Lab

# Lab 2: Manipulating Dataframes

The “dataframe” is one of the most essential data structures used in R. It is conceptually equivalent to a database “relation” and to the typical rectangular dataset with variables as columns and cases as rows. For this activity, you will gain some skill with manipulating a dataframe.

**Task 1**

R offers several built-in dataframes. For this activity we will use the “mtcars” dataset that contains 11 variables and 32 cases representing different models of cars.

The goal is to create a new variable for this dataframe that represents the engine displacement per cylinder in cubic inches for each vehicle. You may not know what displacement is (or maybe even cylinders), but it will suffice to know that values in the column named “disp” divided by values in the column named “cyl” will yield the appropriate quantity.

One fundamental principle of working with data is that you should never overwrite or change your original raw data. Therefore, your very first line of code should be:

my\_mtcars <- mtcars # Copy original dataframe into a new one

From that point forward you can work on my\_mtcars without mucking up the original data. Also note that in order to establish that you have completed the assignment correctly, your last command should summarize your new variable using the summary() function. The output of that final command should look exactly like this:

Min. 1st Qu. Median Mean 3rd Qu. Max.

17.77 26.92 34.48 35.03 43.19 59.00

**Task 2**

Gather some basic “demographic” information from about five friends or family members and then enter those data into a dataframe using the appropriate R commands. Finally, summarize the contents of the dataframe, again using the appropriate R commands. Keep the demographics “light” to avoid getting too personal. For each person report (1) the number of pets that they have (dogs, cats, etc.); (2) their birth order in their family (i.e., 1 for first born, etc.); and (3) the number of siblings they have.

Collect the necessary data from your friends and family members; write, test, and submit the necessary code in R to accomplish the following:

1. Create three vectors of integers as described above, using the c( ) (concatenate) command to store data reported by group members, with these variable names: Pets, Order, and Siblings.

2. Also create a vector of user IDs for the friends and family members.

3. Bind those four vectors together into a dataframe called myFriends.

4. Use the appropriate R command to report the structure of your dataframe as well as a summary of the data (with minimums, means, maximums, etc. as shown on page 32). The result should show, “X obs. Of 4 variables,” where X is the number of friends and family members who reported their data.

5. Use the $ notation explained on page 33 to list all of the values for each of the variables in the myFriends dataframe (e.g. myGroup$Pets).

**Hints**: All of the examples that you need in order to write the necessary R commands are right there in Chapter 5. The most challenging part is getting the data from your friends and family members. Don’t wait too long! It’s okay if not everyone you ask participates. Use the user IDs of friends and family members from item #2 to keep track of who participated.