

```
/*
```

```
Name: Yash Saraiya
```

```
Title: Homework 14
```

```
*/
```

```
/*
```

```
1. Housekeeping
```

```
*/
```

```
TITLE;
```

```
FOOTNOTE;
```

```
ODS NOPROCTITLE;
```

```
RUN;
```

```
/*
```

```
2. Assign a libref to the mylib folder containing your permanent data sets. Create a fileref to the pdf file for output. Create a fileref to the Excel file of employment data. Open the PDF destination to receive your output.
```

```
*/
```

```
LIBNAME mylib "/folders/myfolders/STAT 604/Assignment 7/";
```

```
FILENAME empfile "/folders/myfolders/STAT 604/Assignment 7/county-ests-employment-and-wages.xlsx";
```

```
ODS PDF
```

```
FILE="/folders/myfolders/STAT 604/Assignment 7/YSaraiya_HW14_output.pdf";
```

```
RUN;
```

```
/*
```

```
3. Write a single proc step that uses alltx for input and creates an unduplicated list of FIPS numbers and county names order by ascending FIPS number. The list shall contain only those two variables. Use a second proc step to show the first 25 observations of the list on the PDF output. Supply an appropriate title.
```

*/

TITLE "List of Texas counties with their FIPS numbers";

PROC SORT DATA=mylib.alltx OUT=work.texas (KEEP=COUNTY_NAME COUNTY_FIPS_NUMBER)

NODUPKEY;

BY COUNTY_FIPS_NUMBER;

RUN;

PROC PRINT DATA=texas (OBS=25);

RUN;

/*

4. Import the Excel file into a temporary SAS data set. Ensure the program will overwrite the data set if it already exists.

*/

PROC IMPORT DATAFILE=empfile OUT=work.empdata DBMS=xlsx REPLACE;

RUN;

/*

5. Print the descriptor portion of the new data set. Supply an appropriate title.

*/

TITLE "Descriptor portion of Employment Excel spreadsheet dataset";

PROC CONTENTS DATA=empdata;

RUN;

/*

6. Write a single SAS step that will use the imported employment data as input and create a new temporary data set that is suitable for combining with the Covid data in alltx using the FIPS

number as the common value

*/

DATA newempdata;

SET empdata;

DROP month_3_employment_percent_chang month_3_employment_rank

average_weekly_wages_percent_cha average_weekly_wages_rank;

WHERE SUBSTRN(area_title, LENGTH(area_title) - 1, 2)='TX';

establishments=INPUT(quarterly_establishments, 10.);

employment=INPUT(month_3_employment, 10.);

wages=INPUT(average_weekly_wages, 10.);

COUNTY_FIPS_NUMBER=INPUT(area_fips, 5.);

DROP quarterly_establishments month_3_employment average_weekly_wages

area_fips;

period=CATT('YR', year, 'Q', quarter);

DROP year quarter;

RUN;

/*

7. Without using a DATA step, create a new temporary data set that is a “wide” version of the employment data created in the previous step. There will be one row per county. The value for each year/quarter column will be the employment numbers for that time period. The first row from that data set is shown below as a sample.

*/

PROC SORT DATA=newempdata;

BY DESCENDING period;

BY area_title;

RUN;

PROC TRANSPOSE DATA=newempdata OUT=transposed (KEEP=COUNTY_FIPS_NUMBER YR2019Q1

```

        YR2019Q2 YR2019Q3 YR2019Q4 YR2020Q1);

VAR employment;

ID period;

BY COUNTY_FIPS_NUMBER;

RUN;

```

```
/*
```

8. In a single step, create a temporary data set of all rows from March 31, 2020 order by FIPS number, using alltx as input

```
*/
```

```

DATA mar31;

    SET mylib.alltx;

    WHERE PUT(REPORT_DATE, yymmdd10. -l)='2020-03-31';

RUN;

```

```

PROC SORT DATA=mar31;

    BY COUNTY_FIPS_NUMBER;

RUN;

```

```
/*
```

9. Use a single DATA step to combine the columns from the wide data set created above with the March 31 data set and create two new temporary data sets. Use the IN= option to control processing in this step. Variable names should be created so that they can be used in variable lists. Use variable lists in the code as much as possible.

```
*/
```

```

DATA match nomatch;

    MERGE mar31(IN=mar) transposed(IN=trans);

    BY COUNTY_FIPS_NUMBER;

```

```

IF mar=1 AND trans=1 THEN
    DO;

        ARRAY quarters{5} YR2019Q1--YR2020Q1;

        num_quarters=DIM(quarters);

        ARRAY change{4};

        /* Cannot use num_quarters-1 */
        DO i=2 TO num_quarters;
            change{i-1}=quarters{i} - quarters{i-1};
        END;

        ARRAY changeratio{4};

        /* Cannot use num_quarters-1 */
        DO i=2 TO num_quarters;
            changeratio{i-1}=change{i-1}/PEOPLE_POSITIVE_CASES_COUNT;
        END;

        OUTPUT match;

    END;
ELSE IF mar=1 AND trans=0 THEN
    DO;

        OUTPUT nomatch;

    END;

    DROP num_quarters i;
RUN;

DATA nomatch;

    SET nomatch;

    KEEP PEOPLE_POSITIVE_CASES_COUNT--Fatality_Group;

RUN;

```

```
/*
```

10. Print a sample of the data set of counties with no employment data created in the previous step. The sample will be a series of 25 observations ending with observation 100. Include all columns. Supply an appropriate title.

```
*/
```

```
TITLE "Counties with no employment data found - observations 76 to 100";
```

```
PROC PRINT DATA=nomatch (FIRSTOBS=76 OBS=100) LABEL;
```

```
    LABEL
```

```
        PEOPLE_POSITIVE_CASES_COUNT="Total Cases"
```

```
        COUNTY_NAME="County"
```

```
        REPORT_DATE="Report Date"
```

```
        PEOPLE_DEATH_NEW_COUNT="New Deaths"
```

```
        COUNTY_FIPS_NUMBER="FIPS No."
```

```
        PEOPLE_POSITIVE_NEW_CASES_COUNT="New Cases"
```

```
        PEOPLE_DEATH_COUNT="Total Deaths"
```

```
        pct_fatal_cases="Fatality Rate"
```

```
        Fatality_Group="Fatality Group";
```

```
RUN;
```

```
/*
```

11. Print all rows and selected columns from the county summary data set as shown in the sample below. Use temporary labels as needed. Supply an appropriate first title. The second title must read "Employment Data Represented in Thousands".

```
*/
```

```
TITLE "Impact of COVID-19 on Employment figures in Texas counties";
```

```
TITLE3 "Employment Data Represented in Thousands";
```

```

PROC PRINT DATA=work.match LABEL NOOBS;

    LABEL COUNTY_NAME='County';

    LABEL PEOPLE_POSITIVE_CASES_COUNT='Cases';

    LABEL YR2020Q1='Employment';

    LABEL change4='Change';

    LABEL changeratio4='Change per Case';

    FORMAT changeratio4 5.;

    VAR COUNTY_NAME PEOPLE_POSITIVE_CASES_COUNT YR2020Q1 change4 changeratio4;

RUN;

```

```
/*
```

12. Use the means procedure to create a report of the number of new cases per month for Brazos and McLennan counties using the alltx data set as input. Include the mean and median number of new cases per county for each month. Apply a temporary format to the report date to cause the data to be grouped by month name. Research SAS Help documentation as needed to determine which format to use. Supply temporary labels as needed. The first row of the report is shown below to demonstrate the output you must produce. Supply an appropriate title.

```
*/
```

```
TITLE "Monthwise analysis of COVID-19 cases in Brazos and McLennan counties";
```

```
PROC FORMAT LIBRARY=work;
```

```

    VALUE $mon_name 01='January' 02='February' 03='March' 04='April' 05='May'
                    06='June' 07='July' 08='August' 09='September' 10='October' 11='November'
                    12='December';

```

```
RUN;
```

```
DATA tx;
```

```

    SET mylib.alltx;

    mon_num=MONTH(REPORT_DATE);

```

```

        month_name=PUT(mon_num, z2.);
        FORMAT month_name $mon_name.;
        DROP mon_num;

RUN;

PROC MEANS DATA=tx SUM MEAN MEDIAN NONOBS;

    CLASS month_name COUNTY_NAME;

    VAR PEOPLE_POSITIVE_CASES_COUNT;

    WHERE COUNTY_NAME IN ('Brazos' 'McLennan');

    LABEL month_name="Month" COUNTY_NAME="County";

RUN;

```

```
/*
```

13. Use the FREQ procedure and the alltx data set to show the number of times each county has had a total number of cases above 0. The output is ordered from the most days to the least. Supply an appropriate title.

```
*/
```

```
TITLE "Texas counties & number of days of having an active COVID-19 case (Descending)";
```

```

PROC FREQ DATA=tx ORDER=FREQ;

    TABLES COUNTY_NAME /NOPERCENT NOCUM;

    WHERE PEOPLE_POSITIVE_CASES_COUNT > 0;

RUN;

```

```
/*
```

14. Close the PDF destination

```
*/
```

```

ODS PDF CLOSE;

ODS HTML PATH="%qsysfunc(pathname(work))";

```


ODS HTML CLOSE;

RUN;

/*

15. Questions

a. How many observations and variables were read in from the Excel file?

22818

b. What is the name of the 100th county with no Employment data?

Hudspeth

c. Which counties posted increases in employment during the first quarter of 2020?

Midland, Montgomery, and Smith

d. Which county had the most negative ratio of change to Covid cases and what is the ratio?

Potter;

Ratio = -280

e. Which county had reported the fewest days of Covid cases? How many days?

Sterling - 21 days

*/