BIOL-GA1007 PROGRAMMING FOR BIOLOGISTS
End of term take-home exam 2017

Due Friday 22nd December, 2017.

100 points total.

Part A: (40 points)
The dataset in the file "tetrahymena.tsv" is from a study of growth conditions of Tetrahymena cells. In two groups of cell cultures glucose was either added or not added to the growth medium. The average cell diameter (in microns) and cell concentration (count per ml) were measured for each culture. The measurements were repeated twice for each culture, giving two technical replicates for each culture. It is expected that the cell diameter is affected by the presence of glucose in the medium.

Note: first load the dataset in RStudio and examine it (e.g. using View()) , to understand the columns and their meaning.

Using R and the "tidyverse" packages- dplyr and ggplot- write a processing pipeline to process and visualise the dataset as follows:

(1) read in the dataset
(2) filter out excessively small cells with diameter <= 19.2
(3) use the mean concentration and diameter over the technical replicates (this will involve grouping the data on "culture" and "glucose" columns). This is to remove pseudo-replication.
(4) create new columns "log_conc" and "log_diameter" that have the natural log of concentration and diameter respectively.
(5) create a scatter plot of log concentration versus log diameter (if this is linear then the variables are related by a power law- comment on this); show the glucose status by plotting a different shape and color for points with and without glucose added.
(6) for 5 bonus marks, include a smoothed fitted line separately for the glucose added and not added groups

Upload an R script called "tetrahymena_part_A_NETID.R" and the plot in "tetrahymena_part_A_NETID.pdf".

Part B: (40 points)
Repeat this analysis in python, using the pandas and any plotting package.

Upload an R script called "tetrahymena_part_B_NETID.py" and a pdf file "tetrahymena_part_B_NETID.pdf".

Part C: (20 points)
When writing your code for parts 1 and 2, use git for version control.

Tag the final commit with the name "END_TERM_EXAM_2017".

Save the output of "git log" after you finish the code in part 1 and 2 to a text file called "git_log_part_C_NETID.txt"

Upload both the shell script and the txt file.