

IS 607 – Project 2

Your Project 2 is due end of day on Sunday October 11th. You may work in a small group on this assignment. The goal of this assignment is to give you practice in preparing different datasets for downstream analysis work.

We'll look at some of the most interesting student solutions during our meetup on Thursday October 15th.

Your task is to:

(1) Choose any three of the “wide” datasets identified in the Week 6/7 Discussion item. (You may use your own dataset; please don't use my Sample Post dataset, since that was used in your Week 6 assignment!) **For each of the three chosen datasets:**

- Create a .CSV file (or optionally, a MySQL database!) that includes all of the information included in the dataset. You're encouraged to use a “wide” structure similar to how the information appears in the discussion item, so that you can practice tidying and transformations as described below.
- Read the information from your .CSV file into R, and use *tidyr* and *dplyr* as needed to tidy and transform your data. [Most of your grade will be based on this step!]
- Perform the analysis requested in the discussion item.
- Your code should be in an R Markdown file, posted to *rpubs.com*, and should include narrative descriptions of your data cleanup work, analysis, and conclusions.

(2) Please include in your homework submission, for each of the three chosen datasets:

- The URL to the .Rmd file in your GitHub repository. and
- The URL for your *rpubs.com* web page.