Data Used For the project:

The below link contains all the files that were used for the project.

The source for the data was transtats.bts.gov which contains datasets about the transportations in USA.

**Overview of the datasets**

*All the month files have the following data structure:*

DAY\_OF\_WEEK : contains 1 to 7 values representaing Mon to SUN

FL\_DATE: Date of journey

CARRIER: Carrier Name

FL\_NUM: Flight number

ORIGIN\_AIRPORT\_ID: Origin airport id

ORIGIN\_CITY\_NAME: Origin City Name

DEST\_AIRPORT\_ID: Destination Airport ID

DEST\_CITY\_NAME: Destination City Name

DEP\_TIME: Departure Time

ARR\_TIME: Arrival Time

CANCELLED: 0 or 1 with 1 when flight was cancelled

CANCELLATION\_CODE: code of cancellation

AIR\_TIME: Air Time

DISTANCE: Distance travelled

CARRIER\_DELAY: Delay in minutes by Carrier

WEATHER\_DELAY: Delay in minutes due to Weather

NAS\_DELAY: Delay in minutes due to NAS

SECURITY\_DELAY: Delay in Security due to Security

LATE\_AIRCRAFT\_DELAY: Delay due to Late Aircraft

The 2017-ohare-rain.csv contain the following structure:

Month- Month from 1 to 12

Day-Day of the Month

precipitation(in)- Precipitation details in inches

The holidays.csv contains all the dates of the holidays in the USA which gives a quick view to the major dates that can be looked for while seeing the visualization.

**Downloading of the datasets:**

The datasets were downloaded from the source in a way that only the necessary columns were present in the dataset for faster processing of the application.

**Preprocessing and manipulation:**

The datasets were preprocessed to bring in a proper format to be displayed.

Library such as lubridate used used for dates manipulation and reshape2 was used for reshaping the data for easy manipulation. RbindList() function was used in order to merge all the month datasets in one dataset. Many other function were used to bring the data in the format that the visualizing the data using ggplot became easier.