

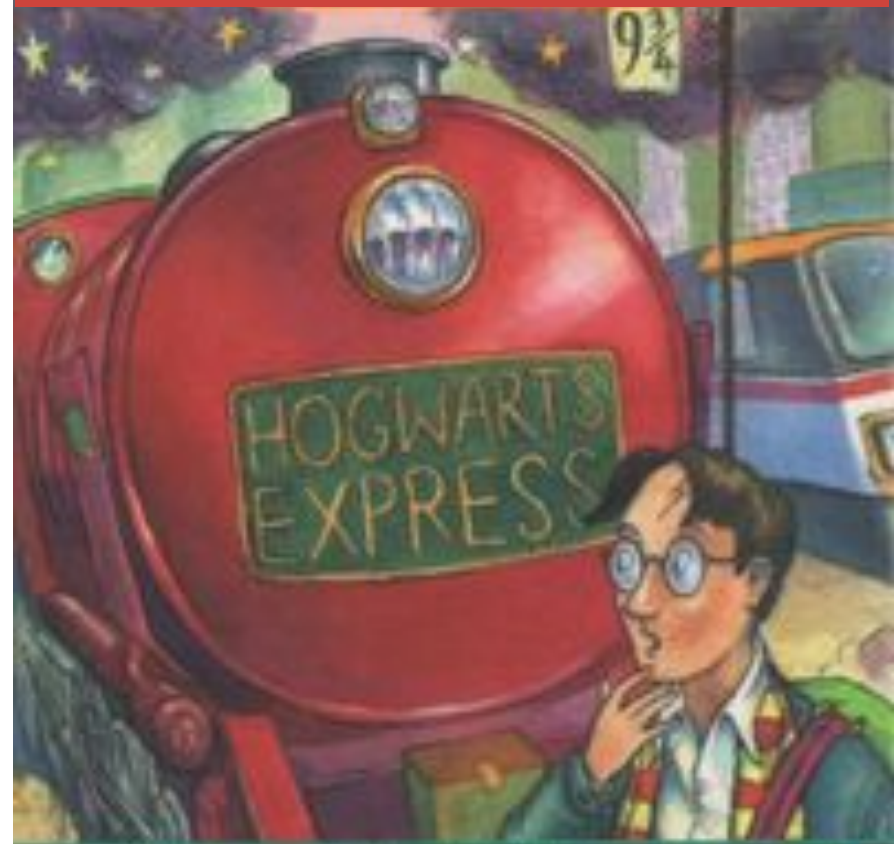
# A very brief introduction to R

**-Andrew Schuh**

*Large portions blatantly plagiarized from: Matthew Keller who cribbed his material from UCLA Academic Technology Services Technical Report Series (by Patrick Burns) and presentations (found online) by Bioconductor, Wolfgang Huber and Hung Chen, & various Harry Potter websites*

R programming language is a lot like magic...  
except instead of spells you have functions.

R, And the Rise of the Best Software Money Can't Buy



"...this is a terrific book." *The Sunday Telegraph*



=



**muggle**

SPSS and SAS users are like muggles. They are limited in their ability to change their environment. They have to rely on algorithms that have been developed for them. The way they approach a problem is constrained by how SAS/SPSS employed programmers thought to approach them. And they have to pay money to use these constraining algorithms.



wizard

R users are like wizards. They can rely on functions (spells) that have been developed for them by statistical researchers, but they can also create their own. They don't have to pay for the use of them, and once experienced enough (like Dumbledore), they are almost unlimited in their ability to change their environment.

# History of R

- S: language for data analysis developed at Bell Labs circa 1976
- Licensed by *AT&T/Lucent* to *Insightful Corp.*  
Product name: *S-plus*.
- R: initially written & released as an open source software by Ross Ihaka and Robert Gentleman at U Auckland during 90s (R plays on name “S”)
- Since 1997: international R-core team ~15 people & 1000s of code writers and statisticians happy to share their libraries! AWESOME!

# What is it?

- R is an interpreted computer language.
  - Most user-visible functions are written in R itself, calling upon a smaller set of internal primitives.
  - It is possible to interface procedures written in C, C+, or FORTRAN languages for efficiency, and to write additional primitives.
  - System commands can be called from within R
- R is used for data manipulation, statistics, and graphics. It is made up of:
  - operators (+ - <- \* %\*% ... ) for calculations on arrays & matrices
  - large, coherent, integrated collection of functions
  - facilities for making unlimited types of publication quality graphics
  - user written functions & sets of functions (packages); 800+ contributed packages so far & growing

# R

## Advantages

- Fast and free.
- State of the art: Statistical researchers provide their methods as R packages. SPSS and SAS are years behind R!
- 2<sup>nd</sup> only to MATLAB for graphics.
- WinBugs, GRASS, and other programs use or will use R.
- Active user community
- Excellent for simulation, programming, computer intensive analyses, etc.
- Interfaces with database storage software (SQL), plr & ajax

## Disadvantages

# R

## Advantages

- Fast and free.
- State of the art: Statistical researchers provide their methods as R packages. SPSS and SAS are years behind R!
- 2<sup>nd</sup> only to MATLAB for graphics.
- WinBugs, GRASS, and other programs use or will use R.
- Active user community
- Excellent for simulation, programming, computer intensive analyses, etc.
- Interfaces with database storage software (SQL), plr & ajax

## Disadvantages

- Not user friendly @ start - steep learning curve, minimal GUI.
- No commercial support; figuring out correct methods or how to use a function on your own can be frustrating.
- Easy to make mistakes and not know.
- Working with large datasets is limited by RAM
- Data prep & cleaning can be messier & more mistake prone in R vs. SPSS or SAS
- Some users complain about hostility on the R listserve



# Learning R....



# R-help listserve....



# R Commercial packages

- Many different datasets (and other “objects”) available at same time
  - Datasets can be of any dimension
  - Functions can be modified
  - Experience is interactive-you program until you get exactly what you want
  - One stop shopping - almost every analytical tool you can think of is available
  - R is free and will continue to exist. Nothing can make it go away, its price will never increase.
- One datasets available at a given time
  - Datasets are rectangular
  - Functions are proprietary
  - Experience is passive-you choose an analysis and they give you everything they think you need
  - Tend to be have limited scope, forcing you to learn additional programs; extra options cost more and/or require you to learn a different language (e.g., SPSS Macros)
  - They cost money. There is no guarantee they will continue to exist, but if they do, you can bet that their prices will always increase

# There are over 1000 add-on packages

(<http://cran.r-project.org/src/contrib/PACKAGES.html>)

- This is an enormous advantage - new techniques available without delay, and they can be performed using the R language you already know.
- Allows you to build a customized statistical program suited to your own needs.
- Downside = as the number of packages grows, it is becoming difficult to choose the best package for your needs, & QC is an issue.

# R packages of particular interest

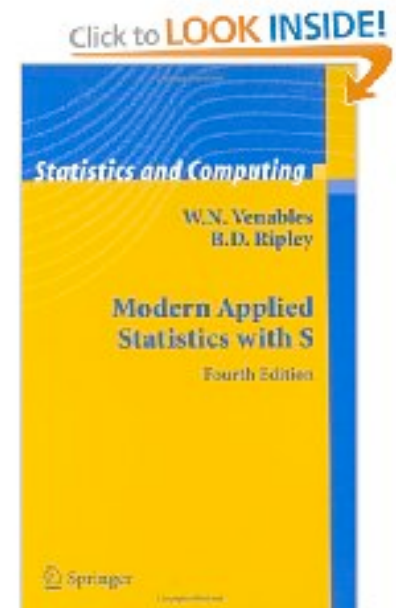
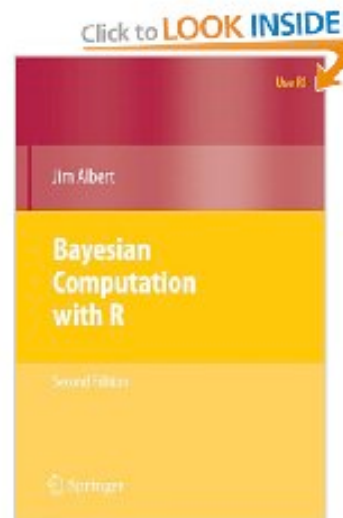
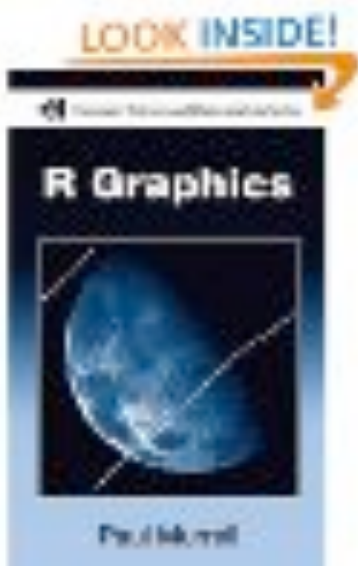
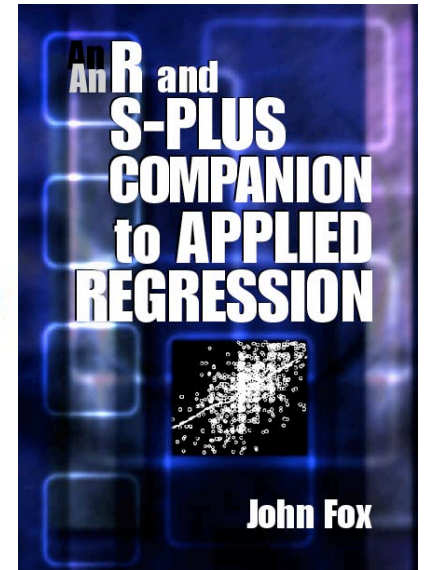
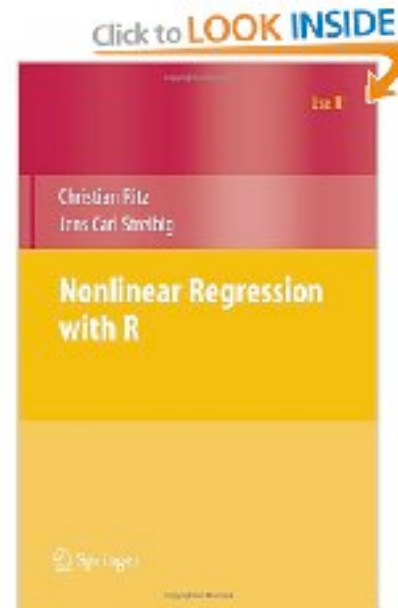
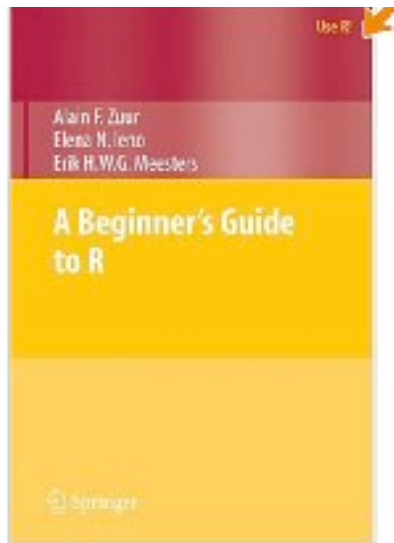
Package	Uses
ncdf/Rnetcdf/hdf5	Wrappers to netcdf libraries, allows easy data access to netcdf files
rgdal	Wrapper to proj4 projections library, allowing reprojecting across many projections
sp	Library of spatial data classes
fields	(NCAR) Doug Nychka's library of spatial analysis functions
foreach, Rmpi, snow,...	Parallel processing
Rgl	Interface to OpenGL capabilities

# R Objects

- Almost all things in R – functions, datasets, results, etc. – are OBJECTS.
  - (graphics are written out and are not stored as objects)
- Script can be thought of as a way to make objects. Your goal is usually to write a script that, by its end, has created the objects (e.g., statistical results) and graphics you need.
- Objects are classified by two criteria:
  - MODE: how objects are stored in R - character, numeric, logical, factor, list, & function
  - CLASS: how objects are treated by functions (important to know!) - [vector], matrix, array, data.frame, & hundreds of special classes created by specific functions

**....break....let's see  
some actual code  
running.....**

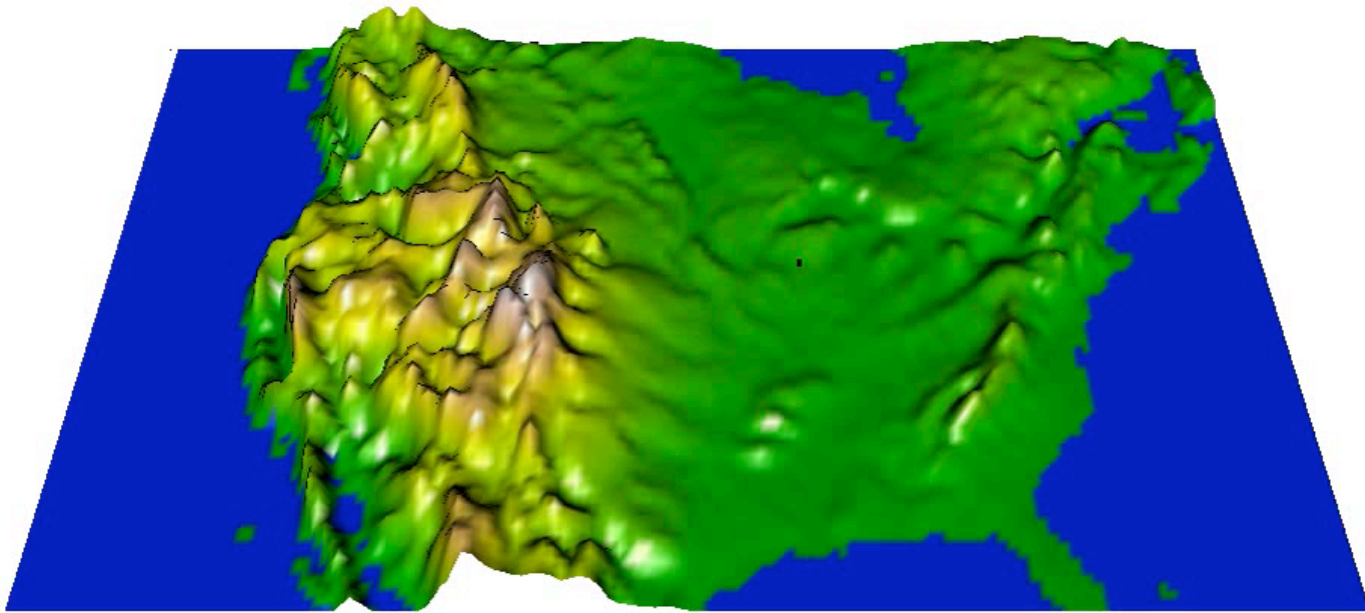
# Books? Tons...



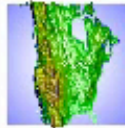


Some eye candy made by  
looping over viewer's  
“perspective” using rgl (OpenGL)  
package and SiB/RAMS/LPDM  
output

20 Days of Backward *Transport*  
*WLEF Tower Park Falls WI, USA*



*Initial Simulation of CO<sub>2</sub>, April – September 2007  
(0 meters to 5000 meters above terrain)*



# Final Words of Warning

- “Using R is a bit akin to smoking. The beginning is difficult, one may get headaches and even gag the first few times. But in the long run, it becomes pleasurable and even addictive. Yet, deep down, for those willing to be honest, there is something not fully healthy in it.” --Francois Pinard

