The problems facing our nation and the world are large. Two big problems in today's society are pollution and obesity. While these two issues may seem unrelated, a large bike sharing program can have an impact on both of the issues in a large city. This is achieved by reducing the need for cars for travel while simultaneously providing a means for city dwellers to exercise. This makes bike sharing programs important if implemented effectively. I propose to use the NYC Citi-Bank bike sharing program data to study how the program is performing since it has started. Although data is fairly sparse in time, it is still important to have a look at the program performance even in its early stage.

I have downloaded all of the available monthly ride data from the link provided. Each file is a zip file containing a csv file of the ride data for each month starting on July 2013 and ending on February 2015. Data contained in the files includes the station id, the starting/ending station id and map coordinates, along with information about the users of the system, such as age and sex.

For an initial look into the data, I have produced two illuminating plots. The first plot simply looks at the number of rides taken in each month, shown separately for male, female and unknown gender (also corresponding to non-subscription users). The first thing one can notice about the plot is that there is a clear cyclical trend in bike usage as a function of weather. That is much fewer bike rides are taken during the cold winter months. Also there are generally more than twice the number of males as females utilizing the rides program. One can also notice that most of the riders are subscription riders. As far as the yearly trend is concerned, unfortunately the span of data in time is not large. But comparing the first peak in usage compared to the second peak in usage indicates a slight decrease in usage of the program since the start.

The second plot illustrates the most popular ride stations used in the city. Only round trip (i.e. trips where the start and end station is the same) are included in this analysis. The plot shows the latitude and longitude map coordinate of each station on the vertical and horizontal axes respectively. The gray scale indicates the number of rides from that station, with a darker gray illustrating less rides and a lighter gray indicating more rides. It is interesting to note that there are very few “hot spots” where many riders like to make round trips from. There really only seems to be one main station that supports this, with a handful of other stations with a more moderate usage. An interesting addition to the figure if time allowed would be to superimpose a map of the city along with the data points to really see physically where these specific stations are located. Another interesting extension would be to study the data of other cities such as San Francisco and Washington DC.

In closing, I propose to look into this data further to look for evidence of trends of usage with the goal of proposing modifications to the program to increase usage by New Yorkers.