R training for
Médecins Sans Frontières

Hello!

We are Didac Fortuny and Paula Subías

We are here because we want to do good through data.

You can find us at in /didacfortuny

in /paula-subías

Hello!

We are part of dataforgoodBCN

You can find us at @dataforgoodBCN

Goal

- You are self-sufficient to use R for your daily computational tasks
- You know about what R can offer

Planning

Day 1: What we need to know about R and statistics

- Introduction to R
- Statistics hands-on

Day 2: Data frames and reporting tools

- Manipulating data frames.
- Reports

Day 3: Beyond basic R

- Plots
- Dashboards
- Other topics

1

What we need to know about R and statistics

R environment, data structures, and statistics

Introduction to R

- What is programming
- What is R
- What is RStudio®
- RStudio basics
- R basics

What is programming

- Create a list of instructions that can be performed automatically by a computer to fulfill a given task.

Introduction to R

What is programming

- Speed.

Introduction to R

What is programming

- Speed.
- Precision.

Introduction to R

What is programming

- Speed.
- Precision.
- Reproducibility.

What is R

- High-level programming language.

What is R

- High-level programming language.
- Interpreted language.

Introduction to R

What is R

- High-level programming language.
- Interpreted language.
- Open source.

Introduction to R

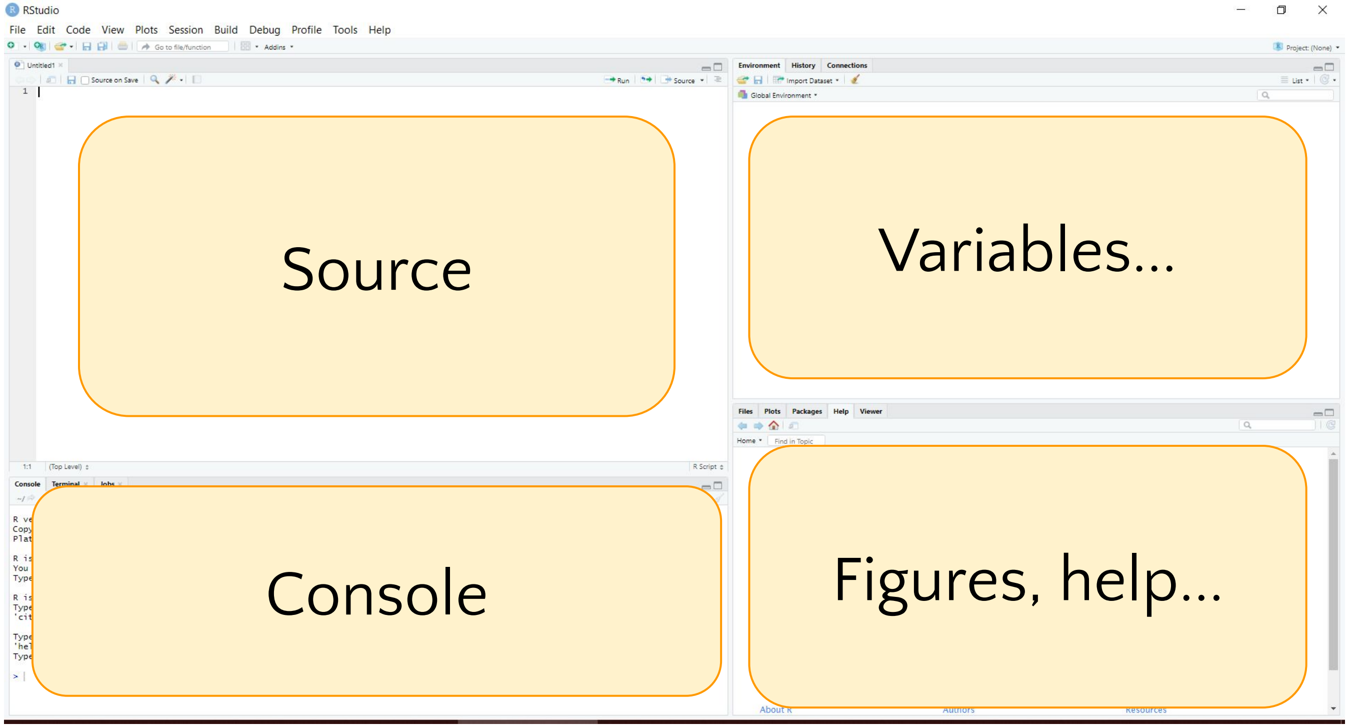
What is R

- High-level programming language.
- Interpreted language.
- Open source.
- Not a visual programming language.

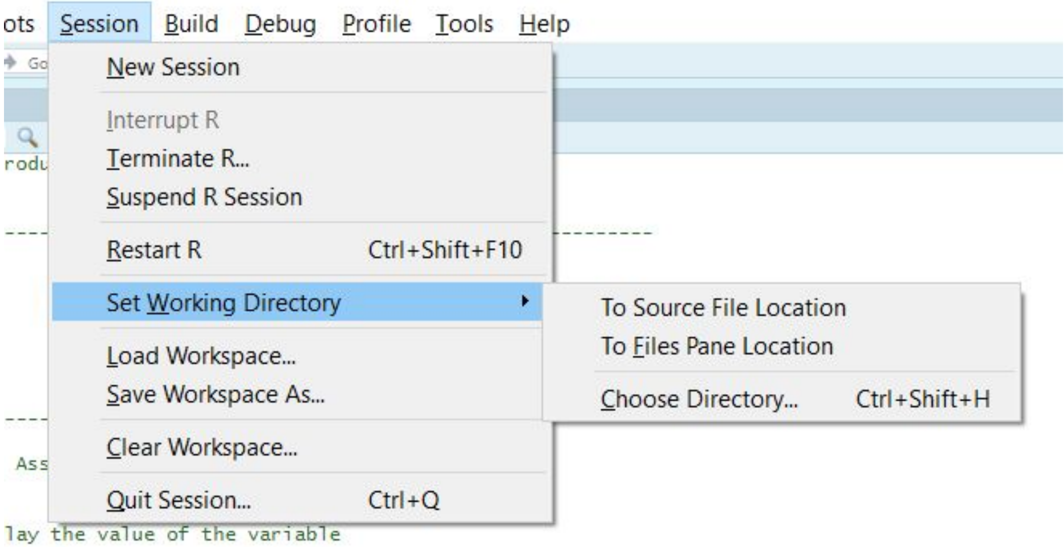
What is RStudio

- Integrated Development Environment (IDE).
- Open source.

RStudio basics



- Working directory:



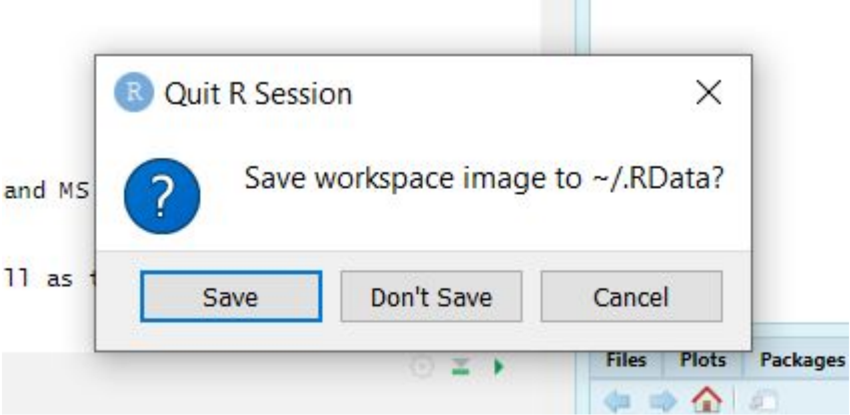
- Keyboard shortcuts:
  - Ctrl + ENTER: Run one cell.
  - Ctrl + SHIFT + F10: Restart session.

Go to first template.

Introduction to R

R basics

- Saving data when exiting the session



Quit R Session

Save workspace image to ~/.RData?

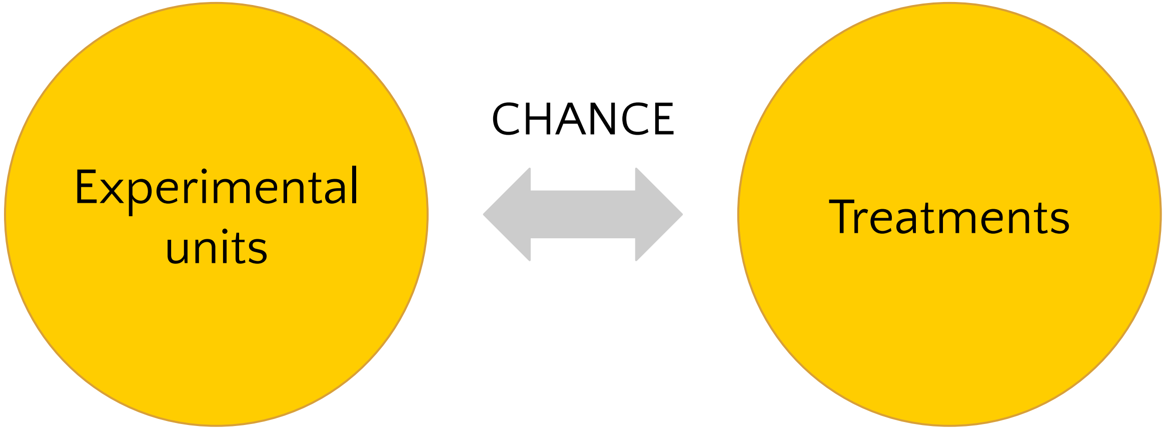
Statistics hands-on

- Types of distribution
- Statistical measures
- Sampling from distributions
- Comparing distributions

Statistics hands-on

Statistics allows to generalise from a sample of data to a true state of nature.

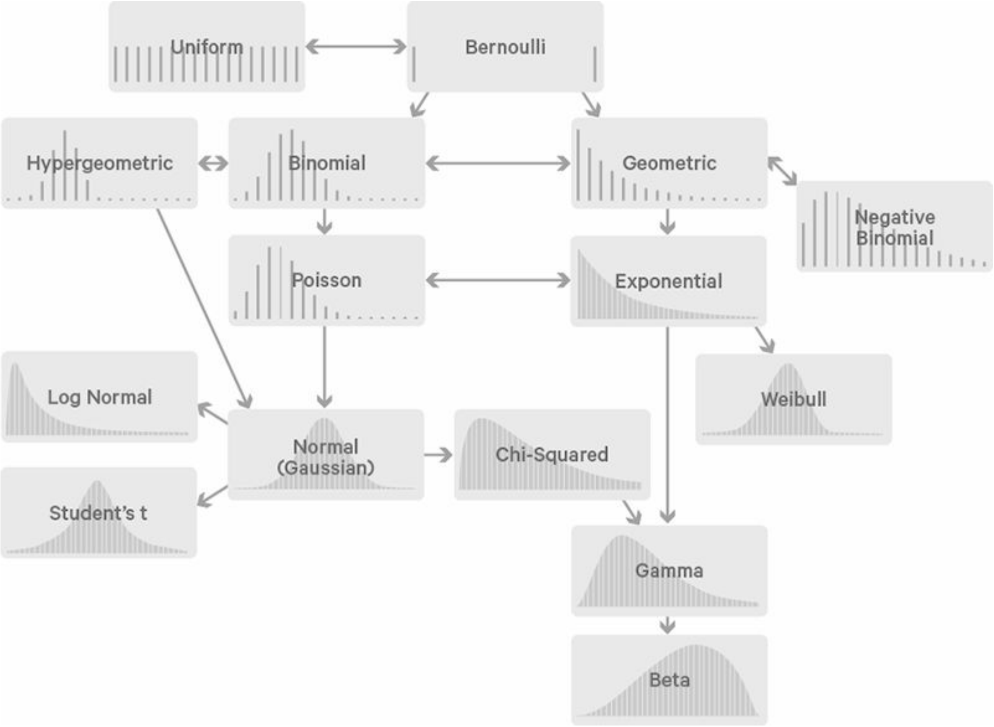
To make this work the data must be obtained by a carefully designed experiment.



Price to pay: we need probability to quantify the chance element in the data.

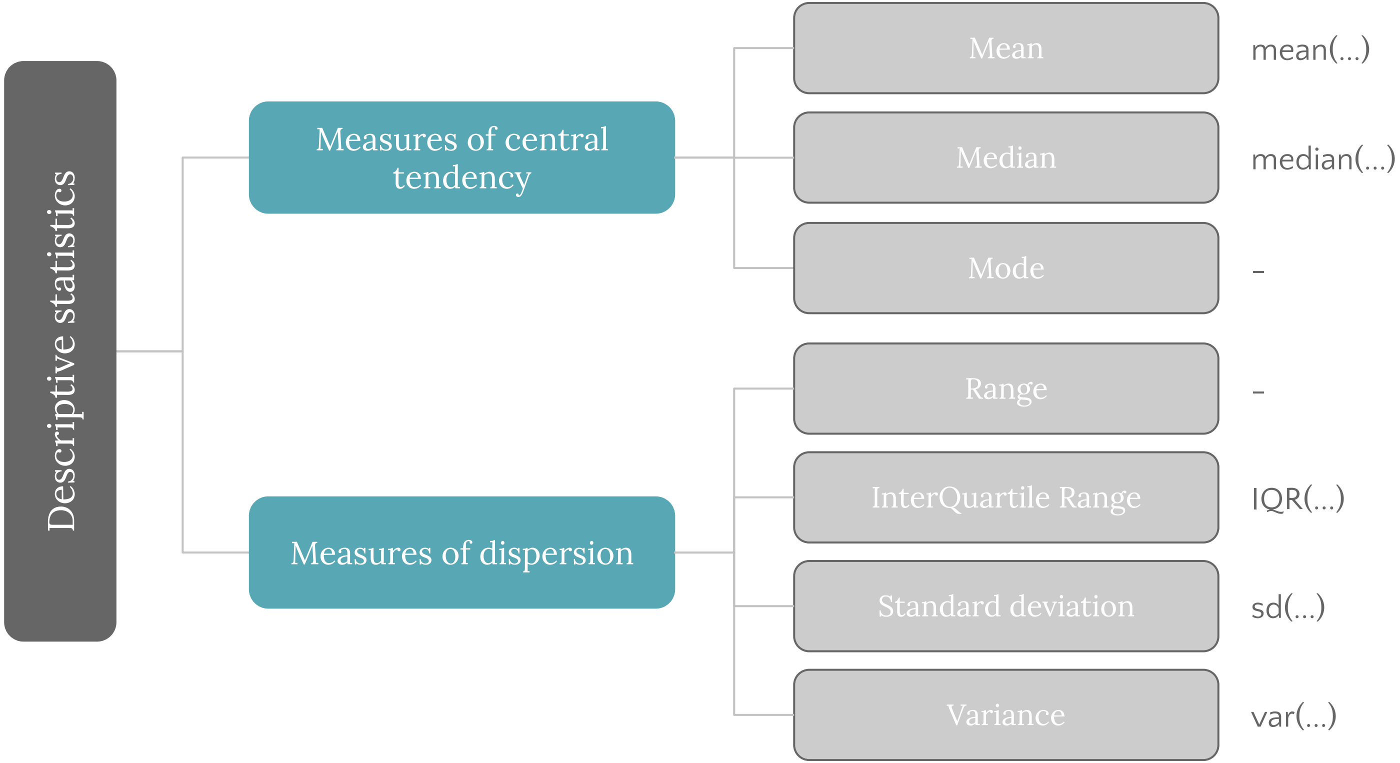
Statistics hands-on

Types of distribution



Statistics hands-on

Statistical measures



Descriptive statistics

Measures of central tendency

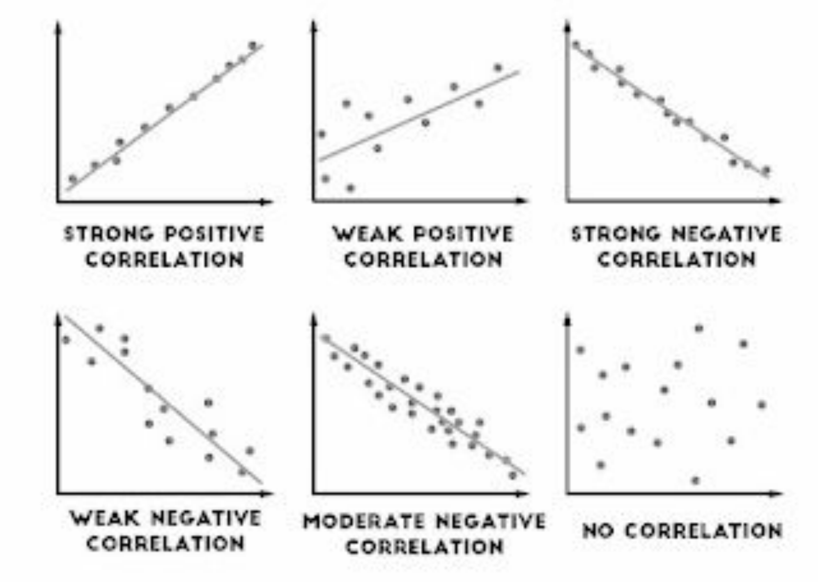
- Mean — mean(...)
- Median — median(...)
- Mode — -

Measures of dispersion

- Range — -
- InterQuartile Range — IQR(...)
- Standard deviation — sd(...)
- Variance — var(...)

Statistics hands-on

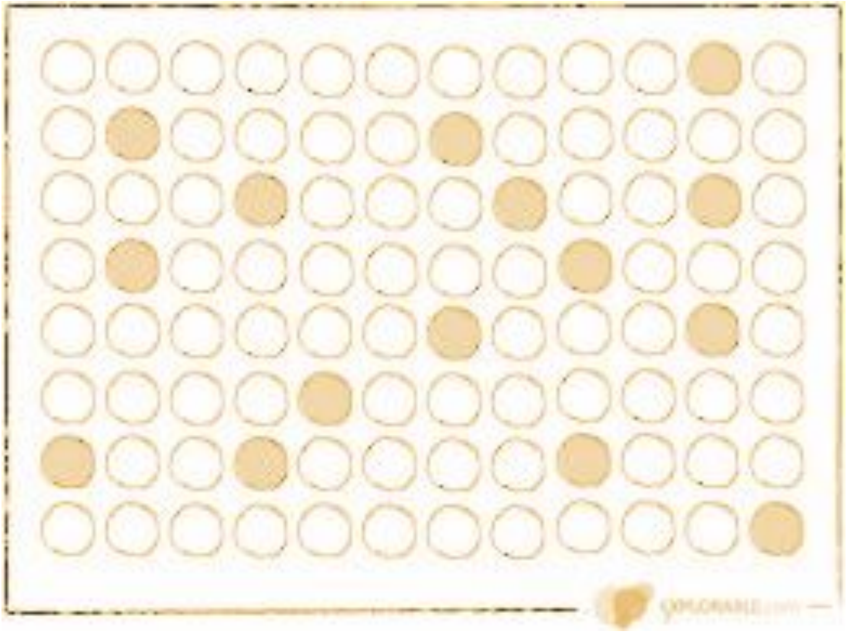
Statistical measures



cor(...)

Statistics hands-on

Sampling distributions



The **sampling distribution** of a statistic is the distribution of that statistic, considered as a random variable, when derived from a random sample of size n.

Statistics hands-on

Go to R.

Statistics hands-on

Descriptive statistics univariate with R			
Variable type		**Graphical representation**	**Numerical summary**
Categorical *qualitative*	Nominal	Pie plot: *pie(table(x))*	Absolute Frequency Table: *table(x)* Relative Frequency Table: *prop.table(x)*
	Ordinal	Bar plot: *barplot(table(x))*	Absolute Frequency Table: *table(x)* Relative Frequency Table: *prop.table(x)* Cumulative Absolute Frequency Table: *cumsum(table(x))* Cumulative Relative Frequency Table: *cumsum(prop.table(table(x)))*
Numerical *quantitative*	Discrete	Histogram: *hist(x)*	Summary: *summary(x)* Central Tendency Statistics: Mean: *mean(x)* Median: *median(x)* Positional Statistics: Minimum, Maximum: *min(x), max(x)* Quartiles (Q1,Q3): *quantile(x,0.25), quantile(x,0.75)* Percentile (alpha): *quantile(x,alpha)* Dispersion Statistics: Variance: *var(x)* Standard deviation: *sd(x)* IQR (Interquartile range=Q3-Q1): *IQR(x)*

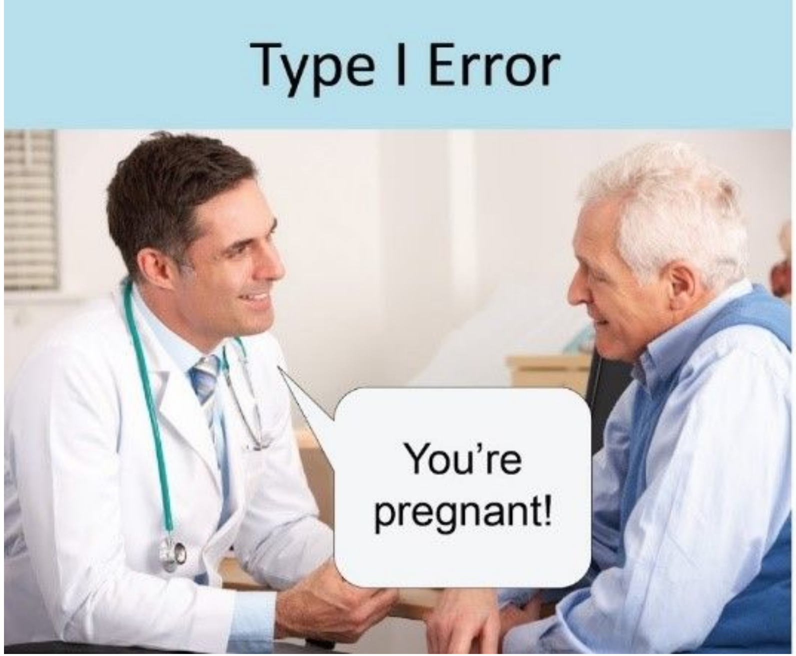
Statistics hands-on

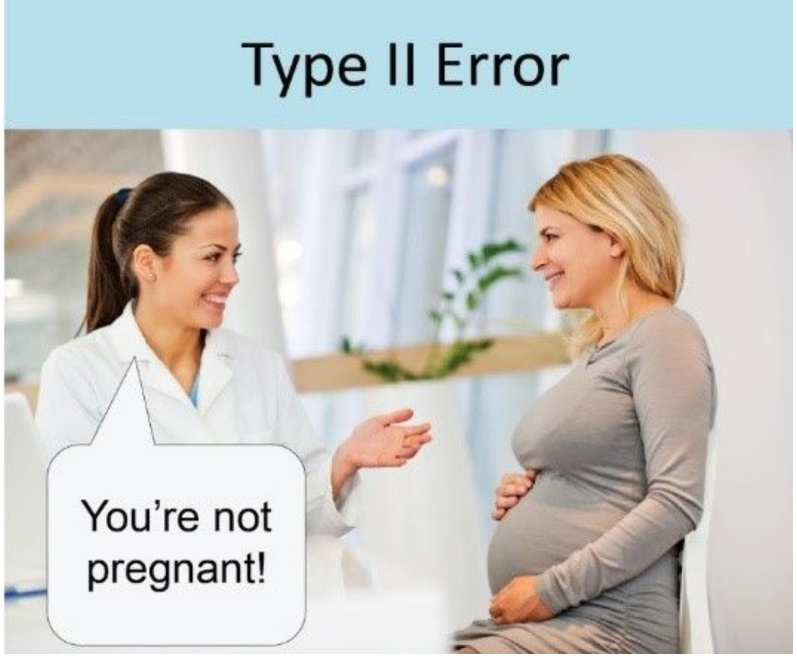
Descriptive statistics bivariate with R		
Variable type	**Graphical representation**	**Numerical summary**
f: categorical g: categorical	Grouped Bar diagram: *barplot(table(f,g))* Mosaic Plot: *mosaicplot(table(f,g))*	Absolute Contingency Table: *table(f,g)* Relative Contingency Table: *prop.table(table(f,g))*
x: numeric g: categorical	Points Diagram: *plot(g,x)* Grouped BoxPlot: *boxplot(x ~ g)* Grouped Histogram: *by(x,g,hist)*	Group Statistics: *by(x,g,summary)*
x: numeric y: numeric	Scatterplot: plot(x,y)	Linear correlation: cor(x,y)

Statistics hands-on

Go to R.

A **statistical test** chooses between two possibilities: the null hypothesis H_0 and the alternative hypothesis H_1 .


Type I Error
You're pregnant!


Type II Error
You're not pregnant!

Statistics hands-on

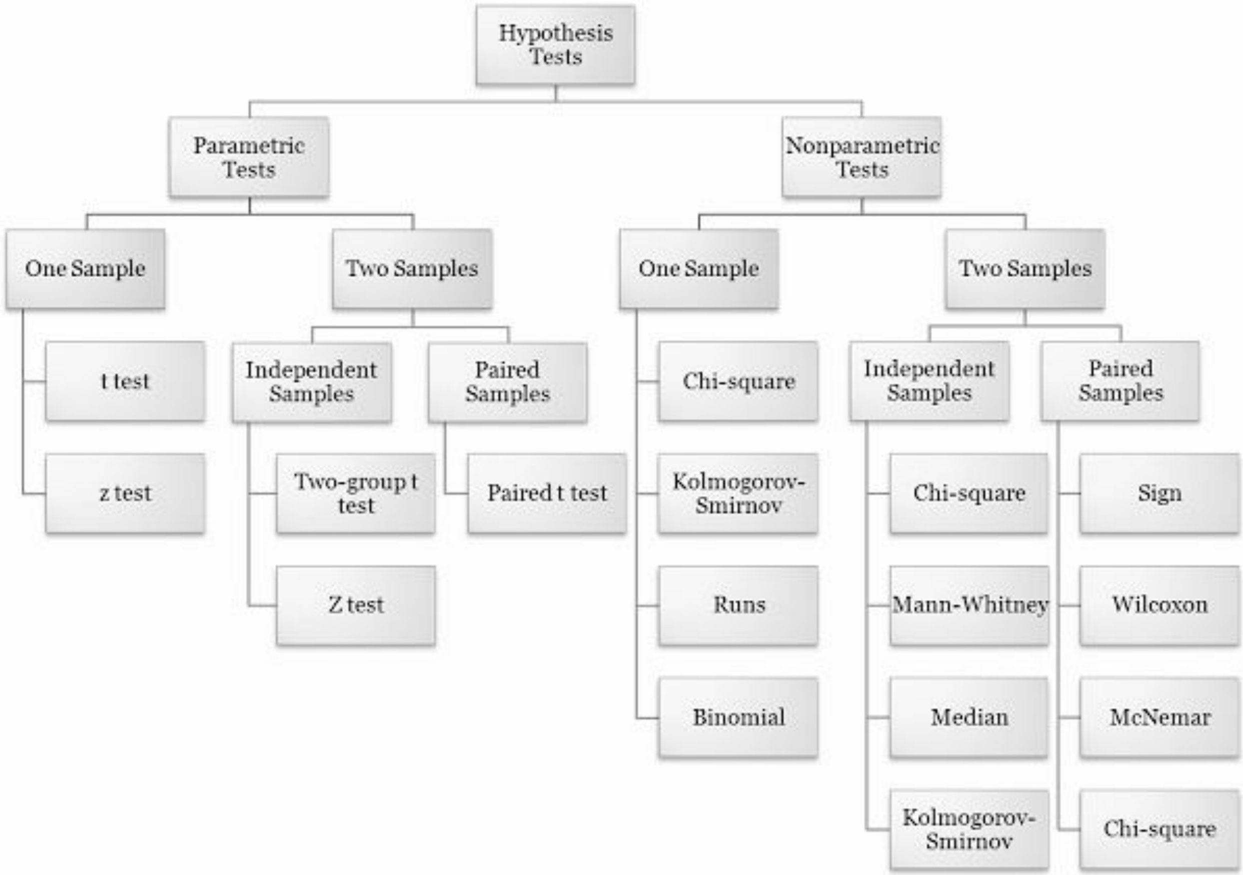
Comparing distributions

Parametric tests assume your data meets some assumptions, which, basically, are:

- Interval level data
- Reasonably normally distributed
- Equal variances
- Reasonable sample size

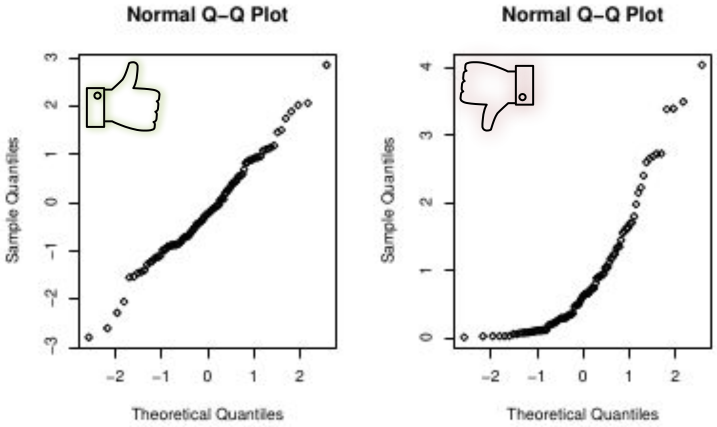
Non-parametric tests:

- Valid for most distributions
- Often easier to compute


Hypothesis Tests
Parametric Tests
Nonparametric Tests
One Sample
Two Samples
One Sample
Two Samples
t test
z test
Independent Samples
Paired Samples
Two-group t test
Z test
Paired t test
Chi-square
Kolmogorov-Smirnov
Runs
Binomial
Independent Samples
Paired Samples
Chi-square
Mann-Whitney
Median
Kolmogorov-Smirnov
Sign
Wilcoxon
McNemar
Chi-square

Checking for normality: QQ-plot

A QQ-plot can reveal whether data (approximately) follows a certain population curve, e.g. the normal curve.



Statistics hands-on

Go to R.

2

Data frames and reporting tools

Let’s create the first analyses and reports

Reports with R Markdown

Reports with R Markdown

R Markdown

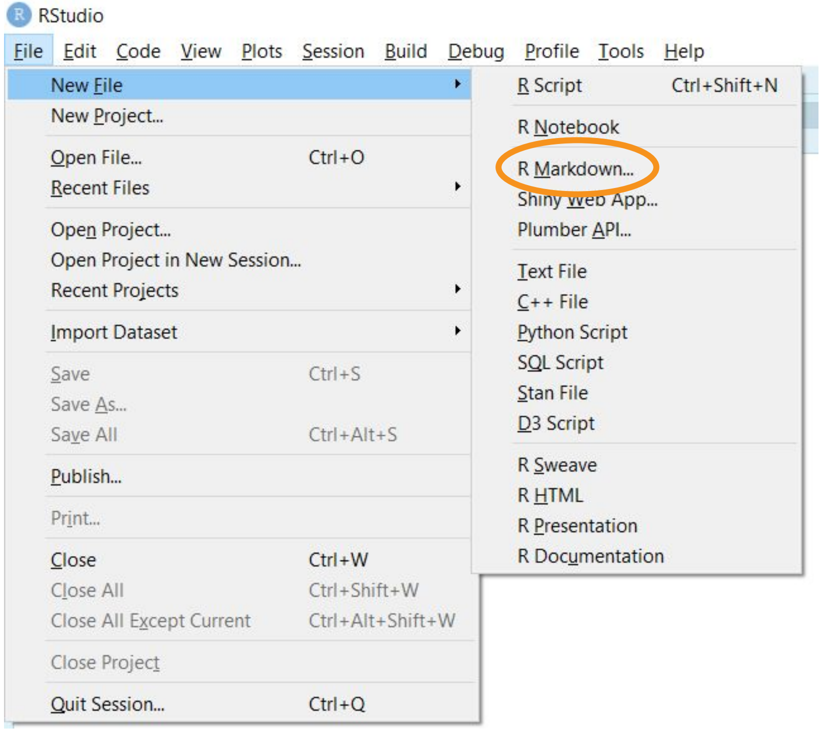
from R Studio

- Tool for automating reports.

Reports with R Markdown

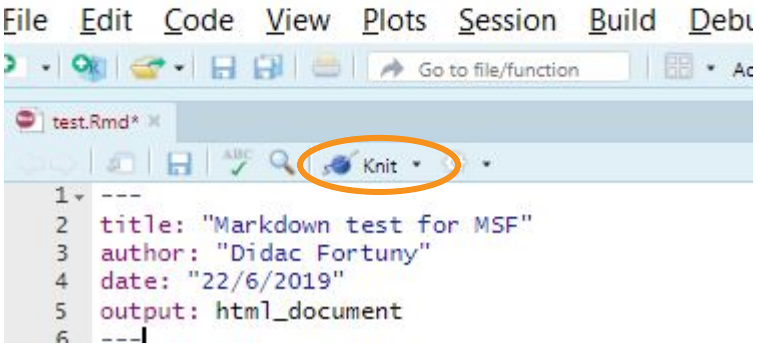
- Tool for automating reports.
- Supporting multiple languages.

Reports with R Markdown



Reports with R Markdown

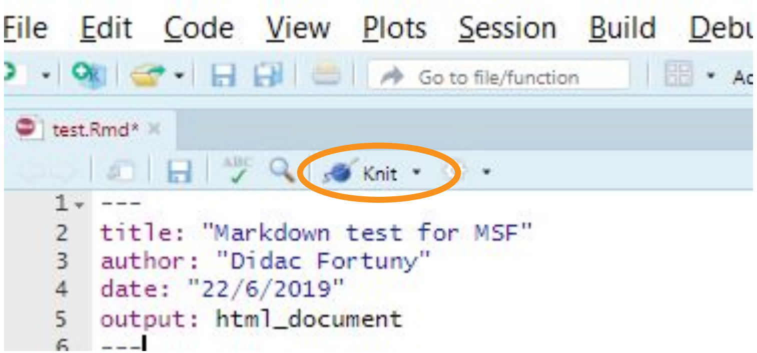
- Publish report



```
---
title: "Markdown test for MSF"
author: "Didac Fortuny"
date: "22/6/2019"
output: html_document
---
```

Reports with R Markdown

- Publish report



```
---
title: "Markdown test for MSF"
author: "Didac Fortuny"
date: "22/6/2019"
output: html_document
---
```

- Publish in PDF:
    - Needs Latex installed: miktex.org/download
    - Needs the tinytext package installed:
        - `install.packages("tinytex")`
        - `tinytex::install_tinytex()`

Reports with R Markdown

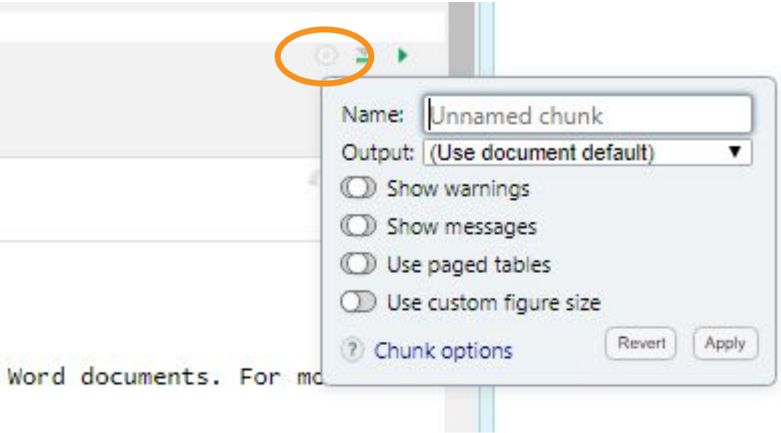
- R Markdown file structure:
  - Header (between ---)
  - Code chunks (between ` ` `)
  - Simple text.

Reports with R Markdown

- Header:
    - Contains basic information about the report.

Reports with R Markdown

- Code chunks



Name: Unnamed chunk

Output: (Use document default) ▼

○ Show warnings

○ Show messages

○ Use paged tables

○ Use custom figure size

Reports with R Markdown

- Simple text:
  - Markdown format: markdown-guide.readthedocs.io
  - Inline code: `r ______`

Reports with R Markdown

- Keyboard shortcuts:
  - Ctrl + Alt + C: Run current chunk.
  - Ctrl + Alt + I: Insert new chunk.
  - Ctrl + Shift + K: Publish document.

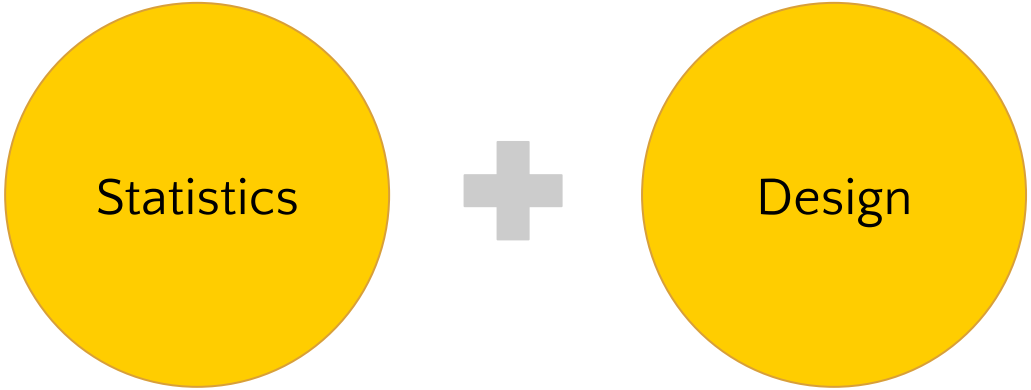
Reports with R Markdown

- Online tutorial: rmarkdown.rstudio.com

3 Beyond basic R

Plots, dashboards and everything else

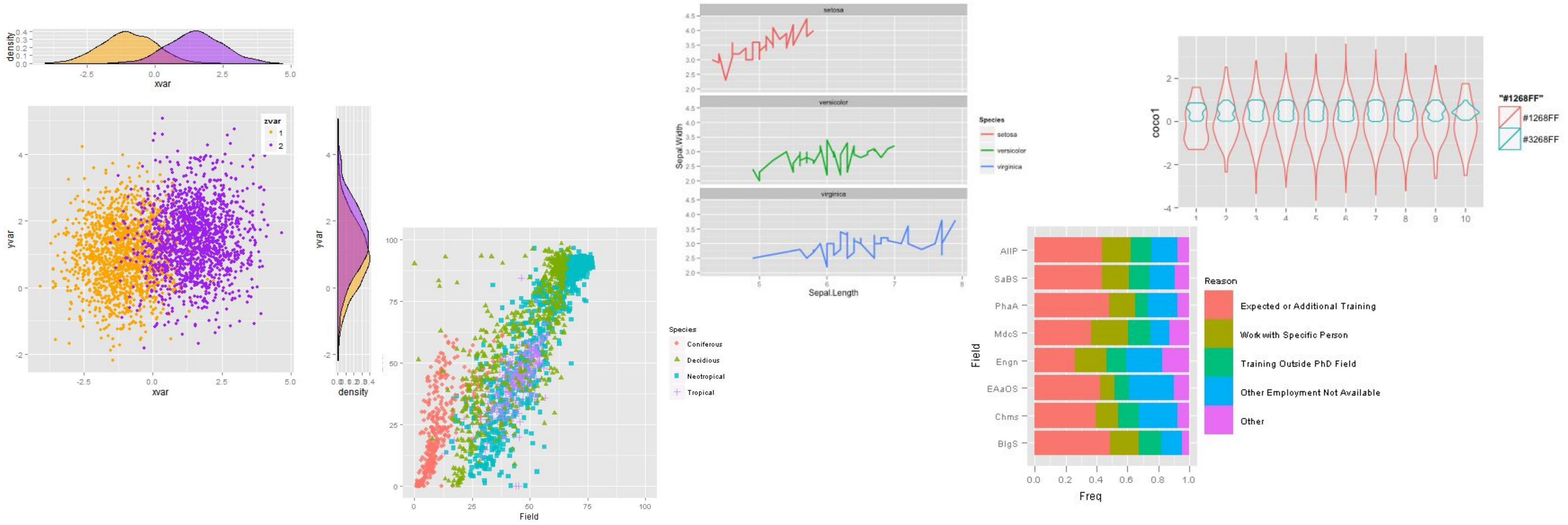
Plots



Statistics + Design

Plots

👉 ggplot2: the most powerful library for plots!



Plots

- Distribution: histogram, boxplot
- Correlation: scatterplot
- Ranking: barplot, spider

👉 Check out the R graph gallery!

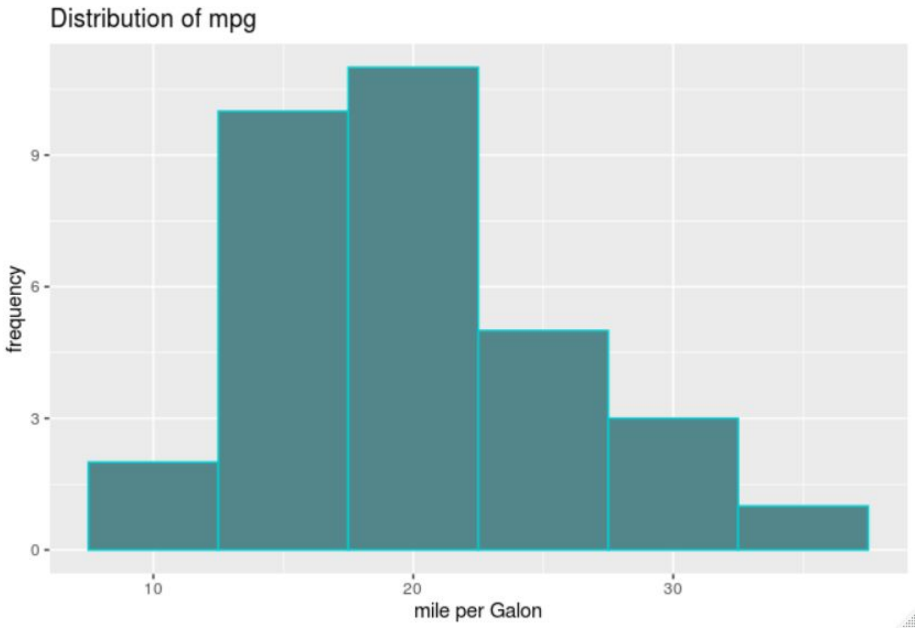
Plots

ggplot()	Create a new ggplot.
aes()	Construct aesthetic mappings.
+ (<gg>)	Add components to a plot.
ggsave	Save a ggplot.
qplot()	Quick plot.

👉 Go to cheatsheet.

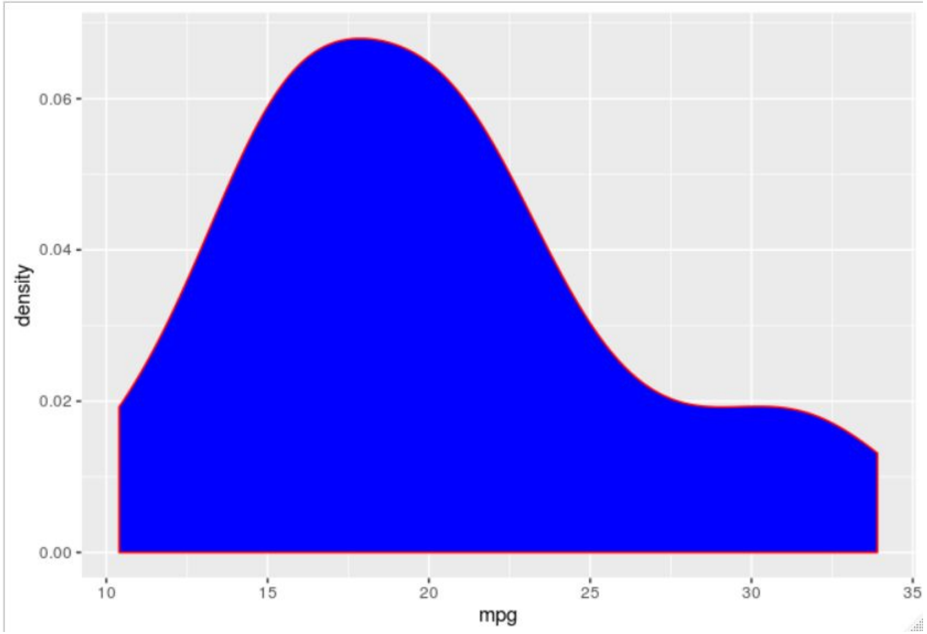
Plots

Histogram



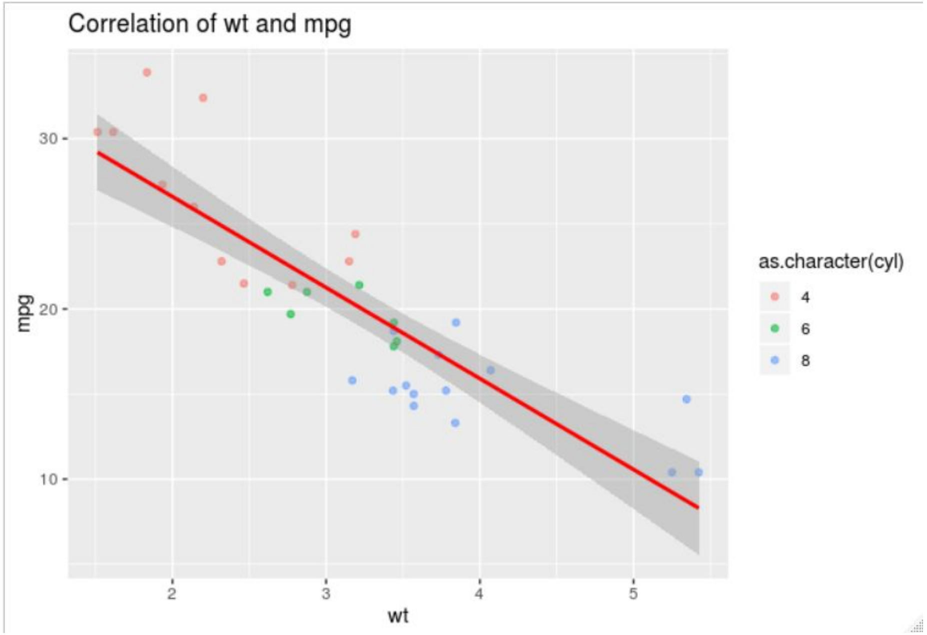
Plots

Density plot



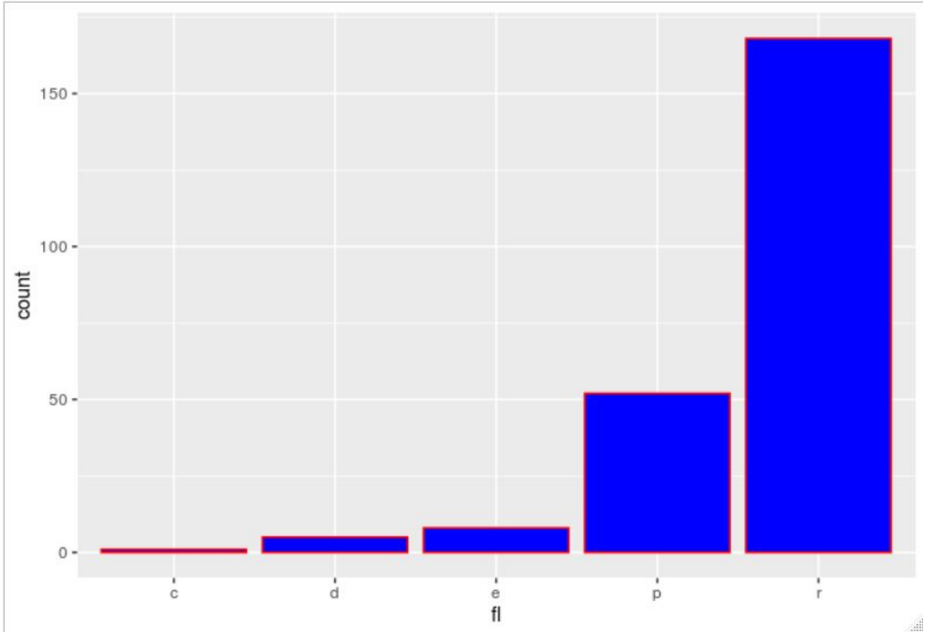
Plots

Scatterplot



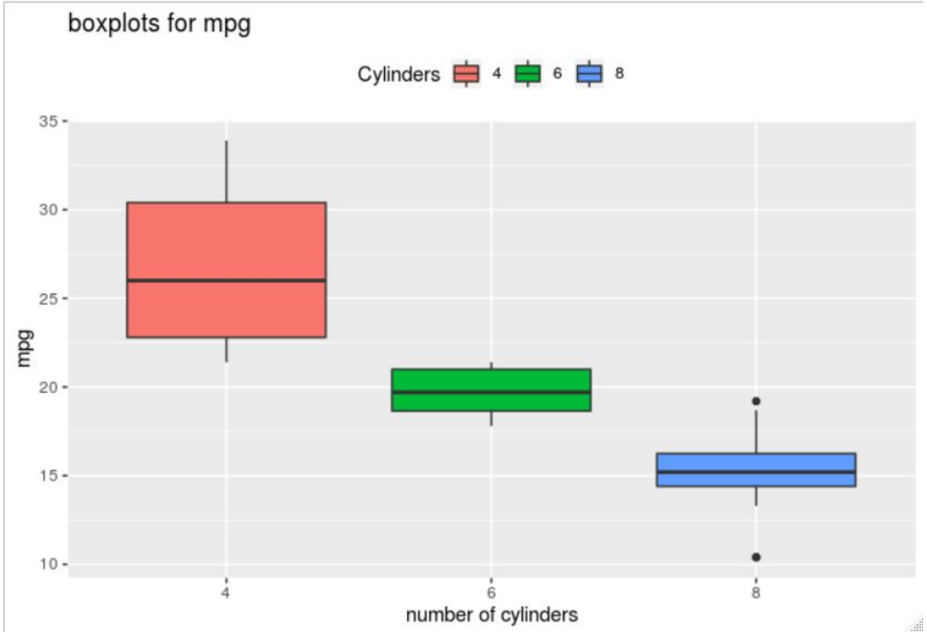
Plots

Barplot



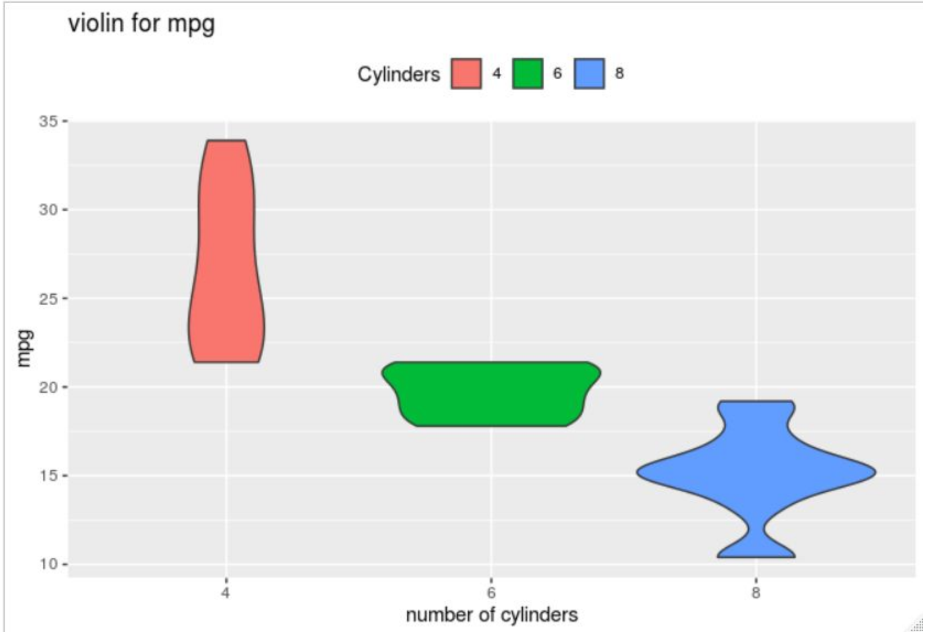
Plots

Boxplot



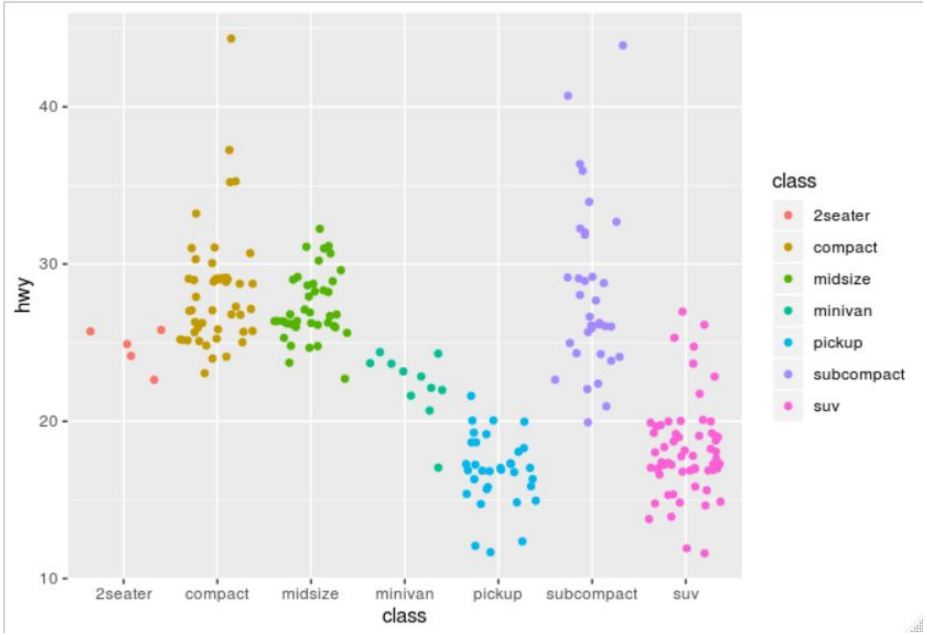
Plots

Violin



Plots

Compare distributions

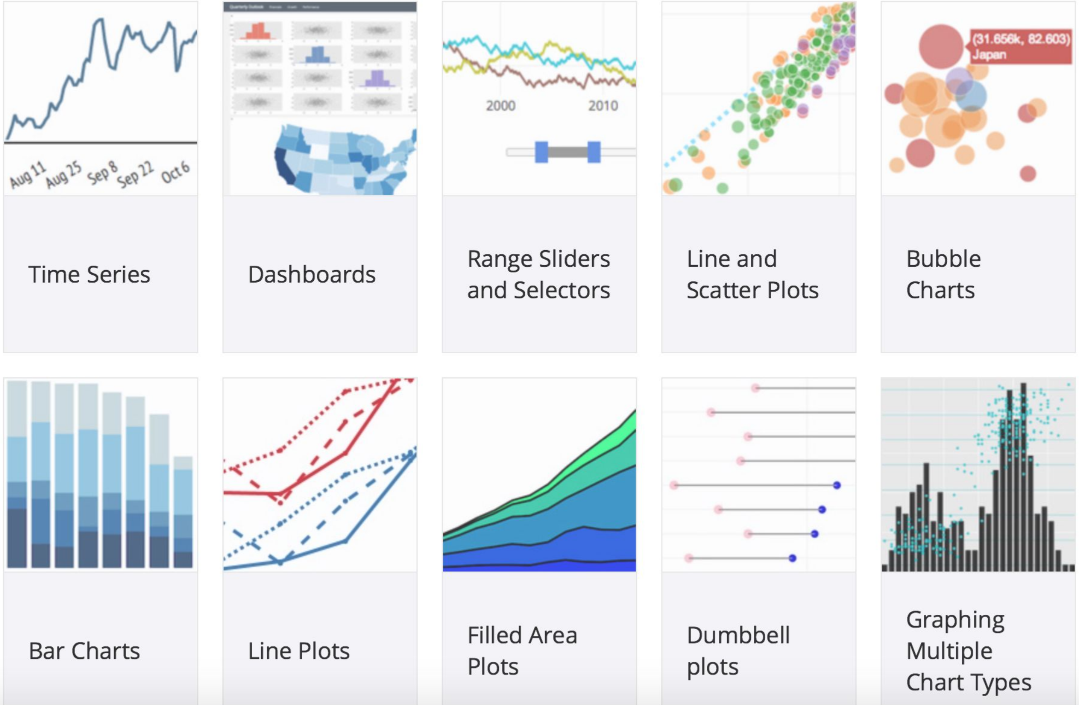


Plots

For more details and information, visit this link: https://plot.ly/r/

Plug-in to include them in Powerpoint: here

Interactive plots



Dashboards with Shiny

What is Shiny?

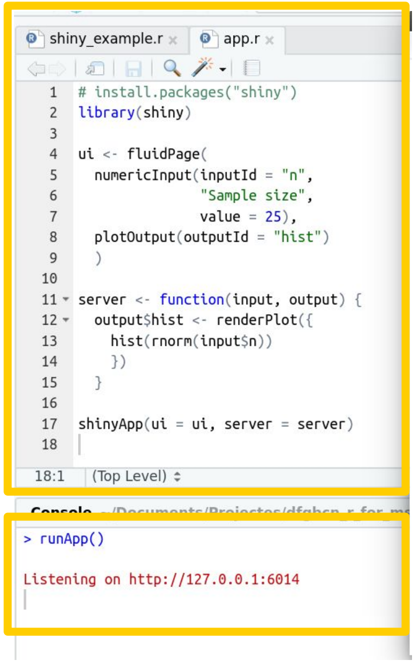
Go to R.

👉 Check out the official tutorial!

Dashboards with Shiny

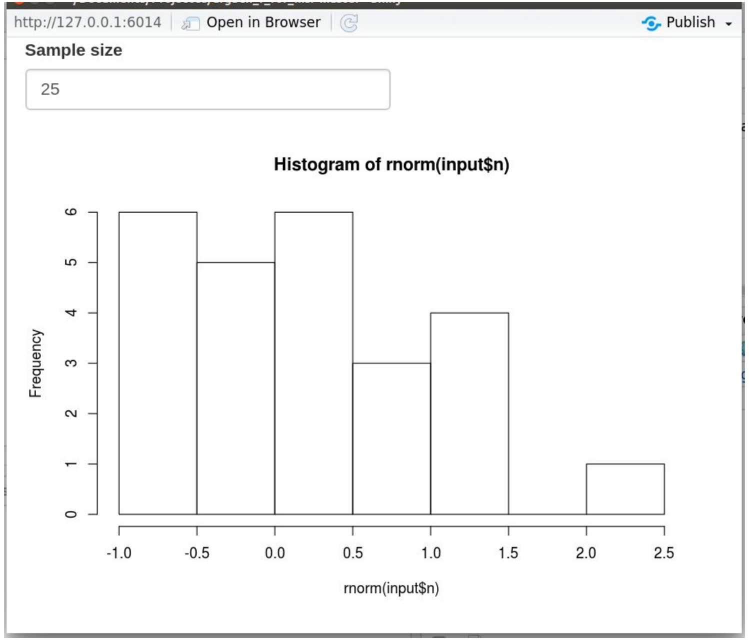
App

Execution



```r
# install.packages("shiny")
library(shiny)

ui <- fluidPage(
  numericInput(inputId = "n",
               "Sample size",
               value = 25),
  plotOutput(outputId = "hist")
  )

server <- function(input, output) {
  output$hist <- renderPlot({
    hist(rnorm(input$n))
  })
  }

shinyApp(ui = ui, server = server)

```

```
> runApp()

Listening on http://127.0.0.1:6014
```

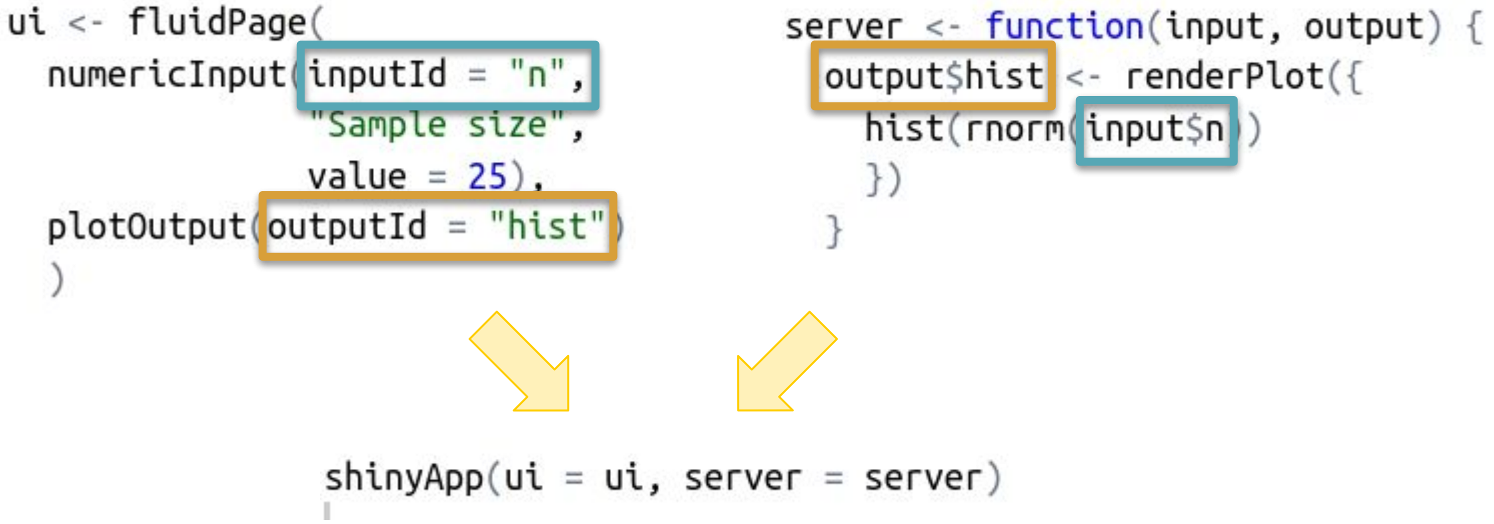


Sample size: 25

Dashboards with Shiny

Input functions

Output functions



```
ui <- fluidPage(
  numericInput(inputId = "n",
               "Sample size",
               value = 25),
  plotOutput(outputId = "hist")
)
```

```
server <- function(input, output) {
  output$hist <- renderPlot({
    hist(rnorm(input$n))
    })
  }
```

```
shinyApp(ui = ui, server = server)
```

📌 Dashboards with Shiny

👉 Go to cheatsheet.

Dashboards with Shiny

Go to R.

Thanks!

Any *questions* ?

You can find us at

- dacfortuny *at* gmail *dot* com
- paulasubias *at* gmail *dot* com

Credits

Special thanks to all the people who made and released these awesome resources for free:

- Presentation template by SlidesCarnival