Assignment #5 - Modelling yeast gene expression under nutrient starvation

This dataset evaluates growth rate and gene expression in yeast for different nutrient starvation conditions (dropout media). Most yeast strains have both a systematic name (ie. YBR158W) and a common name (iw. AMN1). This assignment requires the use of a linear model to determine gene expression as a function of growth rate and nutrient starvation. An increasing rate of growth assumed to be proportional to increased nutrient in the media (less starvation). Functions from dplyr, tidyr, stringr, ggplot2 and limma should be used whenever possible. Plots should have reasonable axis labels and titles.

Data from: *Brauer et al 2018. Coordination of Growth Rate, Cell Cycle, Stress Response, and Metabolic Activity in Yeast.*

Initial data cleaning by David Robinson

Questions:

a) Reshape the data into tidy format. Familiarize yourself with the data and perform any additional data-cleaning that is necessary.(2 marks)

b) Make an exploratory plot for a small subset of the data and describe any trends you find. (3 marks)

c) Filter the data to keep only genes relating to transport in the Biological Process (BP). Plot the data for these genes and explain why the chosen method is informative. What is the data telling you? (3 marks)

d) Model expression as a function of growth rate, nutrient, and gene deletion. What type of linear model is this and why? What is the interpretation of the intercept? What is the interpretation of the slope? (5 marks)

e) Is there an interaction between nutrient starvation and genotype? Is this significantly different from a model without an interaction term? Is this what you would expect? (3 marks)

f) What *biological* assumptions does this model make? (1 mark)

3 marks will be given according to the following rubric:

- 3.0 - code is well-documented and concise
- 1.5 - code is either well-documented or concise, but not both
- 0 - no attempt was made to document code, extra variables are created, code is difficult to read

Total: 20 marks

Submission: Each student will upload a .R file to Quercus. Please include your first and last name, the date of submission, and the assignment number.

Due date: 11:59pm February 20th, 2019