

# Data Acquisition

## Data Sources:

- On-Time Performances: from [United States Department of Transportation](#)
- Aircraft Data: from [Statistical Computing Organization](#)
- Airport Data: from [Statistical Computing Organization](#)
- Weather Data: from [www.wunderground.com](http://www.wunderground.com)

## Data Acquisition Process:

### *On-Time Performances*

We manually downloaded the files of 2014 and 2015 on-time performance data from [http://www.transtats.bts.gov/Fields.asp?Table\\_ID=236](http://www.transtats.bts.gov/Fields.asp?Table_ID=236), and transferred them via filezilla (<https://filezilla-project.org/>) into the Amazon EC2 Instance.

### *Aircraft Data & Airport Data*

We manually downloaded “airport.csv” and “plane-data.csv” from <http://stat-computing.org/dataexpo/2009/supplemental-data.html>, and transferred them via filezilla (<https://filezilla-project.org/>) into the Amazon EC2 Instance.

### *Weather Data*

Please use the python script Wunder\_Weather\_Data\_Scrape.py in the folder to crawl the weather data for major US airports.