Final Project Plan

Data set: US Airport Statistics ( <http://jse.amstat.org/datasets/airport.dat.txt> )

Introduction of the data set: <http://jse.amstat.org/datasets/airport.txt>

Group: Yan Gao (only myself)

My intention: This is a data set of 135 large and medium-sized air hubs in the United States. I want to find if there are any group patterns show up in the different cities, also to see the definition between large and medium size among these airports. There might be some cities have special phenomena or the large airport might suppress other ones in the same city.

Methods: To tell the different sizes among the airports, I’ll definitely use clustering analysis. Besides that, I might use ggqqplot to check the normalrality of the distribution of scheduled departures and try to calculate its variance to check the airports' capability distribution. And use a two-sample test to find whether there is any relation between the performed departures and enplaned revenue tons of freight.