

Vuelos NYC

Hernández Martínez Oscar Gerardo

2/7/2020

```
import os
os.environ['QT_QPA_PLATFORM_PLUGIN_PATH'] = 'C:/Users/jxsje/anaconda3/Library/plugins/platforms'
```

Vuelos de NYC

```
data <- nycflights13::flights
head(data)
```

```
## # A tibble: 6 x 19
##   year month   day dep_time sched_dep_time dep_delay arr_time sched_arr_time
##   <int> <int> <int>   <int>         <int>       <dbl>   <int>         <int>
## 1  2013     1     1     517           515         2     830           819
## 2  2013     1     1     533           529         4     850           830
## 3  2013     1     1     542           540         2     923           850
## 4  2013     1     1     544           545        -1    1004          1022
## 5  2013     1     1     554           600        -6     812           837
## 6  2013     1     1     554           558        -4     740           728
## # ... with 11 more variables: arr_delay <dbl>, carrier <chr>, flight <int>,
## #   tailnum <chr>, origin <chr>, dest <chr>, air_time <dbl>, distance <dbl>,
## #   hour <dbl>, minute <dbl>, time_hour <dtm>
```

```
summary(data)
```

```
##      year      month      day      dep_time      sched_dep_time
## Min.   :2013   Min.   : 1.000   Min.   : 1.00   Min.   : 1      Min.   : 106
## 1st Qu.:2013   1st Qu.: 4.000   1st Qu.: 8.00   1st Qu.: 907    1st Qu.: 906
## Median :2013   Median : 7.000   Median :16.00   Median :1401    Median :1359
## Mean   :2013   Mean   : 6.549   Mean   :15.71   Mean   :1349    Mean   :1344
## 3rd Qu.:2013   3rd Qu.:10.000   3rd Qu.:23.00   3rd Qu.:1744    3rd Qu.:1729
## Max.   :2013   Max.   :12.000   Max.   :31.00   Max.   :2400    Max.   :2359
##
##                                     NA's   :8255
##      dep_delay      arr_time      sched_arr_time      arr_delay
## Min.   : -43.00   Min.   : 1      Min.   : 1      Min.   : -86.000
## 1st Qu.: -5.00    1st Qu.:1104    1st Qu.:1124    1st Qu.: -17.000
## Median : -2.00    Median :1535    Median :1556    Median : -5.000
## Mean   : 12.64    Mean   :1502    Mean   :1536    Mean   : 6.895
## 3rd Qu.: 11.00    3rd Qu.:1940    3rd Qu.:1945    3rd Qu.: 14.000
## Max.   :1301.00   Max.   :2400    Max.   :2359    Max.   :1272.000
## NA's   :8255     NA's   :8713    NA's   :9430
##      carrier      flight      tailnum      origin
## Length:336776   Min.   : 1      Length:336776   Length:336776
## Class :character 1st Qu.: 553   Class :character Class :character
## Mode :character  Median :1496   Mode :character Mode :character
```

```
##           Mean    :1972
##           3rd Qu.:3465
##           Max.    :8500
##
##      dest      air_time      distance      hour
## Length:336776   Min.    : 20.0   Min.    : 17   Min.    : 1.00
## Class :character 1st Qu.: 82.0   1st Qu.: 502   1st Qu.: 9.00
## Mode  :character Median :129.0   Median : 872   Median :13.00
##           Mean    :150.7   Mean    :1040   Mean    :13.18
##           3rd Qu.:192.0   3rd Qu.:1389   3rd Qu.:17.00
##           Max.    :695.0   Max.    :4983   Max.    :23.00
##           NA's    :9430
##      minute      time_hour
## Min.    : 0.00   Min.    :2013-01-01 05:00:00
## 1st Qu.: 8.00   1st Qu.:2013-04-04 13:00:00
## Median :29.00   Median :2013-07-03 10:00:00
## Mean    :26.23   Mean    :2013-07-03 05:22:54
## 3rd Qu.:44.00   3rd Qu.:2013-10-01 07:00:00
## Max.    :59.00   Max.    :2013-12-31 23:00:00
##
```

```
nrow(data)
```

```
## [1] 336776
```

```
pydata = r.data
pydata = pydata[pydata["dest"]=="ORD"]
pydata = pydata[['carrier', 'dep_delay', 'arr_delay', 'origin']]
pydata = pydata[pydata['arr_delay']<6*60]
pydata = pydata.dropna()
print(pydata.head())
```

```
##      carrier dep_delay arr_delay origin
## 5         UA      -4.0      12.0    EWR
## 9         AA      -2.0       8.0    LGA
## 25        MQ       8.0      32.0    EWR
## 38        AA      -1.0      14.0    LGA
## 57        AA      -4.0       4.0    LGA
```

```
print(pydata.shape)
```

```
## (16552, 4)
```

```
summary(pydata)
```

```
##      carrier      dep_delay      arr_delay      origin
## Length:16552   Min.    : -20.00   Min.    : -62.000   Length:16552
## Class :character 1st Qu.:  -5.00   1st Qu.: -20.000   Class :character
## Mode  :character Median :  -2.00   Median :  -8.000   Mode  :character
##           Mean    : 13.04   Mean    :   5.477
##           3rd Qu.: 11.00   3rd Qu.: 13.000
##           Max.    :389.00   Max.    :348.000
```

```
boxplot(arr_delay~origin, data = py$pydata, main = "Retraso de los vuelos hacia Orlando desde NYC")
```

