

8.3 Come analizzare i dati: introduzione a RStudio

Insegnamento di Informatica

Elisabetta Ronchieri

Corso di Laurea di Economia, Università di Ferrara

I semestre, anno 2014-2015



Argomenti

Introduction

Tabs

R markdown

R presentation



Argomenti

Introduction

Tabs

R markdown

R presentation



Introduction

- ▶ RStudio allows the user to run R in a more user-friendly environment.
- ▶ It is open-source (i.e. free) and available at <http://www.rstudio.com/>
- ▶ For R related tutorials and/or resources see the following links:
<http://dss.princeton.edu/training/>
<http://libguides.princeton.edu/dss>
<http://tryr.codeschool.com/>



Argomenti

Introduction

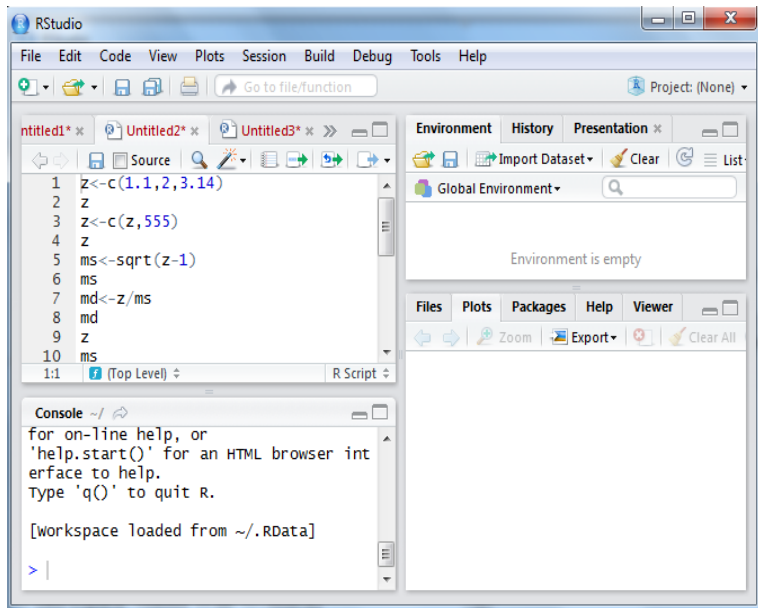
Tabs

R markdown

R presentation



Screen Shot



Screen details

- ▶ The console is where you can type commands and see output.
- ▶ The environment tab shows all the active objects.
- ▶ The history tab shows a list of commands used so far.
- ▶ The presentation tab shows the presentation output.
- ▶ The files tab shows all the files and folders in your default environment as if you were on a PC or MAC window.
- ▶ The plots tab will show all your graphics.
- ▶ The packages tab will list a series of packages or add-ons needed to run certain processes.
- ▶ For additional info see the help tab.



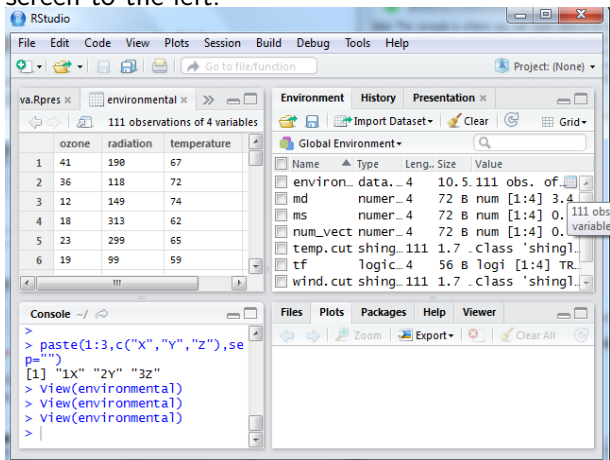
Environment tab 1

- ▶ The environment tab stores any object, value, function or anything you create during your R session.



Environment tab 2

- If you click on the dotted squares you can see the data on a screen to the left.



The screenshot shows the RStudio interface with the Environment tab selected. The 'environmental' dataset is displayed in the top-left pane, showing 111 observations of 4 variables: ozone, radiation, temperature, and an unnamed variable. The bottom-left pane shows the Console with the following commands and output:

```
> paste(1:3,c("X","Y","Z"),sep="")
[1] "1X" "2Y" "3Z"
> view(environmental)
> view(environmental)
> view(environmental)
>
```

The right pane shows the 'Global Environment' list with the following variables:

Name	Type	Length	Size	Value
environ...	data...	4	10.5	111 obs. of...
md	numer...	4	72 B	num [1:4] 3.4
ms	numer...	4	72 B	num [1:4] 0.
num_vect	numer...	4	72 B	num [1:4] 0.
temp.cut	shing...	111	1.7	..class 'shingl..
tf	logic...	4	56 B	logi [1:4] TR
wind.cut	shing...	111	1.7	..class 'shingl..



R script

- ▶ The usual Rstudio screen has four windows:
 1. Console.
 2. Environment, History and Presentation.
 3. Files, Plots, Packages, Help, Viewer.
 4. The R script(s) and data view.
- ▶ The R script is where you keep a record of your work.



R script

- ▶ To create a new R script you can go to File → New → R Script.
- ▶ You can also simply press *Ctrl + Shift + N*.
- ▶ Make sure to save the script.
- ▶ The file extension is *.R*.
- ▶ There is the place where you can type R commands and run them.
- ▶ Just leave the cursor anywhere on the line where the command is and press *Ctrl - R* or click on the Run icon above.
- ▶ Output will appear in the console below.



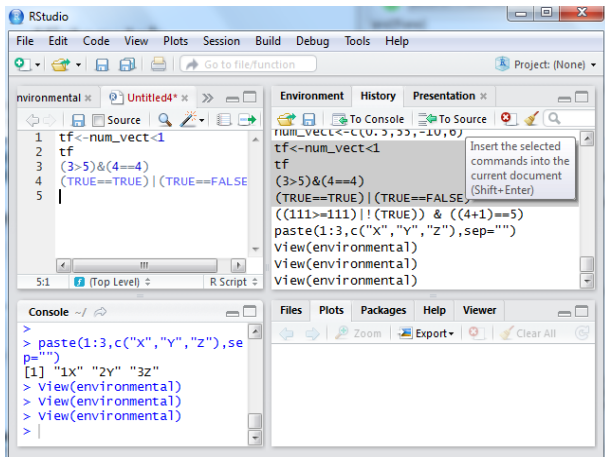
History tab 1

- ▶ The history tab keeps a record of all previous commands.
- ▶ It helps when testing and running processes.
- ▶ Here you can:
 - ▶ save the whole list;
 - ▶ select the commands you want;
 - ▶ send the commands to an R script to keep track of your work.



History tab 2

- ▶ If you select a set of code and click on the *To Source* icon, a window on the left will open with the list of commands.
- ▶ Make sure to save the "Untitled64" file as an *.R script.



Changing the working directory

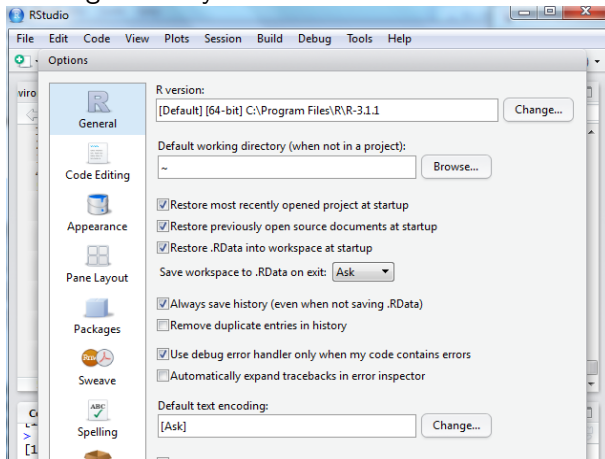
- ▶ If you have different projects you can change the working directory for that session, doing Session → Set Working Directory → Choose Directory.
- ▶ Or you can type:

```
# Shows the working directory (wd)  
> getwd()  
# Changes the wd  
> setwd("C:/myfolder/data")
```
- ▶ More info see the following document:
<http://dss.princeton.edu/training/RStata.pdf>



Setting a default working directory

- ▶ Every time you open RStudio, it goes to a default directory.
- ▶ You can change the default to a folder where you have your datafiles so you do not have to do it every time.
- ▶ In the menu go to Tools → Global Options, then set Default working directory.



Packages tab 1

- ▶ The package tab shows the list of add-ons included in the installation of RStudio.
- ▶ If checked, the package is loaded into R, if not, any command related to that package will not work, you will need select it.
- ▶ Another way to activate a package is by typing, for example, *library(foreign)*.
- ▶ You can also install other add-ons by clicking on the *Install* icon.
- ▶ You can also update already installed packages by clicking on the *Update* icon.



Packages tab 2

Environment		History	Presentation		
Files		Plots	Packages	Help	Viewer
 Install		 Update			
					
Name		Description		Version	
System Library					
	AnnotationDbi	Annotation Database Interface		1.24.0	
	AppliedPredictiveModelin	Functions and Data Sets for 'Applied Predictive Modeling'		1.1-5	
	Biobase	Biobase: Base functions for Bioconductor		2.22.0	
	BiocGenerics	Generic functions for Bioconductor		0.8.0	
	BiocInstaller	Install/Update Bioconductor and CRAN Packages		1.12.1	
	biomaRt	Interface to BioMart databases (e.g. Ensembl, COSMIC , Wormbase and Gramene)		2.18.0	
	Biostrings	String objects representing biological sequences, and matching algorithms		2.30.1	
	bitops	Bitwise Operations		1.0-6	
	boot	Bootstrap Functions (originally by Angelo Canty for S)		1.3-11	
	BradleyTerry2	Bradley-Terry Models		1.0-4	
	brglm	Bias reduction in binomial-response generalized linear models.		0.5-9	
	BSgenome	Infrastructure for Biostrings-based genome data packages		1.30.0	
	car	Companion to Applied Regression		2.0-20	
	caret	Classification and Regression Training		6.0-30	
	caTools	Tools: moving window statistics, GIF, Base64, ROC AUC, etc.		1.17	



Installing a package

- ▶ The package *rgl* is useful to plot 3D images.
- ▶ It does not come with the original R install.
- ▶ Click on *Install* icon, write the name in the pop-up window and click on *Install*.

```
Loading required package: Rcpp
> library("utils", lib.loc="C:/Program Files/R/R-3.
> library("Quandl", lib.loc="E:/R-3.0.2/libraries")
Loading required package: xts
Loading required package: zoo

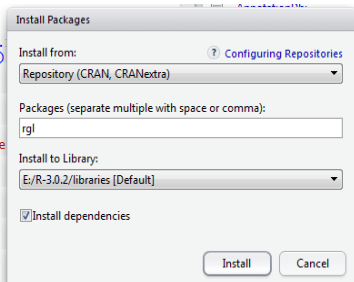
Attaching package: 'zoo'

The following objects are masked from 'package:base':
    as.Date, as.Date.numeric

warning message:
package 'Quandl' was built under R version 3.1.2
>
> install.packages("rgl")
Installing package into 'E:/R-3.0.2/libraries'
(as 'lib' is unspecified)
trying URL 'http://cran.rstudio.com/bin/windows/contrib/3.1/rgl_0.95.1158.zip'
Content type 'application/zip' length 2139462 bytes (2.0 Mb)
opened URL
downloaded 2.0 Mb

package 'rgl' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
C:\Users\ronchieri\AppData\Local\Temp\RtmpuawwUH\downloaded_packages
> |
```



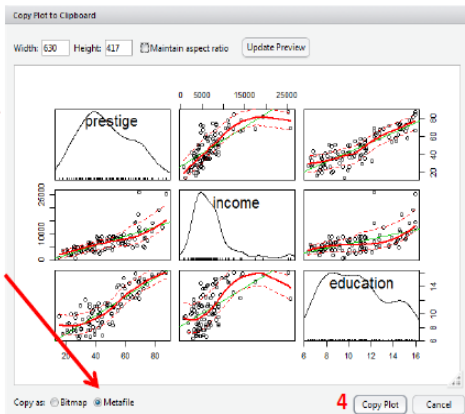
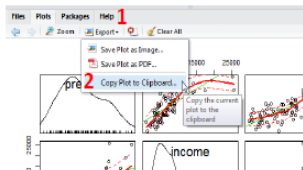
Plot 1

- ▶ The plots tab will display the graphs.
- ▶ It is possible to handle more than one graphs.
- ▶ If you want to see one graph, click on the left-arrow icon.
- ▶ To extract the graph, click on *Export* where you can save the file as an image (PNG, JPG, etc.) or as PDF.
- ▶ These options are useful when you only want to share the graph or use it in a LaTeX document.
- ▶ Probably, the easiest way to export a graph is by copying it to the clipboard and then paste it directly into your Word document.



Plot 2

2. Save plot as PDF



3 Make sure to select 'Metafile'

5 Paste it into your Word document

Plot 3

- ▶ 3D graphs can be display on a separate screen.
- ▶ You cannot be able to save it, but after moving it around, once you find the angle you want, you can screenshot it and paste it to you Word document.



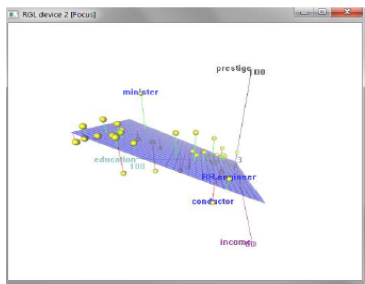
Plot 4

```
RStudio
File Edit Code View Plots Session Project Build Tools Help
[Icons] [Source on Save] [Run] [Source]
HousePets.R | MyScript.R | house.pets | A | 0 | Graphs.R
1
2 library(car) # by John Fox and Sanford Weisberg
3 library(rgl) # by Daniel Adler and Duncan Murdoch
4
5 # Scatterplot per group
6
7 scatterplot(prestige ~ income|type, boxplots=FALSE, span=0.75, data=prestige)
8
9 # Scatterplots in matrix form
10
11 scatterplotMatrix(~ prestige + income + education, span=0.7, data=prestige)
12
13 # 3D graph, scatter3d is from the rgl package. It will open in a separate window.
14
15 scatter3d(prestige ~ income + education, ld.n=3, data=duncan)
```

151 [Top Level] [R Sc]

Console RStudio/MyScript.R / 10

```
> scatterplot(prestige~income|type, boxplots=FALSE, span=0.75, data=prestige)
> scatterplotMatrix(~ prestige + income + education, span=0.7, data=prestige)
> scatter3d(prestige ~ income + education, ld.n=3, data=duncan)
>
```



Argomenti

Introduction

Tabs

R markdown

R presentation



R markdown

- ▶ R Markdown is way to write quick attractive reports that use R output.
- ▶ You write the reports in a markdown document, inserting R code where you like it (left window on RStudio).
- ▶ R then generates a final document that replaces the R code with its results (right window on RStudio).
- ▶ You can automatically update an R Markdown document whenever your data or R code changes.
- ▶ You can use R Markdown to create attractive, fully customizable, HTML, PDF, and MS Word documents as well as Beamer slides.

See <http://rmarkdown.rstudio.com/>



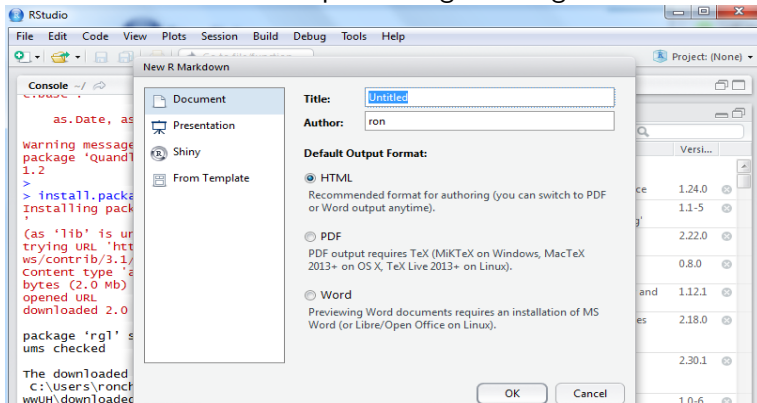
R markdown v2

- ▶ R Markdown v2 is based on knitr and pandoc packages.
- ▶ This implementation includes:
 - ▶ Create HTML, PDF, and MS Word documents as well as Beamer, ioslides and Slidy presentations;
 - ▶ New markdown syntax including expanded support for tables and bibliographies;
 - ▶ Hooks for customizing HTML and PDF output (include CSS, headers, and footers);
 - ▶ Include raw LaTeX within markdown for advanced customization of PDF output;
 - ▶ Compile HTML, PDF, or MS Word notebooks from R scripts;
 - ▶ Extensibility: create custom templates and even entirely new output formats (for example, see the Tufte Handout format);
 - ▶ Create interactive R Markdown documents using Shiny.
- ▶ When you click the "Knit" button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document.



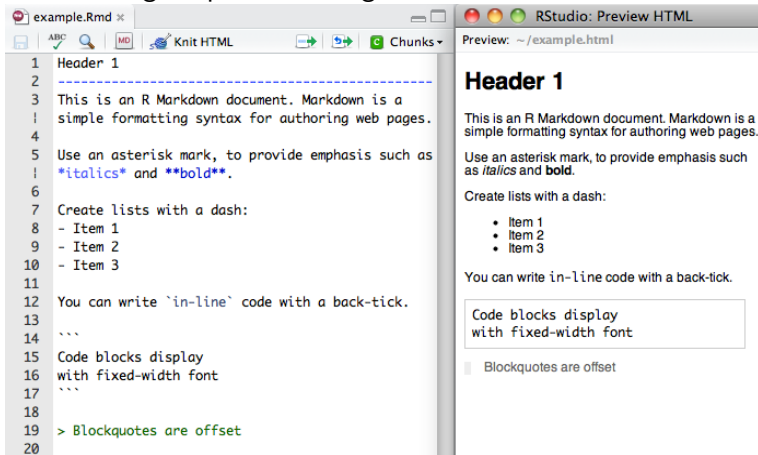
R markdown

- ▶ To create an R markdown you can go to File → New → R Markdown.
- ▶ The file extension is *.Rmd*.
- ▶ The default markdown is for Documentation.
- ▶ The knitr and pandoc packages are automatically installed when missing.
- ▶ Make sure to save the script before generating final document.



Markdown Basics

- ▶ Rather than writing complex markup code (e.g. HTML or LaTeX), Markdown enables the use of a simple syntax.
- ▶ For example the file on the left shows basic Markdown and the resulting output on the right:



The screenshot displays the RStudio interface with a file named 'example.Rmd' open. The left pane shows the source code in Markdown format, and the right pane shows the rendered HTML output.

Source Code (example.Rmd):

```
1 Header 1
2 -----
3 This is an R Markdown document. Markdown is a
4 | simple formatting syntax for authoring web pages.
5
6 Use an asterisk mark, to provide emphasis such as
7 | italics and bold.
8
9 Create lists with a dash:
10 - Item 1
11 - Item 2
12 - Item 3
13
14 You can write `in-line` code with a back-tick.
15
16 Code blocks display
17 with fixed-width font
18
19 > Blockquotes are offset
20
```

Preview (RStudio: Preview HTML):

Preview: ~/example.html

Header 1

This is an R Markdown document. Markdown is a simple formatting syntax for authoring web pages.

Use an asterisk mark, to provide emphasis such as *italics* and **bold**.

Create lists with a dash:

- Item 1
- Item 2
- Item 3

You can write in-line code with a back-tick.

```
Code blocks display
with fixed-width font
```

Blockquotes are offset



Code Chunks 1

- ▶ R Code Chunks can be embedded using the native Markdown syntax for enclosed code regions.
- ▶ For example, the following code chunk computes a data summary and renders a plot as a PNG image:

The screenshot displays the RStudio interface with two main panes. The left pane, titled 'chunks.Rmd', shows an R code chunk with the following content:

```
1 R Code Chunks
2
3
4 With R Markdown, you can insert R code
  chunks including plots:
5
6 ```{r qplot, fig.width=4, fig.height=3,
  | message=FALSE}
7 # quick summary and plot
8 library(ggplot2)
9 summary(cars)
10 qplot(speed, dist, data=cars) +
11   geom_smooth()
12
13 |
```

The right pane, titled 'RStudio: Preview HTML', shows the rendered output of the code chunk. It includes the title 'R Code Chunks', a text description, a code block for the summary, a summary table, another code block for the plot, and the resulting plot.

R Code Chunks

With R Markdown, you can insert R code chunks including plots:

```
# quick summary and plot
library(ggplot2)
summary(cars)
```

##	speed		dist	
##	Min.	: 4.0	Min.	: 2
##	1st Qu.:	12.0	1st Qu.:	26
##	Median :	15.0	Median :	36
##	Mean :	15.4	Mean :	43
##	3rd Qu.:	19.0	3rd Qu.:	56
##	Max.	: 25.0	Max.	: 120

```
qplot(speed, dist, data = cars) + geom_smooth()
```



Code Chunks 2

- For example, the following code chunk computes a data summary and renders a plot:

The screenshot displays the RStudio environment with a code editor on the left and a viewer pane on the right. The code editor shows an R code chunk with the following content:

```
1 ---
2 title: "Code Chunks"
3 author: "ron"
4 date: "Monday, December 15, 2014"
5 output: html_document
6 ---
7
8 You can embed an R code chunk like
9 this:
10
11 ```{r}
12 summary(cars)
13
14 You can also embed plots, for example:
15
16 ```{r, echo=FALSE}
17 plot(cars)
18
19
20 Note that the `echo = FALSE` parameter
21 was added to the code chunk to prevent
22 printing of the R code that generated
23 the plot.
```

The viewer pane on the right shows the rendered output of the code chunk. It includes the title "Code Chunks", the author "ron", and the date "Monday, December 15, 2014". The text "You can embed an R code chunk like this:" is followed by a code block containing the R code `summary(cars)`. Below this, the output of the `summary(cars)` function is displayed as a table:

##	speed	dist
## Min.	: 4.0	Min. : 2
## 1st Qu.:	12.0	1st Qu.: 26
## Median :	15.0	Median : 36
## Mean :	15.4	Mean : 43
## 3rd Qu.:	19.0	3rd Qu.: 56
## Max.	:25.0	Max. :120

Below the table, the text "You can also embed plots, for example:" is shown.



Code Chunks 3

- For example, the following code chunk read data from a file and perform some operations on data:

The screenshot displays the RStudio interface. On the left, the 'ProgettoX.Rmd' script is open, showing a code chunk with the following R code:

```
1 ---  
2 title: "ProgettoX"  
3 author: "Stuente"  
4 date: "Monday, December 15, 2014"  
5 output: html_document  
6 ---  
7  
8 Dopo aver settato la direttory di  
9 lavoro, caricare i dati:  
10  
11 ```{r}  
12 setwd("E:/pandora/Presentation  
13 /PresentationInformatica/corsoinformatica/progetto/dati/cambi")  
14 data <- read.csv("Progetto11.csv")  
15 ```  
16  
17 Determinare la struttura dell'oggetto:  
18  
19 ```{r}  
20 str(data)  
21 ```  
22  
23 Determinare i nomi delle variabili  
24 presenti nell'oggetto:  
25  
26 ```{r}  
27  
28 ```
```

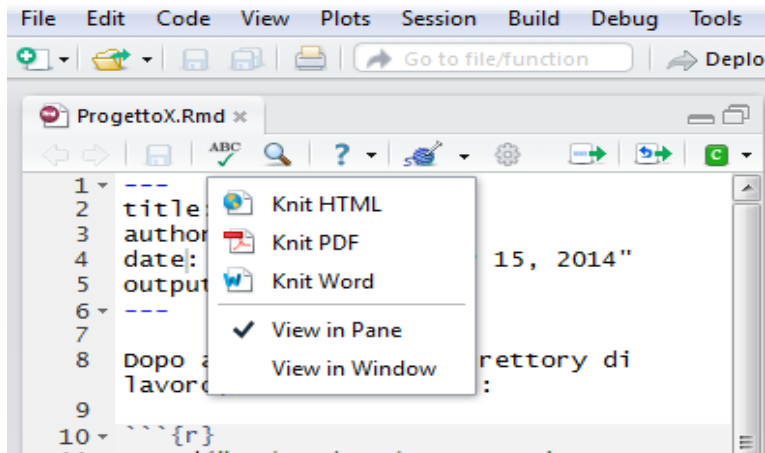
On the right, the 'Viewer' pane shows the rendered HTML output of the code chunk. It features a title 'ProgettoX', an author 'Stuente', and a date 'Monday, December 15, 2014'. The text 'Dopo aver settato la direttory di lavoro, caricare i dati:' is followed by a code block containing the R code for setting the working directory and reading the CSV file. Below this, the text 'Determinare la struttura dell'oggetto:' is followed by another code block showing the R code for printing the structure of the data object. At the bottom, the output of the R code is displayed, showing the structure of the data frame.

- Click on "Show in new window" icon to save the HTML file.



Knit button

- ▶ When you click the "Knit" button you can specify the type of document you want to be generated.



Useful documentation

Markdown basics,

http://rmarkdown.rstudio.com/authoring_basics.html

R code chunks,

[http:](http://rmarkdown.rstudio.com/authoring_rcodechunks.html)

[//rmarkdown.rstudio.com/authoring_rcodechunks.html](http://rmarkdown.rstudio.com/authoring_rcodechunks.html)

R Markdown Cheat Sheet,

<http://rmarkdown.rstudio.com/RMarkdownCheatSheet.pdf>

R Markdown Reference Guide,

[http:](http://rmarkdown.rstudio.com/RMarkdownReferenceGuide.pdf)

[//rmarkdown.rstudio.com/RMarkdownReferenceGuide.pdf](http://rmarkdown.rstudio.com/RMarkdownReferenceGuide.pdf)



Argomenti

Introduction

Tabs

R markdown

R presentation



R presentation

- ▶ R Presentations are a feature of RStudio that enable easy authoring of HTML5 presentations.
- ▶ R Presentations are based on R Markdown, and include the following features:
 - ▶ Easy incorporation of R code and its output (including plots)
 - ▶ Support for LaTeX equations using MathJax
 - ▶ Flexible two column layouts
 - ▶ Many options for slide transitions and slide navigation
 - ▶ Ability to customize the appearance of slides using CSS
 - ▶ Can be played back either inside the RStudio IDE or as standalone HTML5 presentations in a web browser.

See

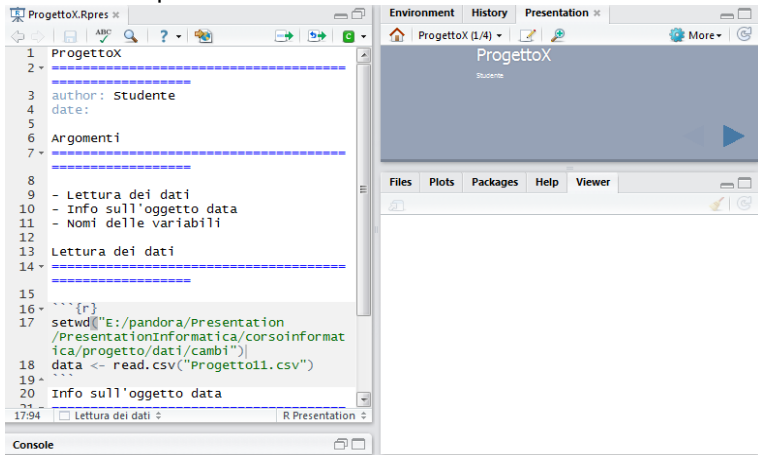
[http:](http://rmarkdown.rstudio.com/ioslides_presentation_format.html)

[//rmarkdown.rstudio.com/ioslides_presentation_format.html](http://rmarkdown.rstudio.com/ioslides_presentation_format.html)



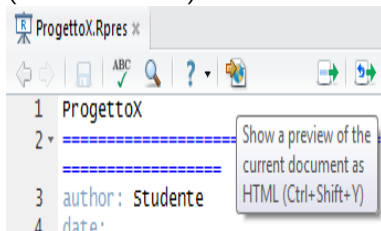
R presentation

- ▶ To create an R presentation you can go to File → New → R presentation.
- ▶ The file extension is *.Rpres*.
- ▶ The default presentation is ioslides.

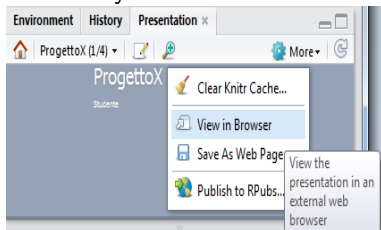


R presentation

- ▶ Click on "Show a preview of the current document as HTML (Ctrl+Shift+Y)" button.



- ▶ Click on More → view in Browser to save it as Web page, Html only.



Per ulteriori letture

-  A. M. Mineo, *Una Guida all'utilizzo dell'ambiente statistico R*
-  Vito M. R. Muggeo, Giancarlo Ferrara, *Il Linguaggio R: concetti introduttivi ed esempi*, 2005, <http://cran.r-project.org/doc/contrib/nozioniR.pdf>
-  Josef Eschgfaller, *Programmare in R*, 2005, <http://cran.r-project.org/doc/contrib/Fondamenti-0405.pdf>