

Q SCI 381: Introduction to Probability and Statistics
S. Scherba, Jr. Assignment #1

Introduction

During this quarter, we will use the statistical package R and the user interfaces RStudio and R Markdown to complete the class assignments. R is a language and environment for statistical computing and graphic display of data and its analysis. Built into R are many functions for a wide range of statistical techniques (the calculation of probabilities; classical statistical tests; linear, nonlinear, and generalized linear modelling; time series analysis, classification, and cluster analysis) created either by the original developers of R (these are the core R techniques) or by vetted contributors who have greatly extended the capabilities of R. It compiles and runs on a wide variety of platforms including UNIX, Linux, Windows, and MacOS.

As a companion text to the class, we will be using:

Andrew P. Beckerman, D. Z. Childs, and O. L. Petchey. 2017(2nd edition). **Getting Started With R: An Introduction for Biologists**. Oxford.

This text will be **referred to as BC&P**. If you wish, you may purchase a bound copy of this text, but pdf copies of the chapters from this text are posted on our course Canvas website (**Files > R Material and Information > BC&P Getting Started with R 2nd Edition > Chapter PDFs**). It is written in a straightforward and conversational manner and it is meant to be read while you simultaneously create the code (often called script in R) you read about and the text discusses. That is one of the reasons I selected it for our class. I worked my way through this text in just this manner as I constructed the tutorials.

During the quarter, and in conjunction with our readings and lectures, we will read and work through all but Chapters 6 and 7 of BC&P, but you should feel free to work your way through these chapters as well. The material in those chapters will be extensively covered Q SCI 482 and 483, if you take them. When you are done, in addition to a basic understanding of statistics, you will have good working knowledge of R and the basic skills to create our own code, manipulate data, and use R functions to analyze and graph it

The keys for all of the assignments will be posted on our courses Canvas website after they are turned in. You may work together on these assignments unless specifically told that you are to do them independently. If you do work with someone else, **please make sure to prominently mention that in your assignment.**

The data sets for these tutorials and all of the other class assignments will be found in **Files > All Assignments > Class Data Sets**. You will notice that the data sets for the tutorials from BC&P are also in **Files > R Material and Information > BC&P Getting Started with R 2nd Edition > Data Sets for BC&P Text**. I did this on purpose. For ease of access, I suggest you **copy the assignment R code and the data sets to your computer desktop and have RStudio read it from there**. Once you start the tutorials you will understand how to enter data into R. Reading data from your desktop will shorten the string of code you will need to do this.

Assignment 1

This first assignment consists of a set of four (4) tutorials that allow you write the code as you read through the assigned chapters of BC&P. Please start with “A First R Tutorial.Rmd” and complete them in order; saving “Dplyr Tutorial.Rmd” for last. The third tutorial focuses on graphing data and is the longest of the four, please plan your time accordingly. Instructions and additional comments appear in the text of the tutorials, which appear as R comment statements. In BC&P, read pages vi – 8 (up to section 1.5), downloading both R and RStudio as you read pages 1 – 6. Open the “A First R Tutorial.rmd” and begin the journey. These tutorials and their companions will instruct you on further pages to read in BC&P.

Start to complete these tutorials now. If you have **questions**, please **first try to figure them out on your own** (your TAs and I have had to do this too as we learned and still do). Follow the suggestions in BC&P about using “Help” in R, and natural language Google searches. **If you still have problems, then please ask one of us and we will help you.** We are always here to back you up! If you will practice this approach in the class, you will be better preparing yourself for life as you will find it in your careers (and probably some of your future classes). As you complete each tutorial, please upload your completed work in the Assignments tab on Canvas.

Final Reminders and Comments

The other assignments in the course will consist of

[1] **Review problems** based on readings from our primary text and course lecture notes, problems that you will be asked to do “by hand”, i.e., without using R., but calculators are permitted. These are **open book, open notes, open other resources**, but **you must do them on your own (not working with anyone else in the course)**.

[2] **Data analysis problems** demonstrating the methods we will be learning using R to do the calculations. These are **open book, open notes, and open other resources including code from prior assignments**. **You may work together, but you must complete these assignments by yourself, e.g., you cannot simply copy classmate’s computer file and upload it. If you work with someone, you must prominently state who you work with.** You will not acquire skill with R if you simply upload a classmate’s code.

[3] Some of the remaining seven (7) assignments will also include a **companion tutorial** that supplements the analysis method we are studying while following the relevant pages in BC&P. **You will benefit yourself by gaining skill in using R by reading the text as you work through the tutorials and the assignments. Since that is one of the goals of the course, make sure to read the assigned pages in BC&P! Simply working through the tutorials and/or the assignments without reading BC&P is a mistake!**

By completing all of these assignments, you will gain an understanding of what the methods are doing, how they are done, why they are done, the assumptions behind them, and knowledge of the current mostly commonly used software package for actually doing data exploration and analysis.

Q SCI 381: Introduction to Probability and Statistics
S. Scherba, Jr. Assignment #1

This skill is something that you may add to your resume when you apply for internships, a position in someone's lab, or an application for a job. **Be sure to ask our TAs if they include a working knowledge of R on their resumes.** Now you are off and running on building skills for your careers.

For those of you who believe that simply running packaged computer software will solve all of your quantitative problems, consider the following **from the FISH 559 website:**

“Most fisheries and other wildlife professionals who are tasked with providing scientific advice to decision makers often find existing software tools insufficient for their needs, requiring the development of purpose-built computer programs... Numerical techniques should not be applied "cookbook" style. It is necessary to understand the mathematical basis for the technique. Furthermore, numerical techniques frequently involve expert judgment regarding choices (e.g. for tolerances, initial values). The best way to learn a numerical technique therefore involves knowing its theoretical basis and spending time practicing it....”

Knowing that theoretical basis of the analyses is absolutely critical to you in your careers. An averagely intelligent high school junior can master the syntax of R (or C, C++, JAVA, Python, etc.) and write code to do an analysis, but **it is the understanding of why you can or cannot do the analysis and what the results mean within the context of the problem**, in words that the person (persons) posing the problem understand **that will set you apart and make you a valuable part of team member.** This class is an introduction to learning those skills.