

# Why R? (CMP595 PPGC/INF/UFRGS)

Lucas Mello Schnorr, Jean-Marc Vincent

INF/UFRGS  
Porto Alegre, Brazil – October 20th, 2017



# Why R?

R is a great language for data analysis and statistics

- ▶ Open-source and multi-platform
- ▶ Very expressive with high-level constructs
- ▶ Excellent graphics
- ▶ Widely used in academia and business
- ▶ Very active community
  - ▶ Documentation, FAQ on <http://stackoverflow.com/questions/tagged/r>
- ▶ Great integration with other tools

# Why is such R a pain for computer scientists?

- ▶ R is **not** really a **programming** language
- ▶ Documentation is for statisticians
- ▶ Default plots are *cumbersome* (meaningful)
- ▶ Summaries are *cryptic* (precise)
- ▶ **Steep learning curve** even for us, computer scientists whereas we generally switch seamlessly from a language to another! That's frustrating! ;)

# Do's and don'ts

*R is high level, I'll do everything myself*

- ▶ CTAN comprises 4,334 T<sub>E</sub>X, L<sup>A</sup>T<sub>E</sub>X, and related packages and tools. Most of you do not use plain T<sub>E</sub>X.
- ▶ Currently, the CRAN package repository features 4,030 available packages.
- ▶ How do you know which one to use??? Many of them are highly exotic (not to say useless to you).

I learnt with <http://www.r-bloggers.com/>

- ▶ Lots of introductions but not necessarily what you're looking for so I'll give you a short tour.

You should quickly realize though that you need proper training in statistics and data analysis if you do not want tell nonsense.

- ▶ Again, you should read Jain's book on The Art of Computer Systems Performance Analysis
- ▶ You may want to follow online courses:
  - ▶ <https://www.coursera.org/course/compdata>
  - ▶ <https://www.coursera.org/course/repdata>

# Install and run R on debian

```
apt-cache search r
```

Err, that's not very useful :) It's the same when searching on google but once the filter bubble is set up, it gets better...

```
sudo apt-get install r-base
```

R

```
R version 3.2.0 (2015-04-16) -- "Full of Ingredients"  
Copyright (C) 2015 The R Foundation for Statistical Computing  
Platform: x86_64-pc-linux-gnu (64-bit)
```

```
R is free software and comes with ABSOLUTELY NO WARRANTY.  
You are welcome to redistribute it under certain conditions.  
Type 'license()' or 'licence()' for distribution details.
```

```
R is a collaborative project with many contributors.  
Type 'contributors()' for more information and  
'citation()' on how to cite R or R packages in publications.
```

```
Type 'demo()' for some demos, 'help()' for on-line help, or  
'help.start()' for an HTML browser interface to help.  
Type 'q()' to quit R.
```

```
>
```

## Install a few cool packages

R has it's own package management mechanism so just run R and type the following commands:

- ▶ `ddply`, `reshape` and `ggplot2` by Hadley Wickham (<http://had.co.nz/>)  
`install.packages("plyr")`  
    # or better: `install.packages("dplyr")`  
`install.packages("reshape")`  
    # or better; `install.packages("tidyr")`  
`install.packages("ggplot2")`
- ▶ `knitr` by (Yihui Xie) <http://yihui.name/knitr/>  
`install.packages("knitr")`

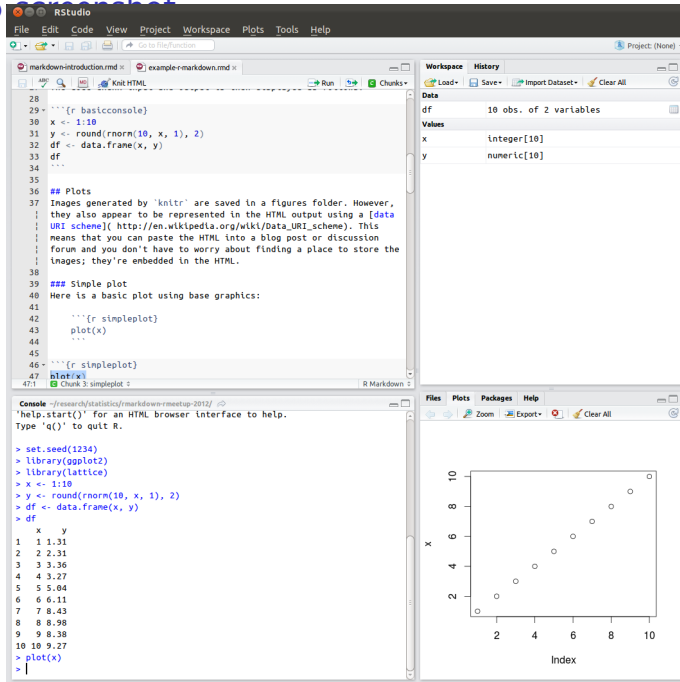
# IDE

Using R interactively is nice but quickly becomes painful so at some point, you'll want an IDE.

Emacs is great but you'll need *Emacs Speaks Statistics*

```
sudo apt-get install ess
```

In this tutorial, I will briefly show you **rstudio**  
(<https://www.rstudio.com/>) and later how to use org-mode





# Reproducible analysis in Markdown + R

- ▶ Create a new **R Markdown** document (Rmd) in rstudio
- ▶ R chunks are interspersed with "`{r}`" and "`"`"
- ▶ Inline R code: `'r sin(2+2)'`
- ▶ You can **knit** the document and share it via **rpubs**
- ▶ R chunks can be sent to the top-level with **Alt-Ctrl-c**
- ▶ I usually work mostly with the current environment and only knit in the end
- ▶ Other engines can be used (use rstudio **completion**)

```
'''{r engine='sh'}  
ls /tmp/  
'''
```
- ▶ Makes **reproducible analysis as simple as one click**
- ▶ Great tool for quick analysis for self and colleagues, homeworks, ...