International Trade vs Pollution 09.13.2024

September 13, 2024

0.1 Starting the Analysis

In order to conduct a thorough analysis of how international trade affects pollution across the world, we require the use of several packages to both find and display relevant information. These include pandas for general data analysis tools and matplotlib, seaborn, and relevant plotly items to create graphs.

[1]:
```
import warnings
warnings.filterwarnings("ignore", message="A NumPy version >=")
```

[2]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
import plotly.graph_objects as go
from plotly.subplots import make_subplots
import plotly.io as pio
pio.renderers.default = 'notebook'
```

0.2 Initial Exploration

We start with the University of Gothenberg's Quality of Government (QoG) Institute's Basic Dataset (2024). To achieve a rudimentary understanding of the relationship between international trade and pollution, we must load the dataset and locate variables that could be used as operational definitions.

[3]:
```
basecross = pd.read_csv('qog_bas_cs_jan24.csv')
basecross.head()
```

[3]:
```
   ccode        cname  ccode_qog    cname_qog ccodealp  ccodecow  \
0      4  Afghanistan          4  Afghanistan      AFG     700.0
1      8      Albania          8      Albania      ALB     339.0
2     12      Algeria         12      Algeria      DZA     615.0
3     20      Andorra         20      Andorra      AND     232.0
4     24       Angola         24       Angola      AGO     540.0

        version  ajr_settmort  atop_ally  atop_number  …  wvs_imprel  \
0  QoGBasCSjan24      4.540098        1.0          1.0  …         NaN
1  QoGBasCSjan24           NaN        1.0          8.0  …    2.869328
```

```
2  QoGBasCSjan24      4.359270        1.0          9.0  …          NaN
3  QoGBasCSjan24           NaN        1.0          2.0  …     2.034930
4  QoGBasCSjan24      5.634790        1.0          8.0  …          NaN

   wvs_pmi12  wvs_psarmy  wvs_psdem  wvs_psexp  wvs_pssl  wvs_relacc  \
0        NaN         NaN        NaN        NaN       NaN         NaN
1        NaN    1.596485   3.849031   3.475513  1.744196         NaN
2        NaN         NaN        NaN        NaN       NaN         NaN
3   2.710393    1.336049   3.681363   2.635721  1.830491    1.751004
4        NaN         NaN        NaN        NaN       NaN         NaN

   wvs_satfin  wvs_subh  wvs_trust
0         NaN       NaN        NaN
1         NaN  3.488758   0.027857
2         NaN       NaN        NaN
3    6.561316  4.089642   0.255744
4         NaN       NaN        NaN

[5 rows x 338 columns]
```

0.3 Variables of Interest

From the QOG Basic Dataset 2024 codebook, multiple variables pertaining to the relationship between international trade and pollution can be ascertained. Two such variables may be used here; international trade may be operationally defined by the variable for economic globalization (dr_eg), which ranks countries on a scale of 1-100 based on the flow of goods and services to other countries. In turn, pollution can be defined by the variable for the Environmental Performance Index (EPI) score (epi_epi), which ranks countries on a scale of 0-100 based on 32 different metrics of environmental health.

[4]:
```
basecross.dr_eg
```

[4]:
```
0      28.830755
1      63.483410
2      33.879074
3      70.329048
4      44.303589
         …
189    51.683723
190    27.821911
191    46.810032
192    40.550980
193    59.640228
Name: dr_eg, Length: 194, dtype: float64
```

[5]:
```
basecross.epi_epi
```

[5]:
```
0      43.599998
1      47.099998
2      29.600000
3            NaN
4      30.500000
         …
189    38.200001
190    46.400002
191    36.400002
192          NaN
193    38.400002
Name: epi_epi, Length: 194, dtype: float64
```

[6]:
```
fig1 = px.scatter(
    data_frame = basecross,
    x = 'dr_eg',
    y = 'epi_epi',
    title = "Economic Globalization vs EPI Score by Country",
    labels={
        'dr_eg': 'Economic Globalization',
        'epi_epi': 'EPI Score'
        },
    trendline = 'ols',
    hover_data = ['cname']
)
fig1
```

0.4 Initial Findings and Re-Testing

Initial findings from the basic dataset show a positive correlation between economic globalization and EPI score in which the greater a country's degree of economic globalization, the greater their performance on the EPI. This suggests that the more international trade a country conducts, the less polluted it is.

However, this seems somewhat counter-intuitive owing to recent research on this connection. According to the Grantham Research Institute on Climate Change and the Environment (2023), international trade is responsible for nearly 30% of global CO_2 emissions, owing to the environmental cost of transporting goods across borders.

How could it be, then, that we are witnessing a positive correlation? It may be that we have chosen a poor metric for trade in economic globalization. For instance, the portion that a country's trade contributes to their gdp (wdi_trade) would serve as a more concrete definition for international trade. Comparing the outputs of using different variables side by side is possible with a scatterplot matrix.

[7]:
```
basecross.wdi_trade
```

[7]:
```
0            NaN
1      59.829731
2      45.330509
3            NaN
4      55.375816
         …
189    61.839191
190          NaN
191    77.483597
192    49.303493
193    79.325485
Name: wdi_trade, Length: 194, dtype: float64
```

[8]:
```
fig2 = px.scatter_matrix(
    data_frame = basecross,
    dimensions = ['dr_eg', 'wdi_trade', 'epi_epi'],
    title = "EPI Score vs Economic Globalization vs Trade % of GDP by Country",
    labels = {
        'dr_eg':'Economic Globalization',
        'wdi_trade':'Trade % of GDP',
        'epi_epi':'EPI Score'
        },
    template = 'seaborn',
    hover_data = 'cname'
)
fig2.update_traces(diagonal_visible = False)
fig2.update_layout(width = 700, height = 700)
fig2
```

0.5 Creating More Specific Analyses

Using the scatterplot matrix, we can view the intersections of our three variables of interest. Unlike how prior research had demonstrated, EPI Score seemed to be positively correlated with both Economic Globalization and the Trade % of GDP. It may be, then, that the other items that comprise a country's EPI score outweigh the negative contribution of international trade on atmospheric pollution. In this sense, international trade may benefit the environment in certain countries except for when it comes to greenhouse emissions.

To find out, and observe the role of time in our analyses, we can take the time-series version of the QOG's Basic Dataset (2024), which takes critical observations of countries throughout the years, and the QOG's Environmental Indicators Dataset (2023), which has variables that allow for more in-depth analyses of pollution.

[9]:
```
basetime = pd.read_csv('qog_bas_ts_jan24.csv')
basetime.head(5)
```

[9]:
```
   ccode        cname  year  ccode_qog    cname_qog ccodealp  ccodecow  \
0      4  Afghanistan  1946          4  Afghanistan      AFG     700.0
```

```
1      4  Afghanistan  1947          4  Afghanistan      AFG     700.0
2      4  Afghanistan  1948          4  Afghanistan      AFG     700.0
3      4  Afghanistan  1949          4  Afghanistan      AFG     700.0
4      4  Afghanistan  1950          4  Afghanistan      AFG     700.0

        version        cname_year ccodealp_year  …  wdi_trade  \
0  QoGBasTSjan24  Afghanistan 1946         AFG46  …        NaN
1  QoGBasTSjan24  Afghanistan 1947         AFG47  …        NaN
2  QoGBasTSjan24  Afghanistan 1948         AFG48  …        NaN
3  QoGBasTSjan24  Afghanistan 1949         AFG49  …        NaN
4  QoGBasTSjan24  Afghanistan 1950         AFG50  …        NaN

   wdi_unempfilo  wdi_unempilo  wdi_unempmilo  wdi_unempyfilo  wdi_unempyilo  \
0            NaN           NaN            NaN             NaN            NaN
1            NaN           NaN            NaN             NaN            NaN
2            NaN           NaN            NaN             NaN            NaN
3            NaN           NaN            NaN             NaN            NaN
4            NaN           NaN            NaN             NaN            NaN

   wdi_unempymilo  wdi_wip  who_sanittot  whr_hap
0             NaN      NaN           NaN      NaN
1             NaN      NaN           NaN      NaN
2             NaN      NaN           NaN      NaN
3             NaN      NaN           NaN      NaN
4             NaN      NaN           NaN      NaN

[5 rows x 252 columns]
```

[10]:
```python
enviro = pd.read_csv('qog_ei_ts_sept21.csv')
enviro.head(5)
```

[10]:
```
   Unnamed: 0        cname  ccode  year    cname_qog  ccode_qog ccodealp  \
0           1  Afghanistan    4.0  1946  Afghanistan          4      AFG
1           2  Afghanistan    4.0  1947  Afghanistan          4      AFG
2           3  Afghanistan    4.0  1948  Afghanistan          4      AFG
3           4  Afghanistan    4.0  1949  Afghanistan          4      AFG
4           5  Afghanistan    4.0  1950  Afghanistan          4      AFG

  ccodealp_year  ccodecow  ccodevdem  … wdi_precip wdi_tpa  wvs_ameop  \
0         AFG46     700.0       36.0  …        NaN     NaN        NaN
1         AFG47     700.0       36.0  …        NaN     NaN        NaN
2         AFG48     700.0       36.0  …        NaN     NaN        NaN
3         AFG49     700.0       36.0  …        NaN     NaN        NaN
4         AFG50     700.0       36.0  …        NaN     NaN        NaN

   wvs_ceom  wvs_deop  wvs_epmip  wvs_epmpp  wvs_imeop  wvs_pedp  wvs_ploem
0       NaN       NaN        NaN        NaN        NaN       NaN        NaN
```

```
1        NaN       NaN        NaN        NaN        NaN       NaN        NaN
2        NaN       NaN        NaN        NaN        NaN       NaN        NaN
3        NaN       NaN        NaN        NaN        NaN       NaN        NaN
4        NaN       NaN        NaN        NaN        NaN       NaN        NaN

[5 rows x 415 columns]
```

0.6 Merging Datasets

From a brief analysis of both datasets, it can be seen that there are commonalities in the data that can be used to perform an outer join. We can join specifically by country name (cname) and year of observation (year).

[11]:
```python
result = pd.merge(basetime, enviro, how = 'outer', on = ['cname', 'year'])
result.head(100)
```

[11]:
```
    ccode_x        cname  year  ccode_qog_x  cname_qog_x ccodealp_x  \
0       4.0  Afghanistan  1946          4.0  Afghanistan        AFG
1       4.0  Afghanistan  1947          4.0  Afghanistan        AFG
2       4.0  Afghanistan  1948          4.0  Afghanistan        AFG
3       4.0  Afghanistan  1949          4.0  Afghanistan        AFG
4       4.0  Afghanistan  1950          4.0  Afghanistan        AFG
..      …            …   …           …            …          …
95      8.0      Albania  1963          8.0      Albania        ALB
96      8.0      Albania  1964          8.0      Albania        ALB
97      8.0      Albania  1965          8.0      Albania        ALB
98      8.0      Albania  1966          8.0      Albania        ALB
99      8.0      Albania  1967          8.0      Albania        ALB

    ccodecow_x      version_x      cname_year_x ccodealp_year_x  …  \
0        700.0  QoGBasTSjan24  Afghanistan 1946           AFG46  …
1        700.0  QoGBasTSjan24  Afghanistan 1947           AFG47  …
2        700.0  QoGBasTSjan24  Afghanistan 1948           AFG48  …
3        700.0  QoGBasTSjan24  Afghanistan 1949           AFG49  …
4        700.0  QoGBasTSjan24  Afghanistan 1950           AFG50  …
..         …             …                …              …  …
95       339.0  QoGBasTSjan24      Albania 1963           ALB63  …
96       339.0  QoGBasTSjan24      Albania 1964           ALB64  …
97       339.0  QoGBasTSjan24      Albania 1965           ALB65  …
98       339.0  QoGBasTSjan24      Albania 1966           ALB66  …
99       339.0  QoGBasTSjan24      Albania 1967           ALB67  …

    wdi_precip  wdi_tpa  wvs_ameop  wvs_ceom  wvs_deop  wvs_epmip  wvs_epmpp  \
0          NaN      NaN        NaN       NaN       NaN        NaN        NaN
1          NaN      NaN        NaN       NaN       NaN        NaN        NaN
2          NaN      NaN        NaN       NaN       NaN        NaN        NaN
3          NaN      NaN        NaN       NaN       NaN        NaN        NaN
```

```
4          NaN      NaN        NaN       NaN       NaN        NaN        NaN
..           …        …          …         …         …          …          …
95         NaN      NaN        NaN       NaN       NaN        NaN        NaN
96         NaN      NaN        NaN       NaN       NaN        NaN        NaN
97         NaN      NaN        NaN       NaN       NaN        NaN        NaN
98         NaN      NaN        NaN       NaN       NaN        NaN        NaN
99      1485.0      NaN        NaN       NaN       NaN        NaN        NaN

    wvs_imeop  wvs_pedp  wvs_ploem
0         NaN       NaN        NaN
1         NaN       NaN        NaN
2         NaN       NaN        NaN
3         NaN       NaN        NaN
4         NaN       NaN        NaN
..        …         …          …
95        NaN       NaN        NaN
96        NaN       NaN        NaN
97        NaN       NaN        NaN
98        NaN       NaN        NaN
99        NaN       NaN        NaN

[100 rows x 665 columns]
```

0.7 International Trade vs CO2 emissions

Now that the datasets are merged, numerous other comparisons can be made using the various metrics of pollution within the environmental dataset. To specifically evalute the impact of international trade on CO2 emissions, we can use a country's total CO2 emissions in kilotons (edgar_co2t) adn compare it with the proportion of GDP produced by trade. Additionally, now that we have longitudinal data, we can conduct this same analysis for all years between 2000 and now.

[12]:
```
fig3 = px.scatter(
    data_frame = result.query('year >= 2000 & year <=2023'),
    x = 'wdi_trade',
    y = 'edgar_co2t',
    animation_frame = "year",
    animation_group = "cname",
    title = "Trade % of GDP vs Total CO2 emissions (kt) from 1995-2020",
        labels={
        'wdi_trade': 'Trade % of GDP',
        'edgar_co2t': 'Total CO2 emissions (kt)'
        },
    trendline = 'ols',
    hover_data = ['cname']
)
fig3
```

0.8 Analyzing International Trade vs CO2 emissions

Despite recent research showing that international trade is responsible for 30% of global CO2 emissions, it seems as though there is not much of a relationship. Given that the least squares line stays mostly horizontal, much like the data for each country throughout the years, it is difficult to say there is much correlation. Most countries appear to keep consistent in both their trading and yearly CO2 emissions, save for some outliers like China. Even zooming into the bottom half of the graph where most of the data is located does not yield much in terms of possible correlation. Therefore, this graph illustrates a lack of connection between international trade and pollution in terms of CO2 emissions.

0.9 International Trade vs Drinking Water Quality

Another important metric of pollution is water scarcity. According to Zhong et al. (2023), the degree of international trade can actually reduce water scarcity for higher-income countries, but exacerbate it for lower-income countries. This is perhaps due to the role of water in manufacturing, the way in which factories pollute water, and how transportation by ship disturbs water. We can assign a color value to countries' GDP (gle_cgdpc) to test this phenomenon.

[13]:
```
fig4 = px.scatter(
    data_frame = result.query('year >= 2000 & year <=2023'),
    x = 'wdi_trade',
    y = 'epi_uwd',
    animation_frame = "year",
    animation_group = "cname",
    title = "Economic Globalization vs Drinking Water Quality by Country and␣
 ↪GDP (1995-2020)",
        labels={
        'wdi_trade': 'Trade % of GDP',
        'epi_uwd': 'Drinking Water Quality'
        },
    trendline = 'ols',
    color = 'gle_cgdpc',
    hover_data = ['cname']
)
fig4
```

0.10 Analyzing International Trade vs Drinking Water Quality

In the years for which GDP was observed, and thus color could be assigned to the graph, it is clearly visible that lower-income countries had worse drinking water quality than higher-income countries, but the relationship between water quality and international trade is more nebulous. Although the least squares line indicates a positive correlation, the points themseelves do not align in a discernable pattern to suggest as much. An interesting outlier in this regard is Qatar, which seems to have the highest drinking water quality of all despite being its middling GDP and trade.

It would also be valuable to identify this relationship in terms of imports and exports. Differentiating between a country's total sum of imports (gle_imp) and exports (gle_exp) in millions of dollars can illustrate the dynamic of water pollution between countries who produce goods for the international market and countries who buy them. However, we can only compare these statistics for the year 2000, given that total import and export were only observed then.

[14]:
```
fig5 = px.scatter(
    data_frame = result.query('year == 2000'),
    x = 'gle_imp',
    y = 'epi_uwd',
    title = "Total Import vs Drinking Water Quality by country (2000)",
        labels={
        'gle_imp': 'Total Import (Millions of USD)',
        'epi_uwd': 'Drinking Water Quality'
        },
    trendline = 'ols',
    color = 'gle_cgdpc',
    hover_data = ['cname']
)
fig5
```

[15]:
```
fig6 = px.scatter(
    data_frame = result.query('year == 2000'),
    x = 'gle_exp',
    y = 'epi_uwd',
    title = "Total Export vs Drinking Water Quality by country (2000)",
        labels={
        'gle_exp': 'Total Export (Millions of USD)',
        'epi_uwd': 'Drinking Water Quality'
        },
    trendline = 'ols',
    color = 'gle_cgdpc',
    hover_data = ['cname']
)
fig6
```

0.11 Conclusion

In conclusion, it would appear that international trade is overall negatively correlated with pollution. However, when it comes to CO2 emissions and drinking water quality specifically, the relationship is more nebulous. In order to discern which factors of pollution international trade benefits and which it exacerbates, in-depth analysis is requried for all 32 items of the EPI in relation to countries' capacity for international trade. This will be a necessary step in determining what actions need to be taken in order to most efficiiently minimize the environmental toll caused by the global market.

Dahlberg, Stefan, Aksel Sundström, Sören Holmberg, Bo Rothstein, Natalia Alvarado Pachon, Cem Mert Dalli, Rafael Lopez Valverde & Paula Nilsson. 2024. The Quality of Government Basic Dataset, version Jan24. University of Gothenburg: The Quality of Government Institute, https://www.gu.se/en/quality-government doi:10.18157/qogbasjan24

Grantham Research Institute on Climate Change and the Environment. 2023, June 12. How does trade contribute to climate change and how can it advance climate action?. London School of Economics and Political Science. https://www.lse.ac.uk/granthaminstitute/explainers/how-does-trade-contribute-to-climate-change-and-how-can-it-advance-climate-action/

Povitkina, Marina, Natalia Alvarado Pachon & Cem Mert Dalli. 2021. The Quality of Government Environmental Indicators Dataset, version Sep21. University of Gothenburg: The Quality of Government Institute, https://www.gu.se/en/quality-government

Zhong, R., Chen, A., Zhao, D., Mao, G., Zhao, X., Huang, H., & Liu, J. (2023). Impact of international trade on water scarcity: An assessment by improving the Falkenmark indicator. Journal of Cleaner Production, 385, 135740. https://doi.org/10.1016/j.jclepro.2022.135740