An Analysis of Relationship among CO2 Emissions, GDP, Population, and Energy Use on a Global Scale

Author: Wei Qi Pang

Advisors: Dr. Jonathan Gratch, Dr. Jian Zhang

Abstract

This project analyzes the relationship between carbon dioxide emissions (CO2) and different global indicators from 2012 to 2014 and explores how energy-related factors affect CO2 emissions. CO2 is one of the drivers that causes the global temperature to rise and leads to climate change. CO2 emissions are different by region according to the situation of the economy and human factors. Therefore, several questions need to be analyzed. First, what is the difference between CO2 emissions in the high gross domestic product (GDP) and low GDP countries, and how to determine the GDP category? Second, do countries with larger populations consume more oil (energy use) and affect CO2 emissions more significantly? Third, what is the difference between life expectancy in countries with high CO2 emissions versus countries with low co2 emissions? This project adopts exploratory data analysis methods and machine learning algorithms to identify the relationship based on the patterns between variables of data sources from The World Bank.

Objectives

1. To verify inferences from past studies and investigate their similarities and differences to improve the gap between the studies.
2. To study direct and indirect determinants of CO2 emissions on life expectancy.
3. To determine the relationship between CO2 emissions and other indicators, such as GDP, population, and energy use.

Data

Data Description:

Country: Country Name

Code: Country Code

EnergyUse: KG(kilogram) of Oil Equivalent Per Capita

CO2Emissions: CO2 Emission by Country (metric kiloton-kt)

Percapita: CO2 Emission Per Capita

UrbanPopulation: Percentage of Total Population (%)

TotalPopulation: Total Population of the Country

GDP: Current US$

LifeExpectancy: Life Expectancy of the Country

Method

-Multi-variable visualization techniques and descriptive statistics: summarize the characteristics

-ANOVA test, T-test, and a regression algorithm: investigate the influence of GDP, energy use, and total population on CO2 emissions

-*Equations (*Gülden and Neşe, 2013) *:*

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n + \varepsilon$$

Result

Table 1: Statistical summary for 2012.

	EnergyUse	CO2Emissions	Percapita	TotalPopulation	GDP	lifeExpectancy
count	266.000000	2.660000e+02	266.000000	2.660000e+02	2.660000e+02	266.000000
mean	2440.925873	1.257054e+06	4.372096	2.876904e+08	2.407733e+12	70.860322
std	2360.604150	3.785682e+06	4.638586	8.999074e+08	8.027217e+12	7.959073
min	63.669707	1.000000e+01	0.035207	1.013600e+04	3.767177e+07	47.416000
25%	950.895499	4.607500e+03	0.910227	1.641155e+06	9.511467e+09	66.490000
50%	2440.925873	4.356000e+04	3.718511	9.925639e+06	5.278262e+10	71.933427
75%	2440.925873	7.210981e+05	6.089142	6.516991e+07	8.626779e+11	76.139189
max	17630.071470	3.244419e+07	33.373132	7.089255e+09	7.535601e+13	85.417073

Table 2: Statistical summary for 2013.

	EnergyUse	CO2Emissions	Percapita	TotalPopulation	GDP	LifeExpectancy
count	266.000000	2.660000e+02	266.000000	2.660000e+02	2.660000e+02	266.000000
mean	2415.142058	1.282564e+06	4.280547	2.914996e+08	2.482236e+12	71.129470
std	2321.420651	3.873042e+06	4.590540	9.110602e+08	8.260102e+12	7.733702
min	65.393340	1.000000e+01	0.042976	1.020800e+04	3.750908e+07	48.663000
25%	917.978242	4.325000e+03	0.922762	1.690462e+06	1.016976e+10	66.755000
50%	2415.142058	4.108000e+04	3.627916	1.002641e+07	5.667069e+10	72.042636
75%	2415.142058	7.418177e+05	5.796861	6.553108e+07	8.628895e+11	76.379791
max	18178.139010	3.305385e+07	31.927018	7.175500e+09	7.742744e+13	83.831707

Table 3: Statistical summary for 2014.

	EnergyUse	CO2Emissions	Percapita	TotalPopulation	GDP	LifeExpectancy
count	266.000000	2.660000e+02	266.000000	2.660000e+02	2.660000e+02	266.000000
mean	2469.420530	1.286440e+06	4.195139	2.953218e+08	2.552297e+12	71.480750
std	2320.263802	3.881771e+06	4.485167	9.222012e+08	8.483493e+12	7.605993
min	66.342002	1.000000e+01	0.039617	1.028900e+04	3.729061e+07	49.891000
25%	1045.981591	4.792500e+03	0.898524	1.722518e+06	1.069201e+10	67.003750
50%	2469.420530	4.042000e+04	3.568216	1.022601e+07	5.634712e+10	72.413048
75%	2469.420530	7.501898e+05	5.603345	6.588462e+07	8.809829e+11	76.665218
max	17922.703790	3.308519e+07	32.693532	7.261847e+09	7.953110e+13	83.980488

- Percentiles are used to determine the labels.
- The standard deviation of life expectancy decreases from 7.959073 years to 7.605993 years over the period from 2012 to 2014.
- Mean life expectancy is increasing approximately from 70.86 years to 71.13 years.
- The increasing minimum value of life expectancy indicates a rising trend in global life expectancy.
- CO2 emissions increase from 1.257054e+06 KT to 1.286440e+06 KT while the total population and GDP increase according to the mean value.
- Energy use shows a decreasing trend from 2012 to 2013, and it increased again in 2014(Table 1, Table 2, and Table 3).

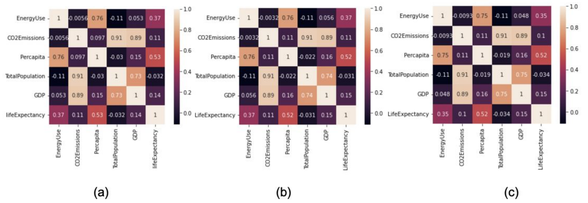
Table 4: Comparisons of coefficient correlations(r) by regression models using life expectancy as the target.

Models	2012	2013	2014
LE~CO2	0.0897	0.0892	0.0853
LE~CO2+GDP	0.1212	0.1318	0.1323
LE~CO2+Population	0.3225	0.3191	0.3248
LE~CO2+Population+GDP	0.318	0.3137	0.3195

-Pearson correlation (Divaris et al., 2012) :
- Range from -1 to 1
- 0.1-0.3: week correlation
- 0.3-0.5: moderate correlation
- 0.5 or greater: strong correlation

-No significant correlation between CO2 emissions and life expectancy in the period from 2012 to 2014 according to the result.

Figure 1: Correlations among variables for 2012, 2013, and 2014.



(a) (b) (c)

- The value of correlation for energy use on life expectancy is higher than the value of correlation for CO2 emissions on life expectancy
- The correlation between per capita and energy use is high, with a value of 0.76 which indicates that the more people, the higher the oil consumption.
- A strong positive correlation between the total population and CO2 emissions was found.

Table 5: Multiple linear regression models using CO2 emissions as target variable under different GDP categories.

	High GDP			Low GDP		
	2012	2013	2014	2012	2013	2014
R^2	0.9173	0.9022	0.9066	0.0624	0.2498	0.2001
R	0.9578	0.9498	0.9522	0.1519	0.3897	0.4473
Intercept	-1491012.0680	-1338832.8017	-1386381.6972	-48437.3494	-39384.1134	-18669.4327
Energy Use	516.9733	481.9333	499.0098	77.6320	93.1180	86.2048
GDP	1.9319e-07	1.8378e-07	1.7459e-07	10.0843e-05	9.1575e-05	8.4735e-05
Population	0.00276	0.00281	0.00283	-0.09125	-0.07815	-0.07588

Three regression models are computed for high GDP countries from 2012 to 2014 in the following:

CO2 emissions = -1491012.0680 + 516.9733*EnergyUse +1.9319e-07*GDP + 0.00276*TotalPopulation + e
CO2 emissions = -1338832.8017 + 481.9333*EnergyUse + 1.8378e-07*GDP + 0.00281*TotalPopulation + e
CO2 emissions = -1338832.8017 + 499.0098*EnergyUse + 1.7459e-07 *GDP +0.00283*TotalPopulation+ e

Table 6: ANOVA test for population and energy use from 2012 to 2014 on a 95% confidence interval.

	$F_{2,263}$	P-value	P-adjusted Medium-Large	P-adjusted Small-Large	P-adjusted Small-Medium
2012	7.729	0.000838	0.3571113	0.0007633	0.0115727
2013	6.899	0.0012	0.3779597	0.0010956	0.0144460
2014	7.057	0.00103	0.4089932	0.0010113	0.0113661

H_0: All population categories have equal mean energy use.
H_1: At least one population category has different mean energy use.

Table 7: t-TEST for life expectancy under high and low emission levels on a 95% confidence interval.

	x_{life_high}	x_{life_low}	$x_{life\_high} - x_{life\_low}$	t	$p-value$
2012	73.71620	65.90277	7.81343	6.192	3.499e-09
2013	73.72222	66.06713	7.65509	6.3543	1.567e-09
2014	73.80574	66.67800	7.12774	5.9981	9.008e-09

H_0: $\mu_{life\_high} \leq \mu_{life\_low}$
H_1: $\mu_{life\_high} > \mu_{life\_low}$

Conclusion

- GDP and total population strongly correlate with CO2 emissions.
- This project has not found a strong correlation between CO2 emissions and life expectancy
- Between high and low CO2 emissions countries, a decreasing trend of the value of the difference from 2012 to 2014 was found.
- Life expectancy is not only affected by a single factor. Different factors play their role simultaneously and will enhance the impact on life expectancy even though it increases year by year.
- Low GDP countries have an increasing trend in CO2 emissions, yet high GDP countries are still the largest emitters.
- The result did not show that the large population countries consume more oil than small population countries, it could affect by multiple factors, such as the industrial process.

References

Divaris, K., Vann F.W., Baker A. D., and Lee Y.J. (2012). Examining the Accuracy of Caregivers' Assessments of Young Children's Oral Health Status. The Journal of the American Dental Association 143, no. 11 (2012): 1237–47. https://doi.org/10.14219/jada.archive.2012.0071.

Gülden K. U. and Neşe G. (2013). A Study on Multiple Linear Regression Analysis,Procedia - Social and Behavioral Sciences,106, 234-240. https://doi.org/10.1016/j.sbspro.2013.12.027.