ASSIGNMENT 2

Analysis of Global CO_2 emissions using SAS, to determine the trend in CO_2 emission and other Greenhouse Gas (GHG) emissions.

Table of Contents

Introduction

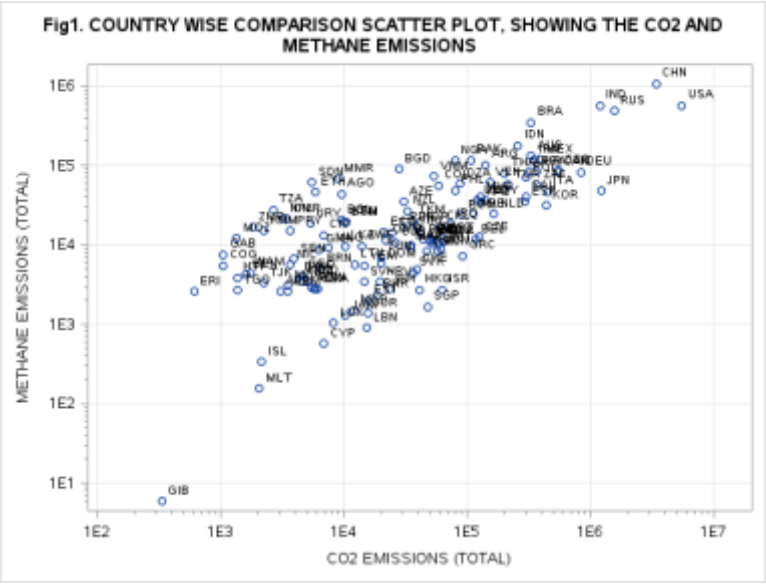
With global warming fast changing the dynamics of climate across the world, the impacts of contributing factors like greenhouse gasses – namely CO_2 - needs be determined on a global scale.
The purpose of this report is to state the findings derived from conducting an analysis of global CO_2 emissions, and determine a trend with other GHG emissions.

Methods

To undertake the analysis, the dataset "Climate_2.CSV" (sourced from http://www.worldbank.org/) has been used. This dataset contains information on the various emissions levels of carbon dioxide, methane etc. in million metric tons. The dataset consists of observations of 216 countries. There are 30 variables in the dataset ranging from AG_LND_EL5M_ZS (Land area below 5m), EN_ATM_CO2E_KT (Total CO2 emissions), SP_POP_TOTL (Total Population) etc.
The analysis was done using SAS STUDIO, which was used to summarize and analyze the findings. ANOVA test was conducted on SAS to test for null hypothesis keeping alpha at 0.05 as a generic rule.

Results

To get a better understanding of the dataset, an overview of the contents was created using code on SAS. From the overview, we could see that there were 216 observations (Countries) and 30 variables (Including CO_2 Emissions) in the dataset. The overview also showed us the type and length of the variables. To understand the effect certain greenhouse gases, have with each other, in Fig 1. below, a scatterplot has been used to see a country-wise comparison, showing the relation between CO_2 and CH4 (Methane) emissions.



The scatterplot shows a linear nature in general, which goes to show that as CO_2 emissions increase, so do Methane emissions. China, USA, Russia and India show the maximum amount of CO_2 and METHANE emissions as per the dataset.
To understand which countries are the biggest contributors towards these emissions, Table 1. was created, to show the 15 highest CO_2 emitting countries and their population.

TABLE 1. SHOWING TOP 15 COUNTRIES POPULATION WISE, AND THEIR CO2 EMISSIONS

Obs	COUNTRY	CO2 EMISSION	POPULATION
1	CHN	3405179.87	1262645000
2	IND	1186663.20	1015923000
3	USA	5512399.42	282172000
4	IDN	258120.13	213395411
5	BRA	330125.34	174425387
6	RUS	1553451.21	146303000
7	PAK	106449.34	144522192
8	BGD	27865.53	129592275
9	JPN	1219592.86	126870000
10	NGA	79181.53	123688536
11	MEX	383021.82	99959594
12	DEU	832100.97	82210000
13	VNM	53582.20	77630900
14	PHL	79111.86	77309965
15	EGY	141326.18	67648419

The above table shows that countries with higher population tends to have higher CO_2 emissions, with China leading the table.

Next, ANOVA test was conducted to compare the CO_2 emissions by the global region.

The following images show the ANOVA results.

ANOVA NULL HYPOTHESIS TEST FOR CO2 EMISSION BY GLOBAL REGIONS

The ANOVA Procedure

Class Level Information		
Class	Levels	Values
Region	7	EA_PAC EUR_CA LAT_AM MENA N_AM SSA S_ASA

Number of Observations Read	216
Number of Observations Used	196

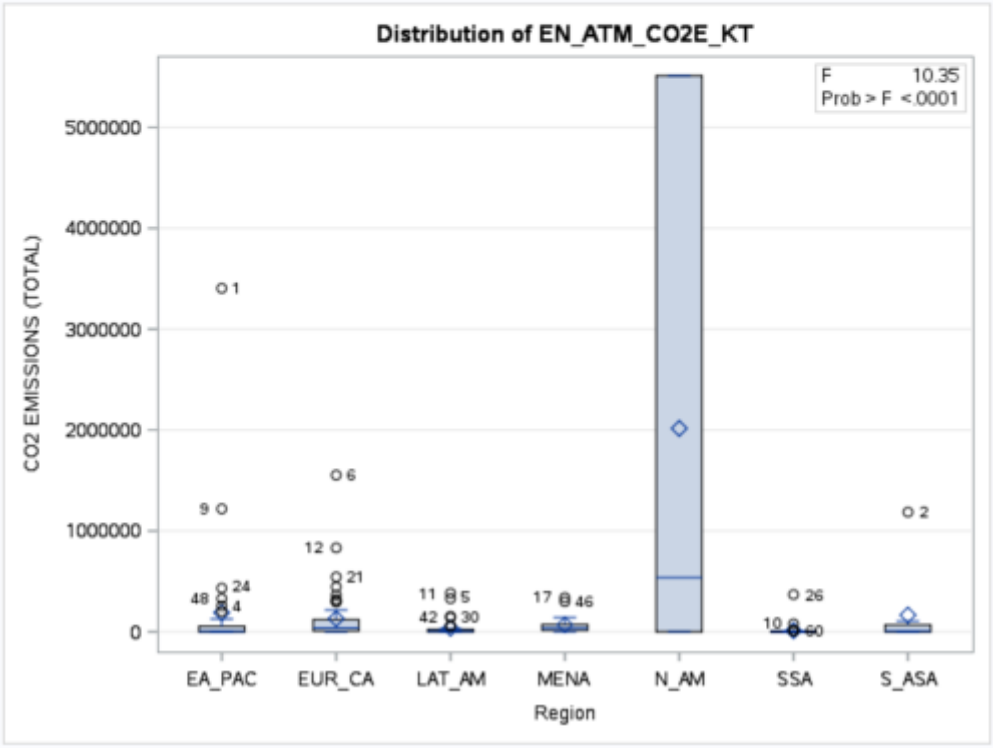
ANOVA NULL HYPOTHESIS TEST FOR CO2 EMISSION BY GLOBAL REGIONS

The ANOVA Procedure
Dependent Variable: EN_ATM_CO2E_KT CO2 EMISSIONS (TOTAL)

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	1.1801855E13	1.9669758E12	10.35	<.0001
Error	189	3.5912729E13	190014440951		
Corrected Total	195	4.7714584E13			

R-Square	Coeff Var	Root MSE	EN_ATM_CO2E_KT Mean
0.247343	363.7436	435906.5	119838.9

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Region	6	1.1801855E13	1.9669758E12	10.35	<.0001



From the ANOVA test, we find the PR value to be <.0001 which rejects the null hypothesis.

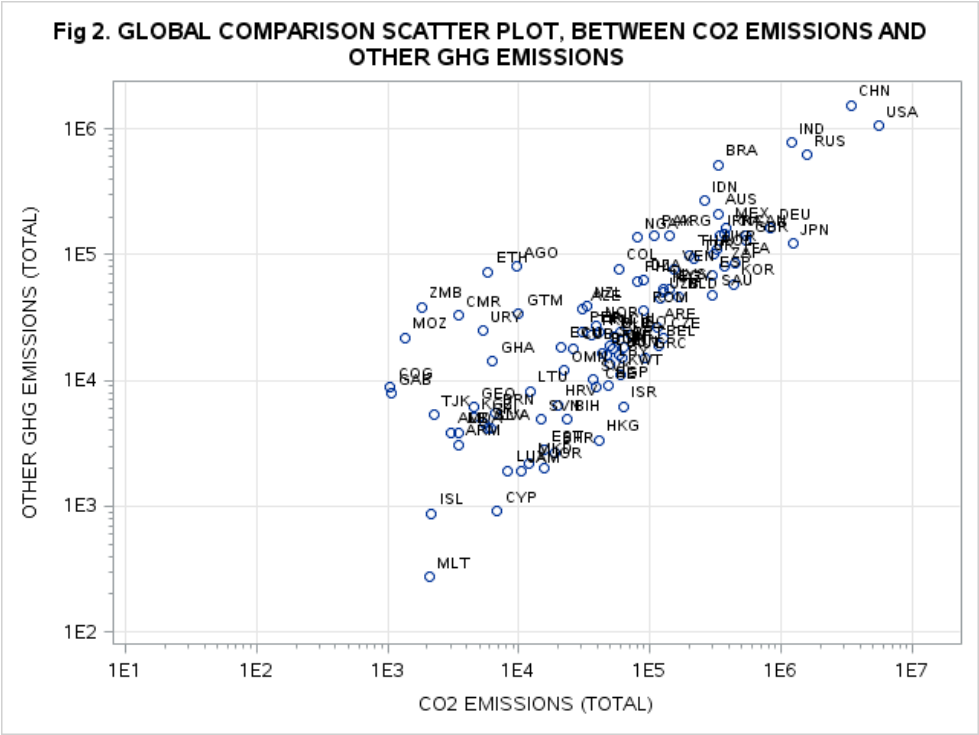
We further created a table to understand which were the top countries responsible for carbon dioxide emissions.

Table 2. shows a list of the top 15 countries emitting CO_2 (in million metric tons).

TABLE 2. SHOWING TOP 15 CO2 EMITTING COUNTRIES

Obs	COUNTRY	CO2 EMISSION
1	USA	5512399.42
2	CHN	3405179.87
3	RUS	1553451.21
4	JPN	1219592.86
5	IND	1186663.20
6	DEU	832100.97
7	GBR	543662.09
8	CAN	537402.52
9	ITA	446156.56
10	KOR	432460.31
11	MEX	383021.82
12	ZAF	368610.51
13	FRA	365559.56
14	IRN	339241.50
15	BRA	330125.34

Further, to better understand the effect of CO_2 emissions on other GHG emissions, we created a new variable in the dataset and added all other GHG emissions to compare with CO_2 emissions using scatterplot in Fig 2.



The scatterplot in Fig. 2, shows that there is an overall linear trend in CO_2 emissions globally compared to other GHG emissions.

Discussion

For the ANOVA test, we have assumed α to be 0.05, hence our Pr (< .001) is less than the α and the null hypothesis gets rejected. This shows that there is an effect of CO_2 emission when compared globally.

Conclusion

From our results and findings, we can see that over a global span, there is a noticeable linear relation between CO_2 emissions and various greenhouse gas emissions like N_2O, CH_4 etc. This goes to show that GHG emissions have a correlation with each other, and there is a high likelihood that if there is increased emission of CO_2 there will also be increased emission of other greenhouse gasses.
We conclude that the trend between CO_2 and GHG emissions are linear in nature. To curb emissions of GHGs, a control on emission of CO_2 will need to take place.

References

Data sourced from http://www.worldbank.org/ (and modified for the purpose of this assignment.)

Appendix

26/04/2017 Code: Worksheet.sas

```
LIBNAME ASSIGN2 '/home/s36235750/sasuser.v94/Assignment 2';

LIBNAME ASS2OUT '/home/s36235750/sasuser.v94/Assignment 2/OUTPUTS';


/* Q1 - Read the data into SAS using code and produce an overview of the datafile using SAS code */
DATA ASSIGN2.WORKSHEET;
    INFILE '/home/s36235750/sasuser.v94/Assignment 2/Climate_2.csv' DELIMITER=',' FIRSTOBS= 2;
    INPUT Country_code$ Region$ AG_LND_EL5M_ZS AG_YLD_CREL_KG
    BX_KLT_DINV_WD_GD_ZS EG_USE_COMM_GD_PP_KD EN_ATM_CO2E_KT
    EN_ATM_CO2E_PP_GD_KD EN_ATM_GHGO_KT_CE EN_ATM_METH_KT_CE
    EN_ATM_NOXE_KT_CE EN_CLC_GHGR_MT_CE EN_POP_EL5M_ZS
    EN_URB_MCTY_TL_ZS ER_H2O_FWTL_ZS ER_LND_PTLD_ZS IE_PPI_ENGY_CD
    IE_PPI_TELE_CD IE_PPI_TRAN_CD IS_ROD_PAVE_ZS NY_GDP_MKTP_CD
    NY_GNP_PCAP_CD SE_PRM_CMPT_ZS SH_H2O_SAFE_ZS SH_MED_PHYS_ZS
    SI_POV_DDAY SP_POP_GROW SP_POP_TOTL SP_URB_GROW SP_URB_TOTL;
RUN;

ODS _ALL_ Close;
ODS PDF file = '/home/s36235750/sasuser.v94/Assignment 2/OUTPUTS/Clim Overview.pdf';
PROC CONTENTS DATA = assign2.worksheet;
RUN;
ODS PDF CLOSE;


/* Q2 - Using SAS code, produce a graph to compare the CO2 emissions to the Methane emissions.
Label the graph appropriately, including the country code label on the data points.
Include this graph in your report. */
PROC SGPLOT DATA = assign2.worksheet;
    SCATTER  X = EN_ATM_CO2E_KT Y = EN_ATM_METH_KT_CE  / DATALABEL = COUNTRY_CODE;
        TITLE 'Fig1. COUNTRY WISE COMPARISON SCATTER PLOT, SHOWING THE CO2 AND METHANE EMISSIONS';
        XAXIS LABEL = 'CO2 EMISSIONS (TOTAL)' TYPE = LOG MINOR GRID MIN = 1E2;
        YAXIS LABEL = 'METHANE EMISSIONS (TOTAL)' TYPE = LOG MINOR GRID;
RUN;



/* Q3 - Sort the data by total population size, and then produce a list of the top 15 countries
(by the largest population) with their CO2 emission and their population size all using SAS code.
Ensure you have an appropriate title. Include this table in your report. */
PROC SORT DATA = assign2.worksheet;
    BY descending SP_POP_TOTL;
RUN;

ODS _ALL_ CLOSE;
ODS PDF FILE = '/home/s36235750/sasuser.v94/Assignment 2/OUTPUTS/TOP15COUNTRIES.pdf';
PROC PRINT DATA = assign2.worksheet (OBS = 15) LABEL;
    TITLE 'TABLE 1. SHOWING TOP 15 COUNTRIES POPULATION WISE, AND THEIR CO2 EMISSIONS';
    VAR COUNTRY_CODE EN_ATM_CO2E_KT SP_POP_TOTL;
    LABEL COUNTRY_CODE = 'COUNTRY' EN_ATM_CO2E_KT = 'CO2 EMISSION' SP_POP_TOTL = 'POPULATION';
RUN;
ODS PDF CLOSE;


/* Q4 - Using SAS code, compare the CO2 emission produced by the global regions using ANOVA.
This will test the null hypothesis that there is no difference in the CO2 emission produced by
all global regions. */
ODS _ALL_ CLOSE;
ODS PDF FILE = '/home/s36235750/sasuser.v94/Assignment 2/OUTPUTS/ANOVA.pdf';
PROC ANOVA DATA = ASSIGN2.WORKSHEET;
    CLASS REGION;
    MODEL EN_ATM_CO2E_KT = REGION;
    TITLE 'ANOVA NULL HYPOTHESIS TEST FOR CO2 EMISSION BY GLOBAL REGIONS';
    LABEL EN_ATM_CO2E_KT = 'CO2 EMISSIONS (TOTAL)';
RUN;
ODS PDF CLOSE;


/* Q5 - Further findings and Analyses */

/* To find top 15 countries with CO2 emissions, we first sort the data by descending
CO2 emissions, then print the results */
PROC SORT DATA = assign2.worksheet;
    BY descending EN_ATM_CO2E_KT;
RUN;

PROC PRINT DATA = assign2.worksheet (obs = 15) LABEL;
    TITLE 'TABLE 2. SHOWING TOP 15 CO2 EMITTING COUNTRIES';
    VAR COUNTRY_CODE EN_ATM_CO2E_KT;
```

1/2

```
    LABEL COUNTRY_CODE = 'COUNTRY' EN_ATM_CO2E_KT = 'CO2 EMISSION'
    ;
RUN;

/* By creating a subset of the data in a table with emission data and country name,
compare the effect of CO2 emissions with regards to other GHG emissions, and create a
new variable OTHEREMI which sums up N2O, CH4 and other GHG emissions*/
PROC SQL;
    CREATE TABLE ASSIGN2.allc AS
    SELECT COUNTRY_CODE, EN_ATM_CO2E_KT, EN_ATM_METH_KT_CE, EN_ATM_NOXE_KT_CE, EN_ATM_GHGO_KT_CE
    FROM ASSIGN2.WORKSHEET
    ;
QUIT;

Data assign2.allc;
    SET ASSIGN2.allc;
    OTHEREMI = EN_ATM_METH_KT_CE + EN_ATM_NOXE_KT_CE + EN_ATM_GHGO_KT_CE;
RUN;

/* using scatterplot to show the comparison of co2 emissions and other ghg emissions*/
ODS _ALL_ Close;
ODS PDF file = '/home/s36235750/sasuser.v94/Assignment 2/OUTPUTS/SCATTER2.pdf';
PROC SGPLOT DATA = assign2.allc;
    SCATTER  X = EN_ATM_CO2E_KT Y = OTHEREMI  / DATALABEL = COUNTRY_CODE;
        TITLE 'Fig 2. GLOBAL COMPARISON SCATTER PLOT, BETWEEN CO2 EMISSIONS AND OTHER GHG EMISSIONS';
        XAXIS LABEL = 'CO2 EMISSIONS (TOTAL)' TYPE = LOG MINOR GRID MIN = 1E1;
        YAXIS LABEL = 'OTHER GHG EMISSIONS (TOTAL)' TYPE = LOG MINOR GRID min = 1e2;
RUN;
ODS PDF CLOSE;
```