

How to Climb the R Learning Curve Without Falling Off the Cliff



Tony Fujs

Latin American Youth Center
tony@layc-dc.org
@tonyfujs



Will Fenn

Innovation Network
wfenn@innonet.org
@wtfenn

Ann Emery

Innovation Network
aemery@innonet.org
@annkemery

Link to the presentation slides and code:

<https://github.com/tonyfujs/AEA2013>

Where do I start?

1. Download and install R: <http://cran.r-project.org/>
2. Install R Studio (A user friendly interface for R): <http://www.rstudio.com/ide/download/>
3. Become familiar with the R environment: <http://tryr.codeschool.com/>
4. Learn how to install packages with R Studio: <http://www.youtube.com/watch?v=u1r5XTqrCTQ>
5. Take a free on-line class:
 - Beginner: <https://www.coursera.org/course/stats1>
 - Intermediate: <https://www.coursera.org/course/dataanalysis>
 - Advanced: <https://www.coursera.org/course/compdata>
6. Get started with your own "toy project" and stick to it!
There is a wealth of resources on-line and it's easy to get overwhelmed.
7. Minimize frustration by copy-pasting existing code for your "toy project"
Don't forget: "Google is your friend!"
8. Expect some frustration!

Additional resources:

- Download: <http://cran.r-project.org/doc/contrib/Short-refcard.pdf>
- Getting starting with R? Chances are what you are trying to do has already been documented <http://www.statmethods.net/> or <http://www.cookbook-r.com/>
- Watch Introduction to R, a video series by Google:
<http://www.youtube.com/playlist?list=PLOU2XLYxmsIK9qQfztXeybpHvru-TrqAP>
- 2 minutes video series: <http://www.twotorials.com/>
- Tutorial: http://www.computerworld.com/s/article/9239625/Beginner_s_guide_to_R_Introduction

Data analysis workflow

Import data

- From a flat file (Excel, .csv, etc.): <http://www.statmethods.net/input/importingdata.html>
- Importing from other statistical system
- Package "foreign": <http://is-r.tumblr.com/post/37181850668/reading-writing-stata-dta-files-with-foreign>
- Relational databases
- Package "RODBC": <http://www.statmethods.net/input/dbinterface.html>
- Grab data from the Web (Web scraping): <http://www.programmingr.com/content/webscraping-using-readlines-and-rcurl/>
- cran.r: <http://cran.r-project.org/doc/manuals/r-release/R-data.html#Image-files>

Data management

- Basic data management tasks: <http://www.statmethods.net/management/variables.html>
- Reshape your data set
- Package "reshape2": <https://dl.dropboxusercontent.com/u/105906695/AEAslides/reshapeMap.png>
- More complex recoding & aggregation
- Package "plyr": <http://www.jstatsoft.org/v40/i01/paper>
- Package "sqldf" (Use SQL statements to manipulate R data set): <http://brusers.tumblr.com/>

Data analysis

- Basic statistics: <http://www.statmethods.net/stats/descriptives.html>
- Exploratory analysis / Data mining
- Package: "Rattle": <http://rattle.togaware.com/>
- Also a very nice tool to learn R. Rattle is a "point and click" interface, making possible to use R without writing code.
- More advanced statistical analysis for social sciences
- Package "Zelig": <http://zeligdev.github.io/files/zelig.pdf>
- "Zelig" is a wrapper for many types of statistical modeling in R. It provides a single interface for modeling, simulation and interpreting the data.

Data visualization

- Basic charts: <http://www.cookbook-r.com/Graphs/>
- Using the grammar for graphics:
- Package "ggplot2": <http://docs.ggplot2.org/current/>
- Going interactive:
- Package "shiny": <http://www.rstudio.com/shiny/>

Other:

- Creating slides with R and Rstudio: <http://www.rstudio.com/ide/docs/presentations/overview>