# Stat 604
# Assignment 13 - SAS

OBJECTIVES: In this assignment you will practice creating and using custom formats, processing groups of data and overriding the default behavior of the DATA step.

You should have all the information you need to complete this assignment by viewing the first 12 SAS lectures. Programming efficiency must be incorporated throughout the program.

This assignment will use the **"All Texas" -** permanent data set that was created in Homework 10 and used in Homework 11. If you had difficulty creating this data set, the professor's version, named **alltx.sas7bdat,** is available on the Week 9 module in Canvas and in the Fall2021 folder on SoDA. Refamiliarize yourself with this data set before you start writing your program code.

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. If you are using the professor's data set, assign a libref to the folder where it is located and add **access=readonly** at the end of the libname statement, before the semicolon, to protect data sets in this folder from being accidentally overwritten. Assign a libref to the **mylib** folder containing your permanent data sets. Create a fileref to the pdf file for output. Ensure that your SAS session can locate any permanent user defined formats that you create.
3. Open a PDF destination to receive your output.
4. Create a permanent custom format in the mylib library. It is to be a numeric format that can be applied to raw percentages and put them in categories. A value of 0 will be displayed as 'None'. Values above 0 through one percent (.01) will be shown as 'Low'. Values above .01 through .04 will be shown as 'Medium'. Values above .04 through 10 percent (.10) will be shown as 'High'. Those with a value above 10 percent through 15 percent (.15) will be in the 'Very High' category. Values over 15 percent will be considered "Extreme". Any values not in these ranges will be displayed as 'N/A'. At the time of creation, send the documentation (listing) of this format to the output destination. If you have been experimenting with permanent formats, either delete any pre-existing formats before doing this assignment or include a statement on this step to ensure that only the format created in this assignment is documented in the output.
5. Write a single SAS step that will use the permanent "All Texas" data set as input and create a temporary data set with the following modifications:
    a. The two variables containing Percentage of Fatal Cases and the Fatality Groups are not useful for this assignment and should be removed.
    b. Give each of the four variables, whose name ends with COUNT, a shorter name like Case_Total or Deaths_New as appropriate. Give them a permanent label that is in proper case and replaces the underscore in the name with a space.
    c. Give the county_name variable a permanent label of County.
    d. Give the County_FIPS_Number variable a permanent label of FIPS.
    e. Create a new variable that contains the month number of the Report_Date. Give this variable a label of "Mo." (including the period).

    f. Use the monname format with the report date to create a new character variable that contains the full name of the month. Give this variable a permanent label of Month.
    g. Create a new variable that contains only the year from the Report_Date. Give this new variable a label of Year.
6. Reorder the new temporary data set so you can use it for by group processing based on the county and the Report Date.
7. Use the new temporary data set as the source for a DATA step that will create a summary of the new cases and new deaths in each county for each month. Since the data span multiple years, the year variable must also be part of the grouping. Note: Even though year and month were not explicitly specified when reordering in the previous step, using a date orders by year, month and day so you can use the year and month variables for the by groups. Store the new data set in the mylib library.
    a. Permanently label these new summary variables "Monthly Cases" and "Monthly Deaths" respectively.
    b. At the end of each month, calculate the "Fatality Rate" for the county by dividing the Monthly Deaths by the Monthly Cases. Use conditional logic to prevent making a calculation that would produce a divide by zero message in the log. Apply a permanent label and the custom format to this variable.
    c. Since new cases and new deaths are daily values, they are not needed in the output data set. Include the county name and FIPS number.
    d. Along with the month number, month name and year number, this will make a total of 8 variables in this data set. This data set should have 5355 rows.
8. Create a list of all objects in the mylib library without displaying the descriptor portions of data sets. Supply an appropriate title.
9. Report the descriptor portions of the permanent data set created above. Supply an appropriate title.
10. Print the monthly summaries for Brazos and McLennan counties. Exclude the FIPS variable, month number, and observation numbers from the report. Show labels instead of variable names. Supply an appropriate title.
11. Print a second report from the monthly summary data except only include rows that are in the Extreme fatality rate category. Since we know the rate is extreme, override the formatted value with a temporary format to show the actual percent. Supply an appropriate title. The format of the report is shown below:

| County | Year | Month | Monthly Cases | Monthly Deaths | Fatality Rate |
|---|---|---|---|---|---|
| Anderson | 2021 | March | 20 | 5 | 25.0% |

12. Close the PDF destination.
13. Use the log and report information contained in your PDF output document to find the answers to the questions below and include the answers in a comment section at the bottom of your program file:
    a. What length is reported for the permanent custom format you created?
    b. How do the summaries for August 2020 for Brazos County compare to the summaries for August 2021?
    c. How does Brazos County compare to McLennan County in January 2021?
    d. When did Hamilton County have an extreme fatality rate and what was the percentage?

e. Consider the overall sizes of the Monthy Cases and Monthly Deaths among the Extreme fatality rates. What observation(s) can be made about the records with an Extreme fatality rate?

14. Save the final version of the program and convert it to a PDF file. Convert the log to PDF.
15. Upload and submit the three PDF documents to the assignment on Canvas.