

Misty Ann Stone
6.2 Exploring Relationships

The Data (Summary)

Dependent Variable: Economic Freedom index from the Fraser Institute

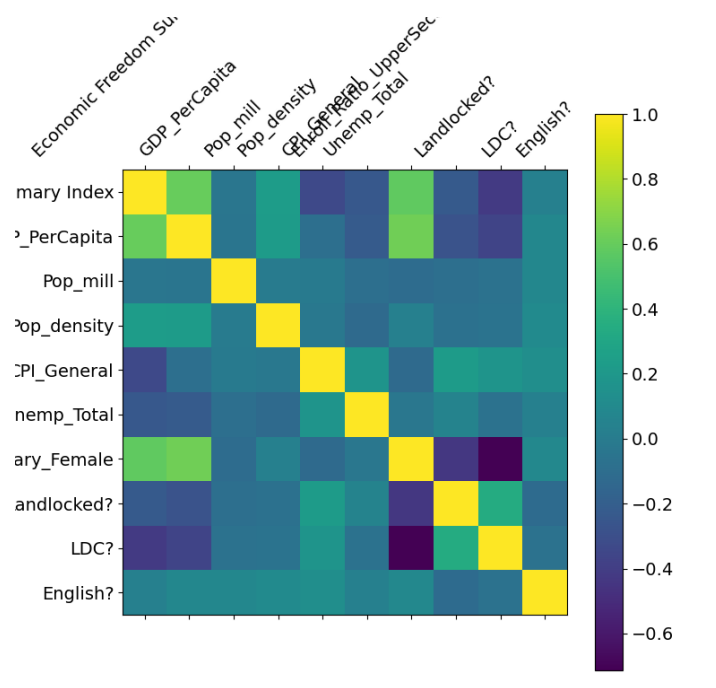
Independent Variables: Economic and social variables from each country:

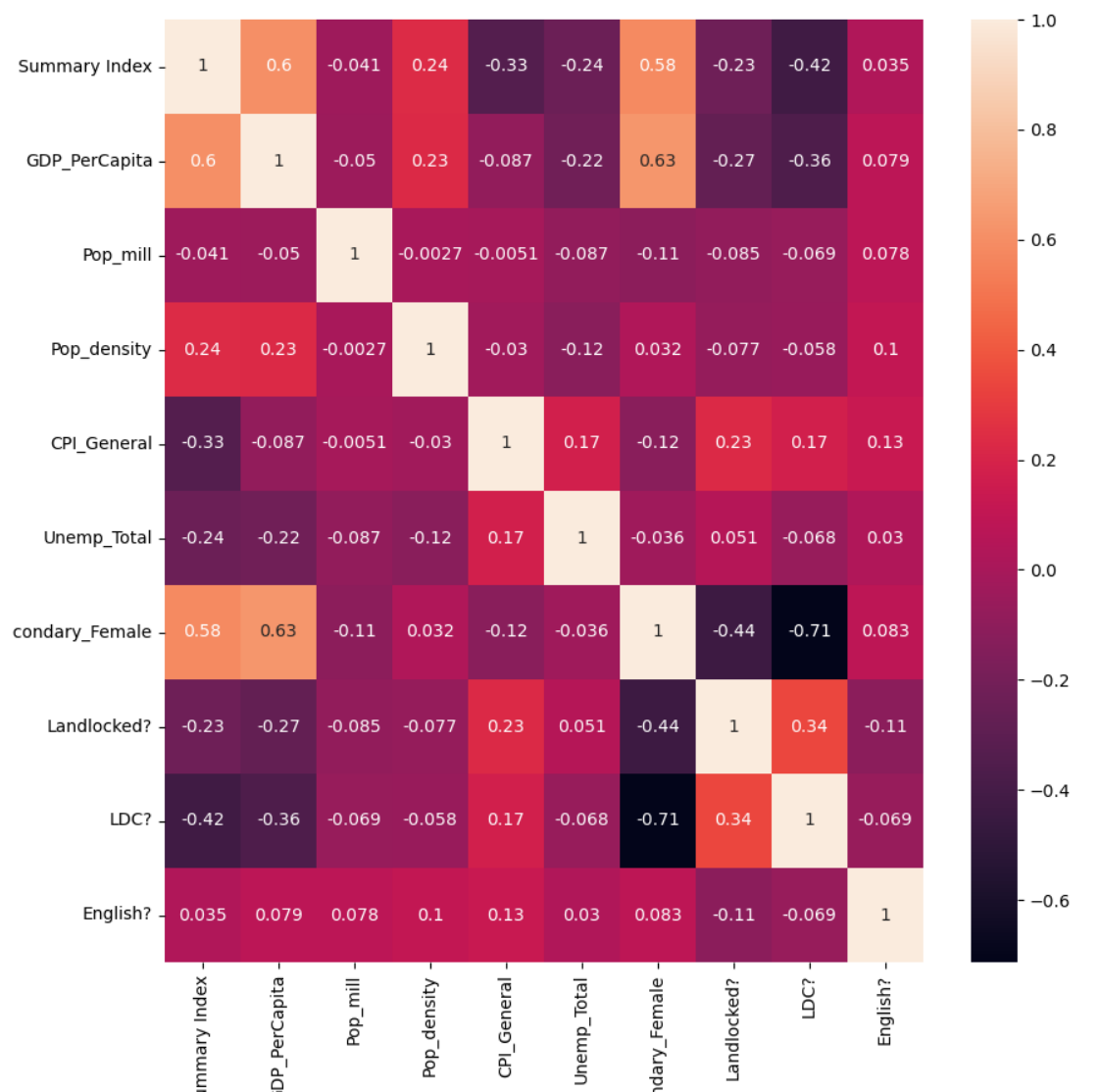
- GDP
- Population
- CPI
- Labor force participation and unemployment
- Education: enrollment rates
- Categories: continents, economies by income
- Indicators: landlocked? English spoken?

The Questions

The principal question that we are going to explore is: Which of the independent variables determine the Economic Freedom of a country? For this portion of the analysis, I have limited the data to the most recent year for each section, or (in the case of the Education data) the year in which we have full data information. A country with a higher GDP per capita is much less likely to have missing data, especially for the older years. This would put much more weight on the “richer” (in terms of GDP) countries. In other words, if we include many years in which the lower GDP countries are missing data, the weight on their observations will be relatively less than the higher GDP countries that are not missing so many data points.

Correlation matrix heatmap:





The coefficients in the plot indicate the relationship between the two variables.

- If the correlation coefficient is 0, there is no correlation between the variables.
- If the correlation coefficient is positive, when one variable is high, the other is likely to be high as well, and vice versa – a positive correlation.
- If the correlation coefficient is negative, when one variable is high, the other is likely to be low, and vice versa – a negative correlation.

If the correlation coefficient is:

- 0: no relationship
- 0.1 - 0.3: weak relationship

- 0.3 - 0.5: moderate relationship
- 0.5 - 1.0: strong relationship

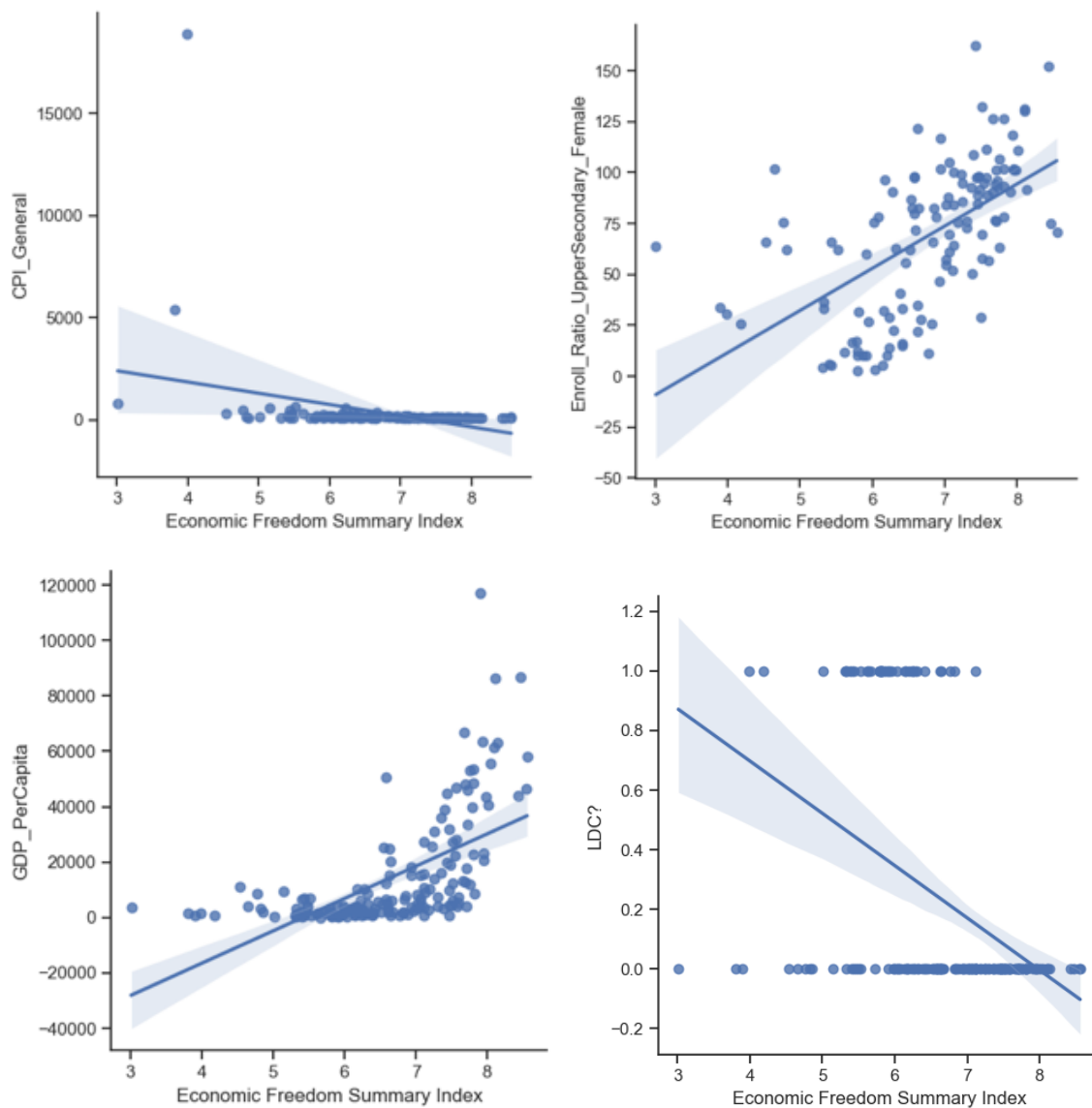
The above heatmaps indicate the correlations between our variables.

- The correlation between the Economic Freedom Summary Index and the GDP per capita of a country is positive and high, with a correlation coefficient of 0.6. A country with a high GDP per capita is very likely to have a high level of Economic Freedom.
- The correlation between the Economic Freedom Summary Index and the enrollment ratio of females in upper secondary education is also positive and high, with a correlation coefficient of 0.58. A country with a high Economic Freedom Index is likely to have a higher percentage of women attend upper secondary school.
- The correlation between the Economic Freedom Summary Index and LDC is strong and negative, with a correlation coefficient of -0.42. Recall that “LDC” refers to the UN’s “Least Developed Country / Nation” category. This is determined based on the net income per capital of a nation. As such, the countries with a low GDP per capita are very likely, nearly *by definition*, to be on the LDC list. Therefore it is not surprising that the a country likely to have a “1” in their LDC column are likely to have a low Economic Freedom Index, given what we have already discussed about the correlation of GDP per capita and the Economic Freedom Index.
 - Note also that the correlation between GDP per capita and the LDC indicator is negative and strong, with a correlation coefficient of -0.36.
 - It is also interesting to note the very strong and negative correlation coefficient between the LDC indicator and the Enrollment Ratio of women in Secondary school. A woman in a country in the LDC category is much less likely to be enrolled in upper secondary school than a woman in a country not in the LDC category. Socially and economically, there are many reasons for this. Several economists have noted in intensive studies that countries which are able to begin their climb *up* the poverty ladder almost always have a higher economic freedom for women in particular. There doesn’t seem to be any other determinant variable that uniquely identifies the beginning of the poverty ladder ascent in the way that female economic freedom does. I wonder about the relationship between women’s education levels and economic freedom in particular, setting aside the fact that a woman from a relatively wealthier family is less likely to need to quit school to seek employment.
- A country with a higher CPI (inflation rate) is likely to have a lower Economic Freedom Index, with a correlation coefficient of -0.33.
- A country with a higher population density is likely to have a higher Economic Freedom Index, with a correlation coefficient of 0.24.
- A country with low unemployment is likely to have a higher Economic Freedom Index, but only with a correlation coefficient of -0.24. This is probably the most surprising correlation to the average person as this correlation is not particularly strong. It is important to note that one of the marks of a country with higher Economic Freedom is the ability to change employment more freely. Therefore, it is (relatively) easy to find a

job. Within the unemployment statistic are people who are structurally unemployed – which tends to be more long term, but also those who are simply between jobs and are likely to find a new job more suited quickly.

- A country that is landlocked is likely to have a lower Economic Freedom Index.
- The relationship between the Economic Freedom Index and whether the country has a tradition of speaking English is not at all strong.
- The relationship between the Economic Freedom Index and the population of the country is not strong at all.

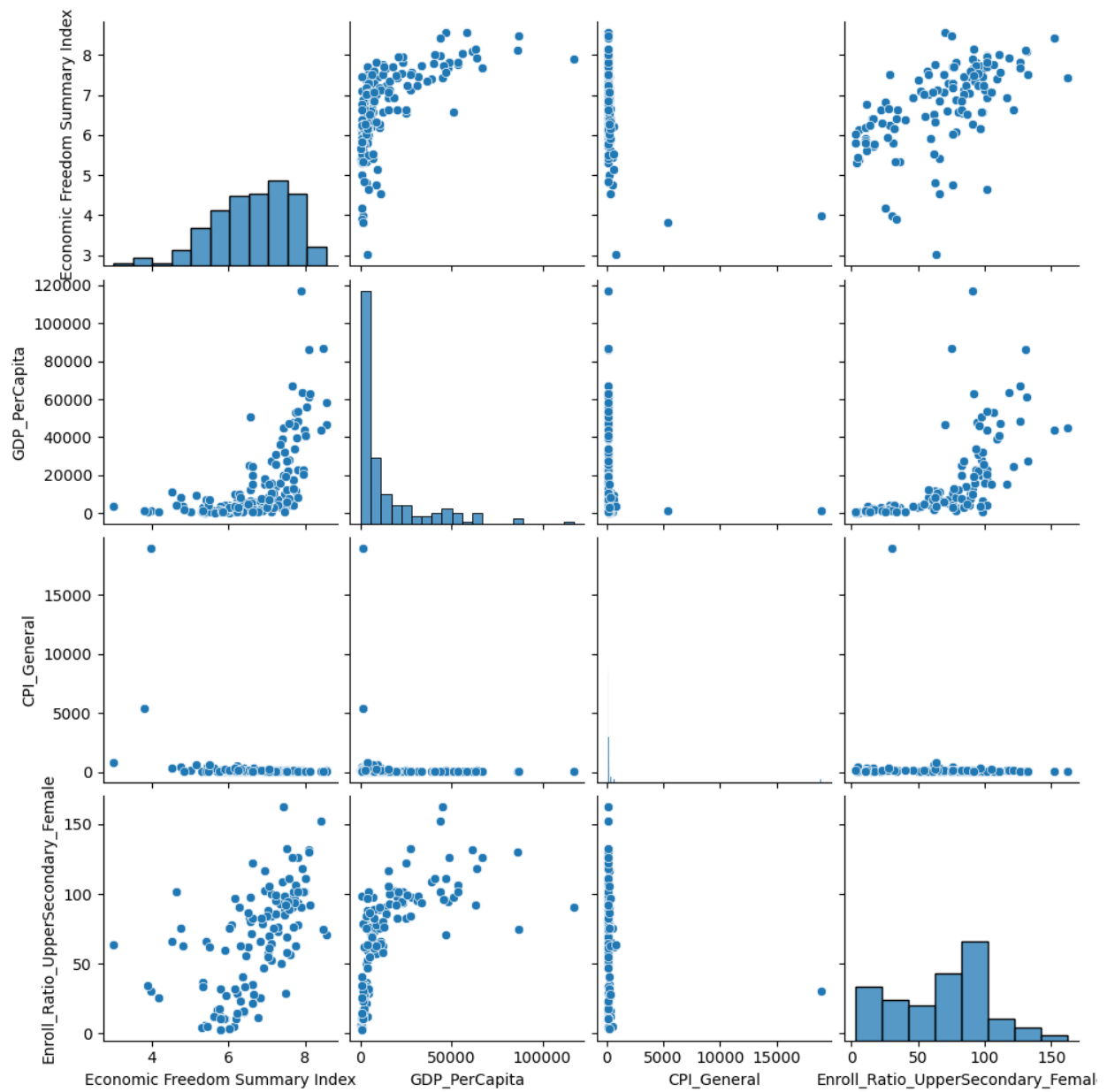
Scatterplots



- Consider the CPI scatterplot. There are several outliers that skew this scatterplot. We are not surprised that the majority of countries have relatively stable inflation rates – a CPI close to 100. Nonetheless, the three dots that represent countries with very, very high inflation rates make it hard to read this chart. Any idea about how to eliminate these from the chart?
- The scatterplot showing upper secondary enrollment ratios for women is strong and positive. There is some variance here, but this is relatively consistent with what we normally see in economic data.
- The scatterplot showing GDP and Economic Freedom Index is strong and positive. There are several notable outliers and I wonder if the trend line among the handful of “richest” countries (in terms of GDP per capita) should be non-linear. Nonetheless, the highest Economic Freedom Index countries do tend to have highest per-capita GDP.
- Regarding the LDC scatterplot, I was curious as to what this data would do in a scatterplot in seaborn. Recall that the variable LDC takes either a 0 if the country is not categorized an LDC and a 1 if it is. We have a negative relationship. I am interested to know which countries are considered LDCs, but have high Economic Freedom, and whether they are more likely to move out of LDC status. Perhaps we can explore that question during the time series analysis.

(Note that the above discussion about the scatterplots is also included in my Jupyter Notebook for this section, in a markdown cell, per the exercise instructions. I included it above for your easy access.)

Pair Plot

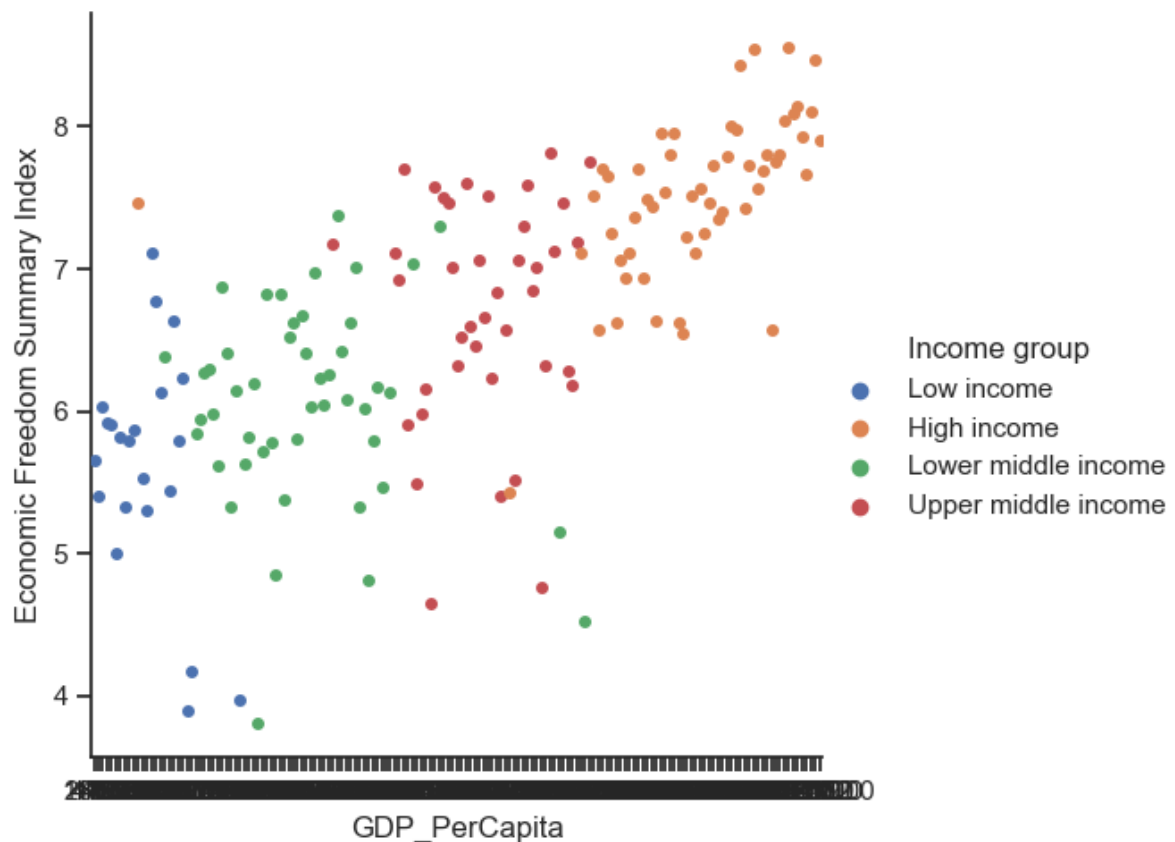


The distribution of the Economic Freedom Index is as expected: the most countries have an index of 7 or 7.5, but the mean is 6.6. There is an obvious positive relationship between GDP and the Economic Freedom index. The outliers skew the CPI graph and it is difficult to see exactly what is happening there. I'd like to isolate the CPI data to see what is happening. Nonetheless, it is unclear at which point to limit the upper bound – too many countries important to this study would be eliminated if we eliminate too many. I will ponder on this. Nonetheless, looking at the cluster of scatterpoints, a negative relationship is evident. A positive relationship between Female enrollment in upper secondary is evident as well, but the variance is higher.

The GDP column is interesting. The skew of the histogram of the GDP variable itself is heavily to the right. The GDP and Female Enrollment in Upper Secondary variables plot is interesting in the very steep slope at the left of the graph. The very “poor” countries tend to have extremely low enrollment rates, but the countries just “up the poverty ladder” have a progressively much higher female enrollment rate.

The CPI plots clearly all have extreme outliers. The Female Enrollment in Upper Secondary charts are all as expected in terms of slope and variation.

Categorical Plot



This plot shows the Economic Freedom Index compared to the Per Capita GDP of each nation. The color of each point is determined by the World Bank's Income Category for each nation.

The low income group countries are concentrated on the lower end of the GDP scale, but several are relatively high on the Economic Freedom scale – I wonder what they are and how their economies have changed over time! The Lower Middle Income group (green) also has a relatively high variation. As we consider the Upper Middle and High income countries, the variance decreases.

Further Questions and Hypotheses

I would love to consider how CPI affects these variables in more detail. Statistically, how many countries can we eliminate and still have a meaningful study? This is definitely something that will require an attentive hand.

I would like to consider the cases of the countries who have a lot of Economic Freedom, but a low GDP per capita. Has their GDP increased in a significant way since their Economic Freedom Index has increased? This would be a question in the time series section.

I am also interested to see whether the Male Enrollment in Upper Secondary ratios show a similar pattern to the female ratios.

- My principle null hypothesis is that each of the independent variables has a statistically significant effect on the Economic Freedom Summary Index.
- My (new) secondary hypothesis is that Female Upper Secondary Enrollment is more affected by Economic Freedom than Male.
- My third hypothesis is that higher Economic Freedom Indices cause higher GDP per capita, especially limited to relatively low income countries.