**Operations and data frame**

Genevals2018.tsv is a file with 395 columns and 5,000 rows – one row per gene. The first column is the Gene\_id, the second column is the gene description, and the 3rd to 395 columns are gene expression data.

Write statements for the following:

1. Assume genevals is a tab separated file in /home/me/mydata/genevals2018.tsv. Read this file and assign to the variable genevals. Assume there is a header and the columns are named Gene\_id, gene description, and s????\_??C.
2. Get the number of rows
3. Get the number of columns
4. Get the value of the 4th column and 100th gene in the list
5. Get all the experiment values for Gene\_id is AT1G01140
6. Get the experiment values for Gene\_id 627 and only experiments between columns s0002\_16C and s0004\_10C.
7. Make a new dataframe in which you have data from genes in rows (this is by row number, not Gene\_id): row numbers are: 45, 32, 178, 195.
8. Your friend tells you that the data in experiment s1074\_16C was bad. Make a new dataframe with all the data except this column.
9. Your friend then says that there are additional problems – not only s1074\_16C was bad, but also s1158\_16C, s1257\_16C, and s1367\_16C. Make a dataframe that eliminates these (and s1074\_16C).
10. Get a “list” of all the column names so that you can check to make sure these columns are gone.
11. Write a statement to verify that s1074\_16C is missing from the new dataframe (in which you eliminated 567).
12. Write a statement to verify that s1074\_16C is in the old original dataframe
13. Write a statement to get the means of each data column
14. Write a statement to get the variances of each data column
15. The really key data is the Gene\_id and the data columns (description is nice but extra). Some software requires that you use a matrix, not a dataframe. Convert your dataframe into a matrix, retaining the Gene\_id information (hint: where would this logically be stored in the matrix?)

**The following assignments will be graded. Please upload your answers to D2L before the next class:**

1. 16C and 10C actually means different temperature. Would you try to say something about the differentially expressed genes between the two temperatures? (Note: this can be a semester-long research project. You can do whatever you can.)

Reading for the next class: The Art of R Programming Chapter 7.