Mini Project 1 Documentation: Creative and Design Process

Dana Bullister, ARTG5330 Advanced Study: Mini Project 1, 2020-10-25

The motivation behind my submission was that I wanted to explore data related to how a particular shared resource provided value for and integrated with the culture of a community. I am especially interested in shared resources at the local/city level, since I hope to one day become more involved in municipal policy making.

I decided to explore data related to the Bluebikes shared bike program, which exists within the greater Boston area. This data set appealed to me as Bluebikes provides community value in some interesting and important ways: It facilitates an accessible, equitable, sustainable, and healthy means of transportation within the greater Boston area and also promotes transport to and from more local venues, facilitating intra community interaction. In addition, a survey of the Bluebikes website showed that there is plentiful, multidimensional, and relatively consistently formatted public data related to Bluebikes (https://www.bluebikes.com/system-data). This makes it convenient to explore and analyze.

I decided to try and communicate the scale of Bluebikes ridership in an accessible way as well as the nature of how people tend to engage with the service. The Bluebikes data I found listed rides taken by day, time, duration, and some other variables, which suited this purpose. Fortunately, because the Bluebikes data was relatively consistently formatted, it did not require extra analysis or external cleaning. I did use JavaScript later on to parse and format the data, however.

I was interested in how people engaged with the service on a daily basis and how this interaction changed over the natural cycle of a year. I decided to encode this interaction as a scatter plot, with each point representing a single ride whose x-coordinate was its start time and y its duration. This way I could show how rides tended to cluster on both of these dimensions as well as how the two characteristics interacted. Scatter plots are also useful in that they offer the possibility to encode many potential characteristics at once through position, color, point size, style, and other ways. This would allow me flexibility to experiment with the data and decide how might be best to represent it.

I decided to look at data by month over the course of the year of 2019. This is because 2019 is the most recent full year, which would provide a great summary of rider patterns. It also occurred before the start of the COVID pandemic, which undoubtedly threw off typical patterns of engagement. For this exploration I wanted to focus on more generalizable patterns of how communities might be expected to engage with a service like this under normal circumstances.

I decided to look at the nature of rides on an average day for each month of the year, since the number of points in such a graph would be descriptive without being overwhelming. I did this by taking a random sample of all rides that occurred over each given month, ensuring the total

subset corresponded to the number of rides that would occur in an average day within that month. This was pretty trivial to do and didn't require R; I simply used Notepad++, my text editor of choice, to do some simple regex that grabbed every 30th row of the csv containing the entire month's rides (or every 28th/29th/31st row, depending on how many days were in that month). Upon doing that the CSV was ready to be parsed by D3.

In addition to parsing the datetime values into objects workable within D3, I also converted "trip duration" values to minutes. I also decided to filter out points falling outside the boundaries of the convenient axes ranges to neaten the chart and focus on overarching trends. In addition to representing ride start time and duration, I also tried representing other characteristics such as rider age. However, I did not see any useful trends related to these other characteristics; the points simply looked like a fairly consistent blend of different ages/different genders/etc. that would not reveal any notable patterns of interest. Therefore, I decided to keep the visualization simple by using just one color.

To explore how patterns in engagement might sustain or differ from each other at different points in the year, I decided to encode ride data also as a small multiple of scatter plots that each showed an average day's data for a different month. This would enable comparison of high-level trends related to level of engagement and temporal patterns. I started with a larger look into a single month's data to demonstrate the scale of Bluebikes rides before then "zooming out" to enable a look at the context of all other months. This would facilitate a multi-scale exploration.

As far as other choices in design composition, color, and typography, I chose to represent each data point in blue to fit with the Bluebike theme. I also emphasized a large title, "A City In Motion," which would emphasize the motif of motion and the way Bluebikes activated the local community. In addition, to help symbolize the idea of motion I decided to animate the scatterplot in such a way that each point would move to its y-coordinate in the order of its start time from the beginning to the end of the day. This would also help show the scale of the sheer amount of motion generated by the service. Sans serif Helvetica font kept the visualization simple and accessible.

My design enabled the viewer to grasp in an accessible way the scale of Bluebikes daily ridership over the course of a year as well as how this differed depending on the season. I ensured that I summarized key takeaways of the data set below the small multiples design, including the sustained patterns of rides on a daily basis throughout the year and subtle differences between months. I ended by connecting the data to the real-life significance of Bluebikes in the lives of those in the community, who all ultimately benefit from its scale, convenience, and benefits in health, sustainability and equity.