CS-171 Project Proposal

Background and Motivation. *Discuss your motivations and reasons for choosing this project, especially any background or research interests that may have influenced your decision.*

Early on we decided to do something related to transportation. Jack Birger works at a transportation consulting company and had the idea of using real-time bus data through the OneBusAway application and MTABustime. We explored ideas of combining MTA data with weather and crime data, but struggled to come up with specific questions that the data, and subsequent visualizations would answer. Therefore, we began to explore new datasets, looking for ways to combine interesting aspects of transportation with data from another source. We sought data that has both an excitement factor and that could be utilized to answer meaningful questions. Along the way we came across a very cool visualization that juxtaposes the London Tube system with second languages spoken, uncovering insights into the cultural fabric of the city. We think it would be interesting to apply a similar approach: using MBTA locations as a basis for Yelp data to explore metro Boston and its culture by mapping the constellation of restaurants within walking distance of train stops by ethnic category. Utilizing this data we aim to expose the clustering of restaurant categories (Italian, Vietnamese, etc) around particular MBTA stops.

Other datasets and we considered using and other visualizations we took for inspiration:

Boston Crime Data

US Climate Data

National Weather Service Forecast

Project Objectives. *Provide the primary questions you are trying to answer with your visualization. What would you like to learn and accomplish? List the benefits.*

Primary question:

What cultural patterns exist around metro Boston that can be exposed through the visualization of restaurant clusters located near MBTA train stops?

Secondary questions:

Which filters can aid in seeing these patterns? Which filters are most accessible and necessary for restaurant goers?

At a practical level, Which filters can aid in selecting a restaurant near a particular stop?

Learn and Accomplish:

We are eager to provide a novel method for the exploration of food and culture around the city of Boston. We hope to allow users to gain new insights into cultural patterns as they relate to the combination of food and culture in relation to public transportation and settlement patterns in the metro area.

Data. *From where and how are you collecting your data? If appropriate, provide a link to your data sources.*

Primary data sources:

Yelp:
Link: https://www.yelp.com/developers

Format: JSON

Sample Data:

{ u'categories': [[u'American (New)', u'newamerican'],[u'Pizza', u'pizza'],[u'Cocktail Bars', u'cocktailbars']],u'display_phone': u'+1-617-500-3055',u'id': u'russell-house-tavern-cambridge',u'image_url': u'http://s3-media4.fl.yelpassets.com/bphoto/M7YViqqBZM7Pl43JSocI1Q/ms.jpg',u'is_claimed': True,u'is_closed': False,u'location': { u'address': [u'14 JFK St'],u'city': u'Cambridge',u'coordinate': { u'latitude': 42.373122,u'longitude': -71.119703},u'country_code': u'US',u'display_address': [u'14 JFK St',u'Harvard Square',u'Cambridge, MA 02138'],u'geo_accuracy': 9.5,u'neighborhoods': [u'Harvard Square'],u'postal_code': u'02138',u'state_code': u'MA'},

Data Processing. *Do you expect to do substantial data cleanup? What quantities do you plan to derive from your data? How will data processing be implemented?*

Yelp data vis will require multiple requests because Yelp's API limits search results to 20 per query. Also, requests of different types will need their data stitched together. Data will need to be collected and transformed into objects containing lat/long, categories, and ratings. We will be exploring two paths to deliver this data: (1) getting the data real-time from yelp via their API and (2) programatically collecting their data, decomposing, and storing it in a RDBMS so that it can then be recomposed and sent to the client via a lightweight REST server (such as Django REST Framework) as needed.

Visualization. *How will you display your data? Provide some general ideas that you have for the visualization design. Include sketches of your design.*

The primary display will be a stylized map of the MBTA train system in metro Boston. Users will be able to see the density of the categories of restaurants (American, Chinese, etc) that make up the majority at each stop (if there was a stop in the North

End it would be Italian, for example) or filter to see the prevalence of a particular category across the system (show density of Vietnamese across the system).

Must-Have Features. *These are features without which you would consider your project to be a failure.*

- Ability to layout stylized MBTA train map
- Ability to place relevant Yelp data around each of the T stops
- Ability to filter Yelp data by several criteria (such as category and rating)

Optional Features. *Those features which you consider would be nice to have, but not critical.*

- Ability to compare multiple T stops in different filters.
- Adjust size of the MBTA stops to reflect the total number of restaurants at a given stop.
- Adjust size of the MBTA stops to reflect average cost or ratings across restaurants at a given stop.
- Provide MBTA service alerts/status that pertain to the location.
- Ability to add MBTA real time data

Project Schedule. *Make sure that you plan your work so that you can avoid a big rush right before the final project deadline, and delegate different modules and responsibilities among your team members. Write this in terms of weekly deadlines.*

(Note to team: Will want to go through and assign dates to milestones, etc.)

April 3-6: To layout an outline/stub for the entire visualization. This will allow us to have a plan for how everything will interact and work together. This portion will be a collaborate effort by the team.

April 6-15: Once the outline is complete, we will divide different portions and views to different individuals. We will work on these primarily independently, but will use each other as resources if we get stuck or for any other purposes.

April 15-17: Combine individual portions to create a working prototype. Reformat process book if necessary.

April 17 Milestone 1: Complete data acquisition, have data structure ready. Create working prototype. Turn in process book.

April 17-30: Update user interface with additional filters and seamlessly combine each individual part.

April 30-May 3: Complete the process book, create screencast, put finishing touches on website.

We are using Asana to organize our project work.

Submission Instructions

Submission will be handled through github. All teams must use a single shared github repository. If we cannot access your work because these directions are not followed correctly, we will not grade your work.

To submit your work ***you must fill out the form that is part of Homework 4*** *where you specify your project URL in ssh format.*

Store the following in your github repository:

- *Code - All web site files and libraries assuming they are not too big to include*

- *Data - Include all the data that you used in your project. If the data is too large for github store it on a cloud storage provider, such as Dropbox or Yousendit.*
- *Process Book- Your Process Book in PDF format.*
- *README - The README file must give an overview of what you are handing in: which parts are your code, which parts are libraries, and so on. The README must contain URLs to your project websites and screencast videos. The README must also explain any non-obvious features of your interface.*