



Pós- Graduação

Data Discovery, OLAP e Visualização de Dados





Pós- Graduação

Visualização de dados com R, Python e Qlik Sense

Bloco 1

Marcelo Tavares de Lima



Objetivos

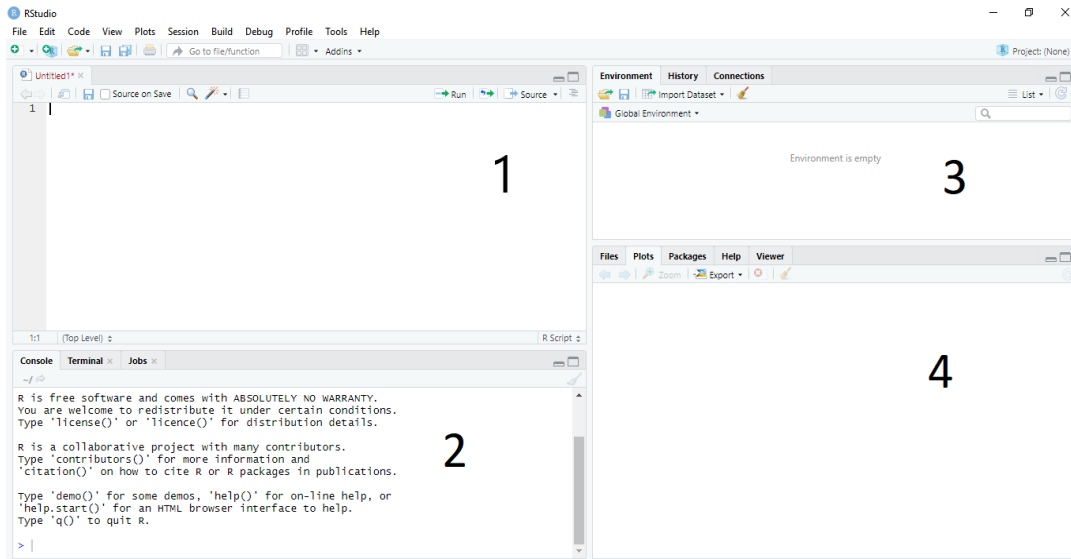
- Apresentar exemplos de visualização de dados com:
 - O R.
 - O Python.
 - O Qlik sense.

Introdução

- Inúmeras ferramentas.
- R e Python exigem conhecimento de programação.
- Qlik Sense não precisa de programação.

Visualização de dados com o R (Rstudio)

Figura 1 – Interface gráfica do RStudio



Fonte: elaborado pelo autor.

Visualização de dados com o R (Rstudio)

- Muitos pacotes para gerar visualizações.
- Um dos mais utilizados é o ggplot2.
- Gera gráficos complexos e cheios de recursos.

Visualização de dados com o R (Rstudio)

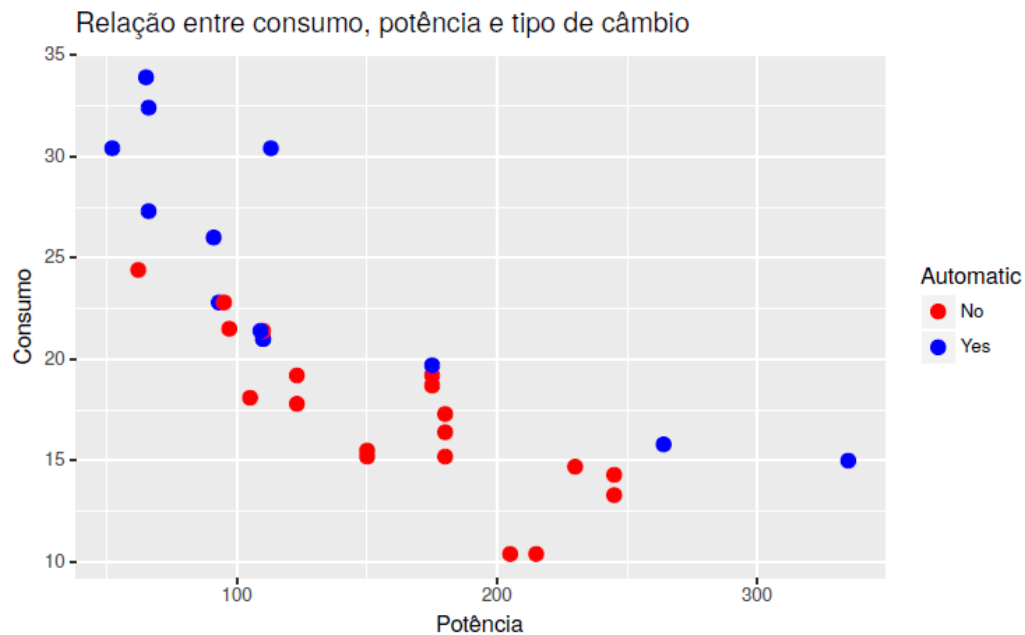
```
install.packages("ggplot2") # instala o  
pacote ggplot2  
library(ggplot2) # carrega o pacote ggplot2  
no Rstudio.
```

Visualização de dados com o R (Rstudio)

```
# Inicia o plot
# Adicionar pontos (geom_point) e
g <- ggplot(data = mtcars) +
  geom_point(mapping = aes(x = hp, y = mpg))
# Rótulos(títulos)
g <- g + labs(title = 'Relação entre
  consumo e potência', y = 'Consumo',
  x = 'Potência')
g
```


Visualização de dados com o R (Rstudio)

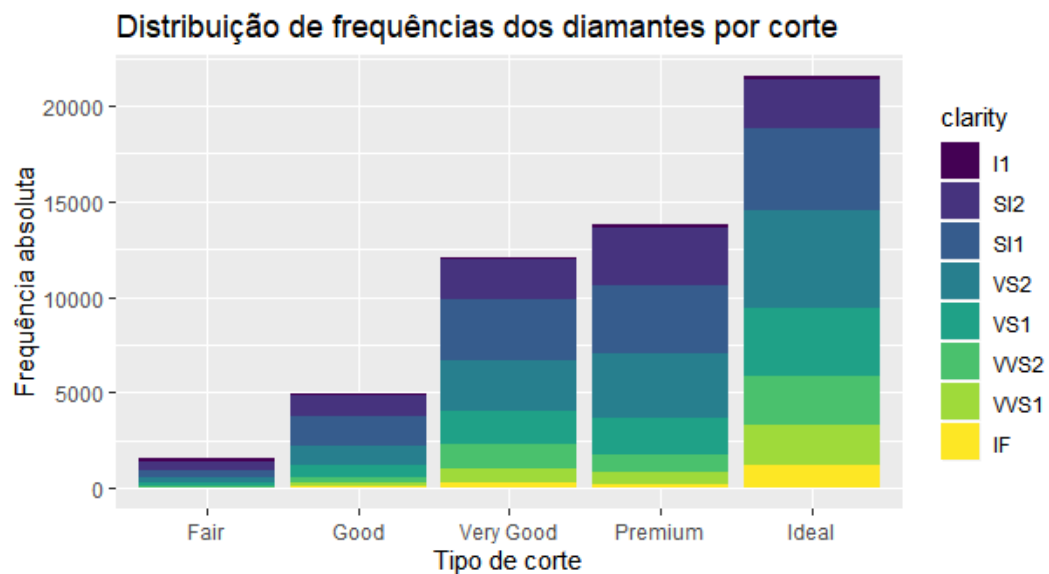
Figura 2 – Diagrama de dispersão



Fonte: elaborado pelo autor.

Visualização de dados com o R (Rstudio)

Figura 3 – Gráfico de barras verticais



Fonte: elaborado pelo autor.

Visualização de dados com o Python

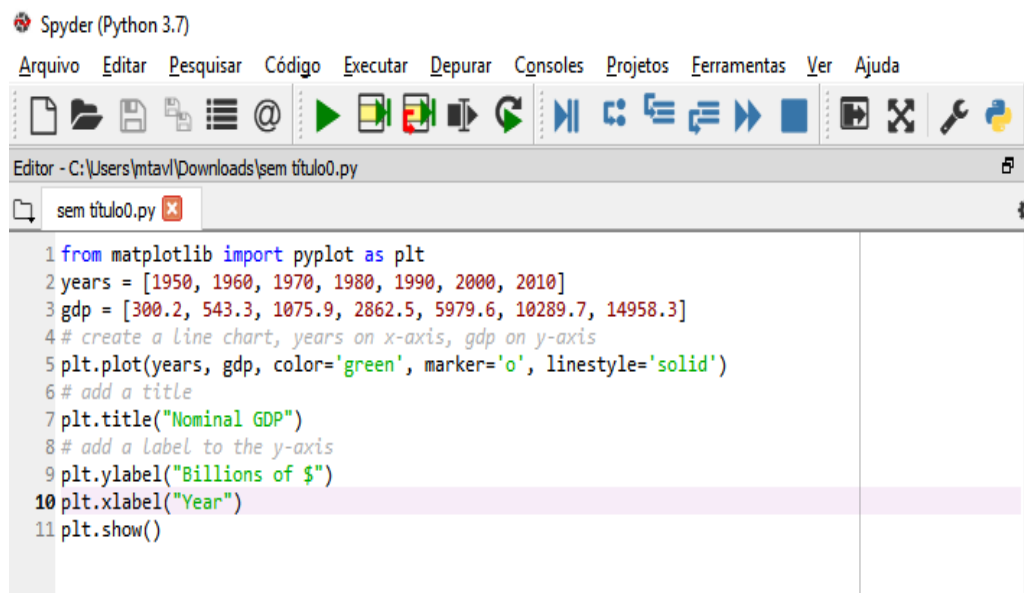
- Criado em 1989, pelo pesquisador Guido Van Rossum, do National Research Institute for Mathematics and Computer Science in Amsterdam – CWI, segundo Santos (2018).
- O nome, Tropa Monty Python, surgiu de um seriado de comédia que existia na época.
- O Python foi desenvolvido em código aberto (*open source*) e a sua primeira versão foi disponibilizada em 1991.

Visualização de dados com o Python

- Criada pra ser versátil.
- Também possui interfaces gráficas para facilitar seu uso.
- A mais usada é o Anaconda-Spyder.
- Trabalha com bibliotecas.

Visualização de dados com o Python

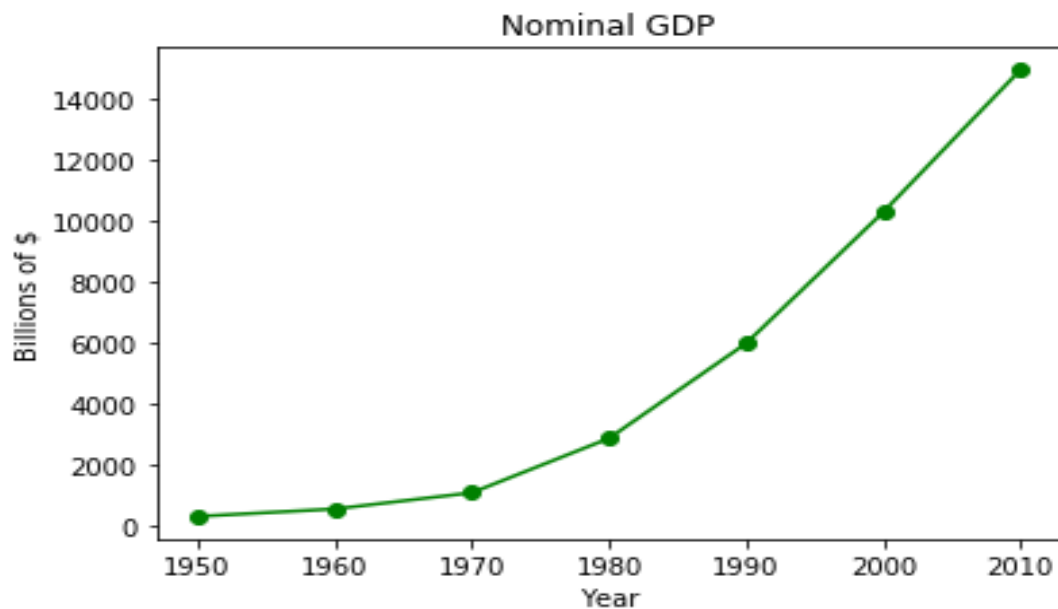
Figura 4 – Interface gráfica do Anaconda-Spyder (Python)



Fonte: elaborado pelo autor.

Visualização de dados com o Python

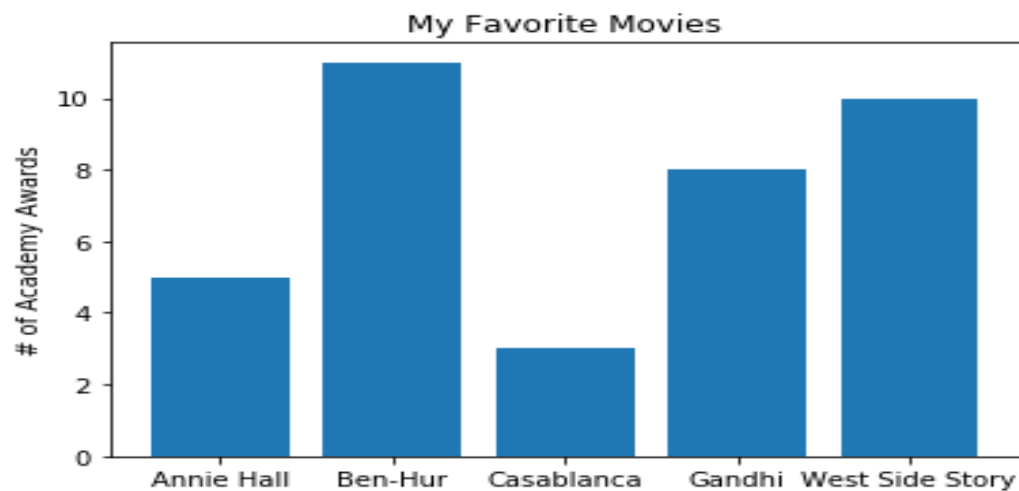
Figura 5 – Gráfico de linhas criado com linguagem Python



Fonte: elaborado pelo autor.

Visualização de dados com o Python

Figura 6 – Gráfico de barras criado com linguagem Python



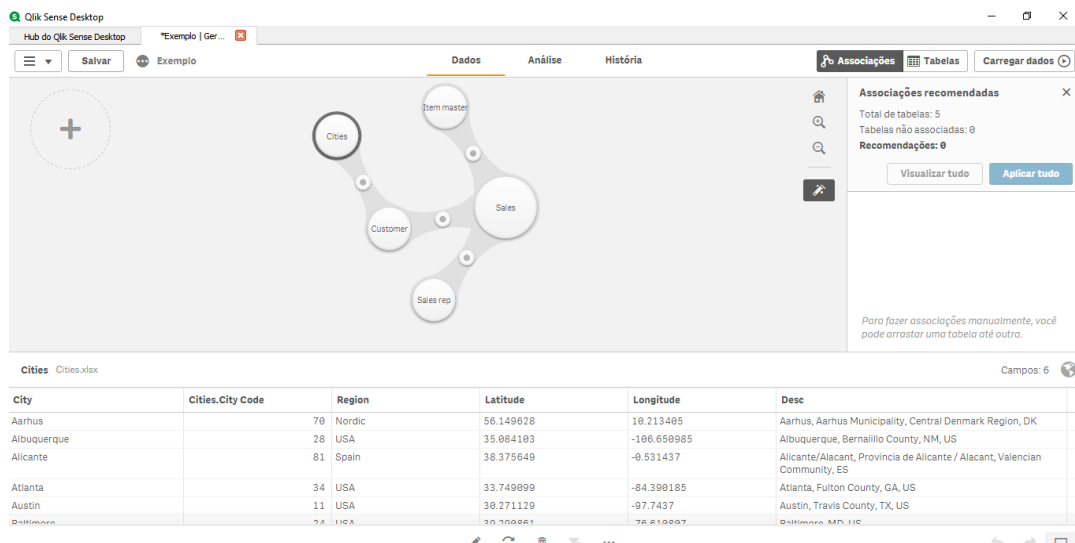
Fonte: elaborado pelo autor.

Visualização de dados com o Qlik Sense

- Pago.
- Código fechado.
- Não precisa de programação.
- É mais intuitivo.

Visualização de dados com o Qlik Sense

Figura 7 – Interface gráfica do Qlik Sense



Fonte: elaborado pelo autor.



Pós- Graduação

Visualização de dados com o R, Python e Qlik Sense

Bloco 2

Marcelo Tavares de Lima

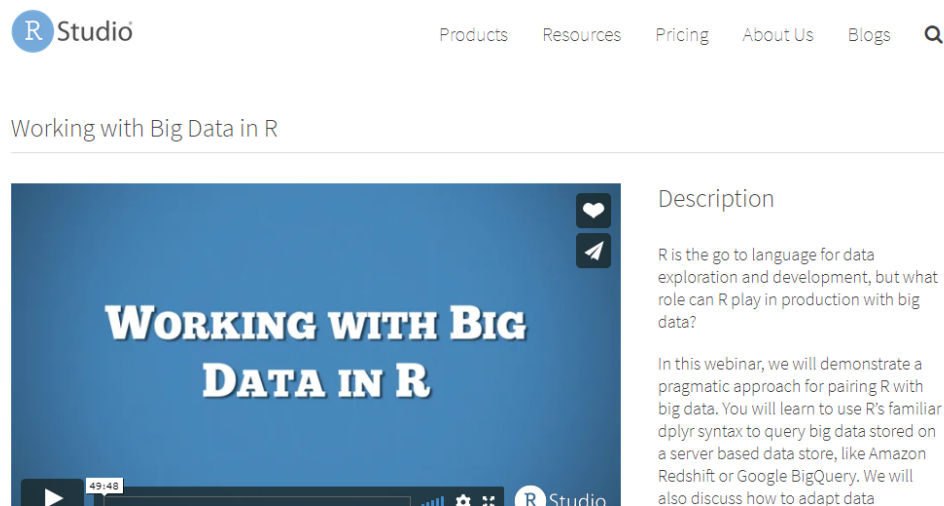


Uso de dados externos

- Todos permitem a importação de dados externos.
- Há inúmeras formas de realizar a importação.
- Na Internet tem muitos fóruns especializados.
- Inclusive de grandes bancos de dados.
- Podem trabalhar com dados em rede.

Uso de dados externos

Figura 8 – *Big Data* no R



The screenshot shows the R Studio website. At the top, there is a navigation bar with the R Studio logo and links for Products, Resources, Pricing, About Us, and Blogs. Below the navigation bar, the page title is "Working with Big Data in R". The main content area features a video player on the left and a description on the right. The video player has a blue background with the text "WORKING WITH BIG DATA IN R" in white. The video player controls show a play button, a progress bar at 49:48, and the R Studio logo. The description on the right is titled "Description" and contains two paragraphs of text.

R Studio

Products Resources Pricing About Us Blogs

Working with Big Data in R

WORKING WITH BIG DATA IN R

49:48

R Studio

Description

R is the go to language for data exploration and development, but what role can R play in production with big data?

In this webinar, we will demonstrate a pragmatic approach for pairing R with big data. You will learn to use R's familiar dplyr syntax to query big data stored on a server based data store, like Amazon Redshift or Google BigQuery. We will also discuss how to adapt data

Fonte: elaborado pelo autor.

Uso de dados externos

Figura 9 – *Big Data* com Python



Fonte: elaborado pelo autor.

Uso de dados externos

Figura 10 – *Big Data* com Qlik Sense



Fonte: elaborado pelo autor.



Pós- Graduação

Teoria em prática

Bloco 3

Marcelo Tavares de Lima



Visualização de dados com RStudio

- Considere que você deseja elaborar um gráfico de colunas (barras verticais).
- Esse gráfico parece simples, mas interessante, pois, revela sutilezas.
- Para exemplificar, vamos utilizar a linguagem R para elaborar um gráfico de coluna,s com o pacote ggplot2, extraído de Wickham e Grolemund (2017).
- Para isso, será utilizado o banco de dados *diamonds*, que faz parte do pacote ggplot2.
- Esse banco de dados contém informações de cerca de 54.000 diamantes, como tipo, preço, característica, cor, corte etc.

Visualização de dados com Rstudio

- Utilizados dados do pacote ggplot2.
- São informações sobre características de 54.000 diamantes.

Visualização de dados com Rstudio

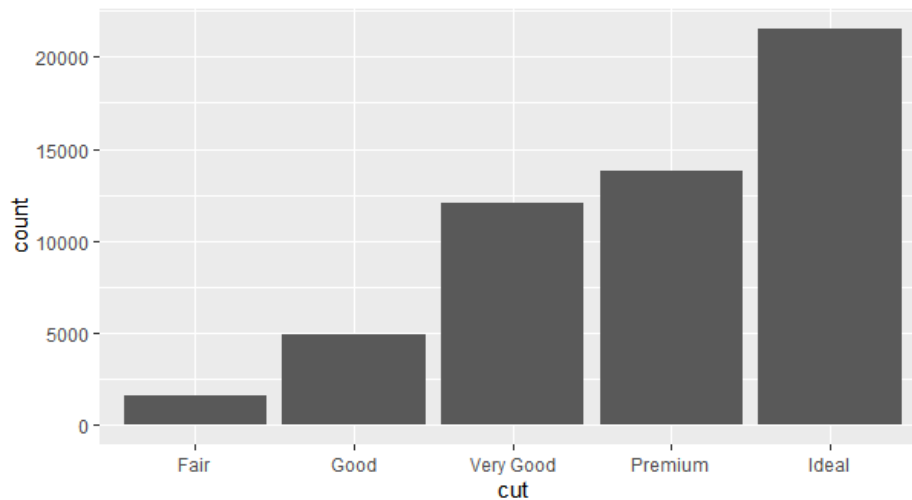
```
library(ggplot2)
```

```
ggplot(data=diamonds)+
```

```
  geom_bar(mapping=aes(x=cut))
```

Visualização de dados com Rstudio

Figura 11 – Gráfico de barras para os dados dos diamantes



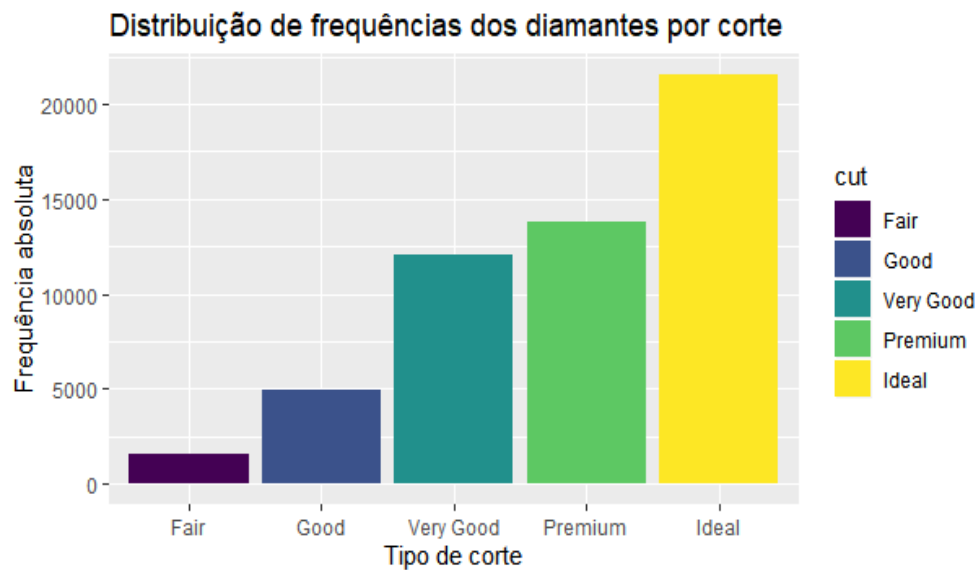
Fonte: elaborado pelo autor.

Visualização de dados com Rstudio

```
ggplot(data=diamonds)+  
  geom_bar(mapping=aes(x=cut,fill=cut))+  
  labs(title = 'Distribuição de  
frequências dos diamantes por corte',  
        y = 'Frequência absoluta',  
        x = 'Tipo de corte')
```

Visualização de dados com Rstudio

Figura 12 – Gráfico de barras para os dados dos diamantes



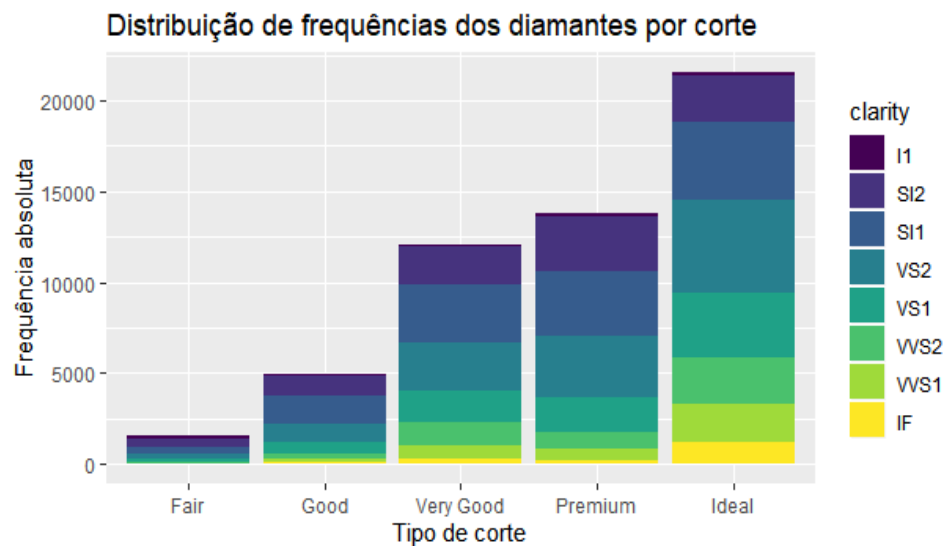
Fonte: elaborado pelo autor.

Visualização de dados com Rstudio

```
ggplot(data=diamonds)+  
  geom_bar(mapping=aes(x=cut,fill=clarity))+  
  labs(title = 'Distribuição de frequências  
dos diamantes por corte',  
        y = 'Frequência absoluta',  
        x = 'Tipo de corte')
```

Visualização de dados com Rstudio

Figura 13 – Gráfico de barras para os dados dos diamantes



Fonte: elaborado pelo autor.



Pós- Graduação

Dica do professor

Bloco 4

Marcelo Tavares de Lima



Fórum sobre ciência de dados

Figura 14 – Fórum on-line sobre ciência de dados

The screenshot shows the Stack Overflow website in Portuguese. The main heading is "Principais perguntas" (Top Questions). There are four questions listed:

Votos	Respostas	Visitas	Título	Tags	Resposta
0	1	12	Função em C que apaga determinada string	c	respondidas 5 minutos atrás Mateus Schroeder da Silva 140
0	0	7	Estou com problemas de repetição no python	python, python-3.x, while	modificada 12 minutos atrás Augusto Vasques 3.026
0	0	2	Unity App error - dalvik.system.BaseDexClassLoader.findClass (BaseDexClassLoader.java:134)	unity3d	perguntada 15 minutos atrás Joao Paulo Fink 1
0	2	190	Acrescentar paragrafo em vba	vba, excel-vba	modificada 31 minutos atrás Comunidade 1

Fonte: elaborado pelo autor.

Fórum sobre ciência de dados

O Stack Overflow é um site americano gratuito de perguntas e respostas sobre desenvolvimento de software de grande força no exterior.

Referências Bibliográficas

SANTOS, R.F.V.C. **Python**: guia prático do básico ao avançado. Série cientista de dados. 2018. E-book Kindle.

WICKHAM, H.; GROLEMUND, G. **R for data Science**: import, tidy, transform, visualize, and model data. Sebastopol: O'Reilly, 2017.



Pós- Graduação

Bons estudos!

