Assignment 8: Iris Statistics and Linear Model

Assignment Objectives

1. Using the iris data set, calculate the min, max, mean, and median for **each variable or trait (each column)** and for **each species**. For example, species *setosa* would have the aforementioned statistics for Sepal.Length, Sepal.Width, Petal.Length, and Petal.Width. Add all of the statistics into a table and format it nicely in your markdown file. This should result in one table containing the statistics for all three species and their four variables (traits).

2. Using the iris data set, determine if Sepal.Width is a good predictor of Sepal.Length across all species (not individually) using a linear model with lm(). Report the r^2, p-value, and slope from your lm() call in a nice table with appropriate column labels. In the table caption mention the type of relationship (positive or negative) and the variables you regressed (including specifying which was the dependent and which the independent variable).

3. Use ggplot to graph the linear model you just created in 2. You can just use ggplot's geom_smooth function with method set to "lm" to add your regression line to the scatter plot. Alternatively, **and this is optional**, you can try and add a line and confidence interval manually using your model outputs with geom_abline() (for the line) and geom_ribbon() (for the confidence interval) as we did in class. In the figure caption mention the type of relationship (positive or negative), the variables you regressed (including specifying which was the dependent and which the independent variable) and the statistics of the regression (r^2, p-value).

What you submit

A document rendered from your .Rmd file as a report. The report should include:

1. Introduction: 2-4 sentences introducing the data and analysis.
2. Code and annotation: show me your code and annotate it as you have in past assignments. Use headers for sections, etc. (NOTE: Except for your final tables and figure, I do not need to see the output from your code, only the code itself).
3. Two high quality tables, one for iris statistics and one for the model output statistics.
4. One high quality figure showing the relationship you modeled with a regression line. A caption should be included complete with statistics. For example: "Figure 1. Scatter plot of variable 1 versus variable 2 from the iris data set for all species. Variable 2 showed a pos/neg and significant/insignificant relationship with variable 1 (r^2 = xx, p-value =)."
5. Conclusion (2-4 sentences summarizing your findings)

Important notes:

- You may render your markdown as a Word, PDF or HTML document. If you decide to render to an html document, you need to publish the document on Rpubs, and submit the link of your published document webpage on Canvas. PDFs and Word docs can be submitted directly to Canvas.
- You should arrange your report clearly using sections and headers. Correctly use the R markdown syntax to set sections.