

## CHAPTER OVERVIEW

### 6: Reading Data into the R Environment

As we have seen, the R environment provides some powerful functions to quickly and relatively easily develop and test regression models. Ironically, simply reading the data into R in a useful format can be one of the most difficult aspects of developing a model. R does not lack good input-output capabilities, but data often comes to the model developer in a messy form. For instance, the data format may be inconsistent, with missing fields and incorrectly recorded values. Getting the data into the format necessary for analysis and modeling is often called *data cleaning*. The specific steps necessary to “clean” data are heavily dependent on the data set and are thus beyond the scope of this tutorial. Suffice it to say that you should carefully examine your data before you use it to develop any sort of regression model. Section 2.2 provides a few thoughts on data cleaning.

In Chapter 2, we provided the functions used to read the example data into the R environment, but with no explanation about how they worked. In this chapter, we will look at these functions in detail, as specific examples of how to read a data set into R. Of course, the details of the functions you may need to write to input your data will necessarily change to match the specifics of your data set.

#### 6.1: Reading CSV files

---

This page titled [6: Reading Data into the R Environment](#) is shared under a [CC BY-NC 4.0](#) license and was authored, remixed, and/or curated by [David Lilja](#) ([University of Minnesota Libraries Publishing](#)) via [source content](#) that was edited to the style and standards of the LibreTexts platform.