

# A Practical Introduction to the R Programming Language

Irucka Embry, EIT  
&  
Jennifer Murphy

Wednesday, 21 August 2013  
Updated 25 October 2013  
Revised 22 June 2014

# Outline

- R Resources page
- Why use R
- Reasons to NOT use Microsoft Excel
- Introduction to R and computer programming
- Applications of R
- Useful R Resources

## R Resources pages

- <http://www.ecoccs.com/RandUSGS.html> Created by Irucka Embry for the USGS (wealth of useful links)
- <http://www.ecoccs.com/tsuresearch.html> Created by Irucka Embry for TSU (wealth of useful links)

# Why use R

- Free (freedom and price)
- Derived from S/S-PLUS and thus built on a powerful programming language
- Publication quality graphics
- Large global community of helpful users
- There are over 5,653 ways to extend base R through packages (22 June 2014)
- Easy to create your own functions

# Reasons to NOT use Microsoft Excel

- Easy to mess up your data and not know how you did it (incorrect keyboard entry, for example)
- No log or record of the work/tasks you've performed on your data (unlike writing a script in R)
- Can be cumbersome to run analysis on subsets of data, especially if the subsets are not uniform
- Output of analysis from Excel can also be cumbersome (the results from the histogram analysis, for example)
- Some of the statistical procedures are not well documented (difficult to determine how something is actually being calculated)

# Reasons to NOT use Microsoft Excel 2

- Lacks publication quality graphics
- Can not non-destructively alter your data
- <http://blog.revolutionanalytics.com/2013/04/more-reasons-not-to-use-excel-for-modeling.html>  
More reasons not to use Excel for modeling
- [http://en.wikibooks.org/wiki/Statistics:Numerical\\_Methods/Numerics\\_in\\_Excel](http://en.wikibooks.org/wiki/Statistics:Numerical_Methods/Numerics_in_Excel)  
Statistics/Numerical Methods/Numerics in Excel
- [https://answers.microsoft.com/en-us/office/forum/officeversion\\_other-excel/excel-r-squared-is-incorrect/1dd48555-5f9e-41db-a0d9-b42d95f85499](https://answers.microsoft.com/en-us/office/forum/officeversion_other-excel/excel-r-squared-is-incorrect/1dd48555-5f9e-41db-a0d9-b42d95f85499)  
Excel R squared is Incorrect - Microsoft Community (2011)

# Introduction to R

- To start R in a \*nix based environment
  - Type R from the command line interface (CLI)
  - Click on the icon for a R GUI [ex. RStudio, etc.]
- To start R in a Microsoft Windows
  - Type R from the cmd.exe prompt
  - Click on the icon for a R GUI [ex. R 32-bit or R 64-bit, RStudio, etc.]

# Introduction to R 2

- R is **case sensitive** (this is not a native Windows application)
- `getwd()`
  - This command provides the current working directory (where R will save all data)
- `setwd("filepath_of_working_directory")`
  - This command allows the user to set the current working directory (where R will save all data)
- `library()`
  - This command allows the user to see a list of the installed R packages



# Introduction to R 3

- `help(topic)`
  - Help for the “topic”
  - `help(plot)` provides help on the `plot` function within the R base package graphics
- `?topic`
  - Help for the “topic”
  - `?plot` provides help on the `plot` function within the R base package graphics
- `help.start()`
  - Starts the HTML version of help

# Introduction to R 4

- `apropos("topic")`
  - The names of all objects in the search list that match the regular expression “topic”
  - `apropos("plot")`

# Introduction to R 5

- `x <- c(2:4) #` this is the variable named x
  - “#” for comments in R
  - “<-” is similar to = in other languages
    - “<-” is left assignment
    - “=” is left assignment (**NOT** recommended)
  - “:” for sequence with a step size of 1
- `x #` in R
- `[1] 2 3 4 #` in R
  - `[1]` is row number 1

# Introduction to R 6: Some Common Mathematical Operations

Mathematical Operation	R expression
$x - y$	<code>x - y</code>
$x + y$	<code>x + y</code>
$xy$	<code>x * y</code>
$x/y$ (fraction)	<code>x/y</code>
$x^y$	<code>x^y</code>
$e^x$	<code>exp(x)</code>
$\log_{10}(x)$	<code>log10(x)</code>
$\ln(x)$	<code>log(x)</code>
$\log_2(x)$	<code>log2(x)</code>
$\cos(x)$	<code>cos(x)</code> [radians]
$\sin(x)$	<code>sin(x)</code> [radians]
$\sqrt{x}$	<code>sqrt(x)</code>

## Introduction to R 7: Example expressions

formula	R expression	Computed Value
$5^2 + 4^2$	<code>5^2 + 4^2</code>	41
$(5 + 4)^2$	<code>(5 + 4)^2</code>	81
$2+3/4-5$ (fraction)	<code>(2+3)/(4-5)</code>	-5
$\log_{10}(100)$	<code>log10(100)</code>	2
$\ln(4 * (2 + 3))$	<code>log(4*(2+3))</code>	2.995732
$\cos(30^\circ)$	<code>cos(30*pi/180)</code>	0.8660254
$\sin(30^\circ)$	<code>sin(30*pi/180)</code>	0.5

## Introduction to R 8: Other useful expressions cont

```
examplex <- c(-1.6, -1.5, -1.4, 1.4, 1.5, 1.6)
```

- `ceiling(examplex)`

- `[1] -1 -1 -1 2 2 2`

- `ceiling` takes a single numeric argument `x` and returns a numeric vector containing the smallest integers not less than the corresponding elements of `x`.

- `floor(examplex)`

- `[1] -2 -2 -2 1 1 1`

- `floor` takes a single numeric argument `x` and returns a numeric vector containing the largest integers not greater than the corresponding elements of `x`.

# Introduction to R 9: Other useful expressions cont

```
examplex <- c(-1.6, -1.5, -1.4, 1.4, 1.5, 1.6)
```

- `trunc(examplex)`

- `[1] -1 -1 -1 1 1 1`

- `trunc` takes a single numeric argument `x` and returns a numeric vector containing the integers formed by truncating the values in `x` toward 0.

- `round(examplex, digits = 0)`

- `[1] -2 -2 -1 1 2 2`

- `round` rounds the values in its first argument to the specified number of decimal places (default 0).

- `signif(examplex, digits = 1)`

- `[1] -2 -2 -1 1 2 2`

- `signif` rounds the values in its first argument to the specified number of significant digits.

# Introduction to R 10

- `summary(x)`
  - Generic function to give a “summary” of `x`, often a statistical one
    - `x <- c(2:4)` # in R
    - `summary(x)` # in R
      - Min. 1st Qu. Median Mean 3rd Qu. Max.
      - 2.0 2.5 3.0 3.0 3.5 4.0
    - Type `library(help = "stats")`
      - Shows the complete list of statistical functions

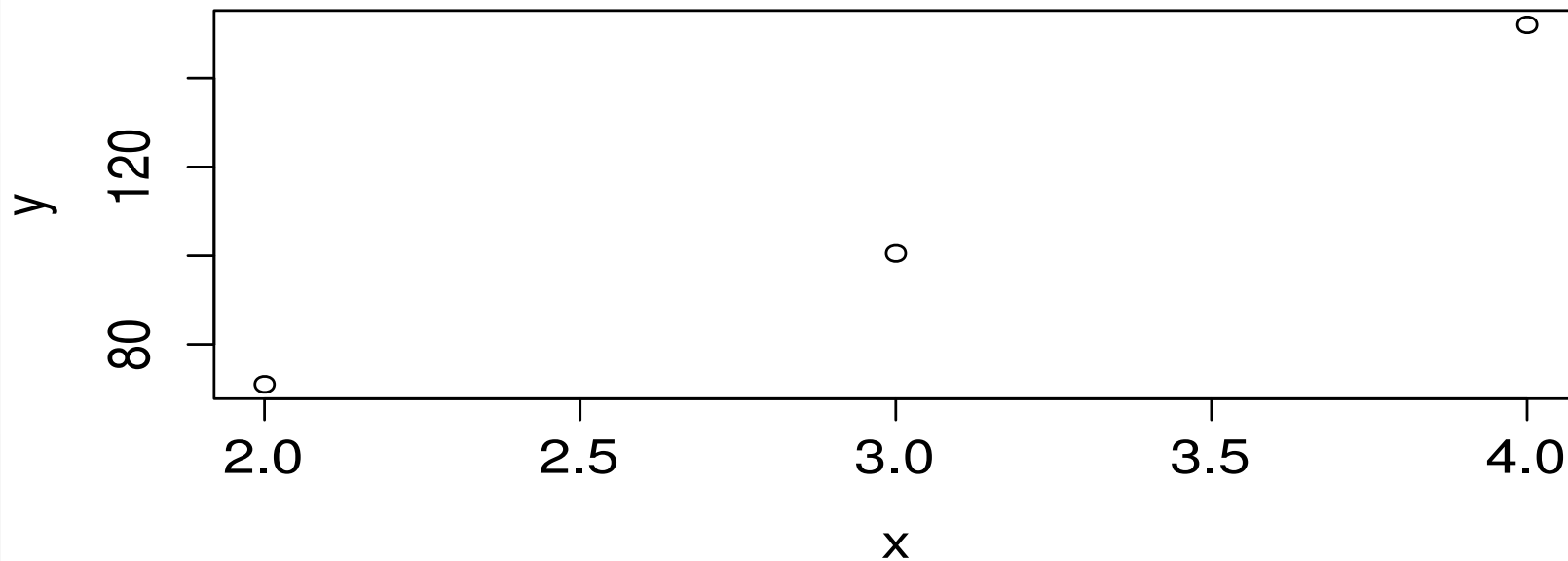


# Introduction to R 11

- `str(x)`
  - Displays the internal structure of an R object
  - `str(x) # in R`
    - `int [1:3] 2 3 4`
      - int is for integer
- `y <- x^3 + 2*x^2 + 0.5*x + 54 # still using the x from before stored in R`
  - `str(y) # in R`
    - `num [1:3] 71 100 152`
      - num is for numeric (includes decimals)

# Introduction to R 12

- `plot(x, y)`
  - generates a plot of x versus y



# Applications of R

- This will be performed within the R environment in RStudio
- The data used in this tutorial will be posted online at <http://www.ecoccs.com/RandUSGS.html#tut> & <http://www.ecoccs.com/tsuresearch.html#tut> along with this tutorial

# Resources used

- <http://www.r-project.org/> R Project for Statistical Computing
- <http://cran.r-project.org/doc/contrib/Baggott-refcard-v2.pdf> [R Reference Card 2.0 by Matt Baggott]
- Chapra, Steven C., *Applied Numerical Methods with MATLAB for Engineers and Scientists*, 2<sup>nd</sup> Edition, Boston, Massachusetts: McGraw-Hill, 2008, p. 22, 24-26, 29-34, 36.
- The MathWorks, Inc, MATLAB R2013b Documentation, “2-D line plot - MATLAB plot”, <<http://www.mathworks.com/help/matlab/ref/plot.html>>, Accessed: 13 October 2013.
- Morrell, D. Freshman Engineering Problem Solving with MATLAB, Connexions Web site. <http://cnx.org/content/col10325/1.18/> , Apr 23, 2007, p. 9-11.

# Useful R Resources

- <http://www.ecoccs.com/RandUSGS.html> [R Resources provided by Irucka Embry]
- <http://www.ecoccs.com/tsuresearch.html> [Research Resources for Tennessee State University (TSU) students and faculty provided by Irucka Embry]
- <http://users.monash.edu.au/~murray/stats/Rmanual.pdf> [R and S-Plus: Basic Instructions by Murray Logan, July 25, 2005]
- <http://www.stat.berkeley.edu/~spector/Rcourse.pdf> [Introduction to R (presentation) by Phil Spector, Department of Statistics, University of California]
- <http://www.stat.berkeley.edu/users/spector/R.pdf> [An Introduction to R by Phil Spector, Department of Statistics, University of California, September 24, 2004]
- <http://pairach.com/2012/02/26/r-tutorials-from-universities-around-the-world/> [R-Uni: (A List of Free R Tutorials and Resources in University webpages) by Pairach on February 26, 2012]

# Useful R Resources 2

- <http://cran.r-project.org/other-docs.html> [R Contributed Documentation]
- <http://www.r-bloggers.com/free-r-book-collection/> [Free R Book Collection By Richard O. Legendi, September 6, 2011]
- <http://stackoverflow.com/questions/tagged/r> [R Tagged Questions: Stack Overflow]
- <http://www.personality-project.org/r/r.commands.html> [A short list of the most useful R commands by William Revelle, Department of Psychology, Northwestern University. A summary of the most important commands with minimal examples. See the relevant part of the guide for better examples. For all of these commands, using the help(function) or ? function is the most useful source of information. Unfortunately, knowing what to ask for help about is the hardest problem.]
- <http://www.gardenersown.co.uk/Education/Lectures/R/> [Using R for statistical analyses</a> by Dr. Mark Gardener. This page is intended to be a help in getting to grips with the powerful statistical program called R. It is not intended as a course in statistics. If you have an analysis to perform I hope that you will be able to find the commands you need here and copy/paste them into R to get going. On this page learn how to create data files, read them into R and generally get ready to perform analyses. Also find out about getting further help and documentation.]

# Useful R Resources 3

- <http://ww2.coastal.edu/kingw/statistics/R-tutorials/index.html> [R Tutorials by William B. King, Ph.D., Coastal Carolina University]
- <http://rtutorialseries.blogspot.com/> [R Tutorial Series By John M Quick: The R Tutorial Series provides a collection of user-friendly guides to researchers, students, and others who want to learn how to use R for their statistical analyses.]
- <http://www.r-bloggers.com/computerworlds-beginners-guide-to-r/> [Computerworld's Beginners Guide to R By David Smith, June 17, 2013]
- <http://pj.freefaculty.org/R/Rtips.html> [Rtips. Revival 2012! by Paul E. Johnson]
- <http://science.nature.nps.gov/im/datamgmt/statistics/r/index.cfm> [Using R Statistical and Graphics Tools for Natural Resource Stewardship Science]
- <http://math.illinoisstate.edu/dhkim/Rstuff/Rtutor.html> [Statistical Computing with R: A tutorial by Dong-Yun Kim]

# Useful R Resources 4

- [http://cran.r-project.org/doc/contrib/Burns-unwilling\\_S.pdf](http://cran.r-project.org/doc/contrib/Burns-unwilling_S.pdf) [The R language - a short companion: This companion is essentially based on the documents “An Introduction to R” and “R language definition” both version 1.7.1, available on the R website <http://www.r-project.org/> . Graphical and statistical functionalities are not considered. Version 1.2. Marc Vandemeulebroecke, July 14th, 2003]
- [http://www.itc.nl/~rossiter/teach/R/RIntro\\_ov.pdf](http://www.itc.nl/~rossiter/teach/R/RIntro_ov.pdf) [Introduction to the R Project for Statistical Computing by D G Rossiter, University of Twente, August 10, 2010]
- <http://cran.r-project.org/doc/contrib/Baggott-refcard-v2.pdf> [R Reference Card 2.0 by Matt Baggott]
- <http://www.ats.ucla.edu/stat/r/> [Resources to help you learn and use R: UCLA: Statistical Consulting Group]
- <http://www.ats.ucla.edu/stat/r/library/> [R Library by Matt Baggott]



# Useful R Resources 5

- [http://www.ats.ucla.edu/stat/r/library/advanced\\_function\\_r.htm](http://www.ats.ucla.edu/stat/r/library/advanced_function_r.htm) [R Library: Advanced function by Matt Baggott]
- <http://phoxis.org/2013/05/04/get-list-of-installed-packages-and-their-details-in-r/> [Get list of installed packages and their details in R: Phoxis]
- <http://www.statmethods.net/interface/packages.html> [Quick-R: R Packages]
- <http://data.princeton.edu/R/> [Introducing R by German Rodriguez, Office of Population Research, Princeton University]
- <http://data.princeton.edu/R/readingData.html> [Reading and Examining Data by German Rodriguez, Office of Population Research, Princeton University]
- <http://msenux.redwoods.edu/math/R/dataframe.php> [Data Frames in R: Department of Mathematics, College of the Redwoods]
- <http://www.r-statistics.com/tag/data-frame/> [Data Frame: R-statistics blog]
- <http://nsaunders.wordpress.com/2010/08/20/a-brief-introduction-to-apply-in-r/> [A brief introduction to “apply” in R: What You’re Doing Is Rather Desperate: Notes from the life of a bioinformatician]

# Closing

- Thank you for your time