R\_Work

Alvaro Cristian Botelho

3 de setembro de 2018

# Trabalho R Excercios.

Neste trabalho foi utilizada o dataset ~countries of the world~ obtido no Kaggle.

<https://www.kaggle.com/fernandol/countries-of-the-world>

**Exercício 1 - Crie um dataframe com 10 colunas e 40 linhas e imprima na tela as primeiras 6 linhas.**

*Task 1 - Create a dataframe with 10 colluns and 40 lines and then print on the screen the 6 firtst lines.*

#Importing data Dataframe and creation of df\_world  
  
library(readxl)  
df\_world <- read\_excel(".\\countries of the world.xlsx")  
head(df\_world)

## # A tibble: 6 x 19  
## Country Region Population `Area (sq. mi.)` `Coastline (coa~  
## <chr> <chr> <dbl> <dbl> <dbl>  
## 1 Afghan~ ASIA ~ 31056997 647500 0   
## 2 Albania EASTE~ 3581655 28748 1.26  
## 3 Algeria NORTH~ 32930091 2381740 0.04  
## 4 Americ~ OCEAN~ 57794 199 58.3   
## 5 Andorra WESTE~ 71201 468 0   
## 6 Angola SUB-S~ 12127071 1246700 0.13  
## # ... with 14 more variables: `Net migration` <dbl>, `Infant mortality  
## # (per 1000 births)` <dbl>, `GDP ($ per capita)` <dbl>, `Literacy  
## # (%)` <dbl>, `Phones (per 1000)` <dbl>, `Arable (%)` <dbl>, `Crops  
## # (%)` <dbl>, `Other (%)` <dbl>, Climate <dbl>, Birthrate <dbl>,  
## # Deathrate <dbl>, Agriculture <dbl>, Industry <dbl>, Service <dbl>

**Exercício 2 - Mostre a que classe pertence cada uma das 10 colunas do seu dataframe.** \* Show all datatypes from all collumns.\*

sapply (df\_world,class)

## Country Region   
## "character" "character"   
## Population Area (sq. mi.)   
## "numeric" "numeric"   
## Coastline (coast/area ratio) Net migration   
## "numeric" "numeric"   
## Infant mortality (per 1000 births) GDP ($ per capita)   
## "numeric" "numeric"   
## Literacy (%) Phones (per 1000)   
## "numeric" "numeric"   
## Arable (%) Crops (%)   
## "numeric" "numeric"   
## Other (%) Climate   
## "numeric" "numeric"   
## Birthrate Deathrate   
## "numeric" "numeric"   
## Agriculture Industry   
## "numeric" "numeric"   
## Service   
## "numeric"

**Exercício 3 - Utilize um comando que faça uma avaliação exploratórias das variáveis do seu dataframe e imprima na tela.** *Use one command line witch makes an explanatory evaluation for all variables on your dataframe.*

summary(df\_world)

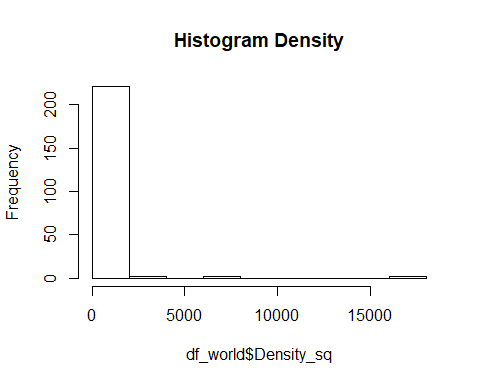
## Country Region Population   
## Length:227 Length:227 Min. :7.026e+03   
## Class :character Class :character 1st Qu.:4.376e+05   
## Mode :character Mode :character Median :4.787e+06   
## Mean :2.874e+07   
## 3rd Qu.:1.750e+07   
## Max. :1.314e+09   
##   
## Area (sq. mi.) Coastline (coast/area ratio) Net migration   
## Min. : 2 Min. : 0.00 Min. :-20.99000   
## 1st Qu.: 4648 1st Qu.: 0.10 1st Qu.: -0.92750   
## Median : 86600 Median : 0.73 Median : 0.00000   
## Mean : 598227 Mean : 21.17 Mean : 0.03812   
## 3rd Qu.: 441811 3rd Qu.: 10.35 3rd Qu.: 0.99750   
## Max. :17075200 Max. :870.66 Max. : 23.06000   
## NA's :3   
## Infant mortality (per 1000 births) GDP ($ per capita) Literacy (%)   
## Min. : 2.29 Min. : 500 Min. : 17.60   
## 1st Qu.: 8.15 1st Qu.: 1900 1st Qu.: 70.60   
## Median : 21.00 Median : 5550 Median : 92.50   
## Mean : 35.51 Mean : 9690 Mean : 82.84   
## 3rd Qu.: 55.70 3rd Qu.:15700 3rd Qu.: 98.00   
## Max. :191.19 Max. :55100 Max. :100.00   
## NA's :3 NA's :1 NA's :18   
## Phones (per 1000) Arable (%) Crops (%) Other (%)   
## Min. : 0.2 Min. : 0.00 Min. : 0.000 Min. : 33.33   
## 1st Qu.: 37.8 1st Qu.: 3.22 1st Qu.: 0.190 1st Qu.: 71.65   
## Median : 176.2 Median :10.42 Median : 1.030 Median : 85.70   
## Mean : 236.1 Mean :13.80 Mean : 4.564 Mean : 81.64   
## 3rd Qu.: 389.6 3rd Qu.:20.00 3rd Qu.: 4.440 3rd Qu.: 95.44   
## Max. :1035.6 Max. :62.11 Max. :50.680 Max. :100.00   
## NA's :4 NA's :2 NA's :2 NA's :2   
## Climate Birthrate Deathrate Agriculture   
## Min. :1.000 Min. : 7.29 Min. : 2.290 Min. :0.00000   
## 1st Qu.:2.000 1st Qu.:12.67 1st Qu.: 5.910 1st Qu.:0.03775   
## Median :2.000 Median :18.79 Median : 7.840 Median :0.09900   
## Mean :2.139 Mean :22.11 Mean : 9.241 Mean :0.15084   
## 3rd Qu.:3.000 3rd Qu.:29.82 3rd Qu.:10.605 3rd Qu.:0.22100   
## Max. :4.000 Max. :50.73 Max. :29.740 Max. :0.76900   
## NA's :22 NA's :3 NA's :4 NA's :15   
## Industry Service   
## Min. :0.0200 Min. :0.0620   
## 1st Qu.:0.1930 1st Qu.:0.4293   
## Median :0.2720 Median :0.5710   
## Mean :0.2827 Mean :0.5653   
## 3rd Qu.:0.3410 3rd Qu.:0.6785   
## Max. :0.9060 Max. :0.9540   
## NA's :16 NA's :15

**Exercício 4 - Adicione mais uma coluna gerada a partir da transformação de uma das 10 originais.**  *Add a new column which is a transformation, using some column of your dataframe.*

#adicionando a Densidade populacional  
#creating column density.  
  
Density\_sq <-round(df\_world$Population/df\_world$`Area (sq. mi.)`)  
  
df\_world <- data.frame(df\_world,Density\_sq)

**Exercício 5 - Faça um gráfico que mostre a distribuição dos dados dessa nova variável do exercício 4.** *Create a graph witch explain the new colunm made up on task 4.*

hist(df\_world$Density\_sq, breaks=10,main="Histogram Density");



plot(df\_world$GDP....per.capita.~df\_world$Density\_sq, type="p")

