R for linguists

Mathieu Avanzi

FNRS & Université catholique de Louvain

Bern Universität

Outline of the class

- Introduction
  - What is R
  - Download, install and open R
- Basic operations in R
  - Use R as a calculator
  - Load and display a dataset
  - Perform basic operations on it
- Basic stats
  - t-tests
  - anova
  - chi-square
- Plots in R: the ggplot2 package
  - Barplots and boxplots
  - Basic maps in R
  - Heatmaps
  - Clustering

What is R?

- R is a language and environment for statistical computing and graphics

What is R?

- R is a **language** [code] and an environment for statistical computing and graphics

What is R?

- R is a language [code] and an **environment** [editor where you write/paste some code] for statistical computing and graphics

What is R?

- R is a language and an environment for **statistical computing** [from t-tests to Generalized Linear Mixed Models] and graphics

What is R?

- R is a language and an environment for statistical computing and **graphics** [realize some nice and high quality plots]

What is R?

- R is a language and environment for statistical computing and graphics

- Why using R (and not SPSS or Stata)?
  - Because everybody uses it
  - Because it's free and open-source
  - Because you can program anything you want (more than 10k packages are available for downloading)
- The cons
  - The programming language is a programming language
  - You need to train a lot before being confident in its use

Download and installation

- Download R here: www.r-project.org

Download and installation

- Download R studio here: www.rstudio.com

R studio environment

This menu reminds one of the Windows' environment – useful if you don't want to use code to perform some basic operations!
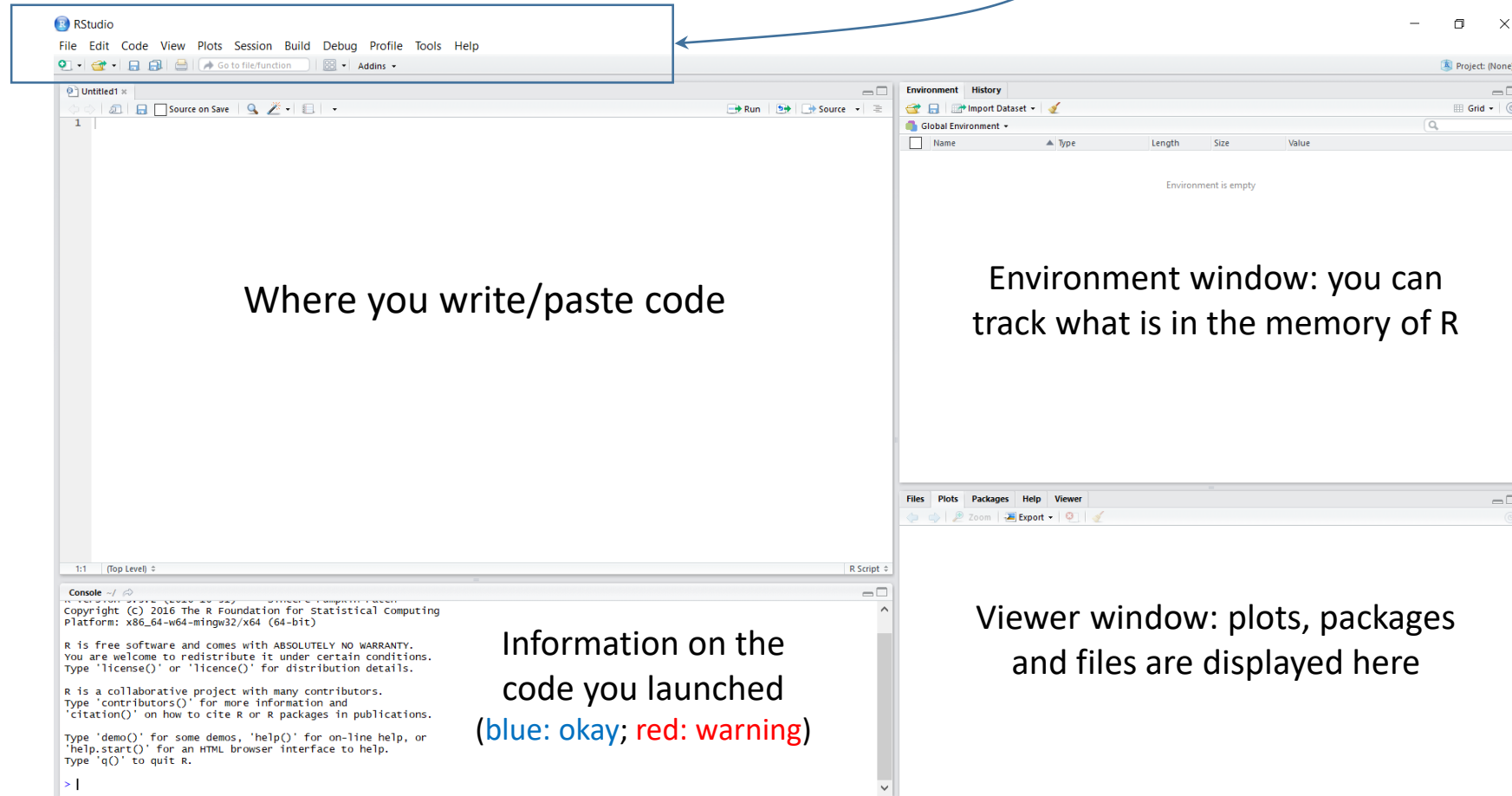
Where you write/paste code

Environment window: you can track what is in the memory of R

Information on the code you launched
(blue: okay; red: warning)

Viewer window: plots, packages and files are displayed here

Outline of the class

- Introduction
  - What is R
  - Download, install and open R
- **Basic operations in R**
  - Use R as a calculator
  - Load and display a dataset
  - Perform basic operations on it
- Basic stats
  - t-tests
  - anova
  - chi-square
- Plots in R: the ggplot2 package
  - Barplots and boxplots
  - Basic maps in R
  - Heatmaps
  - Clustering

R as a calculator

Type the following lines in the main window. At the end of each line, click on Run (CTRL+R or CRTL+enter).
What do you observe?

> 2+2   > 2+3;1+9

> 5+7   > LETTERS

> 40/4   > pi

> 40 / 4   > 10+2*5

> 4 ^2   > (10+2)*5

Run

Create some objects

Type the following lines in the main window. In the end of each line, click on Run (CTRL+R or CRTL+enter). What do you observe?

> 2+2	> 2+3;1+9	> x <- 4	> you = 2
> 5+7	> LETTERS	> x	> x + you
> 40/4	> pi	> x = 5	> y = x+you
> 40 / 4	> 10+2*5	> x*3	> y < 4
> 4 ^2	> (10+2)*5	> x + you	> y == 4

Run

Create some objects

Type the following lines in the main window. In the end of each line, click on Run (CTRL+R or CRTL+enter). What do you observe?

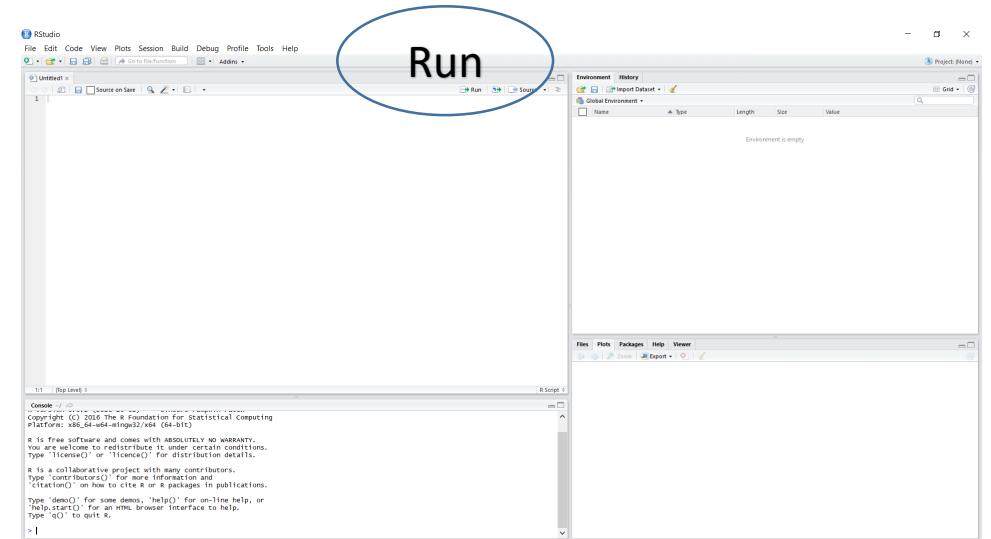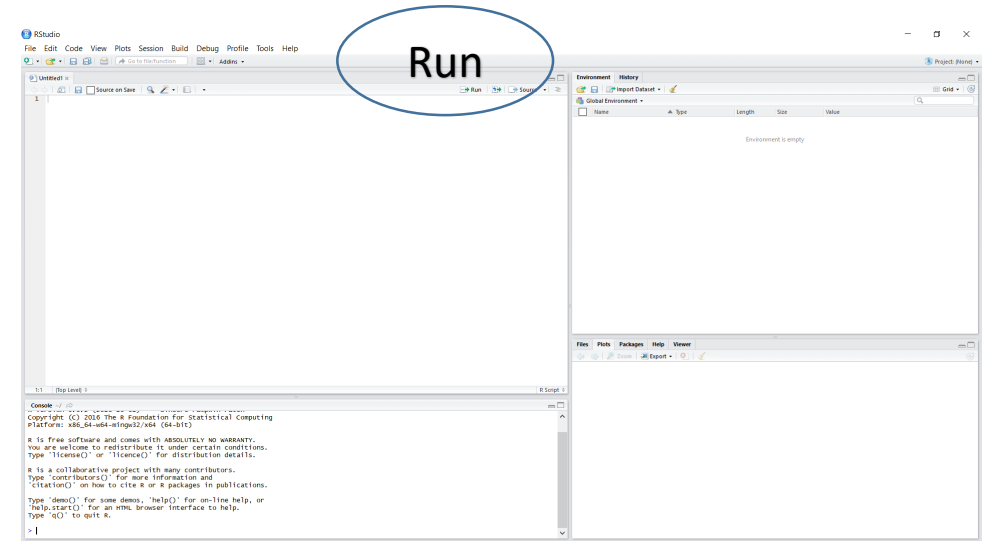|				
> 2+2	> 2+3;1+9	> x <- 4	> you = 2	> y
> 5+7	> LETTERS	> x	> x + you	> X
> 40/4	> pi	> x = 5	> y = x+you	> meann(x+y)
> 40 / 4	> 10+2*5	> x*3	> y < 4	> mean(x+y)
> 4 ^2	> (10+2)*5	> x + you	> y == 4	> sum(xy)
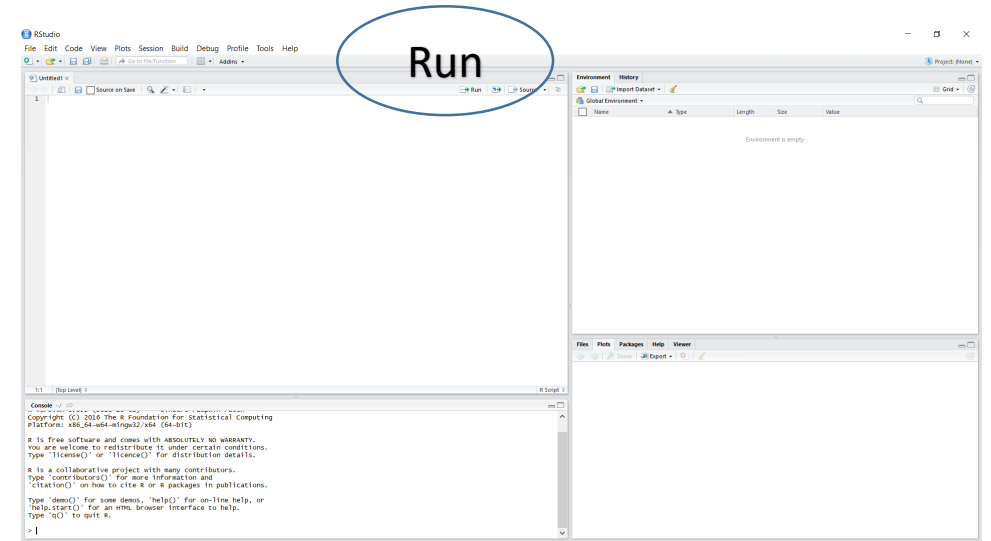
Pit stop

- R as a calculator
- Symbols « + », « * », « - », « / », « ^2 » for basic mathematical operations
- Brackets () have the same use than in mathematics
- Inserting blanks between the elements of a line does not impact the code
- But R is case-sensitive (watch out capital and non-capital letters)
- You can store objects in variables using the operators « = » or « <- »
- Use the command « run » to launch a line (result appears in the window below)

Data types in R

- Numeric vectors
- Character vectors
- Factors
- Matrices
- Data frames

Numeric vectors

> vector1 <- c(1,2,3,4)

> vector1

> vector2 <- c(1:4)

> vector2

> vector3 <- c(1,3,4,2,8:15)

> vector3

➢When you deal with numbers (numerical variables), the elements are blue in R

Character and factor vectors

> char <-c("pink", "is", "beautiful", "red", "is", "not")
> char

> char.fact <-factor(char)
> char.fact

➢When you deal with characters or factors, you need to put them between quotes, and they are green

Data frames

```
> sex <- c("m", "m", "m", "f")
> sex


> weight <-c(75, 83, 65, 52)
> weight


> df <- data.frame(sex,rt)
> df
```
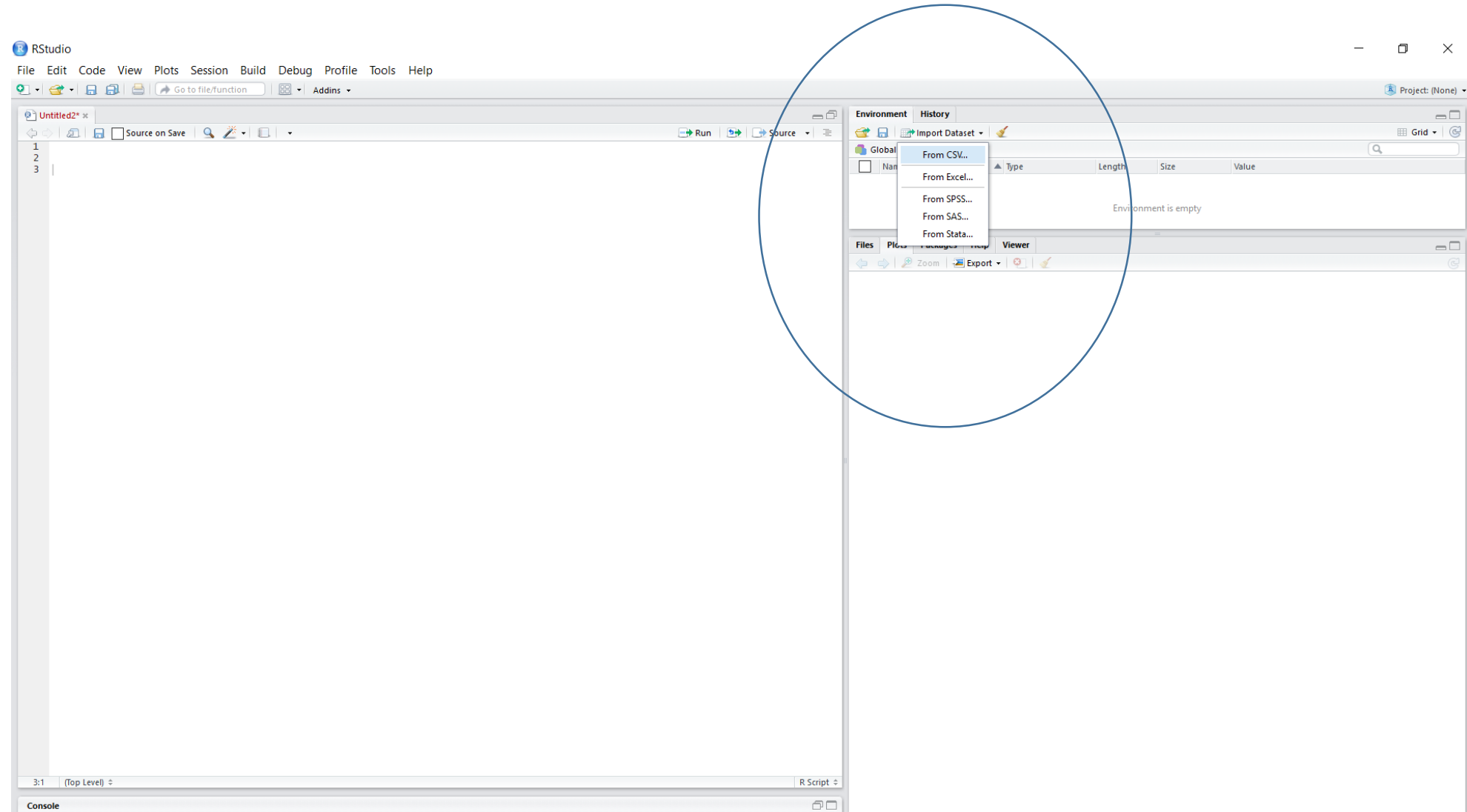
Exercise

- Create a fake dataframe with the names of your friends and their supposed height.
    - First you need to create a vector with the names of your friends
    - Second you need to create a vector with their estimated heights
    - Third you need to combine the two vectors in a dataframe

Load a dataset in R

- Option 1

Load a dataset in R

- Option 2

```
#Specify your working directory
#replace backslashes by forward slashes if needed (as below)
setwd("C:/Users/avanzi/Dropbox")

#create the object
df = read.table("data1.txt", header=T, sep="\t", quote="", dec=".")
df
```

Basic operations on a dataframe

head(df)

#display the first 6 rows

head(df,10)

#display the first 10 rows

tail(df)

#display the last 6 rows

str(df)

#displays the structure

head(df$Age)

#displays the first lines of a specific column

Mean, median and mode

mean(df$Age)

#get mean of a column

median(df$Age)

#get mean of a column

summary(df$Age)

#get everything

Measures of dispersion

range(df$Age)

#min and max

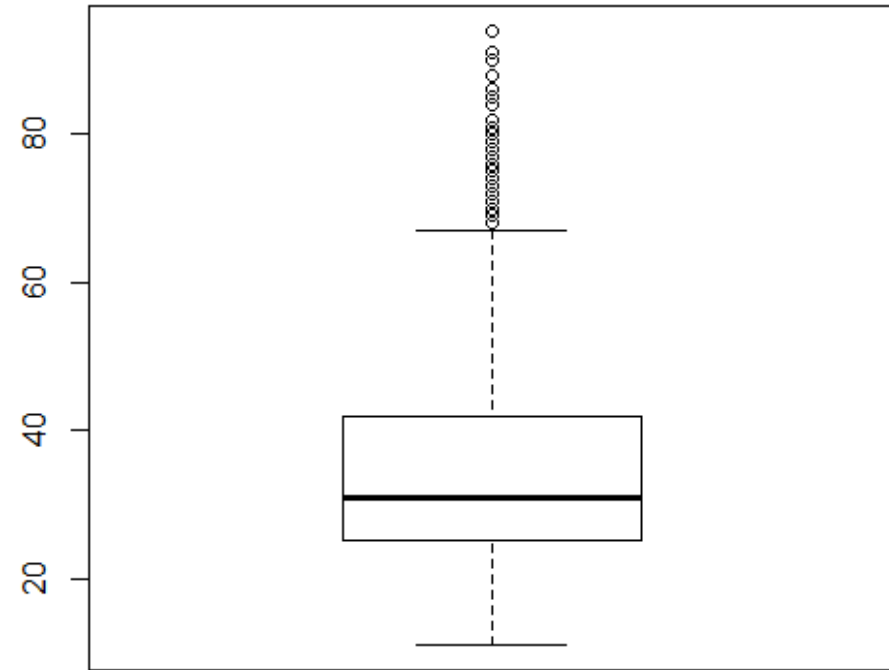#sum of squared deviations from the mean, divided by the number of observations

sd(df$Age)

#squared root of the variance

Basic graphs

boxplot(df$Age)

boxplot.stats(df$Age)

Basic graphs

hist(df$Age)

Histogram of df$age