DATA ANALYSIS

We will be calculating the Pearson correlation to see if there's a correlation between the total number of passenger arrivals in LGA/JFK with the other variables such as Events, Transportation, Weather, and Arrests in NYC.

```
combined <- read.csv("/cloud/project/combined.csv")
head(combined)
```

```
##      Month.Year Number.of.Tourists..jfk. Number.of.Tourists..lga.
## 1  January-2015                12594000                 12531000
## 2 Feburary-2015                12684000                 12608000
## 3    March-2015                12728000                 12640000
## 4    April-2015                12806000                 12693000
## 5     May -2015                12869000                 12738000
## 6     June-2015                12924000                 12802000
##   Total_Tourists Event.ID Transportation Precipitation Number.of.Arrests
## 1       25125000      720      320336485          29.9             25151
## 2       25292000      784      307108527          23.9             26891
## 3       25368000     1031      371333237          38.1             29807
## 4       25499000     2053      365871941          54.3             29769
## 5       25607000     4327      361018084          68.5             28489
## 6       25726000     6627      369622990          71.2             28501
```

```
selected_data <- combined[c("Total_Tourists", "Event.ID", "Transportation", "Precipitation", "Number.of

# Compute the correlation matrix
correlation_matrix <- cor(selected_data, use = "complete.obs")

# Show correlation of total number of passenger arrivals with the other variables
correlation_matrix["Total_Tourists", ]
```

```
##    Total_Tourists          Event.ID    Transportation     Precipitation
##        1.00000000       -0.26790795        0.73181716        0.01153907
## Number.of.Arrests
##        0.55661029
```

```
# Scatterplot matrix
pairs(selected_data,
      main = "Scatterplot Matrix of Tourists and Other Variables",
      pch = 19,      # solid circle
      col = "hotpink")
```

Scatterplot Matrix of Tourists and Other Variables