```
1          OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
NOTE: ODS statements in the SAS Studio environment may disable some output features.
69
70         /*********************************************************************************/
71         /* Program Name: STAT 604 HW#13   */
72         /* Date Created: 11/20/2021   */
73         /* Author: Jack Rodoni   */
74         /* Purpose: STAT 604 HW#13   */
75         /* Date Modified: 11/23/2021    */
76         /* Location: /home/u59649056/Homeworks/JRodoni_Homework13.sas   */
77         /*********************************************************************************/
78
79         /* This assignment will use the “All Texas” - permanent data set that was created in Homework 10 and  */
80         /* used in Homework 11. If you had difficulty creating this data set, the professor’s version, named  */
81         /* alltx.sas7bdat, is available on the Week 9 module in Canvas and in the Fall2021 folder on SoDA.  */
82         /* Refamiliarize yourself with this data set before you start writing your program code. */
83
84         /* 1. Add a header comment section to the beginning of a new program in your SAS session. Be sure  */
85         /* to include a comment line above each section of the program that identifies the associated  */
86         /* assignment step and a brief description of what the section is doing. Include housekeeping  */
87         /* statements to clear titles and footnotes and suppress the printing of procedure titles. */
88
89         title;
90         footnote;
91         ods noProctitle;
92
93
94         /* 2. If you are using the professor’s data set, assign a libref to the folder where it is located and add  */
95         /* access=readonly at the end of the libname statement, before the semicolon, to protect data  */
96         /* sets in this folder from being accidentally overwritten. Assign a libref to the mylib folder */
97         /* containing your permanent data sets. Create a fileref to the pdf file for output. Ensure that  */
98         /* your SAS session can locate any permanent user defined formats that you create. */
99
100        libname mylib "/home/u59649056/Homeworks/mylib";
NOTE: Libref MYLIB was successfully assigned as follows:
      Engine:        V9
      Physical Name: /home/u59649056/Homeworks/mylib
101        filename HW13pdf "/home/u59649056/Homeworks/mylib/JRodoni_HW13_Output.pdf";
102        libname  HWDATA "/home/u59649056/Homeworks/Homework Data" access = readonly;
NOTE: Libref HWDATA was successfully assigned as follows:
      Engine:        V9
      Physical Name: /home/u59649056/Homeworks/Homework Data
103
104        /* 3. Open a PDF destination to receive your output. */
105
106        ods pdf file=HW13pdf;
NOTE: Writing ODS PDF output to DISK destination "HW13PDF", printer "PDF".
107
108
109        /* 4. Create a permanent custom format in the mylib library. It is to be a numeric format that can be  */
110        /* applied to raw percentages and put them in categories. A value of 0 will be displayed as ‘None’.  */
111        /* Values above 0 through one percent (.01) will be shown as ‘Low’. Values above .01 through .04  */
112        /* will be shown as ‘Medium’. Values above .04 through 10 percent (.10) will be shown as ‘High’.  */
113        /* Those with a value above 10 percent through 15 percent (.15) will be in the ‘Very High’ category.  */
114        /* Values over 15 percent will be considered “Extreme”. Any values not in these ranges will be  */
115        /* displayed as ‘N/A’. At the time of creation, send the documentation (listing) of this format to  */
116        /* the output destination. If you have been experimenting with permanent formats, either delete  */
117        /* any pre-existing formats before doing this assignment or include a statement on this step to  */
118        /* ensure that only the format created in this assignment is documented in the output. */
119
120
121        proc format library=mylib fmtlib;
122
122      !  value pct 0 = 'None'
123        0 < -.01 = 'Low'
124        .01 <-.04 = 'Medium'
125        .04 <-.1 = 'High'
126        .1 <- .15 = 'Very High'
127        .15 <- high = 'Extreme'
128         OTHER = 'N/A';
NOTE: Format PCT is already on the library MYLIB.FORMATS.
NOTE: Format PCT has been written to MYLIB.FORMATS.
129
129      !  select pct;
130        run;
```

```
NOTE: PROCEDURE FORMAT used (Total process time):
      real time           0.02 seconds
      user cpu time       0.02 seconds
      system cpu time     0.00 seconds
      memory              1175.87k
      OS Memory           30628.00k
      Timestamp           11/23/2021 07:54:27 PM
      Step Count                        82  Switch Count  0
      Page Faults                       0
      Page Reclaims                     184
      Page Swaps                        0
      Voluntary Context Switches        9
      Involuntary Context Switches      0
      Block Input Operations            8
      Block Output Operations           72


131
132
133        /* 5.) Write a single SAS step that will use the permanent “All Texas” data set as input and create a  */
134        /* temporary data set with the following modifications:  */
135        /* (a) The two variables containing Percentage of Fatal Cases and the Fatality Groups are not */
136        /* useful for this assignment and should be removed.  */
137        /* (b) Give each of the four variables, whose name ends with COUNT, a shorter name like  */
138        /* Case_Total or Deaths_New as appropriate. Give them a permanent label that is in  */
139        /* proper case and replaces the underscore in the name with a space.  */
140        /* (c) Give the county_name variable a permanent label of County. */
141        /* (d) Give the County_FIPS_Number variable a permanent label of FIPS. */
142        /* (e) Create a new variable that contains the month number of the Report_Date. Give this  */
143        /* variable a label of “Mo.” (including the period). */
144        /* (f) Use the monname format with the report date to create a new character variable that  */
145        /* contains the full name of the month. Give this variable a permanent label of Month. */
146        /* (g) Create a new variable that contains only the year from the Report_Date. Give this new  */
147        /* variable a label of Year. */
148
149
150        data temp;
151        set HWDATA.alltx;
152        drop pct_fatal_cases Fatality_Group;
153        rename POSITIVE_CASES_COUNT=Total_Cases
154            DEATH_NEW_COUNT=New_Deaths
155            POSITIVE_NEW_CASES_COUNT=New_Cases
156            DEATH_COUNT=Total_Deaths;
157
158        Month = month(REPORT_DATE);
159        Month_char = put(REPORT_DATE,monname.);
160        year = year(REPORT_DATE);
161
162        label Total_Cases='Total Cases'
163           New_Deaths='New Deaths'
164           New_Cases='New Cases'
165           Total_Deaths='Total Deaths'
166           county_name='County'
167           County_FIPS_Number='FIPS'
168           Month='Mo.'
169           Month_char='Month'
170           year='Year';
171        run;

NOTE: Variable Total_Cases is uninitialized.
NOTE: Variable New_Deaths is uninitialized.
NOTE: Variable New_Cases is uninitialized.
NOTE: Variable Total_Deaths is uninitialized.
NOTE: There were 153255 observations read from the data set HWDATA.ALLTX.
NOTE: The data set WORK.TEMP has 153255 observations and 10 variables.
NOTE: DATA statement used (Total process time):
      real time           0.06 seconds
      user cpu time       0.03 seconds
      system cpu time     0.02 seconds
      memory              3568.96k
      OS Memory           33964.00k
      Timestamp           11/23/2021 07:54:27 PM
      Step Count                        83  Switch Count  3
      Page Faults                       0
      Page Reclaims                     608
      Page Swaps                        0
```

```
      Voluntary Context Switches        25
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           28936


172
173
174
175         /* 6.) Reorder the new temporary data set so you can use it for by group processing based on the  */
176         /* county and the Report Date. */
177
178         proc sort data=temp;
179         by COUNTY_NAME REPORT_DATE;
180         run;

NOTE: There were 153255 observations read from the data set WORK.TEMP.
NOTE: The data set WORK.TEMP has 153255 observations and 10 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.06 seconds
      user cpu time       0.03 seconds
      system cpu time     0.04 seconds
      memory              26145.75k
      OS Memory           55908.00k
      Timestamp           11/23/2021 07:54:27 PM
      Step Count                        84  Switch Count  4
      Page Faults                       0
      Page Reclaims                     5748
      Page Swaps                        0
      Voluntary Context Switches        19
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           28936


181
182
183         /* 7.) Use the new temporary data set as the source for a DATA step that will create a summary of the  */
184         /* new cases and new deaths in each county for each month. Since the data span multiple years,  */
185         /* the year variable must also be part of the grouping. Note: Even though year and month were  */
186         /* not explicitly specified when reordering in the previous step, using a date orders by year, month  */
187         /* and day so you can use the year and month variables for the by groups. Store the new data set  */
188         /* in the mylib library. */
189
190         /* (a) Permanently label these new summary variables “Monthly Cases” and “Monthly  */
191         /* Deaths” respectively.  */
192         /* (b) At the end of each month, calculate the “Fatality Rate” for the county by dividing the  */
193         /* Monthly Deaths by the Monthly Cases. Use conditional logic to prevent making a  */
194         /* calculation that would produce a divide by zero message in the log. Apply a permanent  */
195         /* label and the custom format to this variable.  */
196         /* (c) Since new cases and new deaths are daily values, they are not needed in the output  */
197         /* data set. Include the county name and FIPS number. */
198         /* (d) Along with the month number, month name and year number, this will make a total of 8 */
199         /* variables in this data set. This data set should have 5355 rows. */
200
201         data mylib.newdata;
202         set temp;
203         drop New_Deaths New_Cases Total_Cases Total_Deaths Report_Date;
204         by COUNTY_NAME year Month;
205         if first.Month then do;
206         M_CASES=0;
207         M_DEATHS=0;
208         end;
209         M_CASES+NEW_CASES;
210         M_DEATHS+NEW_DEATHS;
211         if last.Month;
212         FATALITY_RATE = divide(M_DEATHS,M_CASES);
213         options fmtsearch=(mylib.formats);
214         format FATALITY_RATE pct.;
215
216         label M_CASES='Monthly Cases'
217            M_DEATHS='Monthly Deaths'
218            FATALITY_RATE='Fatality Rate';
219         run;

NOTE: There were 153255 observations read from the data set WORK.TEMP.
NOTE: The data set MYLIB.NEWDATA has 5355 observations and 8 variables.
```

```
NOTE: DATA statement used (Total process time):
      real time           0.03 seconds
      user cpu time       0.02 seconds
      system cpu time     0.01 seconds
      memory              2648.96k
      OS Memory           32684.00k
      Timestamp           11/23/2021 07:54:27 PM
      Step Count                        85  Switch Count  4
      Page Faults                       0
      Page Reclaims                     388
      Page Swaps                        0
      Voluntary Context Switches        55
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           1032



220
221
222        /* 8.) Create a list of all objects in the mylib library without displaying the descriptor portions of data  */
223        /* sets. Supply an appropriate title.  */
224
225        title "Mylib contents";
226        proc contents data=mylib._ALL_ nods;
227        run;

NOTE: PROCEDURE CONTENTS used (Total process time):
      real time           0.03 seconds
      user cpu time       0.04 seconds
      system cpu time     0.00 seconds
      memory              1790.50k
      OS Memory           31912.00k
      Timestamp           11/23/2021 07:54:27 PM
      Step Count                        86  Switch Count  0
      Page Faults                       0
      Page Reclaims                     238
      Page Swaps                        0
      Voluntary Context Switches        4
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           8



228
229        /* 9.) Report the descriptor portions of the permanent data set created above. Supply an appropriate  */
230        /* title.  */
231
232        title "Monthly Covid Descriptor";
233        proc contents data=mylib.newdata;
234        run;

NOTE: PROCEDURE CONTENTS used (Total process time):
      real time           0.04 seconds
      user cpu time       0.05 seconds
      system cpu time     0.00 seconds
      memory              1504.59k
      OS Memory           32940.00k
      Timestamp           11/23/2021 07:54:27 PM
      Step Count                        87  Switch Count  0
      Page Faults                       0
      Page Reclaims                     199
      Page Swaps                        0
      Voluntary Context Switches        7
      Involuntary Context Switches      0
      Block Input Operations            288
      Block Output Operations           24



235
236        /* 10.) Print the monthly summaries for Brazos and McLennan counties. Exclude the FIPS variable,  */
237        /* month number, and observation numbers from the report. Show labels instead of variable  */
238        /* names. Supply an appropriate title. */
239
240        proc print data=mylib.newdata label noobs;
241        var COUNTY_NAME Month year M_CASES M_DEATHS FATALITY_RATE;
242        where COUNTY_NAME IN ("Brazos","McLennan");
243        title1 "Brazos and McLennan Monthly COVID Information";
```

```
244        run;

NOTE: There were 42 observations read from the data set MYLIB.NEWDATA.
      WHERE COUNTY_NAME in ('Brazos', 'McLennan');
NOTE: PROCEDURE PRINT used (Total process time):
      real time           0.06 seconds
      user cpu time       0.06 seconds
      system cpu time     0.00 seconds
      memory              1562.21k
      OS Memory           33196.00k
      Timestamp           11/23/2021 07:54:27 PM
      Step Count                        88  Switch Count  1
      Page Faults                       0
      Page Reclaims                     232
      Page Swaps                        0
      Voluntary Context Switches        18
      Involuntary Context Switches      0
      Block Input Operations            768
      Block Output Operations           40



245
246
247
248        /* 11.) Print a second report from the monthly summary data except only include rows that are in the  */
249        /* Extreme fatality rate category. Since we know the rate is extreme, override the formatted value  */
250        /* with a temporary format to show the actual percent. Supply an appropriate title. The format of  */
251        /* the report is shown below: (see pdf of assignment instructions for table) */
252
253        title "Extreme Fatality Rate Data";
254        proc print data=mylib.newdata label noobs split='*';
255        where FATALITY_RATE>.15;
256        format FATALITY_RATE percent7.1;
257        var COUNTY_NAME year Month_char M_CASES M_DEATHS FATALITY_RATE;
258        label COUNTY_NAME ='*County'
259          year='*Year'
260          Month_char='*Month'
261          M_CASES='Monthly*Cases'
262          M_DEATHS= 'Monthly*Deaths'
263          FATALITY_RATE = 'Fatality*Rate';
264        run;

NOTE: There were 198 observations read from the data set MYLIB.NEWDATA.
      WHERE FATALITY_RATE>0.15;
NOTE: At least one W.D format was too small for the number to be printed. The decimal may be shifted by the "BEST" format.
NOTE: PROCEDURE PRINT used (Total process time):
      real time           0.27 seconds
      user cpu time       0.27 seconds
      system cpu time     0.00 seconds
      memory              1598.09k
      OS Memory           33452.00k
      Timestamp           11/23/2021 07:54:28 PM
      Step Count                        89  Switch Count  4
      Page Faults                       0
      Page Reclaims                     254
      Page Swaps                        0
      Voluntary Context Switches        16
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           144



265
266
267        /* 12. Close the PDF destination. */
268
269        ods pdf close;
NOTE: ODS PDF printed 11 pages to /home/u59649056/Homeworks/mylib/JRodoni_HW13_Output.pdf.
270
271        /* 13. Use the log and report information contained in your PDF output document to find the answers  */
272        /* to the questions below and include the answers in a comment section at the bottom of your  */
273        /* program file: */
274        /* (a) What length is reported for the permanent custom format you created? */
275
276        /* 9 */
277
278        /* (b) How do the summaries for August 2020 for Brazos County compare to the summaries  */
```

```
279         /* for August 2021? */
280
281         /*  Monthly cases, monthly deaths and fatality rate both increased in 2021 */
282
283
284         /* (c) How does Brazos County compare to McLennan County in January 2021? */
285
286         /* Although they are both fall under the Medium fatality rate group, the actual fatality rate for McLennan */
287         /* is much higher compared to Brazos */
288
289         /* (d) When did Hamilton County have an extreme fatality rate and what was the percentage? */
290
291         /* May 2020 with 33.3% fatality rate.      */
292
293         /* (e) Consider the overall sizes of the Monthy Cases and Monthly Deaths among the Extreme  */
294         /* fatality rates. What observation(s) can be made about the records with an Extreme */
295         /* fatality rate? */
296
297         /* Extreme fatality rate doesnt always seem to be a good indicator of outliers as most of the countys in the  */
298         /* extreme fatality group have a small number of cases. For instance in May 2020 Hamilton county  */
299         /* had a 33.3% fatality rate but they only had 1 death and 3 cases. These small sample sizes can skew  */
300         /* the fatality percentage greatly */
301
302         /* 14.) Save the final version of the program and convert it to a PDF file. Convert the log to PDF. */
303         /* 15.) Upload and submit the three PDF documents to the assignment on Canvas */
304
305
306
307
308
309
310         OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
320
```