

Instituto Infnet  
Pós Graduação - Data Science  
Estatística para Cientistas de Dados [23E1\_2]

## Sumário

---

Projeto .....	2
Evidências .....	4

## 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"/Environment. Detalhe como você realizou o carregamento dos dados.
5. Instale e carregue os pacotes de R necessários para sua análise (mostre o código necessário):
  - a. tidyverse
  - b. ggplot
  - c. 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:
  - a. a média para todos os eventos

Instituto Infnet  
Pós Graduação - Data Science  
Estatística para Cientistas de Dados [23E1\_2]

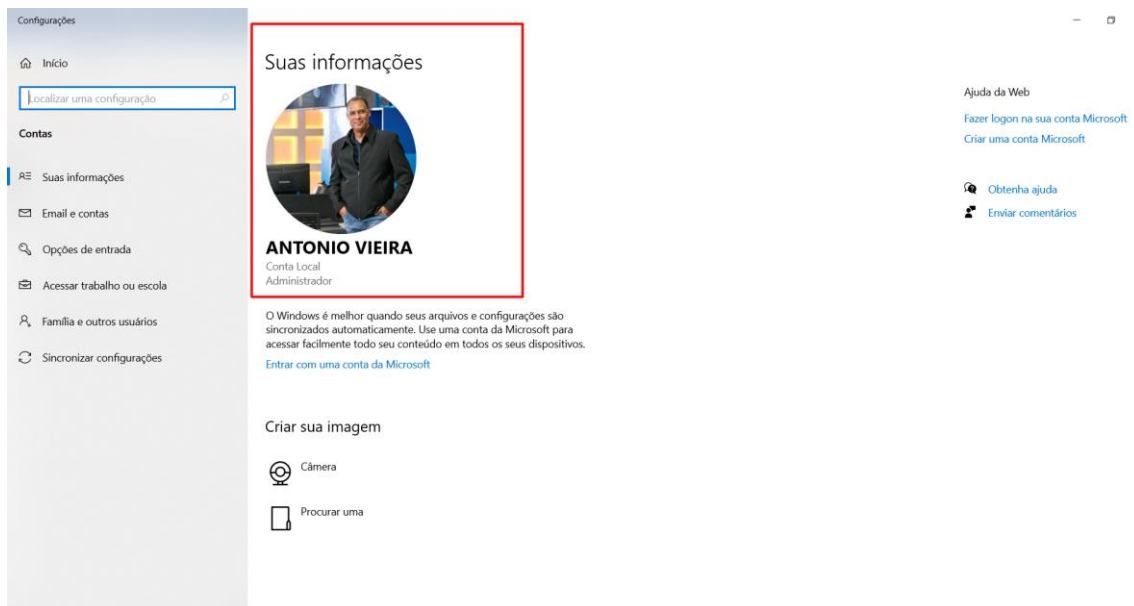
- b. o desvio padrão
  - c. os quantis: 25% e 75%
9. Utilizando o pacote summarytools (função descr), descreva estatisticamente a sua base de dados.
  10. Escolha uma variável e crie um histograma. Justifique o número de bins usados. A distribuição dessa variável se aproxima de uma "normal"? Justifique.
  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".

Instituto Infnet  
Pós Graduação - Data Science  
Estatística para Cientistas de Dados [23E1\_2]

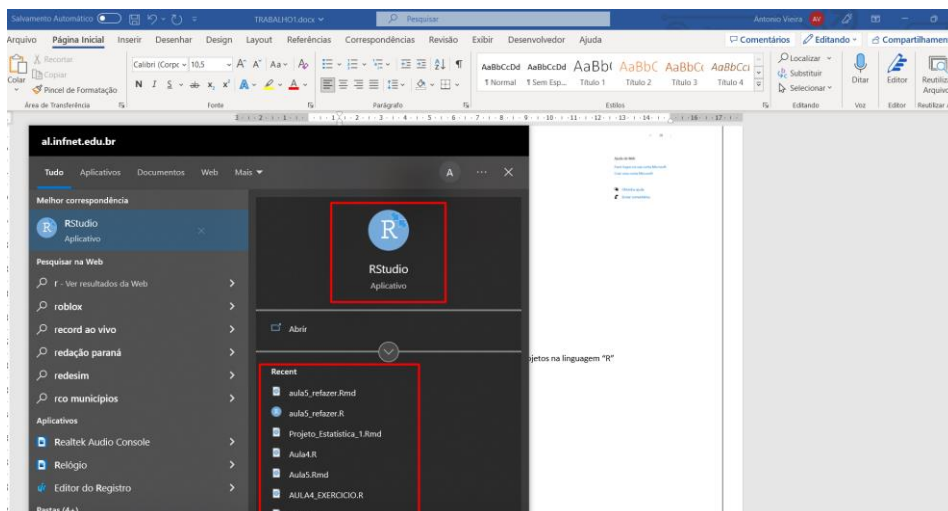
## Evidências

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



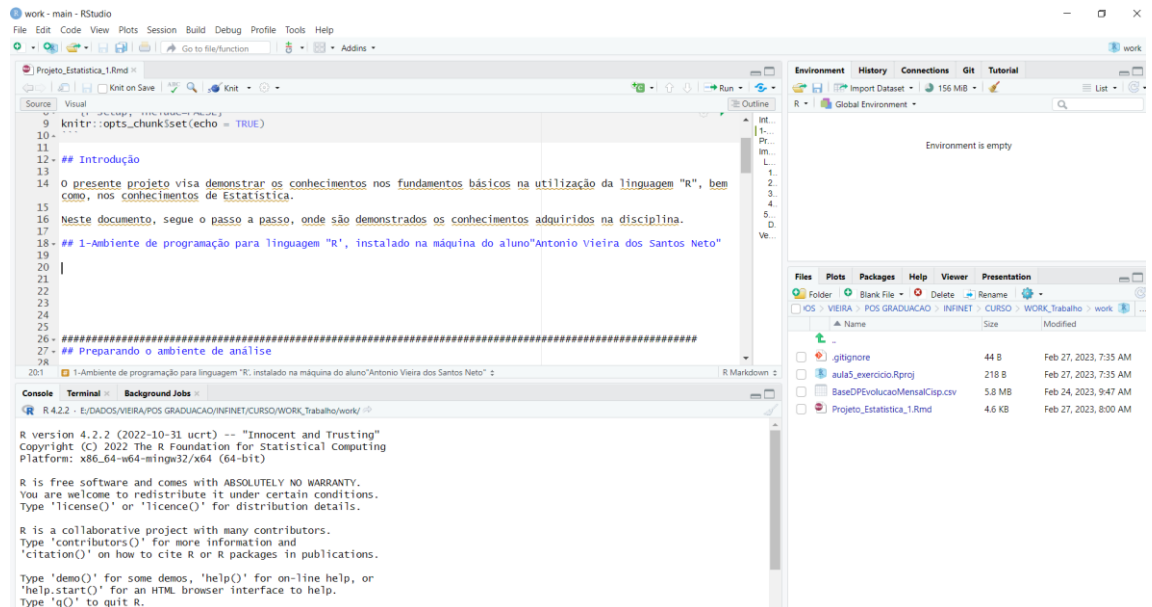
### 2. Evidência do ambiente para desenvolvimento de Projetos na linguagem “R”

#### a. Linguagem “R” e “RStudio” instalados.



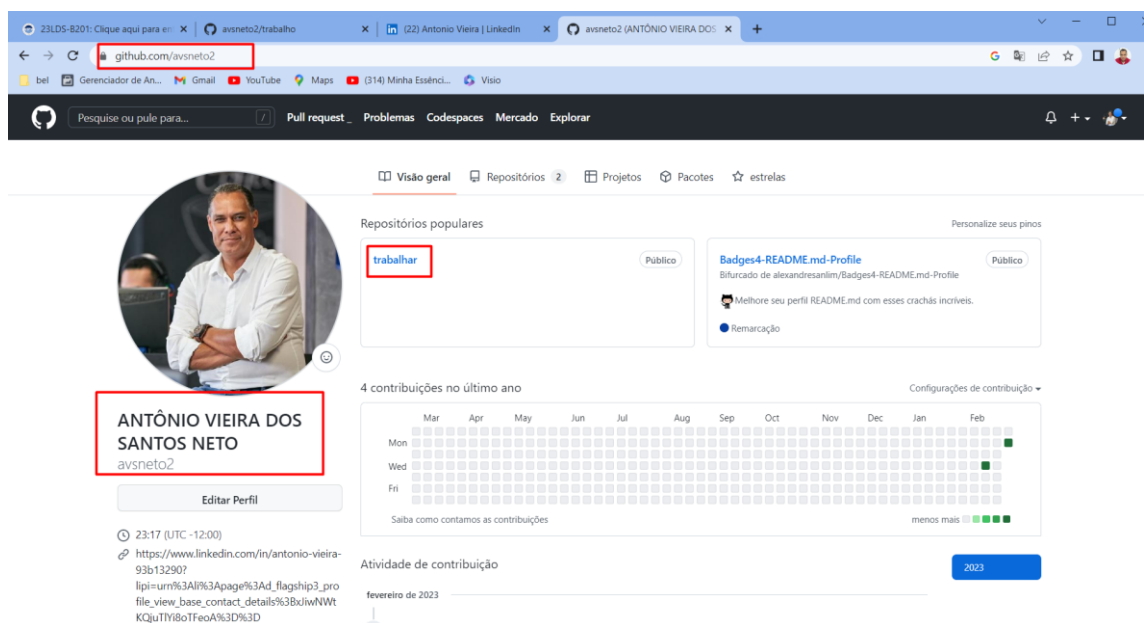
Instituto Infnet  
Pós Graduação - Data Science  
Estatística para Cientistas de Dados [23E1\_2]

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



4. Ambiente “GitHub”

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



# Instituto Infnet

## Pós Graduação - Data Science

### Estatística para Cientistas de Dados [23E1\_2]

**Branch : Projeto\_Estatistica\_1**

The screenshot shows a GitHub repository page for 'Projeto\_Estatistica\_1' by user 'avsneto2'. The repository is in the 'main' branch and has 3 commits. The file list includes: 'cópia de segurança' (projeto estatístico, há 26 minutos), '.gitignore' (atualizando o código com o gráfico da população de dados, há 3 semanas), 'Antonio Vieira dos Santos Neto\_Estat...' (projeto estatística 1, há 3 minutos), 'Antonio Viera dos Santos Neto\_Estat...' (projeto estatístico, há 26 minutos), 'Antonio Vieira dos Santos Neto\_Estat...' (projeto estatístico, há 26 minutos), 'Antonio-Vieira-dos-Santos-Neto\_Est...' (projeto estatística 1, há 3 minutos), 'BaseDPEvolucaoMensalCisp.csv' (Projeto do curso de Estatística, há 4 dias), 'aula5\_exercicio.Rproj' (atualizando o código com o gráfico da população de dados, há 3 semanas), and '~\$tonio Viera dos Santos Neto\_Estat...' (projeto estatístico, há 26 minutos). The right sidebar shows 'Sobre' (No description, site or topics provided), 'Lançamentos' (No release published), and 'Pacotes' (No package published).

Minhas disciplinas x avsneto2/trabalha no Projeto\_Estatistica\_1 Aula 4 - Tidymverse e Manipulação de Dados

github.com/avsneto2/work/tree/Projeto\_Estatistica\_1

bel Gerenciador de Arquivos Gmail YouTube Maps (314) Minha Essência Visio

<> Código Problemas Requisições pull Ações Projetos Wiki Segurança Percepções Configurações

Projeto\_Estatistica\_1 recebeu pushes recentes 3 minutos atrás Solicitação de comparação e pull

Projeto\_Estatistica\_1 3 filiais 0 tags ir para o arquivo Adicionar ficheiro Código

Este branch está 3 commits antes do main. Contribuir

avsneto2 projeto estatístico 1 8fadf34 há 3 minutos 4 confirmações

cópia de segurança	projeto estatístico	há 26 minutos
.gitignore	atualizando o código com o gráfico da população de dados	há 3 semanas
Antonio Vieira dos Santos Neto_Estat...	projeto estatística 1	há 3 minutos
Antonio Viera dos Santos Neto_Estat...	projeto estatístico	há 26 minutos
Antonio Vieira dos Santos Neto_Estat...	projeto estatístico	há 26 minutos
Antonio-Vieira-dos-Santos-Neto_Est...	projeto estatística 1	há 3 minutos
BaseDPEvolucaoMensalCisp.csv	Projeto do curso de Estatística	há 4 dias
aula5_exercicio.Rproj	atualizando o código com o gráfico da população de dados	há 3 semanas
~\$tonio Viera dos Santos Neto_Estat...	projeto estatístico	há 26 minutos

Sobre

Nenhuma descrição, site ou tópicos fornecidos.

0 estrelas

1 assistindo

0 garfos

Lançamentos

Nenhum lançamento publicado

Criar um novo lançamento

Pacotes

Nenhum pacote publicado

Publique seu primeiro pacote