

The aim of this document is to help you, an undergraduate student in an introductory statistics course, learn to use the software R as part of your learning of statistics. If you find that it reads like the rough draft of something that could be more useful and better written, that is because it is a rough draft that could be more useful and better written. This document will evolve on a weekly basis as the semester progresses. I will add new material as we cover it in class and edit old material based on feedback from you to make it clearer for you and future students. I suggest that you do not print new versions but merely replace your electronic copy from time to time. Good luck as you begin your quest to master introductory statistical concepts and their application!

What is R?

R is powerful software for interacting with data. With R you can create sophisticated graphs, you can carryout statistical analyses, and you can create and run simulations. R is also a programming language with an extensive set of built-in functions, so you can, with some experience, extend the language and write your own code to build your own statistical tools. Advanced users can even incorporate functions written in other languages, such as C, C++, and Fortran.

The S language has been around for more than twenty years and has been the most widely-used statistical software in departments of statistics for most of that time, first as S and then as the commercially available S-PLUS. R is an open source implementation of the S language that is now a viable alternative to S-PLUS, and in fact, has many advantages. A core team of statisticians and many other contributors work to update and improve R and to make versions that run well under all of the most popular operating systems. Most importantly to you, R is free, high-quality statistical software that will be useful as you learn statistics even though it is also a first-rate tool for professional statisticians.

Why use R for introductory statistics?

There are several reasons that make R an excellent choice of statistical software for an introductory statistical course. First, R is free and available on the Web. You can use it on your home computers and are not tied to campus labs. Second, R is powerful, widely-used software. The knowledge of R you gain during the course potentially translates to a marketable skill. You will learn to use a tool that has many practical uses outside the classroom. Third, even though it is not the simplest statistical software, the basics are easy enough to master that learning to use R need not interfere overly much with learning the statistical concepts encountered in an introductory course. Fourth, did I mention that it is free and you can use it at home?

The primary drawback to using R in an introductory course is that most existing documentation for R is written for an audience that is knowledgeable about statistics and has experience with other statistical computing programs. In contrast, this document intends to make R accessible to the typical student in an introductory statistics course who is new to both statistical concepts and statistical computing. The aim is to teach you how to install R on your home computer and to teach you to use R to learn the statistical concepts usually included in an introductory course with explanations and examples aimed at the appropriate level. This document purposely does not attempt to teach you about R's advanced features. The intention is to teach you enough R to enhance your learning of introductory statistics and to point you in the direction of more information should you find a desire to learn more.