Aula 1

Beatriz Milz

10/03/2020

Exercícios da aula do dia 06/03/2020 - [Link](https://jonnyphillips.github.io/Ciencia_de_Dados/introducao.html).

# Instalar o pacote nycflights13  
  
# install.packages("nycflights13")  
  
# Carregar o pacote instalado  
library("nycflights13")

## Warning: package 'nycflights13' was built under R version 3.6.3

# Carregar o pacote tidyverse  
  
library("tidyverse")

## -- Attaching packages ----------------------------- tidyverse 1.3.0 --

## v ggplot2 3.2.1 v purrr 0.3.3  
## v tibble 2.1.3 v dplyr 0.8.4  
## v tidyr 1.0.2 v stringr 1.4.0  
## v readr 1.3.1 v forcats 0.4.0

## -- Conflicts -------------------------------- tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

# Abrir o dataframe flights  
  
flights <- flights  
head(flights)

## # 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>

**Quantas observações existem no banco de dados?** 336776 observações.

**Quantas variáveis existem no banco de dados?** 19 variáveis.

**O que representa cada observação no banco de dados?** Cada observação representa um vôo realizado.

**Quais variáveis existem nos dados?**

colnames(flights)

## [1] "year" "month" "day" "dep\_time"   
## [5] "sched\_dep\_time" "dep\_delay" "arr\_time" "sched\_arr\_time"  
## [9] "arr\_delay" "carrier" "flight" "tailnum"   
## [13] "origin" "dest" "air\_time" "distance"   
## [17] "hour" "minute" "time\_hour"

**Qual ‘tipo’ de dado contém cada variável?**

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" ...

siglas\_aeroportos <- c("NWR", "JFK", "LGA")  
nomes\_aeroportos <- c("Newark", "John F Kennedy", "Laguardia")  
  
  
tabela\_aeroportos <- tibble(sigla = siglas\_aeroportos,  
 nome = nomes\_aeroportos)

tabela\_aeroportos

## # A tibble: 3 x 2  
## sigla nome   
## <chr> <chr>   
## 1 NWR Newark   
## 2 JFK John F Kennedy  
## 3 LGA Laguardia

# exercícios no arquivo:  
rstudioapi::navigateToFile("1\_introducao\_exercicios.Rmd")