# 03-Analisis de vuelos con Python

## Vuelos en 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 <dttm>
```

### Filtro de datos con tidyverse

```
pydata = r.data
pydata = pydata[pydata["dest"]=="ORD"]
pydata = pydata[["carrier", "dep_delay", "arr_delay", "origin"]]
# Eliminar outliers
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
```

### Obtener los retrasos de los vuelos hacia Orlanda desde NYC

```
summary(py$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 = "Retrasos de los vuelos por origen")
```