

Combustíveis em Esteio

Média de Preço dos Combustíveis em Esteio 01/2018

Luciano Teixeira

14 de Outubro de 2018

Comandos R de preparação e transformação de dados

Resolva os exercícios a seguir utilizando os comandos no software RStudio através do RMarkdown. Gere um relatório em .doc com os comandos/código e as respectivas saídas e seus comentários.

- Exercício 1 - Crie um dataframe com 10 colunas e 40 linhas e imprima na tela as primeiras 6 linhas;
- Exercício 2 - Mostre a que classe pertence cada uma das 10 colunas do seu dataframe;
- Exercício 3 - Utilize um comando que para avaliação exploratória das variáveis do dataframe;
- Exercício 4 - Adicione mais uma coluna gerada a partir da transformação de uma das 10 originais;
- Exercício 5 - Faça um gráfico que mostre a distribuição dos dados dessa nova variável do exercício 4.

Dados da ANP - Agência Nacional do Petróleo

Preços de Combustíveis

Preços de combustíveis praticados no Brasil referentes ao primeiro semestre de 2018, disponibilizados voluntariamente pelos postos.

Resolução dos Exercícios

1 Crie um dataframe com 10 colunas e 40+ linhas

Carregando as bibliotecas para leitura do arquivo em csv

```
library(readr)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
library(readxl)
library(ggplot2)
library(stringi)
library(stringr)
library(car)
```

```
## Loading required package: carData
```

```
##
## Attaching package: 'car'

## The following object is masked from 'package:dplyr':
##
##      recode
```

Definindo a diretório de trabalho com a função `setwd()`.

```
setwd("~/GitHub/GeneralRepositoriesUnisinos/PosUnisinosIntroducaoPythonR/PrecoVoluntario")
```

Lendo o arquivo "2018-1_CA.csv".

```
EsteioBruto <- read_delim("2018-1_CA.csv", "\t",
                        escape_double = FALSE, locale = locale(decimal_mark = ",",
                                                                encoding = "ISO-8859-1"), trim_ws = TRUE)
```

```
## Parsed with column specification:
## cols(
##   `Região - Sigla` = col_character(),
##   `Estado - Sigla` = col_character(),
##   Município = col_character(),
##   Revenda = col_character(),
##   Produto = col_character(),
##   `Data da Coleta` = col_character(),
##   `Valor de Compra` = col_double(),
##   `Valor de Venda` = col_double(),
##   `Unidade de Medida` = col_character(),
##   Bandeira = col_character()
## )
```

Renomando as Variáveis

```
names(EsteioBruto) <- c("REGIAO", "UF", "CIDADE", "REVENDA", "PRODUTO", "DATA_COLETA", "VALOER_COMPRA", "VALOR_VENDA")
```

Imprimindo na tela as primeiras 6 linhas.

```
head(EsteioBruto)
```

```
## # A tibble: 6 x 10
##   REGIAO UF   CIDADE REVENDA PRODUTO DATA_COLETA VALOER_COMPRA VALOR_VENDA
##   <chr> <chr> <chr>   <chr>   <chr>   <chr>           <dbl>         <dbl>
## 1 CO    DF    BRASI~ AUTO P~ DIESEL  03/01/2018      NA           3.70
## 2 CO    DF    BRASI~ AUTO P~ DIESEL  03/01/2018    3.22         3.56
## 3 CO    DF    BRASI~ AUTO P~ DIESEL  03/01/2018    3.15         3.49
## 4 CO    DF    BRASI~ AUTO P~ DIESEL  03/01/2018      NA           3.20
## 5 CO    DF    BRASI~ BRASAL~ DIESEL  03/01/2018      NA           3.60
## 6 CO    DF    BRASI~ ESTAÇÃ~ DIESEL  03/01/2018      NA           3.43
## # ... with 2 more variables: UNIDADE <chr>, BANDEIRA <chr>
```

2 Mostre a que classe pertence cada uma das 10 colunas do seu dataframe

```
## [1] "REGIAO = character"
## [1] "UF = character"
## [1] "CIDADE = character"
## [1] "REVENDA = character"
## [1] "PRODUTO = character"
```

```
## [1] "DATA_COLETA = character"
## [1] "VALOER_COMPRA = numeric"
## [1] "VALOR_VENDA = numeric"
## [1] "UNIDADE = character"
## [1] "BANDEIRA = character"
```

3 Utilize um comando que faça uma avaliação exploratória das variáveis do seu dataframe e imprima na tela.

```
summary(EsteioBruto)
```

```
##      REGIAO              UF          CIDADE
## Length:469801   Length:469801   Length:469801
## Class :character Class :character Class :character
## Mode  :character Mode  :character Mode  :character
##
##
##
##      REVENDA      PRODUTO      DATA_COLETA      VALOER_COMPRA
## Length:469801   Length:469801   Length:469801   Min.    :1.24
## Class :character Class :character Class :character 1st Qu.:2.92
## Mode  :character Mode  :character Mode  :character Median :3.17
##                                     Mean   :3.24
##                                     3rd Qu.:3.65
##                                     Max.   :4.84
##                                     NA's   :281718
##
##      VALOR_VENDA      UNIDADE      BANDEIRA
## Min.    :1.760   Length:469801   Length:469801
## 1st Qu.:3.290   Class :character   Class :character
## Median :3.579   Mode  :character   Mode  :character
## Mean    :3.648
## 3rd Qu.:4.069
## Max.    :5.599
##
```

4 Adicione mais uma coluna gerada a partir da transformação de uma das 10 originais.

Neste processo, serão criados um dataset para cada combustível, a fim de calcular a média de cada um, no próximo passo, os datasets serão unidos em um só, para então gerar o plot das médias dos valores encontrados.

4.1 Criando os datasets por combustível.

```
# Filtrando a Cidade de Esteio x Produto Diesel

EsteioDiesel <-
  filter(
    select(
      subset.data.frame(EsteioBruto, UF == "RS"),
      REGIAO,
      UF,
      CIDADE,
```

```

    REVENDA,
    PRODUTO,
    DATA_COLETA,
    VALOER_COMPRA,
    VALOR_VENDA,
    UNIDADE,
    BANDEIRA
  ),
  CIDADE == "ESTEIO",
  PRODUTO == "DIESEL"

)

# Filtrando a Cidade de Esteio x Produto Diesel S10

EsteioDieselS10 <-
  filter(
    select(
      subset.data.frame(EsteioBruto, UF == "RS"),
      REGIAO,
      UF,
      CIDADE,
      REVENDA,
      PRODUTO,
      DATA_COLETA,
      VALOER_COMPRA,
      VALOR_VENDA,
      UNIDADE,
      BANDEIRA
    ),
    CIDADE == "ESTEIO",
    PRODUTO == "DIESEL S10"

  )

# Filtrando a Cidade de Esteio x Produto Etanol

EsteioEtanol <-
  filter(
    select(
      subset.data.frame(EsteioBruto, UF == "RS"),
      REGIAO,
      UF,
      CIDADE,
      REVENDA,
      PRODUTO,
      DATA_COLETA,
      VALOER_COMPRA,
      VALOR_VENDA,
      UNIDADE,
      BANDEIRA
    ),
    CIDADE == "ESTEIO",

```

```

    PRODUTO == "ETANOL"

)

# Filtrando a Cidade de Esteio x Produto Gasolina

EsteioGasolina <-
  filter(
    select(
      subset.data.frame(EsteioBruto, UF == "RS"),
      REGIAO,
      UF,
      CIDADE,
      REVENDA,
      PRODUTO,
      DATA_COLETA,
      VALOER_COMPRA,
      VALOR_VENDA,
      UNIDADE,
      BANDEIRA
    ),
    CIDADE == "ESTEIO",
    PRODUTO == "GASOLINA"

)

# Filtrando a Cidade de Esteio x Produto GNV

EsteioGnv <-
  filter(
    select(
      subset.data.frame(EsteioBruto, UF == "RS"),
      REGIAO,
      UF,
      CIDADE,
      REVENDA,
      PRODUTO,
      DATA_COLETA,
      VALOER_COMPRA,
      VALOR_VENDA,
      UNIDADE,
      BANDEIRA
    ),
    CIDADE == "ESTEIO",
    PRODUTO == "GNV"

)

```

4.2 Calculando a média dos valores por combustível e adicionando um coluna com o respectivo valor da média.

```

# Diesel

Media <- mean(EsteioDiesel$VALOR_VENDA)

```

```

EsteioDiesel <- data.frame(EsteioDiesel, Media)

# Diesel S10

Media <- mean(EsteioDieselS10$VALOR_VENDA)

EsteioDieselS10 <- data.frame(EsteioDieselS10, Media)

# Etanol

Media <- mean(EsteioEtanol$VALOR_VENDA)

EsteioEtanol <- data.frame(EsteioEtanol, Media)

# Gasolina

Media <- mean(EsteioGasolina$VALOR_VENDA)

EsteioGasolina <- data.frame(EsteioGasolina, Media)

# GNV

Media <- mean(EsteioGnv$VALOR_VENDA)

EsteioGnv <- data.frame(EsteioGnv, Media)

```

5 Faça um gráfico que mostre a distribuição dos dados dessa nova variável do exercício 4.

Nesta etapa, os datasets anteriormente criados em separado de acordo com cada tipo de combustível, serão base para criação dos seus respectivos gráficos x combustível.

5.1 Unindo os datasets

```
Esteio <- bind_rows(EsteioDiesel, EsteioDieselS10, EsteioEtanol, EsteioGasolina, EsteioGnv)
```

5.2 Boxplo de cada tipo de combustível.

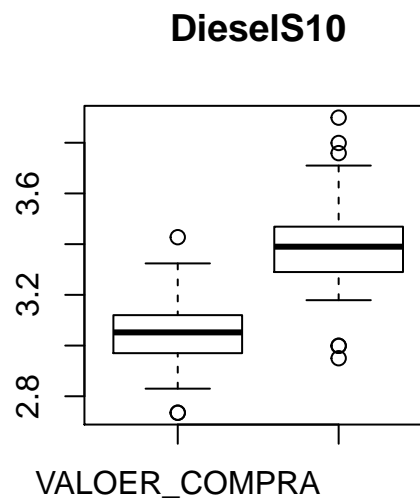
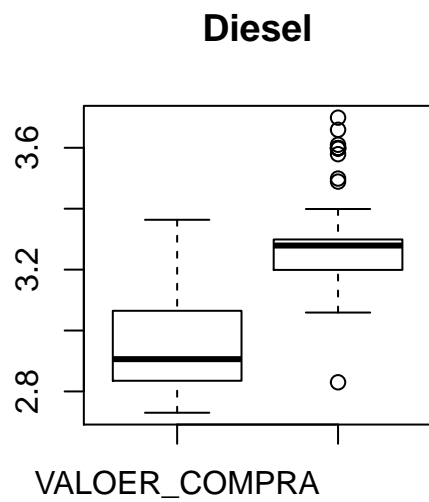
```

par(mfrow = c(1,2), oma = c(4,1,1,1))

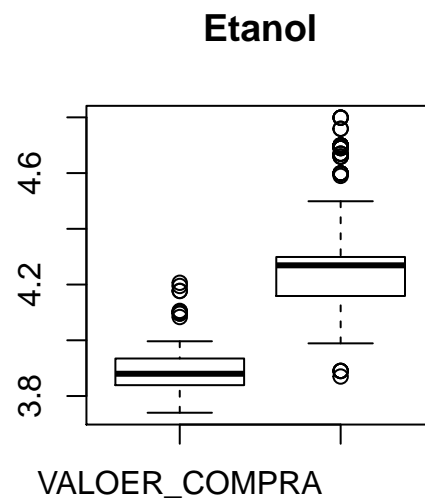
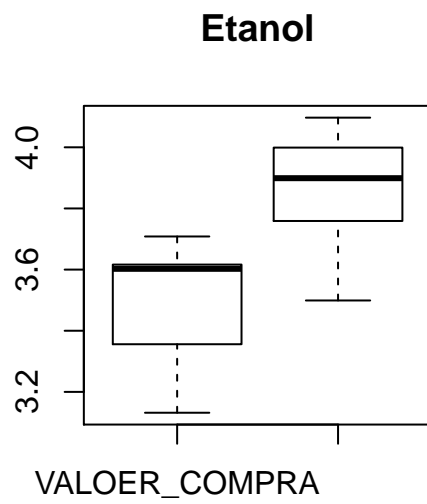
boxplot(EsteioDiesel[,7:8], main = "Diesel")

boxplot(EsteioDieselS10[,7:8], main = "DieselS10")

```



```
par(mfrow = c(1,2), oma = c(4,1,1,1))
boxplot(EsteioEtanol[,7:8], main = "Etanol")
boxplot(EsteioGasolina[,7:8], main = "Etanol")
```



```
par(mfrow = c(1,2), oma = c(4,1,1,1))
boxplot(EsteioGnv[,7:8], main = "GNV")
```