Introduction to Applied Statistical Computing with R

COURSE SCHEDULE

Richard E.W. Berl

Spring 2019

Last Updated: 2019-04-24

Contents

Prior to Week 1

Homework

1. **Install R**

2. **Install RStudio**

3. **Remember to bring your laptop to class!**

Resources

- Torfs, P., & Brauer, C. A (very) short introduction to R. Available: https://cran.r-project.org/doc/contrib/Torfs+Brauer-Short-R-Intro.pdf

- Short, T. R reference card. Available: https://cran.r-project.org/doc/contrib/Short-refcard.pdf

- RStudio. RStudio IDE cheat sheet. Available: https://www.rstudio.com/wp-content/uploads/2016/01/rstudio-IDE-cheatsheet.pdf

  Note: "Resources" are readings and materials available for your benefit if you would like additional detail on a topic. They are not required, but often are very helpful (especially things like cheat sheets and reference documents).

Week 1: Introduction to R programming and data structures

Lectures

- Tuesday, March 26: **Lecture 01** (pdf / Rmd)
- Thursday, March 28: **Lecture 02** (pdf / Rmd)

Homework

1. **Required Reading**

   a. FitzJohn, R. Nice R code: Designing projects. Available: https://nicercode.github.io/blog/2013-04-05-projects/

   b. Navarro, D. Section 2.1: Introduction to psychological measurement, through Section 2.2: Scales of measurement. In *Learning statistics with R*. Available: https://learningstatisticswithr.com/book/studydesign.html

   c. Wickham, H. Welcome, through Section 3: Functions. In *The tidyverse style guide*. Available: https://style.tidyverse.org/

      - Skim and refer back later when writing code

2. **Assignment 1** (pdf / Rmd) (Due April 2)

Resources

Computing

- Wilson, G., et al. (2017). Good enough practices in scientific computing. PLOS Computational Biology, 13(6), e1005510. doi: 10.1371/journal.pcbi.1005510
- Navarro, D. Chapter 8: Basic programming. In *Learning statistics with R*. Available: https://learningstatisticswithr.com/book/scripting.html

Statistics & Measurement

- Navarro, D. Chapter 1: Why do we learn statistics? In *Learning statistics with R*. Available: https://learningstatisticswithr.com/book/why-do-we-learn-statistics.html
- McDonald, J. H. Types of biological variables. In *Handbook of biological statistics*. Available: http://www.biostathandbook.com/variabletypes.html

Markdown & R Markdown

- Pritchard, A. Markdown cheatsheet. Available: https://github.com/adam-p/markdown-here/wiki/Markdown-Cheatsheet
- RStudio. R Markdown reference guide. Available: https://www.rstudio.com/wp-content/uploads/2015/03/rmarkdown-reference.pdf
- Xie, Y., Allaire, J. J., & Grolemund, G. R Markdown: The definitive guide. Available: https://bookdown.org/yihui/rmarkdown/

Git & GitHub

- Bryan, J., et al. Happy Git and GitHub for the useR. Available: https://happygitwithr.com/
- FitzJohn, R., & Falster, D. Nice R code: Introduction to version control using Git. Available: https://nicercode.github.io/git/

- GitHub Guides. Git handbook. Available: https://guides.github.com/introduction/git-handbook/

- GitHub. GitHub Desktop. Available: https://desktop.github.com/

- Zabor, E. C. Creating websites in R. Available: https://www.emilyzabor.com/tutorials/rmarkdown_websites_tutorial.html

Week 2: Importing, working with, and exploring data

Lectures

- Tuesday, April 2: **Lecture 03** (pdf / Rmd)
  - Gombe data

  - Prussian horse-kick data

  - Indo-European folktale data
- Thursday, April 4: **Lecture 04** (pdf / Rmd)
  - Fort Collins shelter service data

Homework

1. **Required Reading**

   a. Wickham, H. (2014). Tidy data. Journal of Statistical Software, 59(10), 1-23. doi: 10.18637/jss.v059.i10

      - "Informal and code heavy" version in the `tidyr` package vignette, available: https://cran.r-project.org/web/packages/tidyr/vignettes/tidy-data.html

   b. Tukey, J. W. (1977). Preface. In *Exploratory data analysis* (pp. v-ix). Reading, MA: Addison-Wesley. Available: here (pdf)

2. **Assignment 2** (pdf / Rmd) (Due April 9)

Resources

Data Wrangling

- RStudio. Data wrangling with dplyr and tidyr cheat sheet. Available: https://www.rstudio.com/wp-content/uploads/2015/02/data-wrangling-cheatsheet.pdf

- Wickham, H., & Grolemund, G. Chapter 12: Tidy data. In *R for data science*. Available: https://r4ds.had.co.nz/tidy-data.html

- Navarro, D. Chapter 7: Pragmatic matters. In *Learning statistics with R*. Available: https://learningstatisticswithr.com/book/datahandling.html

- Chang, W. Converting data between wide and long format. In *Cookbook for R*. Available: http://www.cookbook-r.com/Manipulating_data/Converting_data_between_wide_and_long_format/

Exploratory Data Analysis

- Peng, R. D. Exploratory data analysis with R. Available: https://bookdown.org/rdpeng/exdata/

- Wickham, H., & Grolemund, G. Chapter 7: Exploratory data analysis. In *R for data science*. Available: https://r4ds.had.co.nz/exploratory-data-analysis.html

- Navarro, D. Chapter 5: Descriptive statistics. In *Learning statistics with R*. Available: https://learningstatisticswithr.com/book/descriptives.html

Problems with Data

- Navarro, D. Section 5.8: Handling missing values. In *Learning statistics with R*. Available: https://learningstatisticswithr.com/book/descriptives.html#missing

- van Buuren, S. Section 1.1: The problem of missing data, through Section 1.4: Multiple imputation in a nutshell. In *Flexible imputation of missing data*. Available: https://stefvanbuuren.name/fimd/sec-problem.html

- Prabhakaran, S. Outlier treatment. In *r-statistics.co*. Available: http://r-statistics.co/Outlier-Treatment-With-R.html

- Navarro, D. Section 2.7: Confounds, artifacts and other threats to validity. In *Learning statistics with R*. Available: https://learningstatisticswithr.com/book/studydesign.html#confounds-artifacts-and-other-threats-to-validity

- McDonald, J. H. Confounding variables. In *Handbook of biological statistics*. Available: http://www.biostathandbook.com/confounding.html

Week 3: Information design and data visualization

Lectures

- Tuesday, April 9: **Lecture 05** (pdf / Rmd)
  – Egg data (Excel format)

- Thursday, April 11: **Lecture 06** (pdf / Rmd)
  – Egg data (CSV format)

Homework

1. **Required Reading**

   a. Healy, K. Chapter 1: Look at data. In *Data visualization: A practical introduction*. Available: https://socviz.co/lookatdata.html

   b. Wilke, C. O. Chapter 29: Telling a story and making a point. In *Fundamentals of data visualization*. Available: https://serialmentor.com/dataviz/telling-a-story.html

2. **Assignment 3** (pdf / Rmd) (Due April 16)

3. Start thinking about **Project Proposal** (Due April 23, 11:59pm)

Resources

- Wickham, H., et al. Create elegant data visualizations using the grammar of graphics: ggplot2. Available: https://ggplot2.tidyverse.org/

- Wickham, H., & Grolemund, G. Chapter 3: Data visualization. In *R for data science*. Available: https://r4ds.had.co.nz/data-visualisation.html

- Wickham, H., & Grolemund, G. Chapter 28: Graphics for communication. In *R for data science*. Available: https://r4ds.had.co.nz/graphics-for-communication.html

- RStudio. Data visualization with ggplot2 cheat sheet. Available: https://www.rstudio.com/wp-content/uploads/2015/12/ggplot2-cheatsheet.pdf

- Healy, K. Data visualization: A practical introduction. Available: https://socviz.co/

- Wilke, C. O. Fundamentals of data visualization. Available: https://serialmentor.com/dataviz/

- Rougier, N. P., et al. (2014). Ten simple rules for better figures. PLOS Computational Biology, 10(9), e1003833. doi: 10.1371/journal.pcbi.1003833

- BBC Open Source. BBC visual and data journalism cookbook for R graphics. Available: https://bbc.github.io/rcookbook/

- The Urban Institute. Urban Institute R graphics guide. Available: https://urbaninstitute.github.io/urban_R_theme/

- Geckoboard. Play your charts right: Tips for effective data visualization. Available: https://www.geckoboard.com/learn/data-literacy/data-visualization-tips/

- Rost, L. C. (2018). Your friendly guide to colors in data visualisation. Chartable. Available: https://blog.datawrapper.de/colorguide/

- Tol, P. (2018). Colour schemes. SRON Netherlands Institute for Space Research. Available: https://personal.sron.nl/~pault/data/colourschemes.pdf

Week 4: Hypothesis testing and basic linear models

Lectures

- Tuesday, April 16: **Lecture 07** (pdf / Rmd)
    – PanTHERIA data ("WR05")

- Thursday, April 18: **Lecture 08** (pdf / Rmd)
    – Sage-Grouse data ("S5 Table")

Homework

1. **Required Reading**

    a. McDonald, J. H. Basic concepts of hypothesis testing. In *Handbook of biological statistics*. Available: http://www.biostathandbook.com/hypothesistesting.html

    b. McDonald, J. H. Correlation and linear regression. In *Handbook of biological statistics*. Available: http://www.biostathandbook.com/linearregression.html

        - For information on conducting these tests in R, see: Mangiafico, S. S. Correlation and linear regression. In *An R companion for the handbook of biological statistics*. Available: http://rcompanion.org/rcompanion/e_01.html

    c. Joselson, N. (2016). Eugenics and statistics, discussing Karl Pearson and R. A. Fisher. Available: https://njoselson.github.io/Fisher-Pearson/

2. **Assignment 4** (pdf / Rmd) (Due April 23)

3. **Project Proposal** (Due April 23, 11:59pm)

Resources

Hypothesis Testing

- McDonald, J. H. Choosing a statistical test. In *Handbook of biological statistics*. Available: http://www.biostathandbook.com/testchoice.html

    - For information on conducting these tests in R, see: Mangiafico, S. S. An R companion for the handbook of biological statistics. Available: http://rcompanion.org/rcompanion/

- Navarro, D. Chapter 11: Hypothesis testing. In *Learning statistics with R*. Available: https://learningstatisticswithr.com/book/hypothesistesting.html

Linear Models

- Wickham, H., & Grolemund, G. Chapter 23: Model basics. In *R for data science*. Available: https://r4ds.had.co.nz/model-basics.html

- Wickham, H., & Grolemund, G. Chapter 24: Model building. In *R for data science*. Available: https://r4ds.had.co.nz/model-building.html

- McDonald, J. H. Simple logistic regression. In *Handbook of biological statistics*. Available: http://www.biostathandbook.com/simplelogistic.html

- McDonald, J. H. Multiple logistic regression. In *Handbook of biological statistics*. Available: http://www.biostathandbook.com/multiplelogistic.html

- Navarro, D. Chapter 15: Linear regression. In *Learning statistics with R*. Available: https://learningstatisticswithr.com/book/regression.html

- McDonald, J. H. One-way anova. In *Handbook of biological statistics*. Available: http://www.biostathandbook.com/onewayanova.html

- McDonald, J. H. Two-way anova. In *Handbook of biological statistics*. Available: http://www.biostathandbook.com/twowayanova.html

- Navarro, D. Chapter 14: Comparing several means (one-way ANOVA). In *Learning statistics with R*. Available: https://learningstatisticswithr.com/book/anova.html

- Navarro, D. Chapter 16: Factorial ANOVA. In *Learning statistics with R*. Available: https://learningstatisticswithr.com/book/anova2.html

- McDonald, J. H. Multiple comparisons. In *Handbook of biological statistics*. Available: http://www.biostathandbook.com/multiplecomparisons.html

- Faraway, J. J. Practical regression and anova using R. Available: https://cran.r-project.org/doc/contrib/Faraway-PRA.pdf

- Scholer, F. ANOVA - Type I/II/III SS explained. Available: https://mcfromnz.wordpress.com/2011/03/02/anova-type-iiiiii-ss-explained/

Meta-Science

- Aschwanden, C. (2015). Science isn't broken. FiveThirtyEight. Available: https://fivethirtyeight.com/features/science-isnt-broken/

- Angwin, J., et al. (2016). Machine bias. ProPublica. Available: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

- Lum, K., & Isaac, W. (2016). To predict and serve?. Significance, 13(5), 14-19. doi: 10.1111/j.1740-9713.2016.00960.x

- Rodriguez-Lonebear, D. (2016). Chapter 14: Building a data revolution in Indian country. In Kukutai, T., & Taylor, J. (Eds.), *Indigenous data sovereignty* (pp. 253-272). Available: http://press-files.anu.edu.au/downloads/press/n2140/pdf/ch14.pdf

Week 5: Advanced statistical methods, part I: Ecological analyses, ordinal data, and dimensionality reduction

Lectures

- Tuesday, April 23: **Lecture 09** (pdf / Rmd)
  - Christmas Bird Count data (Fort Collins, 1900-2017)

  - Bob Marshall Wilderness Complex visitor data

- Thursday, April 25: **Lecture 10** (pdf / Rmd)
  - "Best" maximally complete Christmas Bird Coount data

Homework

1. **Required Reading**

   None

2. **Assignment 5** (pdf / Rmd) **(ATTENTION: Due May 7)**

Resources

Ecological Analyses

- Simpson, G. CRAN Task View: Analysis of Ecological and Environmental Data. Available: https://cran.r-project.org/web/views/Environmetrics.html

- Oksanen, J. Vegan: An introduction to ordination. Available: https://cran.r-project.org/web/packages/vegan/vignettes/intro-vegan.pdf

- Guevara, M. R., et al. (2016). `diverse`: An R Package to analyze diversity in complex systems. The R Journal, 8(2), 60-78. doi: 10.32614/rj-2016-033

Ordinal Data

- Mangiafico, S. S. Introduction to Likert data. In *Summary and analysis of extension program evaluation in R*. Available: http://rcompanion.org/handbook/E_01.html

- Barry, D. Do not use averages with Likert scale data. Available: https://bookdown.org/Rmadillo/likert/

- Mangiafico, S. S. One-way permutation test of independence for ordinal data. In *Summary and analysis of extension program evaluation in R*. Available: http://rcompanion.org/handbook/K_02.html

- Holgado–Tello, F. P., et al. (2010). Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. Quality & Quantity, 44(1), 153. doi: 10.1007/s11135-008-9190-y

- Rhemtulla, M., et al. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. Psychological Methods, 17(3), 354. doi: 10.1037/a0029315

Dimensionality Reduction

Multidimensional Scaling (Classical and Nonmetric)

- Boehmke, B. Hierarchical cluster analysis. In *University of Cincinnati Business Analytics R programming guide.* Available: https://uc-r.github.io/hc_clustering

- Boehmke, B. Principal components analysis. In *University of Cincinnati Business Analytics R programming guide.* Available: https://uc-r.github.io/pca

- Boehmke, B. K-means cluster analysis. In *University of Cincinnati Business Analytics R programming guide.* Available: https://uc-r.github.io/kmeans_clustering

Exploratory Factor Analysis

- Revelle, W. (2018). How to: Use the `psych` package for factor analysis and data reduction. Available: http://personality-project.org/r/psych/HowTo/factor.pdf

- Savalei, V. (2011). What to do about zero frequency cells when estimating polychoric correlations. Structural Equation Modeling, 18(2), 253-273. doi: 10.1080/10705511.2011.557339

Week 6: Advanced statistical methods, part II: Structural equation modeling, social network analysis, and geospatial analyses

Lectures

- Tuesday, April 30: **Lecture 11** (pdf / Rmd)

- Thursday, May 2: **Lecture 12** (pdf / Rmd)

Homework

1. **Required Reading**

   None

2. **Assignment 5** (pdf / Rmd) **(ATTENTION: Due May 7)**

3. Start working on **Project Report** (Due May 17, 11:59pm)

4. Start working on **Project Presentation** (Due week of May 13)

Resources

Structural Equation Modeling

Confirmatory Factor Analysis

Social Network Analysis

Geospatial Mapping

Mantel Tests?

- Grossenbacher, T. Beautiful thematic maps with ggplot2 (only). Avalable: https://timogrossenbacher.ch/2016/12/beautiful-thematic-maps-with-ggplot2-only/

tidygraph?

ggraph?

Week 7: Additional topics and workshops for course project

Lectures

- Tuesday, May 7: **Lecture 13** (pdf / Rmd)
- Thursday, May 9: **Lecture 14** (pdf / Rmd)

Homework

1. **Required Reading**

   None

2. **Project Report** (Due May 17, 11:59pm)

3. **Project Presentation** (Due week of May 13)

Resources

Data Science

- Gatto, L. An introduction to machine learning with R. Available: https://lgatto.github.io/IntroMachineLearningWithR/

Random Forests

Text Mining, Scraping, and Sentiment Analysis

Reproducible Research

- rOpenSci. Reproducibility in science. Available: https://ropensci.github.io/reproducibility-guide/

- Hartgerink, C. (2017). Composing reproducible manuscripts using R Markdown. Available: https://elifesciences.org/labs/cad57bcf/composing-reproducible-manuscripts-using-r-markdown

- Lowndes, J. S. S., et al. (2017). Our path to better science in less time using open data science tools. Nature Ecology & Evolution, 1(6), 0160. doi: 10.1038/s41559-017-0160

Science Communication

- Hillier, A., Kelly, R. P., & Klinger, T. (2016). Narrative style influences citation frequency in climate change science. PLOS ONE, 11(12), e0167983. doi: 10.1371/journal.pone.0167983

- Ratliff, W. The David Attenborough style of scientific presentation. Available: https://www.dropbox.com/s/j1vv2baheiduvip/David%20Attenborough%20talk%20technique%202018.pdf

Week 8: Project presentations

Meeting date and time TBD

(pdf / Rmd)