© PracticalStats.com

Nondetects And Data Analysis:
1. Getting Started with R and RStudio

Dennis R. Helsel, Ph.D

Practical Stats

1

1

© PracticalStats.com

Where we're headed

1. What is R?
2. What can R do?
3. Where can I get it?
4. How to get started?
5. R basics
6. Stats with R
7. R packages
8. RStudio

2

2

© PracticalStats.com

1. What is R?

- Free, open-source software
- A programming language
- Modeled after S, a statistics language developed at Bell Laboratories in late 1980s
- Originally developed at Univ. of Auckland in 1995
- Written collaboratively by teams of volunteers
- Broad suite of statistical methods
- A scripting language; you can develop your own routines (scripts)
- Used to quickly distribute new methods to a large audience
- A system for data analysis and statistical modeling

3

3

© PracticalStats.com

2. What can R do? Graphs

4

4
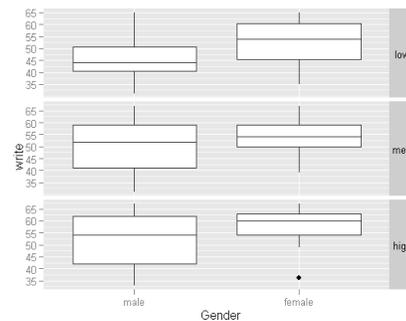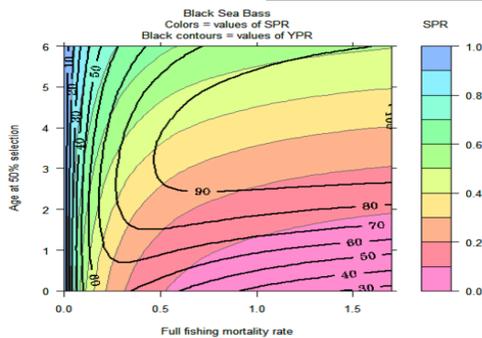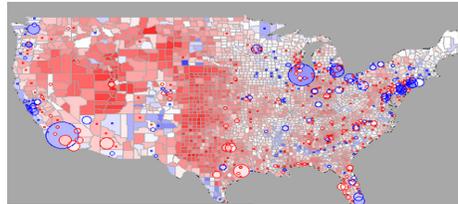
What can R do? Statistical methods

- All basic estimators and hypothesis tests
- Kendall trend tests
- Multivariate methods, including those popular in ecology
- Permutation tests
- Time series methods for data closely-spaced in time
- Methods for data with nondetects (the NADA and EnvStats packages)

5

5

3. Where can I get R software?

http://cran.r-project.org/   the "CRAN site"

- For Linux, Windows and Mac OS
- Download the binary version for your OS
- Comes with guarantee that there are no viruses or malware
- Accuracy assured for base package only. Authors responsible for accuracy of packages
- 'No one to call and yell at' but wikis and mailing lists provide much useful information and support
- Will also find free and low-cost books (pdf) on the CRAN site

6

6

How do I install R?

See the instructions in your student Handouts folder

7

---
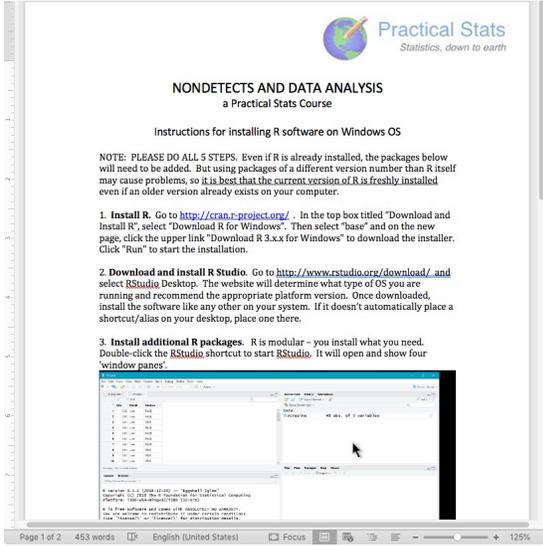
4. How do I get started?

Useful free information is available on the CRAN site. See the Contributed link on the left side. There you'll find, among other items:

1. R for Beginners.pdf

2. R for Biologists.pdf

3. R Commander: An Introduction.pdf

4. Using R for Data Analysis and Graphics - Introduction, Examples and Commentary.pdf

5. Statistics Using R with Biological Examples

Two helpful documents for the first-time R user:

1. A Zero-Level Tutorial for Learning R http://galyardt.myweb.uga.edu/Papers/RTutorial-Level0.pdf

2. Simple R: stats using R by Verzani   http://cran.r-project.org/doc/contrib/Verzani-SimpleR.pdf

8

6. R Basics: the console window

- Lower left window in Rstudio
- Type commands. Can cut and paste from an example file of commands, or insert from the upper left script window in RStudio
- Case sensitive. boxplot ≠ Boxplot
- Assign results of a computation to an object (a 'name')
- Type the object name to see contents

```
> result = log(12) + sqrt(25)
> result
[1] 7.484907
```

9

9

R basics: Entering data by hand and using built-in functions

Use the concatenate function c(…..)
Assign the data vector to an object name with the arrow:

```
> Mydata <- c(2,5,10,16,27,46)
> mean(Mydata)
[1] 17.66667
> sd(Mydata)          (R's standard deviation function)
[1] 16.47625
```

You can combine many functions into an equation:

```
> Mysd <- sqrt(sum((Mydata-mean(Mydata))^2/(length(Mydata)-1)))
> Mysd
[1] 16.47625
```

10

10

© PracticalStats.com

R basics: the Help system

> ?boxplot

Finds any command from a package that is loaded.

Provides HTML files stored by the developers. All options available for commands are listed. Links are provided to additional detail.

> ??bootstrap

Searches for the text string following ?? In all documentation, not just packages that are loaded. Tells you what packages might contain information on the string.

11

11

© PracticalStats.com

6. Stats with R: t-test as example

```
> t.test(Total_N ~ Location)              ⟵         (you type)

        Welch Two Sample t-test

data:  Total_N by Location
t = 0.029, df = 11.555, p-value = 0.9773
alternative hypothesis: true difference in means is not
     equal to 0
95 percent confidence interval:
 -2.081479  2.137479
sample estimates:
 mean in group indust mean in group residen
              1.666                 1.638
```

12

12

© PracticalStats.com

# Stats with R: rank-sum test

```
> wilcox.test(TotN ~ Location)
Wilcoxon rank sum test with
continuity correction


data:  TotN by Location
W = 76.5, p-value = 0.04911


alternative hypothesis: true
location shift is not equal to 0
```


13

13

© PracticalStats.com
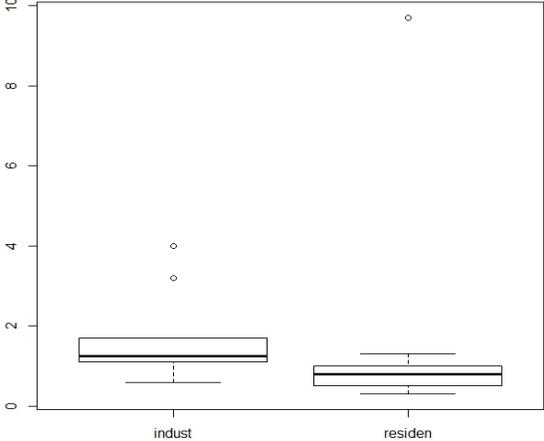
7. R Packages: Adding new methods

Packages expand the capabilities of R. Currently (2021) there are over 10,000 add-on packages for R.

Install packages using the lower right pane of Rstudio, with the Packages tab. Or use the Loadlibs.R script.
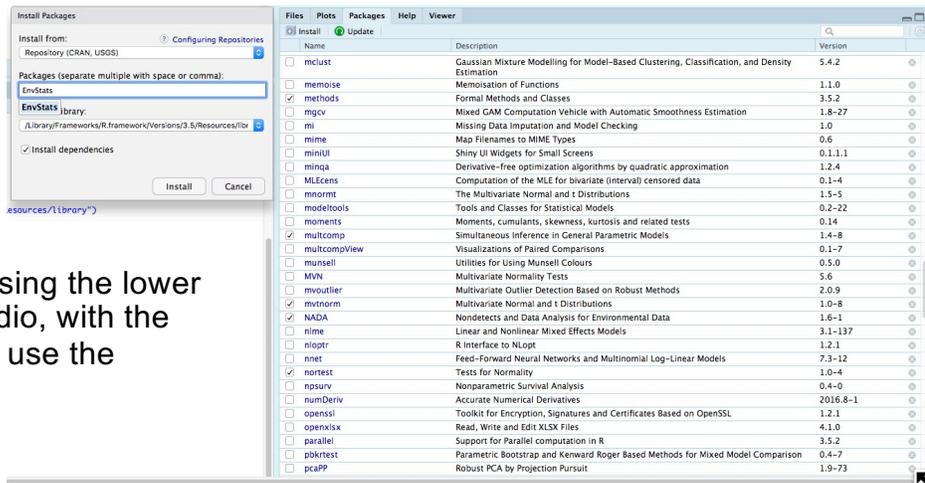
14

14

© PracticalStats.com

The 15 packages to install for this course

bestglm	car	cenGAM	EnvStats
fitdistrplus	Kendall	mgcv	multcomp
NADA	nlme	perm	rms
survminer	vegan	NADA2	

15

15

© PracticalStats.com

8. Rstudio: the Four 'Panes'

Script files, View Data Sets   Data Sets, Variables description
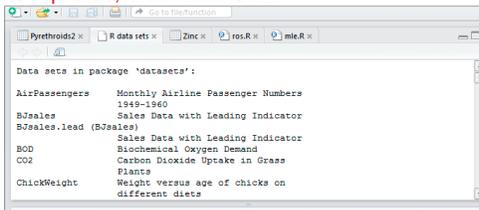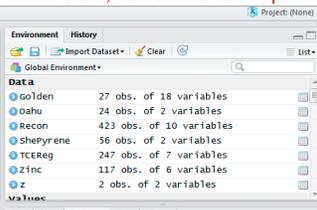
R console
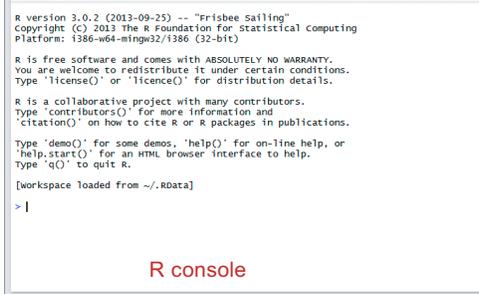
Plots, packages, Help
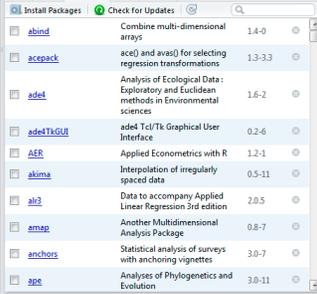
16

16

© PracticalStats.com

Three steps to start using RStudio

1. Start up Rstudio

2. Set the working directory

3. Load the course packages using 1 of 2 methods:

   a) Click the boxes next to each of the 15 packages
      (use the Packages tab in lower right Pane)

   b) Run the Loadlibs.R script (much faster)

17

17

© PracticalStats.com

R basics: Run a Script

R is widely used in academia and industry all around the world.

New scripts (i.e. programs, macros) are written by users. If they go beyond their own use, they can be compiled into packages.

Scripts that come with our courses make complex procedures in R easier to do. To run a script file with RStudio, go to

`Code > Source file`

and select the script file. Here we select Loadlibs.R

18

18

© PracticalStats.com

Remember this!

Attach to the data set. It reduces typing!

```
> attach(DataSetName)
```

Makes a loaded data set available to user-written scripts and typed commands without typing the data set name each time.

Without attaching, a variable is referred to as DataSetName$VariableName

After attaching, just type VariableName

19

19

© PracticalStats.com

The RStudio Demo

The Three Steps to Start Using RStudio

1. Start up RStudio

2. Set the working directory
   ```
   Session > Set working directory
        > Choose directory
   ```

3. Load the course packages
   ```
   Code > Source file
   ```
   and select Loadlibs.R

   OR check the boxes next to all 15 packages required.

Then we will:

1. Read in a data set

2. View the data set

3. Attach to the data set

4. Compute the mean of a variable

20

20