* Geographic Data
  + In order to collect taxi data from the financial firms we are interested in, we need to identify the coordinates of the offices we are interested in. To do this, we went firm by firm using google maps, and located their building that houses their investment banking division. Then, we created a 4-cornered polygon around the building, isolating the area we would expect employees to get taxis from. We noted the latitudinal and longitudinal coordinates for each of these four corners (8 coordinates for each firm). Because we are working with an incredibly large data set, we knew that mathematically, determining if a point is in a square with sides oriented perpendicular to true north and south would be the least computationally intensive. As such, we also added the north-south oriented square coordinates for each firm’s geographical polygon. We did this, so we could first isolate data to the larger square, then take the taxi data from inside that square and isolate it to our determined polygon. In the end, we used our geographical coordinate data for each firm to isolate the taxi data from within those polygon coordinates.
* Financial Data
  + The next data we needed to acquire was financial data for each firm. We are specifically interest in revenues of these firms, particularly investment banking revenues. To acquire these figures for each quarter since 2009, we needed to use a proprietary data aggregator, Bloomberg. We went into Bloomberg for each firm and pulled their quarterly segment financial data since 2009. Since Bloomberg provides data in heavily formatted and less useable way, we needed to clean the data from .xls files to create a useable .csv file. As such, we manually reformatted all the financial data to fit into a csv file with 7 attributes: ticker, quarter, year, date, segment, item and amnt. Each row specifies the firm, the time of the financial metric, the segment of the company it pertains to, the financial metric and the exact amount. Using this data, we will be able to compare the a firm’s financial metrics with their taxi data over a given quarter.