Nicholas J. Horton
Randall Pruim
Daniel T. Kaplan

A Student's Guide to R

Project MOSAIC

Copyright (c) 2018 by Nicholas J. Horton, Randall Pruim, & Daniel Kaplan.

Edition 1.3, June 2018

This material is copyrighted by the authors under a Creative Commons Attribution 3.0 Unported License. You are free to *Share* (to copy, distribute and transmit the work) and to *Remix* (to adapt the work) if you attribute our work. More detailed information about the licensing is available at this web page: `http://www.mosaic-web.org/go/teachingRlicense.html`.

Cover Photo: Maya Hanna.

Contents

1 *Introduction* 13

2 *Getting Started with RStudio* 15

3 *One Quantitative Variable* 27

4 *One Categorical Variable* 39

5 *Two Quantitative Variables* 45

6 *Two Categorical Variables* 55

7 *Quantitative Response, Categorical Predictor* 61

8 *Categorical Response, Quantitative Predictor* 69

9 *Survival Time Outcomes* 73

10 *More than Two Variables* 75

11 *Probability Distributions & Random Variables* 83

12 *Power Calculations* 89

13 *Data Wrangling* 93

14 *Health Evaluation (HELP) Study* 107

15 *Exercises and Problems* 111

16 *Bibliography* 115

17 *Index* 117

About These Notes

We present an approach to teaching introductory and intermediate statistics courses that is tightly coupled with computing generally and with R and RStudio in particular. These activities and examples are intended to highlight a modern approach to statistical education that focuses on modeling, resampling based inference, and multivariate graphical techniques. A secondary goal is to facilitate computing with data through use of small simulation studies and appropriate statistical analysis workflow. This follows the philosophy outlined by Nolan and Temple Lang[1]. The importance of modern computation in statistics education is a principal component of the recently adopted American Statistical Association's curriculum guidelines[2].

Throughout this book (and its companion volumes), we introduce multiple activities, some appropriate for an introductory course, others suitable for higher levels, that demonstrate key concepts in statistics and modeling while also supporting the core material of more traditional courses.

[1] D. Nolan and D. Temple Lang. Computing in the statistics curriculum. *The American Statistician*, 64(2):97–107, 2010

[2] ASA Undergraduate Guidelines Workgroup. 2014 curriculum guidelines for undergraduate programs in statistical science. Technical report, American Statistical Association, November 2014. `http://www.amstat.org/education/curriculumguidelines.cfm`

A Work in Progress

These materials were developed for a workshop entitled *Teaching Statistics Using R* prior to the 2011 United States Conference on Teaching Statistics and revised for USCOTS 2011, USCOTS 2013, eCOTS 2014, ICOTS 9, and USCOTS 2015. We organized these workshops to help instructors integrate R (as well as some related technologies) into statistics courses at all levels. We received great feedback and many wonderful ideas from the participants and those that we've shared this with since the workshops.

CAUTION!
Despite our best efforts, you WILL find bugs both in this document and in our code. Please let us know when you encounter them so we can call in the exterminators.

Click here to download full PDF material