```
/*************************************************************/
/* Program Name: Assignment 10 */
/* Date Created: 20 October 2020 */
/* Author: Yash Saraiya */
/* Purpose: Practice using conditional statements and creating data from
SAS data sets */
/*
/*
1. Housekeeping - Clear titles and footnotes
*/
TITLE;
FOOTNOTE;
ODS NOPROCTITLE;
RUN;

/*
2. Libname, input file name and output file name initialization
*/
LIBNAME mylib "/folders/myfolders/STAT 604/Assignment 3/";

DATA covid19;
    SET mylib.covid19;
    FILENAME covid19 "/folders/myfolders/STAT 604/Assignment 3/covid19.sas7bdat";

    /*  FILENAME output "/folders/myfolders/STAT 604/Assignment 3/YSaraiya_HW10_output.pdf"; */
    /*  ODS PDF (ID=output) FILE="/folders/myfolders/STAT 604/Assignment 3/YSaraiya_HW10_output.pdf"; */
RUN;

/*
3. Formation of the three datasets: 1 temp & 2 perm with COVID19 subsetting including debugging using Putlog
*/
DATA temp;
    PUTLOG "NOTE: PDV Before Set Statement";
    PUTLOG _ALL_;
    SET mylib.covid19(OBS=2);
RUN;

DATA temp;
    SET mylib.covid19(OBS=1);
    WHERE PROVINCE_STATE_NAME="Texas" AND PEOPLE_POSITIVE_CASES_COUNT=0;
    DROP PROVINCE_STATE_NAME CONTINENT_NAME;
    DROP COUNTRY_ALPHA_3_CODE COUNTRY_SHORT_NAME COUNTRY_ALPHA_2_CODE;
    FATALITY_GROUP="Low";
    PUTLOG "NOTE: PDV Before Run Statement";
    PUTLOG _ALL_;

DATA temp;
    SET mylib.covid19;
    WHERE PROVINCE_STATE_NAME="Texas" AND PEOPLE_POSITIVE_CASES_COUNT=0;
    DROP PROVINCE_STATE_NAME CONTINENT_NAME;
    DROP COUNTRY_ALPHA_3_CODE COUNTRY_SHORT_NAME COUNTRY_ALPHA_2_CODE;
    FATALITY_GROUP="Low";
RUN;

DATA mylib.permanent1;
    PUTLOG "NOTE: PDV Before Set Statement";
    PUTLOG _ALL_;
    SET mylib.covid19(OBS=2);
RUN;

DATA mylib.permanent1;
    SET mylib.covid19(OBS=1);
    WHERE PROVINCE_STATE_NAME="Texas" AND PEOPLE_POSITIVE_CASES_COUNT ^=0;
    DROP PROVINCE_STATE_NAME CONTINENT_NAME;
    DROP COUNTRY_ALPHA_3_CODE COUNTRY_SHORT_NAME COUNTRY_ALPHA_2_CODE;
    FATAL_PERCENT_CASES=PEOPLE_DEATH_COUNT / PEOPLE_POSITIVE_CASES_COUNT;

    IF FATAL_PERCENT_CASES < 0.01 THEN
        FATALITY_GROUP="Low    ";
    ELSE IF FATAL_PERCENT_CASES < 0.1 THEN
        FATALITY_GROUP="Medium";
```

```
    ELSE IF FATAL_PERCENT_CASES >=0.1 THEN
        FATALITY_GROUP="High  ";
    PUTLOG "NOTE: PDV Before Run Statement";
    PUTLOG _ALL_;
```

```
DATA mylib.permanent1;
    SET mylib.covid19;
    WHERE PROVINCE_STATE_NAME="Texas" AND PEOPLE_POSITIVE_CASES_COUNT ^=0;
    DROP PROVINCE_STATE_NAME CONTINENT_NAME;
    DROP COUNTRY_ALPHA_3_CODE COUNTRY_SHORT_NAME COUNTRY_ALPHA_2_CODE;
    FATAL_PERCENT_CASES=PEOPLE_DEATH_COUNT / PEOPLE_POSITIVE_CASES_COUNT;

    IF FATAL_PERCENT_CASES < 0.01 THEN
        FATALITY_GROUP="Low   ";
    ELSE IF FATAL_PERCENT_CASES < 0.1 THEN
        FATALITY_GROUP="Medium";
    ELSE IF FATAL_PERCENT_CASES >=0.1 THEN
        FATALITY_GROUP="High  ";
RUN;
```

```
DATA mylib.permanent2;
    PUTLOG "NOTE: PDV Before Set Statement";
    PUTLOG _ALL_;
    SET mylib.covid19(OBS=2);
RUN;
```

```
DATA mylib.permanent2;
    SET mylib.covid19(OBS=1);
    WHERE PROVINCE_STATE_NAME="Texas";
    DROP PROVINCE_STATE_NAME CONTINENT_NAME;
    DROP COUNTRY_ALPHA_3_CODE COUNTRY_SHORT_NAME COUNTRY_ALPHA_2_CODE;
    PUTLOG "NOTE: PDV Before Run Statement";
    PUTLOG _ALL_;
```

```
DATA mylib.permanent2;
    SET mylib.covid19;
    WHERE PROVINCE_STATE_NAME="Texas";
    DROP PROVINCE_STATE_NAME CONTINENT_NAME;
    DROP COUNTRY_ALPHA_3_CODE COUNTRY_SHORT_NAME COUNTRY_ALPHA_2_CODE;
RUN;

/*
4. Open a PDF destination to receive your output.
*/
ODS PDF
    FILE="/folders/myfolders/STAT 604/Assignment 3/YSaraiya_HW10_output.pdf"
    STYLE=Styles.Pearl CONTENTS=YES BOOKMARKLIST=HIDE;
RUN;

/*
5. Report a list of data sets in the mylib library without reporting the
descriptor portion of the data sets.
*/
TITLE "Datasets without descriptor";
```

```
PROC CONTENTS DATA=mylib._ALL_ NODS;
RUN;

/*
6. Report the descriptor portion of the temporary data set
created above.
*/
TITLE "Datasets with descriptor";
```

```
PROC CONTENTS DATA=temp;
RUN;

/*
7. Create two Data steps with Brazos Valley data for 31-May and 31-Aug
*/
%LET text="Brazos Valley Covid Data as of 31May2020";
```

```
TITLE &text;

DATA Brazos_May31;
    SET mylib.permanent1;
    WHERE COUNTY_NAME IN ("McLennan", "Falls", "Robertson", "Brazos") AND
        PUT(REPORT_DATE, yymmdd10. -l)='2020-05-31';
RUN;

PROC PRINT DATA=Brazos_May31;
RUN;

%LET text="Brazos Valley Covid Data as of 31August2020";
TITLE &text;

DATA Brazos_Aug31;
    SET mylib.permanent1;
    WHERE COUNTY_NAME IN ("McLennan", "Falls", "Robertson", "Brazos") AND
        PUT(REPORT_DATE, yymmdd10. -l)='2020-08-31';
RUN;

PROC PRINT DATA=Brazos_Aug31;
RUN;

/*
8. Close the PDF output
*/
ODS PDF CLOSE;
ODS HTML PATH="%qsysfunc(pathname(work))";
ODS HTML CLOSE;
RUN;

/*
9.
a. Describe and explain the differences between the three PDVs written to the log.
The Program Data Vector (PDV) holds all variables whether they be user defined, automatic, or temporary.
And it also holds variables from datasets, input sources, or ones that are created during the execution of
a data step.

In the case of temporary, or permanent data sets, PDV holds both the types of variables so that is something
common between the three PDVs, however, it is not necessary that what PDV holds are written to output datasets.

Therefore, when we compare the three PDVs for a single dataset, we see that the first iteration of "Before Set"
does not match the first iteration of "Before Run" as the first PDV does not have the conditional statements
present in it. This means that PDV is subject to conditional statements.

Also, before execution, we can also see that SAS sets all the variables to "missing" (".") in the PDV at the
start of each iteration of Data step. And as a result, SAS replaces this missing value with the value from dataset
and similarly for each iteration after a current value replaces a previous value in the PDV.


b. How many observations were read from the Covid data set that was created in the
previous assignment?
No. of observations in mylib.covid = 783314


c. How many observations were written out to the temporary data set?
No. of observations in work.temp = 20289


d. Which county had the highest people_death_count on May 31? On August 31?
County with highest death on May 31 - Brazon
County with highest death on August 31 - McLennan


e. Which county had the highest percentage of fatal cases on August 31? Which county
had the lowest people_death_count on August 31? How can that be?
County with highest percentage of fatal cases on August 31 - Falls (1.8%)
```