

Nicholas J. Horton
Randall Pruim
Daniel T. Kaplan

A Student's
Guide to



Project MOSAIC

Copyright (c) 2018 by Nicholas J. Horton, Randall Pruum, & 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

<i>1</i>	<i>Introduction</i>	<i>13</i>
<i>2</i>	<i>Getting Started with RStudio</i>	<i>15</i>
<i>3</i>	<i>One Quantitative Variable</i>	<i>27</i>
<i>4</i>	<i>One Categorical Variable</i>	<i>39</i>
<i>5</i>	<i>Two Quantitative Variables</i>	<i>45</i>
<i>6</i>	<i>Two Categorical Variables</i>	<i>55</i>
<i>7</i>	<i>Quantitative Response, Categorical Predictor</i>	<i>61</i>
<i>8</i>	<i>Categorical Response, Quantitative Predictor</i>	<i>69</i>
<i>9</i>	<i>Survival Time Outcomes</i>	<i>73</i>

10	<i>More than Two Variables</i>	75
11	<i>Probability Distributions & Random Variables</i>	83
12	<i>Power Calculations</i>	89
13	<i>Data Wrangling</i>	93
14	<i>Health Evaluation (HELP) Study</i>	107
15	<i>Exercises and Problems</i>	111
16	<i>Bibliography</i>	115
17	<i>Index</i>	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 Lang¹. The importance of modern computation in statistics education is a principal component of the recently adopted American Statistical Association's curriculum guidelines².

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.

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.

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>

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.

Consider these notes to be a work in progress. We appreciate any feedback you are willing to share as we continue to work on these materials and the accompanying mosaic package. Drop us an email at pis@mosaic-web.org with any comments, suggestions, corrections, etc.

Updated versions will be posted at <http://mosaic-web.org>.

Two Audiences

We initially developed these materials for instructors of statistics at the college or university level. Another audience is the students these instructors teach. Some of the sections, examples, and exercises are written with one or the other of these audiences more clearly at the forefront. This means that

1. Some of the materials can be used essentially as is with students.
2. Some of the materials aim to equip instructors to develop their own expertise in R and RStudio to develop their own teaching materials.

Although the distinction can get blurry, and what works “as is” in one setting may not work “as is” in another, we’ll try to indicate which parts fit into each category as we go along.

R, RStudio and R Packages

R can be obtained from <http://cran.r-project.org/>. Download and installation are quite straightforward for Mac, PC, or linux machines.

RStudio is an integrated development environment (IDE) that facilitates use of R for both novice and expert users. We have adopted it as our standard teaching environment because it dramatically simplifies the use of R for instructors and for students. RStudio can be installed as a desktop (laptop) application or as a server application that is accessible to users via the Internet.

In addition to R and RStudio, we will make use of several packages that need to be installed and loaded separately. The mosaic package (and its dependencies) will

MORE INFO

Several things we use that can be done only in RStudio, for instance `manipulate()` or RStudio’s integrated support for reproducible research).

RStudio server version works well with starting students. All they need is a web browser, avoiding any potential problems with oddities of students’ individual computers.

This is a sample, click download link to get the full Tutorial

CLICK BELOW

