

Relatório de análise dos dados do Cepagri

Integrantes

- **19164** - Bruno Arnone Franchi
- **19191** - Nicolas Denadai Schmidt

Introdução

Este relatório descreve as análises feitas por nosso grupo sobre os dados meteorológicos do Cepagri/Unicamp, no período correspondido de 01/01/2015 a 31/12/2020. Os dados foram lidos de um arquivo CSV, disponível aqui.

```
# ler dataset .csv
names <- c("horario", "temp", "vento", "umid", "sensa")
cepagri <- read.csv("cepagri.csv", header = FALSE, sep = ";", col.names = names)
```

Tratamento de dados

Fizemos um tratamento inicial aos tipos de dados, para que a coluna de `horario` fosse lida como Data; e a coluna `temp` fosse, assim como as outras, lida como Número. Além disso, removemos as entradas com erros de leitura.

```
# criar dataframe, tratar tipos de coluna e erros
df <- data.frame(cepagri)
# ignorar warning de leituras N/A, pois vamos corrigir isso na próxima linha
suppressWarnings({
  df <- df %>%
    mutate(horario = as_datetime(df$horario, format = "%d/%m/%Y-%H:%M")) %>%
    mutate(temp = as.numeric(df$temp))
})
df <- na.omit(df)
```

Depois, extraímos o período de tempo que vamos estudar: do primeiro dia de 2015 ao último dia de 2020.

```
# extrair dados entre 2015-01-01 e 2020-12-31
df <- df[
  df$horario >= "2015-01-01" &
  df$horario <= "2020-12-31",
]
```

Média diária

