REVOLUTION ANALYTICS

Fundamentals in R

August 12, 2014

Fundamentals in R Overview

In this 1-day course we will introduce R, the free interactive programming language and environment for statistical computing and graphics and emphasize the fundamentals and functionality of the language.

You will learn how to get going in R from the beginning, understand file formats and basic R syntax.

Fundamentals in R Overview

After completing Fundamentals in R, you will be able to:

1. Understand the basics of R language syntax.
2. Implement basic data manipulation techniques.
3. Explore data with summary functions.
4. Visualize data in R.

Fundamentals in R Agenda

Module	Description
1	Overview
2	Getting Started
3	Data Management
4	Data Exploration
5	Extended Functionality

Introductions

Poll: What technologies are you currently using?

- R
- Excel
- SQL
- Python
- SAS
- SPSS
- Stata
- MATLAB

Module 1 - Overview

In this module we'll introduce the R language: its history & philosophy, as well as its strengths & weaknesses. We'll introduce some helpful resources and walk you through getting help in R.

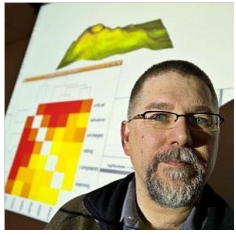
The objectives are

- learn about R's history
- understand the philosophy of R
- know where to go for help
- identify strengths & weakness of R

History

- Inspired by the S environment
  - 1970's
  - John Chambers, Bell Labs

History



- Informal: R Project
  - 1990? -> 1997
  - Ross Ihaka and Robert Gentleman
  - University of Auckland, New Zealand
  - Introductory Statistics course

History

- Formalization: R Core group
  - 1997 -> present
  - nine members -> ~20 members
  - responsible for dev of basic foundational R software and the infastructure supporting it's open source & extendable nature
- R version 1.0 was released on Feb 29, 2000.

R Philosophy

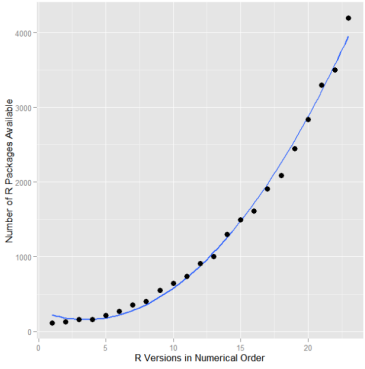
R follows the Unix philosophy:

Write programs that do one thing and do it well. Write programs to work together

The base functionality of R (what you get out-of the box) is rich and flexible. You can extended this functionality with your own functions or with community contributed functions that ship in packages. You can extend this functionality by interfacing R with other tools

Packages in R play well with eachother. R plays well with others.

R Packages



"Number of R packages plotted for each major release of R. The last value on the x-axis represents version 2.15.2, the final release in 2012" - r4stats.com

R Packages

~100 packages in 2002. Now over 6,000 packages!

The Comprehensive R Archive Network (CRAN) is the main repository for R core and its community built libraries. See http://cran.r-project.org/

R Packages

	Package	Title	Downloads
1	plyr	Tools for splitting, applying and combining data	84049
2	digest	Create cryptographic hash digests of R objects	83192
3	ggplot2	An implementation of the Grammar of Graphics	82768
4	colorspace	Color Space Manipulation	81901
5	stringr	Make it easier to work with strings	77658
6	RColorBrewer	ColorBrewer palettes	66783
7	reshape2	Flexibly reshape data: a reboot of the reshape package	64911
8	zoo	S3 Infrastructure for Regular and Irregular Time Series (Z's ordered observations)	60844
9	proto	Prototype object-based programming	59043
10	scales	Scale functions for graphics	58369

Top 10 (most downloaded) R packages in Jan-May 2013. See blog.revolutionanalytics.com

R Resources

- CRAN.R-Project.org/
  - Download R
  - Package Repository
  - Task Views (packages organized by use)
  - Manuals by R-Core & Contributed
- Rseek.org The R Search Engine
- R Bloggers
- R Reference Card
- Stackoverflow.com Q&A Forum where best answers are voted up

Strengths and Weaknesses of R

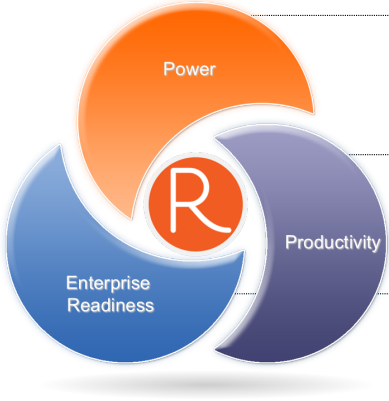
Strengths

- Open source, extendable
- Large user community
- Comprehensive & Cutting Edge
- Advanced data structures and graphics
- Flexible

Weaknesses

- Learning curve
- Lack of technical support
- Single-threaded & Memory bound

Revolution Analytics Scales R to the Enterprise



Power
- Distributed high performance analytics

Productivity
- Build & deploy analytics applications easily

Enterprise Readiness
- Enterprise landscape
- Full-service customer support, consulting and training

REVOLUTION ANALYTICS

Thank you

Revolution Analytics is the leading commercial provider of software and support for the popular open source R statistics language.

www.revolutionanalytics.com, 1.855.GET.REVO, Twitter: @RevolutionR