**Lab 7**  *Some of these problems may be more challenging than others. Please feel free to work with others, attend office hours, or post on the course discussion forum if you need help. While collaboration with other students is encouraged, each student is responsible for submitting his or her own work. This assignment should be submitted in one well-commented SAS program. For any questions that require a written answer, do so in the SAS comments. Be sure to re-name the uploaded SAS scripts according to the naming convention* `LastnameFirstinitial_Lab#.sas` *(e.g.,* `PileggiS_Lab7.sas`*).*

These exercises have been adapted and modified from *Exercises and Projects for The Little SAS Book.*

1. Create a macro variable called `path` that corresponds to the computer location of your STAT 330 data sets.

2. Create a SAS library called `flash` that uses the `path` macro variable.

3. The BabyCentre website publishes the top 10 baby names by gender in various countries. The names are contained in the SAS data sets `australia`, `brazil`, `france`, `india`, `russia`, and `unitedstates`. Use SAS procedures to examine the SAS data sets, paying attention to the variable names, labels, and attributes. In a comment in your SAS code, note features of the data that will need to be addressed when combining the data.

4. Focus on the `unitedstates` data set first. Use a DATA step to:

   (a) Create a temporary data set called `unitedstates`.

   (b) The names of girls and boys are listed in order of ranking (such that the first observation is the most popular name and the $10^{th}$ observation is the $10^{th}$ most popular name). Use the SAS statement $\boxed{\texttt{rank + 1 ;}}$ to create a variable called `rank` for the popularity ranking of the name. (We'll discuss how this code works in a future lecture.)

   (c) Create a variable called `country` which has a value of `unitedstates` for all names in this data set.

   (d) Lastly, print the `unitedstates` data. Your output should match that shown below.

**unitedstates**

| Obs | Girl | Boy | rank | country |
|---|---|---|---|---|
| 1 | Sophia | Jackson | 1 | unitedstates |
| 2 | Emma | Aiden | 2 | unitedstates |
| 3 | Olivia | Liam | 3 | unitedstates |
| 4 | Isabella | Lucas | 4 | unitedstates |
| 5 | Mia | Noah | 5 | unitedstates |
| 6 | Ava | Mason | 6 | unitedstates |
| 7 | Lily | Jayden | 7 | unitedstates |
| 8 | Zoe | Ethan | 8 | unitedstates |
| 9 | Emily | Jacob | 9 | unitedstates |
| 10 | Chloe | Jack | 10 | unitedstates |

5. Copy and paste your code from the previous question. Convert this into a macro called `create_rank`. Execute this macro on all data sets (`australia`, `brazil`, `france`, `india`, `russia`, and `unitedstates`).

6. Combine the data sets by interleaving by popularity ranking (so that all 1's are grouped together, then all 2's, etc.). Make sure the resulting data set only has one variable for boy names and one variable for girl names. Print your final data set. The first 10 observations are shown below so you can verify your output.

| Obs | Girl | Boy | rank | country |
|---|---|---|---|---|
| 1 | Charlotte | Oliver | 1 | australia |
| 2 | Sophia | Miguel | 1 | brazil |
| 3 | Emma | Nathan | 1 | france |
| 4 | Saanvi | Aarav | 1 | india |
| 5 | Sofia | Artem | 1 | russia |
| 6 | Sophia | Jackson | 1 | unitedstates |
| 7 | Ava | Jack | 2 | australia |
| 8 | Julia | Davi | 2 | brazil |
| 9 | Lola | Lucas | 2 | france |
| 10 | Aanya | Vivaan | 2 | india |

7. Information Technology Services (ITS) at Central State University has a computing service called "The Grid," which is offered to faculty, staff, and students. ITS tracks information on registered users in a SAS data set called `users` and information about the projects that they have registered on The Grid in a SAS data set called `projects`. Users are allowed to register more than one project at a time; however, only one

project can actually be processed at a time. Use SAS procedures to examine the SAS data sets, paying attention to the variable names, labels, and attributes. In a comment in your SAS code, (1) note the number of observations in the `users` and `projects` data sets, and (2) state your game plan for combining this data - would it be a stack, interleave, one to one merge, or one to many merge?

8. Combine the `users` and `projects` data sets into one SAS data set. Utilize PROC CONTENTS and verify that you have 7,273 observations in your data set.

9. Utilize a data step to create three SAS data sets: (1) incomplete projects (no end date), (2) completed projects (with an end date), and (3) users with no registered projects (these are users that exist in the `users` data set but have no observations in the `projects` data set). Examine your log to determine the number of observations in the data sets, and note these results as a comment in your SAS code. Utilize PROC CONTENTS on these three data sets and verify that you have 3,146 observations with incomplete projects, 4,127 observations with complete projects, and 7 observations that are users with no registered projects. *(Hint: review Lecture 4 slide 27 for review on subsetting to multiple data sets.)*