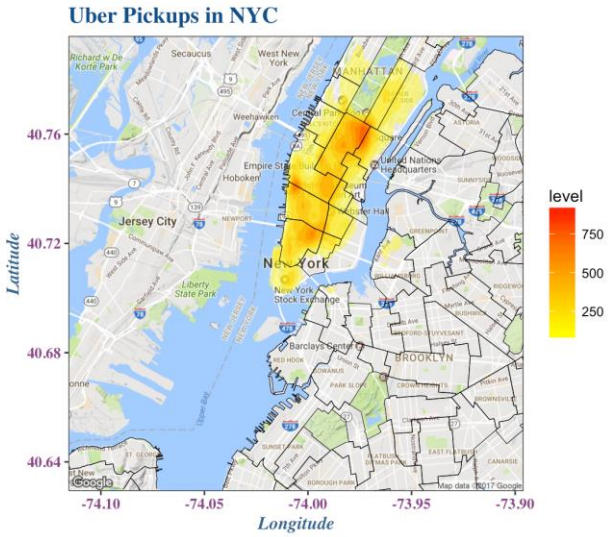
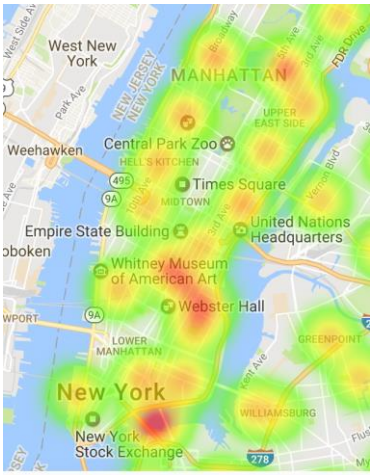
Team 9 Datathon Submission

1. Topic question – Are neighborhood demographics a useful indicator for which transportation services are used more commonly? This is important because we can identify the target audiences for each service and how to better suit their needs. The datasets used for answering this question were the datasets on trip pickup and locations, zones dataset, and demographics dataset.
2. Non-Technical Executive Summary: Consider the below heatmaps of Uber pickup frequency in 2014 (left) and heatmap of median age in the Manhattan area (right).

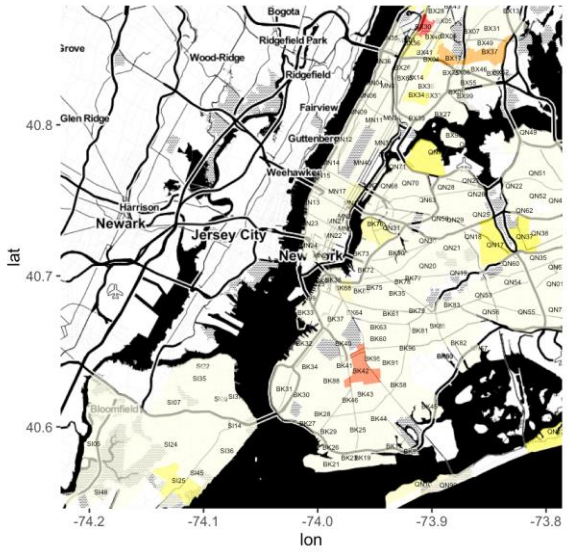


Median Age Heatmap



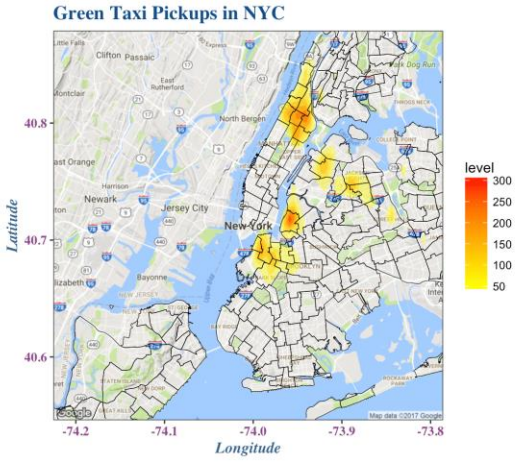
By inspection, we see that areas of Manhattan with higher median age (New York Stock Exchange, Whitney Museum of American Art, upper-right Manhattan) are comparatively less popular pickup spots. This indicates that in 2014, areas of Manhattan with lower median age were the most frequent users of Uber. Furthermore, we note that Uber usage is largely limited to Manhattan.

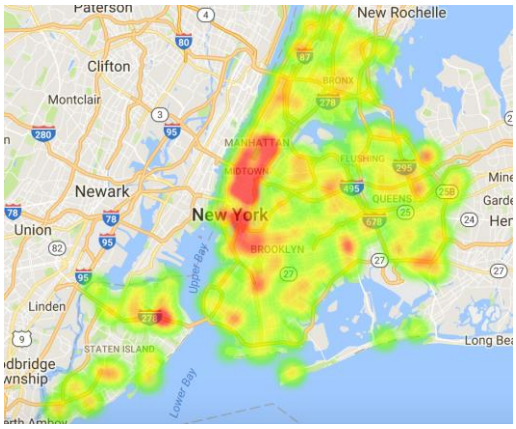
Now consider the heatmap of Uber pickup frequency in 2015:



Immediately we note that Uber usage has expanded outside of Manhattan and into Brooklyn/Queens/Staten Island. These areas are also the main locations of green taxi pickups. (see below left). Furthermore, these locations have a comparatively lower median income (see below right)

Median Income Heatmap





3. Technical Executive Summary – The core approach to answering our question was the use of geographic heat maps. For the Uber 2014 dataset, we mapped the latitude and longitude of pickups onto a map of New York to generate the geographic heatmapping. We also wanted to

capture the bordering regions of the neighborhood areas, so we used the geometric points in the geographic dataset to draw the borders.
For generating the demographic heatmaps, we only had demographic information per NTA code, so we estimated the central latitude/longitude of each NTA code and used those coordinates to generate heatmaps.