**Suggested Naming Conventions**

**Naming Conventions for Datasets**

|  |  |
| --- | --- |
| **Steps** | **Example** |
| Determine a short string of characters for a prefix | BRFSS |
| Add \_a to the first iteration | BRFSS\_a |
| Each time you remake the dataset (either subset it or add a variable), increment the letter suffix | BRFSS\_b |
| If you make a temporary dataset for any reason, such as to test the data or explore some data issues, change the suffix to an explanation | BRFSS\_WomenOnly |
| When you are ready to make your final analytic dataset, rename the last version of your dataset to Analytic. That way, analytic code will operate on a dataset that is always called Analytic. | If BRFSS\_q is our last version:  Analytic <- BRFSS\_q should be the last step |

**Naming Conventions for Code**

|  |  |
| --- | --- |
| **Naming Approach** | **Example** |
| Determine a numbering system to use as a prefix to all code file names. This is to get them to sort in order of how you run them. | I use three digits and this logic:   * Prep or example code starts with 0, followed by two digits. * Transformation code (where data are edited) start with 1 followed by two digits. * Analytic code starts with 2 followed by 2 digits * Subgroup analysis or other post-hoc (unplanned) analysis starts with 6 followed by 2 digits * Exploratory code starts with 7 followed by 2 digits |
| When you develop code files, make them short and modular – only doing one basic operation. | For example, if you are reading in a data table, you might make the code 100\_Read in 2014 BRFSS data and only include this operation. |
| When you make subsequent code, leave space in your numbering system so you can insert other code later. | If you use 100\_Read in 2014 BRFSS data as the name of your first data file, then 105\_Apply exclusions as your second, if you realize later you need to read in an additional year before applying exclusions, you can add 103\_Read in 2013 data in between the two code files. |
| If you are in the analysis phase and find you need to make a new variable, remake a variable, or otherwise alter the analytic data file, go back to the earliest file that needs editing, and edit and rerun the necessary filed going forward. | Imagine I created a variable for age group in my 100 series, but now I am analyzing age in my 200 series, and I want to instead represent it in quartiles.   1. Go to the code under the 100s where I created the age groups 2. Add code (either to that file or a new code file in between that one and the next one) to add the quartiles. 3. Edit the subsequent 100s files to make sure they will all run in order with these additional variables. 4. Update data dictionary with the new variables. 5. Continue with the 200 code for analysis |