

```

/*FINAL PROJECT*/
libname ds 'U:\SAS\datasets\';
x 'cd U:\SAS\extdata\';

/*#####IMPORT THE DATA#####*/

/*DATASET 1: FREEDOM DATASET*/
/*import the freedom dataset*/
proc import
    out=freedom
    datafile= 'human_freedom.csv'
    dbms=dlm replace;
    delimiter=', ';
    getnames=YES;
    guessingrows=20404042;
run;

/*get 2016 data only*/
data freedom;
    set freedom;
    where year = 2016;
run;

/*DATASET 2: HAPPINESS DATASET*/
/*import the happiness dataset*/
proc import
    out=happy
    datafile= '2016_happy.csv'
    dbms=dlm replace;
    delimiter=', ';
    getnames=YES;
    guessingrows=20404042;
run;

/*DATASET 3: ECONOMIC DATASET*/
/*import the happiness dataset*/
proc import
    out=economic
    datafile= 'Economic_Freedom.csv'
    dbms=dlm replace;
    delimiter=', ';
    getnames=YES;
    guessingrows=20404042;
run;

/*get 2016 only*/
data economic;
    set economic;
    where year = 2016;
run;

/*MERGE THE DATASETS TO GET ONE DATASET */
/*to merge happiness and economic*/

```

```

proc sql;
    create table merged as
    select * from happy a left join economic b
    on a.Country = b.Countries;
quit;

/*to merge merged dataset with freedom*/
/*drop repeat columns first*/
data freedom;
    set freedom;
    drop year countries region;
run;

proc sql;
    create table data as
    select * from merged a left join freedom b
    on a.ISO_code=b.ISO_code;
quit;

/*#####REGRESSION & CORRELATION ANALYSIS#####*/

/*### Regression ###*/
proc iml;
    reset log print;

    use data;
    read all var{Happiness_Score pf_expression pf_expression_jailed
pf_expression_influence pf_expression_internet pf_expression_control}
into expression;
    read all var{Happiness_Score pf_religion_estop
pf_religion_harassment pf_religion_restrictions pf_religion} into
religion;
    read all var{Happiness_Score ef_legal_integrity ef_legal_courts
ef_legal_police} into legal;

    close data;

    /*creates a module that performs ordinary least squares*/
    start ols(X, Y, b, yhat, rmse);
        design = j(nrow(X),1,1) || X;
        b = inv(design`*design)*design`*Y;
        yhat = design*b;
        e = Y - yhat;
        rmse = sqrt(e[##] / (nrow(design) - ncol(design)));
    finish;

    /*#####
#####*/
    /*EXPRESSION & INFORMATION*/

    /*create a table comprised of expression variables*/

```

```

        create expression_data from expression[colname = {Happiness_Score
pf_expression pf_expression_jailed pf_expression_influence
pf_expression_internet pf_expression_control}]];
        append from expression;
        close expression_data;

        *remove missing values across all expression variables;
        submit;
            data nmiss_expr;
            set expression_data;
            if nmiss(of _numeric_) > 0 then delete;
        run;
        endsubmit;

        use nmiss_expr;
        read all var{Happiness_Score pf_expression pf_expression_jailed
pf_expression_influence pf_expression_internet pf_expression_control}
into expr;
        close nmiss_expr;

        X_expr = expr[,2:6];
        Y_expr = expr[,1];

        run ols(X_expr, Y_expr, b, yhat, rmse);
        print b[format=6.2] rmse[format=6.2] yhat[format=6.2];

        tot_expr_data = Y_expr || X_expr;
        create ols_expr from tot_expr_data[colname = {Happiness_Score
pf_expression pf_expression_jailed pf_expression_influence
pf_expression_internet pf_expression_control}]];
        append from tot_expr_data;
        close ols_expr;

        submit;
            proc reg data = ols_expr outest = expr_est;
            model Happiness_Score = pf_expression pf_expression_jailed
pf_expression_influence pf_expression_internet pf_expression_control;
            title 'Regression: Happiness vs. Freedom of Expression';
            run;
        endsubmit;

        /*#####*/
        /*RELIGION*/

        /*create a table comprised of expression variables*/
        create religion_data from religion[colname = {Happiness_Score
pf_religion_estop pf_religion_harassment pf_religion}]];
        append from religion;
        close religion_data;

        *remove missing values across all expression variables;

```

```

submit;
    data nmiss_reli;
    set religion_data;
    if nmiss(of _numeric_) > 0 then delete;
run;
endsubmit;

use nmiss_reli;
read all var{Happiness_Score pf_religion_estop
pf_religion_harassment pf_religion} into reli;
close nmiss_reli;

X_reli = reli[,2:4];
Y_reli = reli[,1];

call ols(X_reli, Y_reli, b, yhat, rmse);
print b[format=6.2] rmse[format=6.2] yhat[format=6.2];

tot_reli_data = Y_reli || X_reli;
create ols_reli from tot_reli_data[colname = {Happiness_Score
pf_religion_estop pf_religion_harassment pf_religion}];
append from tot_reli_data;
close ols_reli;

submit;
    proc reg data = ols_reli outest = reli_est;
    model Happiness_Score = pf_religion_estop
pf_religion_harassment pf_religion;
    title 'Regression: Happiness vs. Religious Freedom';
    run;
endsubmit;

/*#####
#####*/
/*LEGAL SYSTEM*/

/*create a table comprised of expression variables*/
create legal_data from legal[colname = {Happiness_Score
ef_legal_integrity ef_legal_courts}];
append from legal;
close legal_data;

*remove missing values across all expression variables;
submit;
    data nmiss_legal;
    set legal_data;
    if nmiss(of _numeric_) > 0 then delete;
run;
endsubmit;

use nmiss_legal;

```

```

        read all var{Happiness_Score ef_legal_integrity ef_legal_courts}
into leg_sys;
        close nmiss_legal;

        X_legal = leg_sys[,2:3];
        Y_legal = leg_sys[,1];

        run ols(X_legal, Y_legal, b, yhat, rmse);
        print b[format=6.2] rmse[format=6.2] yhat[format=6.2];

        tot_legal_data = Y_legal || X_legal;
        create ols_legal from tot_legal_data[colname = {Happiness_Score
ef_legal_integrity ef_legal_courts}];
        append from tot_legal_data;
        close ols_legal;

        submit;
            ods graphics on;
            proc reg data = ols_legal outest = legal_est;
            model Happiness_Score = ef_legal_integrity ef_legal_courts;
            title 'Regression: Happiness vs. Legal Systems';
            run;
            ods graphics off;
        endsubmit;

```

```
quit;
```

```
/*### Correlation ###*/
```

```
/*Religious Freedom*/
```

```

ods graphics on;
title 'Happiness vs. Religious Freedom Correlation';
proc corr data=ols_reli nomiss plots=matrix(histogram);
    var Happiness_Score pf_religion_estop pf_religion_harassment
pf_religion;
run;
ods graphics off;

```

```
/*Legal System*/
```

```

ods graphics on;
title 'Happiness vs. Justice System Correlation';
proc corr data=ols_legal nomiss plots=matrix(histogram);
    var Happiness_Score ef_legal_integrity ef_legal_courts;
run;
ods graphics off;

```

```
/*Additional Analysis*/
```

```
/*remove missing values*/
```

```

data estop;
set data;
if nmiss(of pf_religion_estop) > 0 then delete;

```

```
run;
```

```
/*How does the average religion estop measurement vary across regions?*/  
proc sgplot data=estop;  
vbar Region /response=pf_religion_estop stat=mean categoryorder=RESPDESC  
fillattrs=graphdata3;  
label pf_religion_estop="Average Religion Estop Measurement";  
title "Average Religion Estop across Global Regions (2016)";  
run;
```

```
/*#####*/
```

```
/*DATASET 1: GDP DATASET*/  
/*import the gdp dataset using PROC IMPORT*/  
proc import  
    out=gdp  
    datafile= 'gdp.csv'  
    dbms=dlm replace;  
    delimiter=',';  
    getnames=YES;  
    guessingrows=20404042;  
run;
```

```
/*DATASET 2: SUICIDE DATASET*/  
/*import the suicide dataset using PROC IMPORT*/  
proc import  
    out=suicide  
    datafile= 'suicide.csv'  
    dbms=dlm replace;  
    delimiter=',';  
    getnames=YES;  
    guessingrows=20404042;  
run;
```

```
/*DATASET 3: HAPPINESS DATASET*/  
/*import the happiness dataset using PROC IMPORT*/  
proc import  
    out=happy  
    datafile= '2016_happy.csv'  
    dbms=dlm replace;  
    delimiter=',';  
    getnames=YES;  
    guessingrows=20404042;  
run;
```

```
/*DATASET 4: FREEDOM DATASET*/  
/*import the happiness dataset using PROC IMPORT*/  
proc import  
    out=freedom  
    datafile= 'human_freedom.csv'  
    dbms=dlm replace;
```

```

        delimiter=', ';
        getnames=YES;
        guessingrows=20404042;
run;

/*MERGE THE DATASETS TO GET HAPPINESS AND GDP AND SUICIDES */
/*merge happiness and economic GDP datasets together using a left join
with proc sql*/
proc sql;
    create table merged_part1 as
    select * from happy a left join gdp b
    on a.Country = b.Country;
quit;
/*merge the merged and suicide datasets together using a left join with
proc sql*/
proc sql;
    create table merged as
    select * from merged_part1 a left join suicide b
    on a.Country = b.Country;
quit;

/*filter the freedom dataset for only 2016 data before merging*/
proc sql;
create table freedom2 as
select *
from freedom
where year = 2016;
quit;

/*drop repeat region column first*/
data freedom2;
    set freedom2;
    drop region;
run;

/*merge the merged and suicide datasets together using a left join with
proc sql*/
proc sql;
    create table data as
    select * from merged a left join freedom2 b
    on a.Country=b.Countries;
quit;

/*#####MAKE MAPS#####*/

/*get country names from shape file*/
data maps_to_merge;
set mapsgfk.world;
Keep=1;
drop SEGMENT LONG LAT X Y ISO ISOALPHA2 RESOLUTION DENSITY CONT LAKE;
run;

/*sort by country ID*/

```

```

proc sort data=maps_to_merge out=maps_to_merge2 NODUPKEY;
by ID;
run;

/*use proc sql to merge shape file with happiness scores*/
proc sql;
    create table happiness_data_for_map as
    select * from maps_to_merge2 a left join happy b
    on a.IDNAME = b.Country;
quit;

/*make map of happiness scores*/
proc gmap data=happiness_data_for_map map=mapsgfk.world;
id ID;
choro Happiness_Score / coutline=black;
label Happiness_Score='Happiness Score';
title 'Happiness Score by Country';
run;

/*2-suicide map*/

/*use proc sql to merge shape file with happiness scores*/
proc sql;
    create table suicide_data_for_map as
    select * from maps_to_merge2 a left join suicide b
    on a.IDNAME = b.Country;
quit;

/*make map of happiness scores*/
proc gmap data=suicide_data_for_map map=mapsgfk.world;
id ID;
choro _2016_suicide / coutline=black;
label _2016_suicide='Crude Suicide Rate per 100,000 Residents';
title 'Crude Suicide Rate by Region';
run;

/*3-freedom of expression*/

/*use proc sql to merge shape file with happiness scores*/
proc sql;
    create table freedom_data_for_map as
    select * from maps_to_merge2 a left join freedom2 b
    on a.IDNAME = b.countries;
quit;

proc gmap data=freedom_data_for_map map=mapsgfk.world;
id ID;
choro ef_legal_integrity / coutline=black;
label ef_legal_integrity='Legal Integrity Score (1-10)';
title 'Legal Integrity Score by Country';
run;

/*ANALYSIS #1: WHAT COUNTRIES ARE HAPPIEST ANALYSIS*/
/*1-How does the average happiness level vary depending on the region?*/

```



```

/*use proc sgplot to graph the average response time by region*/
proc sgplot data=merged;
vbar Region /response=Happiness_score stat=mean categoryorder=RESPDESC
fillattrs=graphdata3;
label Happiness_Score="Average Happiness Score";
yaxis values=(0 to 10 by 1) min=0 max=10 valueshint;
title "Average Happiness Scores in Global Regions (2016)";
run;

/*2-Is there a statistically significant difference between regions?*/
/*use proc anova to determine if there is a statistically significant
difference in happiness between regions*/
proc anova data=happy;
    class Region;          * Variable with groups;
    model Happiness_score = Region; * Variable with experimental
results;
    means Region / Scheffe bon;
run;

/*3-how do components vary across regions?*/
/*create new variables to group related categories using datastep*/
data happy2;
set happy;
Economy=Economy_gdpper capita;
Family=Family;
Health=health_lifeexpectancy;
Government=Freedom+trust_govcorruption;
Generosity=Generosity;
Dystopia=dystopia_residual;
drop Happiness_Score Happiness_Rank Lower__Confidence_Interval
Upper_Confidence_Interval Economy_gdpper capita health_lifeexpectancy
trust_govcorruption dystopia_residual;
ID=_n_;
run;

/*convert the data set from wide to long using proc transpose*/
proc transpose data=happy2 out=happy_long(rename=(Coll=Value))
name=Variable;
by ID Country Region;
var Economy Family Health Government Generosity Dystopia;
run;

/*sort the dataset using proc sort*/
proc sort data=happy_long out=happy_long;
    by Region Country Variable;
run;

/*plot happiness by components by region using sgplot*/
proc sgplot data=happy_long;
vbar Region /response=Value group=Variable stat=mean
categoryorder=RESPDESC;
label Value="Average Happiness Score";
title "Average Happiness Scores in Global Regions (2016)";
title2 "Broken up by Components of Happiness";

```

```

run;

/*4-What are the 20 happiest countries?*/
/*sort countries by happiness rating using proc sort*/
proc sort data=happy out=happy;
  by DESCENDING Happiness_Score;
run;

/*get dataset showing top 20 happiest using data step*/
data happy_top20;
set happy(obs=20);
run;

/*create a barplot of 20 happiest countries using proc sgplot*/
proc sgplot data=happy_top20;
vbar Country /response=Happiness_score categoryorder=RESPDESC
fillattrs=graphdata1;
label Happiness_Score="Happiness Score";
yaxis grid values=(0 to 10 by 1);
title "20 Happiest Countries in the World (2016)";
run;

/*5-What are the 20 least countries?*/
/*get dataset showing bottom 20 happiest using data step and where
statement*/
data happy_bottom20;
set happy;
where Happiness_Rank>137;
run;

/*create a barplot of 20 least happy countries using proc sgplot*/
proc sgplot data=happy_bottom20;
vbar Country /response=Happiness_score categoryorder=RESPDESC
fillattrs=graphdata2;
label Happiness_Score="Happiness Score";
yaxis grid values=(0 to 10 by 1);
title "20 Least Happiest Countries in the World (2016)";
run;

/*6- how many countries have a happyness score >5 and <5*/
/*use data step and if and then statement */
data happy_5;
set happy;
if Happiness_Score > 5 then Classification="Greater than 5";
if Happiness_Score le 5 then Classification="Less than 5";
HappinessScore=round(Happiness_Score);
run;

/*sort the dataset using proc sort*/
proc sort data=happy_5;
by Region Classification;
run;

/*use proc freq to get frequency of each region*/

```

```

proc freq data=happy_5;
tables Region*Classification /nocol nopercent;
run;

/*7-what is the frequency of happiness_scores among different
countries/what is the distribution*/
/*use proc freq to make frequency plot of regions*/
proc freq data = happy_5;
tables HappinessScore * Region / plots = freqplot (twoway =
grouphorizontal);
title 'Distribution of Happiness Scores by Region';
run;

/*8-what is the distribution of happiness score*/
/*use proc univariate to create a histogram*/
proc univariate data = happy;
var Happiness_Score;
histogram Happiness_Score / endpoints=0 to 10 by 1 normal;
label Happiness_Score="Happiness Score";
title "Distribution of Happiness Scores among All Countries (2016)";
run;

/*SUICIDE/GDP ANALYSIS*/
/*1-What is the relationship between happiness and gdp*/
/*get logarithmic gdp using a data step*/
data merged;
set merged;
Log_GDP=log(gdp_2016);
run;

/*use proc sgplot to examine the relationship between gdp and happiness*/
proc sgplot data=merged;
reg x=gdp_2016 y=Happiness_Score / DEGREE=2;
label Happiness_Score="Happiness Score";
label gdp_2016="2016 GDP Per Capita (USD)";
yaxis values=(0 to 10 by 1);
title 'Relationship between Happiness Score and GDP among Countries';
title2 'Fitted with a Quadratic Regression Model (Degrees=2)';
run;

/*use proc reg to examine the relationship between log gdp and
happiness*/
proc reg data=merged;
model Happiness_Score=Log_GDP;
label Happiness_Score='Happiness Score';
label Log_GDP='Logarithm of 2016 GDP Per Capita';
run;

/*2-What is the relationship between happiness and suicides?*/
/*use proc sgplot to examine the relationship between suicides and
happiness*/
proc sgplot data=merged;
reg x=Happiness_Score y=_2016_suicide / CLM CLI;
label Happiness_Score="Happiness Score";

```

```

label _2016_suicide="2016 Crude Suicide Rate (per 100,000 persons)";
yaxis values=(0 to 35 by 5);
xaxis values=(0 to 10 by 1);
title 'Relationship between Happiness Score and Suicide Rates among
Countries';
run;

```

```

/*use proc reg to examine the relationship between suicides and
happiness*/
proc reg data=merged;
model _2016_suicide=Happiness_Score;
label Happiness_Score='Happiness Score';
label _2016_suicide='Crude Suicide Rate';
run;

```

```

/*3- is there a difference in relationship for men and women?*/
/*import the men and women dataset*/
/*import the happiness dataset using PROC IMPORT*/
proc import
    out=menwomen
    datafile= 'menvwomen.csv'
    dbms=dlm replace;
    delimiter=', ';
    getnames=YES;
    guessingrows=20404042;
run;

```

```

proc sql;
    create table differenceanalysis as
    select * from happy a left join menwomen b
    on a.Country = b.Country_Name;
quit;

```

```

proc sgplot data=differenceanalysis;
reg x=Happiness_Score y=women_suicides / legendlabel='Womens Suicides';
reg x=Happiness_Score y=men_suicides / legendlabel='Mens Suicides';
label Happiness_Score="Happiness Score";
label women_suicides="Women Suicide Rate";
label men_suicides="Men Suicide Rate";
yaxis label="2016 Crude Suicide Rate (per 100,000 persons)" values=(0 to
35 by 5);
xaxis values=(0 to 10 by 1);
title 'Relationship between Happiness Score and Men and Women Suicide
Rates among Countries';
run;

```

```

/*ANALYSIS OF EXPRESSION ISSUES*/

```

```

/*1-What expression factors are most closely correlated with happiness*/

```

```

/*use proc sql to select expression variables*/
proc sql;
create table expression_happy as
select country, pf_expression_influence, pf_expression_jailed,
pf_expression_cable, pf_expression_newspapers, pf_expression_internet,
Happiness_score
from data;
quit;

/*use proc corr to get correlations and label variables*/
proc corr data=expression_happy nosimple noprob;
label pf_expression_influence='Regulation of Media Content';
label pf_expression_jailed='Media Jailed';
label pf_expression_cable='Access to Cable';
label pf_expression_newspapers='Access to Newspapers';
label pf_expression_internet='Access to Internet';
label Happiness_Score='Happiness Score';
run;

/*2-HOW DOES FREEDOM OF EXPRESSION VARY ACROSS countries overtime */
/*use proc sql to select only variables of interest*/
proc sql;
create table expression as
select year, Countries, pf_expression
from freedom
order by year;
quit;

/*create macro to make sgplots*/
%macro show_result (country1=, country2=,country3=,country4=, country5=);
data expressiondata;
set expression;
where Countries=&country1 or Countries=&country2 or Countries=&country3
or Countries=&country4 or Countries=&country5;
run;
proc sgplot data = expressiondata noautolegend noborder;
series x = Year y = pf_expression / group=Countries
lineattrs=(thickness=2); /*Problem 8.4: plot time series and adjust line
color/thickness*/
xaxis label="Year" interval=year;
yaxis label="Freedom of Expression Score by Country" values=(0 to
10 by 1);
title 'Freedom of Expression Scores from 2008-2016';
keylegend;
run;
%mend show_result;

/*use macro to make plots*/
%show_result(country1="United States",
country2="Denmark",country3="Burundi",country4="",country5="");
%show_result(country1="Syria",
country2="Ukraine",country3="Egypt",country4="",country5="");
%show_result(country1="China", country2="India",country3="United
States",country4="Indonesia",country5="Brazil");

```

```

/*3-WHAT REGIONS HAVE THE HIGHEST FREEDOM OF EXPRESSION*/
/*use proc means to get region */
proc sort data=data;
by region;
run;
proc means data=data;
  class region;
  var pf_expression;
run;

```

```

/* ANALYSIS OF WOMEN'S RIGHTS SCORES */

```

```

/*import the freedom dataset*/
proc import
  out=freedom
  datafile= 'human_freedom.csv'
  dbms=dlm replace;
  delimiter=', ';
  getnames=YES;
  guessingrows=20404042;
run;

```

```

/*get 2016 data only*/

```

```

data freedom;
  set freedom;
  where year = 2016;
run;

```

```

/*import the happiness dataset*/

```

```

proc import
  out=happy
  datafile= '2016_happy.csv'
  dbms=dlm replace;
  delimiter=', ';
  getnames=YES;
  guessingrows=20404042;
run;

```

```

/*import the Freedom dataset*/

```

```

proc import
  out=economic
  datafile= 'Economic_Freedom.csv'
  dbms=dlm replace;
  delimiter=', ';
  getnames=YES;
  guessingrows=20404042;
run;

```

```

/*get 2016 only*/

data economic;
    set economic;
    where year = 2016;
run;

/*to merge happiness and economic*/

proc sql;
    create table merged as
    select * from happy a left join economic b
    on a.Country = b.Countries;
quit;

/*to merge merged dataset with freedom*/
/*drop repeat columns first*/

data freedom;
    set freedom;
    drop year countries region;
run;

proc sql;
    create table data as
    select * from merged a left join freedom b
    on a.ISO_code=b.ISO_code;
quit;

/* Create table womenhappiness that includes variables related to women's
rights, happiness_score, country, and region. */

proc sql;
    create table womenhappiness as
    select region, country, happiness_score label = "Happiness Score",
    pf_movement_women label = "Women's
Movement",
    pf_ss_women label = "Women's
Safety" ,
    pf_ss_women_inheritance label =
"Women's Inheritance",
    pf_ss_women_missing label =
"Missing Women",
    pf_ss_women_fgm label = "Female
Genital Mutilation"
    from data;
quit;

/* Create a table with summary statistics for variables
pf_movement_women,
pf_ss_women, pf_ss_women_inheritance, pf_ss_women_missing, and
pf_ss_women_fgm */

proc tabulate data=womenhappiness;

```

```

var pf_movement_women
    pf_ss_women
    pf_ss_women_inheritance
    pf_ss_women_missing
    pf_ss_women_fgm;
table (pf_movement_women pf_ss_women pf_ss_women_inheritance
pf_ss_women_missing pf_ss_women_fgm)*(N MEAN STD MIN
MAX);
    title "Summary Statistics for Variables Related to Women's Rights";
run;

/* Create dataset womenhappiness2 that has one column pf that includes
all
    values from pf_movement_women, pf_ss_women, pf_ss_women_inheritance,
pf_ss_women_missing,
    pf_ss_women_fgm, and averagewomenscore. */

proc sort data= womenhappiness;
    by region country happiness_score;
run;

/* Transpose womenhappiness data set so that women's rights
variables are in the same column, under the variable women_type */

proc transpose data=womenhappiness
    out=womenhappiness2 (rename=(coll=pf _name_=women_type));
var pf_movement_women
    pf_ss_women
    pf_ss_women_inheritance
    pf_ss_women_missing
    pf_ss_women_fgm;
    by region country happiness_score;
run;

/* Order womenhappiness2 by women_type */

proc sort data = womenhappiness2;
    by women_type;
run;

/* Rename women_type observations */

data womenhappiness2;
    set womenhappiness2;
    if women_type = "pf_movement_women" then women_type = "Women's
Movement";
    if women_type = "pf_ss_women" then women_type = "Women's Safety";
    if women_type = "pf_ss_women_inheritance" then women_type =
"Women's Inheritance";
    if women_type = "pf_ss_women_missing" then women_type = "Missing
Women";
    if women_type = "pf_ss_women_fgm" then women_type = "FGM";
run;

```



```

/* Plot side-by-side vertical boxplots for each women variable */

proc sgplot data = womenhappiness2;
    vbox pf / category = women_type fillattrs=graphdata3;
    title " Boxplots for Women Right's Variables";
    xaxis label = "Women Variables" ;
    yaxis label = "Score";
run;

/* Correlation table */

proc corr data=womenhappiness nomiss nosimple ;
    var pf_movement_women
        pf_ss_women
        pf_ss_women_inheritance
        pf_ss_women_missing
        pf_ss_women_fgm
        happiness_score;
    title "Correlation Table for Variables Related to Women's Rights and
Happiness Score";
run;

/* Scatterplot matrix */

proc sgscatter data=womenhappiness;
    matrix happiness_score pf_movement_women pf_ss_women
pf_ss_women_inheritance pf_ss_women_missing pf_ss_women_fgm /
diagonal=(histogram);
    title "Correlation Matrix for Variables Related to Women's Rights and
Happiness Score";
run;

/* Regression Analysis */

proc reg data = womenhappiness;
    model happiness_score = pf_movement_women
                                pf_ss_women
                                pf_ss_women_inheritance
                                pf_ss_women_missing;
    title 'Results of Regression Analysis';
run;

/* ANALYSIS OF CRIME SCORES */

proc sql;
    create table crimehappiness as
    select region, country,happiness_score, pf_ss_homicide label =
"Homicides",
                                pf_ss_disappearances label =
"Disappearances, Conflict, and Terrorism",
                                pf_ss label = "Safety and
Security",
                                pf_ss_disappearances_disap label
= "Disappearances",

```

```

                                pf_ss_disappearances_violent
label = "Violent Conflict",
                                pf_ss_disappearances_organized
label = "Organized Conflict",
                                ef_legal_police label =
"Reliability of Police",
                                ef_regulation_business_bribes
label = "Bribes"
    from data;
quit;

/* Plot histograms for all variables related to crime. */

%macro crimehist(dataset,variable);
    proc sgplot data=&dataset;
        histogram &variable / scale = count fillattrs=graphdata1;
        title "&variable Histogram";
    run;
%mend crimehist;

%crimehist(crimehappiness,pf_ss_homicide);
%crimehist(crimehappiness,pf_ss_disappearances);
%crimehist(crimehappiness,pf_ss);
%crimehist(crimehappiness,pf_ss_disappearances_disap);
%crimehist(crimehappiness,pf_ss_disappearances_violent);
%crimehist(crimehappiness,pf_ss_disappearances_organized);
%crimehist(crimehappiness,ef_legal_police);
%crimehist(crimehappiness,ef_regulation_business_bribes);

/* Order crimehappiness by region, country, and happiness_score. */

proc sort data= crimehappiness;
    by region country happiness_score;
run;

/* Transpose crimehappiness data set so that crime variables are in the
    same column, under the variable crime_type */

proc transpose data=crimehappiness
    out=crimehappiness2 (rename=(coll=pf _name_=crime_type));
    var pf_ss_homicide
        pf_ss_disappearances
        pf_ss
        pf_ss_disappearances_disap
        pf_ss_disappearances_violent
        pf_ss_disappearances_organized
        ef_legal_police
        ef_regulation_business_bribes;
    by region country happiness_score;
run;

/* order crimehappiness2 by crime_type. */

```

```

proc sort data = crimehappiness2;
    by crime_type;
run;

/* Rename crime_type observations */

data crimehappiness2;
    set crimehappiness2;
    if crime_type = "pf_ss_homicide" then crime_type = "Homicides";
    if crime_type = "pf_ss_disappearances" then crime_type =
"Disappearances,Conflict,Terrorism";
    if crime_type = "pf_ss" then crime_type = "Safety";
    if crime_type = "pf_ss_disappearances_disap" then crime_type =
"Disappearances";
    if crime_type = "pf_ss_disappearances_violent" then crime_type =
"Violent Conflict";
    if crime_type = "pf_ss_disappearances_organized" then crime_type =
"Organized Conflict";
    if crime_type = "ef_legal_police" then crime_type = "Police
Reliability";
    if crime_type = "ef_regulation_business_bribes" then crime_type =
"Bribes";

/* Plot side-by-side vertical boxplots for each crime variable */

proc sgplot data = crimehappiness2;
    vbox pf / category = crime_type fillattrs=graphdata3;
    title " Boxplots for Crime Variables";
    xaxis label = "Crime Variables";
    yaxis label = "Score";
run;

/* Correlation table for crime variables and happiness_score */

proc corr data = crimehappiness nosimple ;
    var happiness_score
        pf_ss_homicide
        pf_ss_disappearances
        pf_ss
        pf_ss_disappearances_disap
        pf_ss_disappearances_violent
        pf_ss_disappearances_organized
        ef_legal_police
        ef_regulation_business_bribes;
    title "Correlation Table for Crime Variables and Happiness Score";
run;

/* Produce scatterplots with regression lines of crime variables
vs. happiness_score, paneled by the type of crime variable. */

proc sgpanel data = crimehappiness2;
    panelby crime_type;
    reg x = pf y = happiness_score/CLM;
    title "Crime Variables vs. Happiness Score";

```

```

    label crime_type = variable;
run;

/* Regression Analysis */

proc reg data = crimehappiness;
    model happiness_score = pf_ss_homicide
                                pf_ss_disappearances
                                pf_ss
                                pf_ss_disappearances_disap
                                pf_ss_disappearances_violent
                                pf_ss_disappearances_organized
                                ef_legal_police
                                ef_regulation_business_bribes;
    title 'Results of Regression Analysis';
run;

```

```

/*FINAL PROJECT*/
libname ds 'U:\SAS\datasets\';
x 'cd U:\SAS\extdata\';

```

```

/*#####IMPORT THE DATA#####*/

```

```

/*DATASET 1: GDP DATASET*/
/*import the gdp dataset using PROC IMPORT*/
proc import
    out=gdp
    datafile= 'gdp.csv'
    dbms=dlm replace;
    delimiter=',';
    getnames=YES;
    guessingrows=20404042;
run;

```

```

/*DATASET 2: SUICIDE DATASET*/
/*import the suicide dataset using PROC IMPORT*/
proc import
    out=suicide
    datafile= 'suicide.csv'
    dbms=dlm replace;
    delimiter=',';
    getnames=YES;
    guessingrows=20404042;
run;

```

```

/*DATASET 3: HAPPINESS DATASET*/
/*import the happiness dataset using PROC IMPORT*/
proc import
    out=happy

```

```

        datafile= '2016_happy.csv'
        dbms=dlm replace;
        delimiter=', ';
        getnames=YES;
        guessingrows=20404042;
run;

/*DATASET 4: FREEDOM DATASET*/
/*import the happiness dataset using PROC IMPORT*/
proc import
    out=freedom
    datafile= 'human_freedom.csv'
    dbms=dlm replace;
    delimiter=', ';
    getnames=YES;
    guessingrows=20404042;
run;

/*MERGE THE DATASETS TO GET HAPPINESS AND GDP AND SUICIDES */
/*merge happiness and economic GDP datasets together using a left join
with proc sql*/
proc sql;
    create table merged_part1 as
    select * from happy a left join gdp b
    on a.Country = b.Country;
quit;
/*merge the merged and suicide datasets together using a left join with
proc sql*/
proc sql;
    create table merged as
    select * from merged_part1 a left join suicide b
    on a.Country = b.Country;
quit;

/*filter the freedom dataset for only 2016 data before merging*/
proc sql;
create table freedom2 as
select *
from freedom
where year = 2016;
quit;

/*drop repeat region column first*/
data freedom2;
    set freedom2;
    drop region;
run;

/*merge the merged and suicide datasets together using a left join with
proc sql*/
proc sql;
    create table data as
    select * from merged a left join freedom2 b

```

```

        on a.Country=b.Countries;
quit;

/*#####MAKE MAPS#####*/

/*get country names from shape file*/
data maps_to_merge;
set mapsgfk.world;
Keep=1;
drop SEGMENT LONG LAT X Y ISO ISOALPHA2 RESOLUTION DENSITY CONT LAKE;
run;

/*sort by country ID*/
proc sort data=maps_to_merge out=maps_to_merge2 NODUPKEY;
by ID;
run;

/*use proc sql to merge shape file with happiness scores*/
proc sql;
    create table happiness_data_for_map as
    select * from maps_to_merge2 a left join happy b
    on a.IDNAME = b.Country;
quit;

/*make map of happiness scores*/
proc gmap data=happiness_data_for_map map=mapsgfk.world;
id ID;
choro Happiness_Score / coutline=black;
label Happiness_Score='Happiness Score';
title 'Happiness Score by Country';
run;

/*2-suicide map*/

/*use proc sql to merge shape file with happiness scores*/
proc sql;
    create table suicide_data_for_map as
    select * from maps_to_merge2 a left join suicide b
    on a.IDNAME = b.Country;
quit;

/*make map of happiness scores*/
proc gmap data=suicide_data_for_map map=mapsgfk.world;
id ID;
choro _2016_suicide / coutline=black;
label _2016_suicide='Crude Suicide Rate per 100,000 Residents';
title 'Crude Suicide Rate by Region';
run;

/*3-freedom of expression*/

/*use proc sql to merge shape file with happiness scores*/
proc sql;
    create table freedom_data_for_map as

```

```

        select * from maps_to_merge2 a left join freedom2 b
        on a.IDNAME = b.countries;
quit;

proc gmap data=freedom_data_for_map map=mapsgfk.world;
id ID;
choro ef_legal_integrity / coutline=black;
label ef_legal_integrity='Legal Integrity Score (1-10)';
title 'Legal Integrity Score by Country';
run;

/*ANALYSIS #1: WHAT COUNTRIES ARE HAPPIEST ANALYSIS*/
/*1-How does the average happiness level vary depending on the region?*/
/*use proc sgplot to graph the average response time by region*/
proc sgplot data=merged;
vbar Region /response=Happiness_score stat=mean categoryorder=RESPDESC
fillattrs=graphdata3;
label Happiness_Score="Average Happiness Score";
yaxis values=(0 to 10 by 1) min=0 max=10 valueshint;
title "Average Happiness Scores in Global Regions (2016)";
run;

/*2-Is there a statistically significant difference between regions?*/
/*use proc anova to determine if there is a statistically significant
difference in happiness between regions*/
proc anova data=happy;
    class Region;          * Variable with groups;
    model Happiness_score = Region; * Variable with experimental
results;
    means Region / Scheffe bon;
run;

/*3-how do components vary across regions?*/
/*create new variables to group related categories using datastep*/
data happy2;
set happy;
Economy=Economy_gdpper capita;
Family=Family;
Health=health_lifeexpectancy;
Government=Freedom+trust_govcorruption;
Generosity=Generosity;
Dystopia=dystopia_residual;
drop Happiness_Score Happiness_Rank Lower__Confidence_Interval
Upper_Confidence_Interval Economy_gdpper capita health_lifeexpectancy
trust_govcorruption dystopia_residual;
ID=_n_;
run;

/*convert the data set from wide to long using proc transpose*/
proc transpose data=happy2 out=happy_long(rename=(Coll=Value))
name=Variable;
by ID Country Region;
var Economy Family Health Government Generosity Dystopia;
run;

```

```

/*sort the dataset using proc sort*/
proc sort data=happy_long out=happy_long;
  by Region Country Variable;
run;

/*plot happiness by components by region using sgplot*/
proc sgplot data=happy_long;
vbar Region /response=Value group=Variable stat=mean
categoryorder=RESPDESC;
label Value="Average Happiness Score";
title "Average Happiness Scores in Global Regions (2016)";
title2 "Broken up by Components of Happiness";
run;

/*4-What are the 20 happiest countries?*/
/*sort countries by happiness rating using proc sort*/
proc sort data=happy out=happy;
  by DESCENDING Happiness_Score;
run;

/*get dataset showing top 20 happiest using data step*/
data happy_top20;
set happy(obs=20);
run;

/*create a barplot of 20 happiest countries using proc sgplot*/
proc sgplot data=happy_top20;
vbar Country /response=Happiness_score categoryorder=RESPDESC
fillattrs=graphdata1;
label Happiness_Score="Happiness Score";
yaxis grid values=(0 to 10 by 1);
title "20 Happiest Countries in the World (2016)";
run;

/*5-What are the 20 least countries?*/
/*get dataset showing bottom 20 happiest using data step and where
statement*/
data happy_bottom20;
set happy;
where Happiness_Rank>137;
run;

/*create a barplot of 20 least happy countries using proc sgplot*/
proc sgplot data=happy_bottom20;
vbar Country /response=Happiness_score categoryorder=RESPDESC
fillattrs=graphdata2;
label Happiness_Score="Happiness Score";
yaxis grid values=(0 to 10 by 1);
title "20 Least Happiest Countries in the World (2016)";
run;

/*6- how many countries have a happyness score >5 and <5*/
/*use data step and if and then statement */

```



```

data happy_5;
set happy;
if Happiness_Score > 5 then Classification="Greater than 5";
if Happiness_Score le 5 then Classification="Less than 5";
HappinessScore=round(Happiness_Score);
run;

/*sort the dataset using proc sort*/
proc sort data=happy_5;
by Region Classification;
run;

/*use proc freq to get frequency of each region*/
proc freq data=happy_5;
tables Region*Classification /nocol nopercent;
run;

/*7-what is the frequency of happiness_scores among diffrent
countries/what is the distribution*/
/*use proc freq to make frequency plot of regions*/
proc freq data = happy_5;
    tables HappinessScore * Region / plots = freqplot (twoway =
grouphorizontal);
    title 'Distribution of Happiness Scores by Region';
run;

/*8-what is the distribution of happiness score*/
/*use proc univariate to create a histogram*/
proc univariate data = happy;
    var Happiness_Score;
    histogram Happiness_Score / endpoints=0 to 10 by 1 normal;
    label Happiness_Score="Happiness Score";
    title "Distribution of Happiness Scores among All Countries (2016)";
run;

/*SUICIDE/GDP ANALYSIS*/
/*1-What is the relationship between happiness and gdp*/
/*get logarithmic gdp using a data step*/
data merged;
set merged;
Log_GDP=log(gdp_2016);
run;

/*use proc sgplot to examine the relationship between gdp and happiness*/
proc sgplot data=merged;
reg x=gdp_2016 y=Happiness_Score / DEGREE=2;
label Happiness_Score="Happiness Score";
label gdp_2016="2016 GDP Per Capita (USD)";
yaxis values=(0 to 10 by 1);
title 'Relationship between Happiness Score and GDP among Countries';
title2 'Fitted with a Quadratic Regression Model (Degrees=2)';
run;

```

```

/*use proc reg to examine the relationship between log gdp and
happiness*/
proc reg data=merged;
model Happiness_Score=Log_GDP;
label Happiness_Score='Happiness Score';
label Log_GDP='Logarithm of 2016 GDP Per Capita';
run;

/*2-What is the relationship between happiness and suicides?*/
/*use proc sgplot to examine the relationship between suicides and
happiness*/
proc sgplot data=merged;
reg x=Happiness_Score y=_2016_suicide / CLM CLI;
label Happiness_Score="Happiness Score";
label _2016_suicide="2016 Crude Suicide Rate (per 100,000 persons)";
yaxis values=(0 to 35 by 5);
xaxis values=(0 to 10 by 1);
title 'Relationship between Happiness Score and Suicide Rates among
Countries';
run;

/*use proc reg to examine the relationship between suicides and
happiness*/
proc reg data=merged;
model _2016_suicide=Happiness_Score;
label Happiness_Score='Happiness Score';
label _2016_suicide='Crude Suicide Rate';
run;

/*3- is there a difference in relationship for men and women?*/
/*import the men and women dataset*/
/*import the happiness dataset using PROC IMPORT*/
proc import
    out=menwomen
    datafile= 'menvwomen.csv'
    dbms=dlm replace;
    delimiter=',';
    getnames=YES;
    guessingrows=20404042;
run;

proc sql;
    create table differenceanalysis as
    select * from happy a left join menwomen b
    on a.Country = b.Country_Name;
quit;

proc sgplot data=differenceanalysis;
reg x=Happiness_Score y=women_suicides / legendlabel='Womens Suicides';
reg x=Happiness_Score y=men_suicides / legendlabel='Mens Suicides';
label Happiness_Score="Happiness Score";
label women_suicides="Women Suicide Rate";

```

```

label men_suicides="Men Suicide Rate";
yaxis label="2016 Crude Suicide Rate (per 100,000 persons)" values=(0 to
35 by 5);
xaxis values=(0 to 10 by 1);
title 'Relationship between Happiness Score and Men and Women Suicide
Rates among Countries';
run;

```

```

/*ANALYSIS OF EXPRESSION ISSUES*/

```

```

/*1-What expression factors are most closely correlated with happiness*/
/*use proc sql to select expression variables*/
proc sql;
create table expression_happy as
select country, pf_expression_influence, pf_expression_jailed,
pf_expression_cable, pf_expression_newspapers, pf_expression_internet,
Happiness_score
from data;
quit;

```

```

/*use proc corr to get correlations and label variables*/
proc corr data=expression_happy nosimple noprob;
label pf_expression_influence='Regulation of Media Content';
label pf_expression_jailed='Media Jailed';
label pf_expression_cable='Access to Cable';
label pf_expression_newspapers='Access to Newspapers';
label pf_expression_internet='Access to Internet';
label Happiness_Score='Happiness Score';
run;

```

```

/*2-HOW DOES FREEDOM OF EXPRESSION VARY ACROSS countries overtime */
/*use proc sql to select only variables of interest*/
proc sql;
create table expression as
select year, Countries, pf_expression
from freedom
order by year;
quit;

```

```

/*create macro to make sgplots*/
%macro show_result(country1=, country2=,country3=,country4=, country5=);
data expressiondata;
set expression;
where Countries=&country1 or Countries=&country2 or Countries=&country3
or Countries=&country4 or Countries=&country5;
run;
proc sgplot data = expressiondata noautolegend noborder;
series x = Year y = pf_expression / group=Countries
lineattrs=(thickness=2); /*Problem 8.4: plot time series and adjust line
color/thickness*/
xaxis label="Year" interval=year;

```

```

        yaxis label="Freedom of Expression Score by Country" values=(0 to
10 by 1);
        title 'Freedom of Expression Scores from 2008-2016';
        keylegend;
run;
%mend show_result;

/*use macro to make plots*/
%show_result(country1="United States",
country2="Denmark",country3="Burundi",country4="",country5="");
%show_result(country1="Syria",
country2="Ukraine",country3="Egypt",country4="",country5="");
%show_result(country1="China", country2="India",country3="United
States",country4="Indonesia",country5="Brazil");

/*3-WHAT REGIONS HAVE THE HIGHEST FREEDOM OF EXPRESSION*/
/*use proc means to get region */
proc sort data=data;
by region;
run;
proc means data=data;
class region;
var pf_expression;
run;

```