

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]

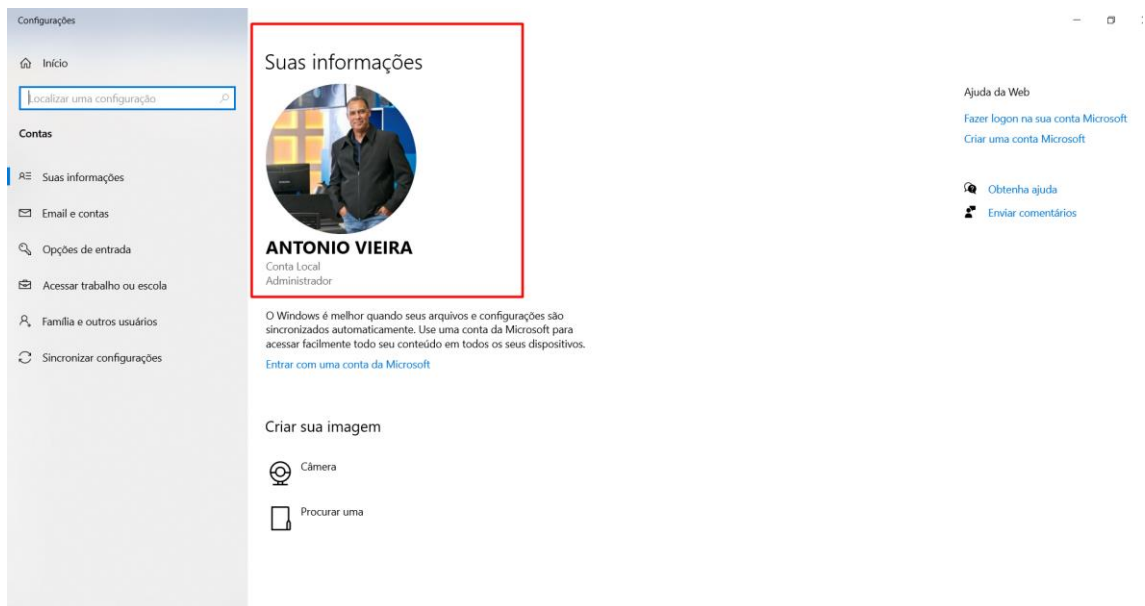
- b. o desvio padrão
 - c. os quantis: 25% e 75%
9. Utilizando o pacote summarytools (função descr), descreva estatisticamente a sua base de dados.
 10. Escolha uma variável e crie um histograma. Justifique o número de bins usados. A distribuição dessa variável se aproxima de uma "normal"? Justifique.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

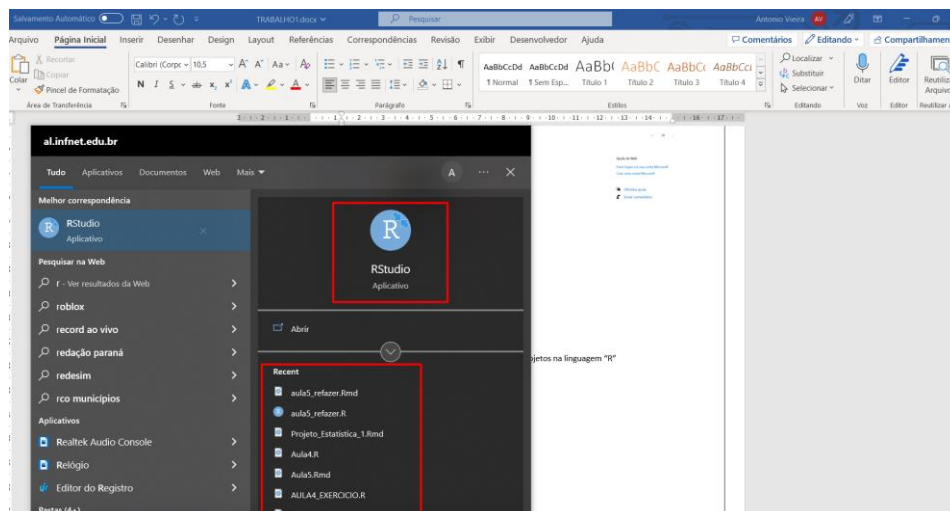
Evidências

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.

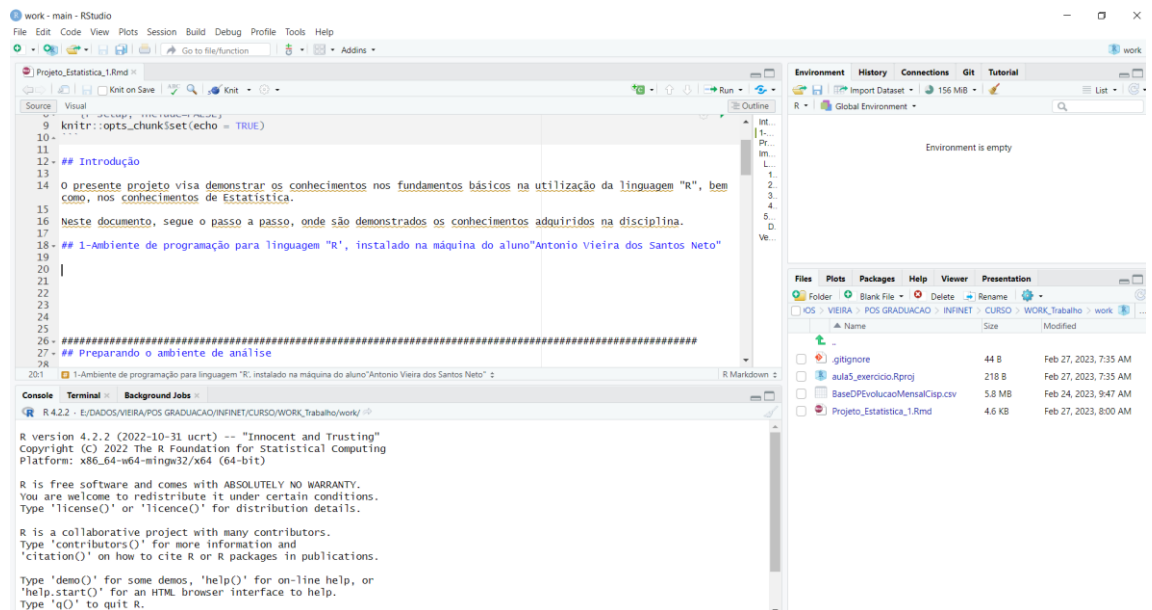


Instituto Infnet

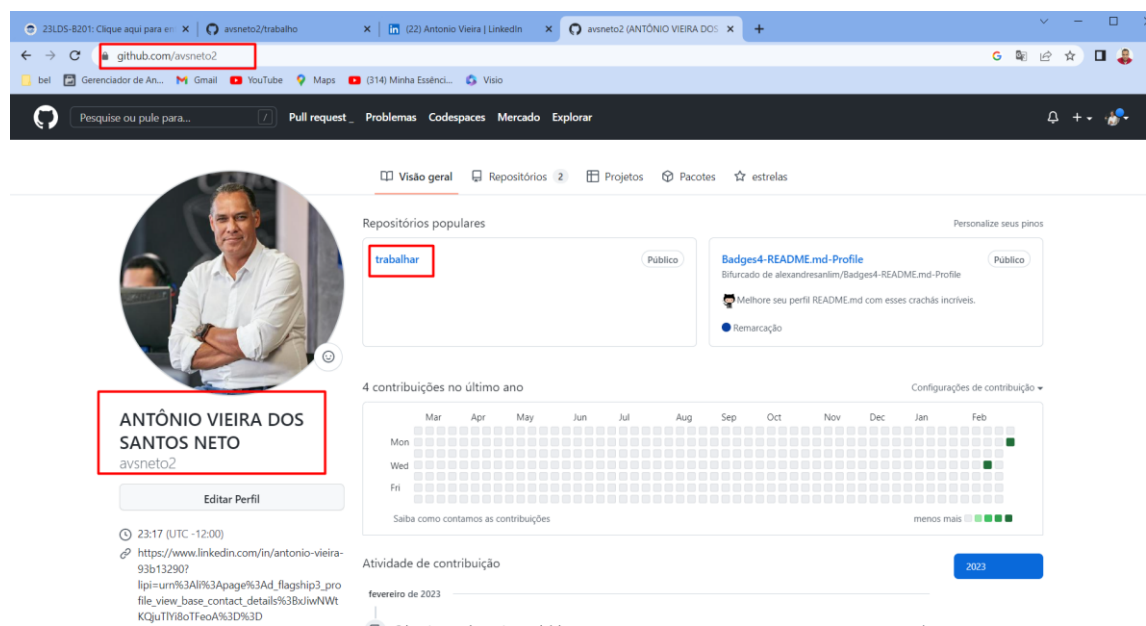
Pós Graduação - Data Science

Estatística para Cientistas de Dados [23E1_2]

3. Interface de desenvolvimento (Enviroment) - "RStudio".



4. Ambiente "GitHub"



Instituto Infnet
Pós Graduação - Data Science
Estatística para Cientistas de Dados [23E1_2]

The screenshot shows a GitHub repository page for 'avsneto2/trabalhar'. The repository is public and has 2 followers and 0 labels. The main branch is 'mestre'. The repository contains a list of files and folders, including 'BASES', '.RData', '.História', 'Aula3.R', 'IMG_GLPI', 'Trabalho_GLPI.Rmd', and 'aula2.R'. The repository is currently at commit 69c8b83, 3 days ago, with 4 confirmations. The repository description is 'Nenhuma descrição, site ou tópicos fornecidos.' The repository has 0 stars, 1 assistant, and 0 forks. The repository is currently at commit 69c8b83, 3 days ago, with 4 confirmations. The repository contains a list of files and folders, including 'BASES', '.RData', '.História', 'Aula3.R', 'IMG_GLPI', 'Trabalho_GLPI.Rmd', and 'aula2.R'.

File/Folder	Description	Last Commit
BASES	teste de alteração	3 dias atrás
.RData	Refazendo aula5	4 dias atrás
.História	Refazendo aula5	4 dias atrás
Aula3.R	Refazendo aula5	4 dias atrás
IMG_GLPI	teste de alteração	3 dias atrás
Trabalho_GLPI.Rmd	teste de alteração	3 dias atrás
aula2.R	Refazendo aula5	4 dias atrás