


# Hands on Lab – Subsetting Datasets

Very often, we're tasked with taking data in one form and transforming it for easier downstream analysis. We'll start with a dataset that does not require a lot of up front clean-up work.




**Machine Learning Repository**  
Center for Machine Learning and Intelligent Systems

## Pittsburgh Bridges Data Set

Download: [Data Folder](#), [Data Set Description](#)

**Abstract:** Bridges database that has original and numeric-discretized datasets



**UCI Machine Learning Repository.** University of California Irvine maintains an archive of datasets.<sup>1</sup>

Your task is to work with the Pittsburgh bridges dataset:

<https://archive.ics.uci.edu/ml/datasets/Pittsburgh+Bridges>

You should first study the data and its associated description (i.e. “data dictionary”). You should take the data, and create an R data frame with a subset of the columns in the dataset. You should provide relevant column names. Your deliverable is the R Markdown code to perform these transformation tasks.

---

<sup>1</sup> 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>