

Data collection is integral to our final project. Since it is statistical analysis and further predictive modeling, data quality is everything. To ensure that, we will be paying close attention to the source of the data. Till date, we have only used credible sources such as the Boston Mayor's website, Massachusetts.gov, etc. Our dataset of coordinates of bike sharing stations in and out of Boston will come from BlueBikes, the authorized bike sharing station coordinator in the area. After our introductory meeting with Jonathon Fanning from the Boston Museum of Science on 24th April, we have been told of possible data and resources influx from the sponsor HubLuv. HubLuv is a philanthropic organization that partners with key players in Biking in Boston to advocate and implement better biking infrastructure and celebrate biking for all. We will also be potentially using the MBTA V3 API to extract station coordinates. Another important point to remember would be the existence of previous ICONS projects in the area. We have contacted the student that worked with Boston Museum of Science last year on a similar project and they have agreed to provide insights, if needed.

We are not concerned with the sample size as much as we are with the accuracy of the sample. Even if we consider only 3 subway stations or let's say 100 bike sharing stations, that is sufficient information if we can apply the mathematical model correctly. The information shared by a single data point is tremendous. The coordinates of the busiest subway station in Boston tells us how many people commute through it on an average day, what stations they commute from, what is the 30-minute window around the busiest and nearby subway stations. When overlapped with the socioeconomic area maps, this will give us a lot of information into transit access inequalities.

To analyze our data, first we will collect it in a spreadsheet. Once we have the necessary data, we will transform all of our data points into vectors. To find which people are in the 30 minute window, we will find the most transited station in the Boston subway system and set that as our goal. Then, we will find the estimated travel distance over every line to that station using the V3 API provided by the MBTA. That will give each station a given "target time", which is the time that people that use that station have to commute to that station before they exceed the 30 minute window. From that, we will draw a circle with radius equal to the distance the average person bikes in the remaining time. We will then identify all of the bike stations within that circle, and measure the time it would take to bike from the bike rack to the station by dividing the distance by the average biking speed. From there, each bike station will have its own "time rating", and we will draw circles around each bike station with radius equal to the distance the average person travels in the time remaining to still be in the 30 minute window. Finally, we will draw a circle around each station with a radius equal to the amount of time people walk in the time remaining in the 30 minute window

after taking the trip to the central station. This process will be done once for each station, and we will thus obtain all of the areas of Boston within the 30 minute window.

The focus from the data that we collect will be specifically based on locations where there are no bike stations or bike sharing stations or the distance to a bike station or bike sharing station is very far for someone to walk if they need to use the bikes that are provided. This is because these locations do not fit within the optimal 30 minute window. There are other aspects that will be considered as well such as the reasons as to why certain areas have less bike stations. It could be based on factors such as race, socioeconomic status or the terrain in that area. Yet by considering these issues it will be a way to view for instance the racial disparities in the greater Boston area around cycling. It can also be useful to then go to these areas and try to understand what type of disconnect has been created and what else could be done besides just putting in bike stations, or bike sharing stations. To then encourage people to use cycling and incorporate it into their lifestyles.