Stat 604
Assignment 9 - SAS

OBJECTIVES: In this assignment you will practice accessing and creating different kinds of data with SAS.

You should have all of the information you need to complete this assignment by viewing the first 4 SAS lectures.

This assignment will use the **COVID Activity.csv** file that was downloaded and used in the R assignments. Re-familiarize yourself with the structure of this file and the data it contains. Because this is such a large file, a copy has been uploaded to the Fall2021 shared folder on the SoDA course for you to access from SAS Studio. Create a new folder that you can use to store permanent data sets that are created in this and upcoming assignments. I will refer to this as the "mylib" folder throughout the assignment instructions.

This assignment will be unique in that there will be four PDF files and one Excel file posted to Canvas – one PDF for the program, one PDF for the log, two PDF output files and an Excel file also containing SAS output. You will create a second Excel file in this assignment, but it will not be submitted to Canvas. Make sure you submit the correct file. It is wise to read through the entire assignment before beginning.

1. Add a header comment section to the beginning of a new program in your SAS session. Be sure to include a comment line above each section of the program that identifies the associated assignment step and a brief description of what the section is doing. Include housekeeping statements to clear titles and footnotes and suppress the printing of procedure titles.
2. Assign a libref named **mylib** to the new folder created above. Assign a libref to a new Excel file with a name of your choosing that will also be located in the mylib folder. Create a fileref to the COVID Activity csv file. Create filerefs to the two PDF output files. The files will have names like FKincheloe_HW09_OutputA.pdf and FKincheloe_HW09_OutputB.pdf. Create another fileref for an xlsx file that will contain your output. Use a name for the file like FKincheloe_HW09_outputx.xlsx. (Use your own initial and name in place of FKincheloe.) Filerefs must contain the full path and name of the file.
3. Write a SAS step that will use the csv file as input and create a permanent data set in the **mylib** library. Write the step so that it will overwrite the data set if it already exists. NOTE: Due to the size of the file, it could take a minute or more for this step to run.
4. Close all active ODS destinations and open the PDF Output A destination that you will use capture the output from all procedures in this assignment. This PDF will include a table of contents page but no bookmarks. Apply a style of your choice. Open the second PDF Output B destination that will contain ONLY the output from the procedure in the next step. This output will use the default style. Since it will contain the output from only one procedure, it is to have no contents or bookmarks. You may want to delay implementing the ODS statements until you have everything else working correctly in your program.
5. Write a PROC step that will report the descriptor portion of the permanent data set created from the CSV file. Supply an appropriate title and an appropriate proc label. If the length of the COUNTY_NAME column is less than 17, you still have work to do on the process that reads the csv file.

6. Use the permanent data set as input to create a temporary data set having only those rows where the county name is Brazos.
7. Use the permanent data set as input to create a table of high death counts in the Excel "library" based on the value of death_new_count being more than 4000.
8. Use the permanent data set as input to create a table of "Corrections" in the Excel "library" based on the value of positive_new_cases_count being less than 0.
9. Use a system option to limit processing to 10 observations to print a sample of the data from the Brazos County data. Supply an appropriate title. Since this output is still going to a PDF file, supply an appropriate proc label. Be sure to reset the option when this step is completed.
10. Open the Excel destination to capture the output from the procedures that follow. You may need to refer to SAS Help documentation to find the option values that will produce the desired output. Your titles must appear within the worksheets so you can see them whenever you open the Excel file.
11. Use a PROC step to list all of the worksheets in the Excel library without printing their descriptor portions. All the output tables from this step must be in a single Excel tab named Covid Table List. Supply an appropriate title and proc label.
12. Print the data portion of the table of high death count to a new sheet named High Covid Deaths. Supply an appropriate title and proc label.
13. Print the descriptor portion of the Corrections table in a single new sheet named Corrections Descriptor. Supply an appropriate title and proc label.
14. Close the Excel destination. It should contain three tabs. Close the PDF destination. Include a line of code that will reopen the default HTML destination. This line of code was supplied in one of the demo programs used in the lectures.
15. After you have finished debugging your program, run it from a new instance of SAS and convert or save the log to a PDF document for submission.
16. Use the log and report information contained in your output documents to find the answers to the questions below and include the answers in a comment section at the bottom of your program file:
    a. Close/Sign Out and then reopen SAS. Use the browsing features in SAS to examine which library references exist. Open the program written for this assignment and execute only the statement to create the mylib library reference. Search the libraries for the data set that was created from the CSV file and the data set of Brazos County data. Explain your findings.
    b. What is the Engine for the new data set created in Mylib?
    c. What is the data type, format and length of the REPORT_DATE variable in the permanent data set?
    d. How many observations are in the Corrections table?
    e. How many lines in the table of contents are created by the first output procedure?
17. Save the final version of the program and convert it to a PDF file with a name like FKincheloe_HW09_prog.pdf. Beware that printing from the SoDA web page will probably not print more than the first page of your code.
18. Upload and submit the five documents to the assignment on Canvas.