

Statistics means never having to say you're certain.

Philip Stark

Data analysis is an aid to thinking and not a replacement for it.
Richard Shillington

Before the curse of statistics fell upon mankind we lived a
happy, innocent life, full of merriment and go, and
informed by fairly good judgment.

Hilaire Belloc
The Silence of the Sea

Quantitative Approaches in Ecological Research and Education Using



Statistical thinking will one day be as necessary for efficient
citizenship as the ability to read and write.

H. G. Wells

“Organic chemist!”, said Tilley disdainfully. “Probably knows no
statistics whatever.”

Nigel Balchin
The Small Back Room

Course Objectives

- Introduce you to R - a freely available and very powerful tool for conducting statistical analyses
- Demonstrate the capabilities and tools available in R that are specifically relevant to ecologists

Where are we going?

8:00-8:15 Introduction

8:15-9:15 Introduction to R

Statistical software and statistical programming
Installation and setup – easy as 1, 2, 3!
Getting data in
Managing data

9:15-10:00 Statistical Modeling

What is a statistical model anyway?
General Linear Model (or – ANOVA, ANCOVA, and REG – What's the difference?)

10:00-10:15 Morning Break

10:15-11:30 Statistical Modeling (Cont.)

Generalized Linear Model (or – how to sidestep those pesky assumptions!)
Mixed-effects models
Questions/Answers

11:30-1:00 Lunch Break

1:00-3:00 Statistical Modeling (Cont.)

Mixed-effects models (Cont.)

3:00-3:15 Afternoon Break

3:15-4:45 Ecological Modeling

Ordinary Differential Equation Models
Final Review – Questions/Answers

4:45-5:00 Wrap-up and Course Evaluations

Who are we?

Why R?

- An open source environment for statistical computing and visualization
 - GNU/GPL version of the S Language from Bell Laboratories
 - Highly extensible (i.e., customizable)
- Integrated suite of software facilities for data manipulation, calculation, analysis, and graphical display
 - Effective data handling and storage facility
 - Large, coherent, integrated collection of tools for data analysis
 - Graphical facilities for data analysis and display
 - A well-developed, simple, and powerful programming language

Why R?

“The term "environment" is intended to characterize it as a fully planned and coherent system, rather than an incremental accretion of very specific and inflexible tools, as is frequently the case with other data analysis software.”

R is free :)

Binaries available for Windows, Mac, Linux, Unix, ...

R is a programming language!

- **Interpreted Language**
 - Issue a command
 - R immediately gives a response (no compiling)
- **Two basic ways to interact with R**
 - Interactive session
 - Type in command – get an answer
 - R commands are functions
 - `output = function_name(input)`
 - R Scripts (text file with name - `file_name.R`)
 - Save a long list of commands in a text file
 - Run the script using `source()`

Scripting!!!!

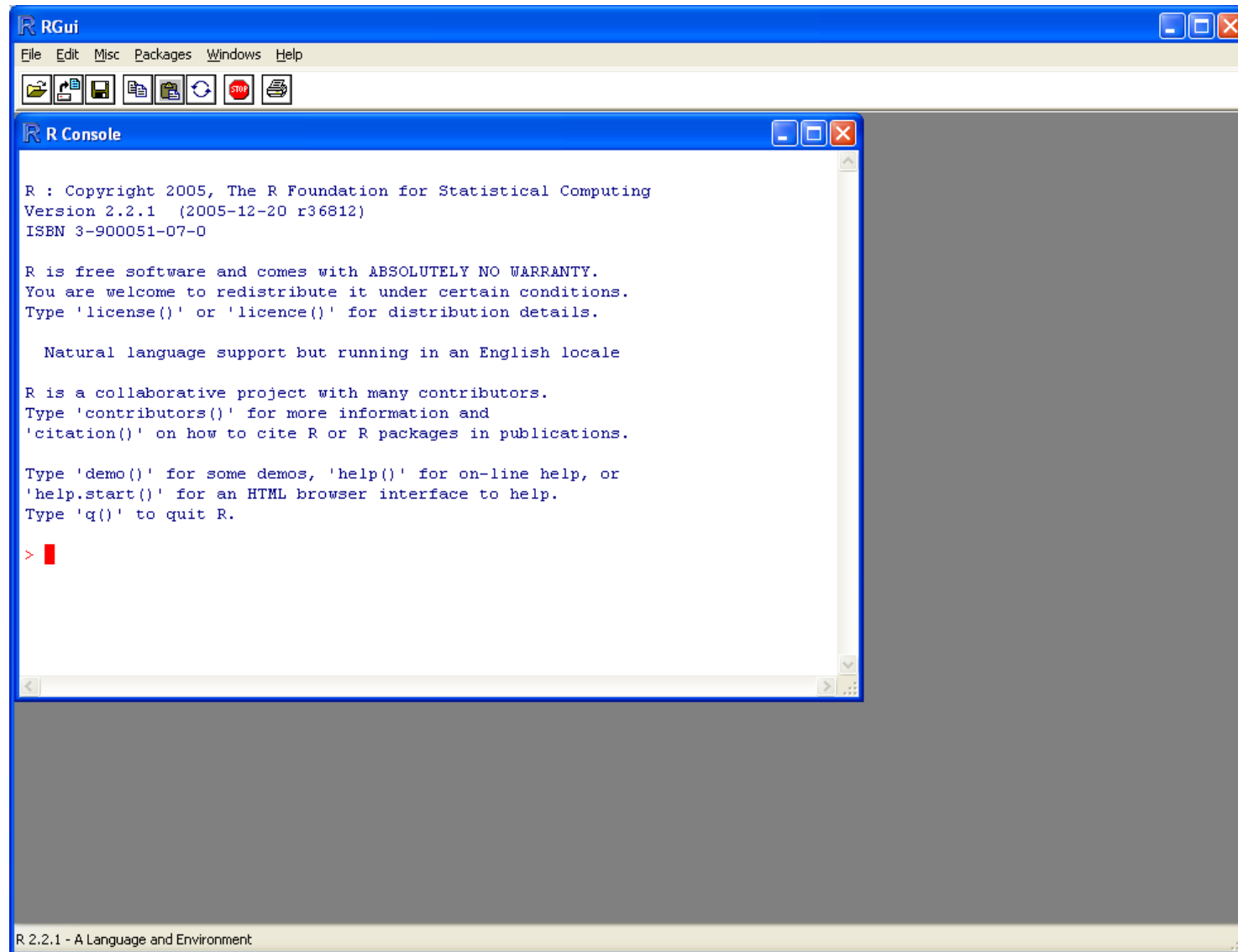
- **Explicit code!**
 - File merges
 - Case deletions
 - Transformations
 - Calculations
 - Analysis
 - Graphics
- **Advantages**
 - Retains integrity of original data
 - All manipulation of raw data is documented
 - Reduces ambiguity and number of data files
 - Reduces chances of mistakes
 - Facilitates unanticipated changes
 - Saves time in the long run!!

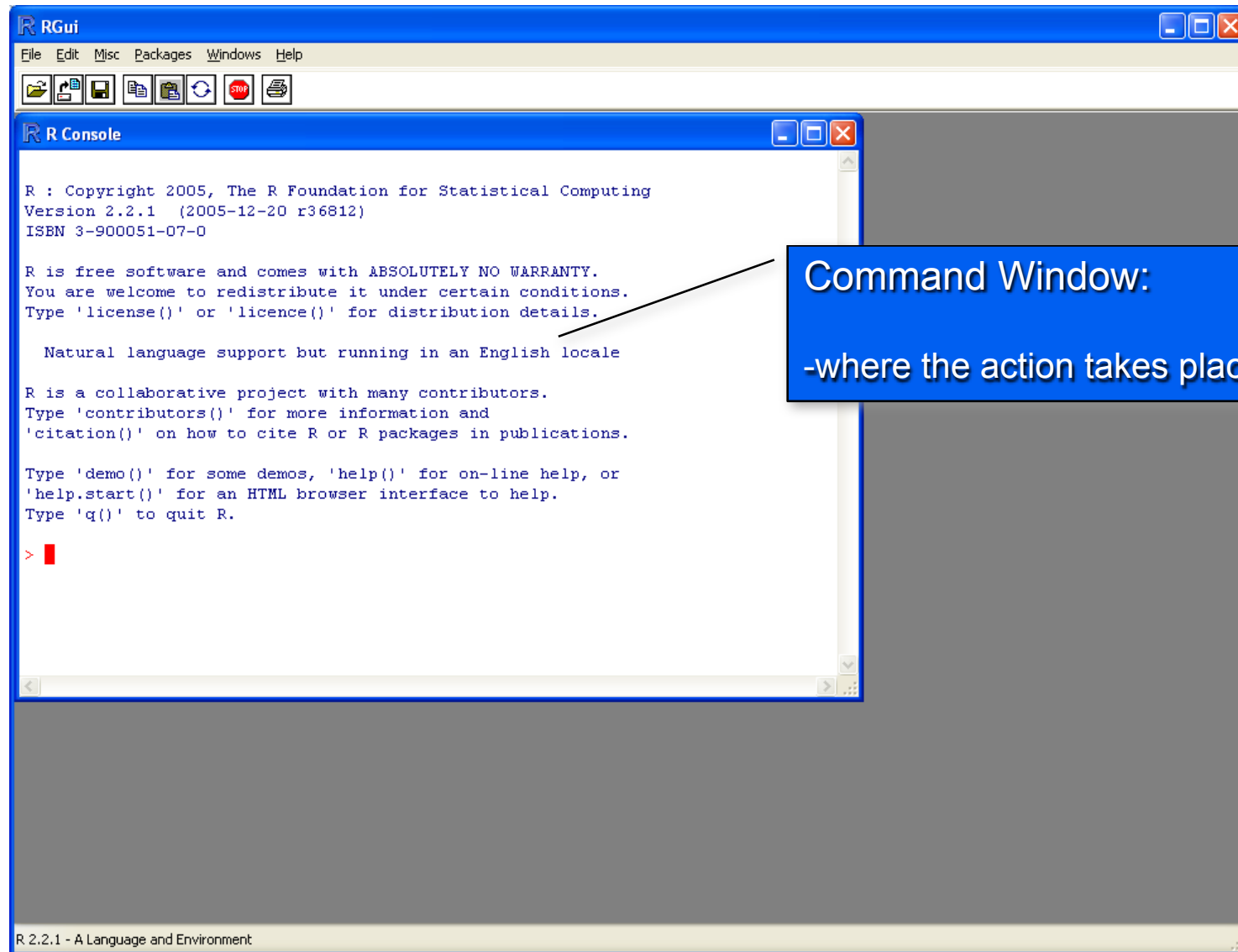
Write your own functions!

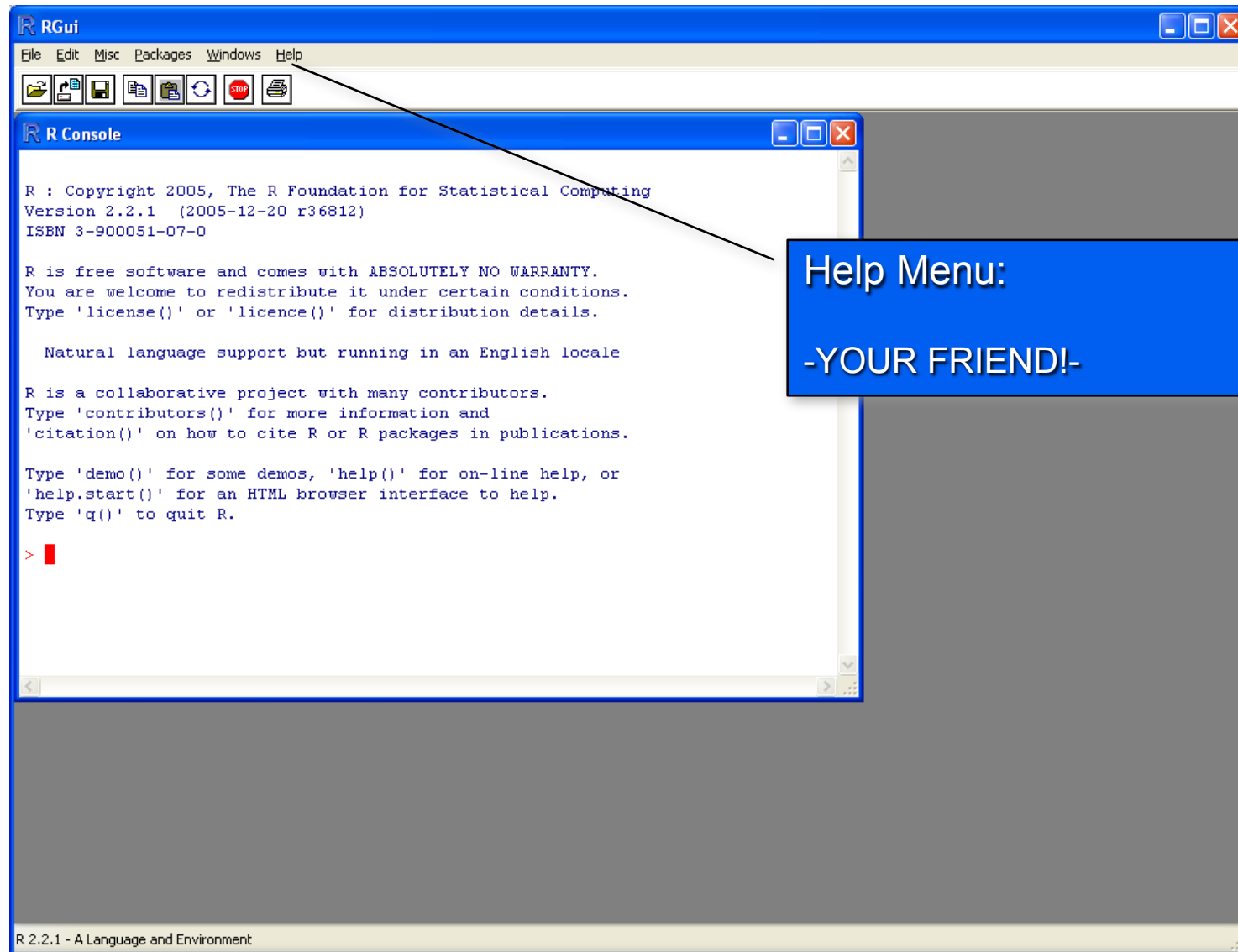
```
EC50.calc<-function(coef,vcov,conf.level=.95) {  
  # calculates confidence interval based upon Fieller's thm.  
  # assumes link is linear in dose  
  call <- match.call()  
  b0<-coef[1]  
  b1<-coef[2]  
  var.b0<-vcov[1,1]  
  var.b1<-vcov[2,2]  
  cov.b0.b1<-vcov[1,2]  
  alpha<-1-conf.level  
  zalpha.2 <- -qnorm(alpha/2)  
  gamma <- zalpha.2^2 * var.b1 / (b1^2)  
  EC50 <- -b0/b1  
  
  const1 <- (gamma/(1-gamma))*(EC50 + cov.b0.b1/var.b1)  
  const2a <- var.b0 + 2*cov.b0.b1*EC50 + var.b1*EC50^2  
  - gamma*(var.b0 - cov.b0.b1^2/var.b1)  
  const2 <- zalpha.2/( (1-gamma)*abs(b1) )*sqrt(const2a)  
  
  LCL <- EC50 + const1 - const2  
  UCL <- EC50 + const1 + const2  
  conf.pts <- c(LCL,EC50,UCL)  
  names(conf.pts) <- c("Lower","EC50","Upper")  
  return(conf.pts,conf.level,call=call)  
}
```

As found in Piegorsch, W. W. & Bailer, A. J. 1997.
Statistics for Environmental Biology and Toxicology.
Chapman and Hall, London.

```
EC50a.calc<-function(obj,conf.level=.95) {  
  # calculates confidence interval based upon Fieller's thm.  
  # modified version of EC50.calc found in P&B Fig 7.22  
  # now allows other link functions, using the calculations  
  # found in dose.p (MASS)  
  # SBC 19 May 05  
  
  call <- match.call()  
  coef = coef(obj)  
  vcov = summary.glm(obj)$cov.unscaled  
  b0<-coef[1]  
  b1<-coef[2]  
  var.b0<-vcov[1,1]  
  var.b1<-vcov[2,2]  
  cov.b0.b1<-vcov[1,2]  
  alpha<-1-conf.level  
  zalpha.2 <- -qnorm(alpha/2)  
  gamma <- zalpha.2^2 * var.b1 / (b1^2)  
  eta = family(obj)$linkfun(.5) #based on calcs in V&R's dose.p  
  EC50 <- (eta-b0)/b1  
  const1 <- (gamma/(1-gamma))*(EC50 + cov.b0.b1/var.b1)  
  const2a <- var.b0 + 2*cov.b0.b1*EC50 + var.b1*EC50^2 -  
  gamma*(var.b0 - cov.b0.b1^2/var.b1)  
  const2 <- zalpha.2/( (1-gamma)*abs(b1) )*sqrt(const2a)  
  LCL <- EC50 + const1 - const2  
  UCL <- EC50 + const1 + const2  
  conf.pts <- c(LCL,EC50,UCL)  
  names(conf.pts) <- c("Lower","EC50","Upper")  
  return(conf.pts,conf.level,call=call)  
}
```








The R Project for Statistical Computing - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://www.r-project.org/

Getting Started Slashdot: News for n... CNET News.com -- Te... The R Project for Sta... R site search s-news Archives SciViews R GUI

The R Project for Statistical Computing



[About R](#)
[What is R?](#)
[Contributors](#)
[Screenshots](#)
[What's new?](#)

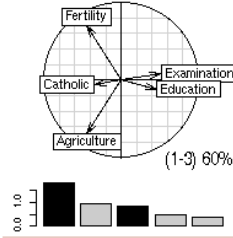
Download
[CRAN](#)

R Project
[Foundation](#)
[Members & Donors](#)
[Mailing Lists](#)
[Bug Tracking](#)
[Developer Page](#)
[Conferences](#)
[Search](#)

Documentation
[Manuals](#)
[FAQs](#)
[Newsletter](#)
[Wiki](#)
[Books](#)
[Other](#)

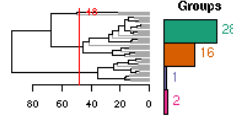
Misc
[Bioconductor](#)
[Related Projects](#)
[Links](#)

PCA 5 vars
princomp(x = data[, col = col])

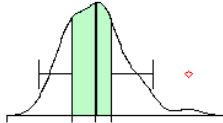


(1-3) 60%

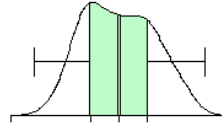
Clustering 4 groups



Factor 1 [41%]



Factor 3 [19%]



Getting started:

- R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To download R, please choose your preferred [CRAN mirror](#).
- If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

News:

- [DSC 2007](#), the 5th workshop on Directions in Statistical Computing, February 15-16, 2007, Auckland, New Zealand.
- Sources of [R 2.4.0 alpha and beta releases](#) (daily snapshot).
- [R News 6/3](#) has been published on 2006-08-24.
- [R version 2.3.1](#) has been released on 2006-06-01.
- The [R Wiki](#) provides an online forum where useRs can help other useRs.

Done

R Libraries (aka Packages)

- Suites of predefined R code
- Available for a wide variety of topics and specific analyses
- Useful examples
 - drc: Analysis of dose-response curves
 - survival: Survival analysis, including penalised likelihood
 - nlme: Linear and nonlinear mixed effects models
 - NADA: Nondetects And Data Analysis for environmental data
 - ade4: Analysis of Environmental Data : Exploratory and Euclidean method
 - Rcmdr: R Commander (GUI)

.... and many, many, more...

Installing R

- Download from CRAN site:
 - <http://www.r-project.org>
- Install the 'base' R package
 - Self-extracting installer
- Find, install R libraries (i.e., extensions)
 - Listing of many contributed packages
 - <http://cran.stat.ucla.edu/src/contrib/packages.html>
 - Use Google!
 - <http://rseek.org/>
 - Windows ...
 - Use the **Packages** menu in the Rgui

Getting data in \ out...

- Generally, two import/export options
 - Exchange via delimited ASCII file
 - R method **read.table()** (and variants)
 - Exchange with external file formats via add-on R package
 - **RDBMS**
 - ROracle: Oracle database interface for R
 - RODB: ODBC database access
 - **Commercial Statistics Packages**
 - RODB: ODBC database access
 - foreign: Read Data Stored by Minitab, S, SAS, SPSS, Stata, Systat, dBase,
 - R.matlab: Read and write of MAT files together with R-to-Matlab connectivity

Getting data in \ out...

- A word (or two) about ASCII as opposed to binary formats
 - Universal access to the data
 - Lifespan is not limited
 - Consider it the “open source” standard for data access

Getting data in \ out...

- **ASCII Data import: the read() method**
 - **read.table()**: reads comma-delimited ASCII file, creates data frame
 - **read.csv()**, **read.delim()**... also create data frame
 - But have different default input parameters
 - **read.fwf()**: reads fixed-width format ASCII file
 - **scan()**: Read data into a vector or list from the console OR file.
- **ASCII Data Export**
 - **write.table()**: writes data to an ASCII text file

Getting data in \ out...

- DEMO

Managing data ...

- The data frame

- > `mydata = read.csv("mydata.csv")`
- > `mydata[i,j]`
- > `mydata[-i,j]`
- > `mydata[[i]]`
- > `mydata$variable`

- Manipulating data

- > `subset()`
- > `merge()`
- > `sort()`
- > `order()`
- many more...

Managing data ...

- DEMO

Useful websites

- NCEAS tutorials and demonstrations
 - <http://www.nceas.ucsb.edu/scicomp/RProgTutorialsLatest.html>
- R labs/tutorials for ecologists
 - <http://ecology.msu.montana.edu/labdsv/R/>
- Vegetation analysis toolbox (lots of useful multivariate analysis and visualization tools)
 - <http://cc.oulu.fi/~jarioksa/softhelp/vegan.html>
- Analysis of bioassays using R
 - <http://www.bioassay.dk/>
- Huge effort for 'omics' data analysis
 - <http://www.bioconductor.org/>