

Assignment 3

Raju Adhikari and Mario Alonso Rodriguez

11/11/2016

Merging databases

This is the third pair assignment of Introduction to Collaborative Social Science Data Analysis. Here, we gathered web based data from the World Bank API, as well as from the Emissions Database for Global Atmospheric Research (EDGAR) from the European Commission website. With both datasets we aim to answer our research question: Does income inequality affect the environment negatively? To operationalize this, we will use CO2 emissions per capita to measure effects on the environment (dependent variable) and the GINI coefficient for inequality (independent variable). We merged the two data sets, and conducted basic descriptive and inferential statistics to address this question.

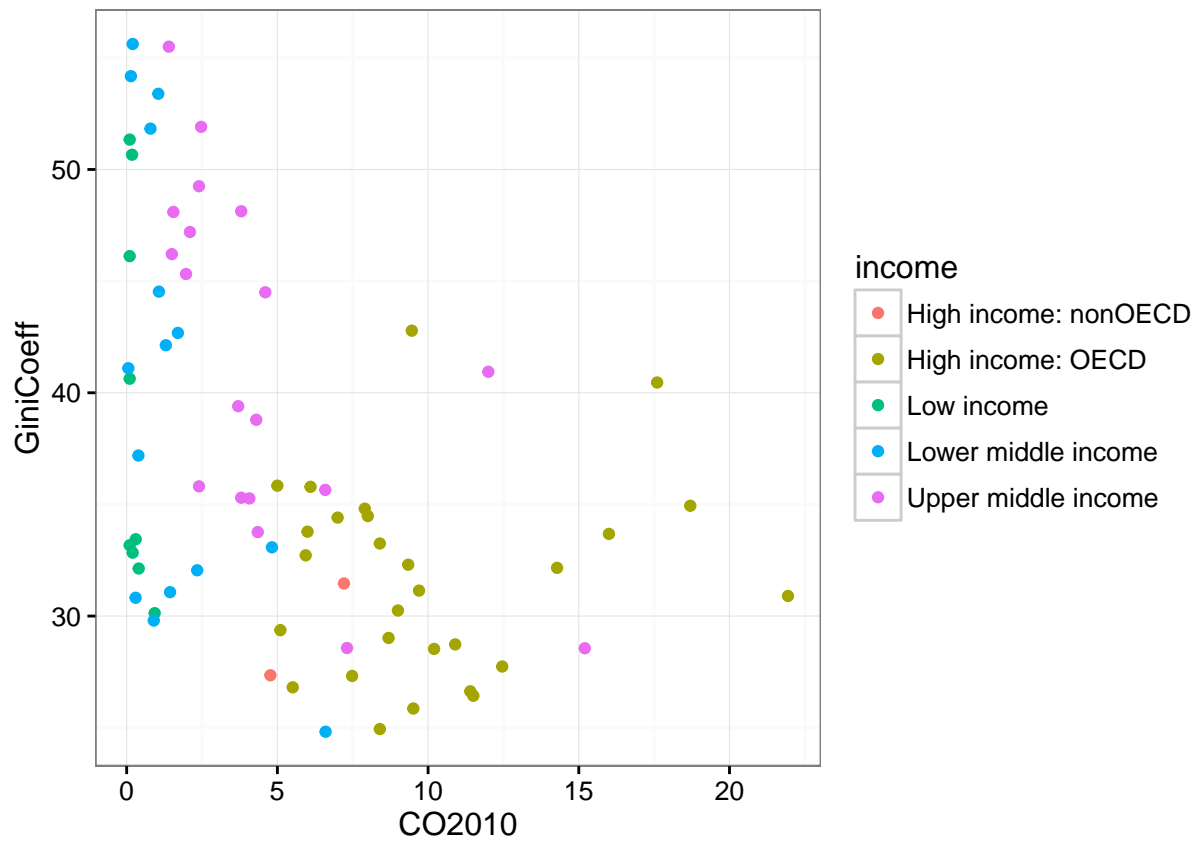
Descriptive statistics

A series of basic statistics show that the mean Gini is 36.81. The minimum value is 24.82 for the Ukraine, which means that it was the less unequal country in terms of income in 2010. The highest inequality was in Zambia with 55.62. The mean value of GDP per capita of our set of countries in 2012 was 19,222.2 USD dollars. The lowest value for this variable is 341.9 of Ethiopia and the highest is 103267.3 of Luxembourg. Finally, the CO2 emissions per capita have a mean of 4.873 metric tons in 2012. The lowest value in 2010 was 0.056 of Tuvalu, and the highest is 21.941 from Luxembourg.

```
##
## =====
## Statistic      N      Mean      St. Dev.      Min      Max
## -----
## GiniCoeff      73     36.805      8.403      24.820      55.620
## GDPperCapita   73  19,222.240  22,562.420  341.859  103,267.300
## CO2010         73     5.541      5.084      0.056      21.941
## -----
```

Inferential statistics

A plot shows that there might not be a linear relationship between inequality and CO2 emissions. We run a first regression model with only one explanatory variable and see that it is statistically significant and that actually, more equal countries (with lower GINI coefficients) are the ones that emit more CO2. A one unit increase in the Gini coefficients means a decrease in the emission of about 0.3 metric tons of CO2 annually. This could be to the fact that countries with less inequality are more developed. On the second model, we include a control variable of GDP per Capita to see if the relationship between our independent and dependent variable is not caused by this third variable. The second model shows that GDP per capita and the Gini coefficient are statistically significant. Nonetheless, our inequality variable is still negative, which means that countries with higher inequality actually have less CO2 emissions, even when controlling for income.



```
##
## =====
##                               Dependent variable:
##                               -----
##                               CO2010
##                               (1)           (2)
## -----
## GiniCoeff                -0.289***      -0.120**
##                          (0.063)        (0.052)
##
## GDPperCapita                                0.0001***
##                                              (0.00002)
##
## Constant                  16.180***      7.126***
##                          (2.380)        (2.154)
## -----
## Observations                73           73
## R2                        0.228          0.572
## Adjusted R2                0.217          0.560
## Residual Std. Error    4.497 (df = 71)    3.372 (df = 70)
## F Statistic            21.004*** (df = 1; 71) 46.819*** (df = 2; 70)
## =====
## Note:                        *p<0.1; **p<0.05; ***p<0.01
```