



LDMI 2023



Linux
Professional
Institute

TRAINING
PARTNER

R Markdown

un'interfaccia notebook per la Literate
Programming in vari linguaggi

Autore: Bibliomike

28 Ottobre 2023

About me



Michele Bonsignore

Ci fu un tempo nostalgico in cui Fedora era solo un cappello e la buffa tuba di Mandrake una distro Linux che faceva lo sgambetto alla neonata RedHat.

Per fortuna oggi sono felicemente debianizzato e ne vado fiero.

[michelebonsignore@mastodon.uno](https://mastodon.uno/@michelebonsignore) - [OpenLinux/bibliomike@telegram](https://t.me/OpenLinux_bibliomike)
 [OpenLinux/bibliomike](https://github.com/OpenLinux/bibliomike)

Acknowledgments



Vimelug: - Vimercate Linux User Group -

👉 telegram.me/vimelug - 🌐 vimelug.org/



[The R Project for Statistical Computing](https://www.r-project.org)



qxaringan developed by Yihui Xie

qxaringanExtra developed by Garrick Anden-Buie

Note di compatibilità delle slides

- Le seguenti slides sono state generate con il pacchetto per **Xaringan** che riproduce localmente su browser un file statico in formato html5.
- È possibile esportare in modo fedele la presentazione completa in formato *.pdf mediante stampa su file da browser **Chromium** o **Brave**.
- Nella conversione si perdono ovviamente alcune funzionalità interattive fruibili solo nel formato html (es. filmanti, widget con mappe, ecc.).
- Un altro piccolo problema che emerge nella conversione in PDF è lo scarso supporto alla visualizzazione degli **emoji** da parte dei diversi visualizzatori PDF (Evince, Okular, ecc.). Da una ricerca in rete pare che il motivo sia questo:
Emojis are basically designed for screen use. They are not fully supported in the PDF standard. It may work or partially work in some applications and not in others (in pratica i visualizzatori PDF non renderizzano bene gli emoji mentre le pagine html non hanno questo problema).
- Per ragioni pratiche ho preferito quindi sostituire nella presentazione gli **emoji** generati dallo script **Rmd** con la controparte in formato *.png (<https://emoji.aranja.com/>).

 The R Project for Statistical Computing



Cos'è R?

R is a language and environment for statistical computing and graphics. It is a GNU project and is similar to the S language and environment that was developed at Bell Laboratories (formerly AT&T, now Lucent Technologies) by John Chambers and his colleagues.

R can be considered as a distinct implementation of S, developed separately from the original implementation at Bell Laboratories. Although there are some important differences between these two implementations, much code written for S runs unaltered under R.

R is available as Free Software under the terms of the Free Software Foundation's GNU General Public License in source code form. It compiles and runs on a wide variety of UNIX platforms and similar systems (including FreeBSD and Linux), Windows and MacOS.



R Foundation and R Consortium

The R Foundation is seated in Vienna, Austria and currently hosted by the Vienna University of Economics and Business. It is a registered association under Austrian law and active worldwide.

The R Foundation is a not for profit organization working in the public interest. It has been founded by the members of the R Development Core Team in order to

- Provide support for the R project and other innovations in statistical computing. We believe that R has become a mature and valuable tool and we would like to ensure its continued development and the development of future innovations in software for statistical and computational research.
- Provide a reference point for individuals, institutions or commercial enterprises that want to support or interact with the R development community.
- Hold and administer the copyright of R software and documentation.



The screenshot shows the R Consortium website. At the top, it says "THE LINUX FOUNDATION PROJECTS". Below that is the "R consortium" logo. A navigation bar includes links for Members, Projects, About, Webinars, R/Medicine, News, Code of Conduct, and social media icons for Twitter and YouTube. A search bar is also present.

R Foundation Member

Platinum Members

Silver Members

© 2023 R Consortium. © R Consortium. A Linux Foundation Project. R Consortium™ is a trademark of The Linux Foundation. All other marks and names mentioned herein may be trademarks of their respective companies.

- Promoting the growth and development of R as a leading platform for data science and statistical computing.
- Supporting and collaborating with the R Foundation, the governing body of the R Project.
- Funding projects to enhance R and support its users.



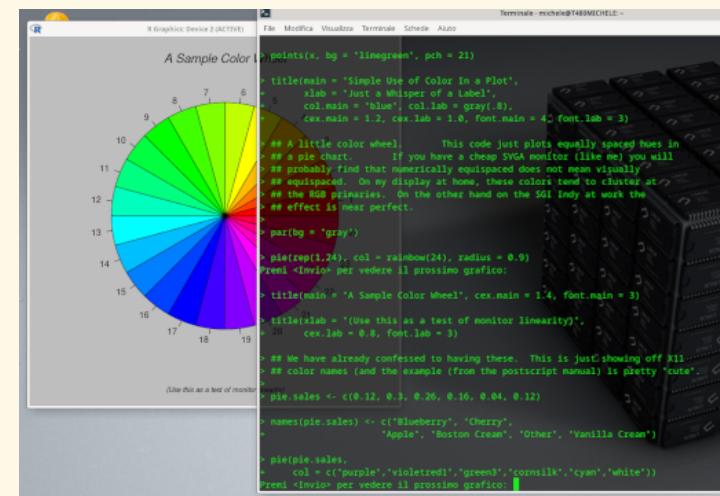
Installiamo R

- Seguire le guide disponibili su <https://cran.r-project.org/>.

A titolo di esempio riportiamo le istruzioni per l'installazione su Linux/Debian:

```
sudo apt update  
sudo apt install r-base r-base-dev
```

Per vedere se tutto funziona correttamente avviamo R da terminale con l'istruzione R, e con la funzione demo(graphics) facciamo girare un paio di esempi grafici:





I Repository di R



(The Comprehensive R Archive Network): repository ufficiale costituito da una rete mondiale di server ftp e web gestiti dalla community R. La fondazione R coordina la pubblicazione e la revisione dei pacchetti.

```
install.packages(c("vioplot", "MASS"))
```



: repository con pacchetti specifici per la Bioinformatica. Ha una community molto attiva che si occupa della revisione dei pacchetti pubblicati.

```
if (!require("BiocManager", quietly = TRUE))
  install.packages("BiocManager")
  BiocManager::install()
  #### Install specific packages, e.g., "GenomicFeatures" and "AnnotationDbi", with
  BiocManager::install(c("GenomicFeatures", "AnnotationDbi"))
```



R packages are extensions of base R that provide additional features or provide alternative functionality.

Tidyverse

Packages Blog Learn Help Contribute

R packages for data science

The tidyverse is an opinionated **collection of R packages** designed for data science. All packages share an underlying design philosophy, grammar, and data structures.

Install the complete tidyverse with:

```
install.packages("tidyverse")
```

<https://r4ds.hadley.nz/data-visualize>

Literate Programming

Donald Knuth



Donald Knuth is the inventor of TeX and literate programming, Professor Emeritus at Stanford University, winner of the ACM Turing award, and author of "The Art of Computer Programming". His favorite programming language is CWEB. Visit [Knuth's website](#) at Stanford University. Read his [biography](#), courtesy of fatbrain.com.



Literate programming

Il literate programming è un paradigma di programmazione introdotto per la prima volta nel 1984 da Donald Knuth.

Nella pratica un *programma literate* si realizza intervallando all'interno del linguaggio naturale frammenti di macro o codice sorgente tradizionale (separati da un markup primitivo), da cui è possibile generare del codice compilabile.

Gli attuali strumenti di literate programming sono indipendenti dal linguaggio di programmazione utilizzato.

Secondo Knuth, il literate programming fornisce programmi di qualità superiore, poiché obbliga i programmatore a dichiarare esplicitamente i pensieri alla base del programma, rendendo più ovvie le decisioni di progettazione scarsamente ponderate.

«Knuth ci ha mostrato come programmare intellegibilmente, ma non saggiamente. Apprezzo la disciplina. Non apprezzo il risultato. Ha creato una sorta di intricato uovo Faberge di qualità industriale, magnificamente lavorato, rifinito oltre ogni ordinario desiderio, un autentico pezzo da museo.» (Douglas McIlroy - 1986)

FREE THOUGHT



Jupyter

FREE WORLD
WWW.FSF.ORG



Jupyter Notebook

Il nome Jupyter deriva da tre linguaggi di programmazione: **Julia, Python ed R**.

È un ambiente di calcolo interattivo che consente la creazione di documenti notebook che includono codice, widget interattivi, grafici, testo narrativo, equazioni, immagini e persino video.

In Jupyter è possibile utilizzare un linguaggio di programmazione distinto per ogni documento, a scelta fra i kernel supportati: ad esempio Python, R, Go e altri ancora
- vedi [pagina wiki](#)

I file di Notebook Jupyter possono essere composti principalmente da due tipi di celle (anche se se ne possono aggiungere altre con i plugin).

- Celle di tipo Markdown per la parte "narrativa" del Notebook.
- Celle di codice: parte di codice eseguibile in una varietà di linguaggi, di default si usa Python.



Jupyter Untitled Last Checkpoint: 3 minuti fa (unsaved changes) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted | Octave O

```
1 #Esempio di utilizzo del kernel Octave in jupyter
```

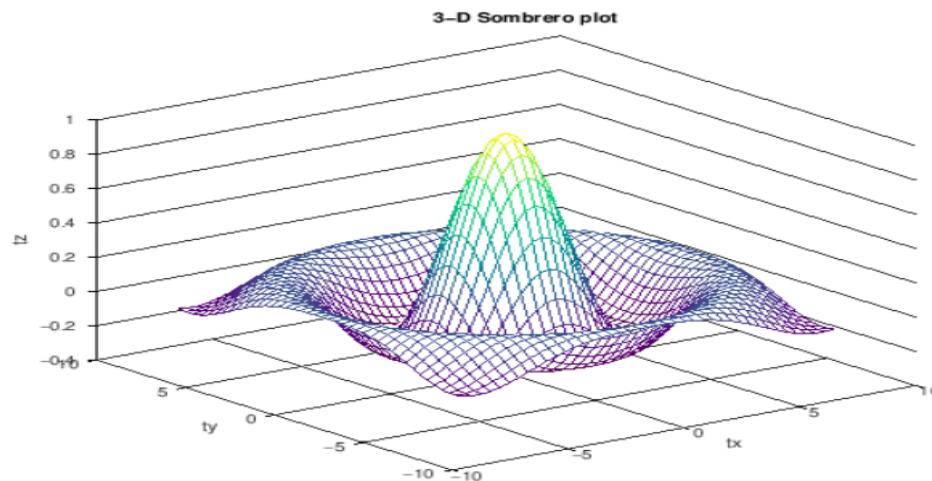
In [3]:

```
1 graphics_toolkit ("gnuplot");
2 figure (1);
3 clf ();
4 tx = ty = linspace (-8, 8, 41)';
5 [xx, yy] = meshgrid (tx, ty);
6 r = sqrt (xx .^ 2 + yy .^ 2) + eps;
7 tz = sin (r) ./ r;
8 mesh (tx, ty, tz);
9 xlabel ("tx");
10 ylabel ("ty");
11 zlabel ("tz");
12 title ("3-D Sombrero plot");
```



Esempio di utilizzo del kernel Octave in jupyter

```
In [3]: graphics_toolkit ("gnuplot");
figure (1);
clf ();
tx = ty = linspace (-8, 8, 41)';
[xx, yy] = meshgrid (tx, ty);
r = sqrt (xx .^ 2 + yy .^ 2) + eps;
tz = sin (r) ./ r;
mesh (tx, ty, tz);
xlabel ("tx");
ylabel ("ty");
zlabel ("tz");
title ("3-D Sombrero plot");
```





RMarkdown

RMarkdown è un altro popolare strumento di *literate programming*.

A differenza di Jupyter è possibile inserire nello stesso notebook codice di altri linguaggi (SQL, bash, Python, ecc.).

In genere, i file RMarkdown vengono modificati utilizzando l'ambiente di sviluppo RStudio.

Il libro [R for Data Science](#) è un buon riferimento per RMarkdown:
<https://r4ds.had.co.nz/r-markdown.html>

Tratteremo ampiamente questo argomento nelle prossime slide.



Markdown ed RMarkdown



Markdown

Markdown è un linguaggio di markup con una sintassi di testo molto semplice progettato per la rapida conversione in altri formati.

È stato progettato per essere leggibile così com'è, senza che sembri marcato con tags o istruzioni di formattazione

Il linguaggio è stato sviluppato nel 2004 da John Gruber con il contributo sostanziale di Aaron Swartz.

Siti come GitHub, Reddit, Diaspora, Stack Exchange, OpenStreetMap, e SourceForge usano varianti di Markdown per facilitare le discussioni fra gli utenti.



MD Markdown in pillole

HEADERS

```
# This is an <h1> tag  
## This is an <h2> tag  
##### This is an <h6> tag
```

EMPHASIS

```
*This text will be italic*  
This will also be italic  
  
**This text will be bold**  
This will also be bold  
  
*You **can** combine them*
```

BLOCKQUOTES

```
As Grace Hopper said:  
  
> I've always been more interested  
> in the future than in the past.
```

```
As Grace Hopper said:  
  
I've always been more interested  
in the future than in the past.
```

LISTS

Unordered

- * Item 1
- * Item 2
 - * Item 2a
 - * Item 2b

Ordered

1. Item 1
2. Item 2
3. Item 3
 - * Item 3a
 - * Item 3b

IMAGES

```
![GitHub Logo](/images/logo.png)
```

Format: ! [Alt Text] (url)

LINKS

<http://github.com> - automatic!

[GitHub](http://github.com)

BACKSLASH ESCAPES

Markdown allows you to use backslash escapes to generate literal characters which would otherwise have special meaning in Markdown's formatting syntax.

```
\*literal asterisks\*
```

```
*literal asterisks*
```

Markdown provides backslash escapes for the following characters:

\ backslash	() parentheses
' backtick	# hash mark
* asterisk	+ plus sign
_ underscore	- minus sign (hyphen)
{ curly braces	. dot
[square brackets	! exclamation mark



USERNAME @Mentions

Typing an `@` symbol, followed by a username, will notify that person to come and view the comment. This is called an “@mention”, because you’re mentioning the individual. You can also @mention teams within an organization.

ISSUE REFERENCES

Any number that refers to an Issue or Pull Request will be automatically converted into a link.

```
#1
github-flavored-markdown#1
defunkt/github-flavored-markdown#1
```

EMOJI

To see a list of every image we support, check out www.emoji-cheat-sheet.com

GitHub supports emoji!
`:+1:` `:sparkles:` `:camel:` `:tada:`
`:rocket:` `:metal:` `:octocat:`

GitHub supports emoji!



FENCED CODE BLOCKS

Markdown converts text with four leading spaces into a code block; with GFM you can wrap your code with ````` to create a code block without the leading spaces. Add an optional language identifier and your code will get syntax highlighting.

```
```javascript
function test() {
 console.log("look ma', no spaces");
}
...
```

```

```
function test() {
  console.log("look ma', no spaces");
}
```

TASK LISTS

- [x] this is a complete item
- [] this is an incomplete item
- [x] @mentions, #refs, [links](), **formatting**, and tags supported
- [x] list syntax required (any unordered or ordered list supported)

- this is a complete item
- this is an incomplete item
- @mentions, #refs, links, formatting, and tags supported
- list syntax required (any unordered or ordered list supported)

TABLES

You can create tables by assembling a list of words and dividing them with hyphens `-` (for the first row), and then separating each column with a pipe `|`:

| First Header | Second Header |
|------------------|------------------|
| Content cell 1 | Content cell 2 |
| Content column 1 | Content column 2 |

| First Header | Second Header |
|------------------|------------------|
| Content cell 1 | Content cell 2 |
| Content column 1 | Content column 2 |

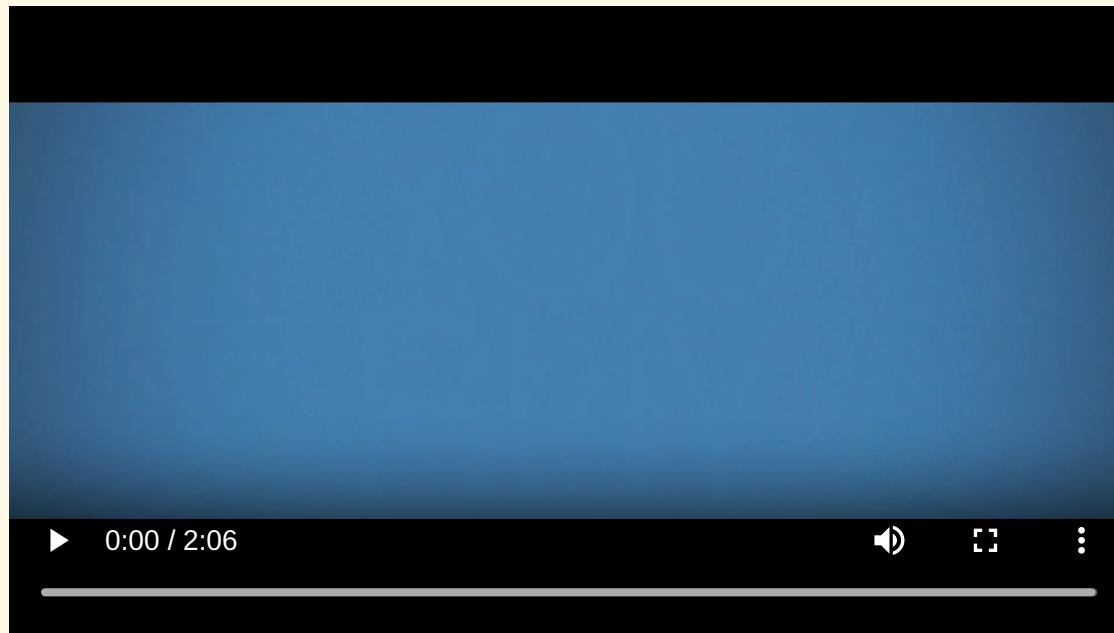
R Markdown





RMarkdown

Rmarkdown è l'interfaccia *notebook* del linguaggio R: aggiunge alla sintassi Markdown funzionalità avanzate di scripting in vari linguaggi (es. R, Python, SQL)



https://rmarkdown.rstudio.com/authoring_quick_tour.html

<https://bookdown.org/yihui/rmarkdown/>



Origini di Rmarkdown

Il formato documentale R markdown venne introdotto per la prima volta nel pacchetto knitr nel lontano 2012. R Markdown si basa sul pacchetto knitr e sul software pandoc. knitr si occupa di eseguire il codice di programmazione e di tradurlo in Markdown. Pandoc trasforma Markdown nel formato di output desiderato (es. PDF, HTML, Word, ecc.).

