

# Assignment – Loading Data into a Data Frame

## FiveThirtyEight

Politics Sports Science & Health Economics Culture

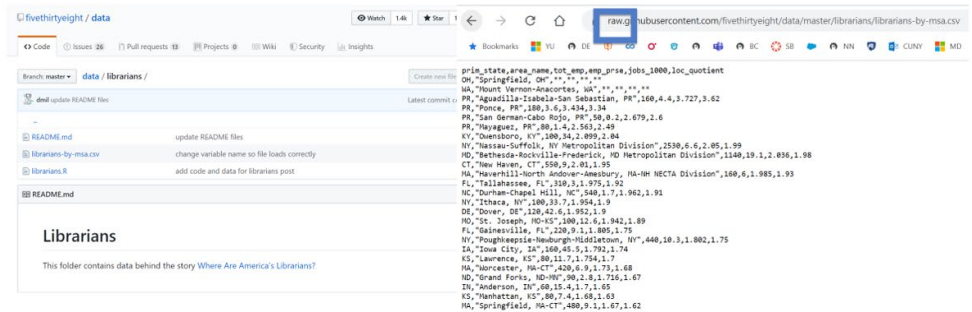
APR. 11, 2014, AT 11:50 AM

### Where Are America's Librarians?

By Mona Chalabi and Andrew Flowers

Filed under [Jobs](#)

Get the data on [GitHub](#)



Above: [fivethirtyeight.com](#) article on librarians with github repo and raw csv file for supporting data

We are often tasked with taking data in one form and transforming it for easier downstream analysis. We will spend several weeks in this course on tidying and transformation operations. Some of this work could be done in SQL or R (or Python or...). Here, you are asked to use R—you may use any base functions or packages as you like.

**Your task** is to first choose one of the provided datasets on [fivethirtyeight.com](#) that you find interesting:

<https://data.fivethirtyeight.com/>

You should first study the data and any other information on the GitHub site, and read the associated [fivethirtyeight.com](#) article.

To receive full credit, you should:

1. Take the data, and create one or more code blocks. You should finish with a data frame that contains a subset of the columns in your selected dataset. If there is an obvious target (aka predictor or independent) variable, you should include this in your set of columns. You should include (or add if necessary) meaningful column names and replace (if necessary) any non-intuitive abbreviations used in the data that you selected. For example, if you had instead been tasked with working with the [UCI mushroom dataset](#), you would include the target column for edible or poisonous, and transform “e” values to “edible.” Your deliverable is the R code to perform these transformation tasks.
2. Make sure that the original data file is accessible through your code—for example, stored in a GitHub repository or AWS S3 bucket and referenced in your code. If the code references data on your local machine, then your work is not reproducible!
3. Start your R Markdown document with a two to three sentence “Overview” or “Introduction” description of what the article that you chose is about, and include a link to the article.
4. Finish with a “Conclusions” or “Findings and Recommendations” text block that includes what you might do to extend, verify, or update the work from the selected article.
5. Each of your text blocks should minimally include at least one header, and additional non-header text.
6. You’re of course welcome—but not required—to include additional information, such as exploratory data analysis graphics (which we will cover later in the course).
7. Place your solution into a single R Markdown (.Rmd) file and publish your solution out to [rpubs.com](#).
8. Post the .Rmd file in your GitHub repository, and provide the appropriate URLs to your GitHub repository and your [rpubs.com](#) file in your assignment link.

Note that Steps #2 through #8 above are baseline requirements for our grading rubric for all subsequent coding assignments in this course.

We’ll look together at some of the most interesting student solutions in next week’s meetup!