

## 5ª LISTA – Manipulação de Objetos usando o Tidyverse

**Disciplina:** R para Ciência de Dados.

**Professor:** Alan Rodrigo Panosso

**PPG:** Agronomia (Ciência do Solo)

---

1) Abrir o arquivo `aula05.Rmd` criado na aula passada. Realizar a importação dos dados de emissão de CO<sub>2</sub> do solo e demais atributos (`data/emissao-co2-solo.rds`) e apresentar os resumos dos dados por meio da função `glimpse`.

Lembre-se:

- para adicionar um `chunk` (pedaço de código em R) use o atalho `CONTROL+ALT+I`
- para evocar o operador PIPE (`|>`) use o atalho `CONTROL+SHIFT+M`

2) Filtrar os dados apenas para avaliações de séries temporais, identificados como a categoria "Temporal" na coluna `experimento` (`filter`).

3) Selecione as colunas `data`, `tratamento`, `id`, `fco2`, `ts` e `us` (`select`).

4) A partir da coluna `data`, crie três outras colunas (`mutate`) para `dia`, `mês` e `ano`, utilize as respectivas funções do pacote `{lubridate}`.

5) Reorganize a ordem das colunas `data`, `dia`, `mes`, `ano`...demais colunas (`relocate`).

6) Extraia a coluna `ano` (`pull`), e indique quantos anos estão nesse subconjunto (`unique`).

7) Filtre todas as observações para o ano de 2017, e salve esse `data.frame` em um novo objeto denomina "`data_set_2017`".

8) Descubra quantos tratamentos foram avaliados nesse ano (`pull`, `unique`).

9) Reorganize os dados dos maiores para os menos valores de emissão de CO<sub>2</sub> do solo (`fco2`).

10) Apresente as 10 primeiras linhas da tabela (`head`) e, em seguida, apresente as 10 últimas linhas da tabela (`tail`).

11) Criar uma nova coluna `classe_emissao`, de forma que:

- `fco2` menor ou igual a 2,92: "Baixa";
- `fco2` entre 2,93 e 5,41: "Média";
- `fco2` entre 5,42 e 7,9: "Alta";
- `fco2` maior que 7,9: "Muito Alta";

Dica: utilize as funções `mutate` e `case_when`.

12) Crie uma tabela apresentando o número de observações classificadas em cada classe (`count`).

13) Calcule a média da emissão de CO<sub>2</sub> do solo por tratamento (`group_by`, `summarise` e `mean`).

14) Descubra a emissão máxima e mínima de CO<sub>2</sub> em cada dia, nos diferentes tratamentos.

- 15) Calcule a média de temperatura do solo ( $t_s$ ) e a umidade do solo ( $u_s$ ) por tratamento e dia.
- 16) Criar uma função para calcular a estatística descritiva de um vetor numérico. `summary()`.
- 17) Aplicar essa função para todas as variáveis numéricas do banco de dados `data_set`, utilize as funções `summarise` e `cross`.
- 18) Salve a tabela da estatística descritiva no formato `.xlsx` (Excel) para isso utilize a função `write_xlsx` do pacote `{writexl}`.
- 19) Utilizando o `ggplot2`, crie um gráfico de dispersão entre as variáveis temperatura e umidade do solo.
- 20) Para o gráfico apresentado anteriormente, mapeie os tratamentos nas cores dos pontos.
- 21) Criar um gráfico de dispersão entre umidade e temperatura do solo para cada tratamento.
- 22) Adicione uma linha de tendência do tipo linear ao gráfico.
- 23) Adicione uma linha de tendência polinomial de ordem 2 (quadrática).
- 24) Para adicionar as equações ao gráfico, explore as opções do pacote `{ggpubr}`.
- 25) Construir o gráfico de colunas para média de emissão de  $CO_2$  do solo para cada tratamento, em cada dia de avaliação.
- 26) Alterar a escala de cores para `viridis`.
- 27) Construir um histograma para `fco2`.
- 28) Construir um boxplot para `fco2`, em seguida construir um boxplot para cada tratamentos nos diferentes dias de avaliação.