

I was exploring the dataset and looking at the distribution of the variables. Apart from the geolocation and labels, the node data doesn't have any useful information. However, the node data combined with the edge data can be a lot of fun. I am dividing this visualization task into three sub visualizations.

1. A spatial visualization
2. A temporal visualization
3. A relational visualization

The idea here is to use all these three to explain the complexity and data flow of the city. With that in mind here is what I propose.

Spatial Viz

The overarching question I will try answering here is 'Which age groups frequent which part of the city?'. The way I am trying to achieve this is by plotting a point for each station in a geolocation based map. The bucketing of age will be based on the classic cohorts of Baby boomers, Gen Y and Millennials. Please see the distribution of age attached to this.

The motivation to do this is to explore the possibility of identifying a clear indication of age group concentration in the city. This can be achieved by normalizing the numbers I report by station, adjusting for bias of reporting and coloring the zip codes based on the numbers.

Temporal Viz

'How many bikes are being used by Casual and Registered users given a time of the day?'

The idea here is to use time as a surface and distance. The distance is the time span of 24 hrs and surface will be split by user type. The data is aggregated per hour to negate sparsity issues and final plot is a stacked surface temporal plot.

This viz can be used to understand the peak times of movement by day of the city. Also, this is super helpful in devising strategies around peak hour pricing, incentivizing registration, demand forecasting etc.

Relational Viz

When I was exploring the dataset I saw some very unique relationships between gender, age and duration of ride. But I wanted to explore a question which a dataset like this should be able to answer.

'Does people use bikes just to cross the river?'

I eyeballed the location of the stations on each side of the Charles river. Once this classification was achieved, I plan to use a radial network plot known as the chord diagram using the [chorddiag](#) package.