

Analisis y Curacion clase 3: Práctico 1

30 de Mayo, 2019, Cohorte Alpha

Análisis exploratorio y curación de datos

- ▶ Gonzalez Nehuen
- ▶ Arja Adel
- ▶ Madoery Pablo

Practico 1: Entregar un Rmd donde se encuentren todos los vuelos que:

- ▶ Que arribaron con un retraso de mas de dos horas.
- ▶ Volaron hacia Houston (IAH o HOU)
- ▶ Fueron operados por United, American o Delta.
- ▶ Salieron en Verano (Julio, Agosto y Septiembre)
- ▶ Arrivaron mas de dos horas tarde, pero salieron bien.
- ▶ Salieron entre medianoche y las 6 am.

Solución

```
library(nycflights13)  
flights<-nycflights13::flights
```

- Que arribaron con un retraso de mas de dos horas.

```
condition <-(flights[, "arr_delay"] > 2*60)
flights[condition, "flight"]
```

```
## # A tibble: 19,464 x 1
##   flight
##   <int>
## 1    4576
## 2    3944
## 3     856
## 4    1086
## 5    4497
## 6     525
## 7      NA
## 8      NA
## 9    4181
## 10   5712
## # ... with 19,454 more rows
```

- Volaron hacia Houston (IAH o HOU)

```
condition <-  
  (flights[, "dest"] == "IAH") |  
  (flights[, "dest"] == "HOU")  
flights[condition, "flight"]
```

```
## # A tibble: 9,313 x 1
```

```
##   flight
```

```
##   <int>
```

```
## 1    1545
```

```
## 2    1714
```

```
## 3     496
```

```
## 4     473
```

```
## 5    1479
```

```
## 6    1220
```

```
## 7    1004
```

```
## 8     455
```

```
## 9    1086
```

```
## 10   1461
```

```
## # with 9,303 more rows
```

- Fueron operados por United, American o Delta.

```
condition <-  
  (flights[, "carrier"] == "UA") |  
  (flights[, "carrier"] == "AA") |  
  (flights[, "carrier"] == "DL")  
flights[condition, "flight"]
```

```
## # A tibble: 139,504 x 1
```

```
##   flight
```

```
##   <int>
```

```
## 1    1545
```

```
## 2    1714
```

```
## 3    1141
```

```
## 4     461
```

```
## 5    1696
```

```
## 6     301
```

```
## 7     194
```

```
## 8    1124
```

```
## 9     707
```

```
## 10    1187
```

- Salieron en Verano (Julio, Agosto y Septiembre)

```
condition <-  
  (flights[, "month"] >= 7) &  
  (flights[, "month"] <= 9)  
flights[condition, "flight"]
```

```
## # A tibble: 86,326 x 1
```

```
##   flight
```

```
##   <int>
```

```
## 1     915
```

```
## 2    1503
```

```
## 3     234
```

```
## 4    1371
```

```
## 5     185
```

```
## 6     165
```

```
## 7     415
```

```
## 8     425
```

```
## 9    1183
```

```
## 10    623
```

```
## # with 86,316 more rows
```


- Arrivaron mas de dos horas tarde, pero salieron bien.

```
condition <-  
  (flights[, "arr_delay"] > 2*60) &  
  (flights[, "dep_delay"] <= 0)  
flights[condition, "flight"]
```

```
## # A tibble: 8,772 x 1  
##   flight  
##   <int>  
## 1     NA  
## 2     NA  
## 3     NA  
## 4     NA  
## 5     NA  
## 6     NA  
## 7     NA  
## 8     NA  
## 9     NA  
## 10    NA  
## # with 8,762 more rows
```

- Salieron entre medianoche y las 6 am.

```
condition <-  
  ((flights[, "hour"] >= 0) & (flights[, "hour"] <= 5)) |  
  ((flights[, "hour"] == 6) & (flights[, "minute"] == 0))  
flights[condition, "flight"]
```

```
## # A tibble: 8,970 x 1  
##   flight  
##   <int>  
## 1    1545  
## 2    1714  
## 3    1141  
## 4     725  
## 5     461  
## 6    1696  
## 7     507  
## 8    5708  
## 9       79  
## 10    301  
## # with 8,960 more rows
```