

Universidade Eduardo Mondlane
Faculdade de Ciências
Departamento de Matemática e Informática
Programação Estatística
Teste I

O presente teste tem duração de 3 horas e 25 minutos (com início as 17:30 e término as 21:00). O teste deve ser feito no Rstudio (abra um novo R script). No seu R script, logo no início escreva o seu nome completo e no seu número de estudante)(Esta informação deve aparecer como um comentário). Não inclua no seu script nenhum output, apenas coloque o código para responder cada questão. Para cada questão, coloque primeiro a pergunta e de seguida a resposta (código). Salve o seu R script com o nome:

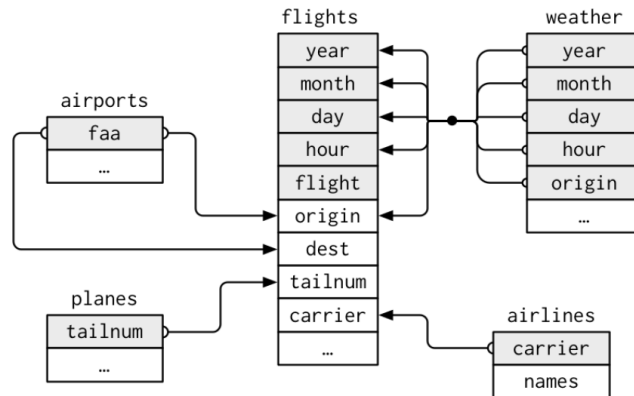
PRIMEIRONOME_ULTIOMONOME_TESTEI. Após terminar o seu teste, submeta o seu teste **AQUI. PARA EVITAR CONSTRANGIMENTOS, SUMETA O SEU TESTE 5 MINUTOS ANTES DAS 21:00. O LINK PARA SUMISSÃO IRÁ FECHAR PONTUALMENTE AS 21:00**

Como o R script deve estar organizado

```
#####  
## MODULO: PROGRAMACAO ESTATISTICA          ####  
## ACTIVIDADE: TESTE I                      ####  
## NOME DO ESTUDANTE: RACHID MULEIA        ####  
## NUMERO DO ESTUDANTE: 00002023          ####  
#####
```

Parte I (8 valores)

Considere as data frames da livreria `nycflights13` e responda as questões a seguir. A várias data frames da livreria `nycflights13` estão relacionadas conforme a figura a baixo.



- Que destino recebeu mais voo no mês de Junho?
- Que operadora teve a maior distância média por voo?
- Que voo viajou mais rápido (milhas totais por hora)?
- Qual dia teve o maior atraso médio de chegada para todos os voos?
- Qual foi a distância total para todos os voos em Janeiro? Qual foi a distância média por voo?
- Em que dia da semana houve mais voos? (Dica, use a coluna “`time_hour`” que é uma coluna de data/hora e use o google (pesquisa na Internet) para encontrar uma função para fornecer o dia da semana para uma data).
- Qual foi o número médio de assentos e motores nos aviões que partiram de Nova York em julho 4? (Os dados de “assentos” e “motores” vêm da tabela de aviões; Aviso: tenha cuidado com suas junções!. Deve aplicar as operações de merge)
- Qual foi o modelo de avião mais comum a sair de Nova York em Outubro (há uma coluna “modelo” na tabela de aviões)?
- Quantos aviões (tailnum) voaram apenas uma rota (voo), mas voaram nessa rota mais

de 10 vezes?

- j) Que hora de partida programada (use a coluna “hora”) teve a maior proporção de voos atrasados (`dep_delay`) por mais de 5 minutos?
- k) Qual(is) voo(s) tiveram a maior duração programada (tempo entre a partida e a chegada programadas)?

Parte II (6 valores)

Para a resolução dos exercícios a seguir considere NHANES disponível na livraria NHANES.

- a) Verifique se a base de dados tem algum valor omisso/missing.
- b) Quantas células vazias existem na base de dados?
- c) Identifique a coluna com maior numero de valores omissos.
- d) Qual das observações/participantes tem a maior taxa de valores omissos em todas a variáveis?
- e) Quantos indivíduos da sexo masculino são negros?
- f) Calcule algumas medidas sumárias (Estatísticas descritivas) para a variável idade?
- g) Categorize a variável idade em intervalos de classe de amplitude 10.
- h) Quantos participantes são obesos cuja idade está acima de 60 anos?
- i) Quantas linhas duplicadas existem na base de dados?
- j) Elimine todos os duplicados na base de dados.
- k) Calcule a pressão arterial média para cada faixa etária
- l) Qual é a pressão arterial média para participantes masculinos, com idade acima do primeiro quartil que sejam casados?
- m) Crie uma nova variável que seja a combinação da variável `Gender`, `Age` e `Race1`.

Parte III (6 valores)

1. Escreva uma função que recebe um vector de inteiros e devolve, a soma, média, o primeiro e terceiro quartil.
2. Escreva uma função que recebe três vectores e devolve um vector concatenado.
3. Escreva uma função que devolve a n-ésima posição de um vector. A função vai receber dois argumentos, um vector e um inteiro que pretende aceder.
4. Crie um vector de comprimento 100: ('label 1', 'label 2', 'label 3', ..., 'label 100').
5. Escreva um código R para importar os dados `Employee_Payroll_Tab`. Os dados estão disponíveis [AQUI](#)
6. Escreva um código R para importar os dados `Employee_Payroll_Pipe`. Estes dados podem ser encontrados [AQUI](#).
7. Salve os dados importados no exercício 6 e 7 em um formato suportado pelo SPSS.
8. Escreva um código para mudar o nome da coluna `Position ID`, na base de dados `Employee_Payroll_Pipe`, para `Identificacao_Posicao`. Escreva um código assumindo que não conhece a posição da coluna que pretende alterar o nome.
9. Escreva um código para alterar na variável `office Name` o valor `STATES ATTORNEY` para `PROCURADOR DO ESTADO`.