

DEPARTAMENTO DE ESTATÍSTICA

09 outubro 2023

Lista 2: forcats, hms e lubridate

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 15/10/23 (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.

Para esta lista, iremos utilizar como inspiração um pedaço do banco de dados sobre corridas de aplicativos de transporte na cidade de Boston (arquivo ‘uber.csv’).

Questão 1:

Com base no banco de dados ‘uber.csv’, responda os itens a seguir.

- a) Acrescente 1 coluna no banco de dados com a informação completa de data e hora de cada corrida.
- b) Acrescente 1 coluna no banco de dados com o dia da semana de cada corrida.
- c) Faça uma tabela com a contagem do número de viagens por dia da semana.
- d) Faça uma tabela com o preço médio da corrida e a temperatura média para cada dia da semana.

Questão 2:

Com base no banco de dados ‘uber.csv’, responda os itens a seguir.

- a) Faça uma tabela com a contagem de corridas para cada hora do dia.
- b) Com o resultado obtido no item anterior, faça um gráfico da contagem de viagens por hora do dia, ordenando da hora com maior contagem observada para a hora com menor contagem observada (utilize o pacote forcats).
- c) Acrescente 1 coluna no banco informando se a corrida foi realizada pela “manhã” (entre 0h e 11h59), “tarde” (entre 12h e 17h59), e o restante como “noite”. (Para isso utilize a função ‘case_when’ do pacote ‘dplyr’).
- d) Faça um gráfico da contagem de viagens por período do dia (manhã, tarde e noite), ordenando do período com maior contagem para o período com menor contagem. Você esperava obter estes resultados após observar os resultados obtidos na questão 2b? (utilize o pacote forcats)
- e) Obtenha o preço e a distância média das viagens em cada turno do dia. Para estes dados, o turno do dia aparenta influenciar o preço médio e a distância média das viagens?