**Kamile Demir**

# CSE 332 Lab 5: Dashboards

Datasets: NYC Open Data on Tree Population, NYC Open Data on Air Quality & Pollution

---

The semester goes by so quickly! At the beginning of the semester, I remember browsing data sets and becoming intrigued by the NYC Open Data sets on Tree Populations and Air Quality. I decided to focus on the relationship between these data sets and see if the amount of trees affected the pollution amounts in the air. My hypothesis was that the more trees there was in a borough, the better the air quality, thus smaller amounts of elements/pollutants in the air.

I decided to divide this data set up by visualizing not only the data set's relationships together, but also separately; thus, brushing and linking was absolutely a great way to see these visualizations.

I created a bar graph with the amounts of trees in hundred thousand per borough; it seems predictable, with Manhattan having the least amount of trees out of the 5 boroughs, and Queens had the most.

Another bar graph shows the amounts of elemental carbon in each borough; for this graph, I decided to only focus on average elemental carbon rates to truly focus on the relationship of air pollution and the number of trees; I did, however, make sure to utilize the rest of the air quality graph by comparing them all in a 5 x 5 Scatter Plot, which I will discuss later.

The first interesting data story to point out is the main one I sought out to discover when I first chose these data sets, and to properly observe this I linked both bar graphs so that when the borough bar of one graph is highlighted, it will highlight the corresponding bar in the other graph. These graphs were a wonderful way to explore this simple yet important question of whether or not the amount of trees present affects the air quality of the area, and my conclusion came to be yes. While Manhattan has the least amount of trees in the Tree Amount Graph, it is highlighted to show the highest mean amount of elemental carbon present in the air. Meanwhile, Queens and Staten Island have the lowest amounts of elemental carbon present in the air, with the highest amount of trees. The distinction between these numbers, presented with the linking and highlighting of these graphs, clearly supports the argument!

To further explore these data sets, I linked these bar graphs to a 5 x 5 Scatter Plot, which contained values from all air quality attributes, including nitrogen, FPM, nitric oxide, ozone, sulfure dioxide, and carbon dioxide. When a borough bar is hovered over in any of the two bar graphs, all of the corresponding air element data points for that borough is presented on the scatter plot. This is nice in that the Air Quality bar graph specifically shows elemental carbon amounts, but if viewers were even more curious to find out what other element rates were and their relationships, they could use this scatter plot. For ease of use, all of the data points in the scatter plot are hidden and are only unveiled when the corresponding bar borough is hovered over. However, I added an additional menu on the bottom to make navigation even easier; I divided the data points on the scatter plot not just by borough, but also by the type of element they were; thus, if a user were to only want to see FPM rates or any other element, they'd simply click the corresponding menu button to unveil it. With this we could discover more data stories between specific data elements, comparing two, to three, to four, to all data points whenever we want. This style of brushing and linking promotes the audience to explore the data through multiple perspectives rather than just one; it eliminates bias, for they may discover more results than the ones they were looking for.

I also included two parallel coordinate graphs, one of which is derived from the main one, to further observe the air quality data in other perspectives. This main parallel graph is brushable, and thus users can select which element ID they'd like to observe.

Although I was looking for a specific result (whether tree amounts relate to air quality) I did discover a lot more through my dashboard. What I discovered was the big range that the elements were present in, and how a said amount of one element may not be the same as another in terms of how impactful it is on air quality. I also realized that there were stark differences between these IDs, which showed that my data was indeed very clearly distributed into their individual categories of elements and boroughs. Furthermore, the data in the 5 x 5 scatter plot also showed the way in which these elements were all condensed towards the center of the graph, although there were some outliers. This showed me again that the range of the elements was very different, but also, in fact, similar in that they were all very distinct in the numbers they corresponded with to their boroughs.

I did not expect to discover so much through these two data sets, and I had never realized that there could be so many more perspectives of data to just one simple question! These data perspectives help reduce bias in data, bias in results and show us many more answers than the ones we seek. Answers lead to more questions, and hopefully, this visualization helps me think of more questions to ask in the future!