```
/************************************************************************/
/* Program Name: STAT 604 HW#13                                         */
/* Date Created: 11/20/2021                                             */
/* Author: Jack Rodoni                                                  */
/* Purpose: STAT 604 HW#13                                              */
/* Date Modified: 11/23/2021                                            */
/* Location: /home/u59649056/Homeworks/JRodoni_Homework13.sas           */
/************************************************************************/

/* 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. */

title;
footnote;
ods noProctitle;


/* 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. */

libname mylib "/home/u59649056/Homeworks/mylib";
filename HW13pdf "/home/u59649056/Homeworks/mylib/JRodoni_HW13_Output.pdf";
libname  HWDATA "/home/u59649056/Homeworks/Homework Data" access = readonly;

/* 3. Open a PDF destination to receive your output. */

ods pdf file=HW13pdf;


/* 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. */

proc format library=mylib fmtlib;
    value pct 0 = 'None'
              0 < -.01 = 'Low'
              .01 <-.04 = 'Medium'
              .04 <-.1 = 'High'
              .1 <- .15 = 'Very High'
              .15 <- high = 'Extreme'
              OTHER = 'N/A';
    select pct;
run;


/* 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. */


data temp;
    set HWDATA.alltx;
    drop pct_fatal_cases Fatality_Group;
    rename POSITIVE_CASES_COUNT=Total_Cases
           DEATH_NEW_COUNT=New_Deaths
           POSITIVE_NEW_CASES_COUNT=New_Cases
           DEATH_COUNT=Total_Deaths;

    Month = month(REPORT_DATE);
    Month_char = put(REPORT_DATE,monname.);
    year = year(REPORT_DATE);

    label Total_Cases='Total Cases'
          New_Deaths='New Deaths'
          New_Cases='New Cases'
          Total_Deaths='Total Deaths'
          county_name='County'
          County_FIPS_Number='FIPS'
          Month='Mo.'
          Month_char='Month'
          year='Year';
run;



/* 6.) Reorder the new temporary data set so you can use it for by group processing based on the  */
/* county and the Report Date. */

proc sort data=temp;
    by COUNTY_NAME REPORT_DATE;
run;


/* 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. */

data mylib.newdata;
    set temp;
    drop New_Deaths New_Cases Total_Cases Total_Deaths Report_Date;
    by COUNTY_NAME year Month;
    if first.Month then do;
```

```sas
        M_CASES=0;
        M_DEATHS=0;
        end;
    M_CASES+NEW_CASES;
    M_DEATHS+NEW_DEATHS;
    if last.Month;
        FATALITY_RATE = divide(M_DEATHS,M_CASES);
    options fmtsearch=(mylib.formats);
    format FATALITY_RATE pct.;

    label M_CASES='Monthly Cases'
          M_DEATHS='Monthly Deaths'
          FATALITY_RATE='Fatality Rate';
run;


/* 8.) Create a list of all objects in the mylib library without displaying the descriptor portions of data  */
/* sets. Supply an appropriate title.   */

title "Mylib contents";
proc contents data=mylib._ALL_ nods;
run;

/* 9.) Report the descriptor portions of the permanent data set created above. Supply an appropriate  */
/* title.   */

title "Monthly Covid Descriptor";
proc contents data=mylib.newdata;
run;

/* 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. */

proc print data=mylib.newdata label noobs;
    var COUNTY_NAME Month year M_CASES M_DEATHS FATALITY_RATE;
    where COUNTY_NAME IN ("Brazos","McLennan");
    title1 "Brazos and McLennan Monthly COVID Information";
run;


/* 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: (see pdf of assignment instructions for table) */

title "Extreme Fatality Rate Data";
proc print data=mylib.newdata label noobs split='*';
    where FATALITY_RATE>.15;
    format FATALITY_RATE percent7.1;
    var COUNTY_NAME year Month_char M_CASES M_DEATHS FATALITY_RATE;
    label COUNTY_NAME ='*County'
          year='*Year'
          Month_char='*Month'
          M_CASES='Monthly*Cases'
          M_DEATHS= 'Monthly*Deaths'
          FATALITY_RATE = 'Fatality*Rate';
run;


/* 12. Close the PDF destination. */

ods pdf close;

/* 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? */

/*                                    9                                        */

/* (b) How do the summaries for August 2020 for Brazos County compare to the summaries  */
/* for August 2021? */

/*          Monthly cases, monthly deaths and fatality rate both increased in 2021      */


/* (c) How does Brazos County compare to McLennan County in January 2021? */

/* Although they are both fall under the Medium fatality rate group, the actual fatality rate for McLennan */
/* is much higher compared to Brazos */

/* (d) When did Hamilton County have an extreme fatality rate and what was the percentage? */

/*                        May 2020 with 33.3% fatality rate.                            */

/* (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? */

/* Extreme fatality rate doesnt always seem to be a good indicator of outliers as most of the countys in the  */
/* extreme fatality group have a small number of cases. For instance in May 2020 Hamilton county  */
/* had a 33.3% fatality rate but they only had 1 death and 3 cases. These small sample sizes can skew  */
/* the fatality percentage greatly */

/* 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 */
```