ME111 - Laboratório de Estatística

Aula 1 - Introdução

Profa. Larissa Avila Matos

Introdução: Estatísticas

*

Objetivo

- Apresentar o R e o RStudio;
  - R: 
  - RStudio: 
- Exportar/Carregar dados;
- Analisar dados;
- Inferir sobre os dados.

Instando o R e o RStudio

- Primeiro você deve baixar o arquivo de instalação correspondente ao seu sistema operacional: `http://cran.r-project.org`.
- Siga as instruções do site para finalizar a instalação do R.
- Agora você pode instalar o RStudio: `http://www.rstudio.com/products/rstudio/download/`.

R: interface

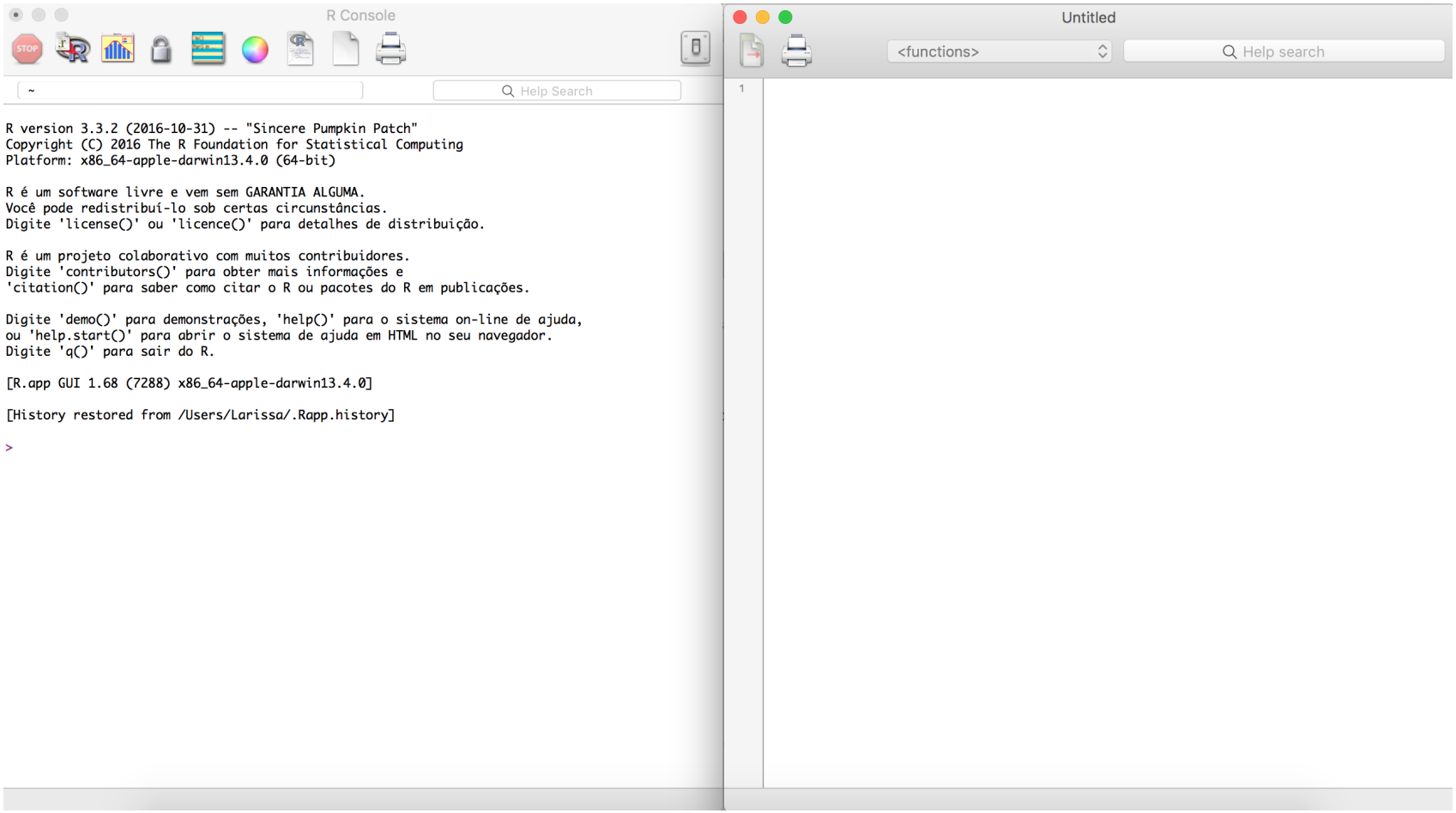


Figure 1: R

RStudio: interface

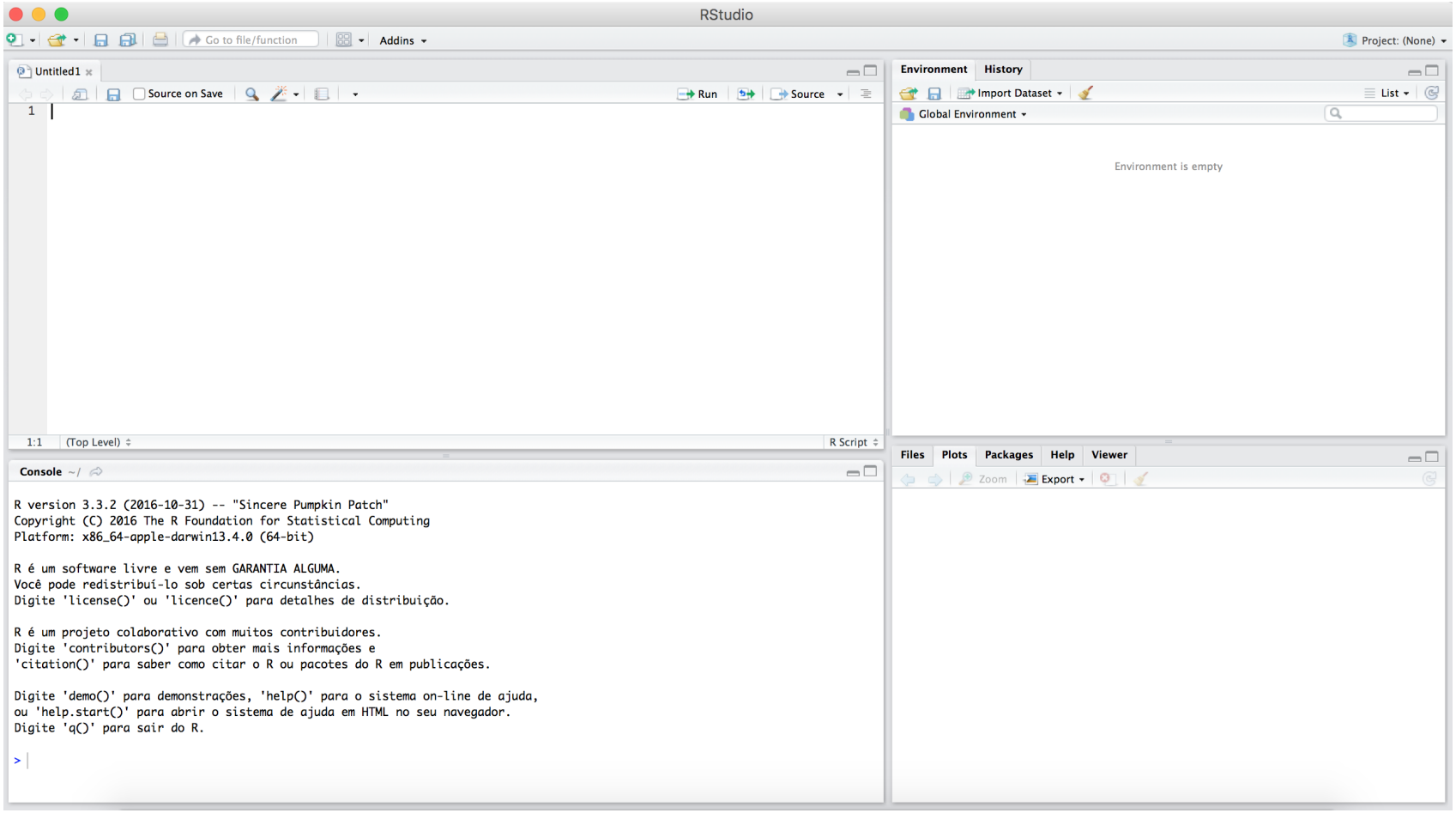


Figure 2: RStudio

Dados

- Podemos carregar os dados no R de diferentes formas.
- O formato mais adequado vai depender do tamanho do conjunto de dados, e se os dados já existem em outro formato para serem importados ou se serão digitados diretamente no R.
- As principais formas são:
  - Entrando com os dados diretamento no R;
  - Lendo os dados de um arquivo texto;
  - Importando os dados de outros programas/sites;
  - Carregando os dados já disponíveis no R;
  - Acessando planilhas de dados.

Entrando com os dados diretamento no R

1. Definindo vetores: Podemos entrar com os dados definindo vetores com o comando `c()`. Exemplos:

```
notasME111<-c(6.0,7.0,6.5,4.0,3.0,10.0,9.0,8.0)
notasME111
```

```
[1]  6.0  7.0  6.5  4.0  3.0 10.0  9.0  8.0
```

```
length(notasME111)
```

```
[1] 8
```

```
mean(notasME111)
```

```
[1] 6.6875
```

```
idade<-c(18,17,22,21,20,18,17,16);idade
```

```
[1] 18 17 22 21 20 18 17 16
```

- Corrigindo e/ou alterando dados

```
idade[3]<-23;idade
```

```
[1] 18 17 23 21 20 18 17 16
```

```
mean(idade)
```

```
[1] 18.75
```

```
idade/10
```

```
[1] 1.8 1.7 2.3 2.1 2.0 1.8 1.7 1.6
```

```
idade+10
```

```
[1] 28 27 33 31 30 28 27 26
```

```r
alunos<-cbind(notasME111,idade)
alunos
```

```
     notasME111 idade
[1,]        6.0    18
[2,]        7.0    17
[3,]        6.5    23
[4,]        4.0    21
[5,]        3.0    20
[6,]       10.0    18
[7,]        9.0    17
[8,]        8.0    16
```

```r
str(alunos)
```

```
 num [1:8, 1:2] 6 7 6.5 4 3 10 9 8 18 17 ...
 - attr(*, "dimnames")=List of 2
  ..$ : NULL
  ..$ : chr [1:2] "notasME111" "idade"
```

```r
dim(alunos)
```

```
[1] 8 2
```

```r
alunos[alunos[,1] >= 7,]
```

```
     notasME111 idade
[1,]          7    17
[2,]         10    18
[3,]          9    17
[4,]          8    16
```

```r
colMeans(alunos)
```

```
notasME111      idade
    6.6875    18.7500
```

Banco de Dados do R

- Para carregar conjuntos de dados que são já disponibilizados com o R use o comando `data()`. Exemplo:

```
data()
data(package="nlme")
data(package="MASS")
```

1. Banco de dados 'cars':

Esse conjunto de dados mostra a velocidade dos carros e as distâncias usadas para parar. Esses dados foram registrados na década de 1920 e contém 50 observações com 2 variáveis (velocidade e distância).

```
data(cars)
?cars
cars$dist
attach(cars)
speed
```

```r
summary(cars)
```

```
     speed           dist       
 Min.   : 4.0   Min.   :  2.00  
 1st Qu.:12.0   1st Qu.: 26.00  
 Median :15.0   Median : 36.00  
 Mean   :15.4   Mean   : 42.98  
 3rd Qu.:19.0   3rd Qu.: 56.00  
 Max.   :25.0   Max.   :120.00  
```

2. Banco de dados 'Iris":

Os dados consistem de 50 unidades amostrais de três espéecies (setosa, virginica, versicolor) de íris (uma espécie de planta), ou seja, temos um total de 150 unidades amostrais (obeservações). De cada unidade mediu-se quatro variáveis:

- comprimento e largura da sépala; e
- comprimento e largura da pétala.

```r
data(iris) # dados iris
dim(iris) # dimensão dos dados
```

```
[1] 150   5
```

```r
names(iris) # nome das variáveis
```

```
[1] "Sepal.Length" "Sepal.Width"  "Petal.Length" "Petal.Width"
[5] "Species"
```

```r
str(iris) # estrutura dos dados
```

```
'data.frame':   150 obs. of  5 variables:
 $ Sepal.Length: num  5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
 $ Sepal.Width : num  3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
 $ Petal.Length: num  1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
 $ Petal.Width : num  0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
 $ Species     : Factor w/ 3 levels "setosa","versicolor",..: 1 1 1 1 1
```

```
iris[1:5,]
```

```
  Sepal.Length Sepal.Width Petal.Length Petal.Width Species
1          5.1         3.5          1.4         0.2  setosa
2          4.9         3.0          1.4         0.2  setosa
3          4.7         3.2          1.3         0.2  setosa
4          4.6         3.1          1.5         0.2  setosa
5          5.0         3.6          1.4         0.2  setosa
```

```
head(iris,5)
```

```
  Sepal.Length Sepal.Width Petal.Length Petal.Width Species
1          5.1         3.5          1.4         0.2  setosa
2          4.9         3.0          1.4         0.2  setosa
3          4.7         3.2          1.3         0.2  setosa
4          4.6         3.1          1.5         0.2  setosa
5          5.0         3.6          1.4         0.2  setosa
```

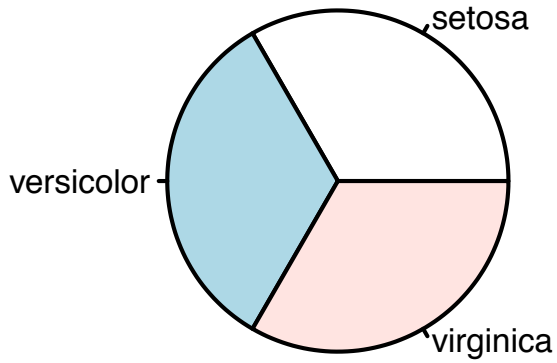
```r
iris[1:10, "Sepal.Length"]
```

```
 [1] 5.1 4.9 4.7 4.6 5.0 5.4 4.6 5.0 4.4 4.9
```

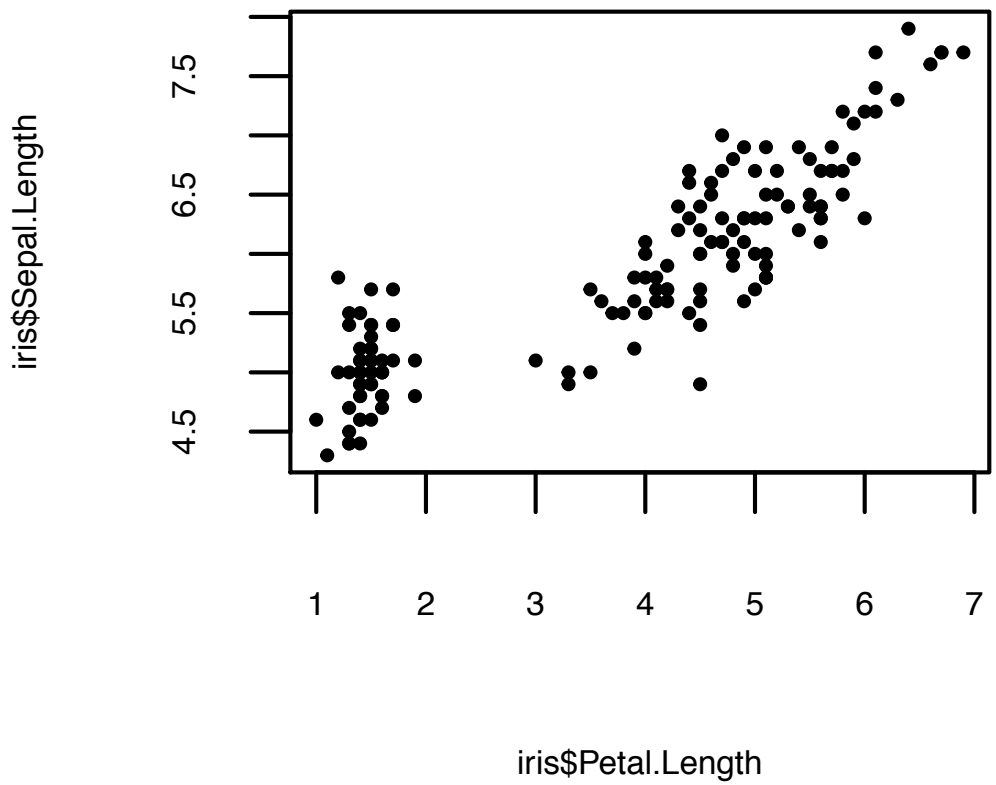
```r
table(iris$Species) # frequência
```

```
    setosa versicolor  virginica
        50         50         50
```

```r
pie(table(iris$Species),cex=0.5)
```

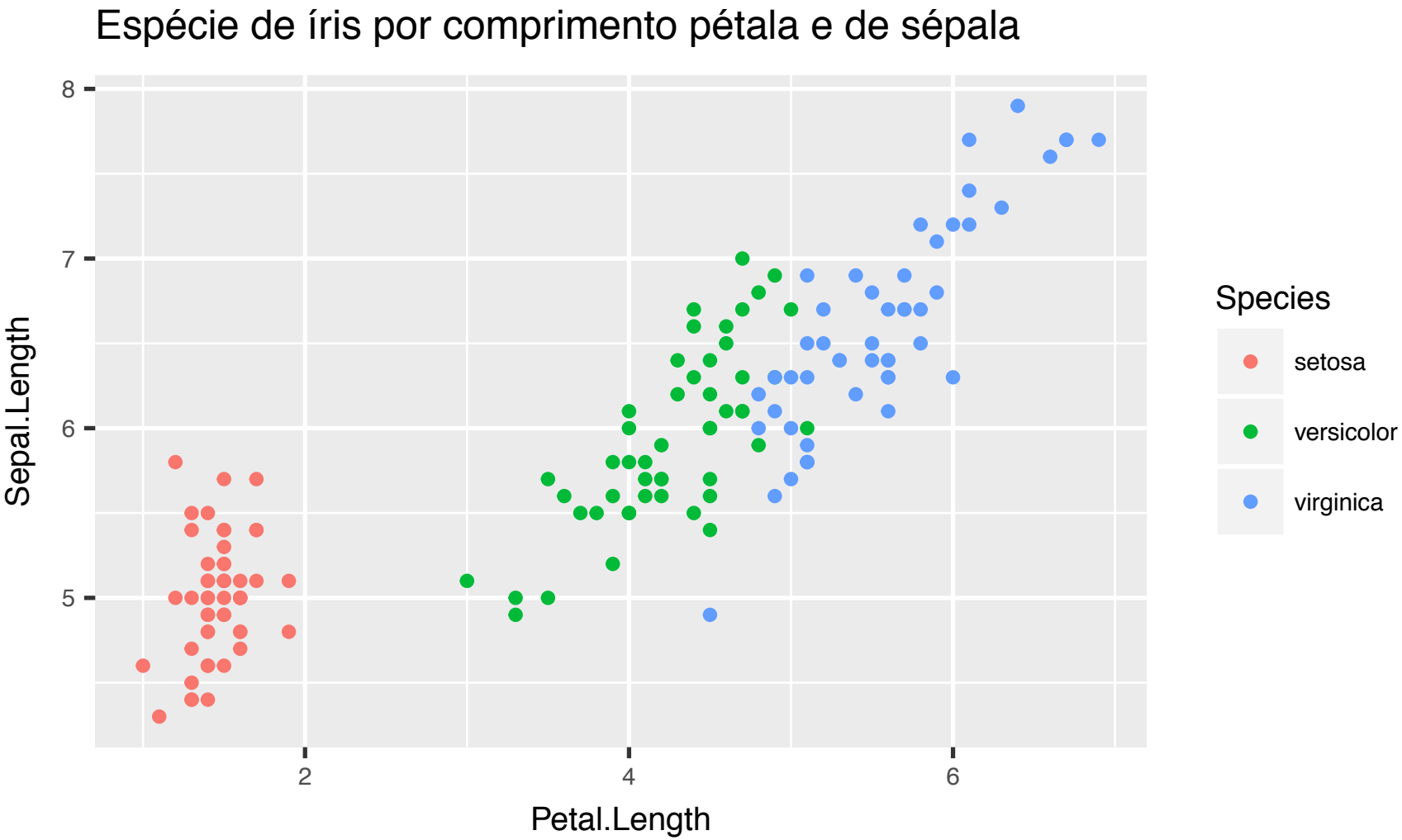


```r
plot(iris$Petal.Length, iris$Sepal.Length,
     cex=0.5,cex.axis=0.5,cex.lab=0.5,pch=20) # scatter plot
```




```r
cor(iris$Sepal.Length, iris$Petal.Length) # correlação
```

```
[1] 0.8717538
```


Espécie de íris por comprimento pétala e de sépala
Sepal.Length
8
7
6
5
2
4
6
Petal.Length
Species
setosa
versicolor
virginica

Lendo os dados de um arquivo texto

- Se os dados já estão disponíveis em formato eletrônico, isto é, já foram digitados em outro programa, você pode importar os dados para o R sem a necessidade de digitá-los novamente, basta salva-los em um arquivo em formato texto. Uma forma de ler os dados em formato texto é usanndo a função `read.table()`. Exemplo:
  1. Copie os dados cars em um arquivo de texto,
  2. Salve o arquivo na sua área de trabalho (working directory do R).
  3. Para importar este arquivo usamos:

```
setwd("/Users/Larissa/Desktop/1S2018/ME_111/Aulas/Aula1")
cars1 <- read.table("cars.txt",header=T)
```

Importação de dados

- Além disto, é comum surgir a necessidade de importar dados de planilhas eletrônicas. Para exportar de sites podemos usar o comando `source()`
- Sites:

```
source("http://www.openintro.org/stat/data/arbuthnot.R")
arbuthnot[1:3,]
```

```
  year boys girls
1 1629 5218  4683
2 1630 4858  4457
3 1631 4422  4102
```

- O conjunto de dados Arbuthnot se refere ao Dr. John Arbuthnot, um médico, escritor e matemático do século 18. Para investigar a razão de meninos e meninas recém-nascidos ele coletou os registros de batismo de crianças nascidas em Londres entre os anos de 1629 e 1710. **Obs.: Dados retirados do site da OpenIntro.**

```r
data(package="HistData")
library("HistData")
data(Arbuthnot)
Arbuthnot[1:8,]
```

	Year	Males	Females	Plague	Mortality	Ratio	Total
1	1629	5218	4683	0	8771	1.114243	9.901
2	1630	4858	4457	1317	10554	1.089971	9.315
3	1631	4422	4102	274	8562	1.078011	8.524
4	1632	4994	4590	8	9535	1.088017	9.584
5	1633	5158	4839	0	8393	1.065923	9.997
6	1634	5035	4820	1	10400	1.044606	9.855
7	1635	5106	4928	0	10651	1.036120	10.034
8	1636	4917	4605	10400	23359	1.067752	9.522

Importando dados de outros programas

- É possível também ler dados diretamente de outros formatos que não seja texto (ASCII). Isto em geral é mais eficiente e requer menos memória do que converter para formato texto. Há funções para importar dados diretamente de EpiInfo, Minitab, S-PLUS, SAS, SPSS, Stata, Systat e Octave. Muitas funções que permitem a importação de dados de outros programas são implementadas no pacote **foreign**.

```
require(foreign)
```

- Para saber mais a respeito destas funções ver

`https://cran.r-project.org/web/packages/foreign/foreign.pdf`

Acessando planilhas de dados

- É comum que dados estajam armazenados em planilhas eletrônicas tais como Excel ou OpenOffice. Neste caso, embora seja possível exportar a partir destes aplicativos os dados para o formato texto para depois serem lidos no R, possivelmente com `read.table()`, pode ser necessário ou conveniente ler os dados diretamente destes formato.

- Em casos em que a estrutura da planilha é simples e sem macros ou fórmulas, pode-se usar o comando
  - `read.xls ()`

```
library("gdata")
cars1 = read.xls ("/Users/Larissa/Desktop/1S2018/ME_111/Aulas
                  /Aula1/cars.xls",
                  sheet = 1   , header = TRUE)
```

- Ou podemos usar a função `xls2cv`, onde podemos usar os comandos
  - `read.csv()`;
  - `read.csv2()`.

- O comando `read.csv2()` é recomendado para dados com vírgula como caractere separados de decimais.

Comandos básicos

- Entrada e Saída:

`q`, `load`, `help.search`, `ls`, `dump`, `library`, `rm`, `source`, `search`, `save`, `history`, `save.image`, `help`

- Manipular objetos:

`c`, `apply/tapply/sapply`, `rep`, `cbind`, `sweep`, `which`, `rbind`, `sort`, `table`, `names`, `seq`

- Tipos de objeto – podemos usar `is.xx()` e `as.xx()`:

`matrix`, `factor`, `logical`, `numeric`, `character`, `logical`

- Indexando: Seja `x` um vetor numérico, `b` uma matriz ou um data frame, então
  - `x[i] = i-ésimo elemento do vetor`,
  - `x[1: n] = vetor com os elemento de 1 a n de x`,
  - `x[c(2,3,5,6,11)] = vetor com elementos 2,3,5,6,11 de x`,
  - `b[i,j] =`
  - `b[i,] =`
  - `b[,j] =`
  - `b$colname =`

Funções Básicas

- Aritimética, Lógica, Matemática:

`+`, `-`, `*`, `/`, `^`, `==`, `!=`, `&`, `|`, `<`, `>`, `<=`, `>=`, `is.na`, `log`, `log10`, `exp`, `sin`, `cos`, `tan`, `asin`, `acos`, `atan`, `t`, `crossprod`, `!`

- Funções resumo:

`sum`, `mean`, `var`, `sd`, `median`, `range`, `min`, `max`, `cor`, `summary`, `quantile`