

# ANP Brasil

Evolução do Preço dos combustíveis em 2012

*Luciano Teixeira*

*11 de setembro 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 faça uma avaliação exploratória das variáveis do seu dataframe e imprima na tela;
- 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

### Evolução de Preço dos Combustíveis no último Bimestre de 2012

#### Resolução dos Exercícios

1. Crie um dataframe com 10 colunas e 40+ linhas

Carregando a biblioteca `readr` para leitura do arquivo em csv

```
library(readr)
```

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

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

Lendo o arquivo "brasil.csv".

```
anpbrasil <- read_delim("brasil.csv", ";", escape_double = FALSE, trim_ws = TRUE)
```

```
## Parsed with column specification:
## cols(
##   DATA_INICIAL = col_character(),
##   DATA_FINAL = col_character(),
##   PRODUTO = col_character(),
##   POSTOS_PESQUISADOS = col_integer(),
##   UNIDADE_MEDIDA = col_character(),
##   PRECO_MEDIO_VENDA = col_number(),
##   DESVIO_PADRAO_REVENDA = col_character(),
##   PRECO_MINIMO_REVENDA = col_number(),
##   PRECO_MAXIMO_REVENDA = col_number(),
##   MARGEM_MEDIA_REVENDA = col_character(),
##   COEF_DE_VARIACAO_REVENDA = col_character(),
##   PRECO_MEDIO_DISTRIBUICAO = col_number(),
```

```
## DESVIO_PADRAO_DISTRIBUICAO = col_character(),
## PRECO_MINIMO_DISTRIBUICAO = col_character(),
## PRECO_MAXIMO_DISTRIBUICAO = col_number(),
## COEF_DE_VARIACAO_DISTRIBUICAO = col_character()
## )
```

Imprimindo na tela as primeiras 6 linhas.

```
head(anpbrasil)
```

```
## # A tibble: 6 x 16
##   DATA_INICIAL DATA_FINAL PRODUTO          POSTOS_PESQUISAD~ UNIDADE_MEDIDA
##   <chr>          <chr>      <chr>          <int> <chr>
## 1 04/11/2012    10/11/2012 ETANOL HIDRATA~      8195 R$/l
## 2 11/11/2012    17/11/2012 ETANOL HIDRATA~      8184 R$/l
## 3 18/11/2012    24/11/2012 ETANOL HIDRATA~      8176 R$/l
## 4 25/11/2012    01/12/2012 ETANOL HIDRATA~      8164 R$/l
## 5 02/12/2012    08/12/2012 ETANOL HIDRATA~      8168 R$/l
## 6 09/12/2012    15/12/2012 ETANOL HIDRATA~      8155 R$/l
## # ... with 11 more variables: PRECO_MEDIO_VENDA <dbl>,
## #   DESVIO_PADRAO_REVENDA <chr>, PRECO_MINIMO_REVENDA <dbl>,
## #   PRECO_MAXIMO_REVENDA <dbl>, MARGEM_MEDIA_REVENDA <chr>,
## #   COEF_DE_VARIACAO_REVENDA <chr>, PRECO_MEDIO_DISTRIBUICAO <dbl>,
## #   DESVIO_PADRAO_DISTRIBUICAO <chr>, PRECO_MINIMO_DISTRIBUICAO <chr>,
## #   PRECO_MAXIMO_DISTRIBUICAO <dbl>, COEF_DE_VARIACAO_DISTRIBUICAO <chr>
```

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

```
## [1] "DATA_INICIAL = character"
## [1] "DATA_INICIAL = character"
## [1] "DATA_FINAL = character"
## [1] "PRODUTO = character"
## [1] "POSTOS_PESQUISADOS = integer"
## [1] "UNIDADE_MEDIDA = character"
## [1] "PRECO_MEDIO_VENDA = numeric"
## [1] "DESVIO_PADRAO_REVENDA = character"
## [1] "PRECO_MINIMO_REVENDA = numeric"
## [1] "PRECO_MAXIMO_REVENDA\t= numeric"
## [1] "MARGEM_MEDIA_REVENDA\t= character"
## [1] "COEF_DE_VARIACAO_REVENDA = character"
## [1] "PRECO_MEDIO_DISTRIBUICAO\t= numeric"
## [1] "DESVIO_PADRAO_DISTRIBUICAO = character"
## [1] "PRECO_MINIMO_DISTRIBUICAO = character"
## [1] "PRECO_MAXIMO_DISTRIBUICAO = numeric"
## [1] "COEF_DE_VARIACAO_DISTRIBUICAO = character"
```

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

```
summary(anpbrasil)
```

```
## DATA_INICIAL          DATA_FINAL          PRODUTO
## Length:40              Length:40             Length:40
## Class :character       Class :character      Class :character
## Mode :character        Mode :character       Mode :character
##
##
##
## POSTOS_PESQUISADOS UNIDADE_MEDIDA          PRECO_MEDIO_VENDA
## Min. : 491          Length:40             Min. : 40
## 1st Qu.:7078        Class :character      1st Qu.: 1758
## Median :8052        Mode :character     Median : 2146
## Mean : 6499                                     Mean : 8612
## 3rd Qu.:8181                                     3rd Qu.: 2749
## Max. : 8685                                     Max. : 40151
## DESVIO_PADRAO_REVENDA PRECO_MINIMO_REVENDA PRECO_MAXIMO_REVENDA
## Length:40              Min. : 18           Min. : 62
## Class :character       1st Qu.:1021          1st Qu.: 295
## Mode :character        Median :1399          Median : 312
##                        Mean : 1298           Mean : 1378
##                        3rd Qu.:1869          3rd Qu.:2449
##                        Max. : 2279           Max. : 3799
## MARGEM_MEDIA_REVENDA COEF_DE_VARIACAO_REVENDA PRECO_MEDIO_DISTRIBUICAO
## Length:40              Length:40             Min. : 157
## Class :character       Class :character      1st Qu.: 1599
## Mode :character        Mode :character     Median : 1927
##                        Mean : 7366
##                        3rd Qu.: 2365
##                        Max. : 29857
## DESVIO_PADRAO_DISTRIBUICAO PRECO_MINIMO_DISTRIBUICAO
## Length:40              Length:40
## Class :character       Class :character
## Mode :character        Mode :character
##
##
##
## PRECO_MAXIMO_DISTRIBUICAO COEF_DE_VARIACAO_DISTRIBUICAO
## Min. : 44.0          Length:40
## 1st Qu.: 242.0        Class :character
## Median : 289.0        Mode :character
## Mean : 938.5
## 3rd Qu.:1846.0
## Max. : 2927.0
```

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

```
anpbrasil$TOTAL_POSTOS<-(anpbrasil$POSTOS_PESQUISADOS/sum(anpbrasil$POSTOS_PESQUISADOS))/100
```

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