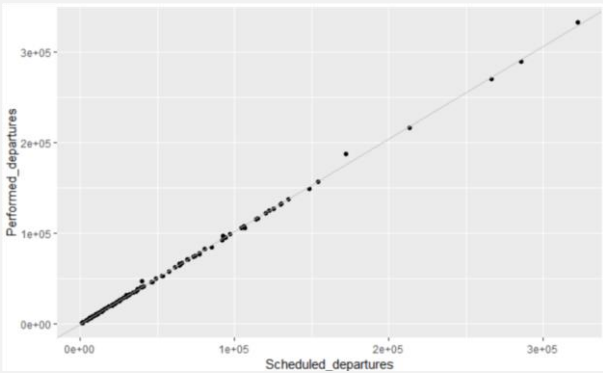


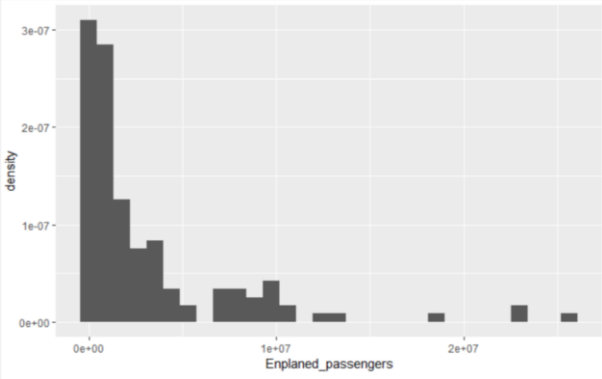
Linearity between scheduled and performed departures



The scheduled departure flights are almost symmetric with the performed ones, which could be considered as same distribution.

Density distribution of enplaned passengers

This is a vital variable to show the potential of clustering as it's the most important air traffic metric. The picture tells me maybe its better to using 3 clutters.

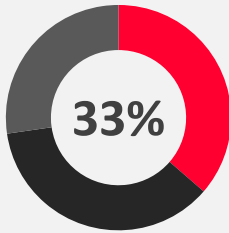


AIRPORT CLUSTERING

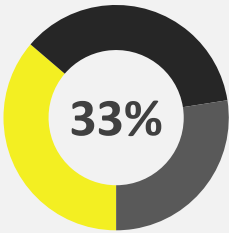
Cluster the airports of medium and big size according to the data\*

Main Features

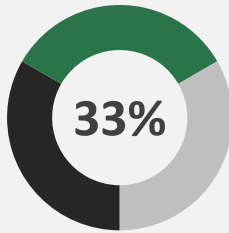
The dataset is 'Airport Activity Statistics' (1990). It is claimed that it consists of all 135 large and medium sized air hubs in the United States as defined by the Federal Aviation Administration. I am interested in the claim that how to clustering the air hubs into large and medium size.



Enplaned Passengers



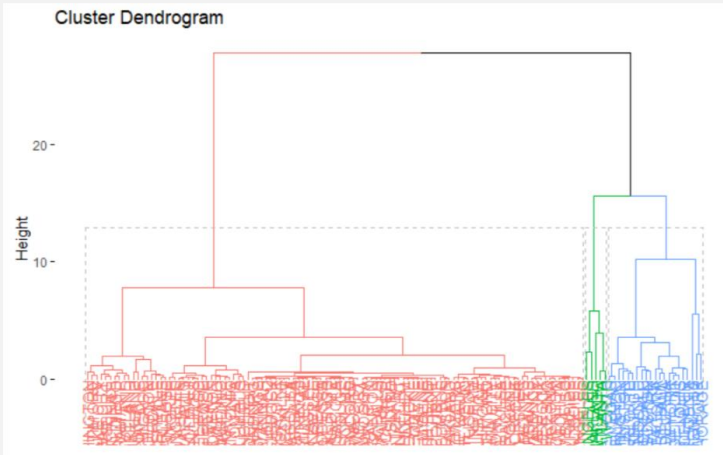
Performed Departures



Enplaned Revenue

Setting Weight\*\*

By using hierarchical clustering, the distance of the clusters is calculated by Euclidean distance, which compute the sum of square of each parameter. In this dataset, the scale of each parameter is quite different. So I have to merge the same meaning columns and define the weight.



The dendrogram is quite obvious to cluster the air hubs. The result shows that the size of this dataset of the air hubs show be devided into 3 part as more precisely distinguishment, not just large and medium.

33%: the Enplaned Passengers is the most important air traffic metric because the majority of airport revenues are generated directly or indirectly from enplaned passengers.

33%: The Performed Departures has the same distribution as the Scheduled Departure. So we only count one column.

33%: The Enplaned Revenue Tons is the sum of all the revenue tons from both the freight and mail.

Clustering Result

The top cluster air hubs is listed as follows:

- HARTSFIELD INTL ATLANTA
- O'HARE INTL CHICAGO
- DALLAS/FT WORTH INTL DALLAS/FT WORTH
- LOS ANGELES INTL LOS ANGELES
- SAN FRANCISCO INTL SAN FRANCISCO/OAKLAND

These results do match with the top largest air hubs showed on the Google

Sources of Information  
\* The dataset is collected from:  
<http://jse.amstat.org/datasets/airport.dat.txt>  
\*\* The parameter of each air hub:

Airport(name), City(location),  
Scheduled\_departures and  
Performed\_departures(number of flights),  
Enplaned\_passengers(number of total  
passengers), Enplaned\_revenue\_tons\_of\_freight

and Enplaned\_revenue\_tons\_of\_mail(total  
revenue been charged).

<http://jse.amstat.org/datasets/airport.txt>  
Yan.Gao@du.edu