

LABORATÓRIO 01

1. Working Directory

Configurando o diretório de trabalho

```
setwd("C:/Users/Utilizador/repos/Formacao_cientista_de_dados/big_data_analytics_R_microsoft_azure_machine_learning")
getwd()
```

2. Imports `install.packages("readr")` `install.packages("data.table")` `install.packages("dplyr")` `install.packages("ggplot2")` `library(readr)` `library(dplyr)` `library(ggplot2)` `library(scales)` `library(data.table)`

3. Data Loading

Usando read.csv2()

```
system.time(df_teste1 <- read.csv2("TemperaturasGlobais/TemperaturasGlobais.csv"))
```

Usando read.table()

```
system.time(df_teste2 <- read.table("TemperaturasGlobais/TemperaturasGlobais.csv"))
```

Usando fread()

Mais rápido

```
?fread system.time(df <- fread("TemperaturasGlobais/TemperaturasGlobais.csv"))
```

4. Feature engineering

Criando subsets dos dados carregados

```
idadesBrasil <- subset(df, Country == 'Brazil') cidadesBrasil <- na.omit(cidadesBrasil) head(cidadesBrasil)
View(cidadesBrasil) nrow(df) nrow(cidadesBrasil) dim(cidadesBrasil)
```

5. Data Cleaning

Preparação e Organização

Convertendo as Datas

```
idadesBrasildt <- as.POSIXct(cidadesBrasil,format='%Y-%m-%d') cidadesBrasilMonth <-  
month(cidadesBrasildt) cidadesBrasilYear <- year(cidadesBrasildt) View(cidadesBrasil)
```

6. Loading Subsets

Palmas

```
plm <- subset(cidadesBrasil, City == 'Palmas') plm <- subset(plm, Year %in% c(1796,1846,1896,1946,1996,2012))
```

Curitiba

```
crt <- subset(cidadesBrasil, City == 'Curitiba') crt <- subset(crt, Year %in% c(1796,1846,1896,1946,1996,2012))
```

Recife

```
recf <- subset(cidadesBrasil, City == 'Recife') recf <- subset(recf, Year %in% c(1796,1846,1896,1946,1996,2012))
```

7. Constrution Plots `p_plm <- ggplot(plm, aes(x = (Month), y = AverageTemperature, color = as.factor(Year))) + geom_smooth(se = FALSE, fill = NA, size = 2) + theme_light(base_size = 20) + xlab("Distribuição ao longo dos meses") + ylab("Temperatura Média") + scale_color_discrete("") + ggtitle("Temperatura Média ao longo dos anos em Palmas") + theme(plot.title = element_text(size = 18))`

`p_crt <- ggplot(crt, aes(x = (Month), y = AverageTemperature, color = as.factor(Year))) + geom_smooth(se = FALSE, fill = NA, size = 2) + theme_light(base_size = 20) + xlab("Distribuição ao longo dos meses") + ylab("Temperatura") + scale_color_discrete("") + ggtitle("Temperatura Média ao longo dos anos em Curitiba") + theme(plot.title = element_text(size = 18))`

`p_recf <- ggplot(recf, aes(x = (Month), y = AverageTemperature, color = as.factor(Year))) + geom_smooth(se = FALSE, fill = NA, size = 2) + theme_light(base_size = 20) + xlab("Distribuição ao longo dos meses") + ylab("Temperatura Média") + scale_color_discrete("") + ggtitle("Temperatura Média ao longo dos anos em Recife") + theme(plot.title = element_text(size = 18))`

8. Plot `p_plm p_crt p_recf`