Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução ao

Estatísticas Descritivas

Tratamento de Dados

Visualização

Gráficos interativos

Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela e-mail: rodolpho.quintela@urca.br

URCA - Iguatu

7 de novembro de 2023

Introdução ao R e Tratamento de Dados

Rodolpho J.
D. Quintela,
e-mail:
rodolpho.qui
ntela@urca.br

Introdução ao

Estatísticas Descritivas

Tratamento de Dados

Visualização de Dados

Gráficos interativo

Introdução ao R

Objetivo

Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução ao R

Estatísticas Descritivas

Tratamento de Dados

Visualização de Dados

- I Introdução ao R: Apresentar o ambiente de programação R, incluindo a instalação, configuração e navegação básica.
- Estatísticas Descritivas: Introduzir conceitos de estatísticas descritivas e mostrar como calcular medidas de tendência central, dispersão e apresentá-las visualmente.
- Tratamento de Dados: Importar, limpar e manipular conjuntos de dados usando pacotes e funções relevantes do R, como dplyr.
- **4 Visualização de Dados**: Explorar as capacidades de visualização do R com pacotes como ggplot2, e mostrar como criar gráficos informativos e esteticamente agradáveis.

Apresentação do R

Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução ao R

Estatísticas Descritivas

Tratamento de Dados

Visualização de Dados

- R é uma linguagem de programação e ambiente estatístico.
- Criado por Ross Ihaka e Robert Gentleman em 1993.
- É uma linguagem interpretada e gratuita.
- Amplamente utilizado para análise de dados, estatísticas e visualização gráfica.
- R é mantido por uma comunidade ativa de desenvolvedores.

Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

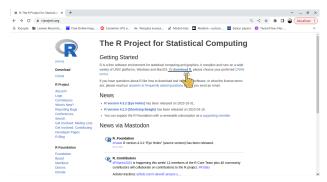
Introdução ao R

Descritivas Descritivas

Tratamento de Dados

Visualização de Dados

- Visite o site oficial do R: https://www.r-project.org/
- Clicar em 'download R' ou em 'CRAN'.



Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução ao R

Estatísticas Descritivas

Tratamento de Dados

Visualização de Dados

- Você será direcionado para a página https://cran.r-project.org/mirrors.html.
- Escolha o espelho de sua preferência



Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução ao R

Estatísticas Descritivas

Tratamento de Dados

Visualização de Dados

Gráficos interativos

■ Clique no sistema operacional do seu computador



Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução ao R

Estatísticas Descritivas

Tratamento de Dados

Visualização

- Clique em 'base'
- Clique em 'Download R-··· for Windows'
- Executar o arquivo após o download

Interface do R

Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução ao R

Descritivas

Tratamento de Dados

Visualização de Dados

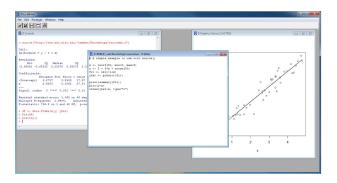


Figura 1: Interface do R.

RStudio: Uma IDE popular para R

Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução ao R

Estatísticas Descritivas

Tratamento de Dados

Visualização de Dados

- O RStudio é uma IDE (Integrated Development Environment) gratuita e de código aberto para R.
- Disponível para Windows, macOS e Linux.
- Oferece uma interface amigável e recursos avançados.
- Facilita a escrita, depuração e visualização de resultados.
- Baixe o RStudio em: https://www.rstudio.com/

Interface do Rstudio

- 0 X Introdução ao File Edit View Project Workspace Plots Tools Help Re 9 diamondPricing R* × 9 formatPlot.R × diamonds × Tratamento Source on Save | 9 2 --+Run S+ Source * de Dados 1 library(ggplot2) diamonds 53940 obs. of 10 variables View(diamonds) Values Rodolpho J. summary(dfamonds) avesize 0.7979 D. Quintela. 6 summary(diamondsSprice) 3-Histórico e Importação rodolpho.aui - Editor de código 10 p de dados ntela@urca.br 11 main="Diamond Pricing"] 14 Introdução ao Files Plots Packages Help Zoom Export* Q dear All [] (Top Level) 0 R Script Diamond Pricing : 0.000 1st Qu.: 4,710 1st Qu.: 4,720 Median : 5,710 - Gráficos e Arquivos 4.040 Console do R 1.800 V92 Price > aveSize <- round(mean(diamonds\$carat), 4) > clarity <- levels(diamondsSclarity) VV92 <- oplot(carat, price, W81 data-diamonds, color-clarity, xlab="Carat", ylab="Price", . 10 main="Diamond Pricing") > format.plot(plot=p, size=23)

Figura 2: Interface do Rstudio.

Primeiros passos

Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Inicie o RStudio e crie um novo script. Comentários no R começam com # e são ignorados pelo interpretador.

Isto é um comentário

A seguir, iremos ver:

- 1 Operações básicas (soma, subtração, multiplicação, divisão, potenciação)
- 2 Variáveis
- Algumas funções básicas

Introdução ao R

Estatísticas Descritivas

de Dados

de Dados

Primeiros passos

Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução ao R

Estatísticas Descritivas

Tratament de Dados

Visualização

Gráficos interativo

Operações básicas

O R pode ser usado como uma calculadora e Suporta as operações matemáticas básicas (soma, subtração, multiplicação, divisão, potenciação): +, -, *, /, $^{\circ}$ (potenciação)

Primeiros passos: Operações básicas

```
Introdução ao
R e
Tratamento
de Dados
```

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução ao R

Estatísticas Descritivas

Tratament de Dados

Visualização de Dados

Gráficos interativos

```
# Soma e subtração
10 - 3
```

[1] 7

10 + 3

[1] 13

Primeiros passos: Operações básicas

```
Introdução ao
   Re
 Tratamento
 de Dados
             # Multiplicação e divisão
Rodolpho J.
D. Quintela.
             8 * 2
ntela@urca.br
             ## [1] 16
Introdução ao
             8 / 2
             ## [1] 4
             # Exponencial
             4^(1/2)
             ## [1] 2
```

Primeiros passos: Operações básicas - Exercícios

Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução ao R

Estatística Descritivas

Tratamento de Dados

Visualização de Dados

Gráficos interativo Imagine que você é um professor de Matemática do ensino básico e um aluno curioso te pergunta:

"Professor, quanto é 13 + 15 + 16 + 6?"

Você não quer fazer essa conta nos dedos, né? Mostre sua sabedoria em tecnologia e recorre ao R para dar uma resposta rápida.

Como deve ser o código para chegar ao resultado?

Primeiros passos: Operações básicas - Exercícios

Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução ao R

Estatística Descritivas

Tratamento de Dados

Visualização de Dados

Gráficos interativo Agora você e mais 3 amigos resolveram sair para comer uma pizza. A conta deu 67 reais e vocês decidiram dividir a conta igualmente. Utilizando o R, escreva um código para saber quanto cada um deve pagar.

Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução ao R

Estatísticas Descritivas

Tratamento de Dados

Visualização de Dados

Gráficos interativo

Variáveis

Variáveis são um dos grandes pilares da programação em R. Ao invés de trabalhar diretamente com valores, atribuímos valores às variáveis e trabalhamos com elas, o que torna a programação muito mais flexível.

Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução ao R

Estatísticas Descritivas

Tratamento de Dados

Visualização de Dados

Gráficos interativos Para atruibuir valores a uma variável, utiliza-se '=' ou <- Vejamos alguns exemplos de como isto funciona:

$$y = 2$$

Com isso, realizar operações com as variáveis. Por exemplo:

Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução ao R

Estatísticas Descritivas

Tratamento de Dados

Visualização

Gráficos interativos

Nomes de variáveis

Um nome de variável deve ser uma sequência de letras e números, e devem sempre começar com uma letra. Não se pode começar o nome com números ou caracteres especiais.

```
Introdução ao
   Re
             # Nomes válidos:
 Tratamento
 de Dados
             num2 = 2
Rodolpho J.
             texto = "Olá"
D. Quintela.
             teste = 10
ntela@urca.br
             # print dos números
Introdução ao
             num2
             ## [1] 2
             texto
             ## [1] "Olá"
             teste
             ## [1] 10
```

Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução ao R

Estatísticas Descritivas

Tratamento de Dados

Visualização de Dados

Gráficos interativos

Funções

Em R, uma função é um bloco de código que realiza uma tarefa específica ou executa um conjunto de operações quando chamado. Funções são uma parte fundamental da linguagem de programação R e são usadas para realizar uma ampla variedade de tarefas, desde cálculos simples até tarefas mais complexas de análise de dados e manipulação.

Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução ao R

Estatísticas Descritivas

Tratamento de Dados

Visualização de Dados

Gráficos interativo

Definição de Função em R

Definição de Função: Antes de usar uma função em R, você deve defini-la. A definição de uma função envolve especificar o nome da função, os argumentos que ela aceita e o conjunto de operações que ela realizará quando chamada.

```
Introdução ao
R e
Tratamento
de Dados
```

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução ao R

Estatísticas Descritivas

Tratamento de Dados

Visualização de Dados

Gráficos interativos

```
Sintaxe de uma função:
```

```
nome_funcao <- function(arg1, arg2,..., argn) {
  comandos da funçao
}</pre>
```

Aqui está um exemplo simples de definição de uma função em R:

```
# Definindo uma função chamada "soma"
soma <- function(a, b) {
  resultado <- a + b
  return(resultado)
}</pre>
```

Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução ao R

Estatísticas Descritivas

Tratamento de Dados

Visualização de Dados

Gráficos interativo

Chamada de Funções

Chamada de Função: Uma vez que uma função esteja definida, você pode chamá-la pelo nome, fornecendo os argumentos necessários. A chamada da função faz com que o R execute as operações especificadas na função

```
Introdução ao
R e
Tratamento
de Dados
```

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução ao R

Estatísticas Descritivas

Tratamento de Dados

Visualização de Dados

Gráficos interativos

```
Exemplo de chamada de uma função:
```

```
# Chamando a função "soma" com os argumentos 5 e 3
resultado_da_soma <- soma(5, 3)
# Isso imprimirá "8" no console
print(resultado da soma)</pre>
```

[1] 8

Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução ao R

Estatística: Descritivas

Tratamento de Dados

Visualização de Dados

Gráficos interativos O R possui muitas funções integradas que executam uma variedade de tarefas comuns em análise de dados, estatísticas e manipulação de objetos. Aqui estão alguns exemplos de funções básicas do R:

Funções Matemáticas:

- sqrt(x): calcula a raiz quadrada de x;
- abs(x): retorna o valor absoluto de x;
- exp(x): retorna o exponencial de x;
- log(x, base = b): Retorna o Logaritmo de x na base b;
- **sum** (x_1, x_2, \dots, x_n) : retorna a soma de $x_1, x_2, \dots x_n$.

```
Introdução ao
   Re
             # Exemplos de funções matemáticas
 Tratamento
             x < -4
 de Dados
Rodolpho J.
             sqrt(x)
D. Quintela.
             ## [1] 2
ntela@urca.br
             exp(x)
Introdução ao
             ## [1] 54.59815
            log(x, 2)
             ## [1] 2
             sum(1,2,3,4)
```

Primeiros passos: Funções - Exercícios

Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução ao R

Estatística: Descritivas

Tratamento de Dados

Visualização de Dados

Gráficos interativos Agora você é um professor do ensino médio, e as perguntas estão mais difíceis. Os alunos querem que você caclule o Logaritmo de 5 na base 8. Como isso pode ser feito no R?

Primeiros passos: Funções - Exercícios

Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução ao R

Estatísticas Descritivas

Tratamento de Dados

Visualização de Dados

Gráficos interativos Utilizando o R, atribua à variável x o valor 45 e calcule:

- Raiz quadrada de x;
- Raiz cúbica de x;
- Logaritmo de x na base 3;
- Exponencial de x
- \mathbf{x}^2

Primeiros passos: Entrada de dados

Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução ao R

Estatísticas Descritivas

Tratament de Dados

Visualização

Gráficos interativos

Entrada de dados no R

A entrada de dados no R pode ser realizada de várias maneiras, dependendo do tipo de dados que você deseja inserir e do contexto em que está trabalhando.

Primeiros passos: Entrada de dados

Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução ao R

Estatísticas Descritivas

Tratamento de Dados

Visualização de Dados

Gráficos interativos Aqui estão algumas das principais maneiras de inserir dados no R:

- 1 Vetores
- Você pode criar vetores diretamente no R, atribuindo valores a variáveis. Para entrar com os vetores utilize o comando c, como descrito a seguir:

vetor_numerico \leftarrow c(1, 2, 3, 4, 5)

Primeiros passos: Entrada de dados

Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução ao R

Estatísticas Descritivas

Tratamento de Dados

Visualização de Dados

Gráficos interativos

2. Leitura de Arquivos de Dados:

Uma maneira comum de inserir dados no R é lendo arquivos de dados, como CSV, Excel, TXT, etc. Você pode usar funções como read.csv(), read.table(), entre outras, para importar dados de arquivos. Por exemplo:

```
dados <- read.csv("dados.csv")</pre>
```

Voltaremos a esse exemplo mais tarde com mais detalhes.

Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução a

Estatísticas Descritivas

Tratamento de Dados

Visualização de Dados

Gráficos interativo

Estatísticas Descritivas

Estatísticas Descritivas

Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução a

Estatísticas Descritivas

Tratament de Dados

Visualização de Dados

Gráficos interativo

Medidas resumo

Após coletar os dados e armazená-los, precisamos interpretálos por meio de medidas que resumem os dados. Estas podem caracterizar a localização (posição), a dispersão ou a forma dos dados.

Estatísticas Descritivas: Medidas resumo

Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução a

Estatísticas Descritivas

Tratament de Dados

Visualização de Dados

Gráficos interativo **Exemplo**: O *mínimo* ou o *máximo* de um conjunto de dados, a *moda* (valor mais frequente), *média*, *mediana* são valores que indicam posições do conjunto de dados.

Já as medidas de dispersão são aquelas que, como o nome sugere, indicam a variabilidade dos dados.

Exemplo: Variância, Desvio padrão e etc.

Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução ao

Estatísticas Descritivas

Tratamento de Dados

Visualização de Dados

Gráficos interativos

Média

A média aritmética é a soma de todos os elementos dividida pelo número de componentes da soma.

Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução a

Estatísticas Descritivas

Tratamento de Dados

Visualização de Dados

Gráficos interativos No R, a função *mean()* calcula a média aritmética dos valores em um vetor.

Exmplo: Suponha que você tenha um vetor chamado idade e deseje calcular a média dessas idades:

```
# Vetor com as idades
idade = c(23,25, 30, 19, 60, 18, 60)
# Utilzando a função mean para cacular a média
mean(idade)
```

[1] 33.57143

Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução ao

Estatísticas Descritivas

Tratament de Dados

Visualização

Gráficos

Mediana

A mediana é o valor que divide o conjunto de dados ordenados ao meio, separando exatamente 50% dos dados abaixo e 50% dos dados acima desse valor.

Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução ao

Estatísticas Descritivas

Tratamento de Dados

Visualização de Dados

Gráficos interativo No R, a função *median()* calcula a mediana de um conjunto de dados.

Exmplo: Considerando o exemplo anterior, temos que:

Utilzando a função median para cacular a mediana median(idade)

[1] 25

Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução ao R

Estatísticas Descritivas

Tratamento de Dados

Visualização de Dados

Gráficos interativos

Variância

A variância é uma medida de dispersão que indica o quão os valores em um conjunto de dados estão dispersos ao redor da média e é definida como:

$$\sigma^2 = \sum_{i=1}^n \frac{(x_i - \overline{x})^2}{n},$$

sendo n o número de observações, x_i os valores individuais no conjunto de dados e \bar{x} a média dos valores.

Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução ao R

Estatísticas Descritivas

Tratamento de Dados

Visualização de Dados

Gráficos interativos

Desvio Padrão

O desvio padrão é outra medida de dispersão, definida como a raiz quadrada da variância. Ele é expresso na mesma unidade de medida dos dados originais, tornando-o mais intuitivo sua interpretação. A fórmula do desvio padrão é:

Desvio padrão =
$$\sqrt{\sigma^2}$$

Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução ao R

Estatísticas Descritivas

Tratament de Dados

Visualização de Dados

Gráficos interativos No R, utilizadmos as funções var() e sd() para cacular, respetivamente, a variância e o desvio padrão de um conjunto de dados.

Exmplo: Considerando o exemplo anterior, temos que:

```
# Cálculo da variância da "idade"
var(idade)
```

```
## [1] 341.619
```

```
# Cálculo do desvio padrão da "idade"
sd(idade)
```

```
## [1] 18.48294
```

Estatísticas Descritivas: Medidas resumo - Exercícios

Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução ao

Estatísticas Descritivas

Tratamento de Dados

Visualização de Dados

Gráficos interativo Considere as notas de 10 alunos da disciplina de estatística dadas a seguir:

```
# Notas de 10 alunos da disciplina de estatística notas<-c(6.0,7.3,9.8,6.7,10,8.9,6.7,2.4,5.4,5.5)
```

Utilizando o R, calcule:

- Média;
- Mediana;
- Variâcia;
- Desvio padrão.

Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução ao

Estatísticas Descritivas

Tratamento de Dados

Visualização de Dados

Gráficos interativos

Tratamento de Dados

Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução a

Estatísticas Descritivas

Tratamento de Dados

Visualização de Dados

Gráficos interativos

Data frame

Um data frame é uma matriz bidimensional em que cada coluna pode conter dados de tipos diferentes, como números, caracteres, fatores, datas, etc. Os data frames são muito semelhantes às tabelas em bancos de dados ou planilhas de Excel e são amplamente utilizados para representar dados no contexto de análise de dados e estatísticas.

```
Introdução ao
R e
Tratamento
de Dados
```

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução ao

Descritivas

Tratamento de Dados

Visualização de Dados

Gráficos interativos Aqui está um exemplo simples de criação de um data frame em R:

```
# Criar um data frame
data_frame <- data.frame(
  Nome = c("João", "Maria", "Carlos"),
  Idade = c(30, 25, 35),
  Sexo = c("M", "F", "M")
)
data_frame</pre>
```

```
## Nome Idade Sexo
## 1 João 30 M
## 2 Maria 25 F
## 3 Carlos 35 M
```

Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução ao R

Estatísticas Descritivas

Tratamento de Dados

Gráficos interativos Para acessar colunas específicas de um data frame, utilize o operador '\$'. Por exemplo:

Acessar a coluna "Idade" no data frame "data_frame" data_frame\$Idade

[1] 30 25 35

Ou então use colchetes [] com o nome da coluna. Por exemplo:

Acessar a coluna "Idade" no data frame "data_frame" data_frame["Idade"]

Idade ## 1 30 ## 2 25

35

Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução ao R

Estatísticas Descritivas

Tratamento de Dados

Visualização de Dados

Gráficos interativos Com isso, podemos realizar análises estatísticos para cada variável em um data frame. Por exemplo, podemos calcular a média das idades para o data frame 'data_frame'

```
# Média da da variável "Idade" no data frame
# "data_frame"
mean(data_frame$Idade)
```

[1] 30

Tratamento de Dados: Data frame - Exercícios

Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução ac

Estatística: Descritivas

Tratamento de Dados

Visualização de Dados

Gráficos interativos Utilizando as notas dos alunos da disciplina de estatística consideradas no último exercício, crie nomes fictícios, atribua um sexo e monte um data frame com essas informações. Dica:

```
# Notas de 10 alunos da disciplina de estatítica notas<-c(6.0,7.3,9.8,6.7,10,8.9,6.7,2.4,5.4,5.5) nomes<- c("João", "Maria", "José", . . .) sexo<-c("M", "F", "M", . . .) dados_notas_df<-data.frame(notas,nomes,sexo)
```

Tratamento de Dados: Data frame - Exercícios

Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução ao

Estatísticas Descritivas

Tratamento de Dados

Visualizaçã

Gráficos interativo Considerando o data frame 'dados_notas_df', por meio do operador '\$', encontre a média das notas e o seu desvio padrão.

Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução ao

Estatística Descritivas

Tratamento de Dados

Visualização de Dados

Gráficos interativos

Leitura de dados

O início de qualquer projeto de análise de dados envolve a aquisição dos dados, que frequentemente são disponibilizados em arquivos locais com formatos como .csv, .xlsx, .txt, entre outros.

Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução ao

Estatística: Descritivas

Tratamento de Dados

Visualização de Dados

Gráficos interativos Um primeiro comando a ser executado e registrado é a configuração da pasta onde estão salvos os arquivos como pasta de trabalho. Para isso, utilize.

setwd()

Por exemplo:

setwd("/home/jordan/Área de Trabalho/Minicurso")

Neste exemplo, é especificado o local onde os aquivos estão salvos.

Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução a R

Estatísticas Descritivas

Tratamento de Dados

Visualização de Dados

Gráficos interativos

Podemos abreviar a pasta pessoal por "~". Então escrevemos:

setwd("~/Área de Trabalho/Minicurso")

Observe que a barra invertida "/" é utilizada para delimitar pastas e arquivos no R. Com ela é possível usar a tecla **Tab** para completar nomes de arquivos.

No Rstudio, ainda podemos utilizar o atalho: $\mathsf{Ctrl} + \mathsf{Shift} + \mathsf{H}$

Uma vez que o diretório está definido, podemos começar a trabalhar com os nossos dados.

Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução a R

Estatística Descritivas

Tratamento de Dados

Visualizaçã de Dados

Gráficos interativo:

Aqruivos .csv

Arquivos do formato CSV (Comma Separated Values) são documentos de texto simples, que podem ser abertos em editores de texto comuns, pois não incluem nenhum tipo de formatação especial, como negrito, itálico, cores ou espaçamento entre linhas.

Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução a

Descritivas

Tratamento de Dados

Visualização de Dados

Gráficos interativos

A fim de importar um arquivo no formato .csv, é comum empregarmos a função *read.table()*. Entretanto, para usar essa função de maneira adequada, é essencial ter ciência de algumas informações cruciais, tais como:

- O caractere utilizado para separar os valores.;
- Se os valores textuais são delimitados por aspas simples, aspas duplas ou não possuem nenhum tipo de delimitação especial;
- 3 Se os nomes das variáveis estão na primeira linha do arquivo.

Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução ao R

Descritivas

Tratamento de Dados

Visualização de Dados

Gráficos interativos Aqui está um exemplo de como ler um arquivo e .csv no R.

Neste contexto, estamos fornecendo à função read.table() as seguintes informações: a primeira linha do arquivo contém os nomes das variáveis (ou seja, há um cabeçalho presente no arquivo, ou header), os valores distintos estão separados por vírgulas, uma vírgula é usada como separador decimal e os valores textuais são delimitados por aspas duplas (quotation marks).

Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução a

Estatísticas Descritivas

Tratamento de Dados

Visualizaçã de Dados

Gráficos interativos Utilize o comando a seguir para vizualizar a tabela com os dados:

View(dados)

Tratamento de Dados: Pacotes (packages) em R

Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução ao R

Estatísticas Descritivas

Tratamento de Dados

Visualização de Dados

Gráficos

Pacotes

Em R, pacotes são extensões de software que contêm funções, dados e documentação adicionais que podem ser usados para estender as funcionalidades da linguagem base do R.

Tratamento de Dados: Pacotes (packages) em R

Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução a R

Estatística: Descritivas

Tratamento de Dados

Visualização de Dados

Gráficos interativos Aqui estão algumas informações importantes sobre pacotes em R:

Instalação de Pacotes: Para usar um pacote em R, você precisa primeiro instalá-lo. Use a função install.packages(). Por exemplo:

install.packages("ggplot2")

Carregamento de Pacotes: Após a instalação, você deve carregar um pacote em sua sessão R usando a função library() ou require(). Isso torna as funções e dados do pacote disponíveis para uso. Por exemplo:

library(ggplot2)

Tratamento de Dados: Pacotes (packages) em R

Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução a R

Estatística Descritivas

Tratamento de Dados

Visualização de Dados

Gráficos interativo **Documentação:** A maioria dos pacotes inclui documentação que descreve como usar as funções e recursos do pacote. Você pode acessar a documentação de um pacote usando a função *help()* ou ? seguida pelo nome do pacote ou função. Por exemplo:

help(ggplot2)

Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução a

Estatística: Descritivas

Tratamento de Dados

Visualização de Dados

Gráficos interativo

dplyr

O pacote dplyr é bastante utilizado para uma ampla gama de operações de manipulação de dados, abrangendo desde agregação, resumo, filtragem, ordenação até a criação de variáveis, junção de dados e muito mais.

Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução ao

Estatística: Descritivas

Tratamento de Dados

Visualização de Dados

Gráficos interativos Instalando o pacote dplyr

install.packages("dplyr")

Carregando o pacote dplyr

library("dplyr")

Vamos usar o dataset 'Dados_empresaMB', com informações sobre os empregados da seção de orçarmento da Companhia MB para demonstrar o uso das principais funções do dplyr.

Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução a R

Estatísticas Descritivas

Tratamento de Dados

Visualização de Dados

Gráficos interativos

select()

Select é uma função muito simples: serve para selecionar ou remover colunas.

Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

```
Introdução ao
R
```

Descritivas

Tratamento de Dados

Visualização de Dados

Gráficos interativos

```
Por exemplo, para selecionar as colunas estado_civil e salário, utilizamos o seguinte comando.
```

```
# Tarefa: selecionar colunas estado_civil e salario
dados %>% select(estado_civil,salario) %>% head()
```

```
##
    estado civil salario
## 1
                    4.00
         solteiro
                    4.56
## 2
          casado
          casado 5.25
## 3
## 4
         solteiro 5.73
## 5
         solteiro
                    6.26
                    6.66
## 6
          casado
```

O comando *head()* retornar as primeiras linhas de um data frame.

Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução ao

Estatísticas Descritivas

Tratamento de Dados

Visualização de Dados

Gráficos interativos

```
# tarefa: remover as colunas estado_civil e salario
dados %>% select(-estado_civil,-salario)
```

Podemos também salvar essa tabela em um outro objeto. Por exmplo:

```
# tarefa: remover as colunas estado_civil e salario
df<-dados %>% select(-estado_civil,-salario)
df
```

```
Introdução ao
R e
Tratamento
de Dados
```

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução a

Estatísticas Descritivas

Tratamento de Dados

Visualização de Dados

Gráficos interativos

Outras possibilidades:

```
# tarefa: selectionar todas as colunas numéricas dados %>% select_if(is.numeric) %>% head()
```

Podemos também salvar essa tabela em um outro objeto. Por exmplo:

```
# tarefa: selecionar apenas as
# colunas cujo nome começa com idade
dados %>% select(starts_with("idade")) %>% head()
```

Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução ao R

Estatística Descritivas

Tratamento de Dados

Visualização de Dados

Gráficos interativos

filter()

Você pode criar filtros para selecionar observações com base em um ou mais critérios. Esses critérios podem ser estabelecidos usando operadores lógicos, como >, >=, <, <=, != e ==, ou através de funções que retornam valores booleanos (ou seja, TRUE ou FALSE)

```
Introdução ao
R e
Tratamento
de Dados
```

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução ao R

Estatísticas Descritivas

Tratamento de Dados

Visualização de Dados

Gráficos

```
# tarefa: filtrar as linhas onde o valor da
# coluna salário é maior que 3.
dados %>% filter(salario > 3) %>% head()
# tarefa: filtrar todas os dados que cujo salário
# é diferente de 3.
dados %>% filter(salario != 3) %>% head()
```

Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução a

Estatísticas Descritivas

Tratamento de Dados

Visualização de Dados

Gráficos interativos

mutate()

Serve para criar novas colunas que são funções de colunas já existentes no data frame

```
Introdução ao
R e
Tratamento
de Dados
```

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução ao

Estatísticas Descritivas

Tratamento de Dados

Visualização

Gráficos interativos

```
# tarefa: criar uma coluna com o valor do salário
# transformado para a escala logarítmica
dados %>%
  mutate(log salario = log(salario)) %>% head()
```

Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução a

Estatística Descritivas

Tratamento de Dados

Visualização de Dados

Gráficos interativos

group_by() e summarise()

A combinação de group_by() e summarise() é uma ferramenta poderosa para agregar e resumir dados. Quando usamos group_by(), as funções aplicadas com summarise(), ou até mesmo com mutate() ou filter(), não são aplicadas a todo o conjunto de dados, mas sim a cada grupo definido pela variável especificada na função group_by. Isso permite análises mais detalhadas e específicas em subconjuntos de dados.

Tratamento de Dados: Limpeza e tratamento

```
Tratamento
de Dados
Rodolpho J.
D. Quintela,
e-mail:
rodolpho.qui
```

Introdução ao R e

Introdução ao

ntela@urca.br

Estatísticas Descritivas

Tratamento de Dados

Visualização de Dados

```
# tarefa: criar uma coluna com o valor médio do salár
# segundo a região de procedência
dados %>%
  group_by(grau_instrucao) %>%
  summarise(media_salario = mean(salario))
```

Introdução ao R e Tratamento de Dados

D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução ao

Estatísticas Descritivas

Tratamento de Dados

Visualização de Dados

Gráficos interativos

Visualização de Dados

O que é o ggplot2?

Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução a R

Estatísticas Descritivas

Tratamento de Dados

Visualização de Dados

Gráficos interativos 1. O ggplot2 é um pacote de visualização de dados no R. 2. Foi desenvolvido por Hadley Wickham. 3. O ggplot2 é altamente flexível e baseado na "Grammar of Graphics." 4. Ele permite criar gráficos complexos com facilidade.

Instalando o pacote

```
install.packages("ggplot2")
```

Carregando o pacote

```
library("ggplot2")
```

Visualização de Dados: esqueleto do ggplot2

Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução a R

Estatística Descritivas

Tratament de Dados

Visualização de Dados

Gráficos interativo Cada camada em um gráfico contém elementos visuais chamados de "geoms," que são objetos geométricos, e possuem propriedades estéticas (aesthetics ou aes) que controlam sua aparência e posicionamento no gráfico, incluindo tamanho, cor, estilo, localização e transparência. Isso permite personalizar elementos gráficos, como barras coloridas e pontos representados por diferentes formas, no ggplot2.

Visualização de Dados: esqueleto do ggplot2

```
Introdução ao
R e
Tratamento
de Dados
```

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

```
Introdução ao
R
```

Estatística Descritivas

Tratamento de Dados

Visualização de Dados

Gráficos interativo

Portanto, o código abaixo representa o "esqueleto" do código usado para produzir um gráfico no ggplot2:

Visualização de Dados: Gráfico de barras - geom_bar()

Introdução ao R e Tratamento de Dados

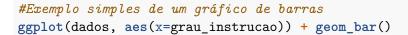
Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

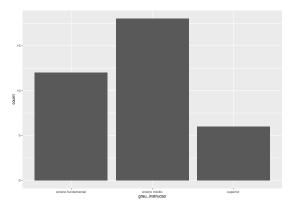
Introdução ao

Estatística: Descritivas

Tratamento de Dados

Visualização de Dados





Visualização de Dados: Histograma - geom_histogram()

Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

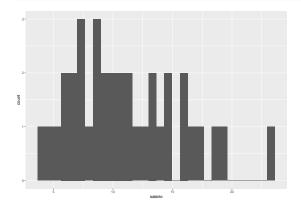
Introdução ao R

Estatísticas Descritivas

Tratament de Dados

Visualização de Dados





Visualização de Dados: Histograma - geom_histogram()

Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução ao

Estatística Descritivas

Tratamento de Dados

Visualização de Dados

```
#Exemplo simples de um gráfico de barras
ggplot(dados, aes(x=salario)) +
  geom_histogram(bins = 5) +
  theme_classic() +
  xlab("Salário") +
  ylab("Frequência")
```

Visualização de Dados: Histograma - geom_histogram()

Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

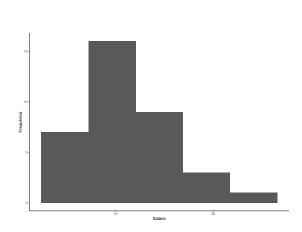
Introdução ao

Estatísticas Descritivas

Tratament de Dados

Visualização de Dados

Gráficos



Visualização de Dados: Gráfico de dispersão - geom_point()

Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução ao

Estatística Descritivas

Tratamento de Dados

Visualização de Dados

```
#Exemplo simples de um gráfico de barras
ggplot(dados, aes(y=salario, x=idade_anos)) +
  geom_point() +
  theme_classic() +
  xlab("Idade") +
  ylab("Salário")
```

Visualização de Dados: Gráfico de dispersão - geom_point()

Introdução ao R e Tratamento de Dados

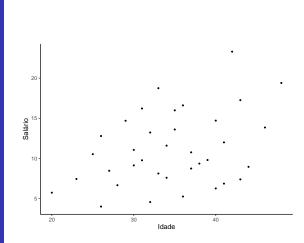
Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução ao R

Estatísticas Descritivas

Tratament de Dados

Visualização de Dados



Introdução ao R e Tratamento de Dados

D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução ao

Estatísticas Descritivas

Tratamento de Dados

Visualização

Gráficos interativos

Pacote esquisse

Introdução ao R e Tratamento de Dados

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução ao R

Estatísticas Descritivas

Tratamento de Dados

Visualização de Dados

Gráficos interativos

O pacote Esquisse possibilita a criação de gráficos interativos com o pacote 'ggplot2', permitindo recursos como a capacidade de arrastar e soltar para mapear suas variáveis de forma intuitiva.

Instalação

install.packages("esquisse")

Carregando o pacote

library("esquisse")

Pacote esquisse - Exemplos

```
Introdução ao
R e
Tratamento
de Dados
```

Rodolpho J. D. Quintela, e-mail: rodolpho.qui ntela@urca.br

Introdução ao

Estatísticas Descritivas

Tratamento de Dados

Visualização

```
esquisser(Dados, viewer = "browser")
```