```
1          OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
NOTE: ODS statements in the SAS Studio environment may disable some output features.
73
74         ********************************************************************************
75         Anto Lourdu Xavier Raj Arockia Selvarathinam
76         ********************************************************************************;
77
78         /* Import the dataset */
79         FILENAME REFFILE '/home/u63739604/Project.csv';
80
81         /* Import the dataset */
82         PROC IMPORT DATAFILE=REFFILE
83             DBMS=CSV
84             OUT=WORK.IMPORT
85             REPLACE;
85       !              /* Add REPLACE option to overwrite the dataset if it already exists */
86             GUESSINGROWS=MAX;
87             GETNAMES=YES;
88         RUN;

NOTE: Unable to open parameter catalog: SASUSER.PARMS.PARMS.SLIST in update mode. Temporary parameter values will be saved to
WORK.PARMS.PARMS.SLIST.
89          /*************************************************************************
90          *   PRODUCT:   SAS
91          *   VERSION:   9.4
92          *   CREATOR:   External File Interface
93          *   DATE:      23APR24
94          *   DESC:      Generated SAS Datastep Code
95          *   TEMPLATE SOURCE:  (None Specified.)
96          *************************************************************************/
97             data WORK.IMPORT    ;
98             %let _EFIERR_ = 0; /* set the ERROR detection macro variable */
99             infile REFFILE delimiter = ',' MISSOVER DSD  firstobs=2 ;
100               informat Year best32. ;
101               informat LocationAbbr $2. ;
102               informat LocationDesc $39. ;
103               informat GeographicLevel $6. ;
104               informat DataSource $4. ;
105               informat Class $23. ;
106               informat Topic $23. ;
107               informat Data_Value best32. ;
108               informat Data_Value_Unit $24. ;
109               informat Data_Value_Type $55. ;
110               informat Data_Value_Footnote_Symbol $1. ;
111               informat Data_Value_Footnote $17. ;
112               informat StratificationCategory1 $6. ;
113               informat Stratification1 $7. ;
114               informat StratificationCategory2 $14. ;
115               informat Stratification2 $34. ;
116               informat TopicID $2. ;
117               informat LocationID best32. ;
118               informat "Location 1"N $27. ;
119               format Year best12. ;
120               format LocationAbbr $2. ;
121               format LocationDesc $39. ;
122               format GeographicLevel $6. ;
123               format DataSource $4. ;
124               format Class $23. ;
125               format Topic $23. ;
126               format Data_Value best12. ;
127               format Data_Value_Unit $24. ;
128               format Data_Value_Type $55. ;
129               format Data_Value_Footnote_Symbol $1. ;
130               format Data_Value_Footnote $17. ;
131               format StratificationCategory1 $6. ;
132               format Stratification1 $7. ;
133               format StratificationCategory2 $14. ;
134               format Stratification2 $34. ;
135               format TopicID $2. ;
136               format LocationID best12. ;
137               format "Location 1"N $27. ;
138            input
139                        Year
140                        LocationAbbr  $
141                        LocationDesc  $
142                        GeographicLevel  $
143                        DataSource  $
144                        Class  $
145                        Topic  $
146                        Data_Value
147                        Data_Value_Unit  $
```

```
148                         Data_Value_Type  $
149                         Data_Value_Footnote_Symbol  $
150                         Data_Value_Footnote  $
151                         StratificationCategory1  $
152                         Stratification1  $
153                         StratificationCategory2  $
154                         Stratification2  $
155                         TopicID  $
156                         LocationID
157                         "Location 1"N  $
158             ;
159             if _ERROR_ then call symputx('_EFIERR_',1);  /* set ERROR detection macro variable */
160             run;

NOTE: The infile REFFILE is:
      Filename=/home/u63739604/Project.csv,
      Owner Name=u63739604,Group Name=oda,
      Access Permission=-rw-r--r--,
      Last Modified=23Apr2024:12:28:55,
      File Size (bytes)=15129888

NOTE: 59076 records were read from the infile REFFILE.
      The minimum record length was 189.
      The maximum record length was 309.
NOTE: The data set WORK.IMPORT has 59076 observations and 19 variables.
NOTE: DATA statement used (Total process time):
      real time           0.09 seconds
      user cpu time       0.08 seconds
      system cpu time     0.02 seconds
      memory              10923.59k
      OS Memory           43300.00k
      Timestamp           04/23/2024 07:15:50 PM
      Step Count                        364  Switch Count  2
      Page Faults                       0
      Page Reclaims                     329
      Page Swaps                        0
      Voluntary Context Switches        14
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           36368


59076 rows created in WORK.IMPORT from REFFILE.



NOTE: WORK.IMPORT data set was successfully created.
NOTE: The data set WORK.IMPORT has 59076 observations and 19 variables.
NOTE: PROCEDURE IMPORT used (Total process time):
      real time           34.37 seconds
      user cpu time       34.20 seconds
      system cpu time     0.17 seconds
      memory              10923.59k
      OS Memory           43816.00k
      Timestamp           04/23/2024 07:15:50 PM
      Step Count                        364  Switch Count  9
      Page Faults                       0
      Page Reclaims                     5614
      Page Swaps                        0
      Voluntary Context Switches        96
      Involuntary Context Switches      979
      Block Input Operations            0
      Block Output Operations           36440


161
162         /* Remove rows with missing values for Data_Value variable */
163         /* Remove rows where Stratification1 and Stratification 2 is "Overall" */
164         DATA WORK.IMPORT_CLEAN;
165             SET WORK.IMPORT;
166             IF NOT MISSING(Data_Value) AND Stratification2 ne 'Overall' AND Stratification1 ne 'Overall'; /* Exclude rows where
166       ! Data_Value is missing or Stratification2 is 'Overall' or Stratification1 is 'Overall' */
167
168             /* Rename variables */
169             RENAME LocationAbbr = State
170                    Stratification1 = Gender
171                    Stratification2 = Ethnicity
172                    Data_Value = Mortality_Rate;
173
174         RUN;

NOTE: There were 59076 observations read from the data set WORK.IMPORT.
NOTE: The data set WORK.IMPORT_CLEAN has 14087 observations and 19 variables.
```

```
NOTE: DATA statement used (Total process time):
      real time           0.01 seconds
      user cpu time       0.01 seconds
      system cpu time     0.01 seconds
      memory              3796.25k
      OS Memory           39092.00k
      Timestamp           04/23/2024 07:15:50 PM
      Step Count                        365  Switch Count  4
      Page Faults                       0
      Page Reclaims                     503
      Page Swaps                        0
      Voluntary Context Switches        18
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           8712


175
176        /* Use PROC MEANS to calculate descriptive statistics for Mortality_Rate variable */
177        title 'Anto Lourdu Xavier Raj Arockia Selvarathinam';
178        TITLE2 'Descriptive Statistics of Mortality_Rate Variable (Excluding Missing Values)';
179        PROC MEANS DATA=WORK.IMPORT_CLEAN maxdec=1;
180            VAR Mortality_Rate; /* Specify the variable */
181        RUN;

NOTE: There were 14087 observations read from the data set WORK.IMPORT_CLEAN.
NOTE: PROCEDURE MEANS used (Total process time):
      real time           0.04 seconds
      user cpu time       0.03 seconds
      system cpu time     0.01 seconds
      memory              10597.00k
      OS Memory           43476.00k
      Timestamp           04/23/2024 07:15:50 PM
      Step Count                        366  Switch Count  1
      Page Faults                       0
      Page Reclaims                     1839
      Page Swaps                        0
      Voluntary Context Switches        24
      Involuntary Context Switches      1
      Block Input Operations            0
      Block Output Operations           8


182
183        /* Additional output */
184        TITLE 'Checking Missing Values in the Dataset';
185        PROC MEANS DATA=WORK.IMPORT_CLEAN NMISS;
186        RUN;

NOTE: There were 14087 observations read from the data set WORK.IMPORT_CLEAN.
NOTE: PROCEDURE MEANS used (Total process time):
      real time           0.02 seconds
      user cpu time       0.02 seconds
      system cpu time     0.01 seconds
      memory              8121.43k
      OS Memory           43460.00k
      Timestamp           04/23/2024 07:15:50 PM
      Step Count                        367  Switch Count  1
      Page Faults                       0
      Page Reclaims                     1744
      Page Swaps                        0
      Voluntary Context Switches        23
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           8


187
188        TITLE 'Contents of the Dataset';
189        PROC CONTENTS DATA=WORK.IMPORT_CLEAN VARNUM;
190        RUN;

NOTE: PROCEDURE CONTENTS used (Total process time):
      real time           0.05 seconds
      user cpu time       0.05 seconds
      system cpu time     0.00 seconds
      memory              5171.12k
      OS Memory           40116.00k
      Timestamp           04/23/2024 07:15:51 PM
      Step Count                        368  Switch Count  0
      Page Faults                       0
      Page Reclaims                     675
      Page Swaps                        0
```

```
      Voluntary Context Switches        0
      Involuntary Context Switches      1
      Block Input Operations            0
      Block Output Operations           48


191
192       /* Calculate descriptive statistics for Mortality Rate by state */
193       PROC MEANS DATA=WORK.IMPORT_CLEAN NWAY noprint maxdec=1;
194           CLASS State;
195           VAR Mortality_Rate;
196           OUTPUT OUT=State_DescriptiveStats
197                  MEAN=Mean_Mortality_Rate
198                  MEDIAN=Median_Mortality_Rate
199                  STD=Std_Mortality_Rate;
200       RUN;

NOTE: There were 14087 observations read from the data set WORK.IMPORT_CLEAN.
NOTE: The data set WORK.STATE_DESCRIPTIVESTATS has 52 observations and 6 variables.
NOTE: PROCEDURE MEANS used (Total process time):
      real time           0.01 seconds
      user cpu time       0.01 seconds
      system cpu time     0.01 seconds
      memory              11394.35k
      OS Memory           49704.00k
      Timestamp           04/23/2024 07:15:51 PM
      Step Count                        369  Switch Count  3
      Page Faults                       0
      Page Reclaims                     2497
      Page Swaps                        0
      Voluntary Context Switches        39
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           264


201
202       /* Display the aggregated descriptive statistics in a single table */
203       PROC PRINT DATA=State_DescriptiveStats;
204           TITLE 'Aggregated Descriptive Statistics of Mortality Rate by State (Rounded to 1 Decimal Place)';
205           VAR State Mean_Mortality_Rate Median_Mortality_Rate Std_Mortality_Rate;
206           FORMAT Mean_Mortality_Rate Median_Mortality_Rate Std_Mortality_Rate 5.1; /* Format variables to display 1 decimal
206      ! point */
207       RUN;

NOTE: There were 52 observations read from the data set WORK.STATE_DESCRIPTIVESTATS.
NOTE: PROCEDURE PRINT used (Total process time):
      real time           0.06 seconds
      user cpu time       0.07 seconds
      system cpu time     0.00 seconds
      memory              4221.00k
      OS Memory           43184.00k
      Timestamp           04/23/2024 07:15:51 PM
      Step Count                        370  Switch Count  1
      Page Faults                       0
      Page Reclaims                     692
      Page Swaps                        0
      Voluntary Context Switches        11
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           64


208
209       /* Calculate average mortality rates for each ethnicity */
210       PROC MEANS DATA=WORK.IMPORT_CLEAN NOPRINT maxdec=1;
211           CLASS Ethnicity;
212           VAR Mortality_Rate;
213           OUTPUT OUT=Ethnicity_Avg_Mortality
214                  MEAN=Avg_Mortality_Rate;
215       RUN;

NOTE: There were 14087 observations read from the data set WORK.IMPORT_CLEAN.
NOTE: The data set WORK.ETHNICITY_AVG_MORTALITY has 6 observations and 4 variables.
NOTE: PROCEDURE MEANS used (Total process time):
      real time           0.01 seconds
      user cpu time       0.01 seconds
      system cpu time     0.01 seconds
      memory              8804.14k
      OS Memory           50888.00k
      Timestamp           04/23/2024 07:15:51 PM
      Step Count                        371  Switch Count  3
      Page Faults                       0
```

```
      Page Reclaims                    1945
      Page Swaps                       0
      Voluntary Context Switches       33
      Involuntary Context Switches     0
      Block Input Operations           0
      Block Output Operations          264


216
217       /* Use PROC MEANS to calculate summary statistics for Mortality_Rate variable by Gender */
218       TITLE 'Summary Statistics of Heart Disease Mortality Rates by Gender';
219       PROC MEANS DATA=WORK.IMPORT_CLEAN MEAN MEDIAN STD MIN MAX maxdec=1;
220           CLASS Gender;
221           VAR Mortality_Rate;
222       RUN;

NOTE: There were 14087 observations read from the data set WORK.IMPORT_CLEAN.
NOTE: PROCEDURE MEANS used (Total process time):
      real time           0.03 seconds
      user cpu time       0.02 seconds
      system cpu time     0.01 seconds
      memory              10662.85k
      OS Memory           53020.00k
      Timestamp           04/23/2024 07:15:51 PM
      Step Count                       372  Switch Count  1
      Page Faults                      0
      Page Reclaims                    2289
      Page Swaps                       0
      Voluntary Context Switches       29
      Involuntary Context Switches     0
      Block Input Operations           0
      Block Output Operations          8


223
224       /* Use PROC MEANS to calculate summary statistics for Mortality_Rate variable by Ethnicity */
225       TITLE 'Summary Statistics of Heart Disease Mortality Rates by Ethnicity';
226       PROC MEANS DATA=WORK.IMPORT_CLEAN MEAN MEDIAN STD MIN MAX maxdec=1;
227           CLASS Ethnicity;
228           VAR Mortality_Rate;
229       RUN;

NOTE: There were 14087 observations read from the data set WORK.IMPORT_CLEAN.
NOTE: PROCEDURE MEANS used (Total process time):
      real time           0.03 seconds
      user cpu time       0.03 seconds
      system cpu time     0.01 seconds
      memory              9490.67k
      OS Memory           51652.00k
      Timestamp           04/23/2024 07:15:51 PM
      Step Count                       373  Switch Count  1
      Page Faults                      0
      Page Reclaims                    1922
      Page Swaps                       0
      Voluntary Context Switches       36
      Involuntary Context Switches     0
      Block Input Operations           0
      Block Output Operations          24


230
231       ods graphics / reset;
232
233       proc template;
234       define statgraph SASStudio.Pie;
235       begingraph;
236       entrytitle "Average Heart Disease Mortality Rates by Gender" /
237       textattrs=(size=14);
238       layout region;
239       piechart category=Gender response=Avg_Mortality_Rate /;
240       endlayout;
241       endgraph;
242       end;
NOTE: Overwriting existing template/link: SASStudio.Pie
NOTE: STATGRAPH 'SASStudio.Pie' has been saved to: WORK.TEMPLAT
243       run;
NOTE: PROCEDURE TEMPLATE used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              333.34k
      OS Memory           44460.00k
      Timestamp           04/23/2024 07:15:51 PM
```

```
      Step Count                        374  Switch Count  2
      Page Faults                       0
      Page Reclaims                     18
      Page Swaps                        0
      Voluntary Context Switches        7
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           80


244
245        ods graphics / reset width=6.4in height=4.8in imagemap;
246
247        proc sgrender template=SASStudio.Pie data=WORK.AVG_MORTALITY_RATE_BY_GENDER;
248        run;

NOTE: There were 3 observations read from the data set WORK.AVG_MORTALITY_RATE_BY_GENDER.
NOTE: PROCEDURE SGRENDER used (Total process time):
      real time           0.18 seconds
      user cpu time       0.09 seconds
      system cpu time     0.02 seconds
      memory              20079.31k
      OS Memory           61360.00k
      Timestamp           04/23/2024 07:15:51 PM
      Step Count                        375  Switch Count  0
      Page Faults                       0
      Page Reclaims                     5198
      Page Swaps                        0
      Voluntary Context Switches        280
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           896


249
250        ods graphics / reset;
251
252        /* Create bar charts for average heart disease mortality rates by racial/ethnic groups */
253        ods graphics / reset width=6.4in height=4.8in imagemap;
254        PROC SGPLOT DATA=Ethnicity_Avg_Mortality;
255            VBAR Ethnicity / RESPONSE=Avg_Mortality_Rate GROUPORDER=DESCENDING fillattrs=(color=CXa0f5e5);
256        yaxis grid;
257            TITLE 'Average Heart Disease Mortality Rates by Racial/Ethnic Groups';
258            YAXIS LABEL='Average Mortality Rate (per 100,000 population)';
259            LABEL Ethnicity = "Ethnicity"; /* Assign longer labels to Ethnicity variable */
260        RUN;

NOTE: PROCEDURE SGPLOT used (Total process time):
      real time           0.13 seconds
      user cpu time       0.06 seconds
      system cpu time     0.01 seconds
      memory              3266.34k
      OS Memory           63412.00k
      Timestamp           04/23/2024 07:15:51 PM
      Step Count                        376  Switch Count  3
      Page Faults                       0
      Page Reclaims                     940
      Page Swaps                        0
      Voluntary Context Switches        283
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           520

NOTE: There were 6 observations read from the data set WORK.ETHNICITY_AVG_MORTALITY.

261
262        ods graphics / reset;
263
264        /* Calculate average mortality rate by state and ethnicity */
265        PROC MEANS DATA=WORK.IMPORT_CLEAN NWAY NOPRINT maxdec=1;
266            CLASS State Ethnicity;
267            VAR Mortality_Rate;
268            OUTPUT OUT=State_Ethnicity_Avg_Mortality
269                   MEAN=Avg_Mortality_Rate;
270        RUN;

NOTE: There were 14087 observations read from the data set WORK.IMPORT_CLEAN.
NOTE: The data set WORK.STATE_ETHNICITY_AVG_MORTALITY has 233 observations and 5 variables.
NOTE: PROCEDURE MEANS used (Total process time):
      real time           0.01 seconds
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              9043.50k
```

```
      OS Memory           70856.00k
      Timestamp           04/23/2024 07:15:51 PM
      Step Count                        377  Switch Count  3
      Page Faults                       0
      Page Reclaims                     2007
      Page Swaps                        0
      Voluntary Context Switches        43
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           264


271
272       /* Box plots of average heart disease mortality rates by racial/ethnic groups within states */
273       ods graphics / reset width=6.4in height=4.8in imagemap;
274       PROC SGPLOT DATA=State_Ethnicity_Avg_Mortality;
275           VBOX Avg_Mortality_Rate / CATEGORY=Ethnicity fillattrs=(color=CXe6cadf);
276       yaxis grid;
277           TITLE 'Box Plots of Average Heart Disease Mortality Rates by Racial/Ethnic Groups';
278           YAXIS LABEL='Average Mortality Rate (per 100,000 population)';
279       RUN;

NOTE: PROCEDURE SGPLOT used (Total process time):
      real time           0.15 seconds
      user cpu time       0.08 seconds
      system cpu time     0.01 seconds
      memory              3516.68k
      OS Memory           64436.00k
      Timestamp           04/23/2024 07:15:51 PM
      Step Count                        378  Switch Count  2
      Page Faults                       0
      Page Reclaims                     641
      Page Swaps                        0
      Voluntary Context Switches        605
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           672

NOTE: There were 233 observations read from the data set WORK.STATE_ETHNICITY_AVG_MORTALITY.

280
281       ods graphics / reset;
282
283       /* Calculate average mortality rate by Gender */
284       PROC MEANS DATA=WORK.IMPORT_CLEAN MEAN noprint maxdec=1;
285           CLASS Gender;
286           VAR Mortality_Rate;
287           OUTPUT OUT=Avg_Mortality_Rate_By_Gender MEAN=Avg_Mortality_Rate;
288       RUN;

NOTE: There were 14087 observations read from the data set WORK.IMPORT_CLEAN.
NOTE: The data set WORK.AVG_MORTALITY_RATE_BY_GENDER has 3 observations and 4 variables.
NOTE: PROCEDURE MEANS used (Total process time):
      real time           0.01 seconds
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              10480.71k
      OS Memory           72392.00k
      Timestamp           04/23/2024 07:15:51 PM
      Step Count                        379  Switch Count  3
      Page Faults                       0
      Page Reclaims                     2223
      Page Swaps                        0
      Voluntary Context Switches        29
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           264


289
290       /* Create bar chart for average heart disease mortality rates by Gender */
291       TITLE 'Average Heart Disease Mortality Rates by Gender';
292       ods graphics / reset width=6.4in height=4.8in imagemap;
293       PROC SGPLOT DATA=Avg_Mortality_Rate_By_Gender;
294           VBAR Gender / RESPONSE=Avg_Mortality_Rate GROUPORDER=DATA fillattrs=(color=CXf5a0f2);
295       yaxis grid;
296           YAXIS LABEL='Average Mortality Rate';
297           XAXIS LABEL='Gender';
298       RUN;

NOTE: PROCEDURE SGPLOT used (Total process time):
      real time           0.12 seconds
      user cpu time       0.05 seconds
```

```
       system cpu time     0.01 seconds
       memory              3149.59k
       OS Memory           64436.00k
       Timestamp           04/23/2024 07:15:51 PM
       Step Count                        380  Switch Count  2
       Page Faults                       0
       Page Reclaims                     722
       Page Swaps                        0
       Voluntary Context Switches        265
       Involuntary Context Switches      4
       Block Input Operations            0
       Block Output Operations           552

NOTE: There were 3 observations read from the data set WORK.AVG_MORTALITY_RATE_BY_GENDER.

299
300
301
302
303        OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
315
```