushing/linking** functionality is a **must** to integrate all the graph together
- The filter is **important** to have to reduce the search size
- The table that displays all the bars is **important** to search, filter, and select bars
- The scatter plot is **important** to have to find the cheapest and best reviewed bars
- The customized line chart is **nice to have** to display the hours of operation of the bars
- The radial graph is **nice to have** to display the distance and direction of the bars
- The paginated view window is **nice to have** to display the written reviews of the bars

Requirements
- The data we are serving is static, several preprocessed csv files. Then we also serve javascript files that use D3 to support user interactions, html and css files to structure and style the website. At this moment we are not planning to include a database into this project which means there isn't a need to expose REST API. However, we will use the Yelp API to obtain written reviews about a particular bar.

7 Visualization Design

Our final visualization tool design consists of 3 main coordinated views: the high-level Boston map with all bar locations denoted with geographically accurate coordinates, the individual bar view with relevant information such as pricing, ratings, and hours, and the location/directional view which provides relevant information regarding the available establishments in the selected bar's vicinity. We chose these visualizations to compose our final tool because it provides the perfect combination of high-level and low-level, detail-oriented information. Users are able to seamlessly browse through large numbers of establishments utilizing the map visualization, and then request and ingest a wealth of information regarding the specific establishments that they select and are most interested in.

The map view offers the most comprehensive, high-level overview of all of Boston's bar options for the user, and invites the user to further explore and query about particular areas or individual bars that will be most suitable for them. In addition, it incorporates brushing features which we can further link to pie charts/histograms to offer the user further insight into the selected area. This view supports the task of gaining insights on a group of bars in a selected area, and it supports data types of geo-location and numerical/quantitative price. Its visual encodings are effective because the map is very intuitive for humans to look at and understand geo-locations and positionings of bars relative to one another, and it also offers a clean and high-level view of the data which the user can easily look more closely into with the interactive features.

The individual bar view clearly and simply informs the user on relevant information such as the weekly hours of operation and reviews of the particular bar in a manner that permits easy comparison and intuitive understanding. This view supports the task of determining whether a particular bar has a suitable schedule for a user's busy vacation plan, as it allows for a

comprehensive view of weekly hours of operation where a user can clearly see if the bar can accommodate them at a particular time. It supports data types of time intervals. The visual encoding is effective because it is so simple and easy for the user to understand just at a glance. The user is also given the ability to quickly understand the distribution of reviews for the location and to read further into the exact comment left by past patrons, if they are interested.

The Location and Direction view offers similar but more detailed information as the higher-level map view, and it can give users more accurate bearings on how they would be able to walk/commute to their target location. This view supports the task of determining how to travel to a particular location, as well as to gain bearings on where nearby locations are so that hopping over to the surrounding establishments is also made simpler. It supports data types of geo-location. The visual encoding is effective because it provides an intuitive and focused view of the particular location and nearby establishments as well which a user can easily interpret and gain information from.

Ultimately, these visualizations support the data and tasks previously identified because they display all relevant data (geo-location, hours of operation, pricing, reviews and review distributions, etc.) Furthermore, their combination allows for a high level overview of all options, filtering and querying to minimize the sample/search space, and various descriptive and summary statistics that aid users in making informed and efficient decisions.

References

[1] Philippe Guelton, “Boozetown Massachusetts”, Feb 2012, https://www.pinterest.com/pin/133982157634646172/

[2] K. B. Schloss, C. C. Gramazio, A. T. Silverman, M. L. Parker and A. S. Wang, "Mapping Color to Meaning in Colormap Data Visualizations," in IEEE Transactions on Visualization and Computer Graphics, vol. 25, no. 1, pp. 810-819, Jan. 2019, doi: 10.1109/TVCG.2018.2865147.

[3] F. Nagel, G. Castiglia, G. Ademaj, J. Buchmller, U. Schlegel and D. A. Keim, "cpmViz: A Web-Based Visualization Tool for Uncertain Spatiotemporal Data," 2019 IEEE Conference on Visual Analytics Science and Technology (VAST), 2019, pp. 140-141, doi: 10.1109/VAST47406.2019.8986941.

[4] Yeon Choi, Joon, et al. “MCORE: A Context-Sensitive Recommendation System for the Mobile Web.” *Expert Systems*, vol. 24, no. 1, 2007, https://doi.org/10.1111/j.1468-0394.2007.00419.x.

[5] Z. Huang and X. Liu, "Designing Guidable Bus Transit Database for Chinese Cities," 2007 International Conference on Wireless Communications, Networking and Mobile Computing, 2007, pp. 5488-5491, doi: 10.1109/WICOM.2007.1344.

Appendix A: Group Charter

- **Group Purpose**

We are a group of collaborative and motivated students who have a collective vision for this project. Our group's purpose is to work in a synergistic manner to produce an informational and practical visualization tool that can benefit our peers and the city of Boston as a whole.

- **Group Goals**

Our goal is to help users make a decision on what bars/clubs/restaurants to go to on a given night. We hope to accomplish this by creating interactive visualizations that are filled with information about the attributes and qualities of nearby locations, but are also simplistic in design and easy to use. Our group is willing to put as much time and effort necessary to produce a product that accomplishes our purpose and that we are proud of. We are collectively aiming to get an A on this project.

- **Group Member Roles/Responsibilities**

Researcher (Decides what data to use and how to present the data visually): All
Data process (Transfers the data from external sources into compatible format): Jason, Zoheb
Paper writing (Writes the paper): Jason, Marco
Data visualization Presentation (Implements the visualization): Ningyu, Marco
Project manager (Keeps track of the timeline, member coordination, and assigns tasks): Zoheb, Ningyu

- **Ground Rules**

We'll primarily meet virtually but in person when necessary. We agree to be cognisant of time management and responsible for our individual tasks. We'll collectively discuss all decision making and employ dialogue to handle conflicting interests. To hold each other accountable, each person is assigned a task each week.

- **Potential Barriers and Coping Strategies**

Barriers would involve poor communication between members. To handle this, we will emphasize informing the whole group of any decisions we make. Group dynamic problems we've experienced in the past are individual members taking control of the whole project by themselves. Once again, to avoid this, we will stress the importance of open communication between all members.

- **Follow-Up**

So far, our team has been abiding by the guidelines and loosely following the predetermined roles. Even though we push the leads onto the respective roles (ie. Marco and Jason drive the paper writing while Jason and Zoheb perform most of the data cleaning), we regularly call in preparation of each project milestone and plan out how we can all pitch into parts. So far we've felt really comfortable with our roles and we regularly take the extra steps to help between roles (such as Jason's check-in texts or Zoheb going in to proofread the paper).

The slight shortcomings, such as Marco's difficulties with tableau, are made up by strengths, such as Marco's python skills helping with data cleaning. Again, no need for hard role changes as we communicate effectively at every milestone. The only other problem is that we need to limit our project scope. Project milestone 5 will be essential to defining exactly which visualizations we will be using, and then we can shift gears into coding them into reality. If we do not have our needed visualizations listed out, we can arrange a decisive meeting to choose out our best designs.

A few positive mentions between group members:

- Marco: I appreciate how Zoheb is always free to call. Especially for the first few milestones, Zoheb organizing the first few meetings prevented me from procrastinating. Additional shoutout to Jason for helping with the Tableau when I was having setup troubles.
- Jason: Ningyu's use case table helped me a lot with creating visualization sketches and understanding the explicit targets our project should hit.
- Zoheb: Jason is really good at data cleaning with python and guided us on how to use the way too complicated yelp API
- Ningyu: During the related work section, Marco looked through all the visualization reports and quickly got back to us with a summary of some good papers.

Appendix B: Data Exploration

- **Data Review**

The data types present in our dataset are attributes (categorical, ordinal, and numerical) and geo-spatial data. The attributes of interest for us are 'price' and 'reviews' from the YELP dataset. The price is ordinal, with '$' representing the cheapest tier and '$$$$' representing the most expensive tier. The price distribution is unimodal with '$$' being the most common by far. The range of review ratings is 1 to 5, with 5 representing the most positive feedback. The reviews distribution is unimodal with a left skew, most reviews were between 3 and 4.
Issues with the data we confronted are as follows:

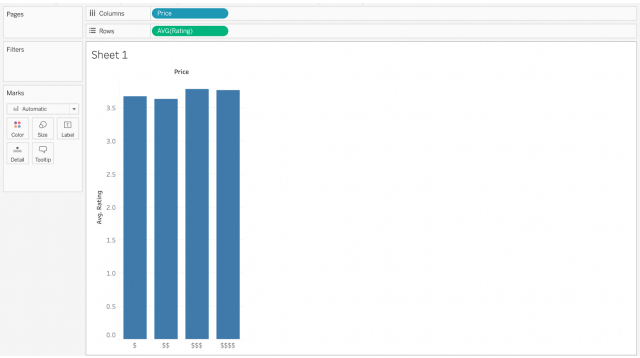
- Our data contained missing/null values in various columns. To address this, we decided to drop records which had missing price data because this attribute is highly relevant to our visualization, and we have a sufficiently large dataset so that dropping incomplete records won't be detrimental. The null values in the other columns will not be a concern because we generally won't be performing queries on 'address2' and 'address3' for example, so we are choosing to leave those values as null.
- Although all of the locations we pulled were recognized as bars, many locations fell under multiple categories (ie. recognizing the location as a 'bar', 'breakfast & brunch', and 'diner). Our solution consists of removing all items that do not have 'bar' as the category. Additionally, we will use the 22 char 'id' category as our primary key to avoid any doubling of locations.

- **Insights**

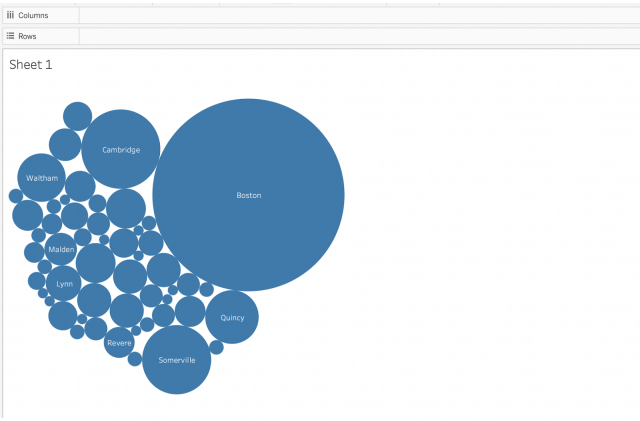
From our data, we noticed that most bars had average reviews of between 3-4, which represents a generally positive sentiment. There were only a few outliers in either extreme end (1-2 and 5 ratings). Furthermore, we were surprised to see that there were more expensive bars than cheap bars, with the count of '$$$' bars at 94 and '$' at 63. We would've assumed that there would have

been more lower cost establishments, particularly in a College town like Boston. In addition, we noticed a pattern that as the number of reviews for an establishment increased, their rating tended to approach a 4. This is likely due to the 'regression towards the mean' phenomenon. Regarding the distribution of review counts, as expected most establishments have <500 reviews, and the distribution is skewed heavily to the right. This makes sense because only a small subset of popular bars would have review counts exceeding 500.

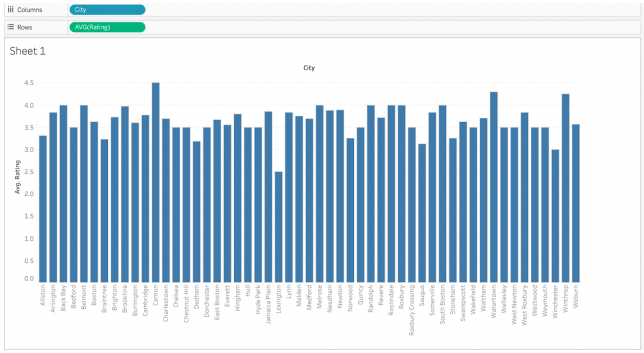
- **Screenshots**



This screenshot portrays the relationship between the average YELP rating and its price classification representing how expensive items at the location are. It seems as if more expensive locations ($$$) and ($$$$) have a slightly higher average rating than the cheaper ones, but it is close. This is pretty expected behavior since on average we would expect places that are more expensive to offer higher quality food and drinks. For this we used a bar plot because it was visually easy to compare the 4 different price points side by side.



This next visualization shows locations of the establishments in our dataset. The circle size represents how many of the organizations in the dataset reside there. The actual plot is interactive and has a tooltip that shows the name of the city for the smaller bubbles on this graph as well as the number of organizations residing there. This plot seems to show that the bigger/ more populated cities have more bars/clubs in our dataset with Boston having by far the most. This is pretty expected since Boston is somewhat known for its nightlife. Boston has ~350 entries for this dataset indicating there will be plenty of entries to use for this project. For this we used this plot because it really emphasized the cities with lots of dataset entries.



This last visualization brings together the purposes of the last two to find the relationship between the city and the average review scores of the organizations within it. It seems as though the biggest cities have some of the lowest average scores while some of the smaller cities have very high average scores. For this we used a bar plot because it differentiated each of the cities well and it's clear and readable for all of the cities in the dataset on one screen.

- **Data Snippet**

rating	price	phone	id	review_count	name	url	latitude	longitude	address1	address2	address3
4	$$	(617) 523-9026	2u_w3rthRzR2u	3719	Giacomo's Risto	https://www.yelp	42.36458	-71.05344	355 Hanover St		
4	$$$	(617) 267-4000	AGR_kBvN__I7c	2769	Atlantic Fish	https://www.yelp	42.34922179	-71.08112729	761 Boylston St		
4	$$$	(617) 536-4300	NYa-JphaaB41E	2294	Toro	https://www.yelp	42.33698788	-71.07592429	1704 Washington St		
4	$$	(617) 456-7849	xOMKjE4xmTgi	2127	The Friendly Toa	https://www.yelp	42.34834793	-71.07327305	35 Stanhope St		
4.5	$$$	(617) 267-0891	sA2gVTJOBH7C	1441	Sattie Girl	https://www.yelp	42.35116	-71.07758	279 Dartmouth St		

city	state	zip_code	country	hours_type	mon_start	mon_end	mon_overnight	tues_start	tues_end	tues_overnight	wed_start	wed_end
Boston	MA	2113	US	REGULAR	1200	2200	FALSE	1200	2200	FALSE	1200	2200
Boston	MA	2116	US	REGULAR	1130	2200	FALSE	1130	2200	FALSE	1130	2200
Boston	MA	2118	US	REGULAR	1700	2100	FALSE	1700	2100	FALSE	1700	2100
Back Bay	MA	2116	US	REGULAR	800	1500	FALSE	800	1500	FALSE	800	1500
Boston	MA	2116	US	REGULAR	1100	2200	FALSE	1100	2200	FALSE	1100	2200

wed_end	wed_overnight	thurs_start	thurs_start.1	thurs_overnight	fri_start	fri_end	fri_overnight	sat_start	sat_end	sat_overnight	sun_start	sun_end
2200	FALSE	1200	2200	FALSE	1200	2230	FALSE	1200	2230	FALSE	1200	2200
2200	FALSE	1130	2200	FALSE	1130	2300	FALSE	1100	2300	FALSE	1100	2200
2100	FALSE	1700	2100	FALSE	1700	2200	FALSE	1700	2200	FALSE	1700	2100
1500	FALSE	800	1500	FALSE	800	2000	FALSE	800	2000	FALSE	800	1700
2200	FALSE	1100	2200	FALSE	1100	2300	FALSE	1100	2300	FALSE	1100	2100

Each image is part of the same 5 rows because the data set has lots of columns.

Appendix C: Interview

- **End User Personas**

I have had experience with using visualizations before in my life. Example contexts are observing trends in the stock market through line graphs derived from time-series data, and visualizations that display how different NBA players are performing in the current season compared to their previous years in the league.

The visualization that this project aims to produce will be highly relevant for the daily-life task of planning a night out with friends in the city. By interacting with the visualization, we will be able to make informed decisions on which subset of locations we should consider, which will save copious amounts of time, and then query further into the selected locations using the tool in order to make a final decision.

Currently, I perform this task by scouring the web, using geolocation features on google maps to search for nearby establishments, and then visiting each location's website to read and learn more about its unique characteristics. This is highly time consuming and I struggle to keep track with all the options let alone ranking them in my favor. I am looking for an interactive

visualization tool that can facilitate the process of finding and ranking nearby establishments.

I envision that the visualization will drastically improve this experience by increasing the quality and efficiency of the search, by providing high levels of relevant information, as well as allowing users to quickly filter for specific characteristics.

- **Interview Script**

Q1. What are the greatest challenges that you face when searching for and deciding on a bar to go to with friends?

Follow up: Are there any areas in the process where you would like to see improvements?

Q2. Where do you think that a visualization can be incorporated to benefit your decision making process?

Follow up: Is there any type of information that you wished would be displayed in a more accessible manner?

Q3. What types of information do you research and consider in order to make a decision?

Q4. What are the various steps that you take to come to a conclusion?

Follow up: Can you walk me through your thought process if you were going to plan a night out with friends for today?

Q5. If you have any ideas, what specific features would you like to see in the visualization?

Follow up: Are there any useful visualizations that you've used in the past? What features of those visualizations were most useful?

- **Interview Notes**

Q1.

- Biggest challenge - takes too long, too hard to keep track of all the options, can hardly rank the option by my favor, get frustrated
- Information is difficult to compile, some bars' information is not easily accessible on their websites
- End up settling for a spot that is not very satisfying out of convenience

Additional Q: Which of these challenges cause the most difficulty, i.e. what would you prioritize for creating a solution

- Clearly displaying relevant information for each bar (hover/click icon) and provide the functionality to rank them by certain attributes

Q2.

- A visualization which can display relevant information in an organized manner
- Want ease of use, low complexity, clean interface with the least amount of background noise
- Most important information is pricing, rating, categories of the establishment
- Ranking tool

Q3.

- Cost is most influential in the decision, also consider type of bar, hours of operation, distance from my location
- Rating is very important also, tend to stick to more favorably reviewed bars

Additional Q: Which of these types of information do you have the most trouble with finding on your own?

- Cost is the most difficult, many bars are not upfront about how much they charge so it can take a long time to locate this data, it would be great if the information was more accessible

Q4.

- I make sure that I consider all information points in order of importance: start with price, then rating, etc.
- I like to select 3 different bars as final candidates and save and compare them at the end for the final decision
- Consult friends and go through their menus before making the final choice

Q5.

- The core functionalities that this tool should implement are querying, ranking, and filtering of bars.
- I also would like the option to like/favorite bars which I can refer back to later
- The most useful visualizations tend to be the most simple, users won't be overwhelmed with features and background noise

- **Interview Results**

Q1. What are the greatest challenges that you face when searching for and deciding on a bar to go to with friends?

A1. My friends and I often spend way too much time and effort trying to decide what bar to go to on a night. Oftentimes we decide to just pick an area such as north end or fenway and go bar to bar to see which one we want to stay at, but that's pretty draining and can get very expensive especially on nights where there are covers. It's honestly just hard to keep track of which bars are on what price level and rank their other attributes in order of my favor, and there's no one resource to use that has all the information I need to decide what bar I would be satisfied going to.

Q2 . Where do you think that a visualization can be incorporated to benefit your decision making process?

A2. A visualization that had all the information I needed to make a decision about what bar to go to would help my situation a lot. In order for the visualization to be helpful I would want all of the information that I need to be available in one place with an interface that is aesthetically pleasing and easy to use. It should also provide the core functionalities to allow me to do querying, ranking, and filtering of bars.

Q3. What types of information do you research and consider in order to make a decision?

A3. Probably the most important information to have for me is price. As a college student, I want to go out and have fun and experience the Boston nightlife, but at the same time I'm on a budget and would like to get the biggest bang for my buck at every outing. Another really important thing for me is distance, since public transportation is often closed by the time I am coming home from a night out this really goes hand in hand since ubers are expensive and walking isn't always feasible.

Q4. What are the various steps that you take to come to a conclusion?

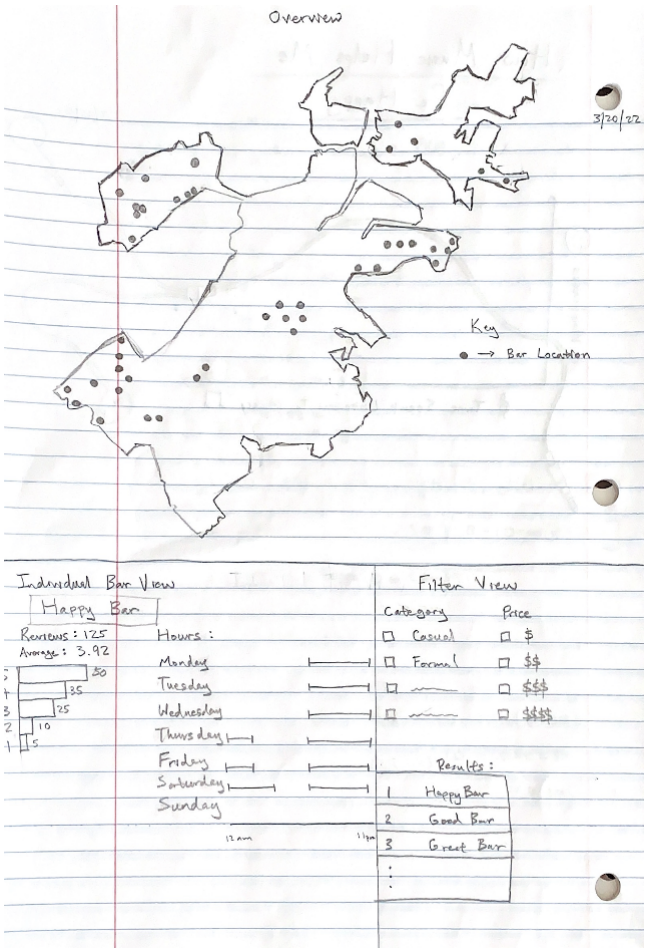
A4. Usually my final conclusion relies on how much I am willing to spend, how far the place I want to go out to is and where the crew I am going with wants to go as well. I usually do this in a process of elimination where I choose areas, then narrow down to bars within the area and have the dealbreaker either be decided by how cheap I can get a drink on that night or where my friends want to go.

Q5. If you have any ideas, what specific features would you like to see in the visualization?

A5. Some extra features that would be cool would be an option to rate some of the bars myself, so that I can remember what bars I liked and disliked. Other than that just the basic functionality i've talked about would be enough, I wouldn't want to complicate the ease of use.

Appendix D: Design Sketches

- **Individual Sketches**



Artist: Jason Zhang

Map

Task: I want to determine if there are suitable locations which will fit our busy vacation schedule. The visualization addresses this task by allowing users to visualize the bars which are in close proximity to locations they plan to visit during their vacation (beach, hotel, museum, etc.) In doing so, users can locate bars which will be convenient for them during their busy schedules.
Marks: **points** - represent bar locations accurately
Channels: **2D position** communicates actual geo-location of the bars displayed
Encodings: This encoding allows users to gain a high level overview of the locational distribution of bars in Boston, where they can quickly identify where large clusters exist as well as to clearly visualize the distance between various locations. By using points to identify each establishment as well as 2D position to model their geo-locations, users gain an intuitive understanding of the distribution of bars in Boston.

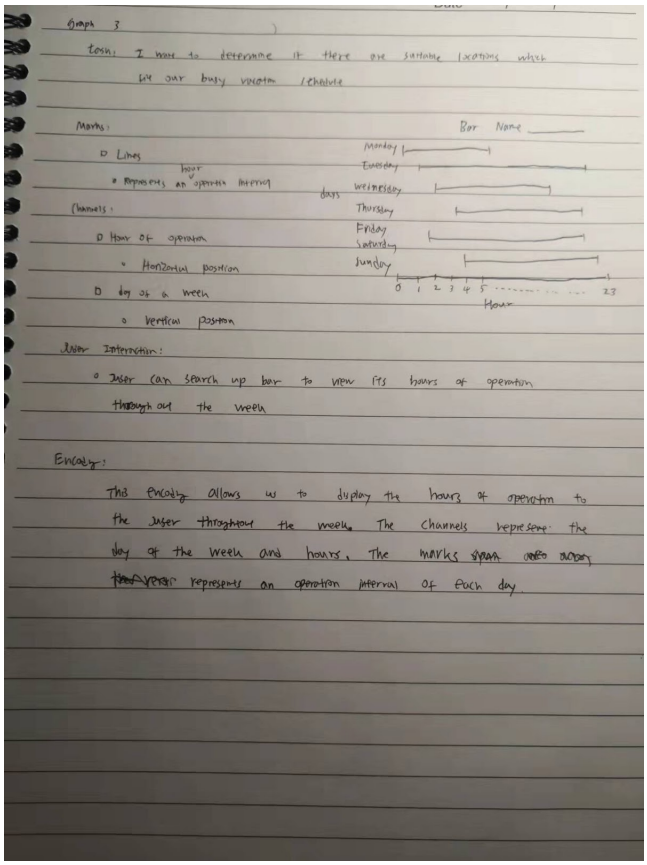
Individual Bar View

Task: I want to just look up a known bar. The visualization addresses this task by allowing users to look through highly detailed description and summary of the particular bar they are researching. They will have access to pricing, hours, reviews, and various other data that may be of interest.

Marks: **areas** - represent distributions well in a bar chart, **lines** - represent intervals well in a line chart
Channels: **vertical position** - represents rating value, **horizontal position** - represents count, **length** - represents time intervals
Encodings: This encoding enables users to quickly access and ingest relevant data about the particular bar they are researching. By using a bar chart which relies on areas for its marks and vertical and horizontal positioning to distinguish between categories, users can gain an accurate picture of what the review distribution truly looks like. In addition, the lines employed with the weekly hours/schedule allow users to quickly compare and contrast the bar's various hours across the week to identify which day(s) align best with their schedule.

Filter View

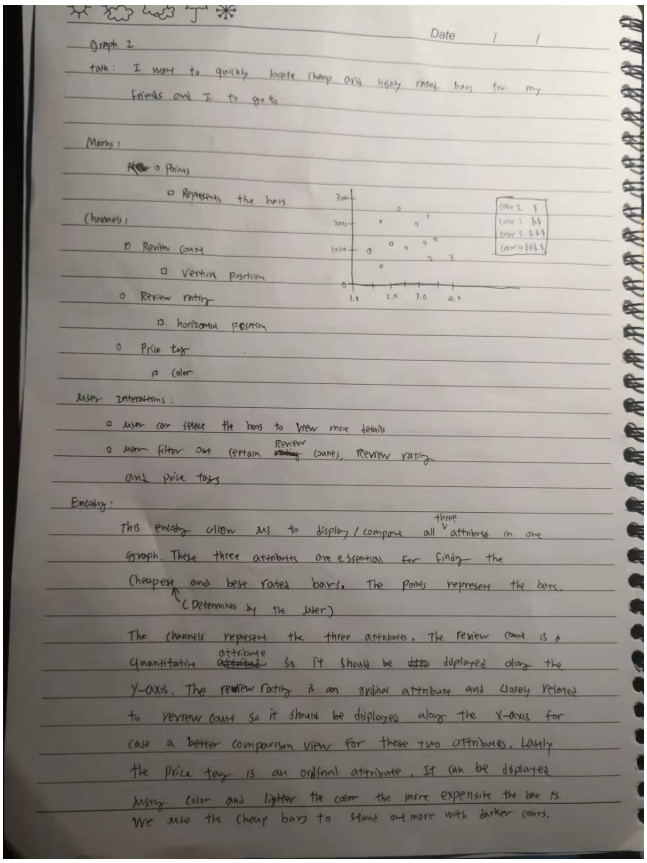
Task: I want to quickly locate cheap and highly rated bars for my friends and I to go to. The visualization addresses this task by allowing the user to filter out expensive bars so that they will only have bars in their acceptable price range to query through and make a decision. This will drastically reduce the amount of bars to consider and thus increase the search efficiency. This visualization will also work in concert with the map visualization to change the points displayed to reflect the particular bars that have been selected with the filter.
Marks: **areas** - represent the different options and characteristics that a user can select to filter through the available bar options in order to optimize their search
Channels: **vertical position** - represents the hierarchical ordering of bars in terms of their fit for the user, the higher the vertical position, the better fit the bar is for the user
Encodings: This encoding allows users to quickly narrow their available search options in order to optimize the bar searching process. By implementing checkboxes for filters, users can efficiently select their parameters and specifications to develop custom queries on the total dataset of bars in Boston. The vertical positionings in the hierarchical list of bars from the resulting query will allow users to first consider the best fit bars in order to further optimize the searching process.



Artist: Ningyu Chen

***FAVORITE* Hours of Operation View**

Task: I want to determine if there are suitable locations that fit our busy vacation schedule.
Marks: **lines** - represent operation intervals
Channels: **horizontal position** - hours of operation (ordinal), **vertical position** - days of week (categorical)
Encoding: this encoding displays the hours of operation of a specific bar to the user throughout the week. The reason that lines are chosen are the marks because this graph is displaying time interval data. Then the two x/y axes represent the two channels days of the week and hours of the day. This encoding is very simple to use and the user should be able to understand the graph at the first glance.



Artist: Ningyu Chen

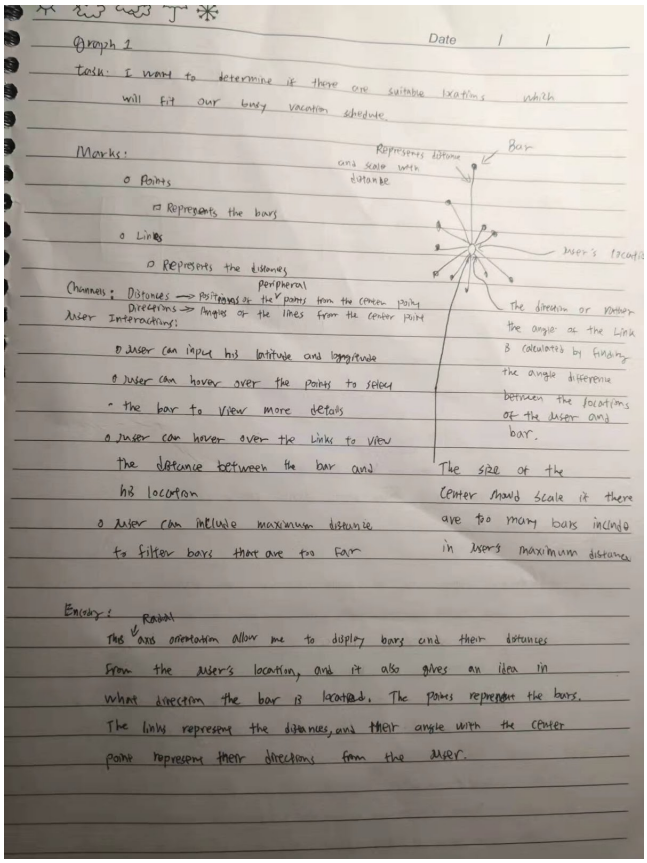
Price vs Review count vs Review :

Task: I want to quickly locate cheap and highly related bars for my friends and I to go to.

Marks: **points**- represent bars

Channels: **horizontal position** - review rating (ordinal), **vertical position** - review count (quantitative), **color -** price tag

Encoding: This encoding displays/compares all three attributes of the bars in one graph. These attributes are essential for finding the cheapest and best rated bars. The points represent the bars. The channels represent the three attributes. The review count is a quantitative attribute. So it is displayed along the y-axis. The review rating is an ordinal attribute and closely related to review count. So it is displayed along the x-axis. Lastly, the price tag is an ordinal attribute. It can be displayed using color and lighter the color the more expensive the bar. We want the cheap bars to stand out more with darker colors.



Artist: Ningyu Chen

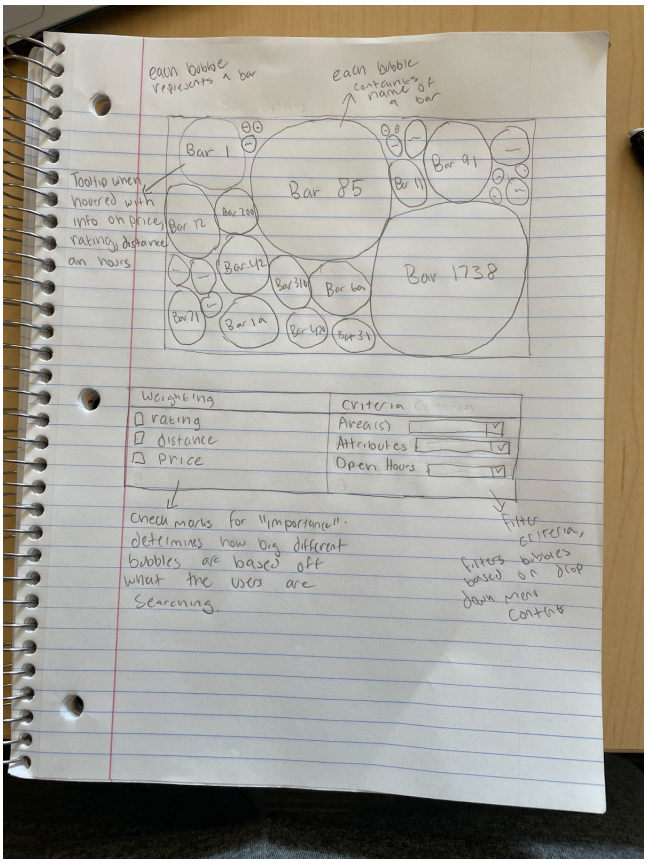
***FAVORITE* Location and Direction**:

Task: I want to determine if there are suitable locations that fit our busy vacation schedule.

Marks: **points** - represents bars, **lines** - represents distance vector

Channels: **position from the center dot** - represents distance of the bar from the user, **angle from the center dot** - represents the direction of the bar from the user

Encoding: This encoding displays bars and their distances/directions from the user's location. The points represent the bars. The links represent the distance vector. The length of the link scales with the distance of the bar from the user. The angle of the link represents the direction of the bar from the user. This graph gives the senses of distance and direction and is easy to understand.



Artist: Zoheb Aziz

Standard View

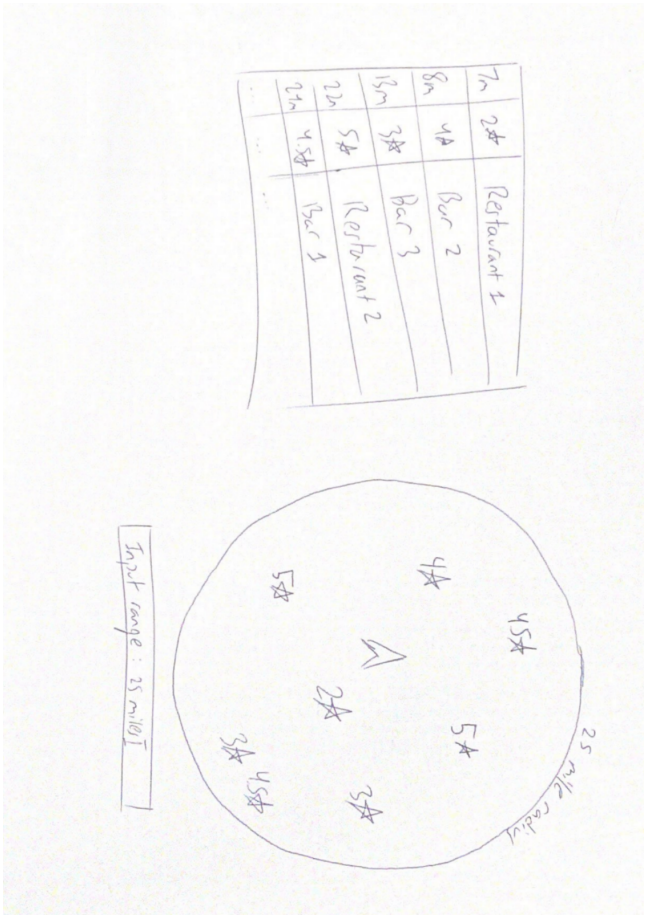
Task: Allows a full overhead view of every bar in the area. This view is for the user who doesn't quite know what they want to do for the night yet and would like to just explore possible options. This allows the user to have a simple unfiltered and unbiased view of every bar our visualizer has to offer and is to be used as the starting point in the search for a bar to visit.
Marks: Each bubble represents a different bar.
Channels: For this view all of the bubbles have the size since they are "on an equal playing field." Since this is just the high level overview of the bars in the area, placement is random and size/color is standardized. This is a very static view.
Encodings: The encoding here is basically a general overview of all the bars in the dataset. Hovering over each bubble will have a tooltip with further information on each bar, information that can be used in some of the other views to narrow down the search. This high level overview of all the bars is for beginning the search for a bar and having all of the options cleanly available without an overwhelming amount of information.

Weighted View

Task: This view allows the user to be given bar recommendations based on some criteria they choose. This is to begin the narrowing down process and see what type of bar is appealing to them in a certain time. The view changes from essentially a static list of all bars to some bubbles being bigger than others based on the criteria the user selects. The user then may hover over the bars they are more interested in to learn further through the use of tooltips. The sizes are chosen comparatively to show a relationship between each other.
Marks: Each bubble represents a different bar.
Channels: The size of each bubble represents how well the bar fits the user criteria. The user has the option of selecting attributes that they find appealing and the bars that best fit the criteria will be the biggest while those that do not as well will shrink in size. The position of each bubble will shift to give the best overall viewing UI, but every bubble will still be on the screen at once.
Encodings: The encodings here are similar to the other view. Hovering over will give a tooltip but the tooltips adjust in size to the weighting on of the bar it represents based on the users inputs. Users are directed towards selecting the bars that fit the criteria they are searching for. Selection of bubbles will also highlight them as a way to "save" for remembering in the future what bar they are currently considering. This medium level overview makes bars that the user is more likely to choose, more visible than the others.

Elimination View

Task: This view is the final step in narrowing down the search for the perfect bar. This view will eliminate bars that do not fit a criteria given by the user. This criteria follows similar attributes as the ones that are used in the weighted view and this is because the user is meant to get ideas from the weighted view and use them with the elimination view to narrow down the search. This view features dropdown menus where the user will select the criteria they wish to keep or eliminate to narrow down the bars.
Marks: Same as in the other views, each visible bubble represents a bar that has not been eliminated.
Channels: This view may be used in accordance with the other views, therefore; the size of each bubble will still represent how recommended it is based on the weightings given in the weighted view. The bubble's visibility is dependent on whether it fits the criteria given in the elimination view. Color is still standardized and used for visual appearance sake.
Encodings: Since this view is compatible with the weighted view, the ecodings here also include that of the last two views with tooltips and weightings if those apply. The extra channel of visibility of filtered out bars eliminate the tooltip hovering and visibility of those bars that do not fit the criteria that the user suggests. The ability to highlight the bars remains and overall visibility for bars that are recommended is further increased in this view. The view dynamically renders with chosen options and reorders itself to offer the best visual experience for the user. The criteria menu offers drop down menus for categorical selection similar to those offered in the tooltips to be used as a natural progression from both the standard view to the weighted view to eventually narrow down to just a few options that the user is looking for.

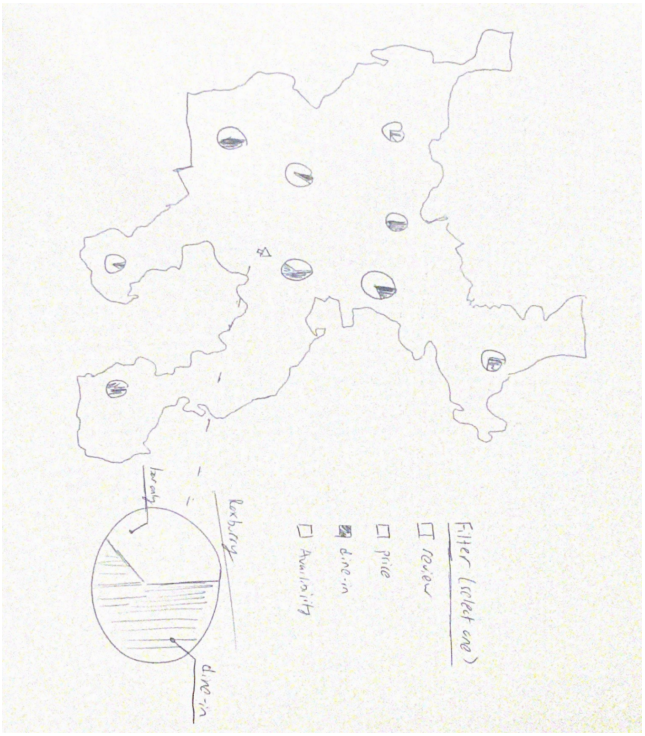


Artist: Marco Tortolani

Radar View

Task: This view fulfills the task as requested by the user “college student” who prioritizes availability and nearness over finding the perfect bar. In order to meet this demand, we present them all bars in a local range and give them the minimal key information (distance and rating) that they’d need to make a quick decision.

Description: The radar visualization consists of a center mark and outer marks. The center 2 dimensional mark contains a fixed shape of a pointer, fixed size, and fixed position in the middle of the radar. The outer marks are 0 dimensional points, visualized through channels of variable position (both horizontal and vertical), and the shape being shaped as an integer representing a rounded average rating for the bar. In addition, the accompanied list on the left provides more information in a table format



Artist: Marco Tortolani

Summary by Area

Task: This visualization best serves the task to identify and consume as requested by the user group “couple looking for an available date area”. The ability for a user to filter by date and find areas that have the most availability is important for making decisions about where opportunities can be found based solely on this one criteria.

Description: This visualization consists of a large static 2 dimensional mark through the channel of a map shape. Nested within this map are multiple 2 dimensional marks with the 2 axis position channel shared with the map. These marks will be channeled as circular colored pie charts, representing filtered data about bars (such as review, price, or availability) categorized by city. Ideally, hovering over a pie chart will display it in an enlarged area reserved with a faceted zoom view.

Artist: Marco Tortolani

***FAVORITE* Map Screening**

Task: This visualization will work best for a normal user that wants to rank bars in an area. The ability to zoom into a certain area will let them make direct comparisons, and the statistical divide in price as determined by the key allows them to see both how bars compare in their area and how that local area could compare to others.

Description: This visualization consists of a large static 2d mark through the channel shape of a map. Within this map there should be a mark for each bar within our database. We will keep the channel of these marks as 2d shapes, varying depending on the price of each bar. Ideally we could implement filtering by other categories at a later date. Additionally, we have a separate side-by-side faceted view. Although there is no shared encoding, the zoom in on the data (with the same marks and channels) gives a far better view of a small subset of local bars that would be impossible to digest from a map of all of Boston.

- **Favorites and Group Sketches**

We selected the Map Screening view, the Hours of Operation view, and the Location and Direction view as our favorites.

We selected the Map Screening view because it offers the most comprehensive, high-level overview of all of Boston's bar options for the user, and invites the user to further explore and query about particular areas or individual bars that will be most suitable for them. In addition, it incorporates brushing features which we can further link to pie charts/histograms to offer the user further insight into the selected area. This view supports the task of gaining insights on a group of bars in a selected area, and it supports data types of geo-location and numerical/quantitative price. Its visual encodings are effective because the map is very intuitive for humans to look at and understand geo-locations and positionings of bars relative to one another, and it also offers a clean and high-level view of the data which the user can easily look more closely into with the interactive features.

We selected the Hours of Operation view because it clearly and simply informs the user on the weekly hours of operation of the particular bar in a manner that permits easy comparison and intuitive understanding. This view supports the task of determining whether a particular bar has a suitable schedule for a user's busy vacation plan, as it allows for a comprehensive view of weekly hours of operation where a user can clearly see if the bar can accommodate them at a particular time. It supports data types of time intervals. The visual encoding is effective because it is so simple and easy for the user to understand just at a glance.

We selected the Location and Direction view because it offers similar but more detailed information as the higher-level map view, and it can give users more accurate bearings on how they would be able to walk/commute to their target location. This view supports the task of determining how to travel to a particular location, as well as to gain bearings on where nearby locations are so that hopping over to the surrounding establishments is also made simpler. It supports data types of geo-location. The visual encoding is effective because it provides an intuitive and focused view of the particular location and nearby establishments as well which a user can easily interpret and gain information from.

Coordinated View Visualization Tool

For the visual encoding choices, we decided to utilize the Map Screening view because it offers a simple and intuitive high-level

summary of the dataset. In addition, we used different shapes to denote various price ranges, and added additional visualizations that are linked to the brushed areas (scatterplot of average review vs. review count and a zoomed image of the selected area with pricing data shown. The hours of operation view was combined with the bar chart showing the distribution of reviews to give a comprehensive look of an individual, with further interactivity enabled when a user clicks on a bar in the chart. There, a table of written reviews will be accessible so a user can read the reviews left by previous customers. Bar charts are a good choice here because they accurately display a distribution and the custom line chart for showing hours of operation is similarly intuitive and quick to convey information. The location and directional view was implemented to give a more detailed view of a selected location with respect to the establishments closest in its proximity. Once again, this visualization is simple and intuitive for the user which achieves our goal of informing the user in an easy and low-effort manner. The lines and geo-spatial visualization are easy to interpret and visually appealing to the user. These visualizations support the data and tasks previously identified because they display all relevant data (geo-location, hours of operation, pricing, reviews and review distributions, etc.) Furthermore, it allows for a high level overview of all options, filtering and querying to minimize the sample/search space, and various descriptive and summary statistics that aid users in making informed and efficient decisions.

Appendix E: Digital Sketch

If a user is planning on locating a cheaper bar in the north of Boston that they can visit with friends on a Friday night, they can begin exploring with the brushing tool in the map overview and observing the pricing data displayed in the linked zoomed view of their brush selection. †he user can also immediately filter out the expensive bars so they won't appear in the queries and add unnecessary time to the searching process. Then, the user can examine the scatterplot to determine which of the locations is the highest rated and/or which locations have the most reviews. From there, the user can click on the most appealing to get more specific information such as hours, review distributions, and written reviews to further inform their decision. The hours will be very helpful in determining whether the particular location will actually be open and available on Friday night. Ultimately, they can also consult the location/directional visualization to identify nearby bars as secondary options that they can also potentially visit in case there is some issue with the current bar that they are considering. The user can continue to click on those neighboring establishments to view their in-depth data as well in order to make the best choice.

Appendix F: Usability Testing

- **Presentation**

Introduction: Our visualization tool is intended to aid users of all kinds in browsing, discovering, and identifying bar establishments to visit in Boston. While Boston is a city rich with diverse bar options for all types of people, it can be very challenging to locate or decide on the particular establishment that most fits one's needs. We recognize that one's desired location involves an abundance of factors such as location, price, style, etc. Compiling all of this data and then drawing comparisons between various establishments is not an easy process, and it can become disorganized and overwhelming. It is common for people to spend lots of time searching for their target location, which they could have instead spent at the bar having a great time. We want to minimize the decision making time by simplifying and streamlining the process. Our visualization addresses every step in the procedure in order to optimize the user experience. It showcases comprehensive data on Boston's bars, which is acquired through the Yelp API. By selecting Yelp as our data source, we can guarantee that our data is trustworthy and reliable. The data includes information on each bar's price category, rating, review count, geolocation, hours of operation, and the link to the bar's Yelp page for further exploration if desired.

Tasks:

- **Task 1**: Identify the highest rated bar that is in the cheapest price category
  Purpose: We are testing this task because it is a common objective for college students to want to save money and visit a location which will be cheap, but one that will afford a great experience.
  Expected Outcome: We expect that the user will be able to quickly identify the target location (3JHookah) by using the filter with the map and the scatter plot visualizations.
- **Task 2**: Identify the bars which have the highest rating. Report their names and price categories, and use this information to determine which is the most appealing to visit.
  Purpose: We are testing this task because it showcases the ability of our map and scatterplot to allow for an ease of comparison between locations. In addition, we want to see if our visualization truly allows users to

make decisions that they believe are strong and informed.
Expected Outcome: We expect that the user will be able to quickly utilize the map and/or scatterplot to gather this information in order to make an informed decision on where they would want to visit. The target locations are: GrandTen Distilling ($$), The Quiet Few ($$), Distraction Brewing Company ($$), and 3JHookah ($).

- **Task 3**: Identify the westernmost bar on the map and report its name, price category, and rating. In addition, determine whether this establishment will be open on Friday night.
Purpose: We are testing this task because it highlights and takes advantage of the strength of our map visualization in intuitively representing geolocational data.
Expected Outcome: We expect that the user will quickly be able to leverage the map visualization to identify the easternmost location and then use the time chart in order to quickly understand the bar's hours of operation in order to determine whether it will be open. In addition, we expect the possibility of the user accessing the bar's website through our built in link in order to acquire more information. The target location is: Rail Stop Restaurant & Bar ($$, 4.0). It will be open on Friday night until 10:00pm.

- **Results**

After completing our testing, we didn't come across any major issues with our visualization tool. However, there were a few minor issues and improvements that were suggested. The tooltips were not extremely responsive and in some cases were lagging and struggling to render. In addition, it was not immediately clear to users that our checkbox filtering system operated on an "or" basis, where any bar matching at least one of the criteria would be displayed. It was recommended that we overcome this by initially rendering all data points and then to filter based on checkbox input. Overall, the participants really liked the multiple views and coordination between them, and they appreciated the wealth of information about the bars that we were able to communicate. The participants disliked the fact that we did not have many annotations or documentation on how to use the visualization, so that is an area where we can add additional resources and content to rectify.

Task Results:

- **Task 1**: Identify the highest rated bar that is in the cheapest price category
Result: This task illuminated the intuitive and functional nature of the scatterplot as well as the map views. The users were able to use both to quickly obtain the information to answer this question.
Analysis: We were looking for the user to be able to quickly identify the target location (3JHookah) by using the filter with the map and the scatter plot visualizations, and as this was achieved, it tells us that our visualization performed to the standards that we expected. The users were able to use the filters to quickly identify the correct bar. These test results indicate that a change in our design is not necessary as the users were able to use the current design adequately. However, there was the suggestion made that we could add titles or labels such as a simple text stating the map is of Boston for those users who are unfamiliar with the area.

- **Task 2**: Identify the bars which have the highest rating. Report their names and price categories, and use this information to determine which is the most appealing to visit.
Result: This task again illuminated the intuitive and functional nature of the scatterplot as well as the map views. The users were able to use both to seamlessly compile the information to answer this question using both the checkbox filtering function as well as by examining the scatterplot.
Analysis: We were looking for the user to be able to quickly utilize the map and/or scatterplot to gather this information in order to make an informed decision on where they would want to visit. The target locations we expected them to identify and choose from were: GrandTen Distilling ($$), The Quiet Few ($$), Distraction Brewing Company ($$), and 3JHookah ($), and they were able to identify these exact locations. Furthermore, they decided that 3JHookah was their most appealing location based on the pricing, reviews, and location. This again tells us that our visualization performed to the standards that we expected, and thus, the test results indicate that a change in our design is not necessary as the users were able to use the current design adequately.

- **Task 3**: Identify the easternmost bar on the map and report its name, price category, and rating. In addition, determine whether this establishment will be open on Friday night.
Result: This task illuminated the ability of the map to show geographical data as well as the strengths of the linking between the map and the time chart. The users were able to use both to quickly obtain the information to answer this question, identifying the location farthest east and then clicking it to access the location's hours of operation in the time chart..
Analysis: We were looking for the user to be able to quickly identify the target location (Rail Stop Restaurant & Bar ($$, 4.0)) by analyzing the map and then requesting further details by clicking on the points. After analyzing the associated time chart, we wanted the user to be able to determine that the bar will be open until 10:00pm on Friday night. The users used the map to find the correct location and identified the closing time correctly, so it tells us that our visualization performed to the standards that we expected. These test results indicate that a change in our design is not necessary as the users were able to use the current design adequately. However, it was noted again that it would be helpful for the map to initially render with all the points showing.

Overall, we don't have too many modifications to perform on our visualization tool. As the users were able to successfully

accomplish each of our tasks, the fundamental functionality of our visualization is all working as desired. However, there were some suggestions that we should add labels, titles, and annotations to make it more clear to the user what they are presently looking at. We will modify our visualization by displaying the currently selected bar's name as a header in order to make this very clear to the user. Furthermore, we are also planning on linking each bar's unique Yelp link to this header, so users can easily access more in depth information if they desire to do so. In addition, we decided to initially render the map with all points shown, and to then filter down options based on user interaction with the checkboxes. This is in order to not present the misleading impression of missing/lacking data. Following the details on demand principle, users are able to query through all the available locations to request specific information about particular establishments.