

GEOG 5680

Introduction to R

00: Class Introduction

Simon Brewer

Geography Department
University of Utah
Salt Lake City, Utah 84112
`simon.brewer@geog.utah.edu`

May 11, 2021

Course Goals

- This course is designed to give an intensive introduction to R for analysis, programming and as a graphical tool. The aims are to:
 - Introduce R as a data analysis and statistical software tool
 - Learn about manipulating data in R
 - Use scripts and functions to help analyzing data
 - Cover basic to more advanced plotting
 - Look at the add-on packages that extend R's functions
 - Introduce basic statistical modeling in R

Course Goals

- This course is designed to give an intensive introduction to R for analysis, programming and as a graphical tool. The aims are to:
 - Introduce R as a data analysis and statistical software tool
 - Learn about manipulating data in R
 - Use scripts and functions to help analyzing data
 - Cover basic to more advanced plotting
 - Look at the add-on packages that extend R's functions
 - Introduce basic statistical modeling in R
- It is aimed at people with little to no prior experience with R or programming, although some basic knowledge of statistics is assumed

Course Goals

- This course is designed to give an intensive introduction to R for analysis, programming and as a graphical tool. The aims are to:
 - Introduce R as a data analysis and statistical software tool
 - Learn about manipulating data in R
 - Use scripts and functions to help analyzing data
 - Cover basic to more advanced plotting
 - Look at the add-on packages that extend R's functions
 - Introduce basic statistical modeling in R
- It is aimed at people with little to no prior experience with R or programming, although some basic knowledge of statistics is assumed
- Designed as a series of modules: short video intro + computer exercises

Syllabus

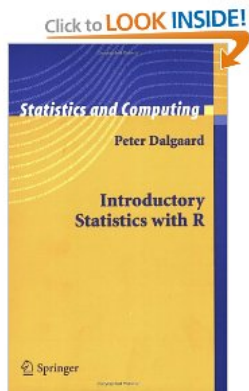
- 00: Class introduction
- 01: Data import and export
- 02: Data manipulation I
- 03: Basic plotting
- 04: Control functions and loops
- 05: R functions
- 06: Writing reports in R
- 07: Data manipulation II

Syllabus

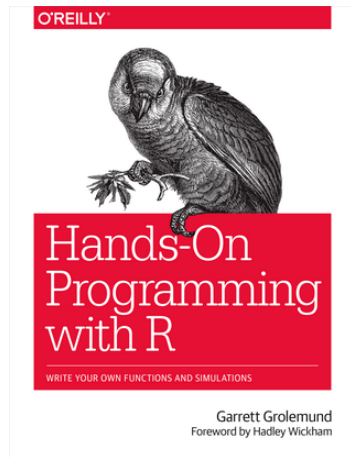
- 08: Extending basic plots with **ggplot2**
- 09: Simple inference tests
- 10: Introduction to statistical modeling in R
- 11: Data manipulation III
- 12: Making maps in R
- 13: Using Github with R
- 14: Web applications with Shiny
- 15: Optional modules (mixed-effects models, interactive maps, ...)

Reading material

Dalgaard, Introductory statistics with R
(Springer)



Grolemund, Hands-on Programming with R
(O'Reilly)



Course assessment

- In-class exercises (75 pts)
 - 2-3 short exercises per class
 - Designed to make you repeat the methods covered
 - Provide R code as a script
 - Copy-paste results and/or figures to Word document
 - Submission through Canvas
 - All exercises to be submitted by Wednesday, June 25

Course assessment

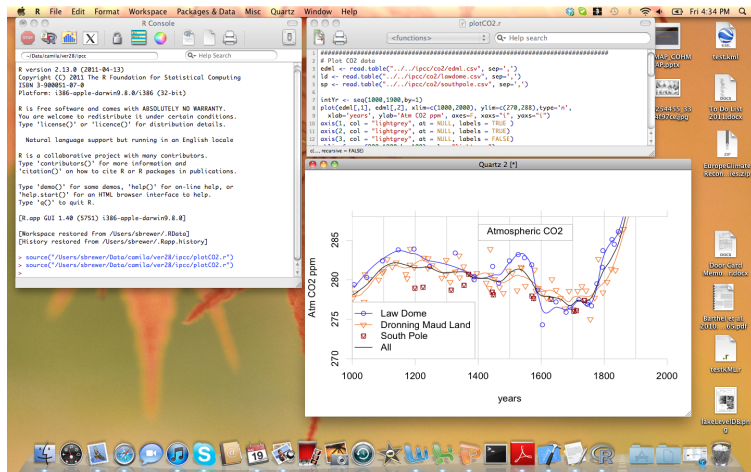
- In-class exercises (75 pts)
 - 2-3 short exercises per class
 - Designed to make you repeat the methods covered
 - Provide R code as a script
 - Copy-paste results and/or figures to Word document
 - Submission through Canvas
 - All exercises to be submitted by Wednesday, June 25
- Course project (25 pts)
 - Analysis of a dataset using one or more of the techniques covered in class (or use R to explore other techniques)
 - Project can be either:
 - One of three predefined datasets
 - Your own dataset

Course Project

- Examples of projects
 - Investigating the link between house characteristics and price
 - Factors influencing companies profits
 - Results of cloud seeding experiment
- Project report - worth 55% of overall grade
 - Projects to be written in R Markdown (covered in module 06)
 - Include:
 - R Code
 - Results including figures
 - Brief discussion of project and results
 - Due Friday, June 25 through Canvas

Introduction to R

- GNU GPL (free) statistical language and environment
- Comprehensive R Archive Network (CRAN)
- www.r-project.org



What is R?

R is an integrated suite of software facilities for data manipulation, calculation and graphical display. Among other things it has

- 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 directly at the computer or on hardcopy, and
- a well developed, simple and effective programming language (called 'S') which includes conditionals, loops, user defined recursive functions and input and output facilities.

What is R?

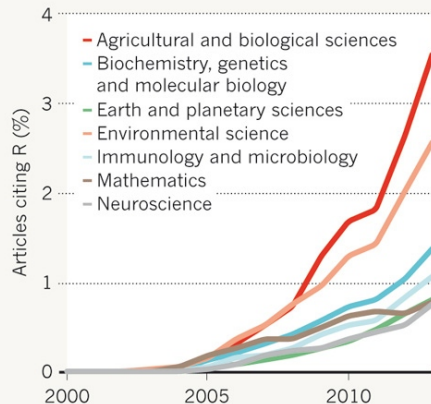
- Programming language + interfaces
- Extensive graphics capability
- Easily transferable — runs on Mac/Windows/Unix
- Free, open source!
- Highly extensible
 - Large number of existing functions/packages
 - Ability to write personal functions
- Capability for mapping data, an asset not generally available in other statistical software
 - Several add-on packages specifically designed for the analysis of time and space data
- Increasingly used in many disciplines

Why learn R?

- “Programming tools: Adventures with R”
- Tippman, Nature Dec. 2014

A RISING TIDE OF R

An increasing proportion of research articles explicitly reference R or an R package.



Why learn R?

- R is the 21st highest paid tech skill (Dice Tech Salary Survey, 2016)
- R second most used data science language after Python (Kaggle, 2016)
- R is #5 on list of most popular analytics jobs (Indeed.com, Feb. 2017)
- R is ranked second in usage in data science articles (Google scholar, Feb. 2017)
- R growing faster than any other data science language used in research (Google scholar, Feb. 2017)
- R is #5 on IEEE spectrum ranking (2019: search ranking, trends, social media, job postings)

Why learn R?



From www.datacamp.com

- R focuses on better, user friendly data analysis, statistics and graphical models
- Used by statisticians who need to do some programming
- Python emphasizes productivity and code readability
- Used by programmers who need to do some statistics

Why learn R?

But:

- No (or limited) graphical interface
- Data manipulation can be complex
- Limited documentation
- Steep learning curve

RStudio

Free integrated development environment (IDE) for R

- Better help functions
- Integrated script editor
- More useful package manager
- Access to current datasets
- Plot history

