

Ficha de trabalho 2

Introdução ao uso do R e do RStudio com visualizações

Exercício 1

1. Obtenha o ficheiro `Dataset_02.xls`
2. Leia-o no RStudio usando a biblioteca `readxl` e a função `read_excel()` (sugestão: use a função `file.choose()`)
3. Realize análise exploratória de dados (EDA)
 - 3.1. Quantas observações e variáveis existem neste dataset?
 - 3.2. Inspeccione as primeiras 25 linhas
 - 3.3. Tente descrever de uma forma geral o que representam as colunas
4. Observe a coluna `Grade` e crie uma visualização no `ggplot` para essa coluna
Nota: reflita sobre o que pretende visualizar!
5. Crie um gráfico de linhas com as variáveis `Similarity` e `Distance_part_percent`
 - 5.1. O que pode concluir?
6. Crie um gráfico que estabeleça o nível de correlação entre as variáveis:
 - `Grade`
 - `String length`
 - `Distance`
 - `Similarity`
 - `Distance_part_percent`
 - 6.1. O que pode concluir?
7. Obtenha o ficheiro `Dataset_03.xls`
8. Leia-o no RStudio usando a biblioteca `readxl` e a função `read_excel()`
9. Como compararia os dois datasets em função das análises anteriores?
10. Exponhas as suas principais conclusões retiradas desta comparação.