



An Introduction to

Presented by Geoffrey Hubona

Overview of R



■ Introduction to R

- What is R ?
- Administration / Organization of R
- Comprehensive R Archive Network (CRAN)
- Packages (libraries) in R
- Reading in Data and Files (with demo script)
- R Workspace
- Help

What is R?



- R is an integrated suite of software facilities for ***data manipulation, calculation*** and ***graphical display***.
- R has:
 - An effective ***data handling and storage*** facility.
 - A ***suite of operators*** for calculations on ***arrays*** and ***matrices***.
 - A large, coherent, ***integrated collection*** of ***tools*** for ***data analysis***.
 - ***Graphical facilities*** for data analysis and display.
 - A well developed, simple and effective ***programming language***.

Organization



- **R Development Core Team** (<http://r-project.org>)
 - They make R; Mostly original R people
- **R Foundation:** Non profit
 - Financial support for R Project
 - Contact point
 - Legal stuff
- **Comprehensive R Archive Network (CRAN)**
 - <http://cran.r-project.org>
 - Pushes R software out to the world
 - “CRAN mirrors”: where you download packages

Comprehensive R Archive Network



Physical network pushes out R software. Here are USA “Mirrors”:

<http://cran.cnr.Berkeley.edu>

University of California, Berkeley, CA

<http://cran.stat.ucla.edu/>

University of California, Los Angeles, CA

<http://streaming.stat.iastate.edu/CRAN/>

Iowa State University, Ames, IA

<http://ftp.ussg.iu.edu/CRAN/>

Indiana University

<http://rweb.quant.ku.edu/cran/>

University of Kansas, Lawrence, KS

http://watson.nci.nih.gov/cran_mirror/

National Cancer Institute, Bethesda, MD

<http://cran.mtu.edu/>

Michigan Technological University, Houghton, MI

<http://cran.wustl.edu/>

Washington University, St. Louis, MO

<http://cran.case.edu/>

Case Western Reserve University, Cleveland, OH

<http://ftp.osuosl.org/pub/cran/>

Oregon State University

<http://lib.stat.cmu.edu/R/CRAN/>

Statlib, Carnegie Mellon University, Pittsburgh, PA

<http://cran.mirrors.hoobly.com>

Hoobly Classifieds, Pittsburgh, PA

<http://mirrors.nics.utk.edu/cran/>

National Institute for Computational Sciences, Oak Ridge, TN

<http://cran.revolutionanalytics.com>

Revolution Analytics, Dallas, TX

<http://cran.sixsigmaonline.org/>

sixsigmaonline.org, Houston, TX

<http://cran.fhcrc.org/>

Fred Hutchinson Cancer Research Center, Seattle, WA

<http://cran.cs.wvu.edu/>

Western Washington University, Bellingham, WA

Packages in R



- ***Experts around the world contribute packages***
 - Promotes rapid deployment of “new” capabilities which are then available to anyone.
- Go to <http://cran.r-project.org>, click on ‘Packages’ link in left margin:
- ***6,215 available packages on CRAN alone ! !***
 - [Table of available packages, sorted by date of publication](#)
 - [Table of available packages, sorted by name](#)
 - [CRAN Task Views](#) allow you to browse packages by topic and provide tools to automatically install all packages for special areas of interest.

Packages in R



- R functionality comes from ***packages***: some packages ‘come with’ the base installation; the rest you must ***install*** from a CRAN mirror and then ***load***.
 - Can do it from the menus, but ***you are better off learning to use script early !***

> # to see what packages are currently in your workspace:

> search() # displays packages and data frames

[1]	".GlobalEnv"	"package:Rcmdr"	"package:car"
[4]	"package:survival"	"package:splines"	"package:nnet"
[7]	"package:MASS"	"package:tcltk"	"package:graphics"
[10]	"package:grDevices"	"package:utils"	"package:datasets"
[13]	"package:stats"	"package:methods"	"RcmdrEnv"
[16]	"Autoloads"	"package:base"	

Packages in R



- Can do it from the menus, but ***you are better off learning to use script early !***

```
> # to see what packages are installed in your working directory
> library()
```

Packages in library 'C:/Users/jeff/Documents/R/win-library/2.13':

abind	Combine multi-dimensional arrays
akima	Interpolation of irregularly spaced data
amap	Another Multidimensional Analysis Package
aplpack	Another Plot PACKage: stem.leaf, bagplot, faces, spin3R, and some slider functions
car	Companion to Applied Regression
coin	Conditional Inference Procedures in a Permutation

Packages in R



- Can do it from the menus, but ***you are better off learning to use script early !***

> ##### TWO-STEP PROCESS TO MAKE PACKAGE ACTIVE:

> ## 1. INSTALL THE PACKAGE (ONLY NEED DO THIS ONE TIME):

```
> install.packages("survey")
```

package 'survey' successfully unpacked and MD5 sums checked

The downloaded packages are in

C:\Users\jeff\AppData\Local\Temp\Rtmp54eD32\downloaded_packages

> ## 2. LOAD THE PACKAGE INTO YOUR WORKSPACE (DO THIS EACH TIME):

```
> library("survey")
```

OR

```
> require("survey")
```

Reading in Data and Importing Files



- **`read.table()`** function reads in text files and converts to a data frame:

```
> # read in a data file from URL and assign to variable:  
> d1 <- read.table("http://www.bio.ic.ac.uk/research/mjcraw/therbook/data/worms.txt",header=T)
```

```
> head(d1) # to print (display) the first six records:
```

	Field.Name	Area	Slope	Vegetation	Soil.pH	Damp	Worm.density
1	Nashs.Field	3.6	11	Grassland	4.1	FALSE	4
2	Silwood.Bottom	5.1	2	Arable	5.2	FALSE	7
3	Nursery.Field	2.8	3	Grassland	4.3	FALSE	2
4	Rush.Meadow	2.4	5	Meadow	4.9	TRUE	5
5	Gunness.Thicket	3.8	0	Scrub	4.2	FALSE	6
6	Oak.Mead	3.1	2	Grassland	3.9	FALSE	2

Reading in Data and Importing Files



- Take 'therbook.zip' and unzip all files into c:/temp folder.

```
> d2 <- read.table("c://temp/worms.missing.txt",header=T)
```

```
> some(d2) # to show a random selection of records
```

	Field.Name	Area	Slope	Vegetation	Soil.pH	Damp	Worm.density
1	Nashs.Field	3.6	11	Grassland	4.1	FALSE	4
4	Rush.Meadow	2.4	5	Meadow	4.9	TRUE	5
6	Oak.Mead	3.1	2	Grassland	3.9	FALSE	2
7	Church.Field	3.5	3	Grassland	NA	NA	NA
9	The.Orchard	1.9	0	Orchard	5.7	FALSE	9
10	Rookery.Slope	1.5	4	Grassland	5.0	TRUE	7
11	Garden.Wood	2.9	10	Scrub	5.2	FALSE	8
14	Observatory.Ridge	1.8	6	Grassland	3.8	FALSE	0
15	Pond.Field	4.1	0	Meadow	5.0	TRUE	6
17	Cheapside	2.2	8	Scrub	4.7	TRUE	4

Reading in Data and Importing Files



- `read.csv()` reads in .csv files and converts to data frame:

```
> d3 <- read.csv("c://temp/bowens.csv") # Do not need header=T
```

```
> tail(d3) # to show last 6 records
```

	place	east	north
728	Wytham	47	8
729	Wytham Meads	47	9
730	Wytham Wood	46	8
731	Yattendon	54	74
732	Yattendon Court	55	74
733	Youlbury	48	3

Reading in Data and Importing Files



- **scan ()** function reads in data from keyboard:

```
x <- scan()
```

```
> 1: 2
```

```
2: 6
```

```
3: 4
```

```
4: 7
```

```
5: ## NOTE: Hit <enter> twice here
```

```
> x
```

```
[1] 2 6 4 7
```

Accessing Data from Other Statistical Packages

- To import data from Minitab, S-PLUS, SAS, SPSS, Stata, Systat, others.....

> library(foreign)

- Always better off working with original data directly, rather than importing data possibly modified by another statistical package.

The R Workspace



- Objects you create in an **R** session are held in memory; the collection of objects one currently has is called the ***workspace***.
- Your workspace is saved when you close R
 - as a *****.RData** file).
 - In menus: **File > Save workspace...**
Mac: Workspace > Save Workspace File
- You can also explicitly load a saved workspace file, even the workspace image of someone else.
 - In menus: **File > Load workspace...**
Mac: Workspace > Load Workspace File

The R Workspace

as working directory



■ Workspace is organized like a **vector**:

```
> # ls() and objects() returns vector of character strings
> # giving names of data sets and functions a user has defined
> ls()
[1] "exprs"          "f1"              "f2"
[4] "f3"             "f4"              "f5"
[7] "far"            "far.global"      "far.inner"
[10] "far.mod"        "far.outer"       "far.res.clus"
[13] "island"         "kyphosis"        "local.rebus2"
[16] "local.rebus3"   "local.rebus4"    "m1"
[19] "men1500m"       "men1500m_cubic"  "men1500m_lm"
[22] "men1500m_lm2"   "men1500m_lowess" "men1500m1900"
[25] "model1"         "model2"          "model3"
[28] "model4"         "model5"          "rebus.far"
[31] "rebus.far.twofunc" "rebus.far2"      "rebus.far3"
```

The R Workspace

as working directory



- **search()** returns **packages** and other objects, usually **data frames**:

```
> # search() returns packages and data frames
```

```
> search()
```

```
[1] ".GlobalEnv"          "package:Rcmdr"        "package:car"
[4] "package:survival"     "package:splines"      "package:nnet"
[7] "package:MASS"         "package:tcltk"        "package:graphics"
[10] "package:grDevices"   "package:utils"        "package:datasets"
[13] "package:stats"       "package:methods"      "RcmdrEnv"
[16] "Autoloads"           "package:base"
```

The R Workspace

as working directory



- `rm(list=ls())` should be used carefully:

```
> rm(list=ls())
```

```
> ls() # listing contents returns empty character string  
character(0)
```

- Standard useful commands for managing the workspace:

```
> # print (list) the current directory
```

```
> getwd()
```

```
[1] "C:/Users/jeff/Documents "
```

```
> # change the working directory
```

```
> setwd(new.dir) # must include entire path
```

```
> # save the workspace to the file ".RData" (default)
```

```
> save.image()
```

```
> # which is equivalent to
```

```
> save(list=ls(all=TRUE), file=".RData")
```

The R Workspace

saving files



```
# save a specific object to a file  
> save(object, file="myfile.RData")
```

```
# load a saved workspace into the current session  
# must have entire path name  
> load(object, file="C:/Users/jeff/Documents/myfile.RData")
```

```
# quit R  
> q()
```

```
# quit without R asking you whether it should save your data  
> q(save="no")
```

The R Workspace

options



- Can use `options()` to view and set options for the R session:

```
> options()
```

```
$add.smooth
```

```
[1] TRUE
```

```
$browserNLdisabled
```

```
[1] FALSE
```

```
$check.bounds
```

```
[1] FALSE
```

```
$continue
```

```
[1] "+ "
```


Online Help in R



- Online help similar to **man** facility in UNIX

```
# Help on function solve()
```

```
> help(solve)
```

```
# or, for short, just:
```

```
> ?solve
```

```
# On most systems have help in html by
```

```
# 'Search Engine and Keywords' link especially useful
```

```
> help.start()
```

```
# help.search() command or alternatively ?? allow searching
```

```
# in various ways - shows anything that contains the word
```

```
> ??solve
```

Help: R Manuals



- **An Introduction to R** (R-intro.pdf)
 - Programming environment for data Analysis and graphics
- **R Data Import / Export** (R-data.pdf)
 - Describes the import and export facilities available either in R itself or via packages which are available from CRAN.
- **The R Language Definition** (R-lang)
 - Documents the **R** language per se.
 - Is a dialect of **S** which was designed in the 1980s and has been in widespread use in statistical community since.
- **R Installation and Administration** (R-admin.pdf)