Assignment #4 - Exploratory Analysis of an Arabidopsis Chemical Genetic Screen

For this assignment, we are using a simulated dataset inpired by a chemical genetic screen in Arabidopsis (doi: 10.1186/1471-2229-12-226). Wildtype and mutant seeds from Arabidopsis were exposed to various concentrations of sulfanilamide family compounds or a dmso control. 3 replicate plates (A, B, C) were seeded for each combination of genotype, concentration and compound. Seedling weight, height (final_growth), and phenotype (whether leaves were Bleached or Green) were measured at the end of a 14 day period.

The purpose of this assignment is to practice plotting with ggplot2. Follow the steps below using functions from the dplyr, tidyr, stringr, and ggplot2 packages to accomplish your tasks whenever possible.

Each question is to be answered with a plot, as well as comments interpreting the plot to state your answer. The use of plot type, color, facetting, and scaling are tools to help emphasize data. Each plot is expected to have reasonable axis labels and titles. Use 2 different themes and 2 different color palettes during the assignment.

Data reshaping (no plot is needed for this question):

a) Read in the data and reshape it into tidy format. Your columns should reflect the above description of the dataset. Your final result should have 1800 rows and 7 columns. You can have different observational units in this table for this assignment. Familiarize yourself with the dataset and perform any additional data-cleaning that is necessary. (3 marks)

Plotting:

b) What is the relationship between seedling weight and final height? (3 marks)

c) Which compound and concentration have the strongest effect overall? (3 marks)

d) What is the relationship between genotype and phenotype? (3 marks)

e) Is there a genotype which is more or less resistant to a compound than wildtype? (3 marks)

f) Are replicates plates comparable? (See the description of replicates above) (3 marks)

g) Take your favorite plot (or create a new one) containing a legend and change the color of the panel background. Make the title larger than default and centered, and remove the legend. (3 marks)

h) Take your favorite plot (or create a new one) containing facetting and change the facet labels (strip text). Let the y-axis be scaled freely. (2 marks)

i) Take your 4 favorite plots (or create new ones) and make a figure for publication using ggpubr. These can be aligned 2x2 or 1x3 depending on your preference. Make any adjustments to alignment or text that is necessary for the plots to appear uniform. Save this as an image using ggsave() to be submitted with your assignment. Check the image to make sure the size and scaling seem appropriate. (4 marks)

A total of 2 marks will be given for the 2 different themes and use of 2 different color palettes.

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: 32 marks

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

Due date: 11:59pm February 13th, 2019