

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



[\[Home\]](#)

Download

[CRAN](#)

R Project

[About R](#)

[Logo](#)

[Contributors](#)

[What's New?](#)

[Reporting Bugs](#)

[Development Site](#)

[Conferences](#)

[Search](#)

R Foundation

[Foundation](#)

[Board](#)

[Members](#)

[Donors](#)

[Donate](#)

Help With R

[Getting Help](#)

The R Project for Statistical Computing

Getting Started

R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To [download R](#), please choose your preferred [CRAN mirror](#).


If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.


News

- [R version 3.5.0 \(Joy in Playing\) prerelease versions](#) will appear starting Friday 2018-03-23. Final release is scheduled for Monday 2018-04-23.
- [R version 3.4.4 \(Someone to Lean On\)](#) has been released on 2018-03-15.
- [useR! 2018](#) (July 10 - 13 in Brisbane) is open for registration at <https://user2018.r-project.org>
- [The R Journal Volume 9/2](#) is available.
- [R version 3.3.3 \(Another Canoe\)](#) has been released on Monday 2017-03-06.
- [useR! 2017](#) took place July 4 - 7 in Brussels <https://user2017.brussels>
- The [R Logo](#) is available for download in high-resolution PNG or SVG formats.

Download and installation


- Download R studio here: www.rstudio.com




Products Resources Pricing About Us Blogs 

Choose Your Version of RStudio

RStudio is a set of integrated tools designed to help you be more productive with R. It includes a console, syntax-highlighting editor that supports direct code execution, and a variety of robust tools for plotting, viewing history, debugging and managing your workspace. [Learn More about RStudio features.](#)



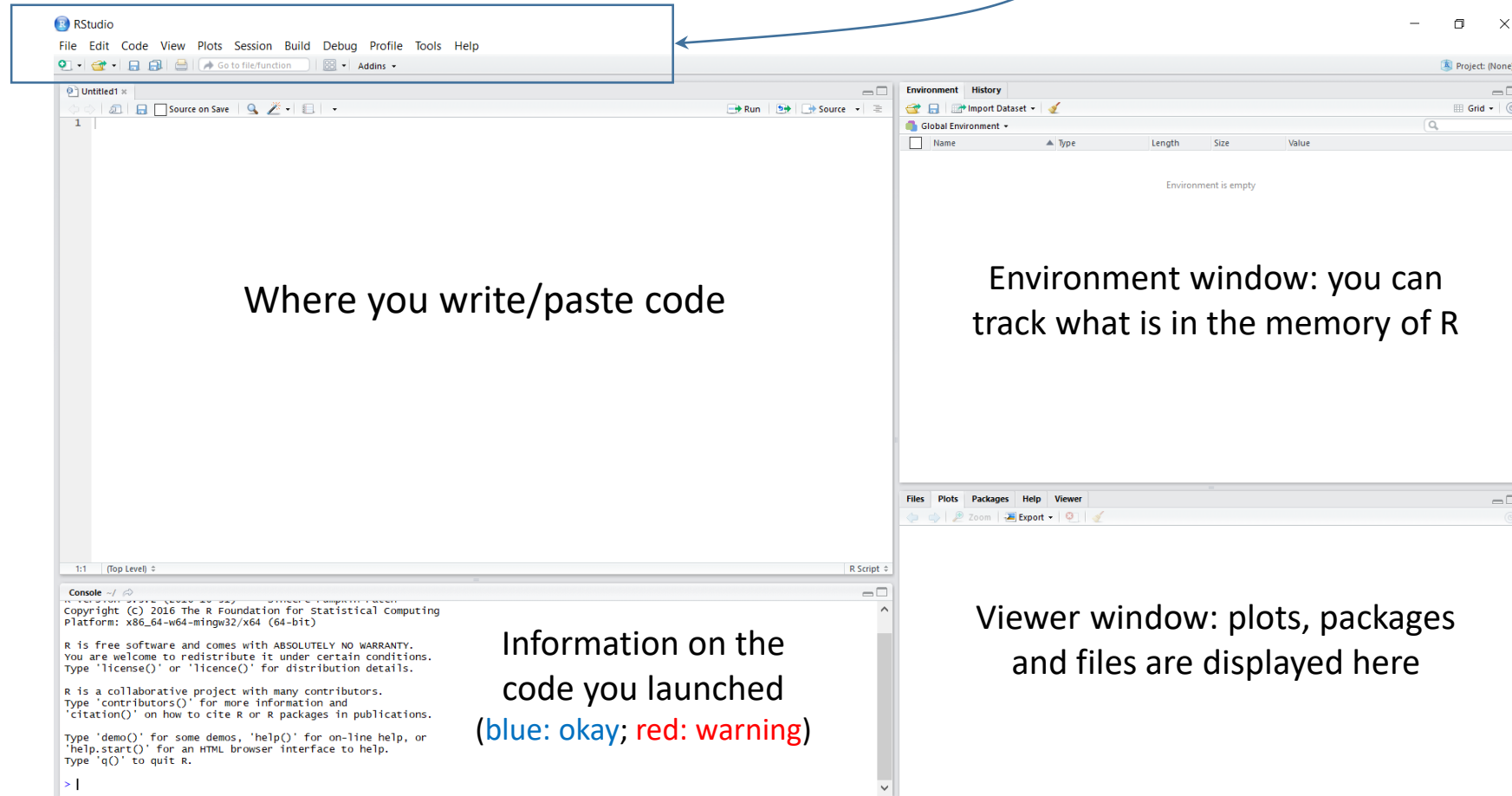


Highly recommended

	RStudio Desktop Open Source License	RStudio Desktop Commercial License	RStudio Server Open Source License	RStudio Server Pro Commercial License	RStudio Server Pro + RStudio Connect Commercial License
	FREE	\$995 per year	FREE	\$9,995 per year	\$29,995 per year
	DOWNLOAD Learn More	BUY Learn More	DOWNLOAD Learn More	DOWNLOAD Learn More	TALK Learn More
Integrated Tools for R	●	●	●	●	●
Priority Support		●		●	●

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!



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 CTRL+enter). What do you observe?

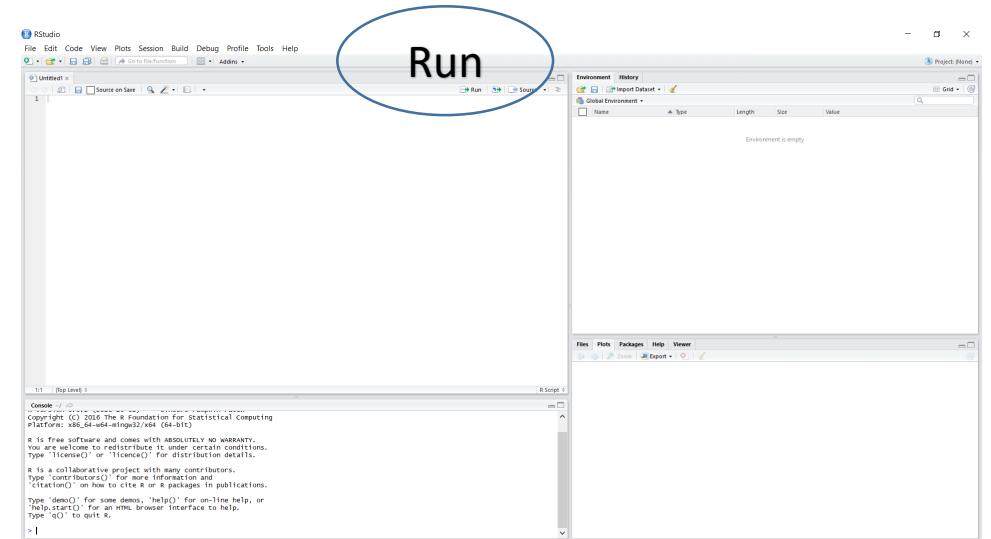
> 2+2 > 2+3;1+9

> 5+7 > LETTERS

> 40/4 > pi

> 40 / 4 > 10+2*5

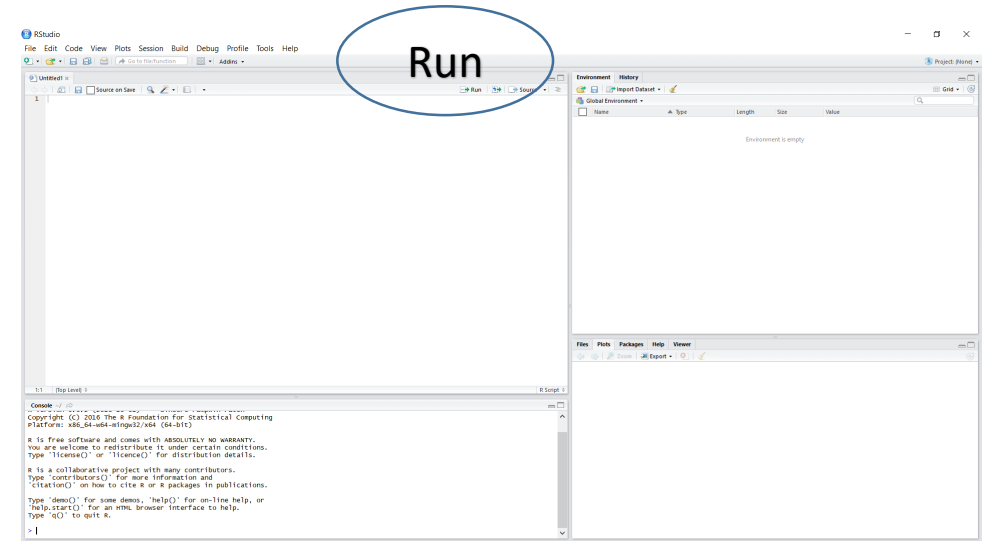
> 4 ^2 > (10+2)*5



Create some objects

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

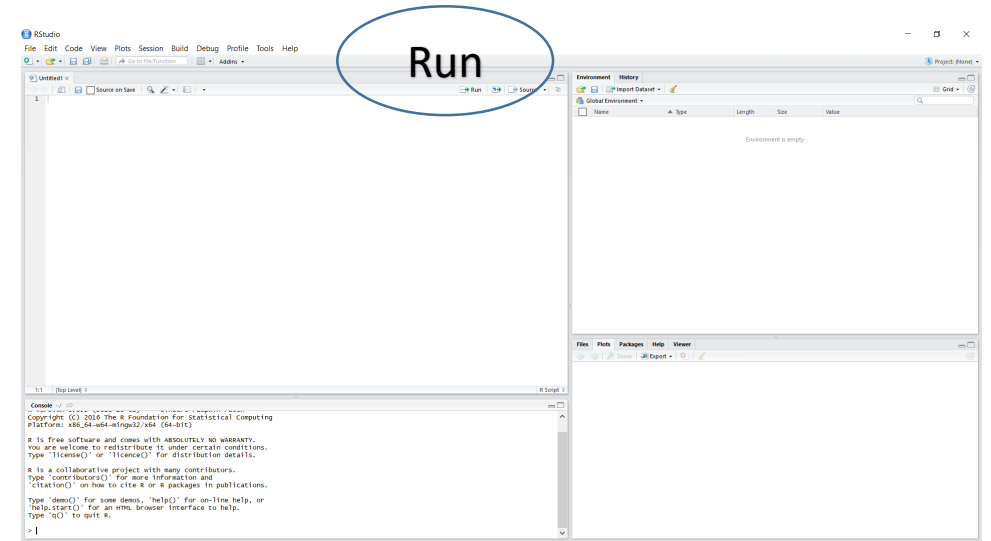
> 2+2	> 2+3;1+9	> x <- 4	> you = 2
> 5+7	> LETTERS	> x	> x + you
> 40/4	> pi	> x = 5	> y = x+you
> 40 / 4	> 10+2*5	> x*3	> y < 4
> 4 ^2	> (10+2)*5	> x + you	> y == 4



Create some objects

Type the following lines in the main window. In the end of each line, click on Run (CTRL+R or CTRL+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)
```

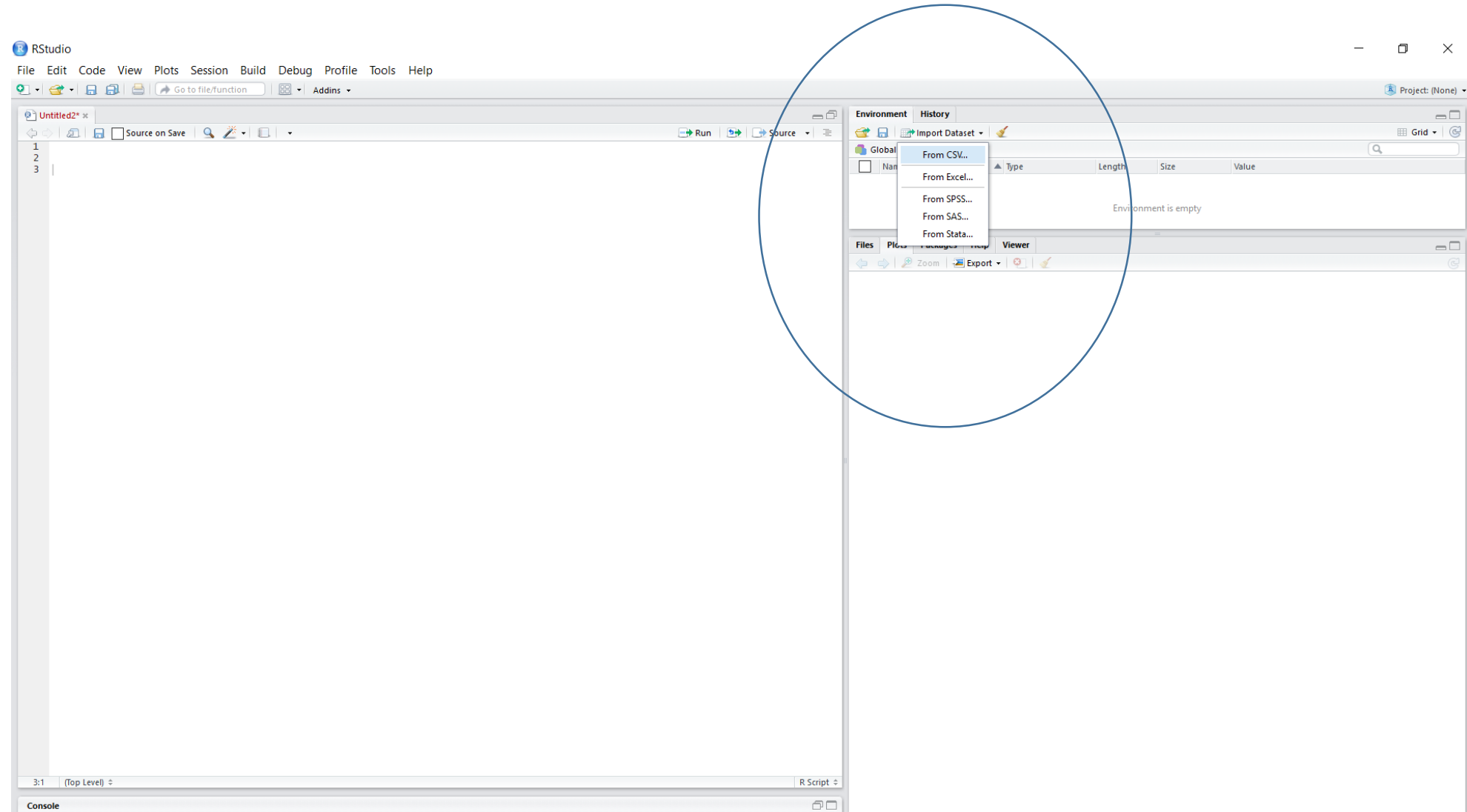
```
> df
```

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

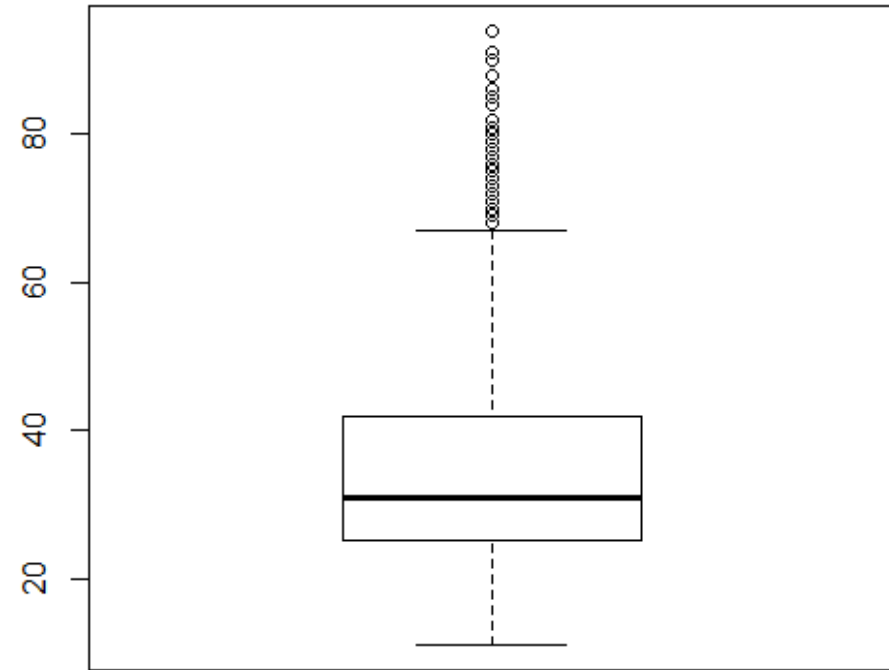
`sd(df$Age)`

#squared root of the variance

Basic graphs

```
boxplot(df$Age)
```

```
boxplot.stats(df$Age)
```



Basic graphs

```
hist(df$Age)
```

