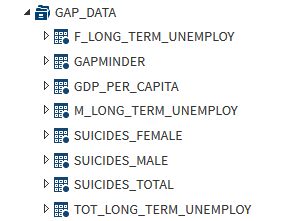
Data Management and Visualization

**Assignment – Week 2**

**Running Your First Program**

**By mapolarbear@gmail.com**

First, I would like to remind that for my research I decided to use GAPMINDER data set, containing some indicators, but not all needed to complete my research. To build extended version with more data, I downloaded and uploaded into SAS Studio additional CSV data file and added needed variables to GAPMINDER. Data from following CSV files/datasets added for the purpose of my research using PROC SQL in Part-2 of program, don’t be scared.



Here is brief description of code:

Part-1 - Retrieve GAPMINDER form provided MYDATA course data

Part-2 - Join/Add needed variables/data from downloaded [www.gapminder.org](http://www.gapminder.org) site

Part-3 - Creation of final GAPMINDER\_RSCH dataset with needed labels in corresponding formats

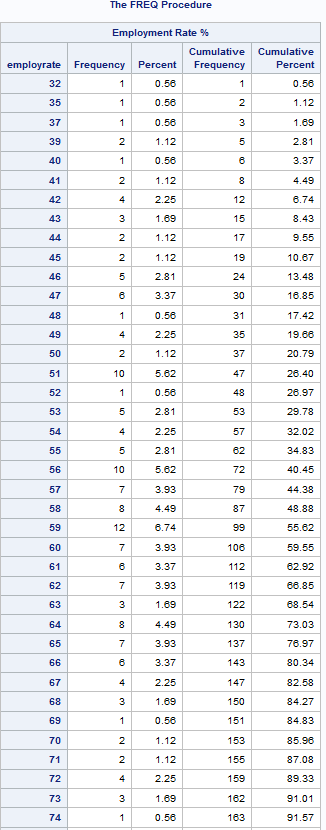
Part-4 - Execution of PROC FEQ to display frequencies for couple variables, including missing values.

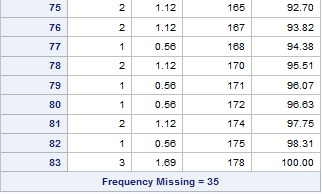
For verification purposes there are originally 213 countries in GAPMINDER dataset and if you add Cumulative Frequency to Missing Freq., you will get same 213, means nothing is skipped or removed.

In addition, based on my preliminary analysis of data I noticed that data is missing some countries, which means that my samples will be smaller than expected. I also noticed that some countries has higher **suicideper100TH** rating, it will be interesting to see if **suicideper100TH** correlates with **employrate** and **long term unemployment** ratings for each category.

1. **Output of program**









1. **Program code**

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\* Part 1 - get data \*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

LIBNAME mydata "/courses/d1406ae5ba27fe300 " access=readonly;

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\* 1. get gapminder data from project library \*/

/\* 2. format and add LABELs to variables \*/

/\* \*/

data gap\_data.gapminder; /\*save gapminder in permanent library gap\_data \*/

set mydata.gapminder;

format suicideper100th 2.1; /\* formatting to 2 decimal points \*/

format employrate 3.;

label polityscore="Country Political Score"

suicideper100th="Suicide Rating Per 100,000 %"

employrate="Employment Rate %";

run;

/\* sort data by joined column - country \*/

proc sort data=gap\_data.gapminder; /\* gapminder sorted by country \*/

by country;

run;

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\* Part 2 - join/build data \*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\* PROC SQL blocs to join/add variables from uploaded datasets \*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\* add tot\_long\_term\_unemploy columns \*/

/\* sort data by joined column - country first \*/

proc sort data=gap\_data.tot\_long\_term\_unemploy;

by country;

run;

proc sql noprint; /\* join column data \*/

create table gapminder\_tmp01 as

select a.\*

,b.\_2002 as ltu2002

,b.\_2004 as ltu2004

from gap\_data.gapminder a

left join

gap\_data.tot\_long\_term\_unemploy b

on a.country = b.country

;

quit;

run;

/\* add f\_long\_term\_unemploy columns \*/

/\* sort data by joined column - country first \*/

proc sort data=gap\_data.f\_long\_term\_unemploy;

by country;

run;

proc sql noprint; /\* join column data \*/

create table gapminder\_tmp02 as

select a.\*

,b.\_2002 as fltu2002

,b.\_2004 as fltu2004

from gapminder\_tmp01 a

left join

gap\_data.f\_long\_term\_unemploy b

on a.country = b.country

;

quit;

run;

/\* add m\_long\_term\_unemploy columns \*/

/\* sort data by joined column - country first \*/

proc sort data=gap\_data.m\_long\_term\_unemploy;

by country;

run;

proc sql noprint; /\* join column data \*/

create table gapminder\_tmp03 as

select a.\*

,b.\_2002 as mltu2002

,b.\_2004 as mltu2004

from gapminder\_tmp02 a

left join

gap\_data.m\_long\_term\_unemploy b

on a.country = b.country

;

quit;

run;

/\* add suicides\_total columns \*/

/\* sort data by joined column - country first \*/

proc sort data=gap\_data.suicides\_total;

by country;

run;

proc sql noprint; /\* join column data \*/

create table gapminder\_tmp04 as

select a.\*

,b.suicides2002

,b.suicides2004

from gapminder\_tmp03 a

left join

gap\_data.suicides\_total b

on a.country = b.country

;

quit;

run;

/\* add suicides\_female columns \*/

/\* sort data by joined column - country first \*/

proc sort data=gap\_data.suicides\_female;

by country;

run;

proc sql noprint; /\* join column data \*/

create table gapminder\_tmp05 as

select a.\*

,b.\_2002 as fsuicides2002

,b.\_2004 as fsuicides2004

from gapminder\_tmp04 a

left join

gap\_data.suicides\_female b

on a.country = b.country

;

quit;

run;

/\* add suicides\_male columns \*/

/\* sort data by joined column - country first \*/

proc sort data=gap\_data.suicides\_male;

by country;

run;

proc sql noprint; /\* join column data \*/

create table gapminder\_combined as

select a.\*

,b.\_2002 as msuicides2002

,b.\_2004 as msuicides2004

from gapminder\_tmp05 a

left join

gap\_data.suicides\_male b

on a.country = b.country

;

quit;

run;

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\* Part 3 - refine data, build final \*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\* 1. create final dataset gapminder\_rsch and \*/

/\* store it in permanent library gap\_data \*/

/\* for futture reference in research \*/

/\* 2. format, add lables to combined dataset \*/

/\* 3. keep only columns needed for research \*/

data gap\_data.gapminder\_rsch (keep=country

incomeperperson

femaleemployrate

polityscore

suicideper100th

employrate

urbanrate

ltu2002

ltu2004

fltu2002

fltu2004

mltu2002

mltu2004

suicides2002

suicides2004

fsuicides2002

fsuicides2004

msuicides2002

msuicides2004

);

set gapminder\_combined; /\* temporary dataset from WORK library \*/

/\* format data to more appropriate \*/

format ltu2002 3.2

ltu2004 3.2

fltu2002 3.2

fltu2004 3.2

mltu2002 3.2

mltu2004 3.2

fsuicides2002 3.2

fsuicides2004 3.2

msuicides2002 3.2

msuicides2004 3.2

;

/\* add lables to newly added/created variable/columns \*/

label ltu2002="2002 Long term unemployment rate - %"

ltu2004="2004 Long term unemployment rate - %"

fltu2002="2002 female long term unemployment rate - %"

fltu2004="2004 female long term unemployment rate - %"

mltu2002="2002 male long term unemployment rate - %"

mltu2004="2004 male long term unemployment rate - %"

fsuicides2002="2002 female suicide rate - %"

fsuicides2004="2004 female suicide rate - %"

msuicides2002="2002 male suicide rate - %"

msuicides2004="2004 male suicide rate - %"

suicides2002="Total suicide deaths in 2002"

suicides2004="Total suicide deaths in 2004"

;

run;

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\* Part 4 - frequency distribution of some data \*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\* perform distribution analysis of variables in question \*/

proc freq data=gap\_data.gapminder\_rsch;

tables employrate suicideper100th ltu2004;

run;

/\* display information about dataset and its variables \*/

proc contents data=gap\_data.gapminder\_rsch;

run;

/\* print contents of gap\_data.gapminder\_rsch \*/

proc print data=gap\_data.gapminder\_rsch;

title "\*\*\* Contents of gap\_data.gapminder\_rsch \*\*\*";

run;

title; /\* initialize title to previous state \*/

Document can be found also at following link at [www.github.com](http://www.github.com):