



Universidade de Brasília

DEPARTAMENTO DE ESTATÍSTICA

18 setembro 2023

Lista 1: dplyr e stringr

Prof^a. Thais Carvalho

Computação em Estatística 2 - 2º/2023

Instruções

- 1) As questões deverão ser respondidas em um único relatório PDF ou HTML, produzido usando as funcionalidades do Rmarkdown.
- 2) O trabalho é individual. Suspeitas de plágio e compartilhamento de soluções serão tratadas com rigor.
- 3) Os códigos R utilizados devem estar presentes no arquivo PDF ou HTML, assim como o resultado do código.
- 4) O aluno deverá enviar o trabalho, impreterivelmente, até 24/09/23 (domingo) às 23h pelo Teams.
- 5) O trabalho será avaliado considerando o nível de qualidade do relatório, o que inclui a precisão das respostas, a pertinência das soluções encontradas, a formatação adotada, dentre outros aspectos correlatos.
- 6) Escreva seu código com esmero, evitando operações redundantes, visando eficiência computacional, comentando os resultados e usando as melhores práticas em programação.

Questão 1:

Peça ao ChatGPT gerar um texto de 4 parágrafos sobre um tema de sua escolha. Copie o texto gerado e salve no objeto “texto” (utilize \n para quebra de linha). Em seguida, faça o que se pede utilizando o pacote *stringr*.

- a) Conte quantas sentenças existem no texto.
- b) Remova todos os sinais de pontuação e quebras de linha do texto.
- c) Separe os parágrafos do texto criado (cada parágrafo ocupa um elemento de um vetor). Em seguida, remova espaços em branco do início e fim de cada objeto.
- d) Encontre no texto todas as palavras que contenham mais de 3 caracteres
- e) Encontre as top 6 palavras mais citadas no texto (que contenham mais de 3 caracteres).

Questão 2:

Carregue o banco de dados imdb.csv que contém informações sobre filmes e responda os itens a seguir.

- a) Qual é a proporção de filmes com orçamento igual ou maior do que 10 milhões?
- b) Liste os filmes que apresentam a variável id_filme com mais de 9 caracteres?
- c) Acrescente 1 variável *CATEGÓRICA* no banco de dados chamada “classificacao_nota” informando o primeiro dígito da nota_imdb.
- d) Escolha o nome de 1 ator do elenco e retorne todos os filmes que este ator atuou.