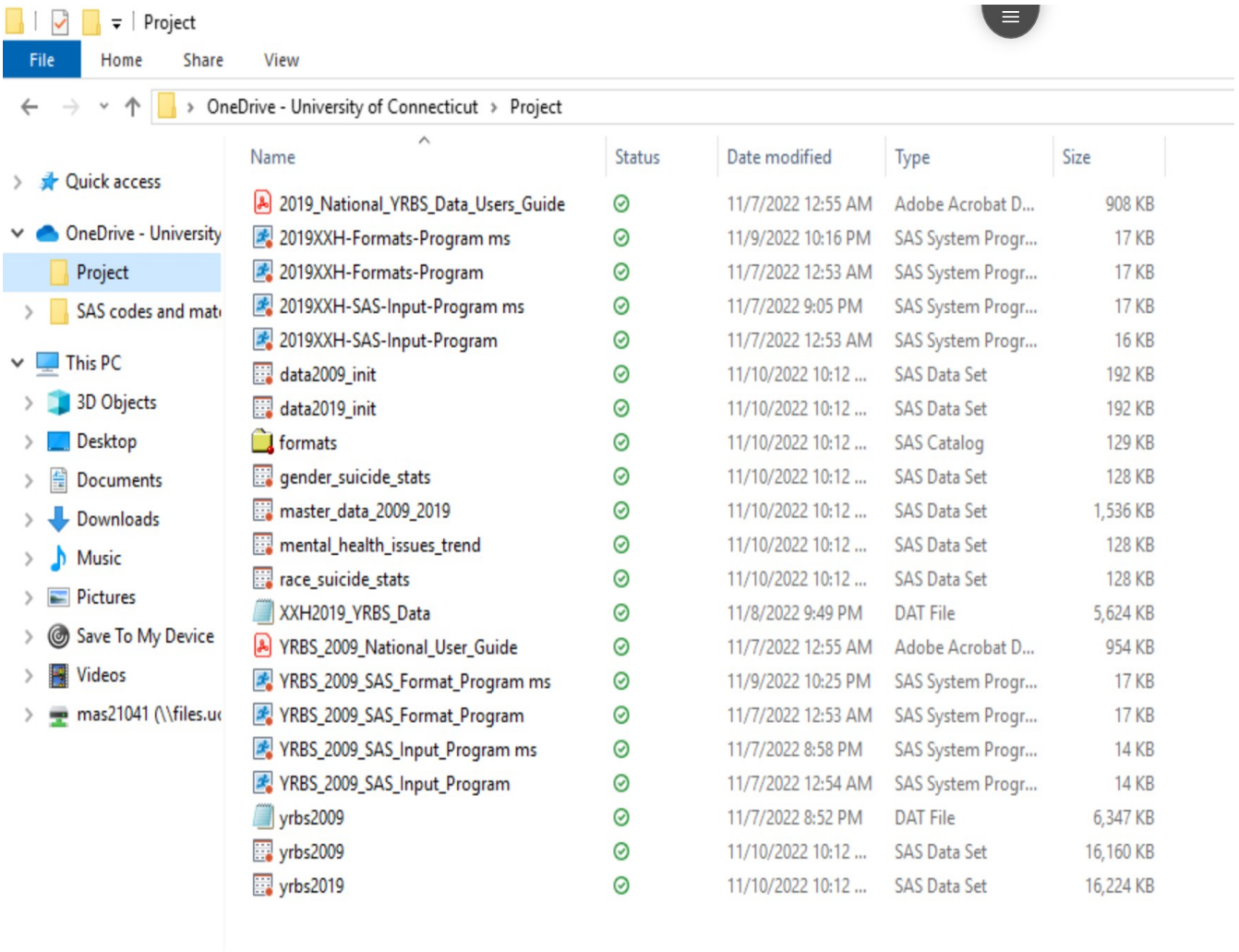
1. Data source(s), i.e., the uniform resource location (URL) of the Public Health Data Repository website you have chosen to download 2 data files with documentation for this project.

I have selected CDC website https://www.cdc.gov/healthyyouth/data/yrbs/data.htm to get data files for YRBSS data for years 2009 and 2019. Please see below the screenshot of the downloaded data sources along with the codes to read the data files and formats to be applied.

Screenshot of the downloaded data sources in UConn any ware :-



2. Describe your selected data files with the help of meaningful SAS outputs.

The datafiles which I have selected is 2009 and 2019 YRBSS datasets which contain responses to surveys provided by high school students from different U.S. schools. The two data files contain information regarding the risky behavior of the youth taken in 2009 & 2019 through variety of questions like Suicide, Weapon Carrying, Fighting, Tobacco Use, Alcohol Use, Marijuana Use, Sexual Behaviors, Sex and Alcohol or Drugs. For the purpose of this project, I have selected Suicide attempts and related mental health questions along with race and gender of the participants. Samples from both the datasets are attached below -

National YRBSS Datasets and Documentation

Sample 5 rows of 2009 YRBSS Dataset

Obs	gender	feel_sad	suicidal_thoughts	Suicide_plans	attempted_suicide	race
1	Male	No	No	No	0 times	White
2	Male	Yes	Yes	Yes	0 times	White
3	Male	Yes	No	No	0 times	White
4	Male	No	No	No	0 times	White
5	Male	No	No	No	Missing	Multiple - Hispanic

National YRBSS Datasets and Documentation
Sample 5 rows of 2019 YRBS dataset

Obs	gender	feel_sad	suicidal_thoughts	Suicide_plans	attempted_suicide	race
1	Male	No	No	No	Yes	Multiple - Hispanic
2	Male	No	No	No	Yes	Multiple - Non-Hispanic
3	Female	No	No	No	Yes	Multiple - Non-Hispanic
4	Male	Yes	No	No	Yes	White
5	Male	No	No	No	Yes	Hispanic/ Latino

3. Explain how these data files are related with each other and create a new SAS dataset by combining them in a useful way.

   Both the data files are from YRBSS and represent the responses from youths. They just represent different years. Questions and answer options for both the surveys and responses are almost similar. I am combining them vertically so that I have a master dataset containing all the responses by year. I am utilizing this master dataset to compare the Suicide rates between 2009 and 2019 by race, gender along with comparison for few other related mental health issues.

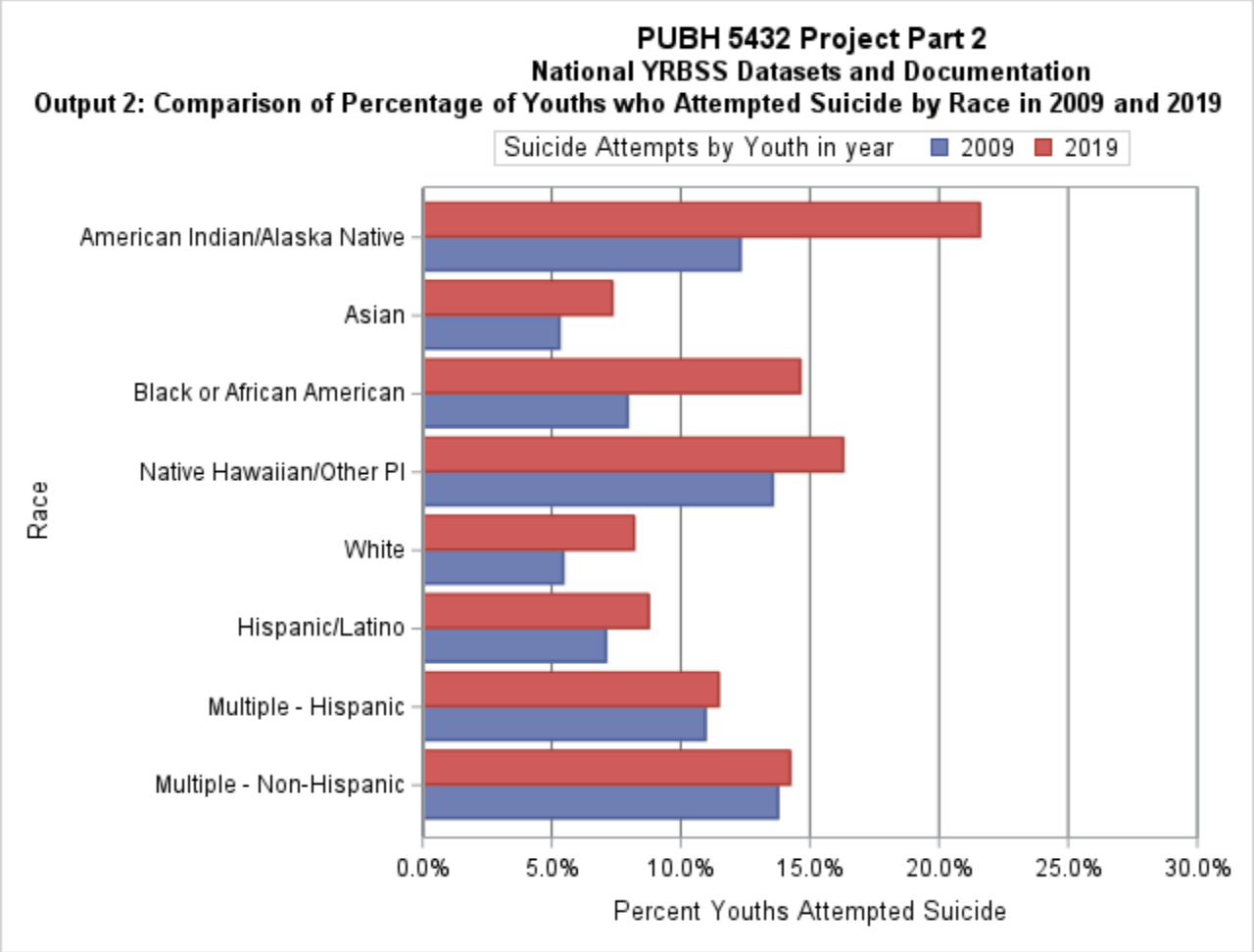
4. Create meaningful SAS outputs using the combined new SAS dataset.

SAS Output 1 – Numerical

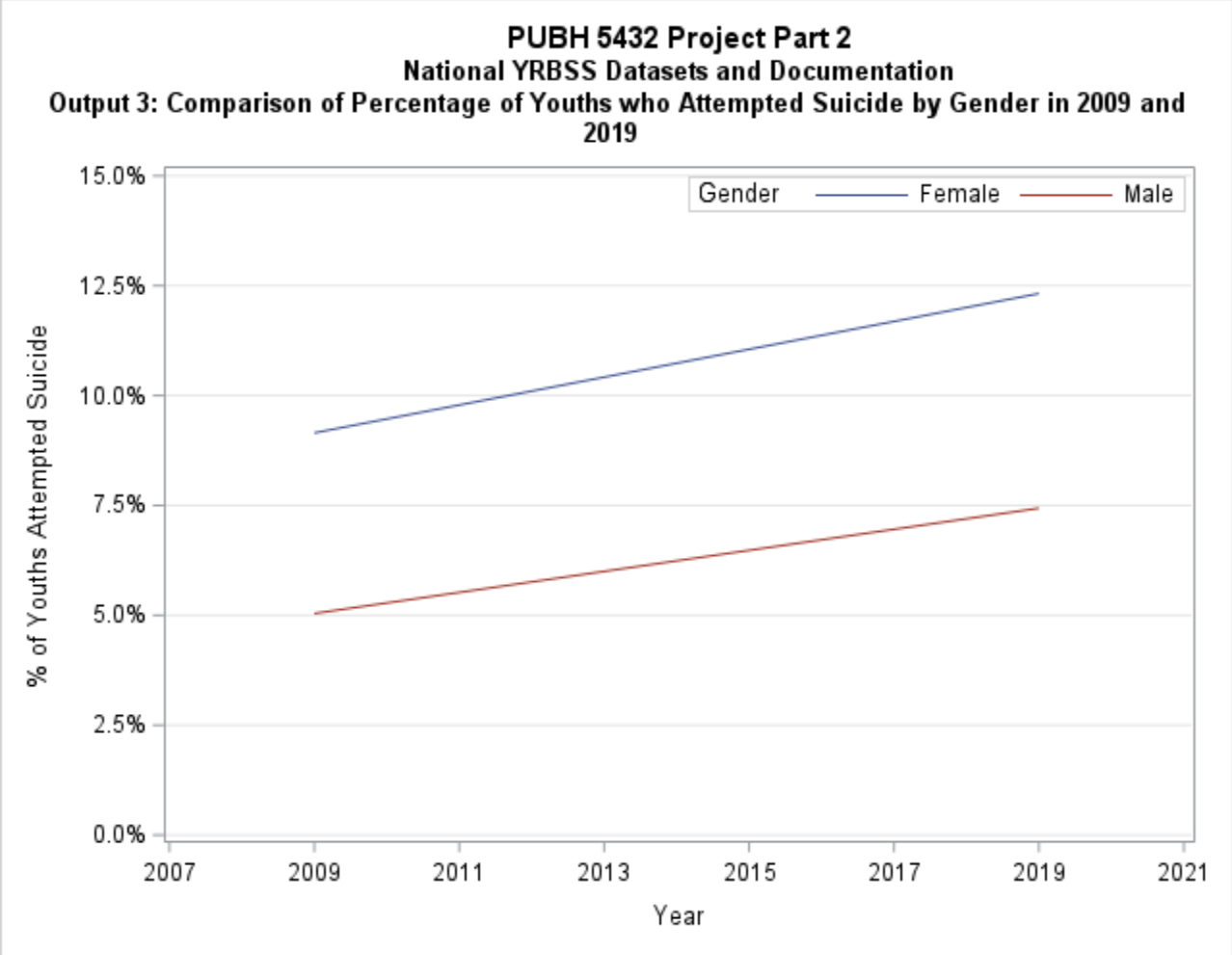
Output 1: Year Over Year Key Mental Health Issues Statistics Among Youths

PERCENTAGE OF YOUTH STUDENTS	2009	2019
Experienced Persistent Feeling of Sadness or Hopelessness	27.9%	36.7%
Serious Considered Attempting Suicide	14.5%	19.6%
Made Suicide Plan	11.6%	16.0%
Attempted Suicide	7.2%	10.1%

SAS Output 2 – Graphical



SAS Output 3- Graphical



5. SAS code you have used to -

   a. Enable access of these data files in SAS,

```
/* File name: pubh5432_Project Part 2.sas

   PUBH 5432 Project Part 2

   Created on November 1, 2022
   Last modified on November 9, 2022
*/

* SAS statements ;
title "PUBH 5432 Project Part 2";
title2 "National YRBSS Datasets and Documentation";


* Set up user defined [permanent] library path ;
%let project = C:\Users\mas21041\OneDrive - University of Connecticut\Project;
libname project "&project";



/*Adding SAS code to get the format for 2009 Dataset*/
%include "&project\YRBS_2009_SAS_Format_Program ms.sas";

/*Adding SAS code to read the 2009 YRBSS dat file and create a permanent SAS
dataset yrbs2009 in the project library*/
%include "&project\YRBS_2009_SAS_Input_Program ms.sas";

/* Selecting and renaming only the columns required for this project from
2009 YRBSS Data */
data project.data2009_init;
set project.yrbs2009;
  keep   q2 raceeth q23 q24 q25 q26;
  format q2 $H2S. raceeth $HRCE. q23 $H23S. q24 $H24S. q25 $H25S. q26 $H26S.;
  rename q2 = gender;
  rename raceeth = race;
  rename q23 = feel_sad;
  rename q24 = suicidal_thoughts;
  rename q25 = Suicide_plans;
  rename q26 = attempted_suicide;
 run;

 /*Printing Sample few rows of sourced 2009 YRBSS dataset*/
 title3 "Sample 5 rows of 2009 YRBSS Dataset";
proc print data=project.data2009_init (obs=5) ;
run;

/*Adding SAS code to get the format for 2019 Dataset*/
%include "&project\2019XXH-Formats-Program ms.sas";
```

```
/*Adding SAS code to read the 2019 YRBS dat file and create a permanent SAS
dataset yrbs2019 in the project library*/
%include "&project\2019XXH-SAS-Input-Program ms.sas";

/* Selecting and renaming only the columns required for this project from
2019 YRBSS Data */

data project.data2019_init; set project.yrbs2019;
  keep   q2 raceeth q25 q26 q27 q28;
  format q2 $H2S. raceeth $HRCE. q25 $H23S. q26 $H24S. q27 $H25S. q28 $H26S.;
  rename q25 = feel_sad;
  rename q26 = suicidal_thoughts;
  rename q27 = Suicide_plans;
  rename q28 = attempted_suicide;
  rename q2 = gender;
  rename raceeth = race;
run;
 /*Printing Sample few rows of sourced 2009 YRBSS dataset*/
 title3 "Sample 5 rows of 2019 YRBS dataset";
proc print data=project.data2019_init (obs=5) ;
run;
```

b. Combine multiple SAS datasets into one new SAS dataset –

```
/*Creating a permanent dataset by vertically combining the two datasets*/
/*Adding a column named Year to specify which year the record belongs to*/
/*Creating Indicators for Mental Health Issues so that it becomes easy to
calculate the mean*/
title3 "Vertical Combination of Datasets";

data project.master_data_2009_2019;
set project.data2009_init (in =inyear2009)
    project.data2019_init (in =inyear2019);

if inyear2009=1 then year=2009;
else year=2019;

if feel_sad='1' then feel_sad_ind=1;
else if feel_sad='2' then feel_sad_ind =0;


if suicidal_thoughts='1' then suicidal_thoughts_ind=1;
else if suicidal_thoughts='2' then suicidal_thoughts_ind =0;

if Suicide_plans='1' then Suicide_plans_ind=1;
else if Suicide_plans='2' then Suicide_plans_ind =0;

if attempted_suicide in ('2','3','4','5') then attempted_suicide_ind=1;
else if attempted_suicide='1' then attempted_suicide_ind =0;


drop inyear2009 inyear2019;
run;
```

c. Produce the included numerical and graphical outputs.

```
 *output 1 percentage of mental health among youth in 2009 and 2019;
 title3 "Output 1: Year Over Year Key Mental Health Issues Statistics Among
Youths";

/* For each year finding the mean values of each Mental Health Issue related
to Suicide*/
proc means data= project.master_data_2009_2019 mean nway maxdec=3 noprint;
class year;
var feel_sad_ind suicidal_thoughts_ind Suicide_plans_ind
attempted_suicide_ind;
output out =means_mental_health (drop= _TYPE_ _FREQ_) mean=;

run;

/*Transposing the output from means data to switch Year field from Rows to
Columns*/
proc transpose data=means_mental_health out=transposed_init prefix=YEAR_
NAME=METRIC;
id year;
var feel_sad_ind suicidal_thoughts_ind Suicide_plans_ind
attempted_suicide_ind;
run;

/*Applying Proper label and formatting to have better display*/
data project.Mental_health_issues_trend ;
set transposed_init;
LENGTH METRIC_FORMATTED $100.;
if METRIC = 'feel_sad_ind' then METRIC_FORMATTED='Experienced Persistent
Feeling of Sadness or Hopelessness';
else if METRIC = 'suicidal_thoughts_ind' then METRIC_FORMATTED='Serious
Considered Attempting Suicide';
else if METRIC = 'Suicide_plans_ind' then METRIC_FORMATTED='Made Suicide
Plan';
else if METRIC = 'attempted_suicide_ind' then METRIC_FORMATTED='Attempted
Suicide';
format year_2009 year_2019 percent7.1;

drop METRIC;
label year_2019 = '2019';
label year_2009 = '2009';
label METRIC_FORMATTED='PERCENTAGE OF YOUTH STUDENTS';
rename METRIC_FORMATTED=METRIC;

run;

/*Printing Table containing Key Mental Health Statistics*/
proc print data=project.Mental_health_issues_trend noobs LABEL;
          var METRIC year_2009 year_2019;
run;

/*Calculating the perecntage of youth attempted suicide in different races in
2009 and 2019;*/
title3 "Output 2: Comparison of Percentage of Youths who Attempted Suicide by
Race in 2009 and 2019";

proc means data= project.master_data_2009_2019 mean nway maxdec=3 noprint;
              by year;
```

```
                class race;
                var attempted_suicide_ind;
                output out=project.race_suicide_stats (drop= _TYPE_ _FREQ_)
mean=;
run;

/*Creating Bar Graph to show the change in Percentage of Youths who attempted
Suicide by Race in 2009 and 2019*/
proc sgplot data= project.race_suicide_stats;

            hbar Race/ response=attempted_suicide_ind group=year

            groupdisplay=cluster;

            xaxis label='Percent Youths Attempted Suicide' grid
gridattrs=(color=gray66)

            values=(0 to .30 by .05) offsetmax=0 valuesformat=percent7.1;

            yaxis label='Race';

            keylegend / position=top title='Suicide Attempts by Youth in
year';

run;

/*Calculating the perecntage of youth attempted suicide by gender in 2009 and
2019;*/
title3 "Output 3: Comparison of Percentage of Youths who Attempted Suicide by
Gender in 2009 and 2019";


proc means data= project.master_data_2009_2019 mean nway maxdec=3 noprint;
                by year;
                class gender;
                var attempted_suicide_ind;
                output out=project.gender_suicide_stats (drop= _TYPE_ _FREQ_)
mean=;
run;

/*Creating Line Graph to show the change in Percentage of Youths who attempted
Suicide by Gender in 2009 and 2019*/
proc sgplot data= project.gender_suicide_stats ;

            series y=attempted_suicide_ind x=year/group=gender ;

            xaxis label="Year" values=(2007 to 2021 by 2) ;

            yaxis label="% of Youths Attempted Suicide " values=(0 to 0.15 by
0.025) valuesformat=percent7.1 grid;

            keylegend /TITLE='Gender'  position=topright location=inside ;

run;
```