Assignment – Subsetting Datasets

Please place your solution in a single R Markdown (.Rmd) file, post your code to GitHub, and publish your markdown to rpubs.com, and then include the URLs of your GitHub repository and your rpubs.com code in your week 3 assignment link. You should also have the original data file accessible through your code—for example, stored in a GitHub repository and referenced in your code.

Very often, we're tasked with taking data in one form and transforming it for easier downstream analysis. Over the next few weeks, we'll work with several packages that help with the tasks of tidying and transforming data.

UCI Machine Learning Repository. University of California Irvine maintains an archive of datasets.[1] **Your task** is to choose one dataset, then study the data and its associated description of the data (i.e. "data dictionary"). You should take the data, and create an R data frame with a subset of the columns (and if you like rows) in the dataset. Your deliverable is the R code to perform these transformation tasks.

If you need help in choosing a dataset, you might choose "Pittsburgh Bridges" here: https://archive.ics.uci.edu/ml/datasets/Pittsburgh+Bridges. [Students that participated in the R Programming Bridge workshop would have done a similar exercise with the mushrooms dataset, so I would prefer that you use any dataset *except* for mushrooms].

---

[1] There are many, many other rich data repositories on the web. Here is one curated list: http://www.inside-r.org/howto/finding-data-internet