

PAE–Programa de aperfeiçoamento de Ensino

Mestrado Modelagem de Sistemas Complexos

TAD – Tratamento e análise de dados / informação
Supervisor: Dr. Fernando Fagundes Ferreira
Estagiário: Ugo Henrique Pereira da Silva

Agenda

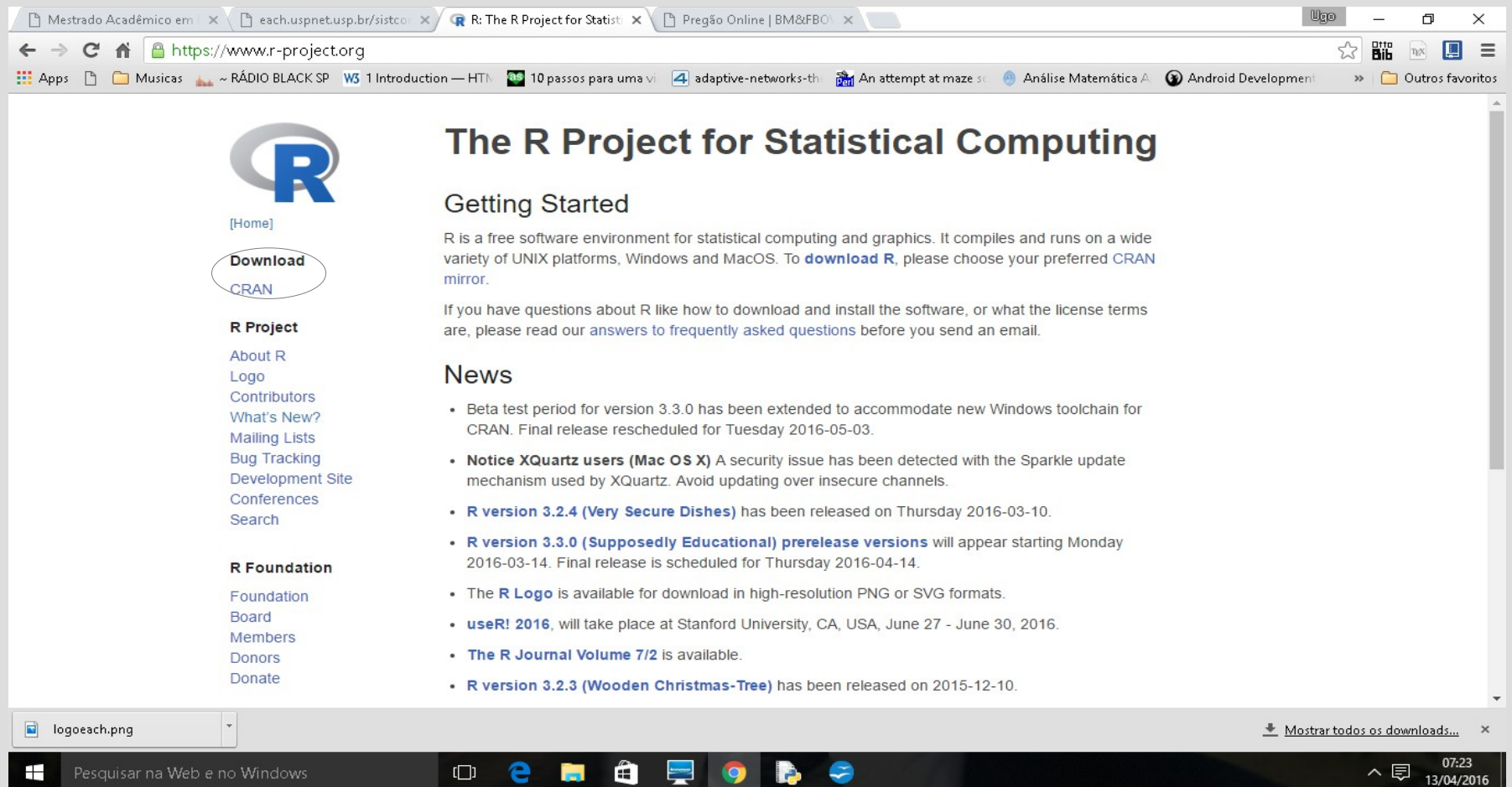
- Definição do R
- Obtenção (Download)
- Requisitos e instalação
- Uso interativo
- Funções e operações básicas
- Scripts

Definições R e RStudio

- R é um ambiente de software livre para estatística computacional e criação de gráficos.
 - Ele compila e roda em vários UNIX, Windows e MacOS.
- RStudio IDE (ambiente de desenvolvimento integrado) é uma interface facilitadora na utilização do R.
 - Livre e open source, roda no Windows, Mac, and Linux.

Obtenção do R

- <https://www.r-project.org/>



The screenshot shows a web browser window displaying the homepage of the R Project for Statistical Computing. The browser's address bar shows the URL <https://www.r-project.org/>. The page features the R logo on the left, a navigation menu with links like 'Download' and 'CRAN', and a main content area with the heading 'The R Project for Statistical Computing' and 'Getting Started'. The 'Getting Started' section includes a paragraph about R being a free software environment and a link to 'download R'. Below this is a 'News' section with several bullet points regarding version updates and security issues. The browser's taskbar at the bottom shows the Windows logo, search bar, and various application icons, along with the system clock showing 07:23 on 13/04/2016.

The R Project for Statistical Computing

Getting Started

R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To [download R](#), please choose your preferred [CRAN mirror](#).

If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

News

- Beta test period for version 3.3.0 has been extended to accommodate new Windows toolchain for CRAN. Final release rescheduled for Tuesday 2016-05-03.
- **Notice XQuartz users (Mac OS X)** A security issue has been detected with the Sparkle update mechanism used by XQuartz. Avoid updating over insecure channels.
- **R version 3.2.4 (Very Secure Dishes)** has been released on Thursday 2016-03-10.
- **R version 3.3.0 (Supposedly Educational) prerelease versions** will appear starting Monday 2016-03-14. Final release is scheduled for Thursday 2016-04-14.
- The **R Logo** is available for download in high-resolution PNG or SVG formats.
- **useR! 2016**, will take place at Stanford University, CA, USA, June 27 - June 30, 2016.
- **The R Journal Volume 7/2** is available.
- **R version 3.2.3 (Wooden Christmas-Tree)** has been released on 2015-12-10.

Obtenção do Rstudio Desktop

- <https://www.rstudio.com/products/rstudio/#Desktop>

The screenshot shows the RStudio website's product page. The browser address bar displays <https://www.rstudio.com/products/rstudio/#Desktop>. The page features a navigation menu with 'Products', 'Resources', 'Pricing', 'About Us', and 'Blog'. Below the navigation, there are two main columns: 'Open Source Edition' and 'Commercial License'. The 'Open Source Edition' column lists features such as local access, syntax highlighting, and integrated help. The 'Commercial License' column lists additional features like priority support and a commercial license for organizations. A table at the bottom compares 'Support', 'License', and 'Pricing' for both editions. The 'DOWNLOAD RSTUDIO DESKTOP' button is circled in blue.

	Open Source Edition	Commercial License
Overview	<ul style="list-style-type: none">• Access RStudio locally• Syntax highlighting, code completion, and smart indentation• Execute R code directly from the source editor• Quickly jump to function definitions• Easily manage multiple working directories using projects• Integrated R help and documentation• Interactive debugger to diagnose and fix errors quickly• Extensive package development tools	<p>All of the features of open source; plus:</p> <ul style="list-style-type: none">• A commercial license for organizations not able to use AGPL software• Access to priority support
Support	Community forums only	<ul style="list-style-type: none">• Priority Email Support• 8 hour response during business hours (ET)
License	AGPL v3	RStudio License Agreement
Pricing	Free	\$995/year

[DOWNLOAD RSTUDIO DESKTOP](#) [BUY NOW](#)

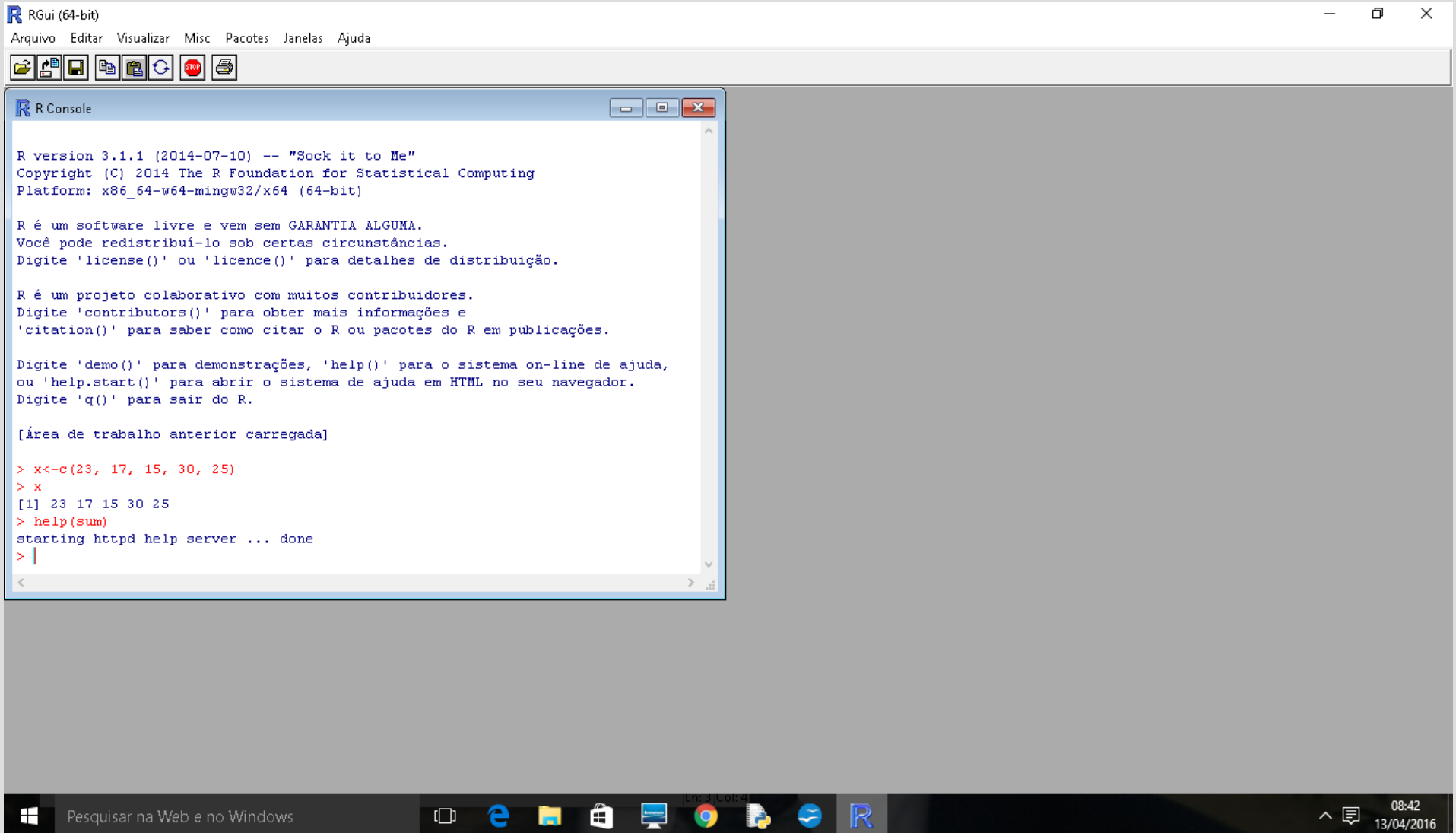
Requisitos e instalação

- <https://cran.r-project.org/doc/manuals/r-release/R-admin.html>

☰ **Uso interativo**

- Prompt de comandos “>”
- Linguagem de expressões com sintaxe simples
- Case sensitivity (maiúscula diferem de minúsculas)
- Histórico de comandos

Uso interativo



RGui (64-bit)

Arquivo Editar Visualizar Misc Pacotes Janelas Ajuda

R Console

```
R version 3.1.1 (2014-07-10) -- "Sock it to Me"
Copyright (C) 2014 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R é um software livre e vem sem GARANTIA ALGUMA.
Você pode redistribuí-lo sob certas circunstâncias.
Digite 'license()' ou 'licence()' para detalhes de distribuição.

R é um projeto colaborativo com muitos contribuidores.
Digite 'contributors()' para obter mais informações e
'citation()' para saber como citar o R ou pacotes do R em publicações.

Digite 'demo()' para demonstrações, 'help()' para o sistema on-line de ajuda,
ou 'help.start()' para abrir o sistema de ajuda em HTML no seu navegador.
Digite 'q()' para sair do R.

[Área de trabalho anterior carregada]

> x<-c(23, 17, 15, 30, 25)
> x
[1] 23 17 15 30 25
> help(sum)
starting httpd help server ... done
> |
```

Pesquisar na Web e no Windows

08:42
13/04/2016

☰ Funções e operações básicas

- Atribuições de valores e variáveis
- Operadores
 - Matemáticos
 - Relacionais
 - Outros
- Funções
 - scan(), c(), length()
 - seq()
 - max(), min()
 - runif(), rnorm()
 - mean(), median(), summary(), sd(), var()
 - log, sqrt, log10, sin, cos , tan

☰ Funções e operações básicas

- Operadores
 - Matemáticos
 - + , - , * , / , %% , ^ , **
 - Relacionais
 - > , < , >= , <= , ++ , !=
 - Outros
 - ← , =

☰ Funções e operações básicas

- Entradas de dados
 - Literais
 - Stdin
 - Arquivos

☰ Funções e operações básicas

- Entradas de dados
 - Literais e expressões
 - `> 1+3`
 - `X<- 10`
 - `Y ← x^2`
 - `z<-sqrt(y)`

☰ Funções e operações básicas

- Entradas de dados
 - Stdin
 - Uso da função scan
 - `x<-scan()`
1 6 4 3.2 9
 - Vetores
 - Uso da função `c()`
 - `Y<- c(2, 7, 3, 9, 11)`

☰ Funções e operações básicas

- Entradas de dados
 - Stdin
 - Uso da função scan
 - `x<-scan()`
1 6 4 3.2 9
 - Vetores
 - Uso da função `c()`
 - `Y<- c(2, 7, 3, 9, 11)`
 - Sequências
 - `1:10`
 - `seq(1,10,1.7)`
 - `req(0,1, length.out=11)`
 - Repetições
 - `rep(1:4,2)`
 - `Rep(3.14,5)`

☰ Funções e operações básicas

- Entradas de dados
 - Importação de dados de arquivos
 - Off-line
 - `HousePrice <- read.table("data1.txt", header=TRUE)`
 - `HousePrice <- read.table("houses.data")`
 - Online
 - `cancer <-
read.csv("http://www.ime.usp.br/~noproest/dados/cancer.txt",
sep="")`

☰ Funções e operações básicas

- Números aleatórios (pseudoaleatórios)
 - `rnorm(n, mean = 0, sd = 1)`
 - `runif(n, min = 0, max = 1)`

☰ Funções e operações básicas

- Funções matemáticas básicas

- $\text{sqrt}(x)$ – retorna a raiz quadrada de x
 - $\text{sqrt}(17)$
- $\text{sqrt}(-x)$ – duas possibilidades (não real ou complexa)
 - $\text{sqrt}(-17)$ NaN
 - $\text{sqrt}(-17+0i)$ complexo
- $\text{abs}(x)$ - retorna o valor absoluto de x (número sem sinal)
- $\log(x)$ – retorna o logaritmo natural de x
- $\log_{10}(x)$ – retorna o logaritmo base 10 de x
- $\log_2(x)$ – retorna o logaritmo base 2 de x
- $\log(x, \text{base}=k)$ – retorna o logaritmo na base k de x
- $\text{exp}(x)$ – retorna x elevada ao expoente $e=2.718282$
- $\sin(x)$, $\cos(x)$, $\tan(x)$ – funções trigonométricas básicas, o argumento x está em radianos por padrão

☰ Funções e operações básicas

- Funções matemáticas básicas

- $\text{sqrt}(x)$ – retorna a raiz quadrada de x
 - $\text{sqrt}(17)$
- $\text{sqrt}(-x)$ – duas possibilidades (não real ou complexa)
 - $\text{sqrt}(-17)$ NaN
 - $\text{sqrt}(-17+0i)$ complexo
- $\text{abs}(x)$ - retorna o valor absoluto de x (número sem sinal)
- $\log(x)$ – retorna o logaritmo natural de x
- $\log_{10}(x)$ – retorna o logaritmo base 10 de x
- $\log_2(x)$ – retorna o logaritmo base 2 de x
- $\log(x, \text{base}=k)$ – retorna o logaritmo na base k de x
- $\exp(x)$ – retorna x elevada ao expoente $e=2.718282$
- $\sin(x)$, $\cos(x)$, $\tan(x)$ – funções trigonométricas básicas, o argumento x está em radianos por padrão

☰ Funções e operações básicas

- Obter valores máximos e mínimos do conjunto
 - `max(x)`
 - `min(x)`
 - Exercício calcular a amplitude do conjunto
- Obter o número de elementos do conjunto
 - `length(x)`

☰ Funções estatísticas

- Média (ex. 4 e 5 da lista 1)
 - `mean(x)`
- Mediana
 - `median(x)`
- Resumindo
 - `summary(x)`
- Histograma
 - `hist(x)`

☰ Funções estatísticas

- Variância
 - $\text{var}(x)$
- Desvio padrão
 - $\text{sd}(x)$
- Coeficiente de variação
 - $Cv = S / \bar{x}$
- Separatrizes
 - $\text{quantile}(x)$

Script de apoio

- #Dados qualitativos
- #conceitos obtidos em determinada disciplina
- conceito<-c('a','a','b','b','a','c')
- #obter frequencia
- tabelafrequencia<-table(conceito)
- #exibir tabela de frequência
- tabelafrequencia
- #mais amigável
- cbind(tabelafrequencia)
- #obter a frequencia relativa
- r<-tabelafrequencia/length(conceito)
- #exibir frequencia relativa
- r
- #menos digitos
- old = options(digits=1)

Script de apoio

- #estatura
- estatura<-c('alto','medio','medio','alto','alto','baixo')
- summary(estatura)
- factor(estatura)
- table(estatura)
- plot(t,main="Estatura dos alunos EACH")

Script de apoio

#frequencia acumulada do conjunto

- `fi<-c(29,40,38)`
- `d<-cumsum(fi)`

Script de apoio

```
x<-scan()
```

- 146 152 168 174 180 178 179 180 178 178 168 165 152
144

Script de apoio

```
x<-scan()
```

- 24 23 25 23 30 29 28 26 33 29
- 24 37 25 23 22 27 28 25 31 29
- 25 22 31 29 22 28 27 26 23 21
- 25 21 25 24 22 26 25 32 26 29

Script de apoio

#dados amostrais armazenados em vetor x

- `x<-c(23, 17, 15, 30, 25)`
- #cálculo da media
- #primeira forma
- `soma<-sum(x)`
- `media1 <- soma / length(x)`
- #segunda forma
- `media2<-mean(x)`
- #calcular variância
- `variância<-var(x)`
- #calculo do desvio padrão
- `desviopadrao <-sd(x)`
- #exibição dos dados
- x
- soma
- media1
- media2
- variância
- desviopadrao

Script de apoio

- `x <- c(11, 0, 36, 21, 31, 23, 24, -11, -11, -21)`
- `y <- c(10, -2, 29, 14, 22, 18, 14, -2, -3, -10)`
-
- `##media das variáveis x e y`
-
- `mx <- mean(x)`
- `my <- mean(y)`
- `mx`
- `my`

Script de apoio

```
#histograma
```

- altura<-scan()
- 179 155 171 170 170 171 172 180 181 185 186 186 165 167 168 167
169 152 210 189 187 188 188
- #ordena o conjunto em ordem crescente
- sort (altura)
- #exibe o histograma
- hist(altura)
- #mudar o numero de classes - melhor ajuste
- hist(altura, breaks = 4)
- #usando a fórmula de Sturges
- sturges<-1+ 3.322*log10(length((altura)))
- #definindo o numero de elementos em cada classe
- r<-max(altura)-min(altura)
- w<-r/sturges

Script de apoio

```
votos<-c(367,137,7,2)
```

- `votacao<-c('sim','nao','abstenção','ausente')`
- `impeachment<-data.frame(votacao,votos)`
- `impeachment`
- `sum(impeachment$votos)`

Script de apoio

#conta de energia elétrica

- kwh<-scan()
- 393 311 348 358 363 339 346 336 287 329 304 326 337
- #cálculo da média
- mean(kwh)
- #cálculo da mediana
- median(kwh)
- #resumo dos dados
- summary(kwh)
- #cálculo da variância
- var(kwh)
- #cálculo do desvio padrão
- sd(kwh)
- #quantile (quartil)
- quantile(kwh)
- #quantile (quartil)
- quantile(kwh,seq(0,1,0.2))
- #quantile (decil)
- quantile(kwh,seq(0,1,0.1))
- #quantile (percentil)
- quantile(kwh,seq(0,1,0.01))
- #frequencia acumulada
- tabela<-cumsum(table(kwh))
- #frequencia relativa acumulada
- tabela<-cumsum(table(kwh))/length(kwh)