**Info for EU effect project**

**Title/Description**

* My project is titled: ‘The EU effect in context: does trade with the EU actually drive CO2 emission reduction in developing countries?

* I will be testing whether export dependence on the EU from developing countries (non-Annex 1) is causing CO2 emission reductions. In general developing country CO2 emission has been increasing however I want to see the whether the year on year increases are affected. In a previous similar study these used a logged version of all variable values.

**Codebook**

* Name: Country Name
* Year: Year
* Country\_Code: Three letter country code
* IPCC\_annex: all countries are non-Annex1. These countries are essentially a proxy for developing countries in international climate agreements
* CO2\_kt: CO2 in kilotons emitted by country
* SO2\_kt: SO2 in kilotons emitted by country
* PM10\_kt: PM10 in kilotons emitted by country
* Urban\_Pop\_%of\_total: Precent of total population living in urban areas
* Merchandise\_Exported\_to\_EU\_thousandsUSD: Value of goods exports to the EU in 000USD, current prices
* GDP\_cur\_USD: GDP in current years prices
* Electricity\_from\_coal\_%oftotal: electricity from coal as a percent of total electricity
* Electricity\_from\_renewables\_%oftotal: electricity from renewables as a percent of total electricity.
* Manufacturing\_value\_added\_%ofGDP: manufacturing as a percent of gdp in current prices.
* democ: factor variable, 1-10, how democratic a country in given year

I have not cleaned all my columns yet. Sometimes there are “..”, “NA” or “-88”. I also need to ensure which variables are coded as numeric, factor etc

**Hypothesis**

* 1: Greater export dependence on the EU from developing countries (Annex 2 countries) will be associated with lower CO2 emissions

* 2a: Greater export dependence on non-EU Annex 1 countries from developing countries will not be associated with lower CO2 emissions.
* 2b: Greater export dependence on non-EU and non-Annex 1 countries from developing countries will not be associated with lower CO2 emissions.

* 3a: Developing countries dependence on EU export markets will be associated with smaller reductions in SO2 emissions.
* 3b: Developing countries dependence on EU export markets will be associated with no reductions in PM10 emissions

I am still missing two columns of data that will be needed for hypothesis 2a and 2b. I am having to retrieve data from the [UNCTADSTAT](https://unctadstat.unctad.org/wds/ReportFolders/reportFolders.aspx?sCS_ChosenLang=en) so it takes some time- it’s a very old data centre.

**Analysis**

I would want to run a panel regression to test these hypotheses. I think I will be using regression for inference.

I tried to run a ggpairs to start with. It said I had reached the cardinality\_threshold parameter. I increased it but it was taking too long to run.

It would be good to talk about potential problems with the data or any weaknesses you spot, and what the approach would like for this sort of problem. I am comfortable with the methods and code we have used last semester, and I’m getting more comfortable with the Data Science modelling content week by week. I think this project requires different techniques though.