Sumário

[Projeto 2](#_Toc129291490)

[Evidências 4](#_Toc129291491)

# Projeto

Antonio,

Nessa disciplina, aprendemos os fundamentos básicos da linguagem R e de estatística, que serão fundamentais para realizar análises descritivas de bases de dados, tarefa que é extremamente importante para o dia-a-dia de um cientista de dados. Agora iremos validar nosso conhecimento. Todas as respostas desse projeto devem ser feitas em um arquivo RMarkdown.

1. Mostre através de prints que você tem acesso a uma plataforma RStudio (instalado localmente ou nuvem).
2. Escolha uma base de dados para realizar esse projeto. Essa base de dados será utilizada durante toda sua análise. Essa base necessita ter 4 (ou mais) variáveis de interesse. Caso você tenha dificuldade para escolher uma base, o professor da disciplina irá designar para você.
3. Explique qual o motivo para a escolha dessa base e explique os resultados esperados através da análise.
4. Carregue a base para o RStudio e comprove o carregamento tirando um print da tela com a base escolhida presente na área "Ambiente"/Enviroment. Detalhe como você realizou o carregamento dos dados.
5. Instale e carregue os pacotes de R necessários para sua análise (mostre o código necessário):
   1. tidyverse
   2. ggplot
   3. summarytools
6. Escolha outros pacotes necessários, aponte sua necessidade e instale e carregue (mostrando o código necessário).
7. Aplique uma função em R que seja útil para sua análise e mostre.
8. Escolha uma variável de seu banco de dados e calcule:
   1. a média para todos os eventos
   2. o desvio padrão
   3. os quantis: 25% e 75%
9. Utilizando o pacote summarytools (função descr), descreva estatisticamente a sua base de dados.
10. Escolha uma variável e crie um histograma. Justifique o número de bins usados. A distribuição dessa variável se aproxima de uma "normal"? Justifique.
11. Calcule a correlação entre todas as variáveis dessa base. Quais são as 3 pares de variáveis mais correlacionadas?
12. Crie um scatterplot entre duas variáveis das resposta anterior. Qual a relação da imagem com a correlação entre as variáveis.
13. Crie um gráfico linha de duas das variáveis. Acrescente uma legenda e rótulos nos eixos.

Assim que terminar, salve o seu arquivo PDF e poste no Moodle. Utilize o seu nome para nomear o arquivo, identificando também a disciplina no seguinte formato: “nomedoaluno\_nomedadisciplina\_pd.PDF”.

# Evidências

1. Notebook do aluno – Antonio Vieira dos Santos Neto – CPF 077.523.948-82

Interface gráfica do usuário, Texto, Aplicativo

Descrição gerada automaticamente

1. Evidência do ambiente para desenvolvimento de Projetos na linguagem “R”
   1. Linguagem “R” e “RStudio” instalados.

Interface gráfica do usuário, Aplicativo

Descrição gerada automaticamente

1. Interface de desenvolvimento (Enviroment) - “RStudio”.

Tela de computador com texto preto sobre fundo branco

Descrição gerada automaticamente

1. Ambiente “GitHub”

<https://github.com/avsneto2/work.git>

Tela de computador com texto preto sobre fundo branco

Descrição gerada automaticamente

**Branch : Projeto\_Estatistica\_1**

Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente