Mathematics SL: Applications and Interpretations IA

An Investigation into Net Carbon Emissions and Development Indices

What has a stronger relationship to net Carbon emissions: Human Development Index or the Gender Inequality Index?

Page Count: 20

CANDIDATE CODE: 0000

Contents

Introduction

The idea of "development" in terms of human geography can be defined in myriad different ways – from life expectancy to educational attainment. Most high-income countries, defined as having a "gross national income per capita of US $12,376" (World Population Review, 2020), are regarded as being more developed in almost all these counterparts compared to lower-income ones. I've noticed that most high-income nations tend to be large emitters of carbon dioxide, though, and it led me to wonder if countries could really develop both economically and socially without polluting the environment and using up fossil fuels – a topic of heated debate (OMOJU, 2014). This is a topic of particular interest to me as a keen Geographer because I have always questioned whether development on a global scale could really be sustainable. Although correlation is not causation, I believe this investigation will help to reveal if the variables follow a similar pattern.

In order to do this, I decided to compare net Carbon Dioxide emissions (shortened to Carbon from now on, though there are other sources of Carbon) with the Human Development Index (a composite index made up of separate income, quality of life and qualifications indices). However, using just the Human Development Index (HDI) I realised another important indicator of social development was missing – gender equality (United Nations Development Programme, 2020) , which could be measured using the Gender Inequality Index, GII. This led me to the aim of this investigation: to explore which one of either HDI or GII (as the independent variables, x_1, x_2) had greater correlations to net Carbon emissions (the dependent variable, y).

Personally, I predict that net carbon dioxide emissions will generally increase as HDI increases – this is because I expect more developed countries to usually have a high transport, imported goods, road usage, energy-intensive tertiary economic sector activity. I expect any relationship between net carbon dioxide emissions and the gender inequality index to be less clear-cut and have more exceptions to the general trend, because I anecdotally know some countries are very developed with high per capita Carbon emissions, but have less gender equality – such as Iran. Indeed, I chose to compare Gender Inequality because I feel it would be interesting to see if there were any major outliers. To begin with I needed to collect raw data and I decided to use sampling to obtain an appropriate number of data points.

Sampling – Strategy and Distribution

Instead of trying to find data for *all* the "countries" (loosely defined) of the world, where many sources have different opinions, I decided to use random sampling to generate 50 countries that all had HDI (United Nations Human Development Reports , n.d.), GII (United Nations Human Development Reports , n.d.) and net Carbon emissions (EDGAR, 2017) data readily available. The data for each variable was collected for the same year, 2017 (otherwise comparisons would be unrepresentative). The major advantage of random sampling is that it removes selection bias that might be present in convenience sampling, as well as being relatively simplistic to carry out (each country was assigned a number and a random number generator was used to pick 50 of them). When I conducted the test, though, the nature of differences in data availability meant I had to do the sampling multiple times – some island states did not have the necessary data and the sources sometimes combined data for two "countries" that might have been regarded as one "country" by the other. However, as I repeated the entire sampling procedure and it was still random, this still meant all the countries available still had an equal probability of being chosen.

Next, I wanted to find out whether there were any major outliers to my data samples – if there were, they could affect the validity of any further conclusions because of their potential to skew results in one direction. “A convenient definition of an outlier is a point which falls more than 1.5 times the interquartile range above the third quartile or below the first quartile” (Renze, 1999-2020). A good way to represent this visually is through the use of box plots (generated using GeoGebra). Outliers will lie outside the “whiskers”.

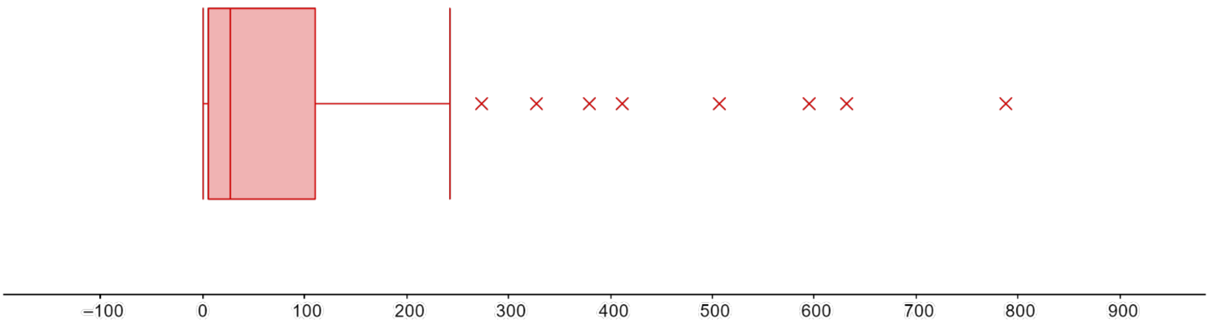


Figure 1 - Net Carbon Emissions (Megatonnes)

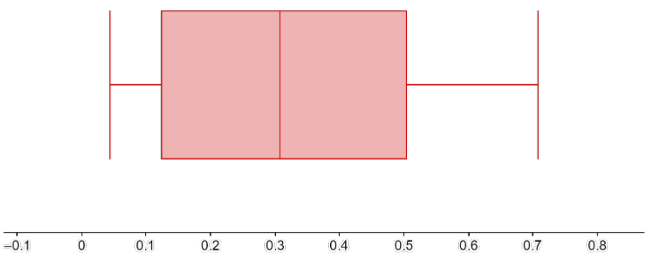


Figure 3 – GII (Relative Index Units)

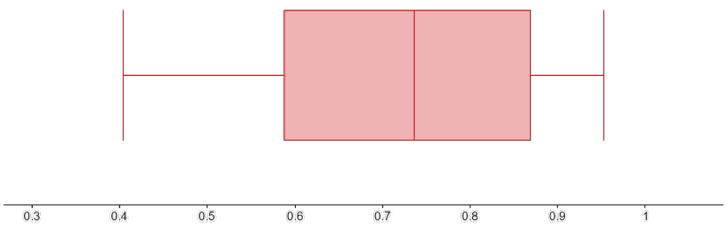


Figure 2 - HDI (Relative Index Units)

Firstly, it is apparent that the data for figure 1 is considerably skewed towards a lower value and, to a lesser extent, skewed in the same direction for GII, but in the opposite direction for HDI – none of the data sets strictly follow a normal distribution and this tells us we cannot simply model each variable in this way and gives us a reason to instead compare how one variable affects the other.

Moreover, it is clear that in figures 2 and 3 there are no outliers. However, there are multiple outliers that all exceed the upper quartile in figure 1. We can confirm this mathematically using interquartile range and upper/lower quartiles. If the data is ordered in ascending order for each variable: $Q_1 = \frac{1}{4}(n+1)^{\text{th}}$ term and $Q_3 = \frac{3}{4}(n+1)$th term. Here n, the number of data points, is 50 for all sets so $Q_3 = \frac{3\times 51}{4} = 38.25^{\text{th}}$ term and $Q_1 = \frac{51}{4} = 12.75^{\text{th}}$ term. These can be rounded up to the 39th and 13th term, respectively. The interquartile range, IQR, can then be calculated by subtracting the values of these two terms: $IQR = Q_3 - Q_1$. Adding and subtracting $1.5 \times IQR$ from the value of the upper and lower quartile, respectively, will give the maximum and minimum values to the data before it can be considered an outlier.

Figure 1 (Units in Mton):

39th term: Philippines, 143.420

13th term: Uganda, 5.380

IQR (39th – 13th term): 138.040

39th term + $1.5 \times IQR$ = 350.260

13th term - $1.5 \times IQR$ = -201.680

Figure 2 :

39th term: Greece, 0.870

13th term: Zambia, 0.588

IQR (39th – 13th term): 0.282

39th term + $1.5 \times IQR$ = 1.293

13th term - $1.5 \times IQR$ = 0.165

Figure 3:

39th term: Iraq, 0.506

13th term: Croatia, 0.124

IQR (39th – 13th term): 0.382

39th term + $1.5 \times IQR$ = 1.079

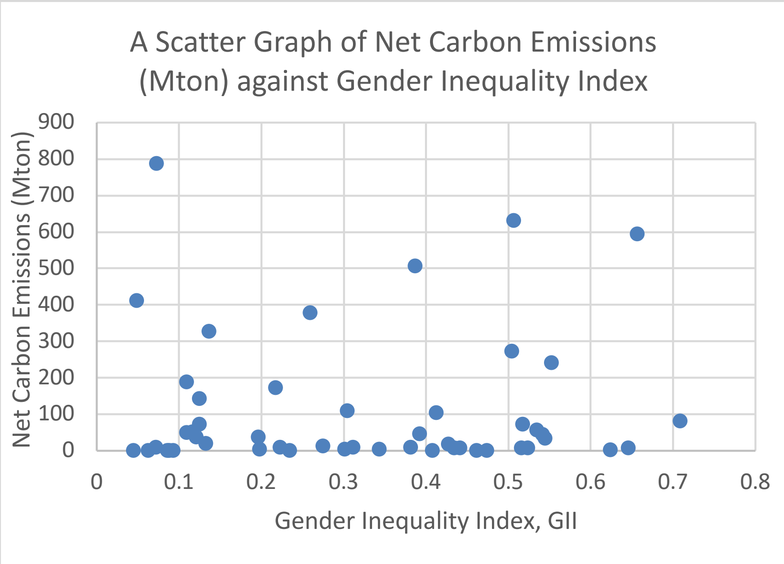
13th term - $1.5 \times IQR$ = -0.449

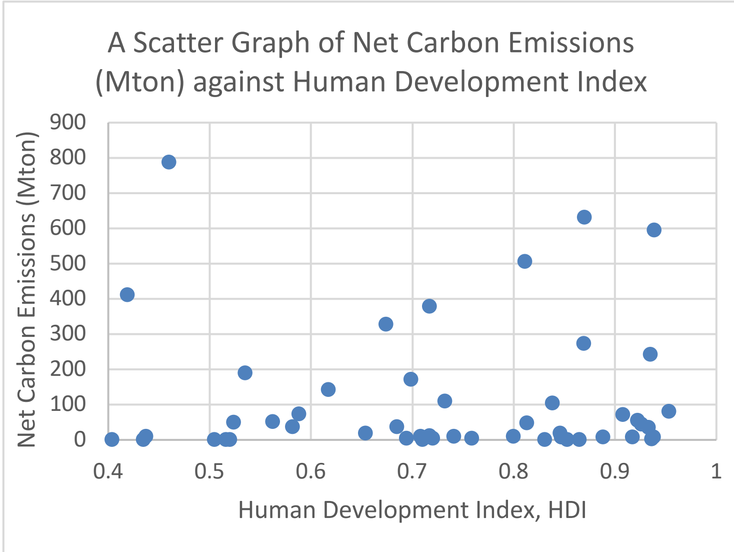
The calculations for figure 2 and 3 confirm there are no outliers outside the needed range, and the figure 1 calculations indicate countries emitting above 350.26 Mton, of which there are 6, are indeed outliers to the rest of the Carbon dioxide emissions data. The box plot shows 8 outliers, though, and I suspect this is because I rounded up the upper quartile to be the 39th term when in reality it could be considered somewhere between the 38th and 39th (which had a large difference).

I believe determining where the outliers are or, indeed, whether they were present in the data sets was a useful thing to do – they can cause some statistical values, such as the mean, to be skewed in a way that can be considered unrepresentative of the entire set (it would be vastly different if the mean was re-calculated excluding the outliers) and the box plots showing the skewed distributions both help us to determine which tests can and can't be done/altered to draw meaningful conclusions from the data.

How Strong are the Relationships?

Now that I have evaluated the sampling, I will construct scatter graphs to visually show the relationship between the variables we are investigating as well as getting an idea for the strength and direction (positive or negative) of the correlation if there is any. Again, I predict that generally countries with higher HDI values and lower GII (i.e. more gender equality) will have higher net carbon emissions (purely based on anecdotal observations) and expect HDI-Carbon to have a stronger trend than GII-Carbon. Using the previous section, we can predict the outliers for net Carbon emissions will result in a series of data points abnormally high on both of the y-axes.





Generally, it seems that there is a slight positive correlation between HDI and Net Carbon emissions. However, although it is clear the strength of correlation between GII and Net Carbon Emissions is undeniably weaker it is unclear whether any correlation present is negative or positive. I think the scatter graphs were useful to visually show the relationships, but now I think we need a more mathematical determination of the strength of correlation – the Pearson's Product Moment Correlation Co-efficient (PMCC).

$Country$	HDI (x)	$Total\ CO2$ Emissions (Mton) (y)	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})(y-\bar{y})$	$(x-\bar{x})^2$	$(y-\bar{y})^2$
Chad	0.404	0.71	-0.3280	-111.6624	36.6230	0.1076	12468.4916
Sierra Leone	0.419	1.40	-0.3130	-110.9724	34.7321	0.0980	12314.8736
Liberia	0.435	1.20	-0.2970	-111.1724	33.0160	0.0882	12359.3025
Mozambique	0.437	9.59	-0.2950	-102.7824	30.3188	0.0870	10564.2218
Gambia	0.46	0.58	-0.2720	-111.7924	30.4053	0.0740	12497.5407
Senegal	0.505	10.37	-0.2270	-102.0024	23.1525	0.0515	10404.4896
Uganda	0.516	5.38	-0.2160	-106.9924	23.1082	0.0467	11447.3737
Lesotho	0.520	0.71	-0.2120	-111.6624	23.6702	0.0449	12468.4916
Rwanda	0.524	1.10	-0.2080	-111.2724	23.1424	0.0433	12381.5470
Zimbabwe	0.535	12.81	-0.1970	-99.5624	19.6118	0.0388	9912.6715
Pakistan	0.562	189.31	-0.1700	76.9376	-13.0779	0.0289	5919.3943
Cambodia	0.582	10.45	-0.1500	-101.9224	15.2863	0.0225	10388.1756
Zambia	0.588	5.050	-0.1440	-107.3224	15.4523	0.0207	11518.0975
Honduras	0.617	10.05	-0.1150	-102.3224	11.7650	0.0132	10469.8735
Guyana	0.654	1.86	-0.0780	-110.5124	8.6178	0.0061	12212.9906
El Salvador	0.674	7.77	-0.0580	-104.6024	6.0648	0.0034	10941.6621
Iraq	0.685	172.63	-0.0470	60.2576	-2.8309	0.0022	3630.9784
Vietnam	0.694	242.46	-0.0380	130.0876	-4.9407	0.0014	16922.7837
Philippines	0.699	143.42	-0.0330	31.0476	-1.0239	0.0011	963.9535

Belize	0.708	0.44	-0.0240	-111.9324	2.6841	0.0006	12528.8622
Uzbekistan	0.710	104.08	-0.0220	-8.2924	0.1823	0.0005	68.7639
Maldives	0.717	0.85	-0.0150	-111.5224	1.6706	0.0002	12437.2457
Botswana	0.717	7.68	-0.0150	-104.6924	1.5683	0.0002	10960.4986
Suriname	0.720	2.17	-0.0120	-110.2024	1.3202	0.0001	12144.5690
Jamaica	0.732	7.89	0.0000	-104.4824	-0.0021	0.0000	10916.5719
Mongolia	0.741	18.40	0.0090	-93.9724	-0.8476	0.0001	8830.8120
Brazil	0.759	506.90	0.0270	394.5276	10.6601	0.0007	155652.0272
Kazakhstan	0.800	273.62	0.0680	161.2476	10.9681	0.0046	26000.7885
Romania	0.811	81.08	0.0790	-31.2924	-2.4727	0.0062	979.2143
Bulgaria	0.813	47.47	0.0810	-64.9024	-5.2584	0.0066	4212.3215
Croatia	0.831	19.52	0.0990	-92.8524	-9.1942	0.0098	8621.5682
Hungary	0.838	51.60	0.1060	-60.7724	-6.4431	0.0112	3693.2846
Bahrain	0.846	34.94	0.1140	-77.4324	-8.8288	0.0130	5995.7766
Latvia	0.847	7.820	0.1150	-104.5524	-12.0256	0.0132	10931.2044
Portugal	0.847	56.52	0.1150	-55.8524	-6.4241	0.0132	3119.4906
Saudi Arabia	0.853	631.74	0.1210	519.3676	62.8539	0.0147	269742.7039
Poland	0.865	327.44	0.1330	215.0676	28.6083	0.0177	46254.0726
Cyprus	0.869	7.53	0.1370	-104.8424	-14.3655	0.0188	10991.9288
Greece	0.870	73.14	0.1380	-39.2324	-5.4149	0.0191	1539.1812
Czechia	0.888	110.41	0.1560	-1.9624	-0.3062	0.0243	3.8510
Austria	0.908	72.46	0.1760	-39.9124	-7.0254	0.0310	1592.9997
New Zealand	0.917	37.35	0.1850	-75.0224	-13.8806	0.0342	5628.3605
UK	0.922	379.38	0.1900	267.0076	50.7368	0.0361	71293.0585
Canada	0.926	595.02	0.1940	482.6476	93.6433	0.0376	232948.7058
Sweden	0.933	45.14	0.2010	-67.2324	-13.5151	0.0404	4520.1956
Iceland	0.935	4.07	0.2030	-108.3024	-21.9876	0.0412	11729.4099
Germany	0.936	787.95	0.2040	675.5776	137.8313	0.0416	456405.0936
Ireland	0.938	37.76	0.2060	-74.6124	-15.3716	0.0424	5567.0102
Australia	0.939	411.57	0.2070	299.1976	61.9399	0.0429	89519.2039
Norway	0.953	49.83	0.2210	-62.5424	-13.8231	0.0489	3911.5518

Pearson's Moment Correlation Coefficient

The full table is shown above, but the method is clarified below. Correlation is defined as the measure of strength of the linear association between two variables. The PMCC calculation gives us an *r-value*, which is a more concrete (less subjective) indicator of correlation than qualitative descriptors of the scatter graph.

It is calculated by substituting appropriate values into this equation:

$$r = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2\sum(y-\bar{y})^2}}$$

E.g. Let HDI be x and let net Carbon emissions be y. $\bar{x}$ and $\bar{y}$ are the mean values of the independent and dependent sets, respectively. So, for example, the $(x - \bar{x})$ for Liberia is its HDI subtract the mean HDI of the 50 sampled countries:

0.435 – 0.7319 = -0.2969 = -0.297 (3sf).

$\bar{x} =$ 0.73198

$\bar{y} =$ 112.3724

$\sum(x - \bar{x})(y - \bar{y})$ = 620.5736324 (numerator)

$\sum(x - \bar{x})^2$ = 1.35066

$\sum(y - \bar{y})^2$ = 1698527.239

$$\sqrt{\sum(x - \bar{x})^2 \sum(y - \bar{y})^2} = \sqrt{\sum(1.35066)(1698527.239)}$$

$$= 1514.642287 \text{ (denominator)}$$

$$r_{HDI\ \&\ Carbon} = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2 \sum(y-\bar{y})^2}}$$

$$= \frac{620.5736324}{1514.642287}$$

$$= 0.409716299 = 0.410\ (3sf)$$

The same process was used between GII and Net Carbon Emissions, resulting in $r_{GII\ \&\ Carbon} =$ -0.35571 = -0.356 (3sf). This tells us this correlation is very slightly negative.

As both (absolute) r values > 0, we do not accept the null hypothesis. It is clear to see that the correlation between HDI and net Carbon Dioxide Emissions is indeed stronger than between GII and Net Carbon Emissions – though both are still very weak: a value much closer to -1 to +1 would indicate a much stronger and significant relationship.

I think performing this calculation was valuable because now we can actually tell the direction of correlation between GII and Net Carbon Emissions; and we have determined we probably cannot expect to get a regression line and predict accurate values with interpolation because the correlation is so weak. A limitation here is that the means of the actual values were used, but the data is not truly normally distributed - this decreases the validity of the test. A non-parametric alternative like Spearman's could be used, though I believe the coefficient will have a similar value to the Pearson's r-value so wouldn't be very useful.

Are the variables independent of each other?

I think it is appropriate to do a Chi squared test for independence here – this will help us determine whether the "outcome" or value of one variable (e.g. HDI/GII) *does not* affect the other (e.g. Net Carbon Emissions). In this way we can assume whether one variable *is* likely to actually have a significant effect on the other:

Chi-squared Testing for Independence

$H_0 =$ "HDI and Net Carbon Dioxide Emissions are Independent"

$H_1 =$ "HDI and Net Carbon Dioxide Emissions are Not Independent

In order to accept/reject the null hypothesis, we must carry out a few steps – I shall illustrate this with data for the using net Carbon Dioxide emissions and HDI, though exactly the same method was followed where GII was the independent variable, x.

1) Split the data into categories. In order to do sufficient processing, I decided to split the independent variable data (HDI) into 2 categories [high or low] and split the Net Carbon emissions into 3 categories [high, medium or low]. This was done by ordering the data from each variable into ascending order and

determining cut-off points where the data would roughly be split into half or thirds and/or if there were large differences in two adjacent data points nearby these values e.g. for Carbon emissions data: to illustrate the difference between "medium" and "high" here is approximately 8Mton.

72.46 Mton	"Medium Emissions"
73.14 Mton	"Medium Emissions"
81.08 Mton	"High Emissions"

Here is the full ranking of data sets for HDI & total CO2 emissions (GII-CO2 rankings are in the Appendix): *__Key__*: Green data is "low", Orange data is "medium" (where applicable) and Red data is "high" for Chi-Squared testing purposes.

$Country$	HDI (x)	$Total\ CO2$ Emissions (Mton) (y)
Chad	0.404	0.71
Sierra Leone	0.419	1.40
Liberia	0.435	1.20
Mozambique	0.437	9.59
Gambia	0.460	0.58
Senegal	0.505	10.37
Uganda	0.516	5.38
Lesotho	0.520	0.71
Rwanda	0.524	1.10
Zimbabwe	0.535	12.81
Pakistan	0.562	189.31
Cambodia	0.582	10.45
Zambia	0.588	5.050
Honduras	0.617	10.05
Guyana	0.654	1.86
El Salvador	0.674	7.77
Iraq	0.685	172.63
Vietnam	0.694	242.46
Philippines	0.699	143.42
Belize	0.708	0.44
Uzbekistan	0.710	104.08
Maldives	0.717	0.85
Botswana	0.717	7.68
Suriname	0.72	2.17
Jamaica	0.732	7.89
Mongolia	0.741	18.40
Brazil	0.759	506.90
Kazakhstan	0.800	273.62
Romania	0.811	81.08
Bulgaria	0.813	47.47
Croatia	0.831	19.52
Hungary	0.838	51.60
Bahrain	0.846	34.94
Latvia	0.847	7.820
Portugal	0.847	56.52
Saudi Arabia	0.853	631.74
Poland	0.865	327.44
Cyprus	0.869	7.53

Greece	0.870	73.14
Czechia	0.888	110.41
Austria	0.908	72.46
New Zealand	0.917	37.35
UK	0.922	379.38
Canada	0.926	595.02
Sweden	0.933	45.14
Iceland	0.935	4.07
Germany	0.936	787.95
Ireland	0.938	37.76
Australia	0.939	411.57
Norway	0.953	49.83

2) Using the different categories, we can now fill out a 3x2 contingency table with *observed* values and determine the degrees of freedom. This is the minimum number of values we need to fill out the other values in the table and can be calculated with the formula:

3) $number\ of\ df = (number\ of\ rows - 1)(number\ of\ columns - 1)$

$$= (3 - 1)(2 - 1)$$

$$= 2\ degrees\ of\ freedom$$

Observed Values 3x2 Table (2 degrees of freedom)	**Low HDI**	**High HDI**	**Sum of Rows**
Low Carbon Emissions	*13*	*5*	**18**
Medium Carbon Emissions	*5*	*12*	**17**
High Carbon Emissions	*5*	*10*	**15**
Sum of Columns	**23**	**27**	$\boldsymbol{n = 50}$

4) Next, we must determine the expected frequencies for each pairing – this is done by assuming the results would be independent such that $p(A \cap B) = p(A) \times p(B)$.

5) Therefore, each expected value would be the probability of being in that row multiplied the probability of it being in that column multiplied by the total

number of pairings (50). For example, determining the expected frequency of *Low Carbon emissions* and *Low HDI*:

$$\frac{p(Low\ Carbon)}{n} \times \frac{p(Low\ HDI)}{n} \times\ n = \frac{18}{50} \times \frac{23}{50} \times 50 = 8.28$$

If this is done with each pairing the expected frequencies in this case will look like this:

	Low HDI	High HDI	Sum of Rows
Low Carbon Emissions	$\frac{18}{50} \times \frac{23}{50} \times 50 = 8.28$	$\frac{18}{50} \times \frac{27}{50} \times 50 = 9.72$	18
Medium Carbon Emissions	$\frac{17}{50} \times \frac{23}{50} \times 50 = 7.82$	$\frac{17}{50} \times \frac{27}{50} \times 50 = 9.18$	17
High Carbon Emissions	$\frac{15}{50} \times \frac{23}{50} \times 50 = 6.90$	$\frac{15}{50} \times \frac{27}{50} \times 50 = 8.10$	15
Sum of Columns	23	27	**50**

6) The chi squared value is calculated with the equation:

$$\chi^2 = \sum \frac{\left(\boldsymbol{f}_{\boldsymbol{observed}} - \boldsymbol{f}_{\boldsymbol{expected}}\right)^2}{\boldsymbol{f}_{\boldsymbol{expected}}}$$

[where f is frequency or count - from now on $\boldsymbol{f}_{\boldsymbol{observed}}$ and $\boldsymbol{f}_{\boldsymbol{expected}}$ are shortened to $\boldsymbol{f}_{\boldsymbol{o}}$and $\boldsymbol{f}_{\boldsymbol{e}}$, respectively]

7) In order to do this easily, I constructed an excel table and used sub-formulae e.g. for HDI and net Carbon Emissions, matching up observed and expected frequencies:

$\boldsymbol{f}_{\boldsymbol{o}}$	$\boldsymbol{f}_{\boldsymbol{e}}$	$\boldsymbol{f}_{\boldsymbol{o}} - \boldsymbol{f}_{\boldsymbol{e}}$	$(\boldsymbol{f}_{\boldsymbol{o}} - \boldsymbol{f}_{\boldsymbol{e}})^2$	$\frac{(\boldsymbol{f}_{\boldsymbol{o}} - \boldsymbol{f}_{\boldsymbol{e}})^2}{\boldsymbol{f}_{\boldsymbol{e}}}$
13	8.28	4.72	22.2784	2.690628
5	9.72	-4.72	22.2784	2.292016
5	7.82	-2.82	7.9524	1.016931
12	9.18	2.82	7.9524	0.866275
5	6.90	-1.90	3.6100	0.523188
10	8.10	1.90	3.6100	0.445679

Summing the final column, we calculate:

$\chi^2 = \sum(2.690628, 2.292016, 1.016931, 0.866275,0.523188, 0.445679) = 7.834717 = 7.83$ (3sf)

8) In order to conclude we must compare the chi squared value, 7.83, to the critical value at 2 degrees of freedom with 5% significance level:

Critical values of the Chi-square distribution with *d* degrees of freedom

	Probability of exceeding the critical value						
d	0.05	0.01	0.001	*d*	0.05	0.01	0.001
1	3.841	6.635	10.828	11	19.675	24.725	31.264
2	5.991	9.210	13.816	12	21.026	26.217	32.910
3	7.815	11.345	16.266	13	22.362	27.688	34.528
4	9.488	13.277	18.467	14	23.685	29.141	36.123
5	11.070	15.086	20.515	15	24.996	30.578	37.697
6	12.592	16.812	22.458	16	26.296	32.000	39.252
7	14.067	18.475	24.322	17	27.587	33.409	40.790
8	15.507	20.090	26.125	18	28.869	34.805	42.312
9	16.919	21.666	27.877	19	30.144	36.191	43.820
10	18.307	23.209	29.588	20	31.410	37.566	45.315

***INTRODUCTION TO POPULATION GENETICS*, Table D.1**

Table 1 - Critical Values Table (Sinauer Associates, Inc, 2013)

$5.991\ (Critical\ Value) < 7.83\ (Chi\ Squared\ Value)$, so, the null hypothesis is rejected: HDI and net Carbon emissions are *not* independent (the alternative hypothesis is not rejected).

Performing the identical process for GII and net Carbon emissions (see Appendix and contingency table below) with the same significance level and degrees of freedom yields $5.991\ (Critical\ Value) < 9.093486\ (Chi\ Squared\ Value),$ so, the null hypothesis is rejected: Gender Inequality Index and net Carbon emissions are *not* independent.

Observed Values 3x2 Table (2 degrees of freedom)	Low GII	High GIII	Sum of Rows
Low Carbon Emissions	5	13	**18**
Medium Carbon Emissions	13	5	**18**
High Carbon Emissions	10	4	**14**
Sum of Columns	**28**	**22**	$n = 50$

We have confirmed in both of the examples that one variable could have meaningfully affected the other as the statistics show the likelihood of one variable being any number is affected by (i.e. not independent of) the outcome of the other variable's corresponding value. Here the GII-Carbon relationship has a higher chi squared value than the one between HDI & Carbon, indicating the Gender Inequality Index values are less close to the frequencies we would expect if the two variables were independent.

The major limitation to this test is that I separated the data into categories (e.g. high/low) by myself and didn't follow any strict definitions when doing so. However, I strongly believe the conclusion drawn from the test was a good indication to independence nevertheless as the data was split into roughly equal parts each time where there was a significant incremental difference between two adjacent numbers for a variable.

Overall Conclusion and Evaluation

To answer my original question: I think this investigation has shown that HDI has a stronger relationship with net Carbon dioxide emissions more so than when GII was the independent variable. However, surprisingly according to the Chi-squared values HDI aligns much more closely to what we would predict if the variables were independent of one another. I believe this investigation has given a good insight into how development and net Carbon emissions are related – despite only finding weak correlations between the variables, the Chi-squared test confirmed that development indices and net Carbon emissions are not statistically independent of each other so there may even be some causation in reality. However, such weak correlations may suggest there are other variables that have a much greater association to Carbon Dioxide emissions – for example the level of industrialisation of countries.

There were a few limitations to the investigation. First of all, the use of sampling limited the sample size to 50 in this case out of almost 200 total countries. Perhaps a t-test between the variables of my sample and all the countries of the world would have given a useful insight into how representative the sample was and to what extent my conclusions could be made for all the world's countries, though this would probably require a lot more data that is yet to be collected.

Within the actual investigation there were a few more limitations. All statistical tests make assumptions – many assume a normal distribution within the variables, which was simply not the case in my data (especially for net Carbon emissions), so this could have limited the validity of e.g. my Pearson's coefficient calculations. Although a Spearman's ranking could have been calculated here it was unlikely to change the outcome of what hypothesis we rejected as the values likely would have been very similar (and indeed were when calculated using technology). Furthermore, as aforementioned, for tests like Chi-squared, although I tried to split the data into appropriate categories, this could be seen as entirely arbitrary and the fact that I chose where to split the data dependent on where I personally saw large differences introduces subjectivity into the statistical test.

Despite these limitations, I believe this was a valuable investigation and makes me question what other variables may affect both net Carbon emissions and development indices.

In the future, comparing different relationships may give us a bigger insight into the complex relationships that go alongside increases in development, perhaps splitting the HDI into its constituent indices (educational, life quality etc). Or perhaps the relationship between carbon emissions and the KOF index of Globalisation. I want to explore this because I know globalisation of trade is strongly associated with more imported goods from abroad and, therefore, I predict there is likely to be a stronger correlation between these variables than the indices I compared, whilst still being indicative of some form of development (though potentially more economic than social, like the HDI/GII indices).

Appendix

Data Sources: GII sourced from (United Nations Human Development Reports , n.d.)*, HDI sourced from* (United Nations Human Development Reports , n.d.)*, Net Carbon Emissions sourced from* (EDGAR, 2017)*.*

Pearson's r-value calculations for GII-CO2 Emissions

Country	*Gender* (x)	*Total CO2 Emissions (Mton)* (y)	$x-\bar{x}$	$y-\bar{y}$	$(x-\overline{x})(y-\bar{y})$	$(x-\overline{x})^2$	$(y-\bar{y})^2$
Sweden	0.044	45.14	-0.2754	-67.2324	18.5185	0.0759	4520.1960
Norway	0.048	49.83	-0.2714	-62.5424	16.9765	0.0737	3911.5520
Iceland	0.062	4.07	-0.2574	-108.302	27.8814	0.0663	11729.4100
Austria	0.071	72.46	-0.2484	-39.9124	9.9158	0.0617	1593.0000
Germany	0.072	787.95	-0.2474	675.5776	-167.1650	0.0612	456405.1000
Cyprus	0.085	7.53	-0.2344	-104.842	24.5793	0.0550	10991.9300
Portugal	0.088	56.52	-0.2314	-55.8524	12.9265	0.0536	3119.4910
Canada	0.092	595.02	-0.2274	482.6476	-109.7730	0.0517	232948.7000
Australia	0.109	411.57	-0.2104	299.1976	-62.9631	0.0443	89519.2000
Ireland	0.109	37.76	-0.2104	-74.6124	15.7014	0.0443	5567.0100
UK	0.116	379.38	-0.2034	267.0076	-54.3200	0.0414	71293.0600
Greece	0.120	73.14	-0.1994	-39.2324	7.8245	0.0398	1539.1810

Croatia	0.124	19.52	-0.1954	-92.8524	18.1471	0.0382	8621.5680
Czechia	0.124	110.41	-0.1954	-1.9624	0.3835	0.0382	3.8510
Poland	0.132	327.44	-0.1874	215.0676	-40.3123	0.0351	46254.0700
New Zealand	0.136	37.35	-0.1834	-75.0224	13.7621	0.0337	5628.3610
Latvia	0.196	7.82	-0.1234	-104.552	12.9060	0.0152	10931.2000
Kazakhstan	0.197	273.62	-0.1224	161.2476	-19.7432	0.0150	26000.7900
Bulgaria	0.217	47.47	-0.1024	-64.9024	6.6486	0.0105	4212.3220
Bahrain	0.222	34.94	-0.0974	-77.4324	7.5450	0.0095	5995.7770
Saudi Arabia	0.234	631.74	-0.0854	519.3676	-44.3748	0.0073	269742.7000
Hungary	0.259	51.60	-0.0604	-60.7724	3.6731	0.0037	3693.2850
Uzbekistan	0.274	104.08	-0.0454	-8.2924	0.3768	0.0021	68.7639
Mongolia	0.301	18.40	-0.0184	-93.9724	1.7329	0.0003	8830.8120
Vietnam	0.304	242.46	-0.0154	130.0876	-2.0086	0.0002	16922.7800
Romania	0.311	81.08	-0.0084	-31.2924	0.2641	0.0001	979.2143
Maldives	0.343	0.85	0.0236	-111.522	-2.6275	0.0006	12437.2500
Rwanda	0.381	1.10	0.0616	-111.272	-6.8499	0.0038	12381.5500
Belize	0.386	0.44	0.0666	-111.932	-7.4502	0.0044	12528.8600
El Salvador	0.392	7.77	0.0726	-104.602	-7.5900	0.0053	10941.6600
Brazil	0.407	506.90	0.0876	394.5276	34.5448	0.0077	155652.0000
Jamaica	0.412	7.89	0.0926	-104.482	-9.6709	0.0086	10916.5700
Philippines	0.427	143.42	0.1076	31.0476	3.3395	0.0116	963.9535
Botswana	0.434	7.68	0.1146	-104.692	-11.9936	0.0131	10960.5000
Suriname	0.441	2.17	0.1216	-110.202	-13.3962	0.0148	12144.5700
Honduras	0.461	10.05	0.1416	-102.322	-14.4848	0.0200	10469.8700
Cambodia	0.473	10.45	0.1536	-101.922	-15.6512	0.0236	10388.1800
Guyana	0.504	1.86	0.1846	-110.512	-20.3962	0.0341	12212.9900
Iraq	0.506	172.63	0.1866	60.2576	11.2417	0.0348	3630.9780
Senegal	0.515	10.37	0.1956	-102.002	-19.9476	0.0382	10404.4900
Zambia	0.517	5.05	0.1976	-107.322	-21.2026	0.0390	11518.1000
Uganda	0.523	5.38	0.2036	-106.992	-21.7794	0.0414	11447.3700
Zimbabwe	0.534	12.81	0.2146	-99.5624	-21.3621	0.0460	9912.6710
Pakistan	0.541	189.31	0.2216	76.9376	17.0463	0.0491	5919.3940
Lesotho	0.544	0.71	0.2246	-111.662	-25.0749	0.0504	12468.4900
Mozambique	0.552	9.59	0.2326	-102.782	-23.9031	0.0541	10564.2200
Gambia	0.623	0.58	0.3036	-111.792	-33.9357	0.0921	12497.5400
Sierra Leone	0.645	1.40	0.3256	-110.972	-36.1282	0.1060	12314.8700
Liberia	0.656	1.20	0.3366	-111.172	-37.4162	0.1133	12359.3000
Chad	0.708	0.71	0.3886	-111.662	-43.3875	0.1510	12468.4900

Chi-Squared (Categories and Method) for GII-CO2 Emissions

Key: Green data is “low”, Orange data is “medium” (where applicable) and Red data is “high” for Chi-Squared testing purposes.

Country	*Gender* (x)	*Total CO2 Emissions (Mton)* (y)
Sweden	0.044	45.14
Norway	0.048	49.83
Iceland	0.062	4.07
Austria	0.071	72.46
Germany	0.072	787.95
Cyprus	0.085	7.53
Portugal	0.088	56.52
Canada	0.092	595.02
Australia	0.109	411.57
Ireland	0.109	37.76
UK	0.116	379.38
Greece	0.120	73.14

Croatia	0.124	19.52
Czechia	0.124	110.41
Poland	0.132	327.44
New Zealand	0.136	37.35
Latvia	0.196	7.82
Kazakhstan	0.197	273.62
Bulgaria	0.217	47.47
Bahrain	0.222	34.94
Saudi Arabia	0.234	631.74
Hungary	0.259	51.60
Uzbekistan	0.274	104.08
Mongolia	0.301	18.40
Vietnam	0.304	242.46
Romania	0.311	81.08
Maldives	0.343	0.85
Rwanda	0.381	1.10
Belize	0.386	0.44
El Salvador	0.392	7.77
Brazil	0.407	506.9
Jamaica	0.412	7.89
Philippines	0.427	143.42
Botswana	0.434	7.68
Suriname	0.441	2.17
Honduras	0.461	10.05
Cambodia	0.473	10.45
Guyana	0.504	1.86
Iraq	0.506	172.63
Senegal	0.515	10.37
Zambia	0.517	5.05
Uganda	0.523	5.38
Zimbabwe	0.534	12.81
Pakistan	0.541	189.31
Lesotho	0.544	0.71
Mozambique	0.552	9.59
Gambia	0.623	0.58
Sierra Leone	0.645	1.40
Liberia	0.656	1.20
Chad	0.708	0.71

$f_{observed}$	$f_{expected}$	$f_{observed} - f_{expected}$	$(f_{observed} - f_{expected})^2$	$\frac{(f_{observed} - f_{expected})^2}{f_{expected}}$
5 (Both Low)	10.08	-5.08	25.8064	2.560159
13 (Medium Carbon, Low GII)	10.08	2.92	8.5264	0.845873
10 (High Carbon, Low GII)	7.84	2.16	4.6656	0.595102
13 (Low Carbon, High GII)	7.92	5.08	25.8064	3.258384
5 (Medium Carbon, High GII)	7.92	-2.92	8.5264	1.076566
4 (Both High)	6.16	-2.16	4.6656	0.757403

Bibliography (APA)

EDGAR. (2017, October 30). *Fossil CO2 and GHG emissions of all world countries*. Retrieved from Emissions Database for Global Atmospheric Research: https://edgar.jrc.ec.europa.eu/overview.php?v=booklet2019#

OMOJU, O. (2014, September 23). *Environmental Pollution is Inevitable in Developing Countries*. Retrieved from Breaking Energy: https://breakingenergy.com/2014/09/23/environmental-pollution-is-inevitable-in-developing-countries/

Renze, J. (1999-2020). *"Outlier"*. Retrieved from MathWorld--A Wolfram Web Resource: https://mathworld.wolfram.com/Outlier.html

Sinauer Associates, Inc. (2013). Critical values of the Chi-square distribution at p = 0.05, 0.01, & 0.001 for d = 1 - 20 degrees of freedom. *Introduction to Population Genetics, Table D.1*. Retrieved September 12, 2020, from https://www.mun.ca/biology/scarr/4250_Chi-square_critical_values.html

United Nations Development Programme. (2020). *Why does the HDI not include dimensions of participation, gender and equality?* Retrieved from United Nations Development Programme - Human Development Repots: http://hdr.undp.org/en/content/why-does-hdi-not-include-dimensions-participation-gender-and-equality

United Nations Human Development Reports . (n.d.). *Table 1. Human Development Index and its components*. Retrieved from United Nations Human Development Reports : http://hdr.undp.org/en/composite/HDI

United Nations Human Development Reports . (n.d.). *Table 5: Gender Inequality Index*. Retrieved from United Nations Human Development Reports : http:/hdr.undp.org/en/composite/GII

World Population Review. (2020). *High Income Countries 2020*. Retrieved from World Population Review: https://worldpopulationreview.com/countries/high-income-countries/