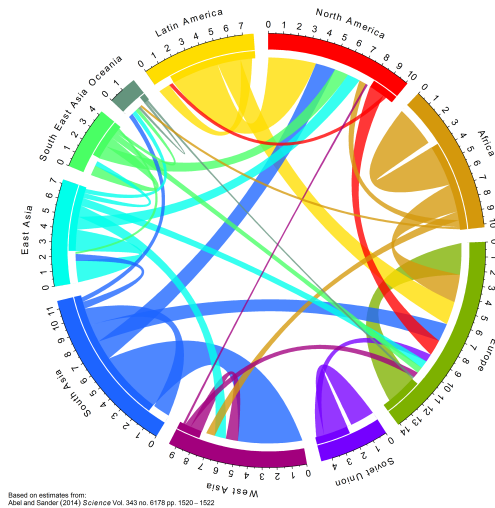# VISUALIZATION
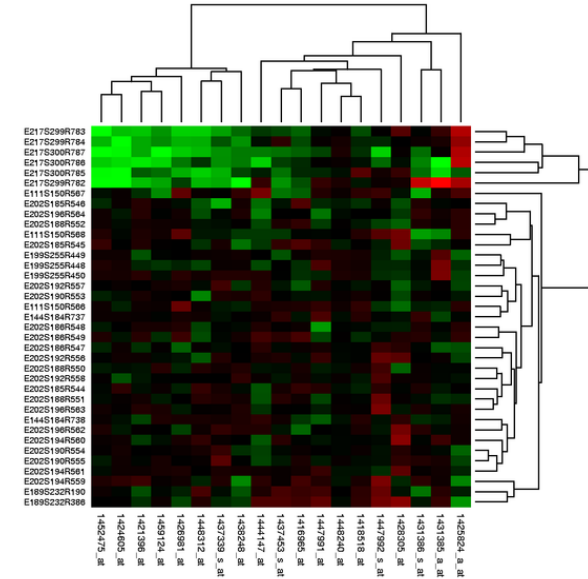
Introducción a la Ciencia de Datos

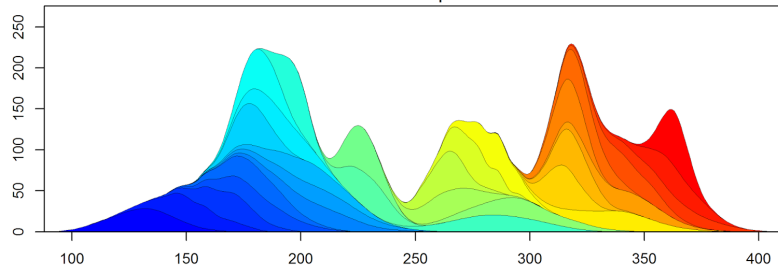# Gallery of plots
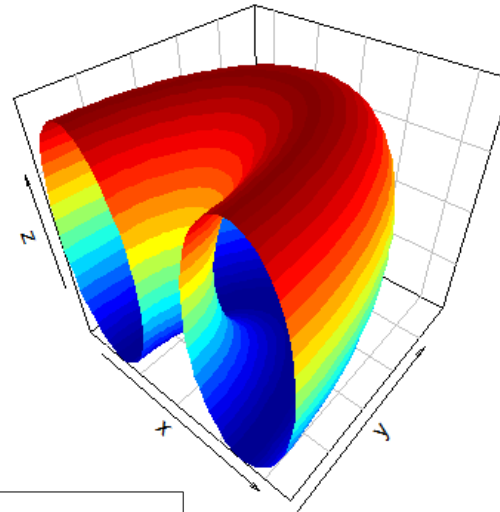
R plots
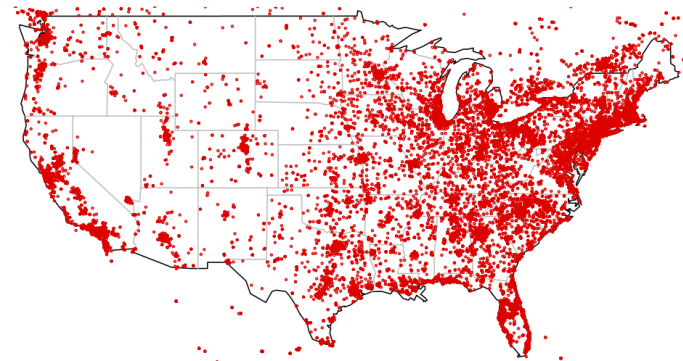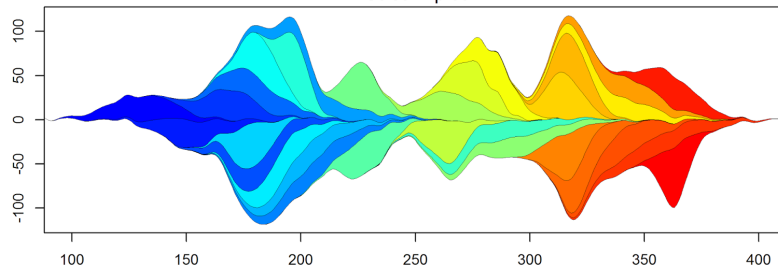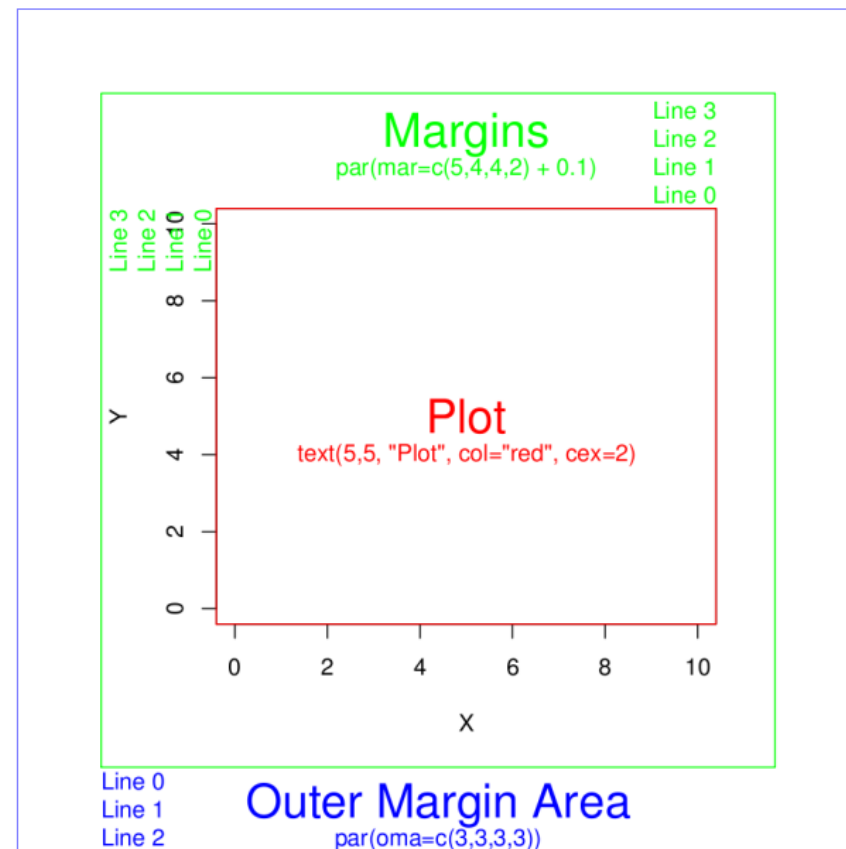
- We will see a small set of them:
  - Plots
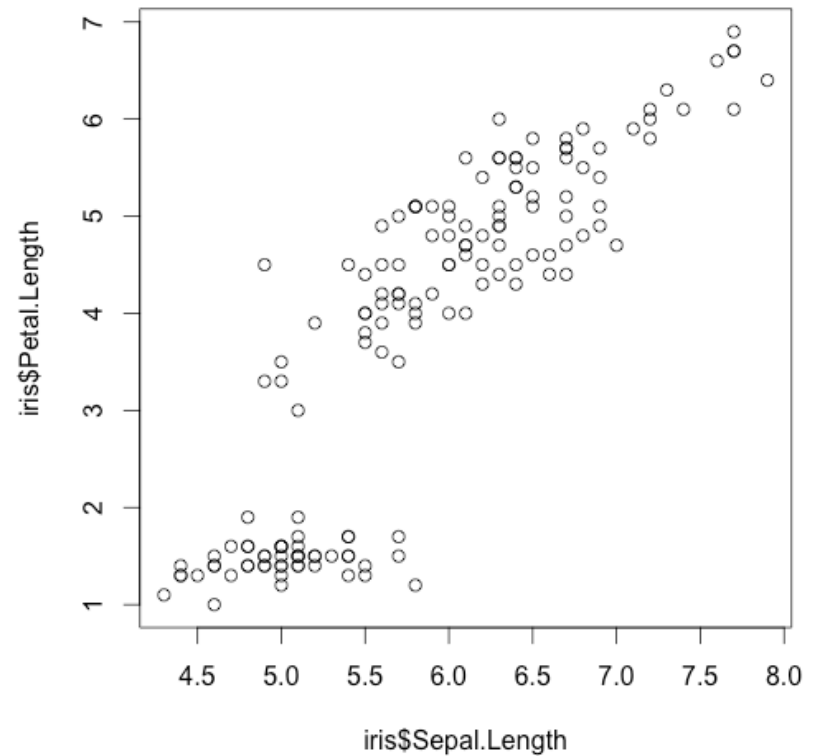 - Histograms
  - Boxplots
  - Barplots
  - Pies

Plot

- To change parameters for margins and plot areas use the `par` function.
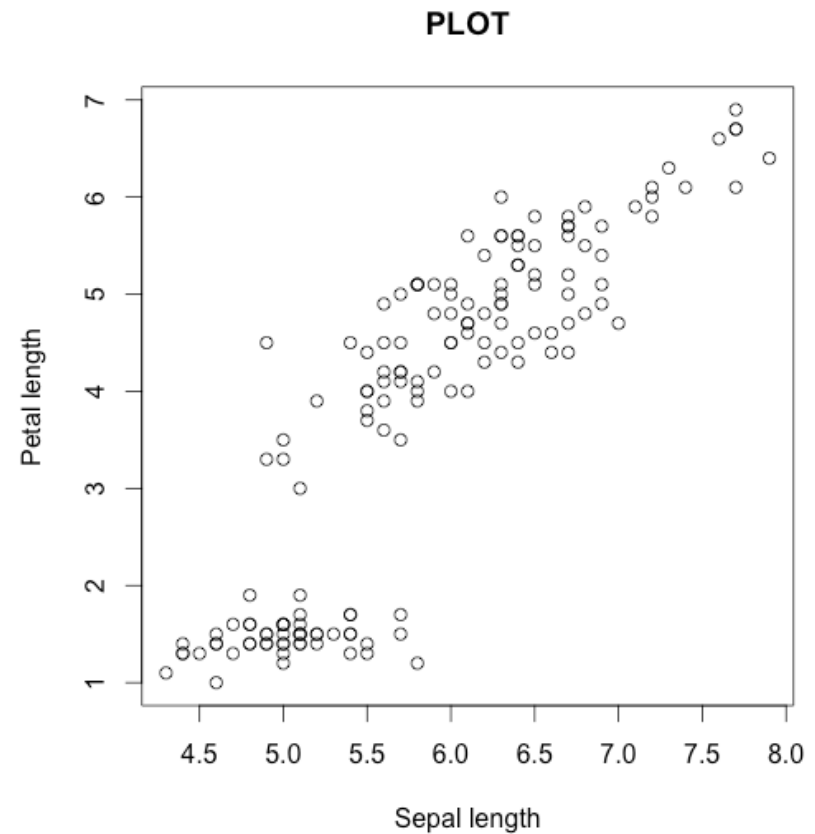
- Link to R manual

Plot

- `?iris`
- `iris`
- `class(iris)`
- `names(iris)`
- `?plot`
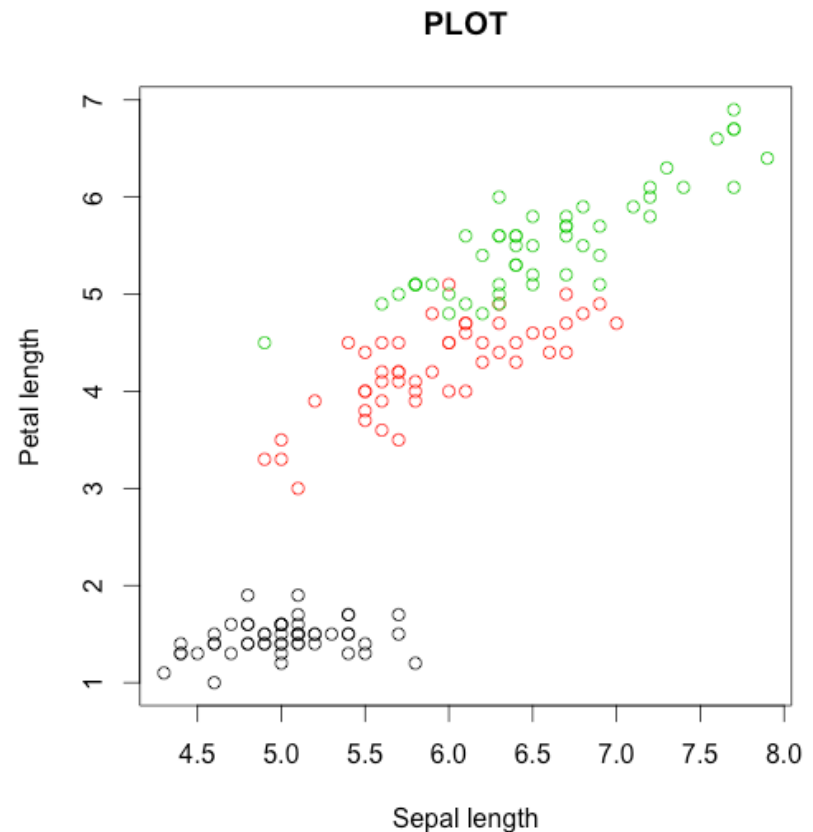- `example(plot)`
- `plot(iris$Sepal.Length, iris$Petal.Length)`

Plot

- ```
  plot(iris$Sepal.Length,
  iris$Petal.Length,
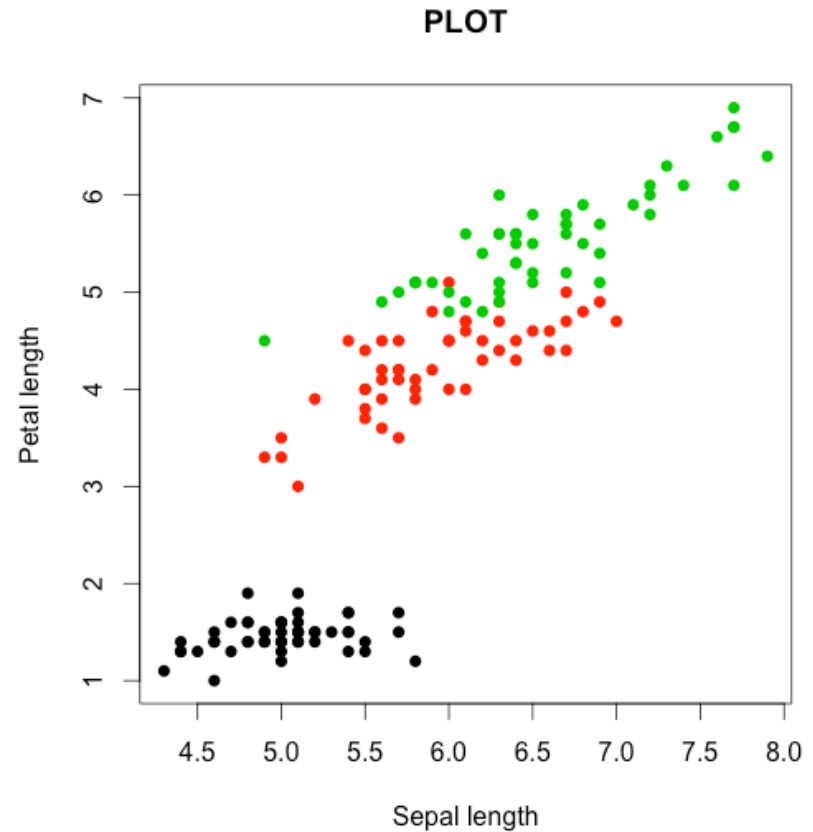 main="PLOT", xlab="Sepal
  length", ylab="Petal
  length")
  ```

# Plot

- ```
  plot(iris$Sepal.Length,
  iris$Petal.Length,
  col=iris$Species,
  main="PLOT", xlab="Sepal
  length", ylab="Petal
  length")
  ```

Plot

- ```
  plot(iris$Sepal.Length,
  pch=16,
  iris$Petal.Length,
  col=iris$Species,
  main="PLOT", xlab="Sepal
  length", ylab="Petal
  length")
  ```

Plot

pch

0: □	10: ⊕	20: ●	A: **A**
1: ○	11: ⚹	21: ●	a: **a**
2: △	12: ⊞	22: ■	B: **B**
3: +	13: ⊗	23: ◆	b: **b**
4: ×	14: ⊠	24: ▲	S: **S**
5: ◇	15: ■	25: ▼	`: `
6: ▽	16: ●	@: **@**	.: ·
7: ⊠	17: ▲	+: **+**	,: ,
8: ✳	18: ◆	%: **%**	?: **?**
9: ⊕	19: ●	#: **#**	*: ∗

Plot

- ```
  plot(iris$Sepal.Length,
  pch=16,
  iris$Petal.Length,
  col=iris$Species,
  main="PLOT", xlab="Sepal
  length", ylab="Petal
  length")
  ```
- `?abline`
- `abline(h=2.5)`

Plot

- ```
  plot(iris$Sepal.Length,
  pch=16,
  iris$Petal.Length,
  col=iris$Species,
  main="PLOT", xlab="Sepal
  length", ylab="Petal
  length")
  ```
- ```
  abline(6.5, -0.8, lty=2)
  ```

Plot

pch

0: □	10: ⊕	20: ●	A: **A**
1: ○	11: ⚹	21: ●	a: **a**
2: △	12: ⊞	22: ■	B: **B**
3: +	13: ⊗	23: ◆	b: **b**
4: ×	14: ⊠	24: ▲	S: **S**
5: ◇	15: ■	25: ▼	`: `
6: ▽	16: ●	@: @	.: .
7: ⊠	17: ▲	+: +	,: ,
8: ✳	18: ◆	%: %	?: ?
9: ⊕	19: ●	#: #	*: *

lty

0. 'blank'	
1. 'solid'	————————————
2. 'dashed'	– – – – – – – – –
3. 'dotted'	·············
4. 'dotdash'	–·–·–·–·–·–
5. 'longdash'	— — — — — —
6. 'twodash'	–––·–––·–––

Plot

- ```
  plot(iris$Sepal.Length,
  pch=16,
  iris$Petal.Length,
  col=iris$Species,
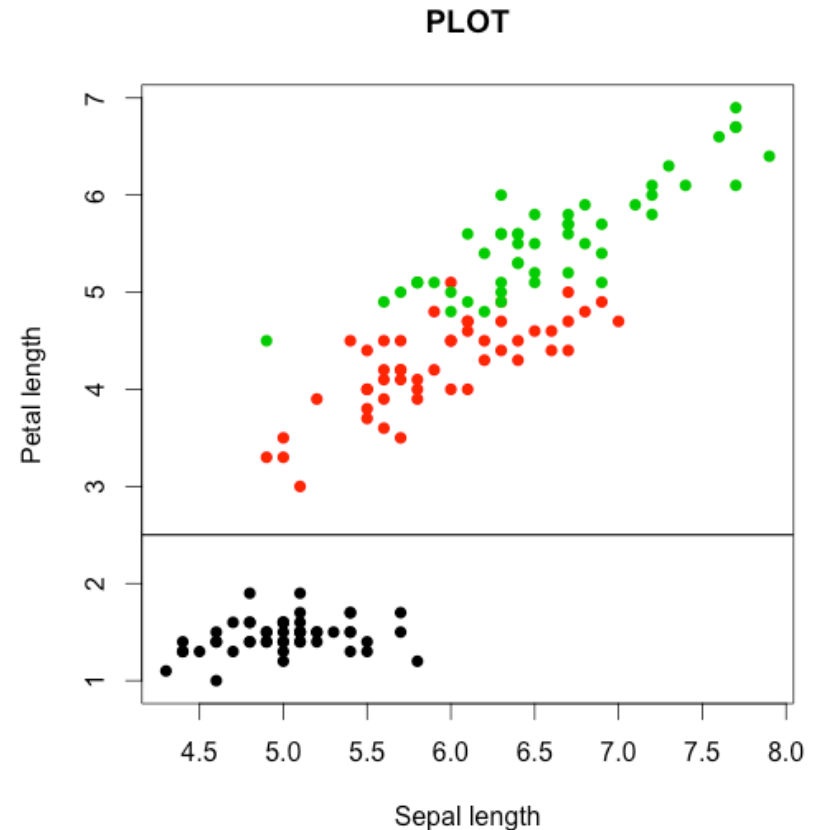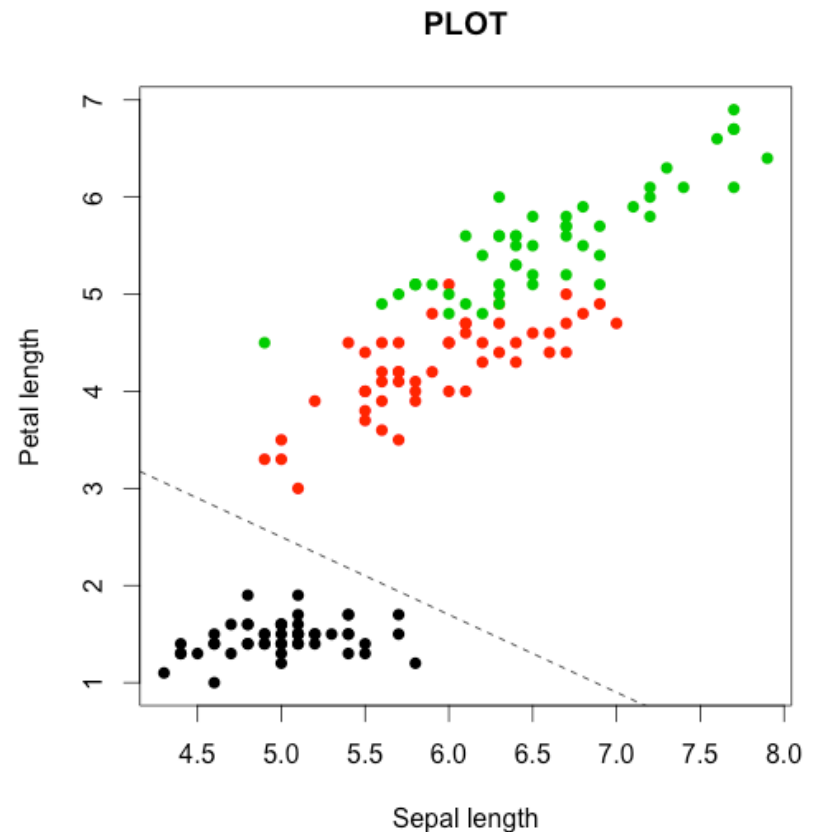 main="PLOT", xlab="Sepal
  length", ylab="Petal
  length")
  ```

- ```
  legend("topleft",
  legend=levels(iris$Speci
  es),
  col=1:length(levels(iris
  $Species)), pch=16)
  ```
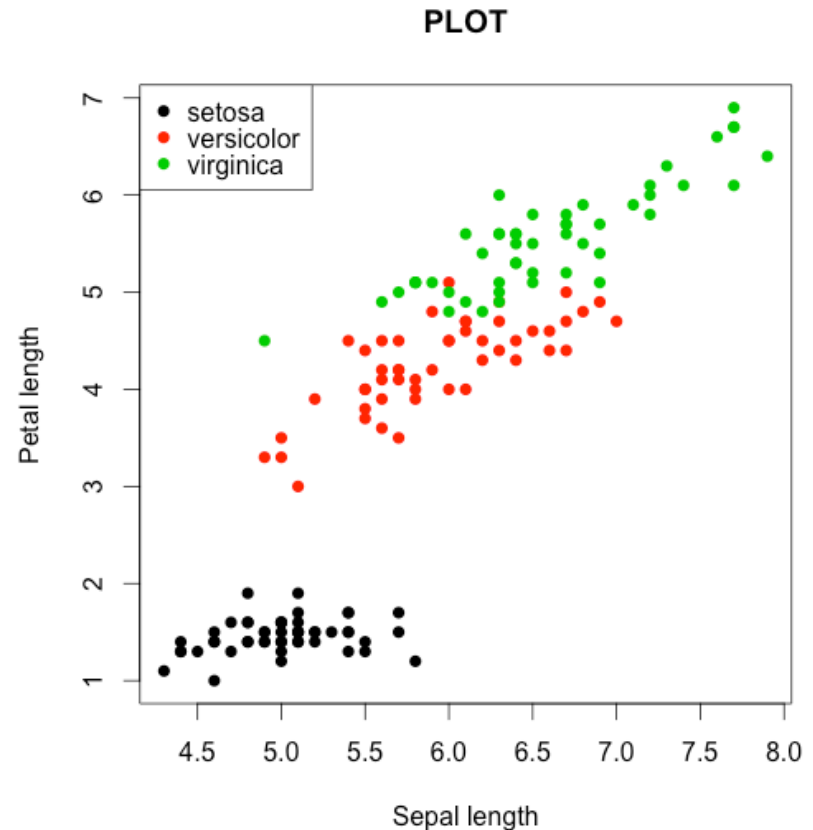
Curves

- `plot(0:10, sin(0:10), type="l")`
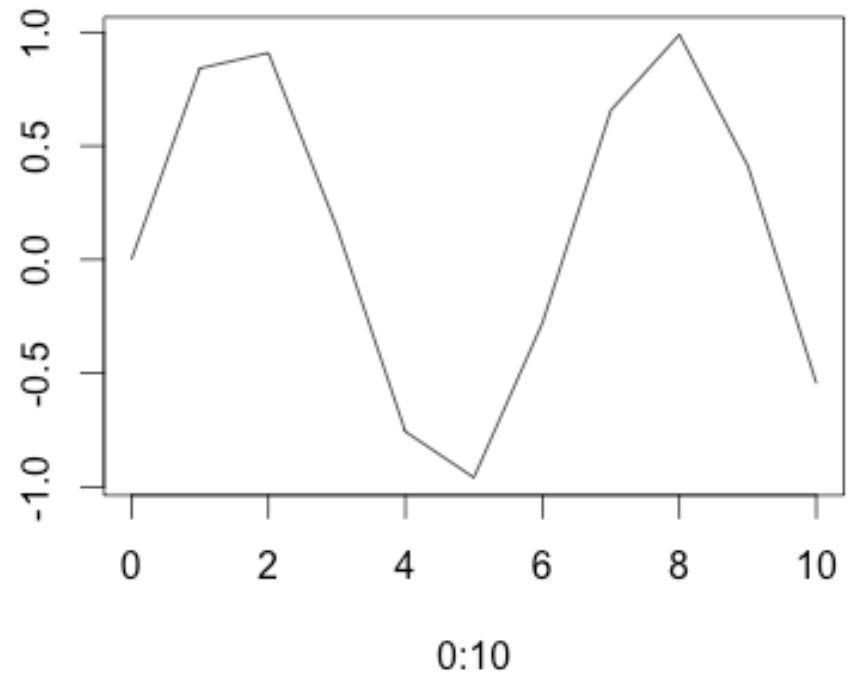
Curve

- `curve(sin(x), 0, 10, n=100)`
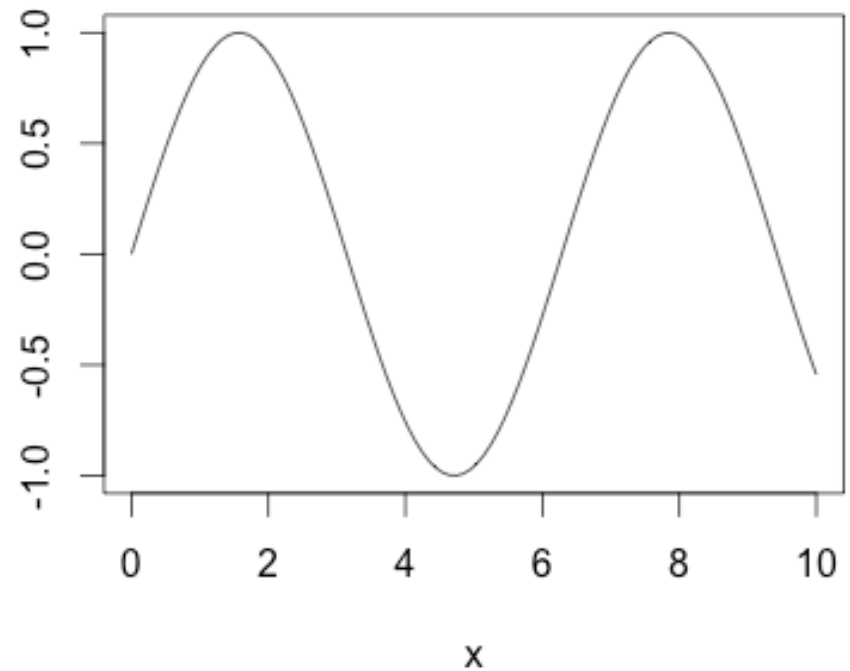
Piecewise functions

- ```
  myfun <- function(x)
  {ifelse(x > 0,7,-2)}
  ```
- ```
  curve(myfun(x), -5, 5)
  ```
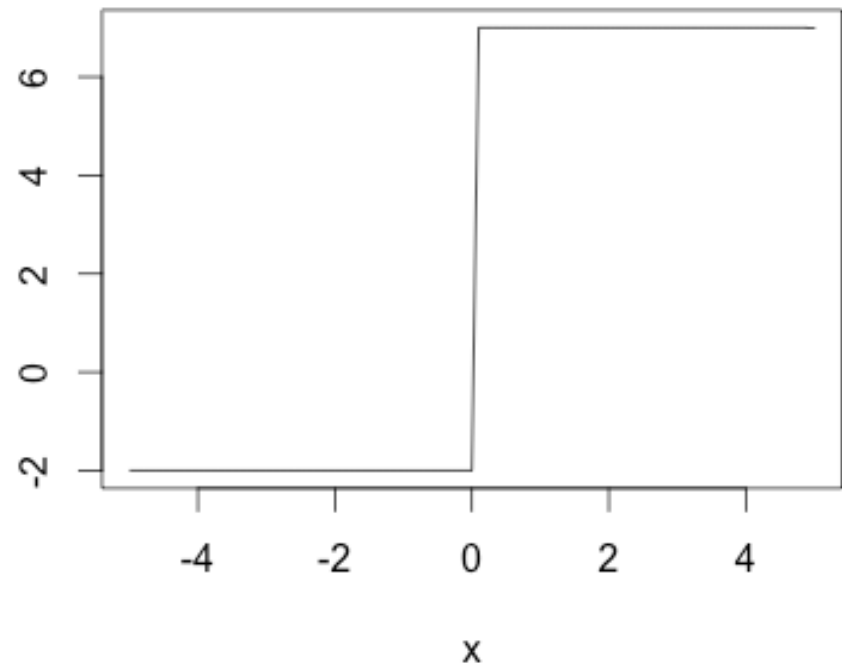
Scatterplot

- `plot(iris)`
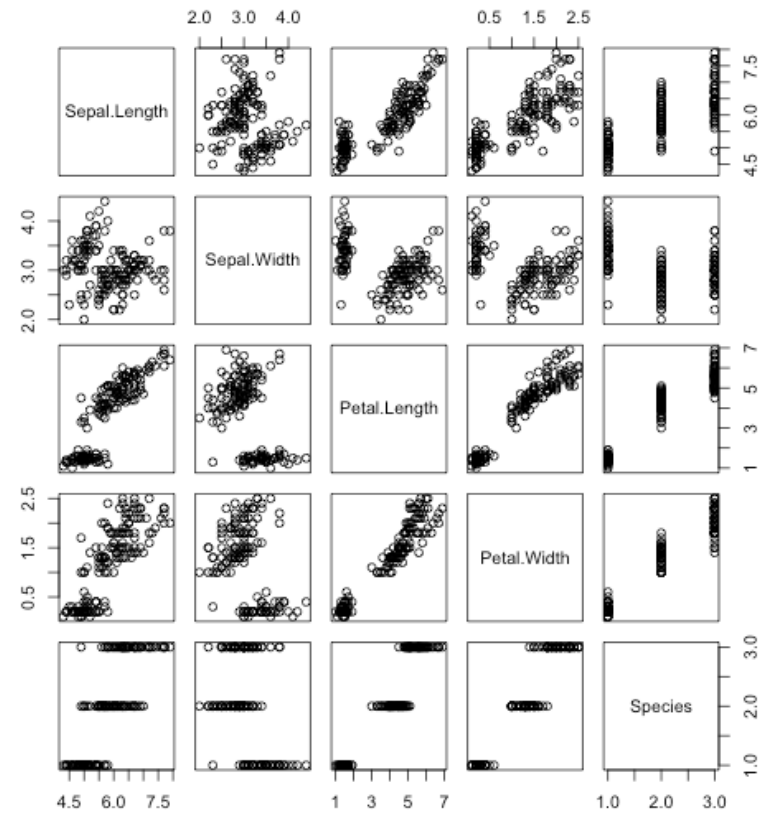
Scatterplot

- ```
  plot(iris[,-5],
  col=iris$Species)
  ```
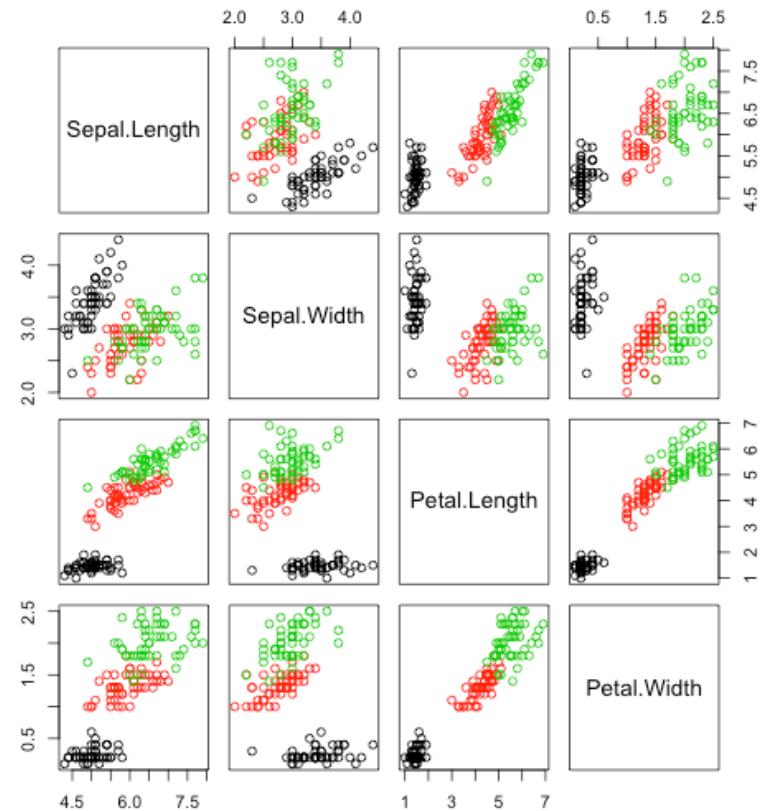
Histogram

- `?hist`
- `hist(iris$Petal.Length,`
  `main="Histogram",`
  `xlab="Petal length")`
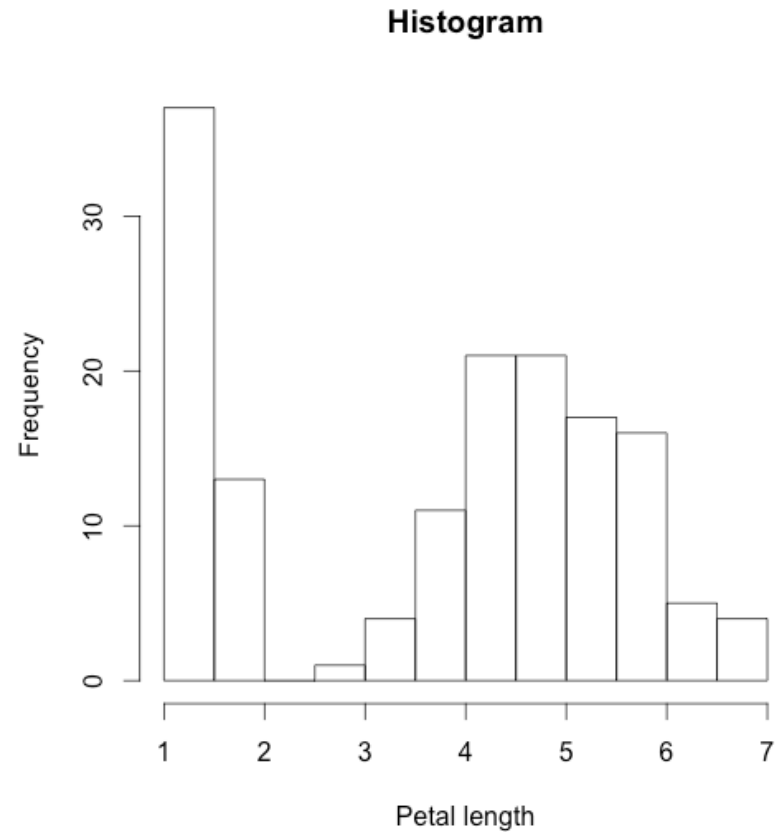
Histogram

- ```
  hist(iris$Petal.Length,
  main="Histogram",
  xlab="Petal length",
  breaks=5)
  ```
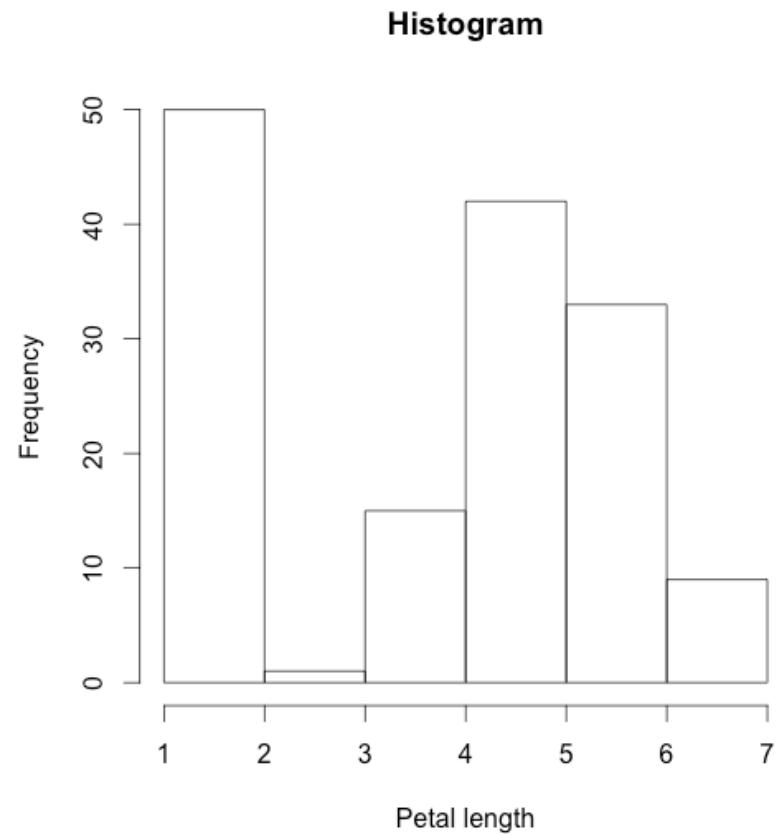
Histogram

- ```
  minimo <-
  min(iris$Petal.Length,ir
  is$Sepal.Length)
  ```
- ```
  maximo <-
  max(iris$Petal.Length,ir
  is$Sepal.Length)
  ```
- ```
  hist(iris$Petal.Length,
  main="Histogram",
  xlab="Variable",
  col="lightblue",
  xlim=c(minimo,maximo))
  ```
- ```
  hist(iris$Sepal.Length,
  main="Histogram",
  xlab="Variable",
  col="pink", add=TRUE)
  ```
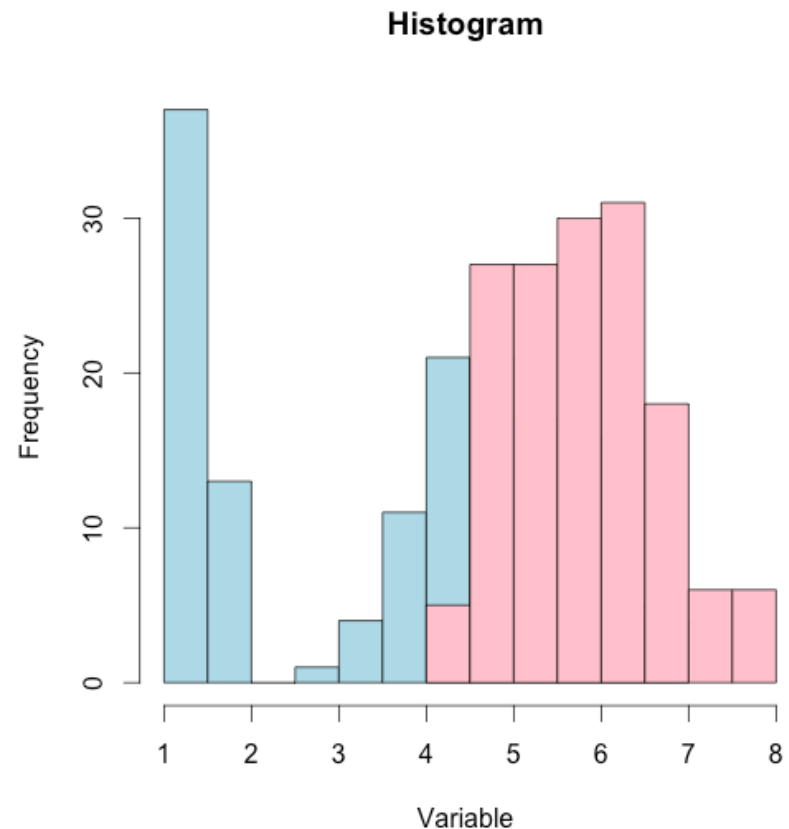
Histogram

Histogram

- ```
  hist(iris$Petal.Length,
  main="Histogram",
  xlab="Variable",
  col=rgb(1,0,0,0.5),
  xlim=c(minimo,maximo))
  ```
- ```
  hist(iris$Sepal.Length,
  main="Histogram",
  xlab="Variable",
  col=rgb(0,0,1,0.5),
  add=TRUE)
  ```

Boxplot

Atípico $\longrightarrow$ o

$\max(x)\big|_{x \le Ls}$

$Ls = Q3 + RIC \cdot 1.5$

Q3 (75%)

Mediana (50%)

$RIC = Q3 - Q1$
(50% de datos)

Q1 (25%)

$Li = Q1 - RIC \cdot 1.5$

$\min(x)\big|_{x \ge Li}$
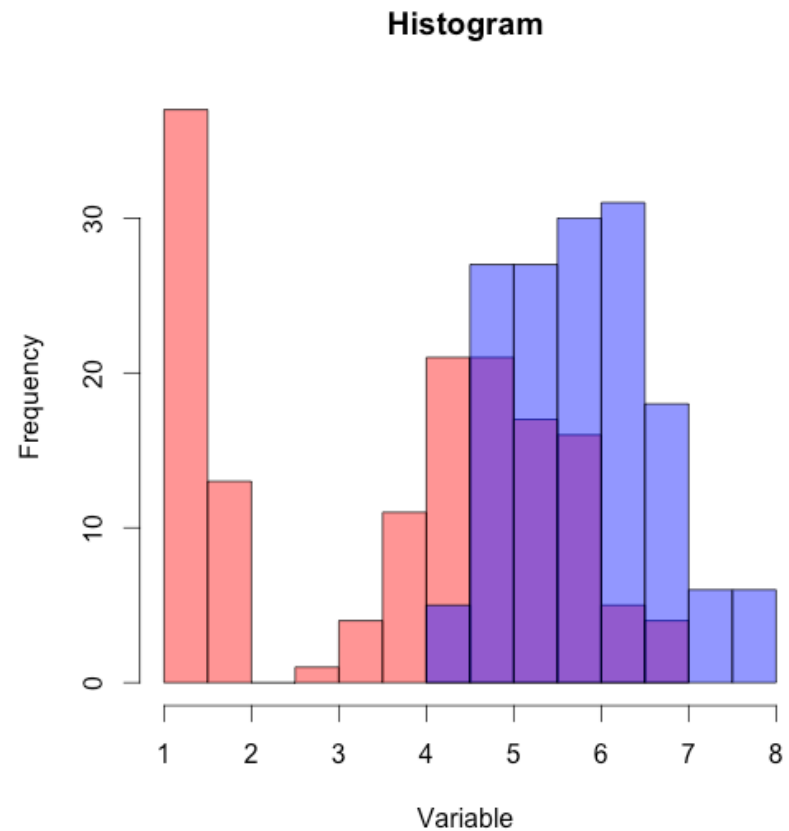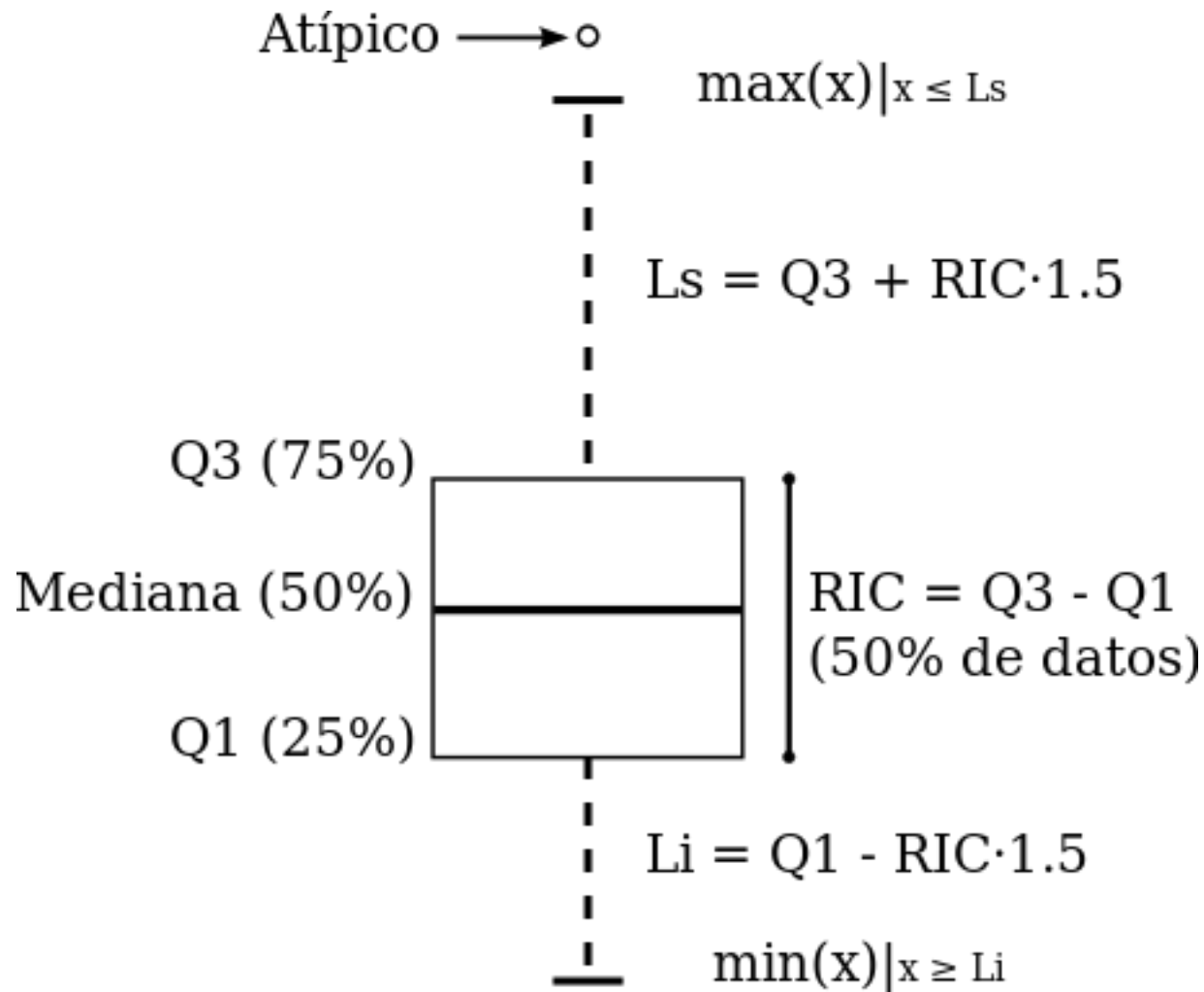
Boxplot

- ```
  data<-
  data.frame(Stat11=rnorm(100,mean=3,
  sd=2),
  Stat21=rnorm(100,mean=4,sd=1),
  Stat31=rnorm(100,mean=6,sd=0.5),
  Stat41=rnorm(100,mean=10,sd=0.5),
  Stat12=rnorm(100,mean=4,sd=2),
  Stat22=rnorm(100,mean=4.5,sd=2),
  Stat32=rnorm(100,mean=7,sd=0.5),
  Stat42=rnorm(100,mean=8,sd=3),
  Stat13=rnorm(100,mean=6,sd=0.5),
  Stat23=rnorm(100,mean=5,sd=3),
  Stat33=rnorm(100,mean=8,sd=0.2),
  Stat43=rnorm(100,mean=4,sd=4))
  ```

- `summary(data)`
- `boxplot(data)`
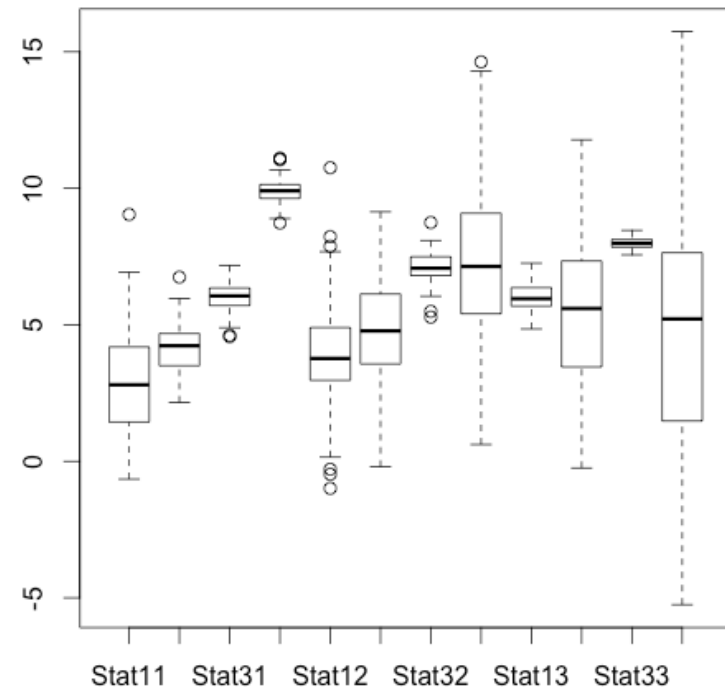
Boxplot

- `boxplot(data, las = 2)`
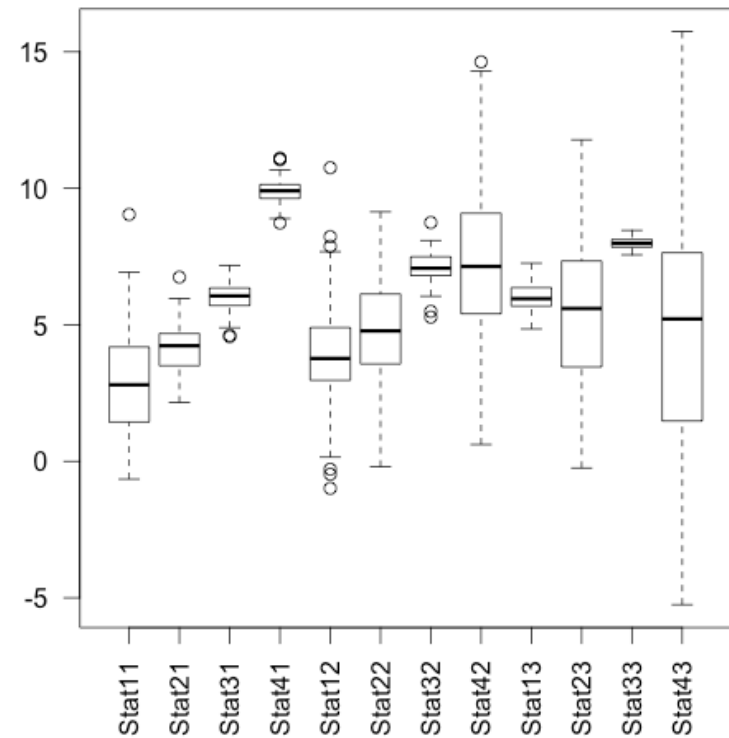
Boxplot

- ```
  boxplot(data, las = 2,
  names = c("Station
  1","Station 2","Station
  3","Station 4","Station
  1","Station 2","Station
  3","Station 4","Station
  1","Station 2","Station
  3","Station 4"))
  ```
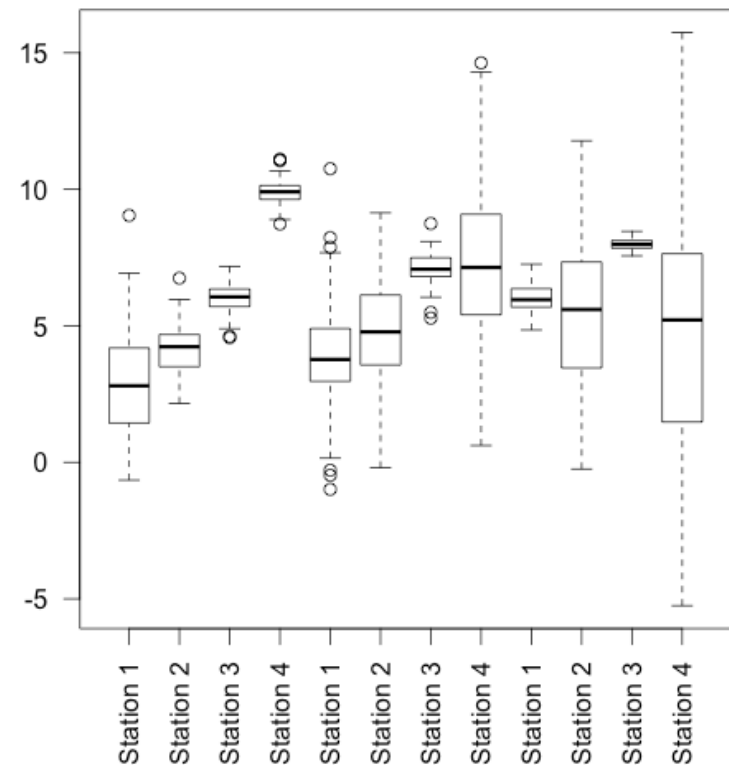
Boxplot

- ```
  boxplot(data, las = 2,
  at =
  c(1,2,3,4,6,7,8,9,11,12,
  13,14))
  ```
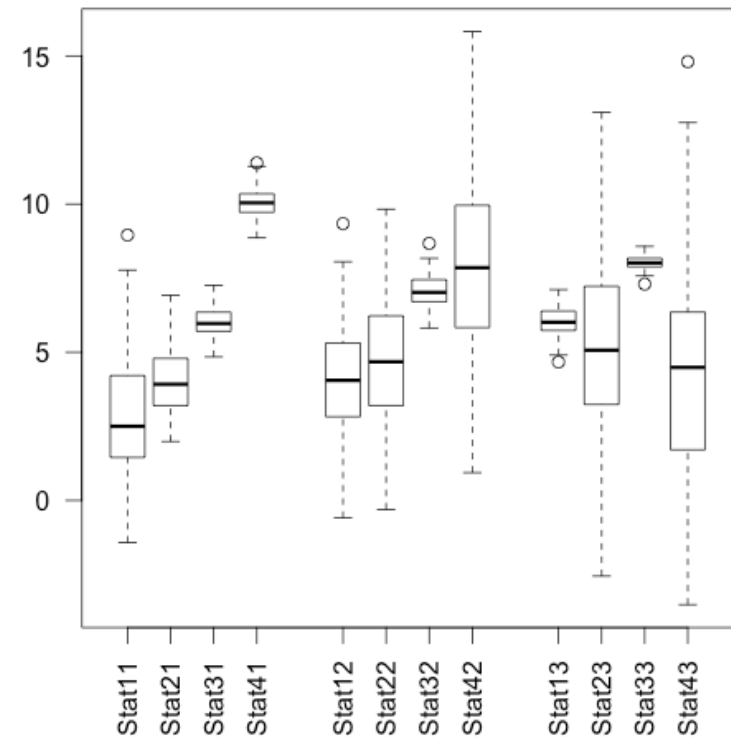
Boxplot

- ```
  boxplot(data, col =
  c("red","sienna","palevi
  oletred1","royalblue2","
  red","sienna","paleviole
  tred1","royalblue2","red
  ","sienna","palevioletre
  d1","royalblue2"))
  ```

- ```
  colors()
  ```
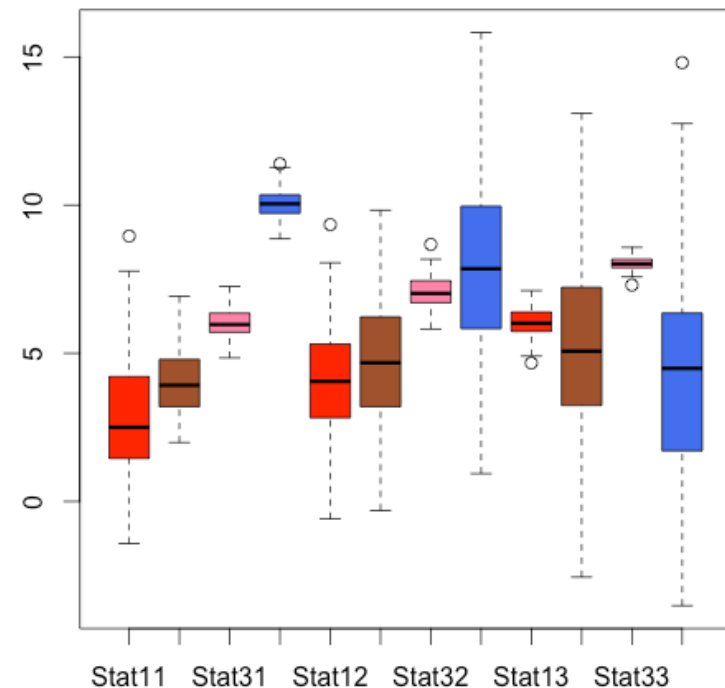
Barplot

- ```
  counts <-
  table(mtcars$gear)
  ```
- ```
  barplot(counts,
  main="Car Distribution",
  xlab="Number of Gears")
  ```

Car Distribution

Number of Gears

Barplot

- ```
barplot(counts,
main="Car Distribution",
horiz=TRUE,
names.arg=c("3 Gears",
"4 Gears", "5 Gears"))
```
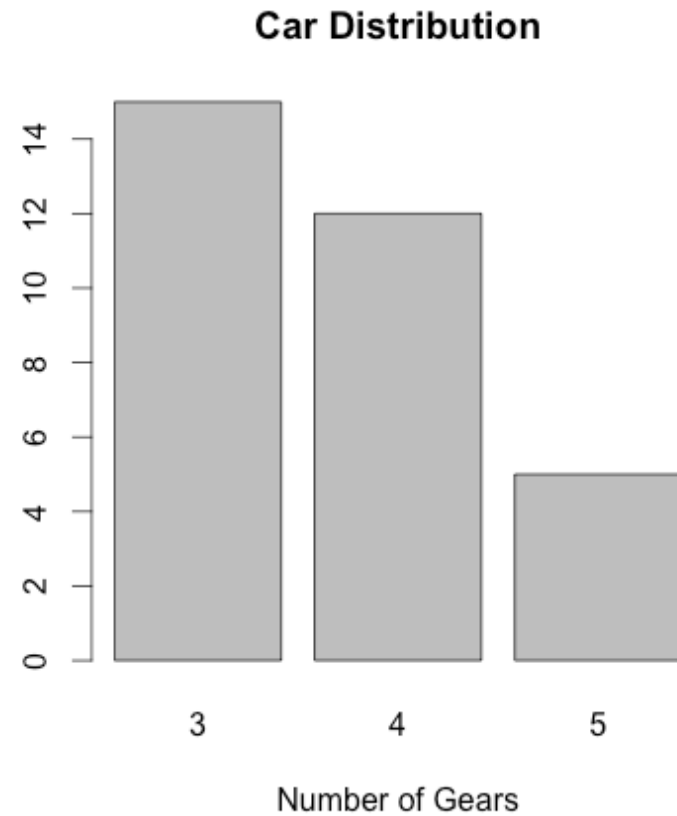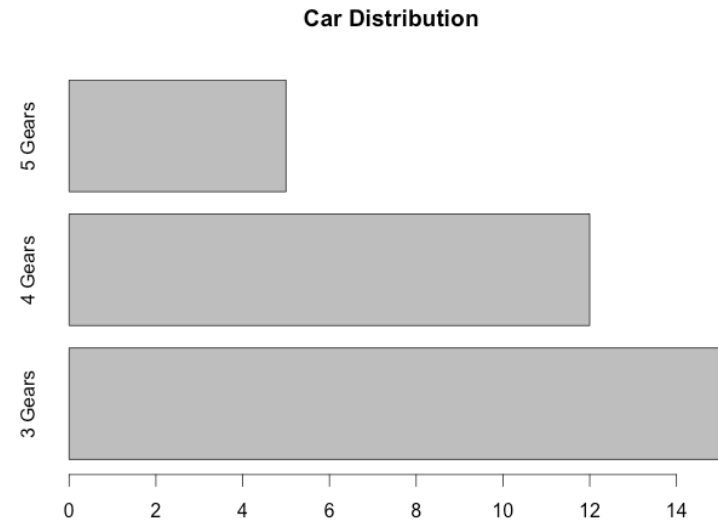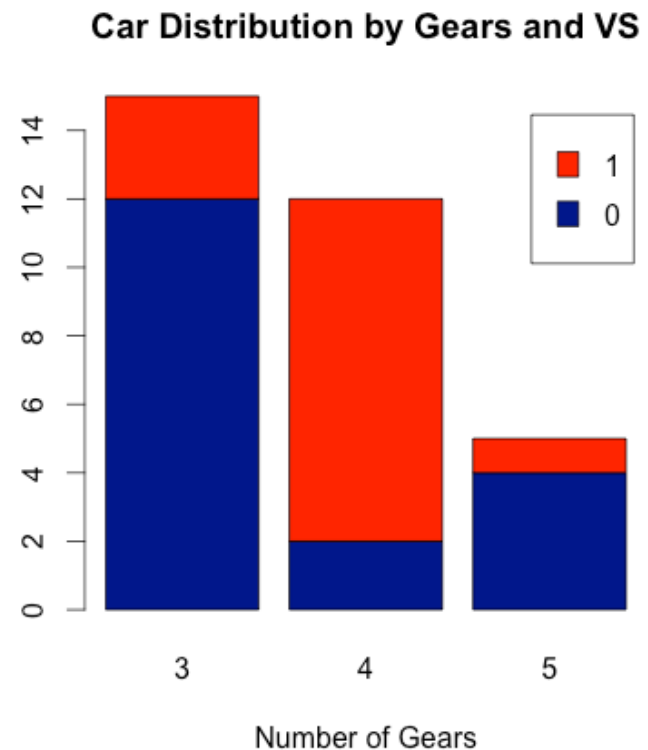
# Barplot

- ```
counts <-
table(mtcars$vs,
mtcars$gear)
```
- ```
barplot(counts,
main="Car Distribution
by Gears and VS",
xlab="Number of Gears",
col=c("darkblue","red"),
legend =
rownames(counts))
```

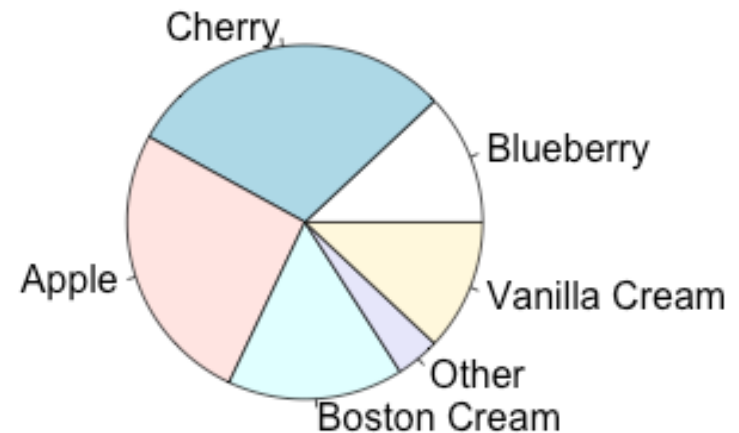Car Distribution by Gears and VS

Barplot

- ```
  barplot(counts,
  main="Car Distribution
  by Gears and VS",
  xlab="Number of Gears",
  col=c("darkblue","red"),
  legend =
  rownames(counts),
  beside=TRUE)
  ```


Car Distribution by Gears and VS

Number of Gears
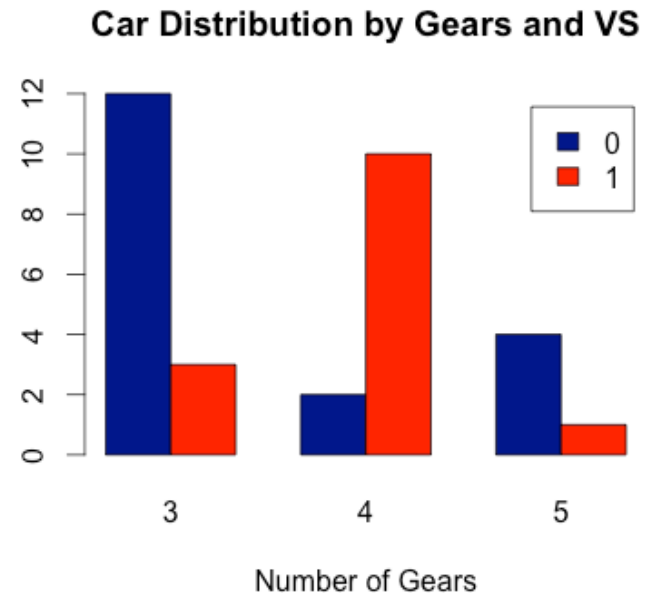
Pie

- `pie.sales <- c(0.12, 0.3, 0.26, 0.16, 0.04, 0.12)`
- `names(pie.sales) <- c("Blueberry", "Cherry", "Apple", "Boston Cream", "Other", "Vanilla Cream")`
- `pie(pie.sales)`

Pie

- `pie(pie.sales, col = rainbow(6))`

Export plots

- PDF
- PNG
- SVG
- JPEG
- BMP
- TIFF

- `pdf("archivo.pdf")`
- *…R sentences…*
- `dev.off()`

- `?png`
- `?pdf`

Bibliography

- R graphics. Paul Murrell. Computer Science and Data Analysis Series. Chapman & Hall/CRC. 2006.
- http://www.r-bloggers.com
- http://www.statmethods.net

EXERCISES (VISUALIZATION)

Introducción a la Ciencia de Datos

Exercise 1

- Given this data:

stretch	distance
46	148
54	182
48	173
50	166
44	109
42	141
52	166

① Plot distance against stretch.

Exercise 2

The table on the right have ten observations, taken during the years 1970-79, are on October snow cover for Eurasia (snow cover is in millions of square kilometers).

①   Plot snow.cover versus year.

②   Plot a histogram of the snow.cover values.

year	snow.cover
1970	6.5
1971	12.0
1972	14.9
1973	10.0
1974	10.7
1975	7.9
1976	21.9
1977	12.5
1978	14.5
1979	9.2

Exercise 3

Given the data in NY.xls

① (Optional) convert ºF to ºC and in to mm.

② Plot Year vs. Warmest Minimum Temperature

③ Plot Year vs. Warmest Minimum Temperature and Coldest Minimum Temperature. Don't forget to add a legend!