Loading Data from the Course Github Repository

Dr. Christopher Brown

December 18, 2018

Introduction

In this document we provide details on downloading data sets from the course repository, archived on Github at https://github.com/cbrown-clu/Math-331-Data-Analysis.

You can also check out the data sets that are included with R, by entering the command **data()** in R.

Obtaining a data set the straightforward way

The first step is to create a new project and project folder in RStudio. From RStudio you can do this from the File menu. Make sure you can locate this project folder later; it may be easiest to put the project folder on your desktop.

Next, visit the course repository data sets at https://github.com/cbrown-clu/Math-331-Data-Analysis/blob/master/B.%20Data%20sets/BookstorePriceData.csv. Find the data set you want.

Now, *right-click* on the link for the data set, and select *Save link as....* This will open a dialog menu for you to save the data file. Make sure to save the file in your project folder. At this point you can read it with **read.csv()**.

Using RCurl

You can also install the package RCurl on your machine and then use that to download the data set. This can be more involved, but also gives enough flexibility to perform webscraping pretty easily. For more on that, visit http://www.programmingr.com/content/webscraping-using-readlines-and-rcurl/.