Live Session Assignment

Units 3 and 4

Creating projects in RStudio

The live session assignment is due in 10 days (Wednesday, June 1).

Here is the assignment:

* I will have code for an analysis of Brooklyn housing data on the course website (from pages 49 and 50 in the O’Neil and Schutt text). The code is not available yet. You can use the code on pages 49-50 of O’Neil and Schutt if you want to work ahead, but it probably doesn’t work in places. It’s old R Code.
* Using the Rolling Data Sales website, download and examine another housing sales data set (not the one we will use in live session, which is linked here - [Rolling Housing Sales for NYC](http://www1.nyc.gov/home/search/index.page?search-terms=Rolling+sales+update)).
* Goal: Create an RStudio project for the analysis of this data set. Your file structure within the project should include the following:
  + A README file in the project root directory that includes an explanation of the purpose of the project and the other files
  + A data directory containing files to load in and clean up the data. The clean up should include finding out where there are outliers or missing values, deciding how you will treat them, making sure the dates are formatted correctly, making sure values you think are numerical are being treated as such (correct R class), etc.
  + An Analysis directory containing a file (or files) for exploratory data analysis on the clean data to visualize and make comparisons (i) across neighborhoods, and (ii) across time.
  + A Paper directory containing a file (plain text or Markdown) that explains any meaningful patterns in this dataset.
* Note: I am indifferent to the names of the directories as long as they a generally descriptive of the contents and follow naming conventions discussed in Gandrud.
* Deliverable: A link to a repository on GitHub containing the above. I need only one link per group. Since this is a group project, I expect you to divide the labor.