



R for data analysis

QAC Boot camp 2018

Rob Kabacoff, Ph.D.



SEEEMS LEGIT



PART I. INTRODUCTION

Part 1 Topics

- Overview
- IDEs and GUIs
- Syntax
- Workspace
- Help
- Interactive and Batch
- Packages
- Getting Unstuck

What is R

- A language and environment for statistical computing and graphics
- Based on the "S" Language developed at Bell Labs
- R was first created by **R**oss Ihaka and **R**obert Gentleman at the Univ. of Auckland in 1993

Why use R?

- Free
- Open source
- State-of-the-art graphics and data analysis
- Platform for programming new methods
- Runs on Windows, Linux, Mac OS X
- Enormous user base

Statistical Methods

Descriptive Statistics

Experimental Design

Linear , Generalized, Nonlinear,
and Hierarchical Models

Analysis of Categorical Data

Nonparametric Analysis

Survival Analysis

Latent Variable Models

Bayesian Models

Missing Values Analysis

Cluster Analysis

Decision Trees

Data Mining

Classical Test Theory

Item Response Theory

Correspondence Analysis

Multidimensional Scaling

Meta Analysis

Structural Equation Modeling

Complex Survey Design

Time Series Analysis

Longitudinal Analysis

Social Network Analysis

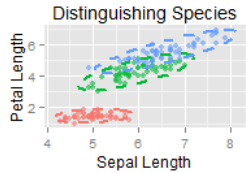
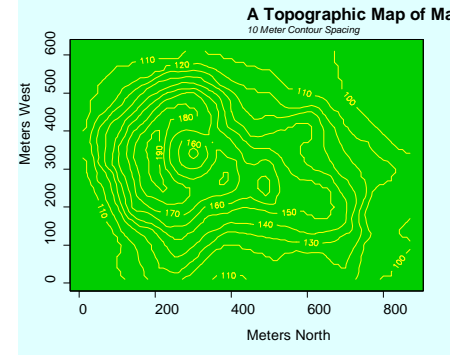
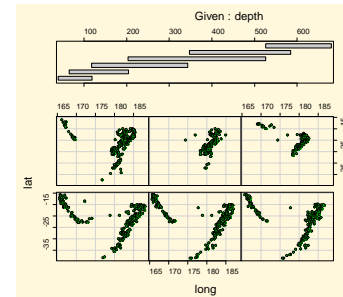
Study of Mediation and Moderation

Power Analysis

Clinical Trials

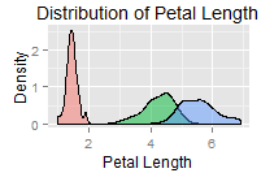
and ...

Graphs!



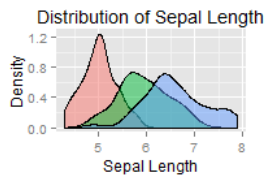
Species

- setosa
- versicolor
- virginica



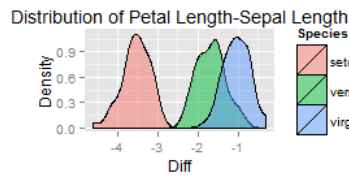
Species

- setosa
- versicolor
- virginica



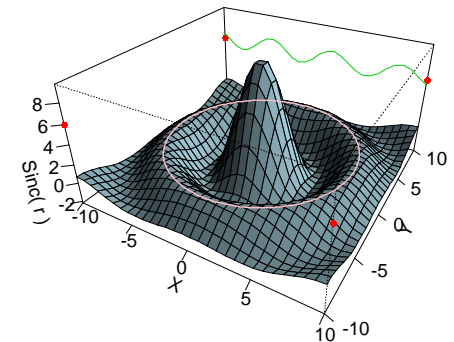
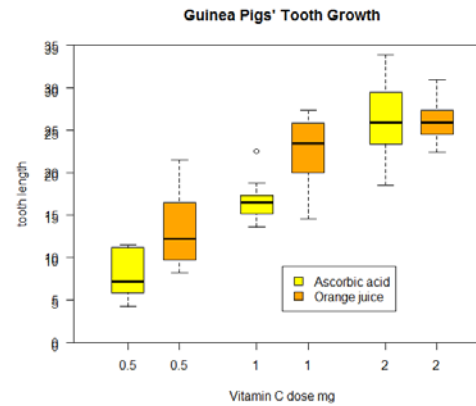
Species

- setosa
- versicolor
- virginica

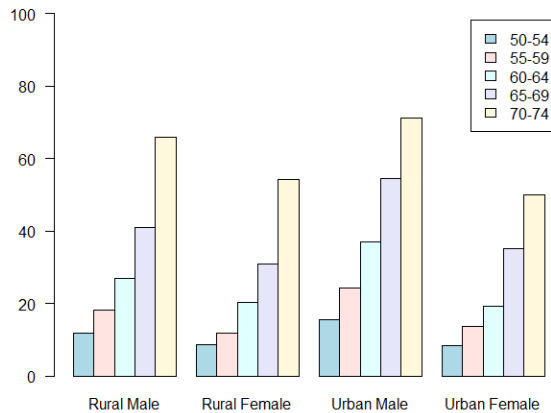


Species

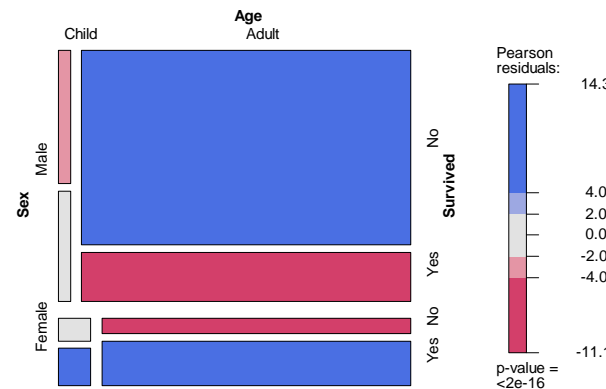
- setosa
- versicolor
- virginica



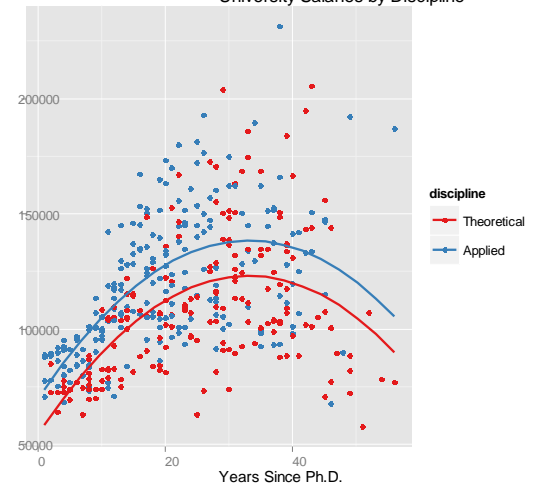
Death Rates in Virginia



Survival on the Titanic




University Salaries by Discipline



Obtaining R

- R Homepage - <http://www.r-project.org/>
- CRAN Mirrors – <http://cran.r-project.org/>



The Comprehensive R Archive Network

CRAN

- [Mirrors](#)
- [What's new?](#)
- [Task Views](#)
- [Search](#)

About R

- [R Homepage](#)
- [The R Journal](#)

Software

- [R Sources](#)
- [R Binaries](#)
- [Packages](#)
- [Other](#)

Download and Install R

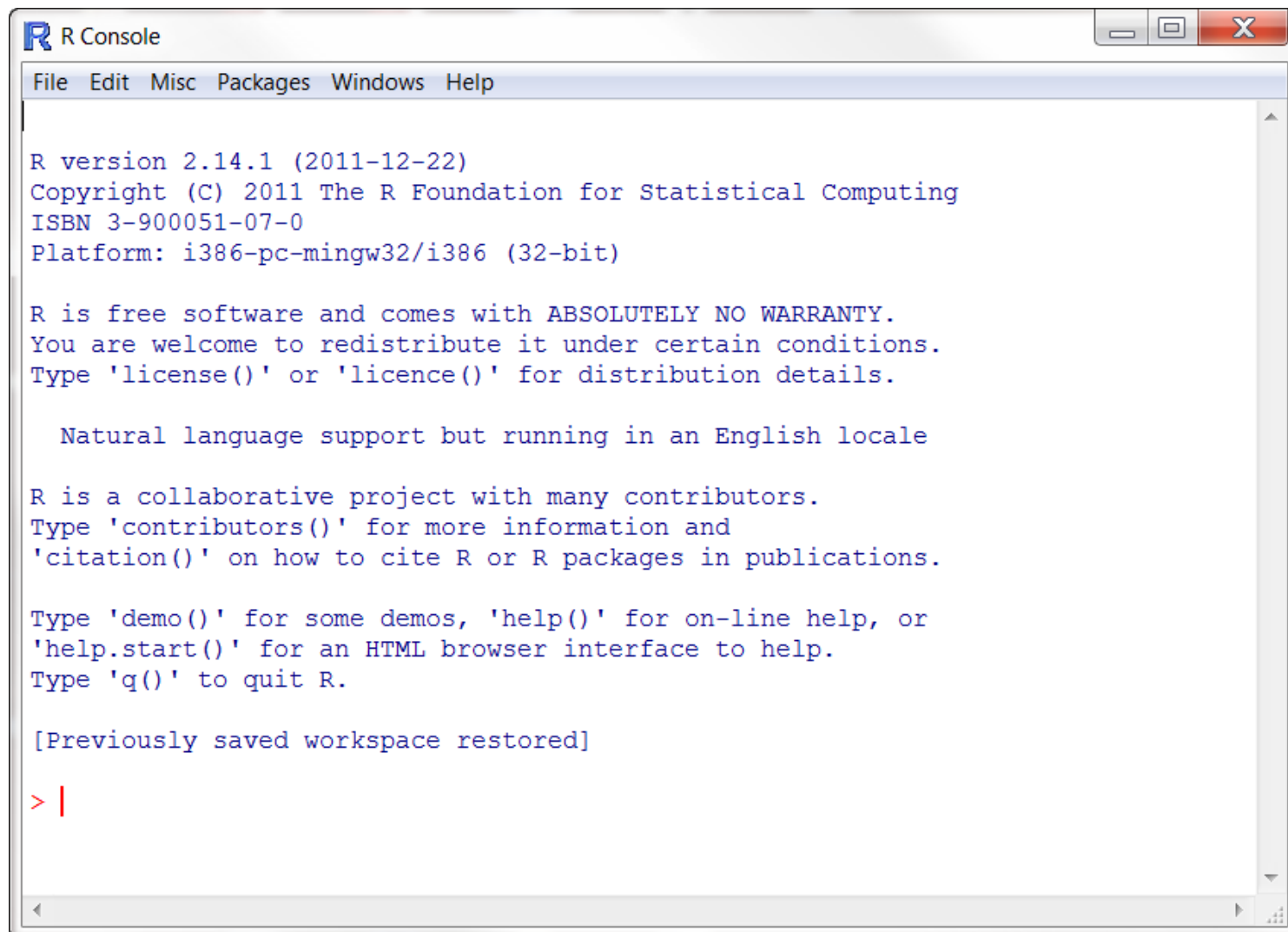
Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- [Download R for Linux](#)
- [Download R for MacOS X](#)
- [Download R for Windows](#)

Source Code for all Platforms

Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!

Default Console



The screenshot shows the R Console window with the following text:

```
R version 2.14.1 (2011-12-22)
Copyright (C) 2011 The R Foundation for Statistical Computing
ISBN 3-900051-07-0
Platform: i386-pc-mingw32/i386 (32-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

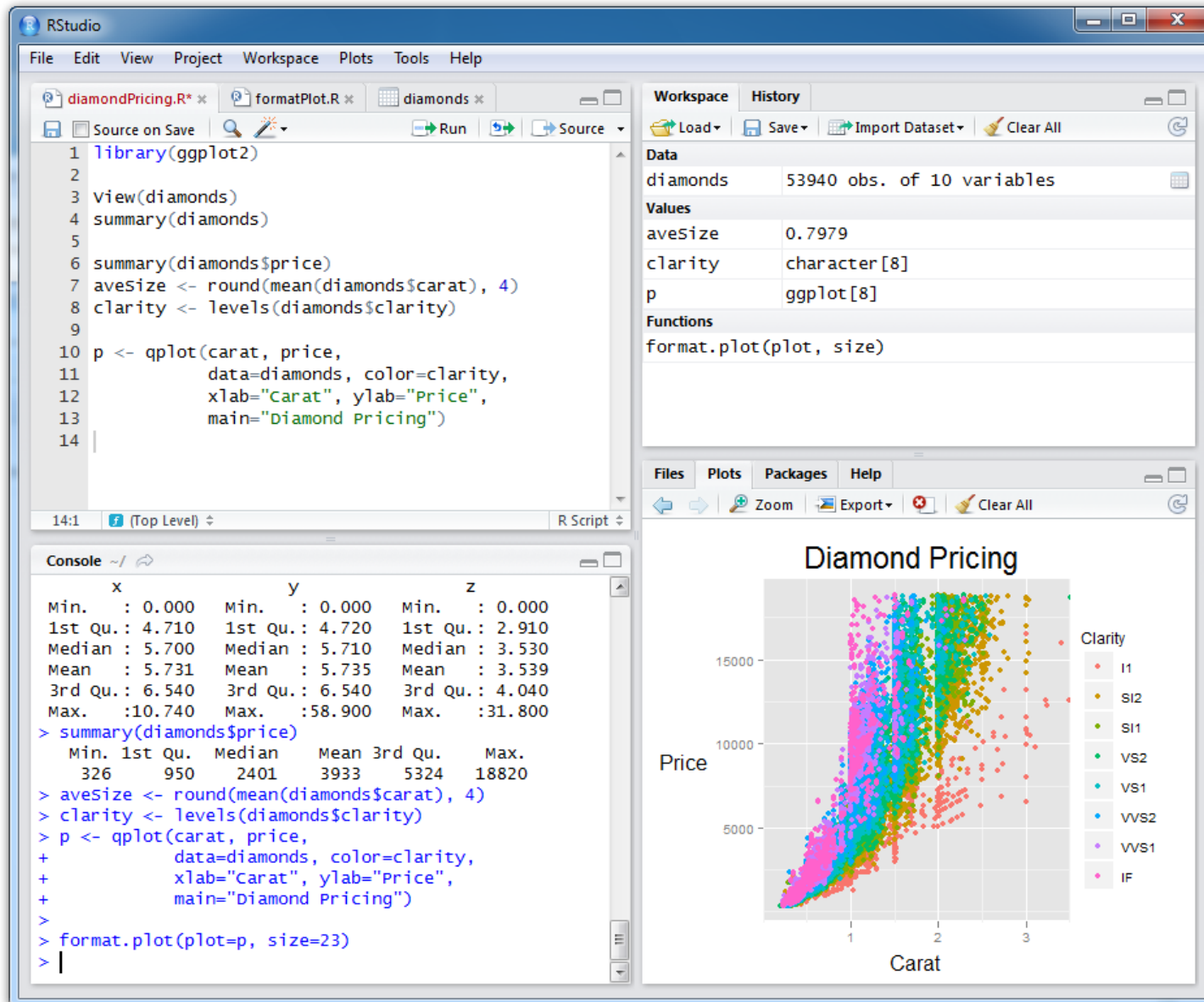
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Previously saved workspace restored]

> |
```

Rstudio (<http://rstudio.org/>)



Working with R

- Case-sensitive interpreted language
- Enter commands at prompt (>) or in batch
- Statements consist of functions and assignments
- Comments are preceded by #
- semi-colon separates statements on a line

```
x <- c(4, 4, 5, 6, 7, 2, 9)
length(x) ; mean(x)
plot(x) # plot the vector
```

Some definitions

- An **object** is anything that can be assigned to a variable.
 - Includes constants, data structures, functions, and even graphs
- Objects have a **mode** - how the object is stored.
- Objects have a **class** - which tells generic functions like `print()` how to handle it.

R Workspace

- Your current R working environment
- Includes any user-defined objects (vectors, matrices, functions, data frames, lists)
- At end of session you can save image of the workspace and it will be automatically loaded next time R starts

Functions for managing workspace

Function

Action

`getwd()`

List current working directory

`setwd("mydirectory")`

Change the current working directory to mydirectory

`rm(object)`

Delete object

`save(objectlist, file="myfile")`

Save specific objects to a file

`load("myfile")`

Load a workspace into the current session
(default = .RData)



Getting Help

Function

Action

`help("foo")`
`?foo`

Help on function foo (the quotation marks are optional)

`help.search("foo")`
`??foo`

Search the help system for instances of the string foo

`help(package="foo")`

Help on package foo

`example("foo")`

Examples of function foo (the quotation marks are optional)

`data()`

List all available datasets for currently loaded packages

Packages

- Collections of R functions, data, and compiled code in well-defined format
- Massively extend the functionality of R
- Thousands of user written packages on CRAN
<http://cran.r-project.org/web/packages>

Working with Packages

- `install.packages("packagename")`
- `update.packages()`
- `library(packagename)`
- `help(package="packagename")`

Sample session

```
install.packages("vcd")
```

```
help(package="vcd")
```

```
library(vcd)
```

```
help(Arthritis)
```

```
Arthritis
```

```
example(Arthritis)
```

Common Mistakes

- **Using the wrong case.**
help(), Help(), and HELP() are three different functions (only the first will work).
- **Forgetting to use quote marks when they are needed.**
install.packages("gclus") will work, while
install.packages(gclus) will generate an error.
- **Forgetting to include the parentheses in a function call.**
help() rather than help.
Even if there are no options, you still need the ().

Common Mistakes (2)

- **Using the \ in a path name on Windows.**

R sees the backslash character as an escape character.

`setwd("c:\mydata")` will generate an error. Use `setwd("c:/mydata")` or `setwd("c:\\mydata")` instead.

- **Using a function from a package that is not loaded.**

The function `order.clusters()` is contained in the `gclus` package. If you try to use it before loading the package, you will get an error.



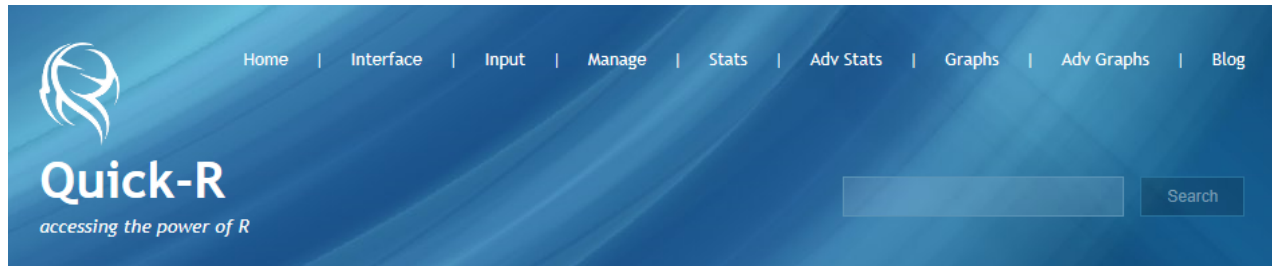
RESOURCES

Look it up on Google

- “summary statistics in r”
- “specifying colors in a barchart in r”
- “rotating labels in ggplot2”
- Error messages



www.statmethods.net



Top Menu

[Home](#)

[The R Interface](#)

[Data Input](#)

[Data Management](#)

[Basic Statistics](#)

[Advanced Statistics](#)

[Basic Graphs](#)

[Advanced Graphs](#)

[Blog](#)

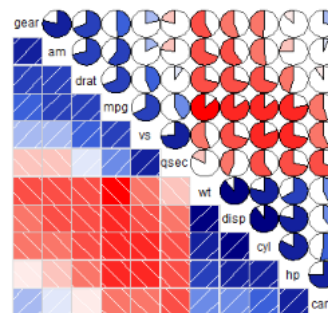
R in Action



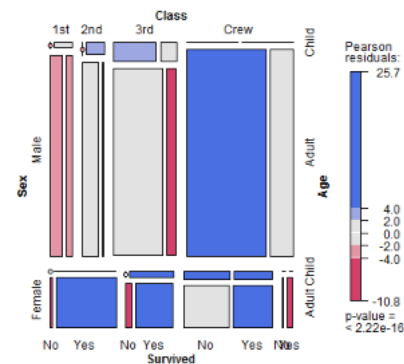
After two years of continuous

About Quick-R

Correlations Among Auto Characteristics



Who Survived the Titanic?



R is an elegant and comprehensive statistical and graphical programming language. Unfortunately, it can also have a [steep learning curve](#). I created this website for both current R users, and experienced users of other statistical packages (e.g., SAS, SPSS, Stata) who would like to transition to R. My goal is to help you quickly access this language in your work.

I assume that you are already familiar with the [statistical methods](#) covered and instead provide you with a roadmap and the code necessary to get started quickly, and orient yourself for future learning. I designed this web site to be an easily accessible reference. Look at the [sitemap](#) to get an overview.

Find the functions you need

CRAN Task Views

Bayesian	Bayesian Inference
ChemPhys	Chemometrics and Computational Physics
ClinicalTrials	Clinical Trial Design, Monitoring, and Analysis
Cluster	Cluster Analysis & Finite Mixture Models
DifferentialEquations	Differential Equations
Distributions	Probability Distributions
Econometrics	Computational Econometrics
Environmetrics	Analysis of Ecological and Environmental Data
ExperimentalDesign	Design of Experiments (DoE) & Analysis of Experimental Data
Finance	Empirical Finance
Genetics	Statistical Genetics
Graphics	Graphic Displays & Dynamic Graphics & Graphic Devices & Visualization
HighPerformanceComputing	High-Performance and Parallel Computing with R
MachineLearning	Machine Learning & Statistical Learning
MedicalImaging	Medical Image Analysis
Multivariate	Multivariate Statistics
NaturalLanguageProcessing	Natural Language Processing
OfficialStatistics	Official Statistics & Survey Methodology
Optimization	Optimization and Mathematical Programming
Pharmacokinetics	Analysis of Pharmacokinetic Data
Phylogenetics	Phylogenetics, Especially Comparative Methods
Psychometrics	Psychometric Models and Methods
ReproducibleResearch	Reproducible Research
Robust	Robust Statistical Methods
SocialSciences	Statistics for the Social Sciences
Spatial	Analysis of Spatial Data
SpatioTemporal	Handling and Analyzing Spatio-Temporal Data
Survival	Survival Analysis
TimeSeries	Time Series Analysis

CRAN
Task Views

Finding Functions : <http://www.rdocumentation.org>

Q Start searching the documentat

TASK VIEWS

Bayesian

ChemPhys

ClinicalTrials

Cluster

DifferentialEquations

Distributions

Econometrics

Environmetrics

ExperimentalDesign

Finance

Genetics

gR

Graphics

HighPerformanceComputing

MachineLearning

MedicalImaging

MetaAnalysis

Multivariate

NaturalLanguageProcessing

NumericalMathematics

OfficialStatistics

Optimization

Pharmacokinetics

Phylogenetics

Psychometrics

ReproducibleResearch

Robust

SocialSciences


Spatial


SpatioTemporal


R Documentation

Discussion About Rdocumentation package

Search the R documentation of **6693** R packages and **136470** R functions:

 CRAN

 Bioconductor
OPEN SOURCE SOFTWARE FOR BIOINFORMATICS

 GitHub

Rdocumentation is a tool that helps you easily find and browse the documentation of all current and some past packages on CRAN. Click on the search bar at the top left for instant search or fill out the forms below for advanced search!

All Fields

Package Name

Function Name

Title

Description

Author(s)

Start search

Top Ranked CRAN Packages

Week | Month | All time

#	Package	#
1	- ggplot2	121570
2	- DBI	73928
3	⬆ rJava	67279
4	⬇ Rcpp	65982
5	⬇ stringr	65607
6	⬇ plyr	64181
7	⬆ LPCM	61484
8	⬇ digest	51510
9	⬆ foreach	52336
10	⬆ iterators	51326

New Packages

#	Package
1	cleanUpdTSeq
2	meme
3	LakeMetabolizer
4	derfinder
5	RchyOptimyx
6	ShinyHelper
7	soepR
8	googleCharts
9	FieldSpectra
10	animaker



I STILL DON'T SEE IT

When you're really stuck

- Stack Overflow

<http://stackoverflow.com/questions/tagged/r>