

Practico1

Daniel Robins - Fernando Palacios

9 de mayo de 2018

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

Including Plots

You can also embed plots, for example:

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

Visualizacion es una herramienta muy importante para la generacion de intuicion, pero raramente uno tiene los datos en la forma necesaria. Frecuentemente se necesitara crear nuevas variables o simplemente reordenarlas.

Exploraremos ahora la manipulacion basica utilizando un conjunto de datos sobre los vuelos en Nueva York en 2013.

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

```
## # A tibble: 336,776 x 19
##   year month   day dep_time sched_dep_time dep_delay arr_time
##   <int> <int> <int>   <int>         <int>         <dbl>   <int>
## 1  2013     1     1     517             515           2     830
## 2  2013     1     1     533             529           4     850
## 3  2013     1     1     542             540           2     923
## 4  2013     1     1     544             545          -1    1004
## 5  2013     1     1     554             600          -6     812
## 6  2013     1     1     554             558          -4     740
## 7  2013     1     1     555             600          -5     913
## 8  2013     1     1     557             600          -3     709
## 9  2013     1     1     557             600          -3     838
## 10 2013     1     1     558             600          -2     753
## # ... with 336,766 more rows, and 12 more variables: sched_arr_time <int>,
## #   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>
```

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.

```
str(flights)
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame':  336776 obs. of  19 variables:
## $ year      : int  2013 2013 2013 2013 2013 2013 2013 2013 2013 2013 ...
## $ month     : int  1 1 1 1 1 1 1 1 1 1 ...
## $ day       : int  1 1 1 1 1 1 1 1 1 1 ...
## $ dep_time  : int  517 533 542 544 554 554 555 557 557 558 ...
## $ sched_dep_time: int  515 529 540 545 600 558 600 600 600 600 ...
## $ dep_delay : num  2 4 2 -1 -6 -4 -5 -3 -3 -2 ...
## $ arr_time  : int  830 850 923 1004 812 740 913 709 838 753 ...
## $ sched_arr_time: int  819 830 850 1022 837 728 854 723 846 745 ...
## $ arr_delay : num  11 20 33 -18 -25 12 19 -14 -8 8 ...
## $ carrier   : chr  "UA" "UA" "AA" "B6" ...
## $ flight    : int  1545 1714 1141 725 461 1696 507 5708 79 301 ...
## $ tailnum   : chr  "N14228" "N24211" "N619AA" "N804JB" ...
## $ origin    : chr  "EWR" "LGA" "JFK" "JFK" ...
## $ dest      : chr  "IAH" "IAH" "MIA" "BQN" ...
## $ air_time  : num  227 227 160 183 116 150 158 53 140 138 ...
## $ distance  : num  1400 1416 1089 1576 762 ...
## $ hour      : num  5 5 5 5 6 5 6 6 6 6 ...
## $ minute    : num  15 29 40 45 0 58 0 0 0 0 ...
## $ time_hour : POSIXct, format: "2013-01-01 05:00:00" "2013-01-01 05:00:00" ...
```

Que arribaron con un retraso de mas de dos horas.

```
subset(flights, flights["arr_delay"]>120)
```

```
## # A tibble: 10,034 x 19
##   year month   day dep_time sched_dep_time dep_delay arr_time
##   <int> <int> <int>   <int>         <int>      <dbl>   <int>
## 1  2013     1     1     811             630        101    1047
## 2  2013     1     1     848            1835        853    1001
## 3  2013     1     1     957             733        144    1056
## 4  2013     1     1    1114             900        134    1447
## 5  2013     1     1    1505            1310        115    1638
## 6  2013     1     1    1525            1340        105    1831
## 7  2013     1     1    1549            1445         64    1912
## 8  2013     1     1    1558            1359        119    1718
## 9  2013     1     1    1732            1630         62    2028
## 10 2013     1     1    1803            1620        103    2008
## # ... with 10,024 more rows, and 12 more variables: sched_arr_time <int>,
## #   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>
```

Volaron hacia Houston (IAH o HOU)

```
subset(flights, flights["dest"]=="IAH" | flights["dest"]=="HOU" )

## # A tibble: 9,313 x 19
##   year month   day dep_time sched_dep_time dep_delay arr_time
##   <int> <int> <int>   <int>         <int>         <dbl>   <int>
## 1  2013     1     1     517           515           2     830
## 2  2013     1     1     533           529           4     850
## 3  2013     1     1     623           627          -4     933
## 4  2013     1     1     728           732          -4    1041
## 5  2013     1     1     739           739           0    1104
## 6  2013     1     1     908           908           0    1228
## 7  2013     1     1    1028          1026           2    1350
## 8  2013     1     1    1044          1045           -1    1352
## 9  2013     1     1    1114           900        134    1447
## 10 2013     1     1    1205          1200           5    1503
## # ... with 9,303 more rows, and 12 more variables: sched_arr_time <int>,
## #   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>
```

Fueron operados por United, American o Delta.

```
subset(flights, carrier=="DL" | carrier=="AA" | carrier=="UA")

## # A tibble: 139,504 x 19
##   year month   day dep_time sched_dep_time dep_delay arr_time
##   <int> <int> <int>   <int>         <int>         <dbl>   <int>
## 1  2013     1     1     517           515           2     830
## 2  2013     1     1     533           529           4     850
## 3  2013     1     1     542           540           2     923
## 4  2013     1     1     554           600          -6     812
## 5  2013     1     1     554           558          -4     740
## 6  2013     1     1     558           600          -2     753
## 7  2013     1     1     558           600          -2     924
## 8  2013     1     1     558           600          -2     923
## 9  2013     1     1     559           600          -1     941
## 10 2013     1     1     559           600          -1     854
## # ... with 139,494 more rows, and 12 more variables: sched_arr_time <int>,
## #   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>

# Airlines DL, AA, UA
```

Salieron en Verano (Julio, Agosto y Septiembre)

```
subset(flights, month >= 7 & month <= 9, select=c(month))

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

```
##      <int>
## 1      7
## 2      7
## 3      7
## 4      7
## 5      7
## 6      7
## 7      7
## 8      7
## 9      7
## 10     7
## # ... with 86,316 more rows
```

Salieron entre medianoche y las 6 am.

```
subset(flights, dep_time<=600 )
```

```
## # A tibble: 9,344 x 19
##   year month   day dep_time sched_dep_time dep_delay arr_time
##   <int> <int> <int>   <int>         <int>       <dbl>   <int>
## 1  2013     1     1     517             515         2     830
## 2  2013     1     1     533             529         4     850
## 3  2013     1     1     542             540         2     923
## 4  2013     1     1     544             545        -1    1004
## 5  2013     1     1     554             600        -6     812
## 6  2013     1     1     554             558        -4     740
## 7  2013     1     1     555             600        -5     913
## 8  2013     1     1     557             600        -3     709
## 9  2013     1     1     557             600        -3     838
## 10 2013     1     1     558             600        -2     753
## # ... with 9,334 more rows, and 12 more variables: sched_arr_time <int>,
## #   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>
```