Library Book Use

Data Computing

A First Small Project

In this small project, you will be exploring some of the records on books from the Macalester College library. You will be creating an `.Rmd` file that summarises certain aspects of the data. That file will contain three things:

1. A narrative explaining (briefly) what you are looking at and your conclusions.
2. The computer commands that allow you to look inside the data.
3. The output of your computer commands.

You only need to write (1) and (2). The outputs (3) will be generated automatically when you *compile*[1] the `.Rmd` to `.html`.

The `.Rmd` file is called the *source* document. In general, you write, revise, and update source documents. Compilation of the source document to html (or other formats) is done automatically. The point of compilation is to bring together your narrative and computer commands with the computer output into one easy-to-read document.

[1] "Translate" is a reasonable English equivalent to the technical jargon "compile."

Getting Started

In RStudio, open the "project" holding the assignments for this course.[2] There will be a file named `Week-1-project.Rmd`. Open that file in the RStudio editor. You'll be entering your R commands in that file.

You will need to get access to the library-book data. The full collection of books is too large to download in a short time, so you will be using a small sample of books.

[2] **Setting up an RStudio Project.** These instructions assume that you have already created on GitHub your own copy ("clone") of the assignment project for this course. If you haven't done that yet, this is a good time to do so.

- Step 1. Download the library-book data to your computer. To do this, give the following command in your RStudio console:

```
download.file("http://tiny.cc/dcf/Library-small.rda",
    dest = "Library-small.rda")
```

You only need to do this once. You can confirm that the data file has been downloaded by going to the "Files" tab in RStudio.

- Step 2. If you haven't already done so, open the `Week-1-project.Rmd` file in the RStudio editor.
- Step 3. Before you add anything to the `.Rmd` file, press the "Knit" button in the editor to compile the document from `.Rmd` to `.html`.

Depending on how your RStudio system is set up, the compiled document will be displayed in either a new window or in the "Viewer" tab. On some computers, particularly Windows, the new window may be *behind* RStudio, so if you don't see the new window immediately, look back there.

Compile early and often

Why compile the document before you have added anything to it? So that you can confirm that you are starting with a working document.[3] Then, after you add a little bit, compile the document again. If the compilation works, then continue on. If it doesn't, you have a huge hint about where the problem is: somewhere in the stuff you just added to the document.

Compilation doesn't cost anything. And don't think of it as the *final step*. Compilation is part of the work process and you will be doing it many, many times as you construct your document.

[3] The idea of a document that "works" may be unfamiliar. After all, in word processors or email, you type whatever you want, even if it makes no sense. But the `.Rmd` document has to make sense to the computer because the computer is going to be carrying out the instructions in the document and translating it to `.html`.

In your document …

1. Add a sentence or two describing where you downloaded the data from, including the URL that you used to access the data file.
2. Put in a section header: **Basics**
3. Under that header, put in a *chunk* that will read the data into R from the file you had already downloaded from the console, namely `Library-small.Rda`. The contents of the chunk will be:

   ```
   load("Library-small.Rda")
   ```

   This will create two data table objects, `Inv` and `Bks`. The library's collection is in `Inv`. The `Bks` data table is about individual books that might or might not be in the library's collection.
4. Create new chunks to calculate the number of cases in each file, the names of the variables, and whatever else might occur to you. In addition, write a narrative with a short description of the contents of each file.
5. Create a chunk to look at the number of books with each different `Current.Status`. Your command will look like this:

```
Inv %>% group_by(Current.Status) %>% tally()
```

6. Create a chunk to look at how many times books have been checked out. Use this command:

```
Inv %>% group_by(Issued.Count) %>% tally()
```

Try to figure out what the results mean, and write your interpretation in a narrative following the chunk.

Congratulations! You've completed your data project.