STAT 540: Final Exam

All data sets you need for the exam can be found as separate worksheets in the Excel workbook; import them as needed. Any data manipulation must take place within your SAS programs. You may consult online documentation, but you cannot seek external help, other than myself.

1. Read the Excel worksheet Q1 into SAS. It lists collection size and annual circulation at eight high school libraries for graphic novels, along with total collection size ad total circulation.

   (a) Transpose the data so that it has the following layout. It helps to transpose the circulation columns and collection columns separately, and then match merge the two data sets. Print the data set so that it appears similar to the layout below.

School	Graphic Novel Coll	All Coll	Graphic Novel Circ	All Circ
1	950	6570	3030	12290
2	650	7040	2450	12480
...	...	...	...	...
8	830	5980	3120	11590

2. Use the variable labeled **Flow** in Excel worksheet Q2 to re-create Plot 1 on the website using PROC SGPLOT. You may need some data transformations to generate the plot. In addition, please use the following lines of code to convert the character variable **Date** to the date variable **RealDate** and format **RealDate**.

   ```
   format realdate mmddyy10.;
   realdate=input(date,mmddyy10.);
   ```

   This extra step is necessary since Excel interprets date formats prior to 1900 as character variables.

   Grad students only: Replicate Plot 2 as well.

3. The data in Excel worksheet Q3Q4 contains counts of some common bird species at four different Christmas Bird Count (CBC) locations in South Carolina for the years 2010-2012. The data were collected for a single sector of each count circle over the course of a full day.

   (a) Use PROC MEANS to print the maximum count for each species for each year; no other statistics (mean, etc) should be printed.

   (b) Use PROC MEANS to compute the mean count for each species for each circle. The output should display the means with only two significant digits.

   (c) Save the means from the above step and match merge them with the original data set.

   (d) *Graduate students only*: Write code to identify and print the circles associated with each high count from step (a); multiple data steps are allowed.

4. Use the Excel worksheet Q3Q4 for the following.

   (a) Use %LET to select a given circle and plot all species over **Year** for that circle on the same plot with SGPLOT. Demonstrate your code for the Santee NWR circle, being sure to include an appropriate title.

   (b) Use CALL SYMPUT to select the circle with the highest mean for Ruby-crowned Kinglet, then plot Ruby-crowned Kinglet over **Year** for that circle using SGPLOT. Include an appropriate title.

   (c) *Graduate students only*: Write a macro to complete the task in (b) that uses the species name as an argument. Demonstrate the macro for Red-bellied Woodpecker.