

# Curso de Especialização em Data Science e Estatística Aplicada

## Módulo I - Introdução à Linguagem R

### Atividade Avaliativa

Profa. Dra. Juliana Félix

26/10/2024

## Instruções

- O desenvolvimento desta atividade deve ser realizada de forma individual.
- Deve-se completar o arquivo Rmd enviado na atividade.
- É necessário devolver o arquivo em Rmd e em pdf.
- Valor da atividade: 10 pontos.

**Data disponível: 26/10/2024**

**Data máxima para entrega: 22/11/2024**

## Atividade

Nesta atividade avaliativa, o objetivo é reforçar alguns conceitos relacionados a Big Data e explorar as ferramentas de manipulação e visualização de dados apresentadas na disciplina. As 2 primeiras questões são discursivas. As demais questões referem-se a manipulação de dados, e o tipo de resposta esperada é identificado em cada questão.

### Questões discursivas

**(2 pontos) Questão 1: Comente brevemente sobre a importância do processamento paralelo para a manipulação de Big Data.**

**(1 ponto) Questão 2: Como o uso de consultas SQL com o BigQuery pode facilitar a manipulação de Big Data?**

### Questões práticas

Para as próximas questões, o conjunto de dados

`bigquery-public-data.covid19_tracking.city_level_cases_and_deaths`,

já apresentado nos exemplos durante as aulas deverá ser utilizado. Você deve, preferencialmente, acessar a tabela a partir do BigQuery para facilitar suas consultas e análises. Entretanto, a tabela completa também foi

disponibilizada para download na página do curso e pode ser utilizada em outras ferramentas de manipulação de dados de sua preferência.

A partir da tabela de dados, faça o que se pede:

**(2 pontos) Questão 3:** Identifique o total de óbitos por estado. Apresente uma screenshot de uma visualização em forma de mapa, feita no PowerBI, apontando os estados e a quantidade de óbitos em forma de bolhas (quanto mais casos, maior a bolha).

**(2 pontos) Questão 4:** Identifique a tendência de casos e de mortes ao longo do tempo, considerando a informação de novos casos semanais. Apresente o código em R que obtém essas informações e que gera uma visualização dessas tendências ao longo do tempo a partir de um gráfico de linha.

**(3 pontos) Questão 5:** Identifique a quantidade de casos e de mortes, por etnia, no estado da Califórnia. Apresente a consulta SQL que gera a tabela desejada, e uma screenshot de um dashboard com pelo menos 3 visualizações distintas para suportar essa análise.