STAT 540 Final Exam-Fall 2020

This is a take-home exam; all the usual course resources are available, but you are not to discuss the exam with other students. Contact me if you have questions.

We will use a COVID-19 dataset from Johns Hopkins GitHub website (worksheet Q3Q4 in Final2020.xlsx Excel workbook) and a dataset from an evergreen tree growth study (workshett Q1Q2 in Final2020.xlsx Excel workbook). The first data set is comprised of a week's worth of COVID deaths from South Carolina reported at the county level. The second data set is comprised of the growth history of individual trees. Information for Q2 includes Site (3 levels), Species (BS=Black Spruce, EL=European Larch, JL=Japanes Larch), Tree, and height and diameter measurements; the measurements are irregular and incomplete, as you can tell by inspecting the worksheet, though the missing values actually do not present any special issues for the questions below.

1. (25 points) Use worksheet Q1Q2 for the following questions
    (a) Transpose the height records by Site, Species and Tree.
    (b) After inspecting the results in (a), remove _LABEL_, and replace _NAME_ and COL1 with more appropriate variable names.
    (c) Now transpose the diameter records by Site, Species and Tree.
    (d) Match merge the two transposed data sets in (b) and (c) by Site, Species and Tree. Don't worry about cleaning up the former _NAME_ variable(s); you will be tested on that in Question 2.
    (e) Graduate students should re-transpose the data set in (b).

2. (25 points) Use the data set from 1(b) for the following questions.
    (a) Use scan() and input() to create a numeric variable Year (with values 5, 10, 15, 16, and 27) from YearLabel.
    (b) Print output data set with full names of species.
    (c) Write code to save the data set from 2(a) in an Excel workbook.

3. (25 points) Use the worksheet Q3Q4 to answer the following questions
    (a) Provide commands to generate a plot similar to Plot 1 on the website. Graduate students should provide ODS commands to save the graph with an appropriate name as a JPG file on a directory of your choice.
    (b) Provide commands to generate a plot similar to Plot 2 on the website.

4. (25 points) Use Q3Q4 to answer the following questions.
    (a) Report the maximum number of DailyDeaths for each county by using PROC MEANS.
    (b) Use ODS SELECT and ODS OUTPUT in PROC MEANS to save the table generated by PROC MEANS as SAS data set COUNTYMAX.

(c) Based on the following code, use %LET to allow the user the flexibility to specify the county for the WHERE and OUT= statements; be sure to rename the output file appropriately. Test your code for Richland County.

```
proc means data=q2 median; where county="Richland";
var DailyDeaths;
output out=Richlandmedian (drop=_type_ _freq_)
  median(dailydeaths)=MedianDeaths;
run;
```

(d) Write a macro named CountyMedian to carry out the same task as the %LET command in 4(c). Test your code using Richland and Lexington counties.