

### Homework 3

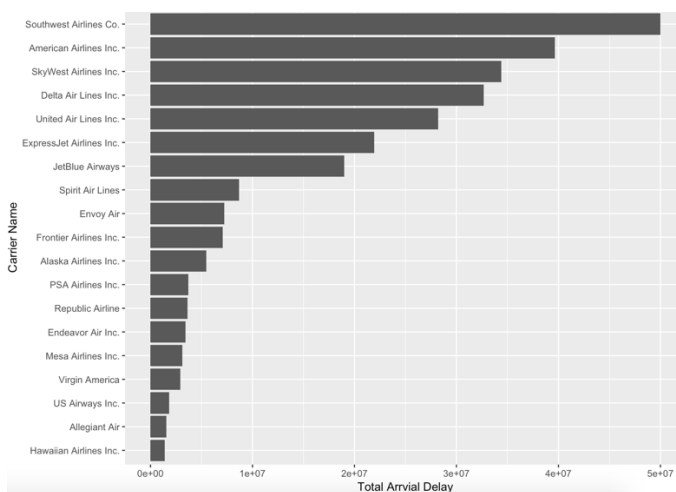
The attached dataset contains data related to aircraft arrival delays (in mins) from 2015 to 2018. It contains details regarding airline services, airports, and reasons for the delays. The column `arr_delay` is the total delays in mins. A delay event is calculated only when an aircraft arrival is delayed by more than 15 mins. For example, an aircraft arrival delayed by 10 mins is not recorded but an aircraft delayed by 18 mins is registered as 18 mins. So, the column `arr_delay` is the total delay in mins for a particular airline at a particular airport. Note that  $arr\_delay = carrier\_delay + weather\_delay + nas\_delay + security\_delay + late\_aircraft\_delay$ .

The detailed explanation of the variables are given as follows

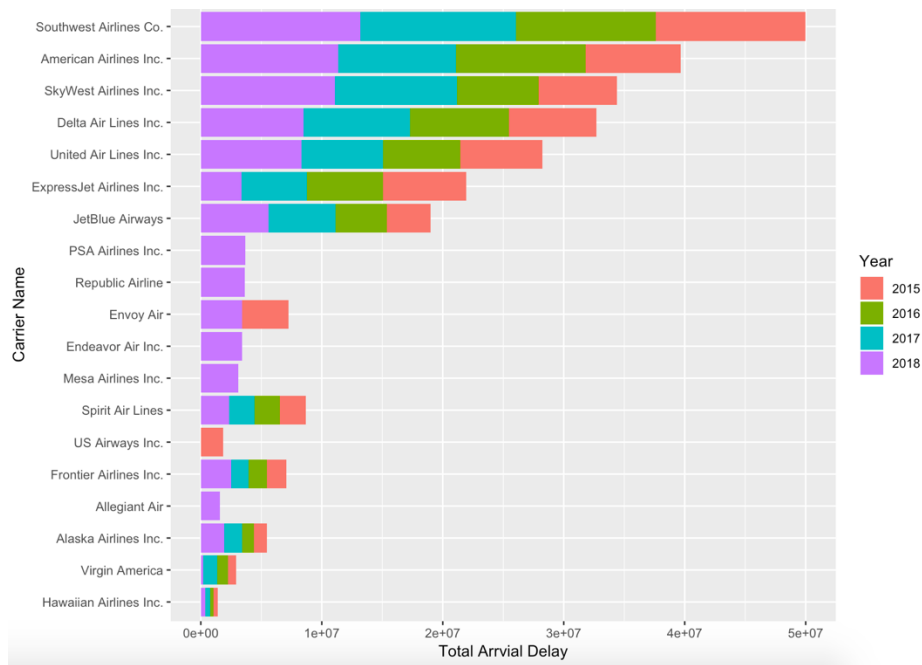
- **Air Carrier:** The cause of the cancellation or delay was due to circumstances within the airline's control (e.g. maintenance or crew problems, aircraft cleaning, baggage loading, fueling, etc.).
- **Extreme Weather:** Significant meteorological conditions (actual or forecasted) that, in the judgment of the carrier, delays or prevents the operation of a flight such as tornado, blizzard or hurricane.
- **National Aviation System (NAS):** Delays and cancellations attributable to the national aviation system that refer to a broad set of conditions, such as non-extreme weather conditions, airport operations, heavy traffic volume, and air traffic control.
- **Late-arriving aircraft:** A previous flight with same aircraft arrived late, causing the present flight to depart late.
- **Security:** Delays or cancellations caused by evacuation of a terminal or concourse, re-boarding of aircraft because of security breach, inoperative screening equipment and/or long lines in excess of 29 minutes at screening areas.

Answer the following questions for the aforementioned dataset

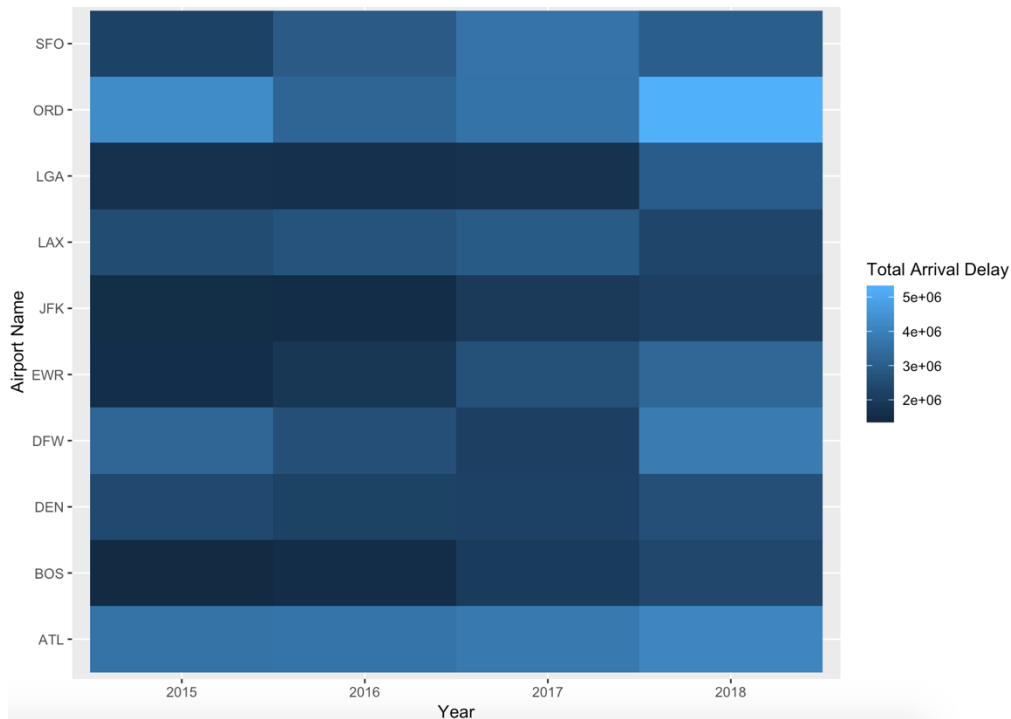
1. Plot the bar plot as shown below. The y axis is **carrier name**, the x axis is the **total arrival delay** (sum of `arr_delay` column) for the entire data set. (10 points)



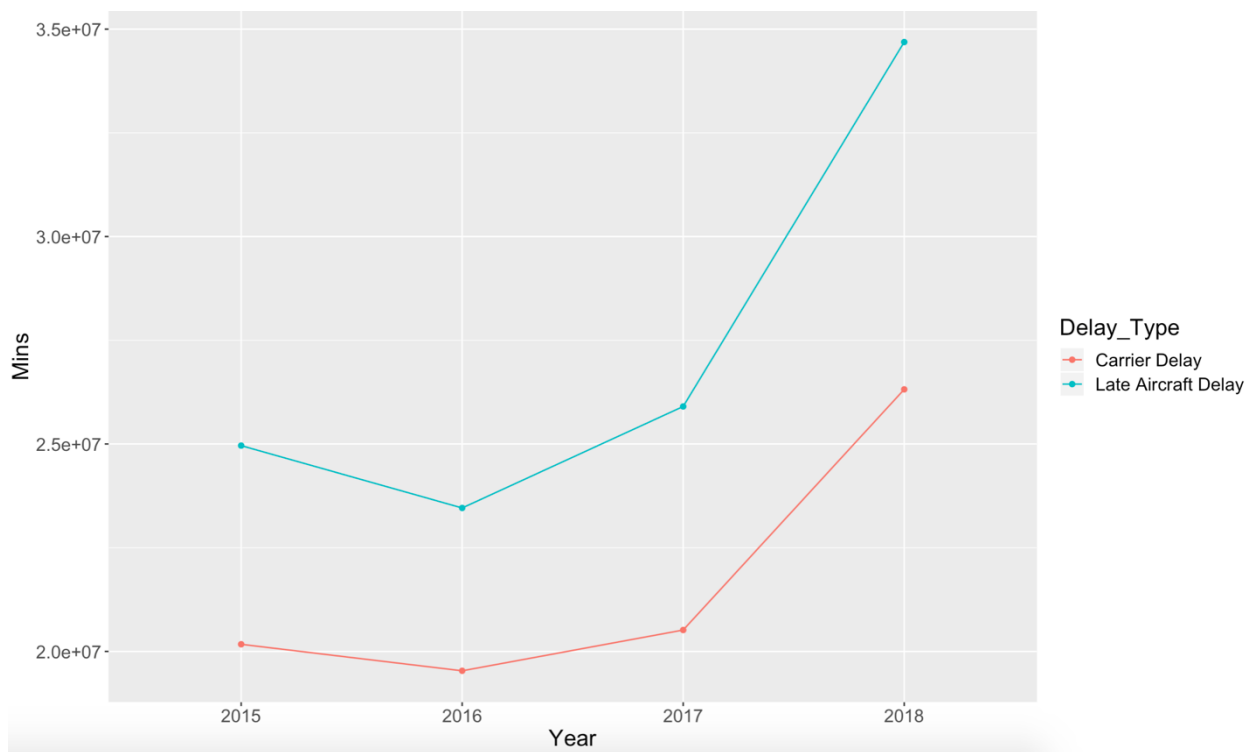
- Plot the stack bar plot as shown below. The y axis is **carrier name**, the x axis is the **total arrival delay** (sum of arr\_delay column) for the entire data set. The stacks represent the variable **year**. (10 points)



- Plot the following heat map. The y axis should contain only the **airports** shown in the figure below. The x axis represents the **year**. The heat map color should represent **total arrival delay** (sum of arr\_delay column). (10 points)

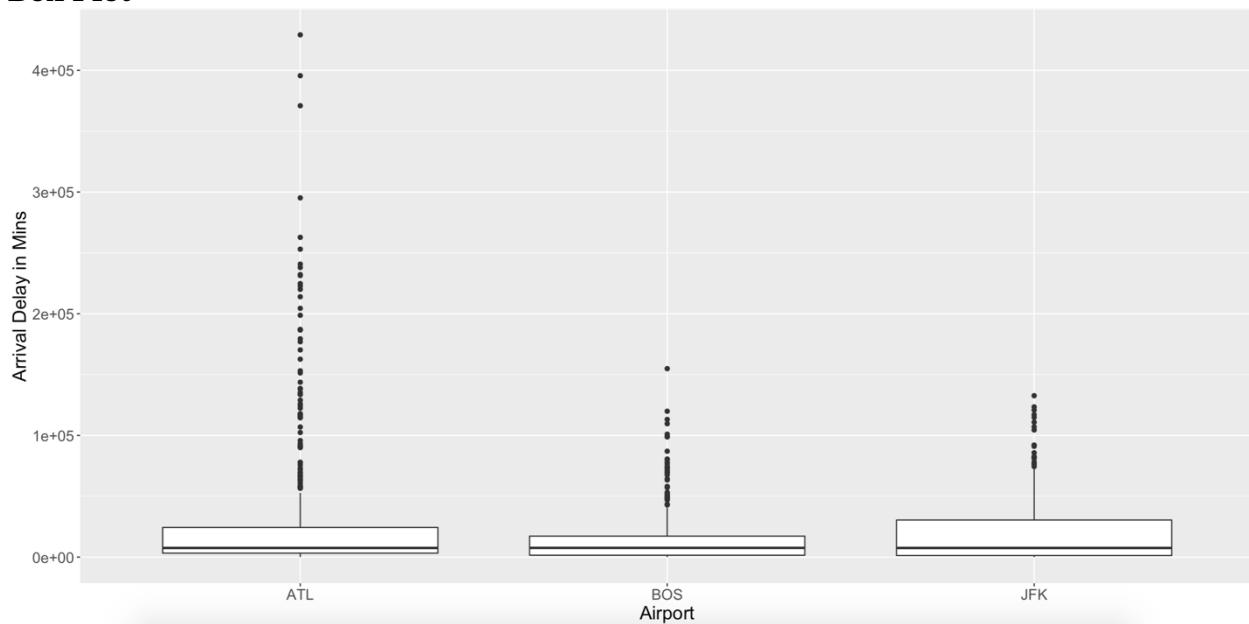


4. Plot the following line plot (10 points)

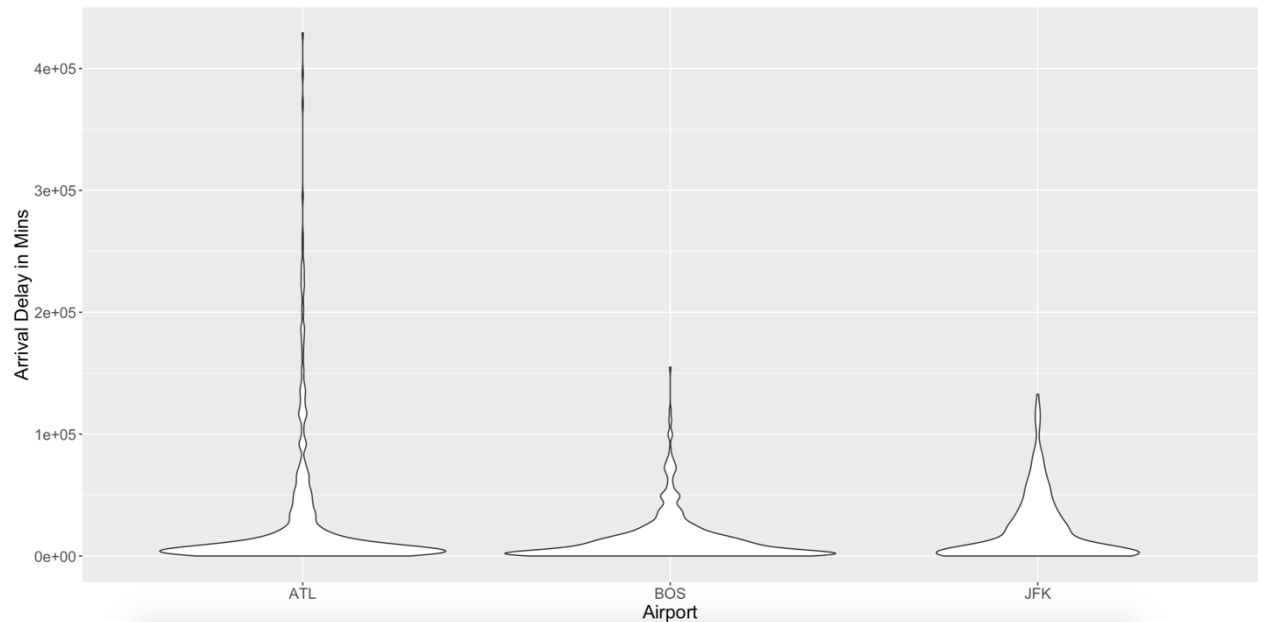


5. Plot the following distributions (30 points)

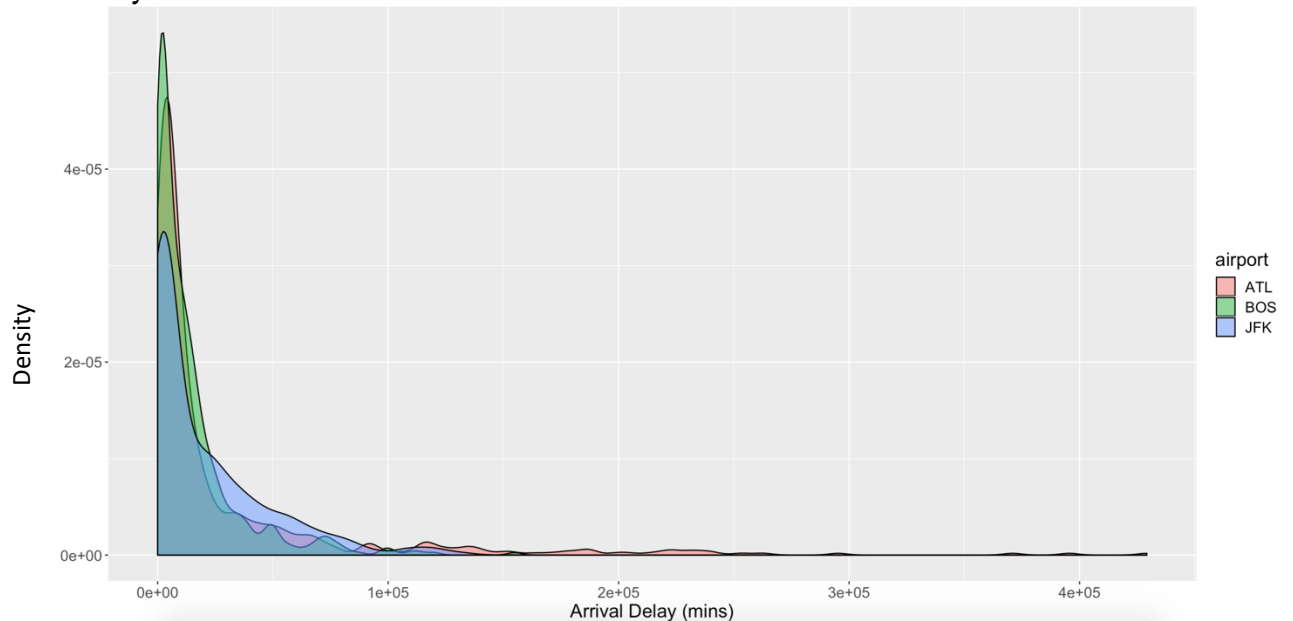
### Box Plot



### Violin Plot



### Density Plot



6. Answer the following (30 points)

The link below gives the data for passengers enplanement

([https://www.faa.gov/airports/planning\\_capacity/passenger\\_allcargo\\_stats/passenger/previus\\_years/](https://www.faa.gov/airports/planning_capacity/passenger_allcargo_stats/passenger/previus_years/)). Enplanement is the term used for passenger boarding.

- For the year 2017 download the data titled “Enplanements at All Commercial Service Airports (by Rank)”

- Create a new table as follows

Airport	Enplanement	Arrival Delay (mins)
JFK	5000000	30004
..	..	..

- The enplanement data is from the downloaded data while the arrival delay data can be obtained using the data attached with the homework
- Plot a scatter plot with enplanement on x axis and arrival delay in y axis.
- Plot the same figure above with x and y axis in log scale.