R Resources

- Ujjwal Karn
  - DataScienceR
  - xda
- Tim Smith
  - aRrgh: a newcomer's (angry) guide to R
- David Robinson
  - Data analysis and visualization using R
- Rob Kabacoff
  - Introduction to R
- Tony Ojeda
- Getting started with R
  - Learning R in Seven Simple Steps
- Institute for Digital Research and Education
  - Resources to help you learn and use R.
- An Introduction to Statistical Learning
- Swirl
- Antony Unwin (2015).
- Jared Lander (2014).
- John Maindonald (2008).
- Jason Brownlee
- Microsoft
- Twitter #rstats
- Discovering Statistics Using R
- Quick Guide to Learn Statistics for R Users (with Titanic Data Set)
- From 0 to R with Google Analytics
- From level[1] to level[2] with R and Google Analytics: ggplot2
- Statistical Rethinking: A Bayesian Course with Examples in R and Stan.
- Bartomeus Lab
- ggplot2
  - ggplot2 Main Site
  - ggplot2 Documentation
  - The R Graphics Cookbook
  - The Best ggplot2 Cheat Sheet
  - ggplot2's Wikipedia Page
  - Data visualisation for Social Science
- Dplyr
  - Data manipulation with R
- Enhance Data Science
  - R packages for interactive plot and visualisation 1
  - R packages for interactive plot and visualisation 2
- Shiny
  - Tutorial

- Articles
- Interactive
- Examples
  - New Zealand Tourism Data Dashboard
  - Googly
  - R Shiny packages

Ploty
- Carson Sievert

Using R packages and education to scale Data Science at Airbnb
Ilia Karmanov
- Neural networks from scratch (in R)

Burakhimmetoglu
- Web scraping with R

R Tutorials (Douglas Wiig)
Renata de Souza
- Getting started with regression in R

Rahul Saxena
- K-Nearest Neighbor Classifier in R

Manish Saraswat
- SQL Tutorial on Data Analysis in R

Analytics Vidhya
- Python and R codes for common machine learning algorithms

Text Mining with R
Jake Russ
- A Stargazer Cheatsheet

Thomas Pedersen
- ggraph: a grammar of graphics for relational data

Andrea Spano
- Building views with R

Brian Ripley
- Nnet
- Visualization neural nets
- Update

Richard Iannone
- GitHub
- Validate data

Jeroen Ooms
- Advanced graphics and image processing in R

Matt Upson
- Dealing with dates

Ilya Kashnitsky
- Arranging subplots with ggplot2

- Analytics Vidyha
  - Tree Based Modeling
- Mara Averick
  - Tidyeval resource round up
  - R workflow fun
- Represtools
  - Overview
- Christian Robert & George Casella
  - Introducing Monte carlo Methods with R
- Jan Holtz
  - The R Graph Gallery

- Some Sport Links
  - Stephanie Kovalchik
    - UseRSportTutorial
    - Sports Analytics with R
  - Tony Hirst
    - Sports Data and R - Scope for a Thematic View? (2015)
  - Thomas Lorigan
    - Teouch Analytics
  - James Curley
    - GitHub Pages
    - English soccer data
    - Shinyapps
    - Analyzing Basketball Data with R.
  - Tinniam Ganesh
    - Introducing cicketr!
  - Manuel Eugster's Infrastructure for Sports Analytics.
  - Guillermo Vinue's R Package for Analysis of Anthropometric Data
  - Revolutions Analytics
  - Max Marchi and Jim Albert
    - Analyzing Baseball Data with R (2013).
  - Norm Matloff
    - Open Textbook: From Algorithms to Z-Scores.
  - Reddit Sports Analytics
  - Julien Jacques, Quentin Grimonprez and Christophe Biernacki
  - Isaac Petersen
  - Thomas Heslop
    - R Resources
    - R Package: fantasysocceR
    - R Package: servevolleyR
  - Simon Muller

Nick Zani
Scottish Hill Race Data
Mara Averick
    Stapig PuzzleR: NBA #1
    Point(s) of No Return - Statpig PuzzleR #2
Rajiv Shah
    Exploring NBA SportVu Movement Data
Tanya Cashorali
    NBA Player Movement Data in R
    Exploring Baseball Data with R
    Jim Albert (2016) Six Home Runs at Riverfront
Andy Field
    The referee's a …
Open Football
Aaron Miles
    Game Changers: Assessing QB Win Probability Added Using R
Mark Horowitz
    nflscrapR
CRAN
    Sharon Machlis
    Sean Lahman's Baseball Database
    Stattleship
    Yorkr
Christopher Long
Chris Mann
    Tidy Baseball data
    GitHub
David Robinson
    Ebbr package for empirical Bayes estimation (Baseball)
DexLab
    NFL
Dan Riskin
    How much harder is it to win a best-of-seven compared to one game?
RStrava
Brian Boulay
    Sports Data Analytics, Spring 2017
Emmanuel Perry
    The CoRsica Package for Hockey Analysis in R (0.0 An Introduction)
    The CoRsica Package for World Hockey Analysis in R (0.1 Hello World)
    The CoRsica Package for Hockey Analysis in r (0.2 Fundamentals)
Culture of Insight
    Animated Maps in R

Introduction

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.

Information about the R project can be found here.

Information about downloading and installing R can be found here. There is an introduction to R in this article from Volume 1(1) of R News (2001).

The R Environment

Is an integrated suite of software facilities for data manipulation, calculation and graphical display. It includes:

- an effective data handling and storage facility,
- a suite of operators for calculations on arrays, in particular matrices,
- a large, coherent, integrated collection of intermediate tools for data analysis,
- graphical facilities for data analysis and display either on-screen or on hardcopy, and
- a well-developed, simple and effective programming language which includes conditionals, loops, user-defined recursive functions and input and output facilities.

It is 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 Documentation

This page has a range of R manuals. They can be downloaded as PDF files, EPUB files, or directly browsed as HTML.

An Introduction to R

R Data Import/Export

R Installation and Administration

Other R Resources

R-bloggers & DataCamp

How to Learn R.

- Getting started: The basics of R
- Setting up your machine
- R packages
- Importing your data into R
- Data Manipulation
- Data Visualization
- Data Science & Machine Learning with R
- Reporting Results in R
- Next steps

Why you should learn R first

RStudio

- RStudio Start Here
- RStudio Cheat Sheets

- R Markdown Cheat Sheet
- Shiny
- Data Visualisation
- 2017 Conference

Hadley Wickham

R for Data Science

http://hadley.nz/

- R for Data Science (with Garrett Grolemund) introduces the key tools for doing data science with R.
- ggplot2: elegant graphics for data analysis use ggplot2 to create graphics that help you understand your data.
- Advanced R
- R packages teaches good software engineering practices for R, using packages for bundling, documenting, and testing your code.
- RStudio workshops

The tidyverse

Swirl

You're a student

Storybench

Tutorials

How to explore and manipulate a dataset from the fivethirtyeight package in R

Marcos Vital

Teach-R Project

Ujjwal Karn

DataScienceR

xda

Tim Smith

aRrgh: a newcomer's (angry) guide to R

David Robinson

Data analysis and visualization using R

Rob Kabacoff

Introduction to R

Tony Ojeda

Guide to R for Excel users.

Getting started with R

http://www.ats.ucla.edu/stat/r/

Martjin Theuwissen

Learning R in Seven Simple Steps

Institute for Digital Research and Education

Resources to help you learn and use R.

An Introduction to Statistical Learning

Swirl

A collection of interactive courses for Swirl

Antony Unwin (2015).

Graphical Data Analysis with R. Boca Raton: Taylor & Francis.

Jared Lander (2014).

R for Everyone. New Jersey: Pearson Education.

John Maindonald (2008).

Using R for Data Analysis and Graphics: Introduction, Code and Commentary. Canberra: Centre for Mathematics and Its Application, Australian National University.

Jason Brownlee

Jason Brownlee (2016a). Better understand your data in R using descriptive statistics.
Jason Brownlee (2016b). Better understand your data in R using visualization.

- Visualization packages
- Univariate visualization
- Multivariate visualization

Jason Brownlee (2016c). Use R for machine learning.
Jason Brownlee (2016d). How to use R for machine learning.

Microsoft

R Application Network (2016).

- What is R?
- Download
- Packages

Twitter #rstats

Discovering Statistics Using R

Andy Field, Jeremy Miles and Zoe Field (2012). London: Sage Publishing.
(Video introduction)

Quick Guide to Learn Statistics for R Users (with Titanic Data Set)

Manish Saraswat (2015)

From 0 to R with Google Analytics

Tim Wilson (17 January 2016)

From level[1] to level[2] with R and Google Analytics: ggplot2

Tom Miller (8 February 2016)

Statistical Rethinking: A Bayesian Course with Examples in R and Stan.

Richard McElreath (2015). Boca Raton: CRC Press.

Bartomeus Lab

I am looking for fun (public) datasets to use in data manipulation and visualization. I would like to use a single dataset that has some easy variables for the first days, but also some more challenging ones for the final days. And I want that when I put exercises, the students* are curious about finding out the answer.

ggplot2

ggplot2 Main Site

ggplot2 Documentation

The R Graphics Cookbook

The Best ggplot2 Cheat Sheet

ggplot2's Wikipedia Page

Data visualisation for Social Science

Dplyr

Data manipulation with R

Enhance Data Science

R packages for interactive plot and visualisation 1

R packages for interactive plot and visualisation 2

Shiny

Tutorial

Articles

Interactive

Examples

New Zealand Tourism Data Dashboard

Googly

The Googly Shiny app is based on my R package 'yorkr' which is now available in CRAN. The R package and hence this Shiny app is based on data from Cricsheet.

R Shiny packages

Ploty

Carson Sievert

Using R packages and education to scale Data Science at Airbnb

Ilia Karmanov

Neural networks from scratch (in R)

Burakhimmetoglu

Web scraping with R

R Tutorials (Douglas Wiig)

Renata de Souza

Getting started with regression in R

Rahul Saxena

K-Nearest Neighbor Classifier in R

Manish Saraswat

SQL Tutorial on Data Analysis in R

Analytics Vidhya

Python and R codes for common machine learning algorithms

Text Mining with R

Jake Russ

A Stargazer Cheatsheet

Thomas Pedersen

ggraph: a grammar of graphics for relational data

Andrea Spano

Building views with R

Brian Ripley

Nnet

Visualization neural nets

Update

Richard Iannone

GitHub

Validate data

Jeroen Ooms

Advanced graphics and image processing in R

Matt Upson

Dealing with dates

Ilya Kashnitsky

Arranging subplots with ggplot2

Analytics Vidyha

Tree Based Modeling

Mara Averick

Tidyeval resource round up

R workflow fun

Represtools

Overview

Christian Robert & George Casella

Introducing Monte carlo Methods with R

Jan Holtz

The R Graph Gallery

Some Sport Links

Stephanie Kovalchik

UseRSportTutorial

Sports Analytics with R

Tony Hirst

Sports Data and R - Scope for a Thematic View? (2015)

Thomas Lorigan

Teouch Analytics

James Curley

GitHub Pages

English soccer data

Shinyapps

- engsoccerbeta - Exploring historical trends in English soccer data https://jalapic.shinyapps.io/engsoccerbeta/
- cricket - Interactive ggvis charts of every test match batsman https://jalapic.shinyapps.io/cricket/
- soccerteams - Interactive ggvis chart of historical soccer teams/seasons https://jalapic.shinyapps.io/soccerteams/

- gamebygame - Comparing EPL teams season by season performances - using shinydashboard https://jalapic.shinyapps.io/gamebygame/
- basicstats - Using shinydashboard to explore data viz and basic stats https://jalapic.shinyapps.io/basicstats/

Twitter account.

Columbia university website Curley Lab.

- R packages
- R instruction

Analyzing Basketball Data with R.

Max Marchi & Jim Albert (2014).Boca Raton: CRC Press.

Tinniam Ganesh

- Introducing cicketr!
- cricketr digs the Ashes (July, 2015)
- cricketr plays the ODIs (August, 2015)
- cricketr adapts to the Twenty20 international (August, 2015)
- Sixer cricketr's Shiny App (November, 2015) (see also Andrew Clark)
- A short video tutorial on cricketr (December, 2015)

Manuel Eugster's Infrastructure for Sports Analytics.

"a selection of data sets, functions to fetch sports data, examples, and demos --
with the ambition to develop bit by bit a set of classes to represent general concepts of sports analysis." Link to R package. See also Archetypal Athletes (2011), Performance Profiles (2012), Archetypal Analysis (2014) and Probabilistic Archetypal Analysis (2014).

Guillermo Vinue's R Package for Analysis of Anthropometric Data

This paper (2015) presents a new R package, called Anthropometry, which is available on the Comprehensive R Archive Network. It brings together some statistical methodologies concerning clustering, statistical shape analysis, statistical archetypal analysis and the statistical concept of data depth, which have been especially developed to deal with anthropometric data. They are proposed with the aim of providing effective solutions to some common anthropometric problems, such as clothing design or workstation design (focusing on the particular case of aircraft cockpits). Thesis (2014): *Development of Statistical Methodologies Applied to Anthropometric Data Oriented Towards the Ergonomic Design of Products*. See also, Guillermo Vinue, Irene Epifanio and Sandra Alemany (2015) on archetypoids.

Revolutions Analytics

What is R?

Andrie de Vries (2015). Using segmented regression to analyse world record running times.
David Smith (2015). Using R to analyze pro motorcycle racing.
Mark Malter (2015). Baseball Game Wins Probability Calculator.
Mark Malter (2015). Situational Baseball: Analyzing Runs Potential Statistics.

Max Marchi and Jim Albert

Analyzing Baseball Data with R (2013).

Norm Matloff

The Freqparcoord Package for Multivariate Visualization (2014). More on Freqparcoord (2014). (See also Novel Methods for Parallel Coordinates, 2014, and this updated version.)
Simpson's Paradox is Back (2014).

Open Textbook: From Algorithms to Z-Scores.

(For additional R resources see Chapter 1 of The Art of R Programming R, a rough and partial draft of that book.
Boosting (2015).

Reddit Sports Analytics

Eduardo Maia (2015) Visualising the 2015 NBA Draft in R.
Eduardo Maia (2015) What AFL stadiums pull the biggest crowds?
Eduardo Maia (2015) How to create NBA shot charts in R.

Julien Jacques, Quentin Grimonprez and Christophe Biernacki

Rankcluster: An R Package for Clustering Multivariate Partial Rankings (2014).
In this example, we analyse the rankings of these four teams (Manchester United, Liverpool, Arsenal and Chelsea) in the English championship (Premier League) and their rankings according to the UEFA coefficient, a European statistic on football teams based on the results of European football competitions and used for ranking and seeding teams in international competitions. R> data(big4)

Isaac Petersen

Why R is Better Than Excel for Fantasy Football (and most other) Data Analysis (2014).

Although there are R packages for importing Excel data directly it is generally best to export the Excel data to a .csv file and then import the .csv file in R using the read.csv() function. Here's the code for importing a .csv file in R and storing it in an object called 'mydata': mydata <- read.csv("C:/excelData.csv").

Thomas Heslop

- R Resources
- R Package: fantasysocceR
- R Package: servevolleyR

Simon Muller

R: Analysis of a Sport Event (2011)
I started R and analysed the LAC Degerloch Volkslauf 2010; a 10km race near Stuttgart-Hoffeld. A post written in response to Two Castles Run (2011)

Nick Zani

Scraping Data from Cricinfo (2015) Batting and bowling performances for individual players can be broken down by innings and the XML package in R provides a convenient method to collect this, since the data is stored in HTML tables. The readHTMLTable function as part of this package will return all the tables present on a web page. Github.

Scottish Hill Race Data

- StaSci.org
- Finzi
- Anthony Atkinson (1986). [Influential Observations, High Leverage Points, and Outliers in Linear Regression]: Comment: Aspects of Diagnostic Regression Analysis. *Statistical Science*, 397-402.
- Anthony Atkinson (1988). Transformations unmasked. *Technometrics 30*, 311-318.
- Gunther Sawitzki, G. (1990). Tools and Concepts in Data Analysis 1.
- Niall Adams & David Stephens (2005). Introduction to R.
- John Maindonald (2007). Hill Race Data.
- John Maindonald & John Braun (2010). An example based approach.
- Thomas O'Gorman (2012). Adaptive Tests.
- SOCR (2012). Multiple linear regression.
- David Hoaglin (2013). Making Sense of Coefficients in Multiple Regression.
- R programming: worksheet 2
- Introduction to R
- Zhang: linear regression
- BMA (2015). p.40

Mara Averick

Stapig PuzzleR: NBA #1

Point(s) of No Return - Statpig PuzzleR #2

Rajiv Shah

Exploring NBA SportVu Movement Data

Tanya Cashorali

NBA Player Movement Data in R

Exploring Baseball Data with R

Jim Albert (2016) Six Home Runs at Riverfront

Andy Field

The referee's a …

I collected data from football-lineups.com for the past 10 years (2005-6 up to the Chelsea fixture on 19th Sept 2015). I am only looking at English Premier League (EPL) games. I have filtered the data to include only the opponents mentioned above, so we'll get a fair comparison of Arsenal's results under MD against their results against those same clubs when MD is not officiating. The data are here as a CSV file.

A scientist essentially sets up two competing hypotheses. The first is that there is no effect of MD, that is Arsenal's results under MD are exactly the same as under other referees. This is known as the *null hypothesis*. The second is the opposite: that there is an effect of MD, that is Arsenal's results under MD are

different compared to other referees. This is known as the *alternative hypothesis*. We can use statistical tests to compare these hypotheses.

A frequentist approach
To grab the file into R run:

arse<-read.csv(file.choose(), header = T)

Open Football

A free and open public domain football database & schema for use in any (programming) language (e.g. uses plain datasets).

Aaron Miles

Game Changers: Assessing QB Win Probability Added Using R

Up until now, most of my sports analytics have focused on basketball (my favorite sport). But this blog should be more sport-agnostic, and I just got my hands on some data from Armchair Analysis (which I would HIGHLY recommend) and figured I'd give some football analytics a spin! One metric that has intrigued me across multiple sports is Win Probability Added (WPA). Essentially, this metric measures the effect a player has on the chances of their team winning the game. (Github)

Mark Horowitz

nflscrapR

Will scrape NFL play-by-play and boxscore data across full seasons

CRAN

Sharon Machlis

Packages that are on CRAN can be installed on your system by using the R command **install.packages("packageName")** -- you only need to run this once.

GitHub packages are best installed with the devtools package -- install that once with **install.packages("devtools")** and then use that to install packages from GitHub using the format **devtools::install_github("repositoryName/packageName")**. Once installed, you can load a package into your working session once each session using the format **library("packageName")**.

Some of the sample code below comes from package documentation or blog posts by package authors. For more information about a package, you can run **help(package="packageName")** in R to get info on functions included in the package and, if available, links to package vignettes (R-speak for additional documentation). To see sample code for a particular function, try **example(topic="functionName" , package="packageName")** or simply **?functionName** for all available help about a function including any sample code (not all documentation includes samples).

Sean Lahman's Baseball Database

(See also, Kris Eberwein's blog post Hacking the New Lahman Package)

Stattleship

(See also, Tanya Cahorali's blog post Stop ScrAPIng For Sports Data, Start Stattleshiping)

Yorkr

Analyzing performances of cricketers and cricket teams based on 'yaml' match data from Cricsheet <http://cricsheet.org>

Christopher Long

Github (Football, Basketball, Ice Hockey, Volleyball, Web scrapers)

Chris Mann

Tidy Baseball data

GitHub

David Robinson

Ebbr package for empirical Bayes estimation (Baseball)

DexLab

NFL

Dan Riskin

How much harder is it to win a best-of-seven compared to one game?

NHL seven-game series discussion.

RStrava

Brian Boulay

Sports Data Analytics, Spring 2017

(Baseball)

Emmanuel Perry

The CoRsica Package for Hockey Analysis in R (0.0 An Introduction)

The CoRsica Package for World Hockey Analysis in R (0.1 Hello World)

The CoRsica Package for Hockey Analysis in r (0.2 Fundamentals)

Culture of Insight

Animated Maps in R