Conhecendo o 

Pollyanna Gonçalves

Data Scientist & Business Intelligence Analyst at Hotmart

pollyannaogoncalves@gmail.com

pollyannaogoncalves

Ladies BH

Um pouco sobre mim...

- Graduação em Ciência da Computação @ UFOP
- Mestrado em Ciência da Computação @ UFMG
  - Linha de pesquisa: Análise de Sentimentos na Web
- Especialização em Estatística @ UFMG
- Cientista de Dados & Analista de BI @ Hotmart
  - Construção de dashboards para monitoramento de KPIs
  - Implementação de modelos estatísticos para forecast (R e Python)
  - Implementação de modelos de ML para classificação, recomendação, predição (R e Python)

Ladies BH

Agenda

- História do R
- Por que aprender R?
- Download & Instalação do R
- Download & Instalação do Rstudio
- Características
  - Variáveis
  - Vetores
  - Listas
  - Matrizes
  - Funções / Funções de Grupos
  - Importação de dados / Gráficos
  - Packages
- Problemas reais resolvidos com R

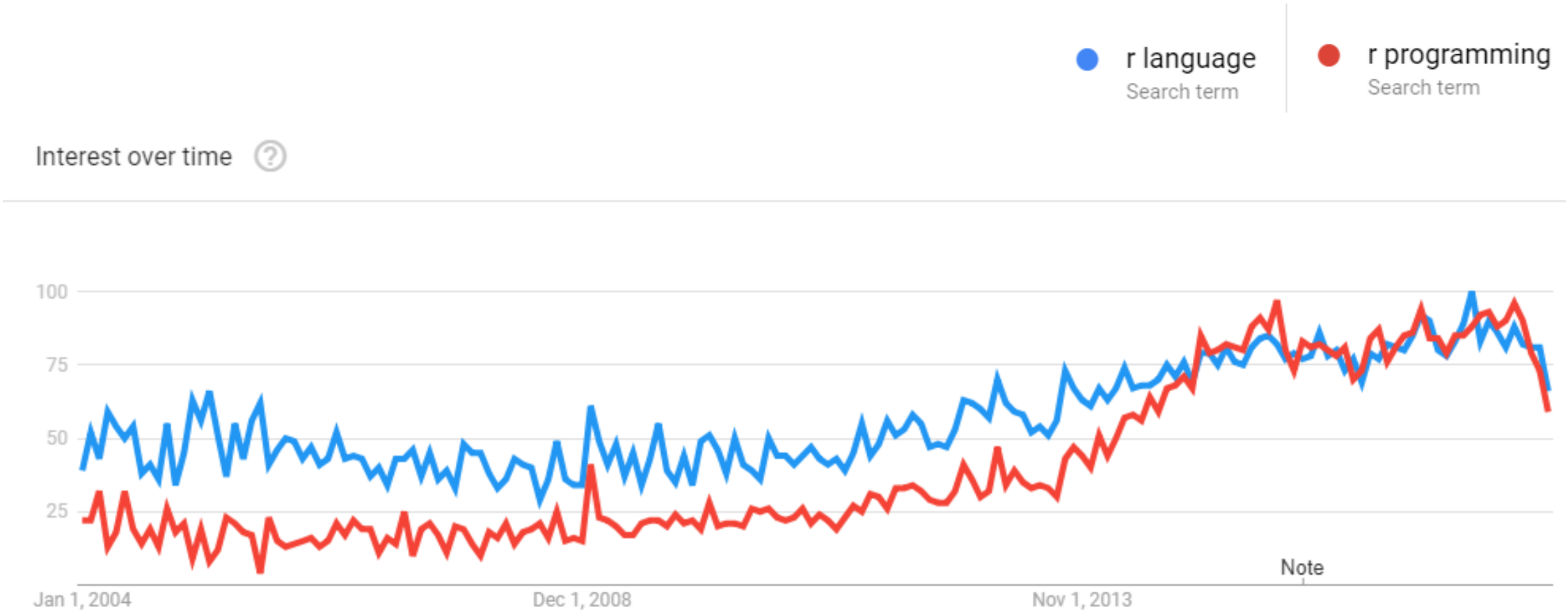
Ladies BH

História do R

História do R

- Criação em 1993
  - Professores de estatística: Ross Ihaka e Robert Gentleman
  - Universidade de Auckland / Nova Zelândia
  - Problema: bons softwares de estatística eram pagos
  - Baseada na linguagem de programação S
- Objetivo: facilitar análises estatísticas e visualização de dados
- Inicialmente foi adotada por acadêmicos
- Open source (desde 1995)

Por que aprender R?


r language
Search term
r programming
Search term
Interest over time
100
75
50
25
Note
Jan 1, 2004
Dec 1, 2008
Nov 1, 2013

Ladies BH

Por que aprender R?

- R vem se tornando principal linguagem utilizada por profissionais de Analytics, Big Data e Data Science

- A linguagem é mantida por uma comunidade colaborativa com grupos de contribuidores formados por pesquisadores de renome internacional

- Tem pacote para tudo! (E se achar, então alguém já está desenvolvendo!)

- Integrações recentes com diversos players de mercado (Oracle, Microsoft, etc.)

Ladies BH

Por que aprender R?

- Top 5 no ranking de 2017 da IEEE

Language Rank	Types	Spectrum Ranking
1. Python		100.0
2. C		99.7
3. Java		99.5
4. C++		97.1
5. C#		87.7
6. R		87.7
7. JavaScript		85.6
8. PHP		81.2
9. Go		75.1
10. Swift		73.7

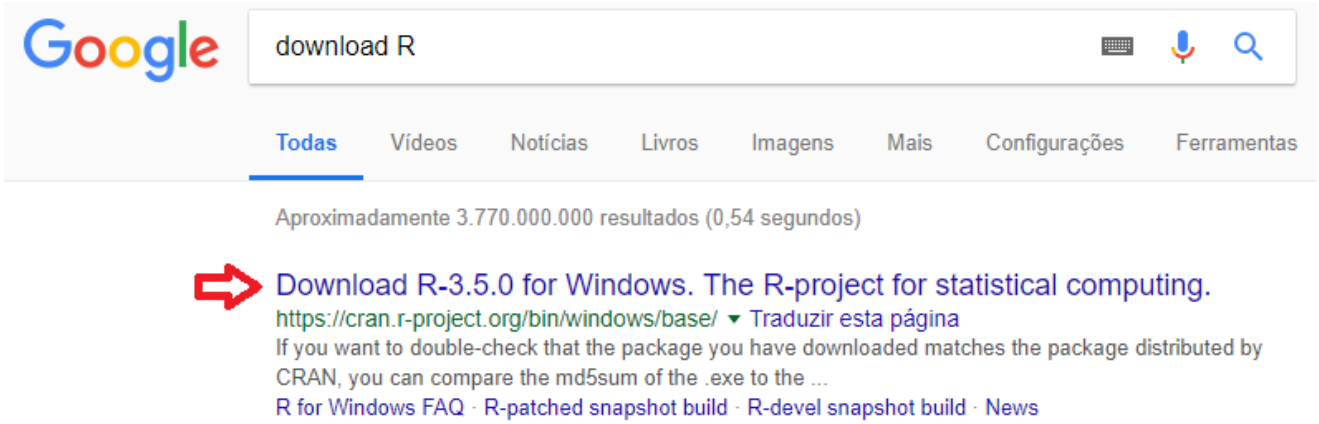
Ladies BH

Download & Instalação

Ladies BH

Download e Instalação do R

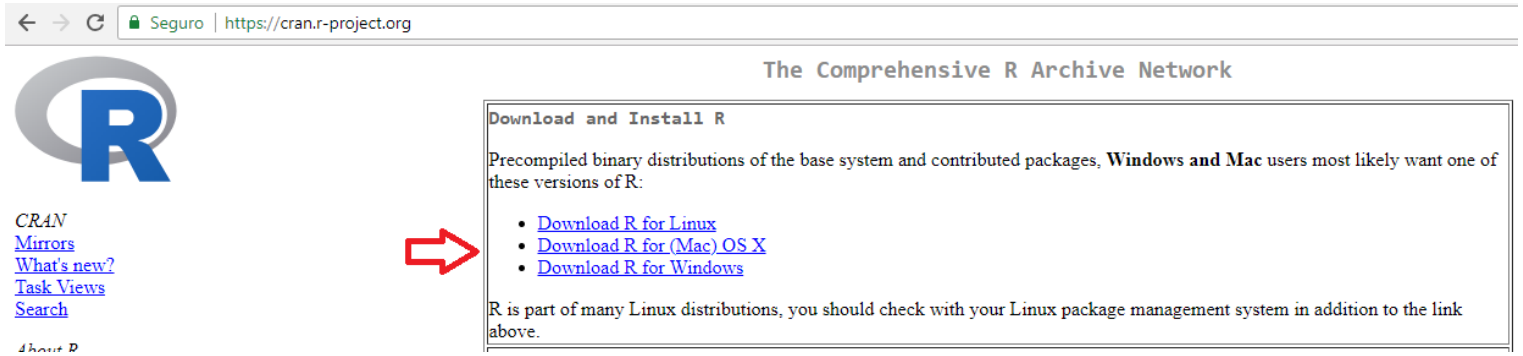
- CRAN: Rede de FTP e Web servers mundiais que armazena versões atualizadas da linguagem, libraries e documentações.



Ladies BH

Download e Instalação do R

- Multiplataforma.
- Após a instalação, já será possível começar a desenvolver na linguagem R.



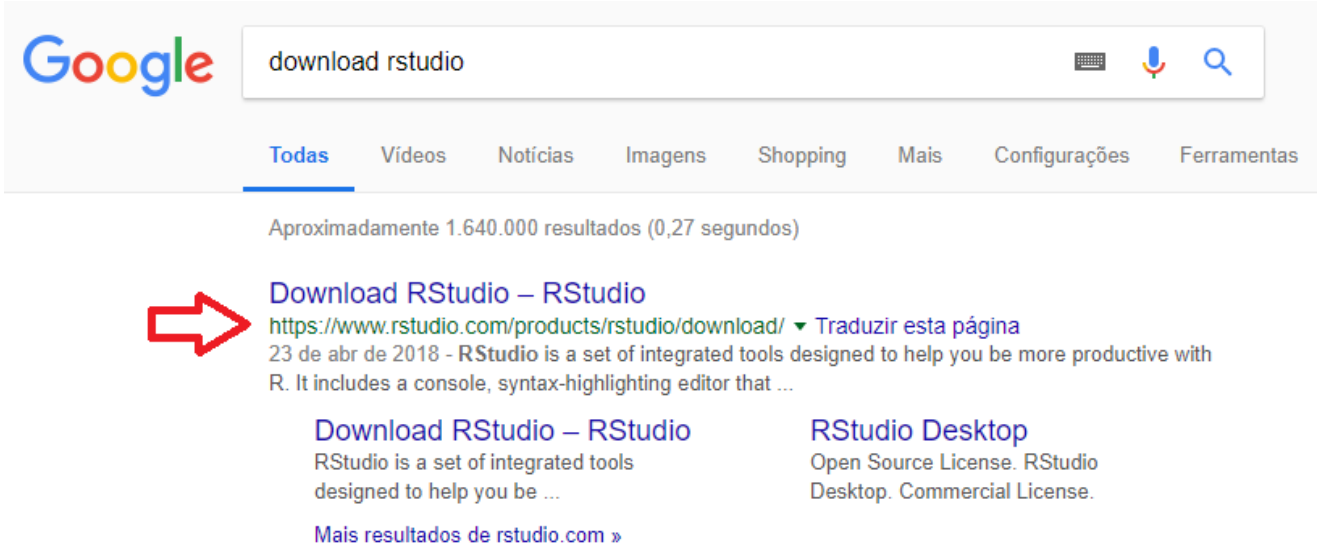
Ladies BH

Download & Instalação

Ladies BH

Download e Instalação do RStudio

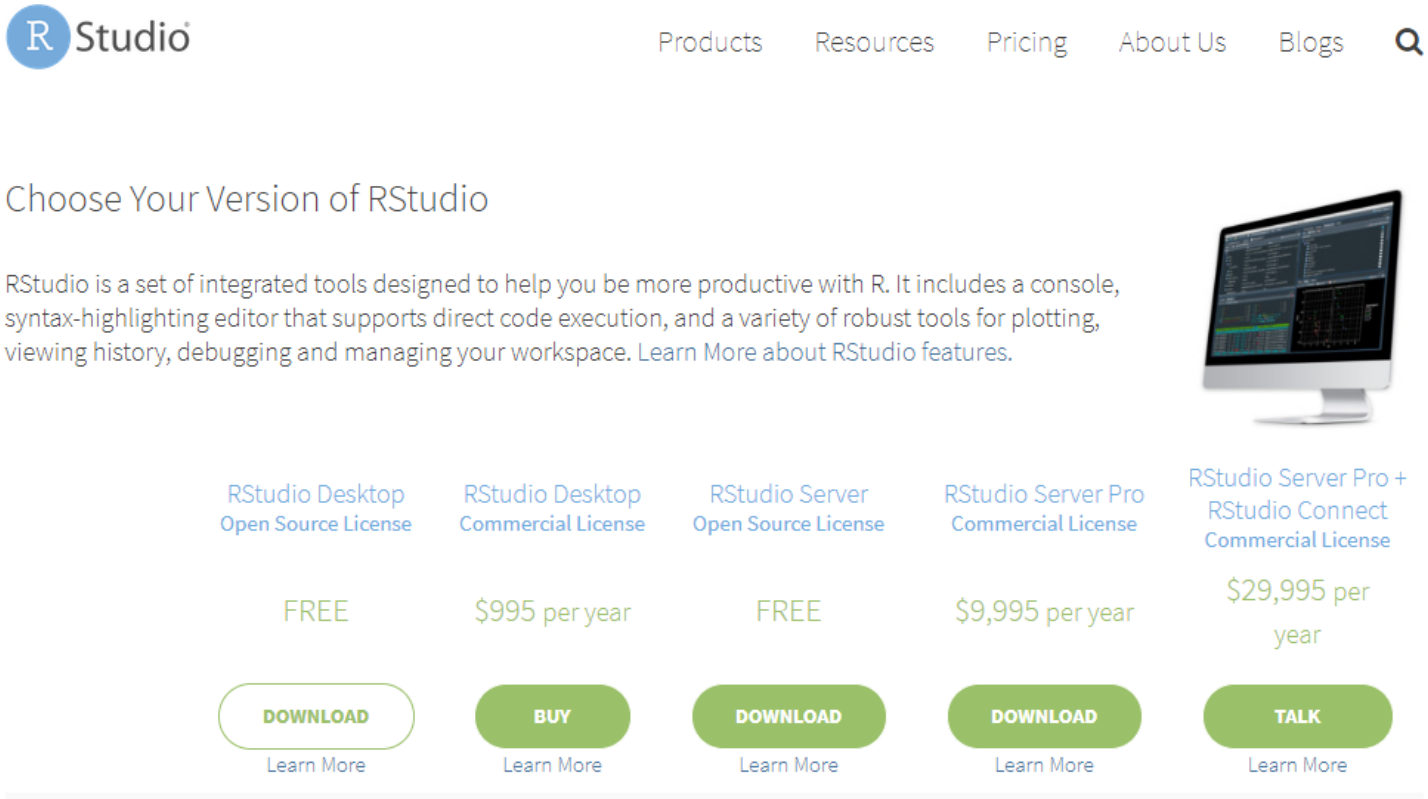
- IDE que facilita e auxilia a programação
- Open Source



Ladies BH

Download e Instalação do RStudio

- Versões:
  - Rstudio Desktop (Free)
  - Rstudio Server (possibilidade de criar schedules de scripts, etc.)



Ladies BH

Download e Instalação do RStudio

- Multiplataforma

RStudio

Products Resources Pricing About Us Blogs

Installers for Supported Platforms

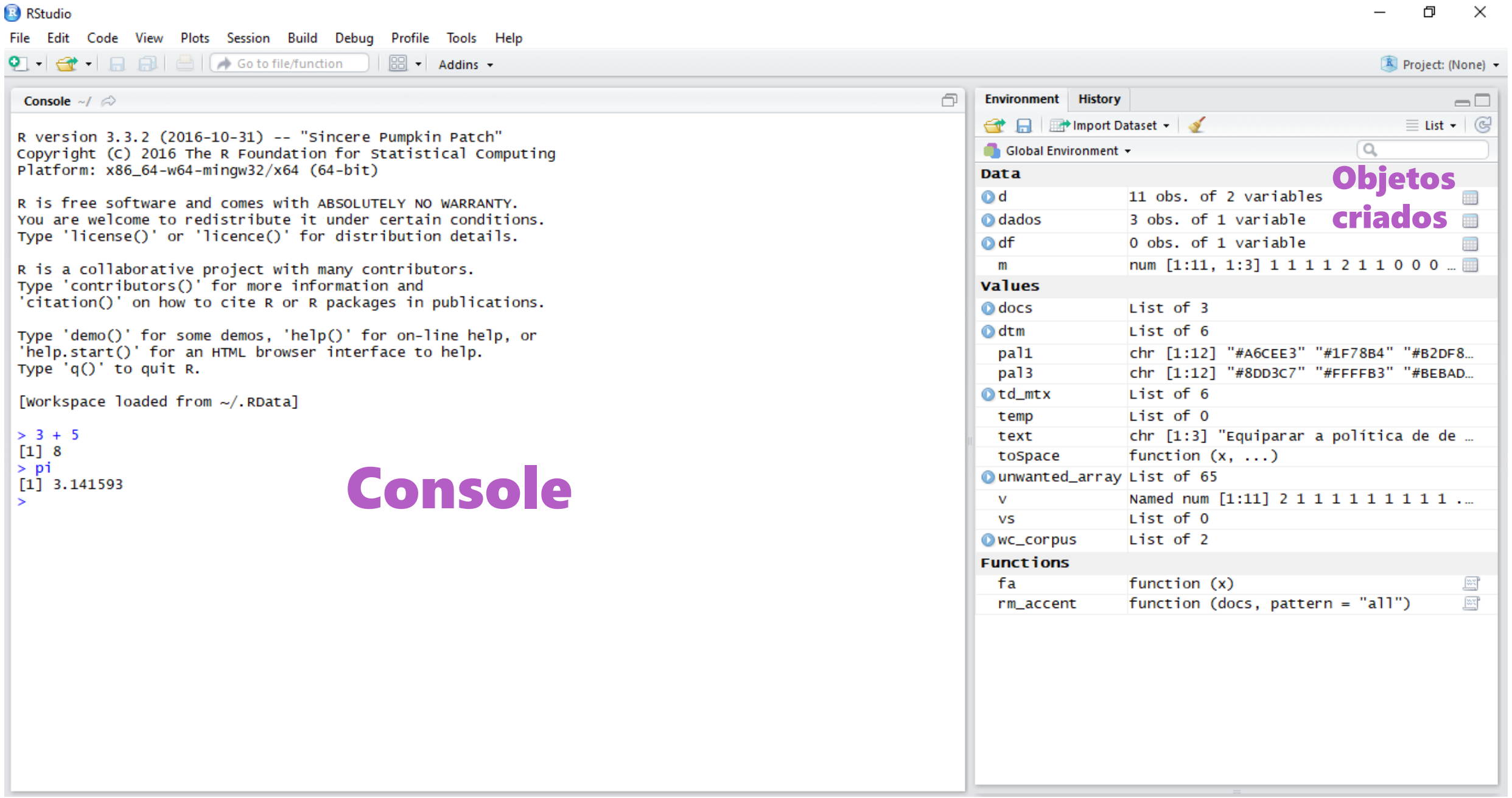
Installers	Size	Date	MD5
RStudio 1.1.453 - Windows Vista/7/8/10	85.8 MB	2018-05-16	bf287e385aef53829204023087e98735
RStudio 1.1.453 - Mac OS X 10.6+ (64-bit)	74.5 MB	2018-05-16	00a0088424ed06ac434f7a966f602b9c
RStudio 1.1.453 - Ubuntu 12.04-15.10/Debian 8 (32-bit)	89.3 MB	2018-05-16	6cfd86770c7b6dbc13e66f4f59c299ce
RStudio 1.1.453 - Ubuntu 12.04-15.10/Debian 8 (64-bit)	97.4 MB	2018-05-16	63e36e8138e369d19f9aaf4b0e995bbc
RStudio 1.1.453 - Ubuntu 16.04+/Debian 9+ (64-bit)	64.4 MB	2018-05-16	85b3e76c9fad4613bc9cf0de1f34b183
RStudio 1.1.453 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (32-bit)	88.1 MB	2018-05-16	37cade7e162eab62483e6556e39dedee
RStudio 1.1.453 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (64-bit)	90.6 MB	2018-05-16	44cddd285bc31c41e4eaec1d74b8eebb

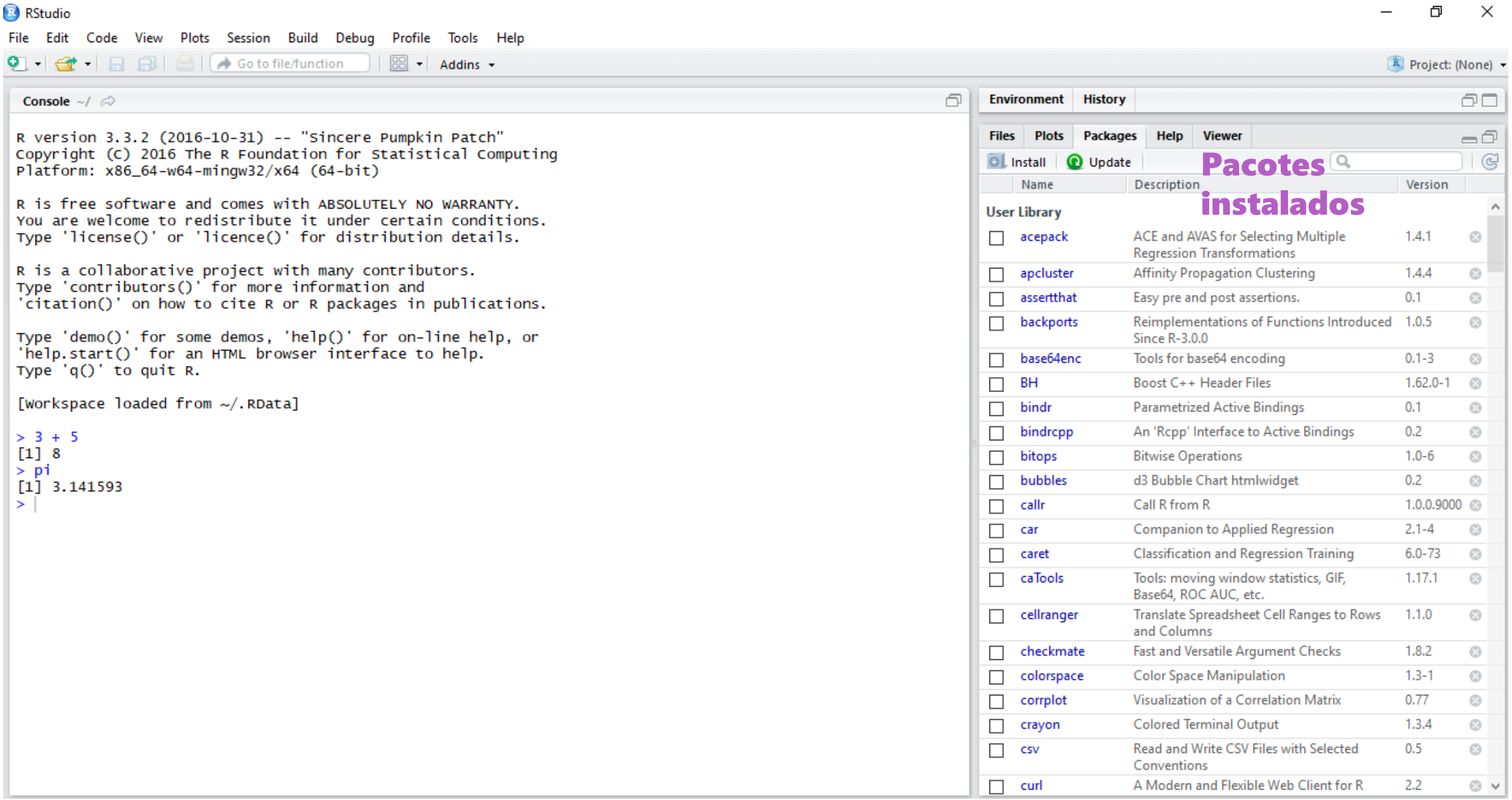
Zip/Tarballs

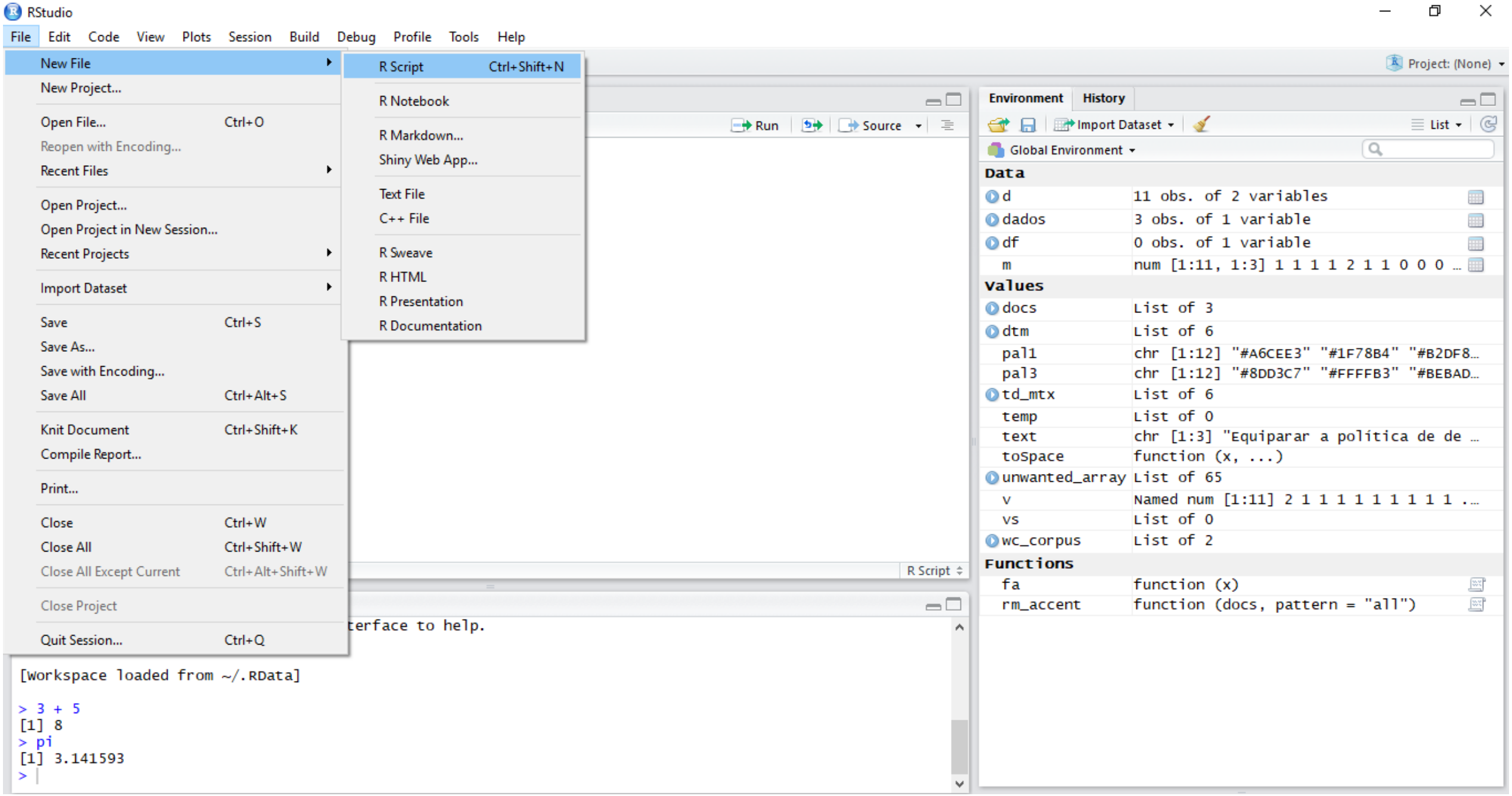
Zip/tar archives	Size	Date	MD5
RStudio 1.1.453 - Windows Vista/7/8/10	122.9 MB	2018-05-16	c933e87a1cf1a4852e67ebebd867c2bb
RStudio 1.1.453 - Ubuntu 12.04-15.10/Debian 8 (32-bit)	90 MB	2018-05-16	16e3347064194ed2a27f4f40139329e5
RStudio 1.1.453 - Ubuntu 12.04-15.10/Debian 8 (64-bit)	98.3 MB	2018-05-16	db60b00e15d235202466524cd0fbfc25
RStudio 1.1.453 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (32-bit)	88.8 MB	2018-05-16	546642e1fe826f09082f
RStudio 1.1.453 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (64-bit)	91.4 MB	2018-05-16	8df70b969a1786d65982

Try RStudio Server Pro f

Ladies BH


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
Project: (None)
Console ~/
R version 3.3.2 (2016-10-31) -- "Sincere Pumpkin Patch"
Copyright (C) 2016 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
[Workspace loaded from ~/.RData]
> 3 + 5
[1] 8
> pi
[1] 3.141593
>
Console
Environment History
Import Dataset
List
Global Environment
Objetos criados
Data
d 11 obs. of 2 variables
dados 3 obs. of 1 variable
df 0 obs. of 1 variable
m num [1:11, 1:3] 1 1 1 1 2 1 1 0 0 0 ...
Values
docs List of 3
dtm List of 6
pal1 chr [1:12] "#A6CEE3" "#1F78B4" "#B2DF8...
pal3 chr [1:12] "#8DD3C7" "#FFFFB3" "#BEBAD...
td_mtx List of 6
temp List of 0
text chr [1:3] "Equiparar a política de de ...
toSpace function (x, ...)
unwanted_array List of 65
v Named num [1:11] 2 1 1 1 1 1 1 1 1 1 ...
vs List of 0
wc_corpus List of 2
Functions
fa function (x)
rm_accent function (docs, pattern = "all")

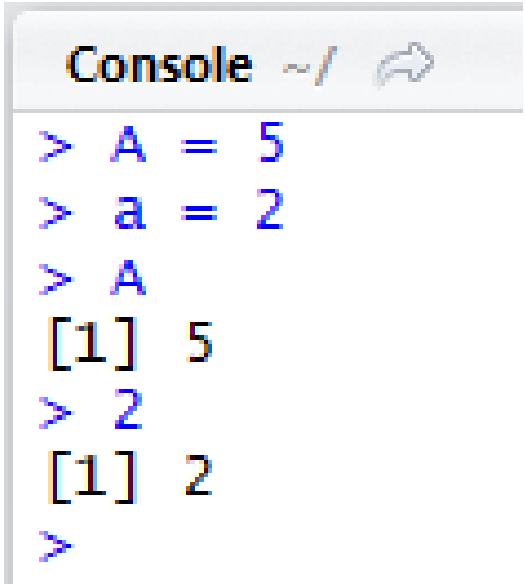

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
Project: (None)
Console ~/
R version 3.3.2 (2016-10-31) -- "Sincere Pumpkin Patch"
Copyright (C) 2016 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
[Workspace loaded from ~/.RData]
> 3 + 5
[1] 8
> pi
[1] 3.141593
>
Environment History
Files Plots Packages Help Viewer
Install Update
Pacotes instalados
Name Description Version
User Library
acepack ACE and AVAS for Selecting Multiple Regression Transformations 1.4.1
apcluster Affinity Propagation Clustering 1.4.4
assertthat Easy pre and post assertions. 0.1
backports Reimplementations of Functions Introduced Since R-3.0.0 1.0.5
base64enc Tools for base64 encoding 0.1-3
BH Boost C++ Header Files 1.62.0-1
bindr Parametrized Active Bindings 0.1
bindrcpp An 'Rcpp' Interface to Active Bindings 0.2
bitops Bitwise Operations 1.0-6
bubbles d3 Bubble Chart htmlwidget 0.2
callr Call R from R 1.0.0.9000
car Companion to Applied Regression 2.1-4
caret Classification and Regression Training 6.0-73
caTools Tools: moving window statistics, GIF, Base64, ROC AUC, etc. 1.17.1
cellranger Translate Spreadsheet Cell Ranges to Rows and Columns 1.1.0
checkmate Fast and Versatile Argument Checks 1.8.2
colorspace Color Space Manipulation 1.3-1
corrplot Visualization of a Correlation Matrix 0.77
crayon Colored Terminal Output 1.3.4
csv Read and Write CSV Files with Selected Conventions 0.5
curl A Modern and Flexible Web Client for R 2.2


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
New File
New Project...
Open File... Ctrl+O
Reopen with Encoding...
Recent Files
Open Project...
Open Project in New Session...
Recent Projects
Import Dataset
Save Ctrl+S
Save As...
Save with Encoding...
Save All Ctrl+Alt+S
Knit Document Ctrl+Shift+K
Compile Report...
Print...
Close Ctrl+W
Close All Ctrl+Shift+W
Close All Except Current Ctrl+Alt+Shift+W
Close Project
Quit Session... Ctrl+Q
R Script Ctrl+Shift+N
R Notebook
R Markdown...
Shiny Web App...
Text File
C++ File
R Sweave
R HTML
R Presentation
R Documentation
Project: (None)
Run
Source
R Script
terface to help.
[Workspace loaded from ~/.RData]
> 3 + 5
[1] 8
> pi
[1] 3.141593
>
Environment History
Import Dataset
List
Global Environment
Data
d 11 obs. of 2 variables
dados 3 obs. of 1 variable
df 0 obs. of 1 variable
m num [1:11, 1:3] 1 1 1 1 2 1 1 0 0 0 ...
Values
docs List of 3
dtm List of 6
pal1 chr [1:12] "#A6CEE3" "#1F78B4" "#B2DF8...
pal3 chr [1:12] "#8DD3C7" "#FFFFB3" "#BEBAD...
td_mtx List of 6
temp List of 0
text chr [1:3] "Equiparar a política de de ...
toSpace function (x, ...)
unwanted_array List of 65
v Named num [1:11] 2 1 1 1 1 1 1 1 1 1 ...
vs List of 0
wc_corpus List of 2
Functions
fa function (x)
rm_accent function (docs, pattern = "all")

Características

Variáveis

- Linguagem não fortemente tipada
- Não há necessidade de definir variáveis previamente
- *Case sensitive* (minúsculas e maiúsculas são consideradas diferentes)



Ladies BH

Variáveis

- Atribuições podem ser feitas de mais de uma forma

```
Console ~/
> a = 5
> a
[1] 5
> a <- 5
> a
[1] 5
> 5 -> a
> a
[1] 5
> assign("a", 5)
> a
[1] 5
```

Variáveis

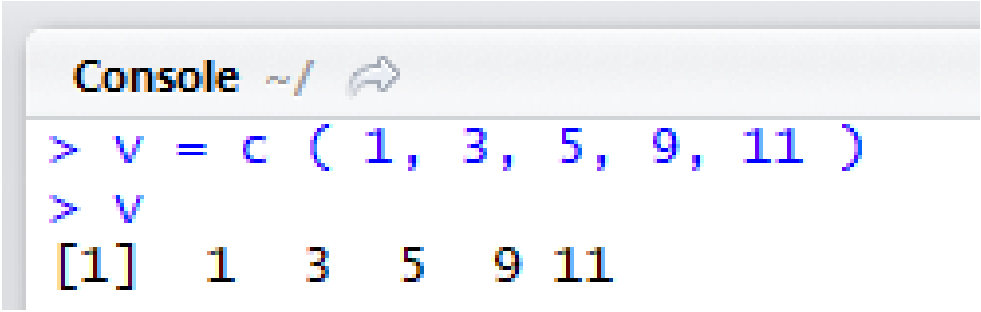
- Atribuições múltiplas

```
Console ~/
> x <- y <- z <- 15
> x
[1] 15
> y
[1] 15
> z
[1] 15
```

Ladies BH

Vetores

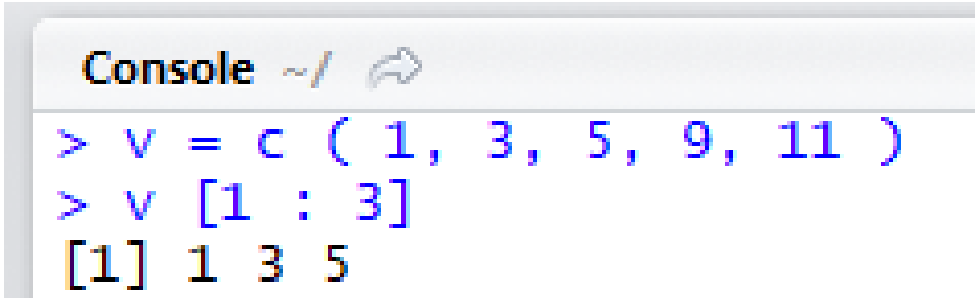
- Uso da função *c()*
- Primeira posição é 1, e não 0
- Armazena objetos de tipos iguais



Ladies BH

Vetores

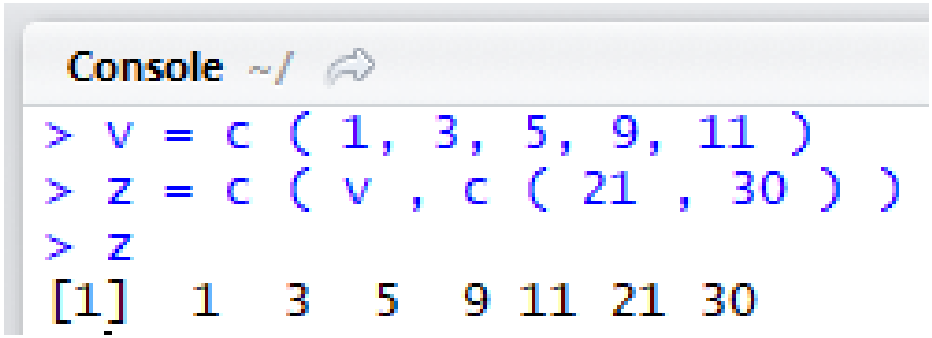
- Permite acesso consecutivo a elementos (assim como NumPy)



Ladies BH

Vetores

- Permite concatenação de vetores (já criados ou não)



Ladies BH

Matrizes

- Uso da função *matrix()*
- Parâmetros:
  - Vetor que será transformado em matriz
  - Quantidade de linhas ou colunas
  - Identifica se matriz será de linha ou coluna

```
Console ~/
> vec <- c(1, 2, 3, 4, 5, 6, 7, 8, 9)
> M <- matrix(vec, nrow=3)
> M
     [,1] [,2] [,3]
[1,]    1    4    7
[2,]    2    5    8
[3,]    3    6    9
```

Matrizes

- Acesso a posição da matriz

- Todas as operações matriciais em R são *element-wise*, ou seja, elemento por elemento

```
Console ~/
> vec <- c(1, 2, 3, 4, 5, 6, 7, 8, 9)
> M <- matrix(vec, nrow=3)
> M
     [,1] [,2] [,3]
[1,]    1    4    7
[2,]    2    5    8
[3,]    3    6    9
> M[1:2,2:3]
     [,1] [,2]
[1,]    4    7
[2,]    5    8
```

Ladies BH

Listas

- Permitem armazenar objetos de tipos diferentes
- Inicializadas com função *list()*

```
Console ~/
> obj1 = c(1,2) #objeto de vetor
> obj1
[1] 1 2
> obj2 = matrix(c(4,5,6)) # objeto de matrix
> obj2
     [,1]
[1,]    4
[2,]    5
[3,]    6
> lista = list(obj1, obj2)
> lista
[[1]]
[1] 1 2

[[2]]
     [,1]
[1,]    4
[2,]    5
[3,]    6
```

Listas

- Acesso a uma posição específica da lista

```
Console ~/
> obj1 = c(1,2) #objeto de vetor
> obj2 = matrix(c(4,5,6)) # objeto de matrix
> lista = list(obj1, obj2)
> lista
[[1]]
[1] 1 2

[[2]]
     [,1]
[1,]    4
[2,]    5
[3,]    6

> lista[[2]]
     [,1]
[1,]    4
[2,]    5
[3,]    6
```

Ladies BH

Funções

- Bastante útil em qualquer linguagem de programação

Sintaxe:

nome_da_função <- function(parâmetros)

```
Console ~/
> soma_dois_valores <- function(a,b) {
+ return(a+b) }
> soma_dois_valores(5,7)
[1] 12
```

Ladies BH

Funções de Grupo

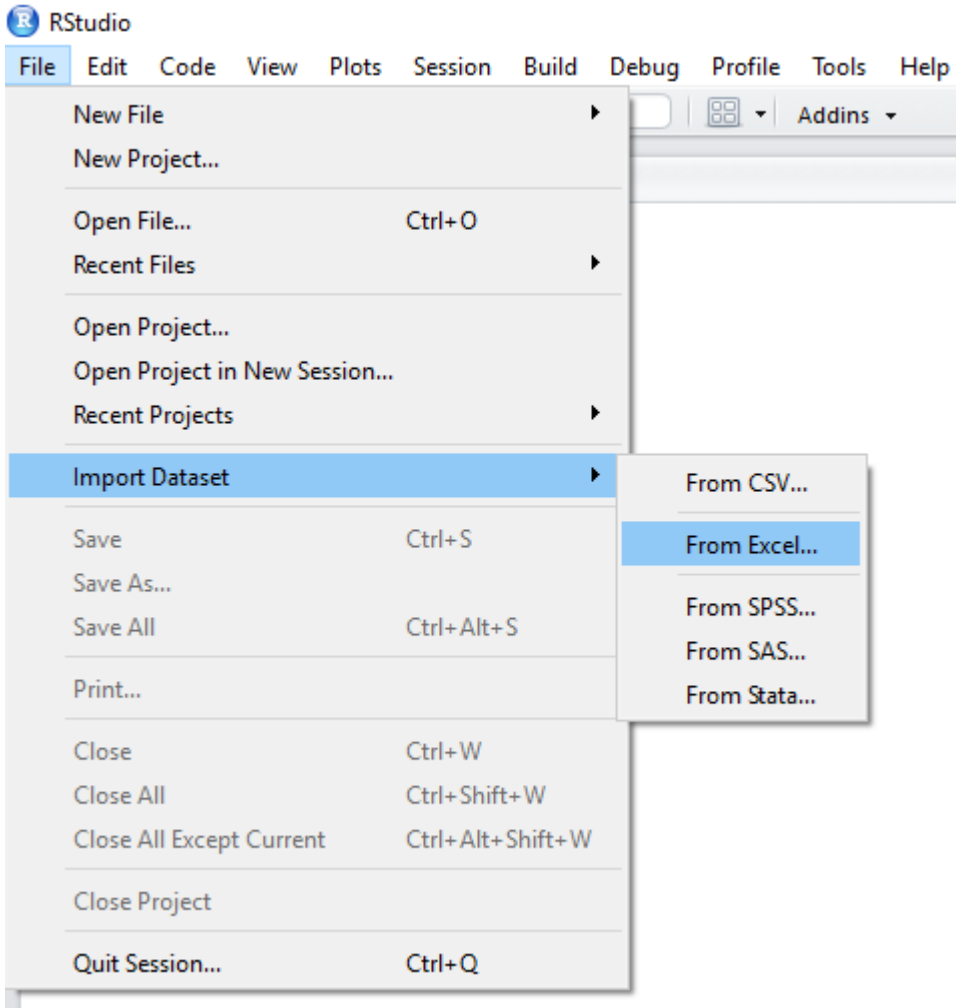
- Sum()
- Mean()
- Sd()

```
Console ~/
> v <- c(1,2,3,4,5,6,7,8,9)
> sum(v)
[1] 45
> mean(v)
[1] 5
> sd(v)
[1] 2.738613
```

Ladies BH

Importação de dados

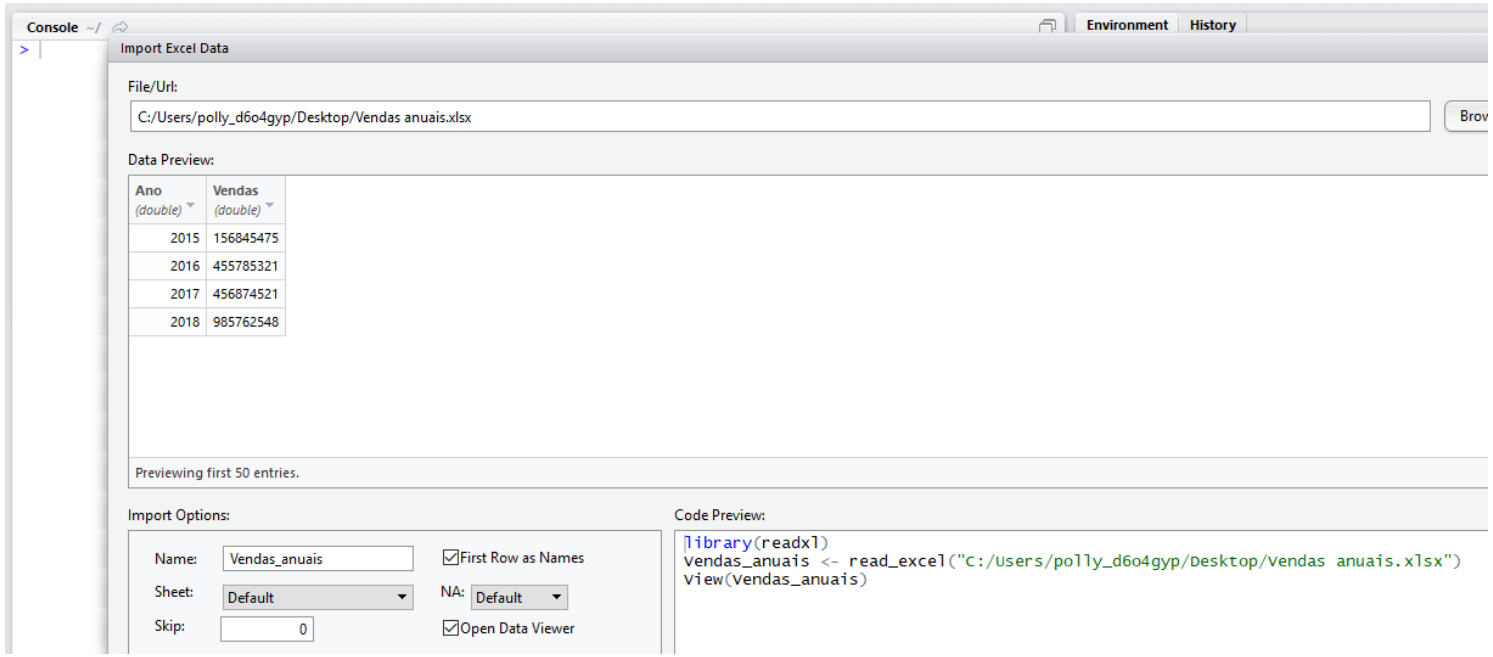
- Via assistente do RStudio
  - Utilização da biblioteca readr por padrão
- Via código
  - Escolha de biblioteca que mais te agrade



Ladies BH

Importação de dados

- Via assistente do RStudio
  - Utilização da biblioteca readr por padrão
- Via código
  - Escolha de biblioteca que mais te agrade



Ladies BH

Importação de dados

- Via assistente do RStudio
  - Utilização da biblioteca readr por padrão
- Via código
  - Escolha de biblioteca que mais te agrade

Vendas_anuais

Filter

	Ano	Vendas
1	2015	156845475
2	2016	455785321
3	2017	456874521
4	2018	985762548

Showing 1 to 4 of 4 entries

Console ~/

```
> library(readxl)
Warning message:
package 'readxl' was built under R version 3.3.3
> Vendas_anuais <- read_excel("C:/Users/polly_d6o4gyp/Desktop/Vendas anuais.xlsx")
> View(Vendas_anuais)
```

Importação de dados

- Permite acessar dados de colunas específicas

Vendas_anuais

Filter

	Ano	Vendas
1	2015	156845475
2	2016	455785321
3	2017	456874521
4	2018	985762548

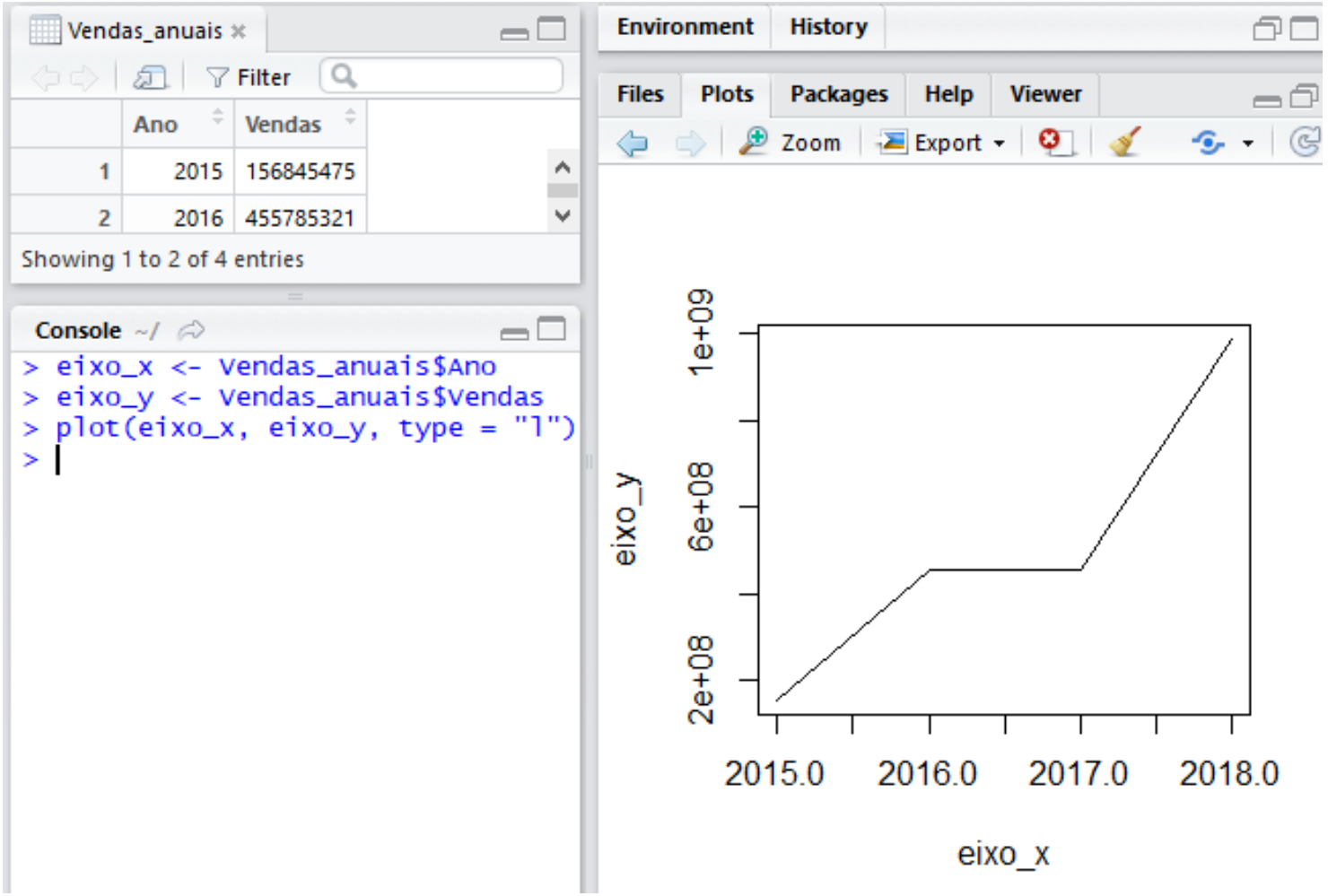
Showing 1 to 4 of 4 entries

Console ~/

```
> eixo_x <- Vendas_anuais$Ano
> eixo_x
[1] 2015 2016 2017 2018
> eixo_y <- Vendas_anuais$Vendas
> eixo_y
[1] 156845475 455785321 456874521 985762548
```

Gráficos

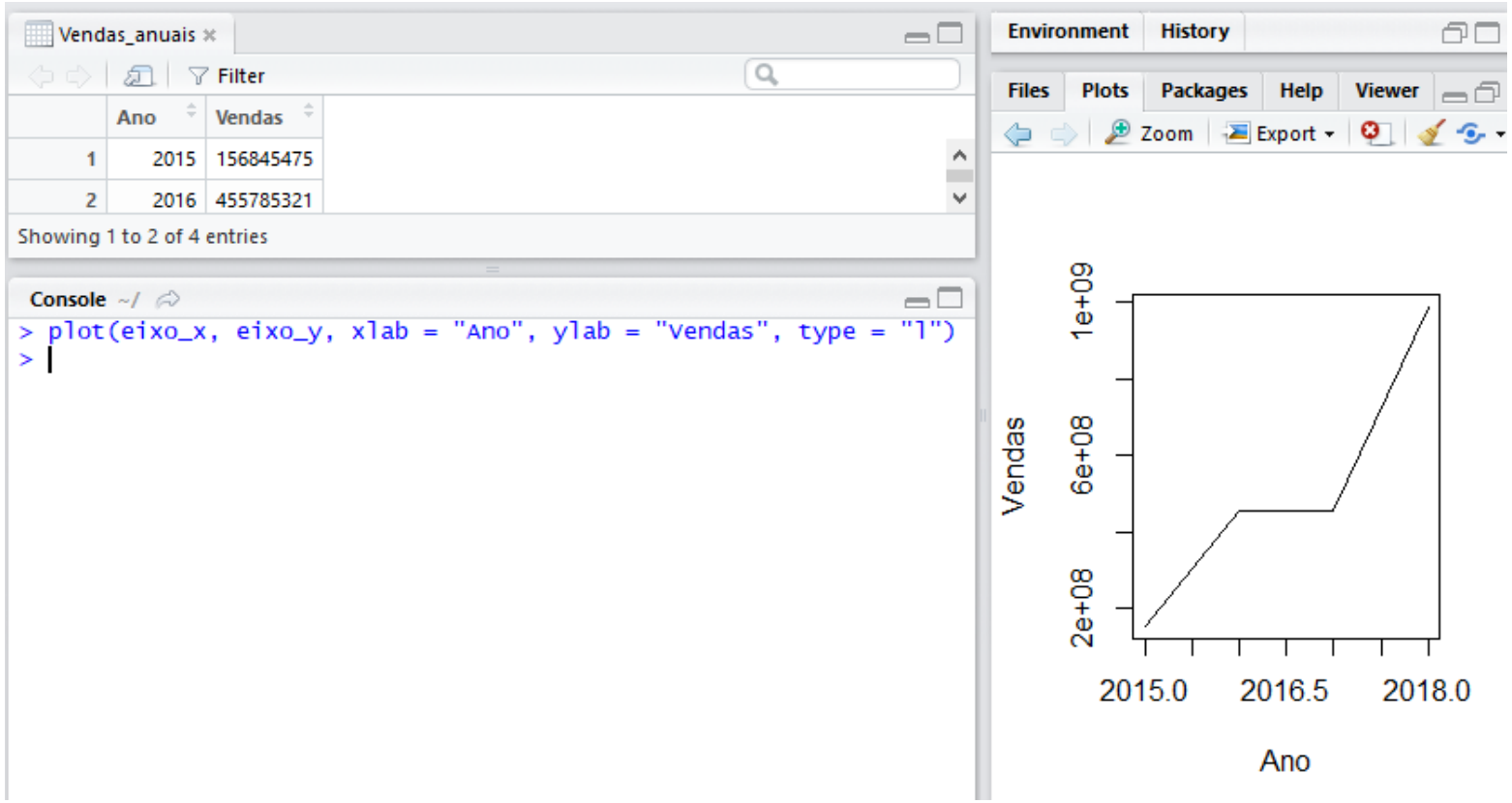
- Resultados dentro da aba *Plots*



Ladies BH

Gráficos

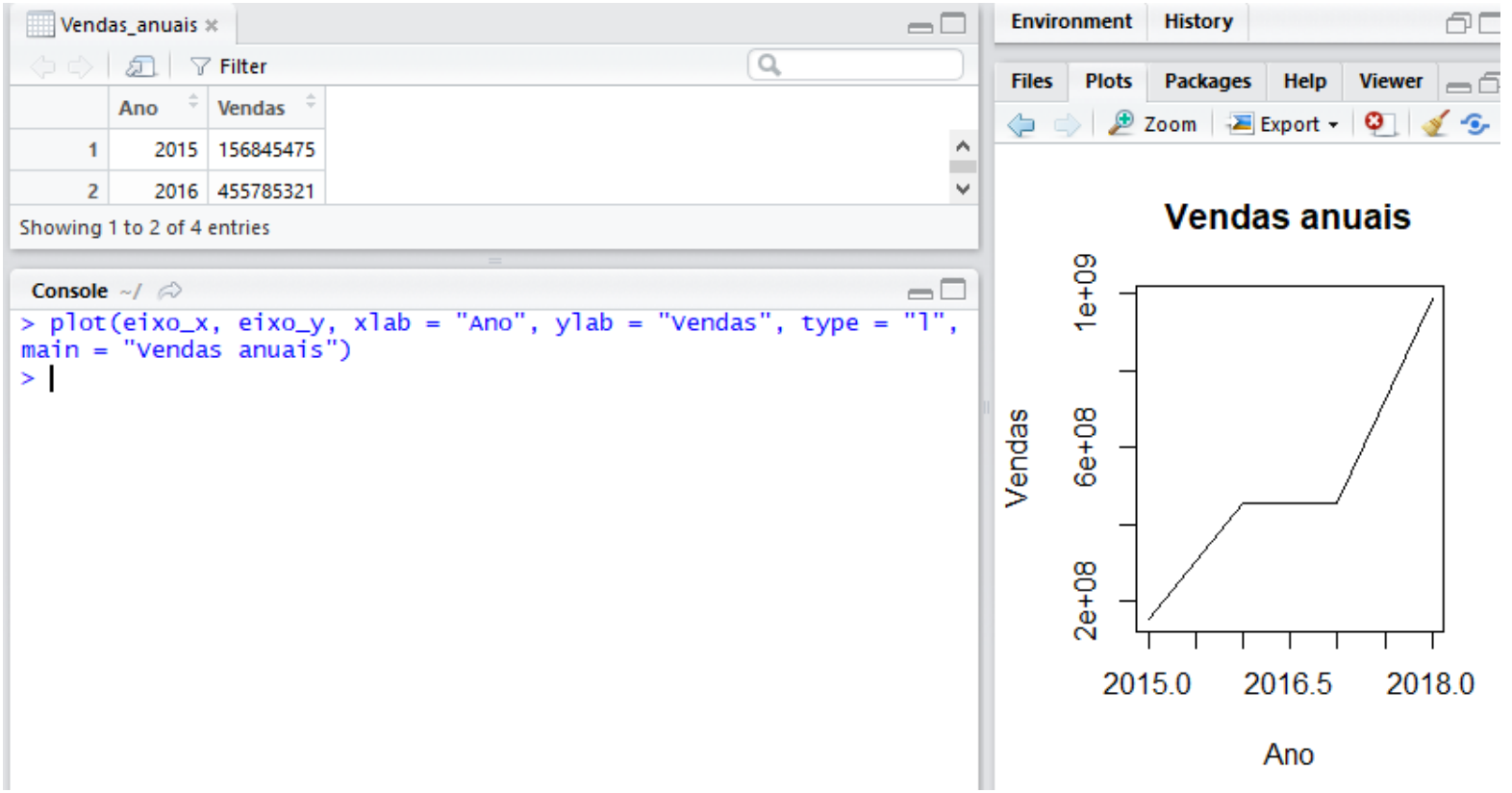
- Alteração de nomes dos eixos



Ladies BH

Gráficos

- Adição de título

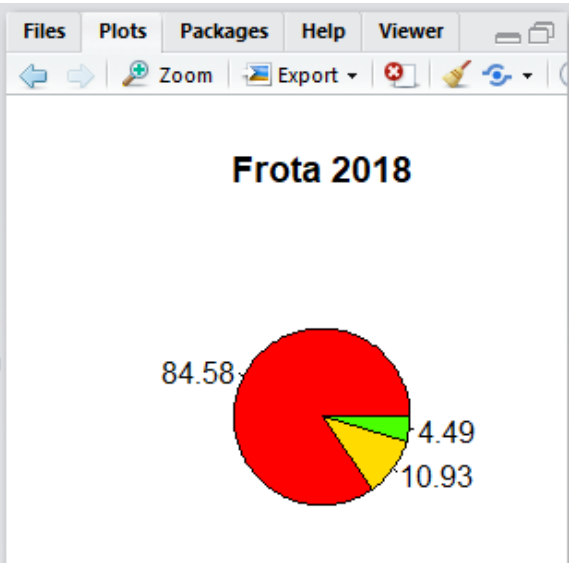


Ladies BH

Gráficos

- Gráfico de Pizza

```
Console ~/
> frota <- c(151882, 19627, 8058)
> names(frota) <- c("automóveis","motocicletas","caminhonetes")
> porc <- round( frota*100 / sum(frota), 2)
> pie(frota, main ="Frota 2018", labels=porc, col=rainbow(7))
>
```

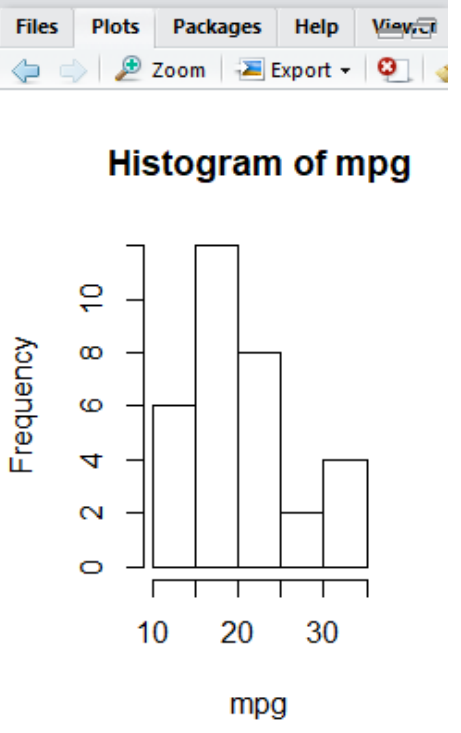


Ladies BH

Gráficos

- Gráfico de Histograma

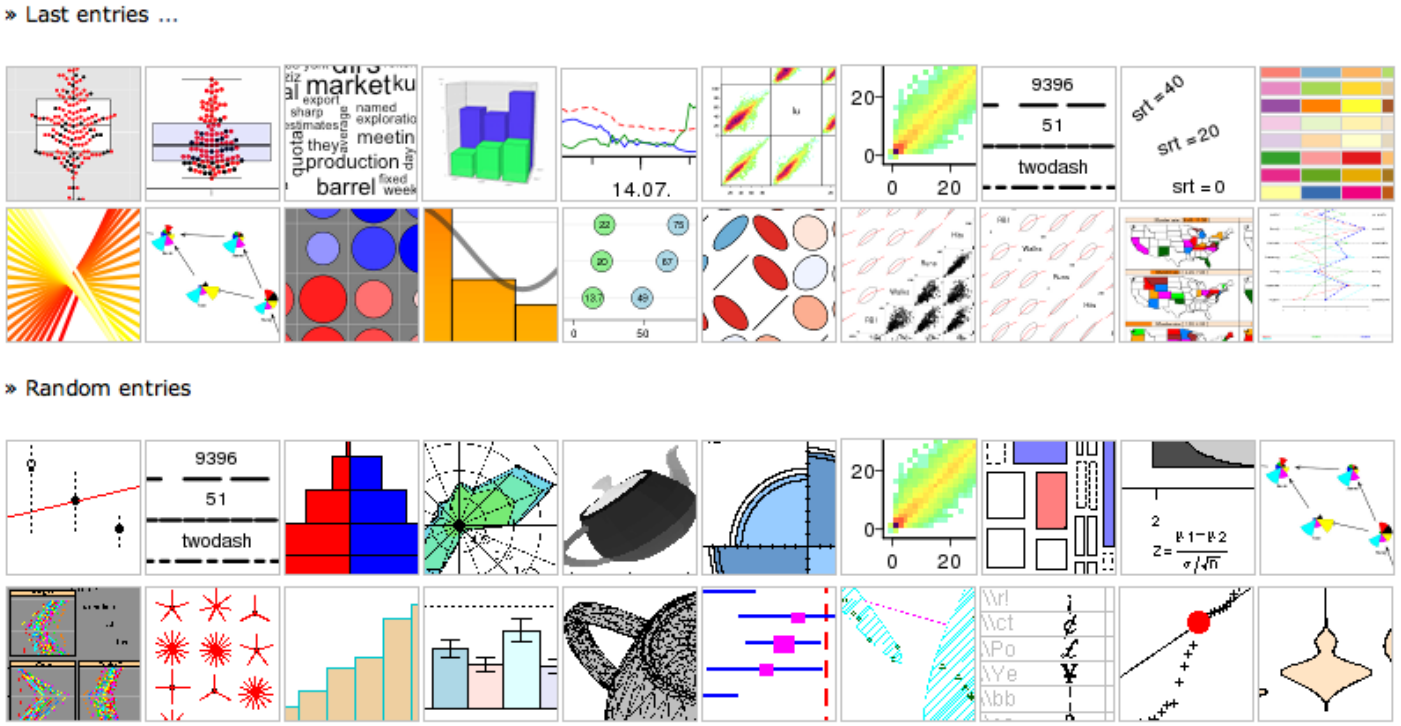
```
Console ~/
> data(mtcars) #Carregando dados disponivel no R
> mtcars[1:3,] #Visualizando parte dos dados
               mpg cyl disp  hp drat    wt  qsec vs am gear carb
Mazda RX4     21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
Mazda RX4 Wag 21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
Datsun 710    22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
> attach(mtcars) #Anexa conjunto de dados de modo a poder chamar colunas
The following objects are masked from mtcars (pos = 3):

    am, carb, cyl, disp, drat, gear, hp, mpg, qsec, vs, wt

> hist(mpg)
>
```



Ladies BH

Gráficos

- Bibliotecas:
  - PlotLy
  - HighCharter
  - Dygraphs
  - Ggiraph
  - rBokeh
  - chart.js
  - rChart

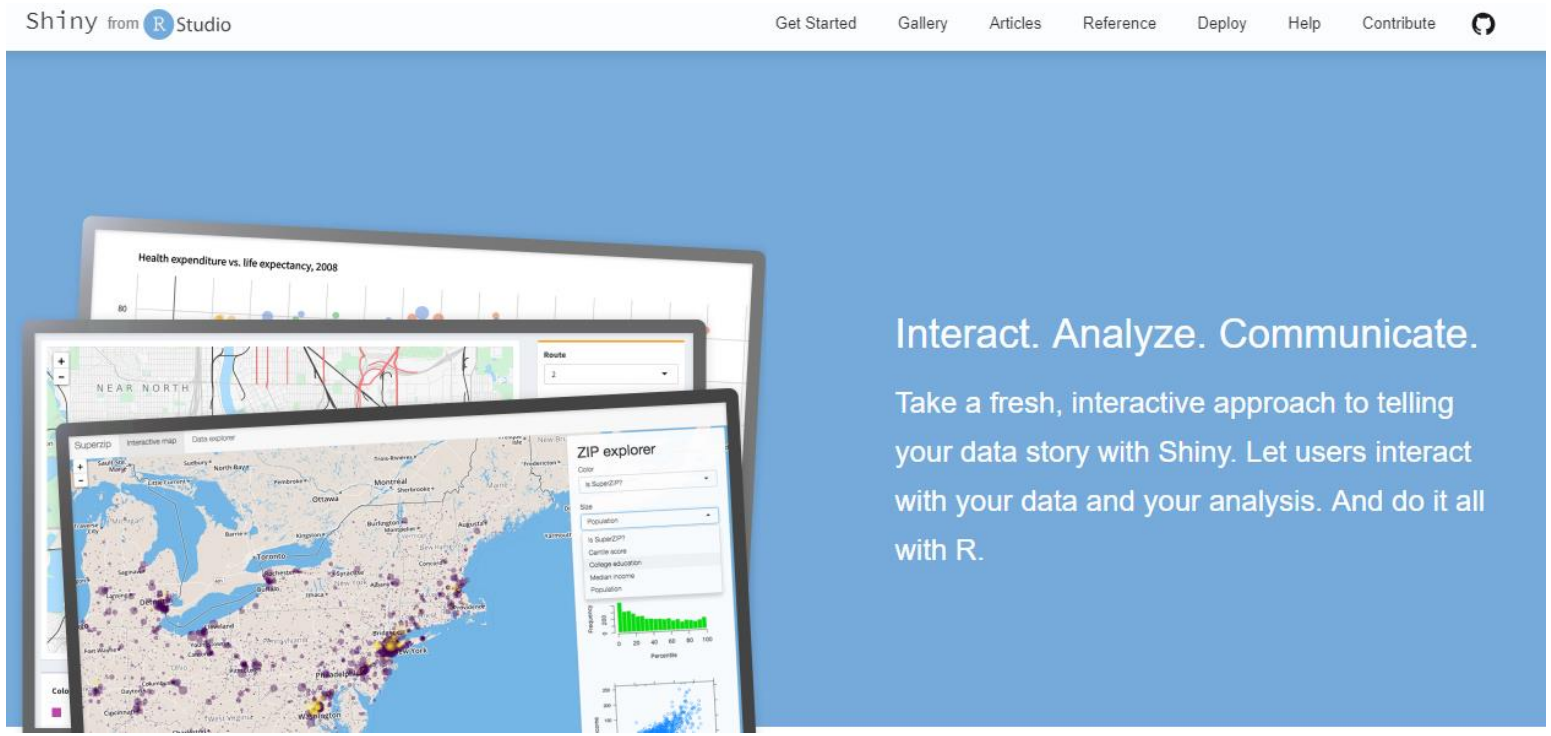


Ladies BH

Shiny

- Desenvolvido para aplicações Web
- Permite construir páginas Web interativas diretamente do R

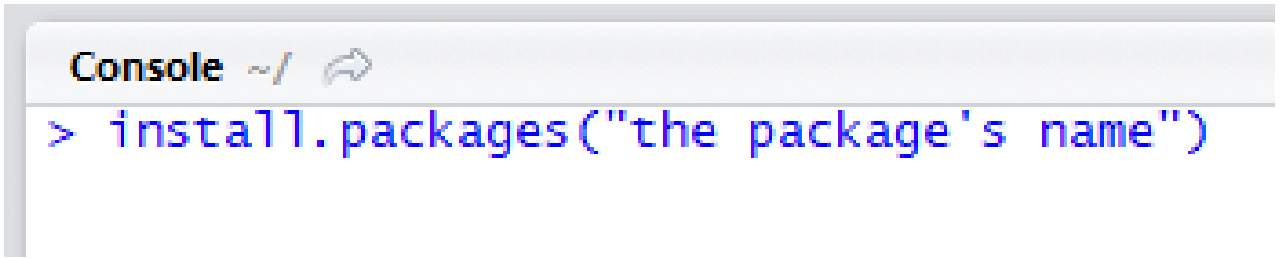
http://shiny.rstudio.com

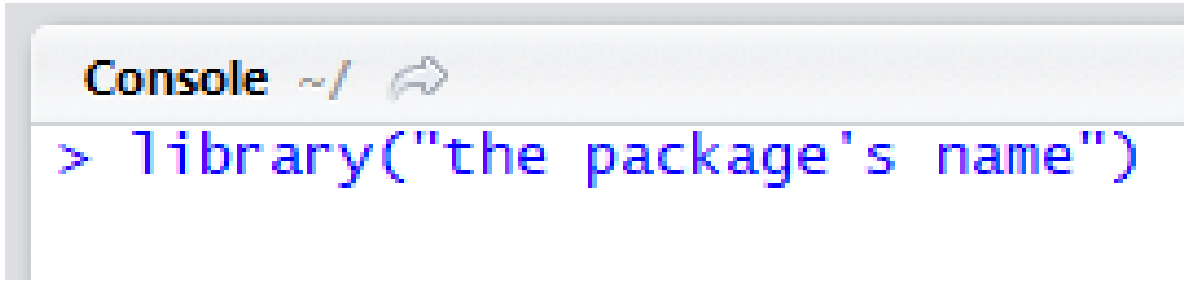


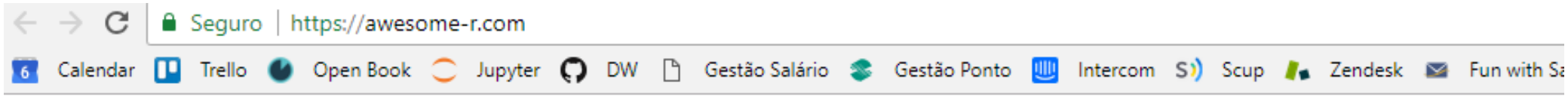
Ladies BH

Pacotes

- Desenvolvidos por uma comunidade fortemente ativa







AWESOME R

A curated list of awesome R packages and tools. Inspired by awesome-machine-learning.

For better navigation, see https://awesome-r.com

💛 for Top 50 CRAN downloaded packages or repos with 400+ ⭐

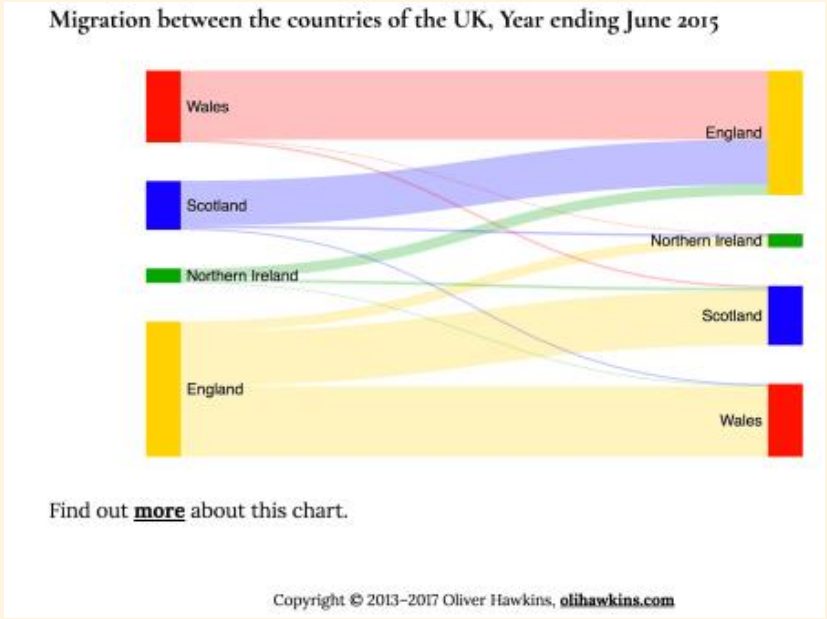
2018

- Readings in Applied Data Science - These readings reflect Hadley's personal thoughts about applied data science.
- promises - Abstractions for Promise-Based Asynchronous Programming
- tinytex - A lightweight and easy-to-maintain LaTeX distribution

Problemas reais resolvidos com R

Problema 1 – Entender como era o caminho dos visitantes de um site até o momento da compra

- Solução:
  1. Gerar dataset no formato de “fluxo” (grafo com peso nas arestas)
  2. Importar dataset no R
  3. Usar biblioteca d3Network para gerar um gráfico desse tipo:

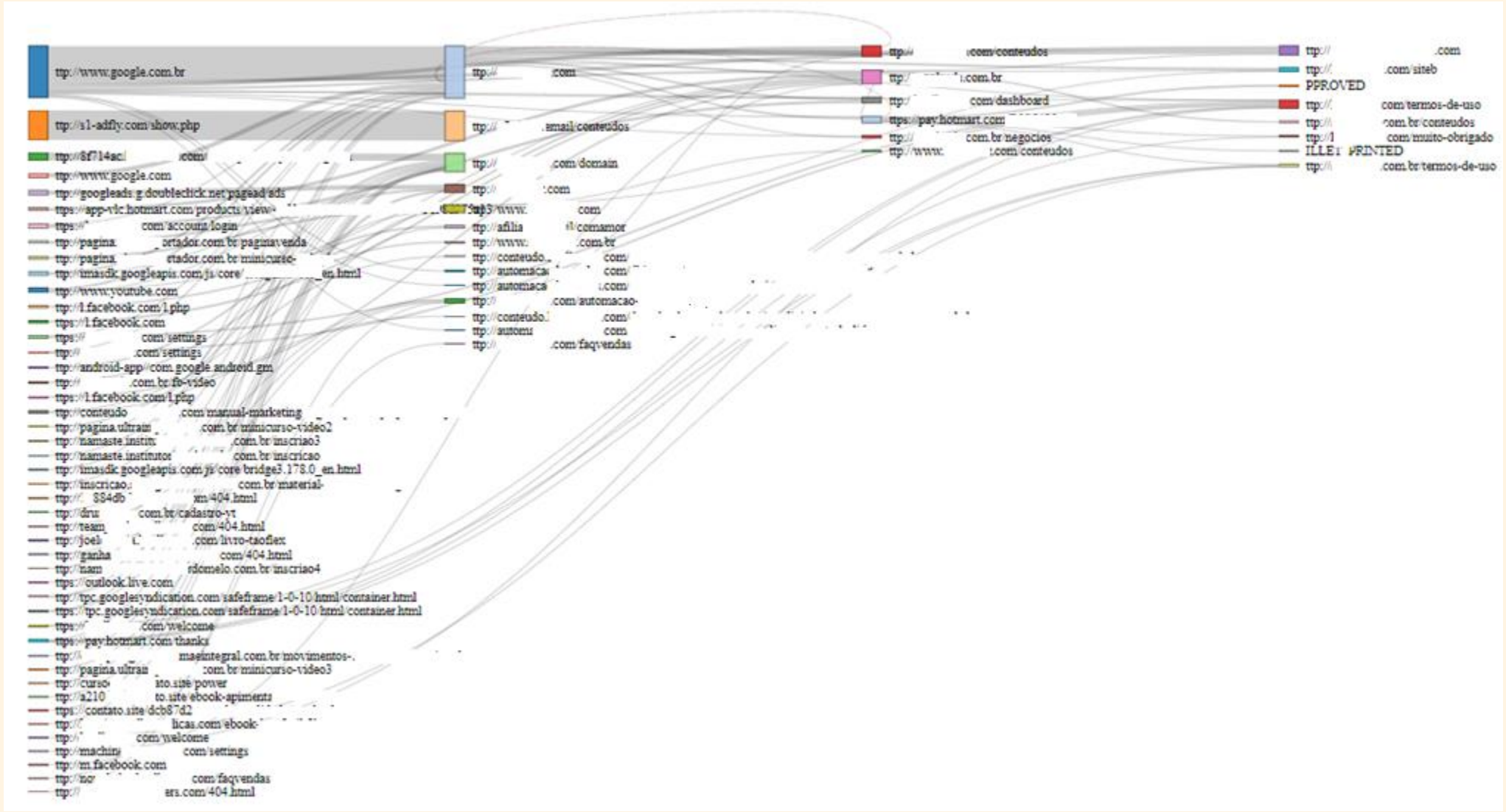


Problema 1 – Entender como era o caminho dos visitantes de um site até o momento da compra

```r
library(devtools)
library(d3Network)
library(networkD3)
library(jsonlite)
library(plyr)

nodes <- JSONtoDF(file = paste0('C:/Users/polly_d6o4gyp/Desktop/Problemas reais R/sankey.json'), array = 'nodes')
array = 'nodes'
MainList <- fromJSON(paste0("C:/Users/polly_d6o4gyp/Desktop/Problemas reais R/sankey.json"))
ArrayList <- MainList[[array]]
MainDF <- ldply(ArrayList, data.frame)
nodes = MainDF

nodes$id <- substring(nodes$id, 2)

links <- JSONtoDF(file = paste0('C:/Users/polly_d6o4gyp/Desktop/Problemas reais R/sankey.json'), array = 'links')
array = 'links'
MainList <- fromJSON(paste0("C:/Users/polly_d6o4gyp/Desktop/Problemas reais R/sankey.json"))
ArrayList <- MainList[[array]]
MainDF <- ldply(ArrayList, data.frame)
links = MainDF

colnames(nodes)<- c("name")
colnames(links)<- c("qtd_users","source","target")

d3Sankey(Nodes = nodes, Links = links, Source = 'source',Target = 'target', Value = 'qtd_users', NodeID = 'id',width = 1000,
         height = 1000, fontsize = 12,standAlone = FALSE, iframe = FALSE,
         file = 'C:/Users/polly_d6o4gyp/Desktop/Queries/Analytics/Sankey - Copia/sankey.html')

```

Ladies BH

Problema 1 – Entender como era o caminho dos visitantes de um site até o momento da compra



Ladies BH

Problema 2 – Predizer o tipo de um produto de acordo com sua performance de vendas

- Solução:
  1. Machine Learning – Classificação
  2. Modelo de Random Forest implementado em R
  3. Criação de dataset com produtos já classificados em tipo E e L

Mostrar código dentro do Rstudio

- Resultado: modelo treinado com taxa de assertividade de 85%

Problema 3 – Criação modelo de forecast para faturamento mensal

- Solução:
  1. Estatística – ARIMA
  2. Modelo implementado em R

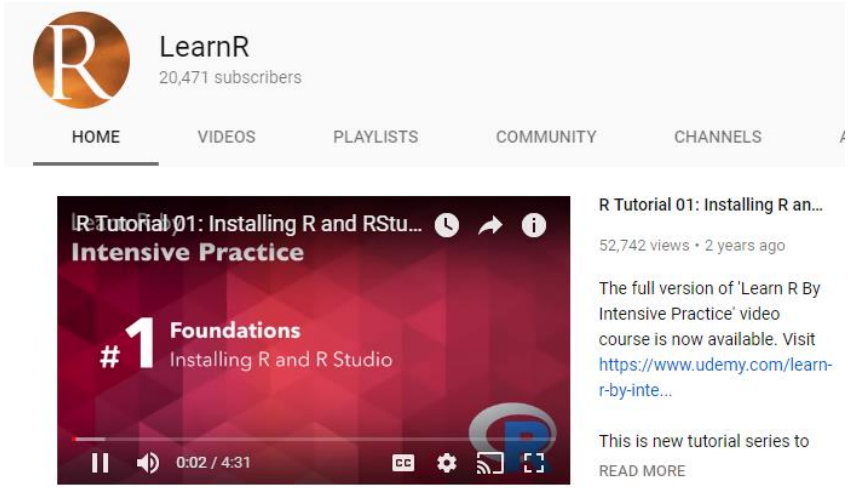
Mostrar código dentro do Rstudio

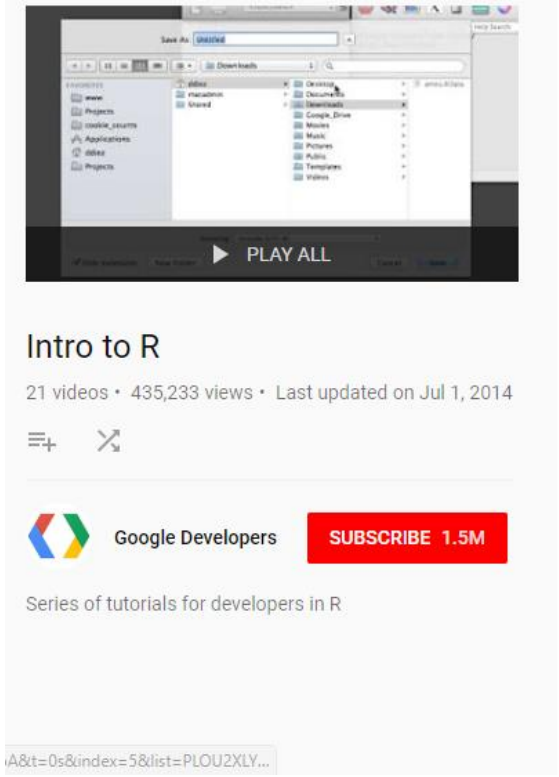
- Resultado: modelo de forecast mensal com erro médio de ~3%

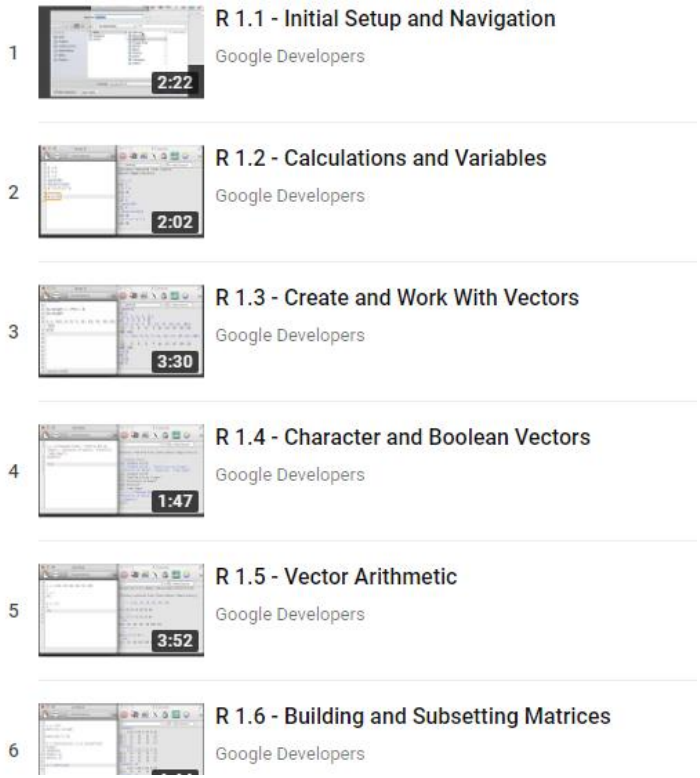
Dicas para acelerar o aprendizado!

Canais no Youtube

- LearnR
- Google Developers
- MarinStatsLectures
- ...



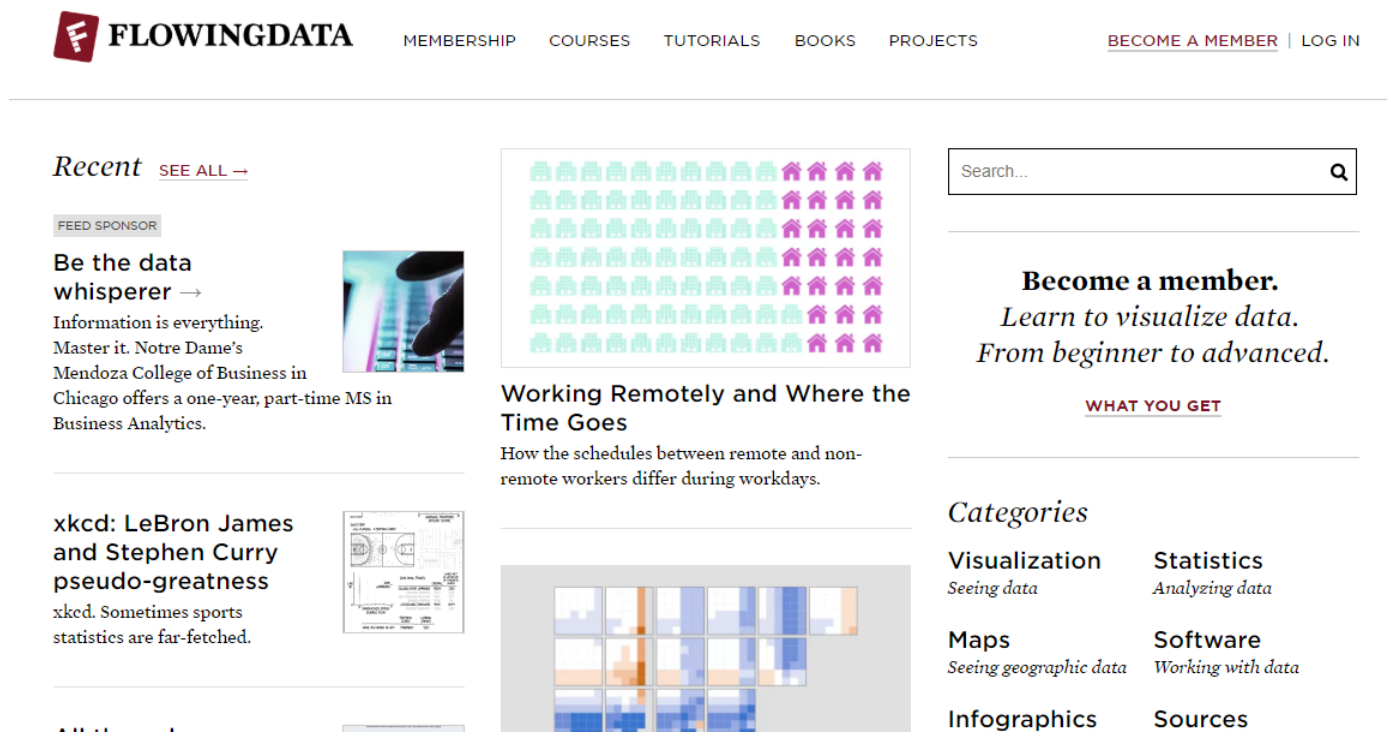




Ladies BH

Blogs

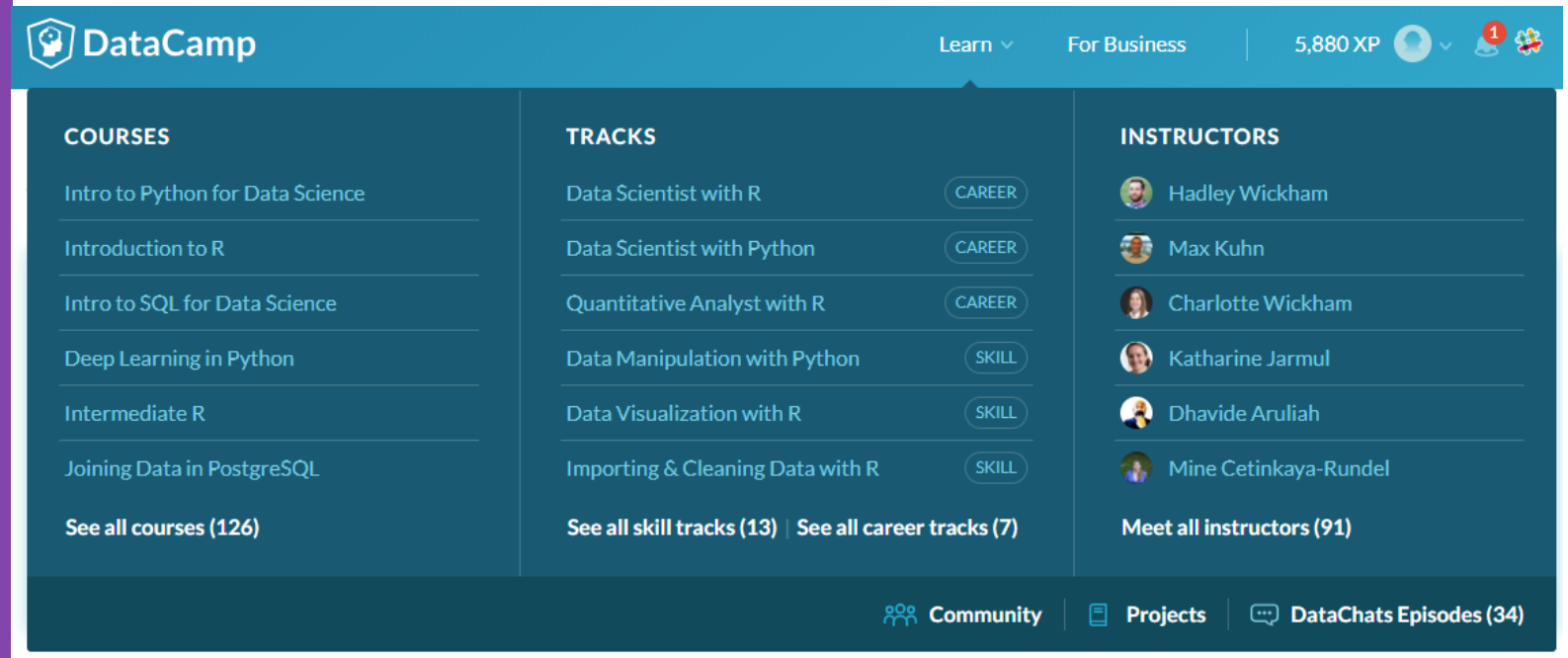
- Revolutions (blog da Microsoft sobre R)
- Civil Statistician
- Flowing Data
- Datazar Blog
- ...



Ladies BH

Cursos Online

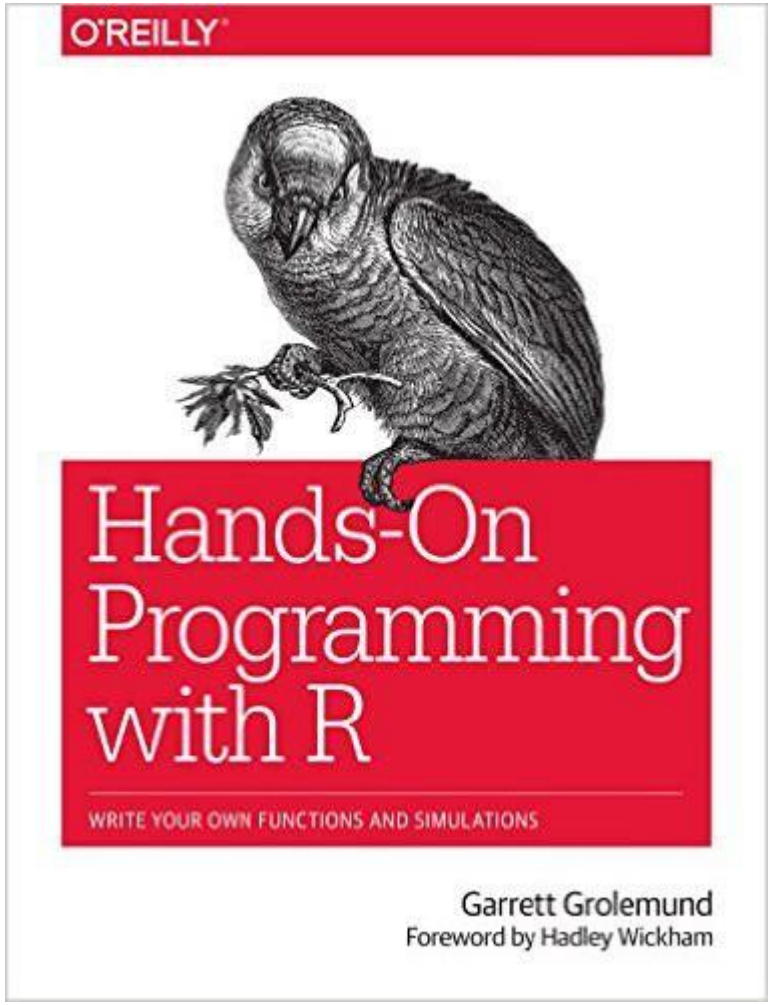
- Udemy
- Udacity
- DataCamp
- Coursera
- ...



Ladies BH

Livros

- Editora O’Reilly
- R Cookbook]
- R in a Nutshell
- Introduction to Data Science with R
- ...



Ladies BH

Pratique!

- Editora O’Reilly
- R Cookbook]
- R in a Nutshell
- Introduction to Data Science with R
- ...

Ladies BH

Why

Craig Dennis

Here's the thing: You don't really have a choice.

Obrigada! Dúvidas?

@ pollyannaogoncalves@gmail.com

in pollyannaogoncalves

Ladies BH

Estamos com vagas abertas!

https://www.hotmart.com/pt/trabalhe-conosco

Android/iOS Developer

Front-end Developer

Back-end Developer

...

Ladies BH