Team members:
Isabel Garcia Pietri
Lucas Barbosa
Jamie Smith

W200 Project 2 Final Report

Income inequality trend around the world and its relationship with economic growth and global trade

August 4th 2020

Introduction

The main goal in this study is to analyze income inequality across the world and its relationship with economic growth and global trade. The study focuses on three main research questions:

1. What is the trend of income inequality across the world in the last decade? Which countries/regions are doing better/worse?
2. What is the relationship between economic growth and income inequality?
3. What is the relationship between exposure to global trade and income inequality? (E.g. Do higher levels of exposure to global trade translate into lower inequality?)

Data

Data Sources

This study uses five datasets:

1. World Income Inequality (published by the academic and research arm of the UN). Link: https://www.wider.unu.edu/project/wiid-world-income-inequality-database
2. GDP growth (published by the International Monetary Fund) Link: https://www.imf.org/external/datamapper/NGDP_RPCH@WEO/OEMDC/ADVEC/WEOWORLD
3. Trade as % of GDP and nominal GDP by country (published by the World Bank). Links: https://data.worldbank.org/indicator/NE.TRD.GNFS.ZS https://data.worldbank.org/indicator/NY.GDP.MKTP.CD
4. Human Development Index (HDI) by country used to classify countries into developed and developing economies (published by UN). Link:

http://hdr.undp.org/en/indicators/137506

Data preparation

The data preparation phase of the project involved a series of steps and iterations to ensure data consistency and quality. As expected, this represented one of the most time consuming tasks in the project.

World Income Inequality dataset

The first step in the data preparation process was to **explore** the **World Income Inequality** dataset and understand the meaning of all variables available (columns). The variable that contains reported gini indexes ("gini_reported") was identified as the main variable of interest in this dataset (see definition of Gini Index in Appendix A).

The second step was to filter out all **non-relevant data**. The World Income Inequality dataset offers a great variety of gini indexes from diverse sources and with different levels of granularity (country, regions inside countries, specific demographics, etc.). As this study focuses on understanding income inequality at the country level, all gini indexes that do not represent an entire country and its population were filtered out (using variables "areacovr" and "popcovr"). Data was also filtered to consider years from 2009 to 2018 (last year available in the dataset). Although 2018 contains less data than any other year, it was kept to preserve the latest recorded gini indexes in some countries.

Next focus was the **quality** of the data. There is a quality variable attached to all gini indexes in the dataset ("quality"). This variable reflects the level of confidence in the reported gini indexes using four categories: high, average, low and not known. Observations with low and unknown quality were filtered out.

After that, the **structure** of the table was manipulated. Gini index observations in the original dataset have the form:

...	**Country**	...	**Year**	**Gini reported**	...
...	Canada	...	2009	value 1	...
...	Canada	...	2011	value 2	...

Also, there can be more than one gini index per year in a given country (reported by different sources).

Data was reorganized to have countries listed as rows and reported gini index per year as columns. For countries with more than one gini index per year, an average value was calculated:

Country	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**
Canada	value1	value2	value3	value4	value5	value6	value7	value8	value9	value10
Zambia	value1	value2	value3	value4	value5	value6	value7	value8	value9	value10

...	...	...	...	...	...	...	...	...	...	...	...

Next a check for **undefined values** (NaN) was done. There are several countries with none or few gini indexes reported between 2009 and 2018. As one of the objectives of this study is to determine whether the income inequality has increased or decreased around the world, all countries with less than two gini indexes reported between 2009 and 2018 were filtered out.

After all the considerations mentioned above, 115 countries remained in the table.

Finally a **trend in income inequality** was calculated. Gini index values for several countries were scattered from 2009 to 2018. This made it difficult to calculate a representative trend in income inequality. After some iterations, the approach that worked best was to calculate the best-fitted trend line of gini indexes available between 2009 and 2018. A positive slope of the trend line indicates that income inequality has increased and a negative slope indicates that the income inequality has decreased. The magnitude of the slope represents the average yearly percentage point change in the gini index.

Other variables extracted from this dataset were: regions and subregions that the countries belong to, and 3-digits and 2-digits country codes. Checks for uniqueness and undefined values (NaN) were performed on all these variables.

The resulting gini index table is used to answer research question number 1. This table was later merged with GDP and trade datasets to help answer research questions number 2 and 3. This process is explained in the following sections.

Trade dataset

For the trade analysis, we included a **new country classification** column, splitting countries into developed vs. developing economies. To do so, our first step was to import a dataset with Human Development Indexes (HDI) by country from the United Nations. We picked only the most recent indexes available (2018) and applied a conditional rule such that for HDI equal or higher than 0.85 the country is considered developed, and for HDI lower than 0.85 the country is considered developing. To be able to apply this conditional rule we had to replace some string values '. .' for NaN, and also to make the column numeric, as the data was imported as string.

Second step was to **merge** this new country classification with our original World Income Inequality dataset. We used a left-merge to preserve the set of countries previously selected for the analysis. From 115 countries, we were able to match 101. From the 14 remaining, 5 were countries with data not available in the HDI dataset, but other 9 were countries with different spelling (for example, 'Gambia' vs. 'Gambia, The'). For these 9 cases, we edited the names one by one using the loc function before performing the merge. This increased the final number of matches to 110.

Next step involved the **import of two datasets from the World Bank**, one with the Trade as % of GDP per country, and the other with the nominal GDP per country. Both datasets extended from 1960 to 2019, so we selected only the years of interest for us (2009 to 2018). We also had to standardize column names such as 'country' or 'code3' to make them aligned with our original dataset and facilitate the merging.

Because the number of variables increased in the data and a number of years were considered (2009-2018), we switched back to a format where each row represented a country in a given year, with the indicators (gini, trade, gdp, etc.) as columns. As the datasets from the World Bank, and our Income Inequality dataset had only one indicator each and 'years' as columns, we had to **flip the format of all of them using the melt method**. We also used the to_datetime method to convert the new 'year' column into datetime format.

Next step was to perform the **merges**, using a left merge again. Merges were more successful this time, with only one country missing (Taiwan).

Next we included a **new column with the trade in nominal terms**. This was done by multiplying the trade as a % of GDP by the column GDP in nominal terms. It proved handy to have this column as everytime we grouped our dataset, we needed to recalculate the Trade Ratio dividing the trade by the GDP in nominal terms.

Last step was to add a **new column with the trend of the Trade Ratio**. To keep consistency, we applied the same approach used with the gini indexes, using the best-fitted trend line available between 2009 and 2018. A positive slope indicates that the Trade Ratio has increased, while a negative slope indicates that the Trade Ratio has decreased. The magnitude of the slope represents the average yearly percentage point change in the trade ratio.

GDP dataset

In order to better understand the relationship between Real GDP and gini, we leveraged a dataset from the IMF. The first phase of data clean up included trimming the dataset down to a span of 10 years from 2009 to 2018. Similar to the gini trend calculation, a **trend in the real gdp growth rate** was calculated to allow us to better understand if the real gdp growth for a given country trended positively or negatively over the observed 10 year period. This calculation was done using a similar method as the gini trend highlighted above.

Once we had a column which captured the best fit for the real gdp growth rate, we **collapsed the 10 columns** associated to years (2009 - 2018) down to a single column **leveraging the melt method**. Once this was completed, we were able to then perform a **left merge** on our master data set to allow us to include two new columns. The real gdp growth rate trend that we

calculated as well as the real GDP growth rate per year which was part of the original IMF data file.

Once the real gdp growth rate trend and the gini trend were captured in the same dataset, we were able to create a final column capturing whether a country was increasing or decreasing either gini or their gdp growth rate. A function “trends” was created which intakes a dataframe and will apply a result to each row of the dataframe based on whether gini or the gdp growth rate trend either increased or decreased. This function allowed for us to quickly create a new column to indicate if the country had increased/decreased gini/real gdp growth over the 10 year period. This would then be leveraged as part of the GDP analysis

One final column that was used as part of our analysis was **GDP per capita**. The World Bank data set, primarily leveraged to analyze trade, also included population. This made it easy to quickly create a new column by dividing the country’s total GDP by the population.

Income inequality analysis

Income inequality trend around the world

The first part of this study focuses on understanding how income inequality has changed around the world over the last decade[1].

To measure income inequality the gini index is used. This index ranges from 0 to 100, where 0 indicates that income is shared equally among all people and 100 indicates that one person accounts for all income (See Appendix A for more information about the Gini Index).

The results of the analysis show that 71% of the countries[2] decreased income inequality between 2009 and 2018. Figure 1 shows how the trend is distributed around the world. This figure shows the average yearly percentage point change in the gini index. A positive value indicates an increase in income inequality, and a negative value a decrease in income inequality.

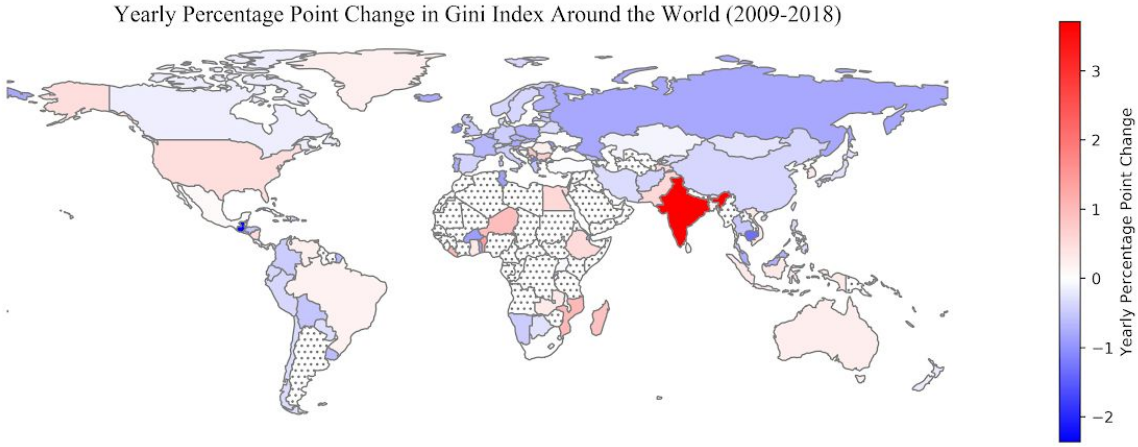


Figure 1. Gini index yearly percentage point change around the world between 2009 and 2018. Blank dotted areas indicate that no data is available for the countries to determine the change in inequality.

Although the majority of countries reduced income inequality, some countries increased it drastically in Africa and Asia. The tables below show a list of the countries with more drastic changes (in both ways).

	%PointChange
India	3.71
Benin	1.44
Mozambique	1.03
Liberia	1.02
Niger	0.94

Table1. Top 5 countries that increased income inequality.

	%PointChange
Guatemala	-2.36
Barbados	-1.86
Gambia, The	-1.75
Cambodia	-1.34
North Macedonia	-1.23

Table2. Top 5 countries that decreased income inequality.

[1] The analyzed data goes from 2009 to 2018. More recent data was not available at the time of the study.
[2] Considering data of 115 countries. See the Data section for more details about the data.

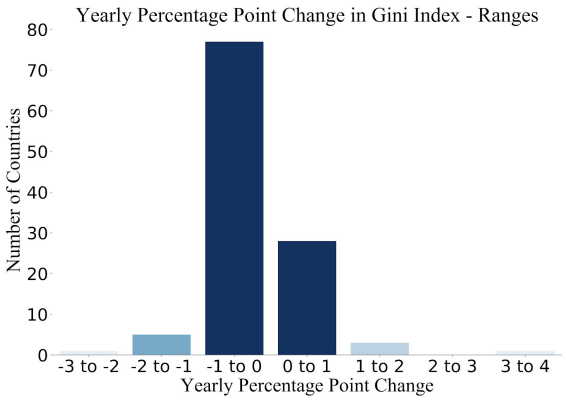


Figure 2. Average yearly gini index change ranges between 2009 and 2018.

Majority of countries have percentage point changes between -1 to 0 (Figure 2), which means that over a given year the majority of countries decreased their gini index by about 1 percentage point. Most of the countries that increased income inequality also did it by about a percentage point yearly. Figure 3 shows the income inequality change categorized by five major regions around the world. As can be observed in the figure, Africa is a region that exhibits a heterogeneous behavior, with about the same number of countries increasing and decreasing income inequality. In all other regions the majority of countries reduced income inequality and the behavior within the region is relatively homogeneous, except in America and Asia, both regions with countries that exhibit extreme changes in income inequality.

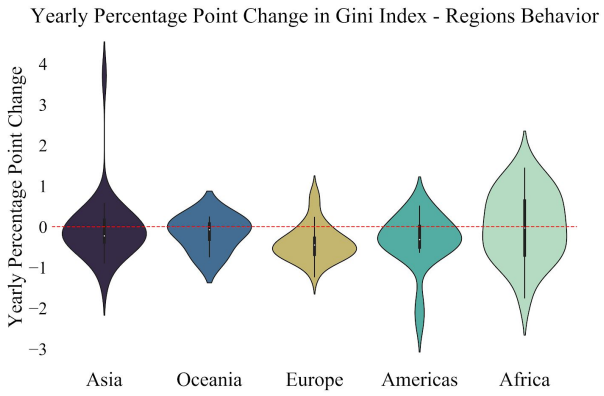


Figure 3. Average yearly gini index change distributions in the five major regions of the world between 2009 and 2018.

So far data has shown that income inequality decreased around the world. But did countries with severe inequality move towards a more equal distribution of income or the reduction of income inequality was concentrated in the less affected countries? Table 3 shows the percentage of countries that increased/decreased income inequality with respect to to their inequality level in 2009[3].

As can be seen, the group where more countries decreased inequality is the group of countries that already had good equality levels. But the second best group is the one with countries with severe inequality. On the other hand, countries

	Decreased	Increased
Adequate equality	83.3%	16.7%
Low inequality	68.4%	31.6%
High inequality	71.4%	28.6%
Severe inequality	81.8%	18.2%

Table3. Percentage of countries that increased/decreased income inequality between 2009 and 2018. The income inequality categorization is with respect to the gini index value countries had in 2009.

[3] Gini index values closest to 2009 were considered for countries with no available gini index value in 2009.

with low inequality is the group where more countries increased income inequality, putting them at a higher risk. Figure 4 shows the relationship between percentage point change in the gini index and the gini index of countries in 2009.

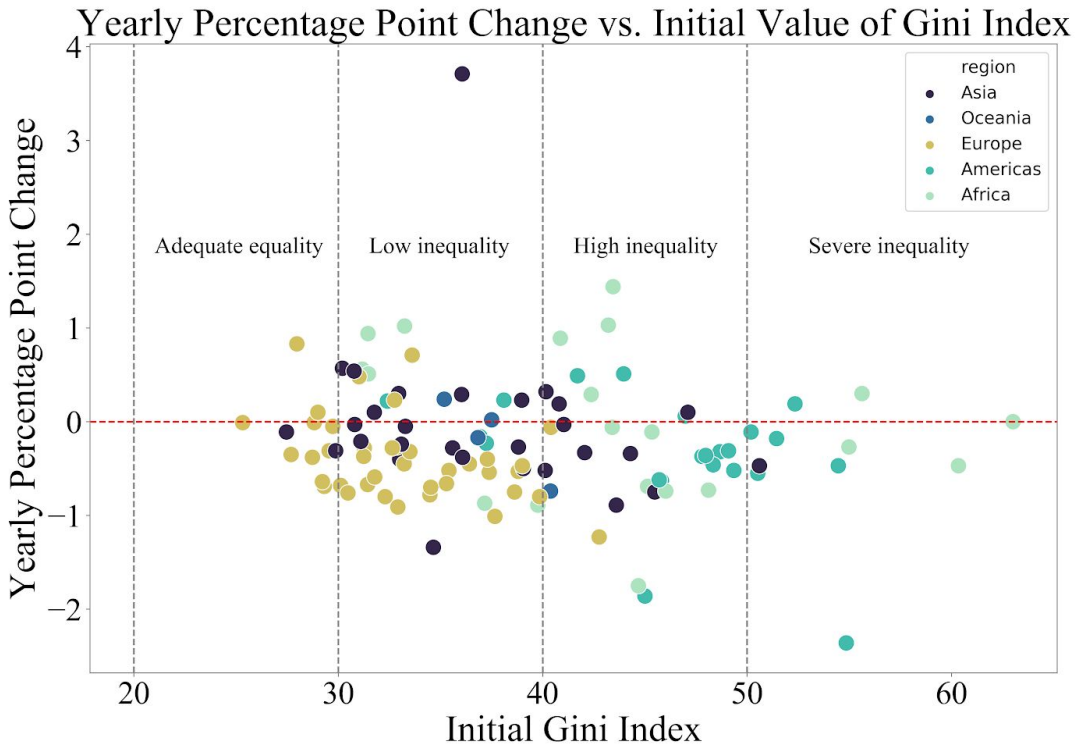


Figure 4. Relationship between yearly percentage point change in gini index and gini index values at the beginning of the analyzed 10 years period (2009-2018).

As can be seen in Figure 4, most European countries further decreased their already low income inequality. Oceania shows a similar behavior. Asia, America and Africa, with countries that have a wide variety of income inequality levels, exhibit a more heterogeneous behavior in inequality percentage point changes. This diverse behavior is probably the reflection of low levels of economic and political integration in these regions. Europe, with many countries belonging to the European Union, managed to move almost all its countries to lower levels of income inequality.

The objective of the next section is to investigate how the income inequality trends relate to the performance of countries' economies.

Income inequality and economic growth

Though the gini coefficient serves as a reliable indicator of income inequality, in isolation, it does not paint the full picture of the economic situation for a country and its citizens. In order to get a better measure, it's important to also review the country's GDP.

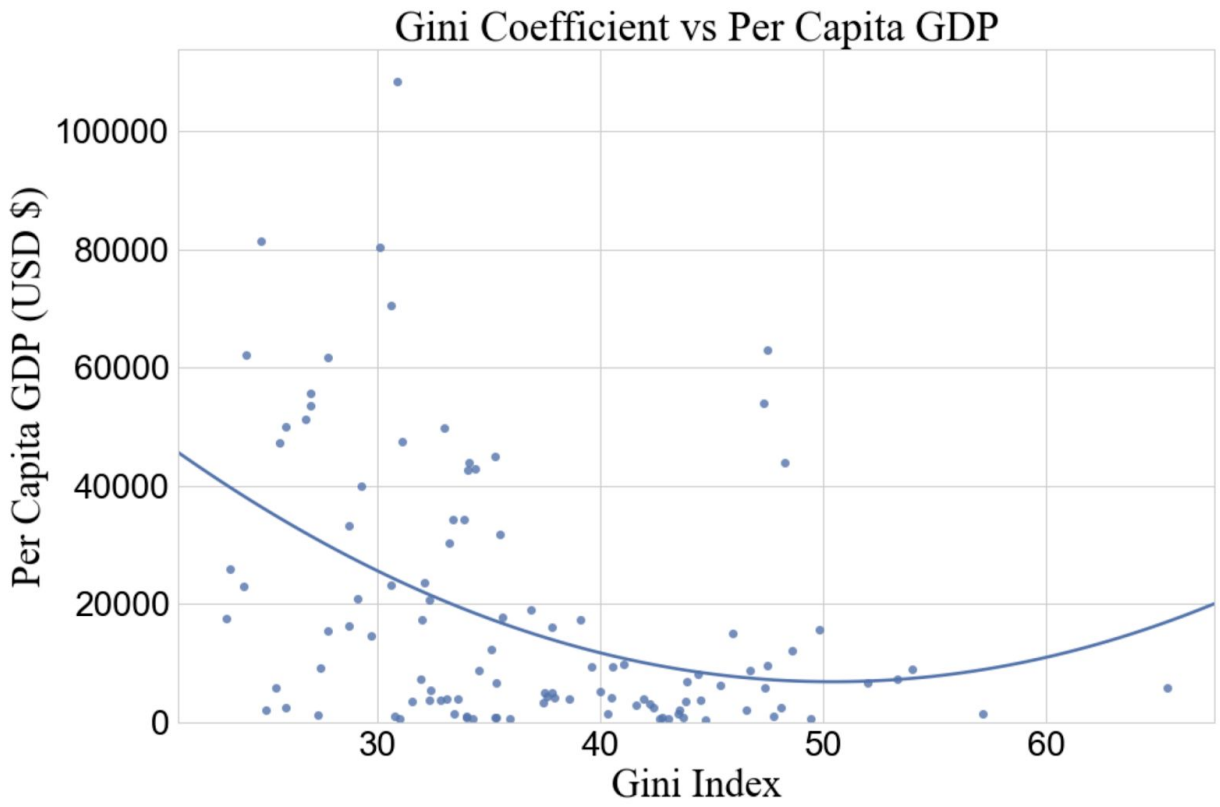


Figure 5. Scatter plot of countries based on the relationship between Gini and the Per Capita GDP. The GDP and Gini scores represent the latest data available for a given country between 2008 and 2018

With the exception of a few data points, there's a general insight that the distribution of wealth amongst nations whose per capita GDP is below $20,000, varies dramatically. There appears to be very little correlation between the gini index and nations whose per capita GDP is below $20,000.

For nations whose per capita GDP is greater than $20,000 we see that the gini coefficient tends to be lower than 35%, with the exception of just a few nations. Though these facts are interesting, they only provide a single point in time. It's also important to understand the trends as well.

Below, we've calculated the 10 year average rate of change for the gini coefficient as well as the 10 year average rate of change for the Real GDP growth rate for the nations in our data set.

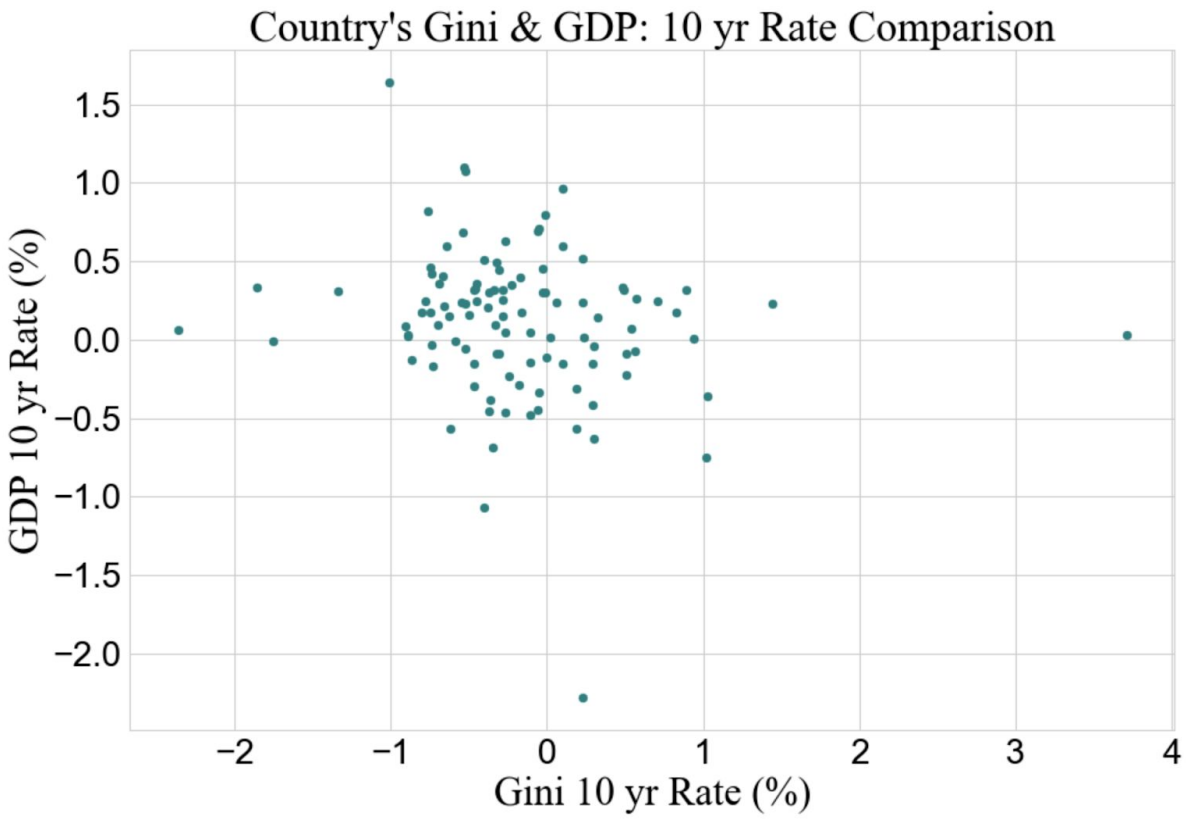


Figure 6. Scatter plot of Gini 10 year trend with the GDP 10yr trend rate. Highlights if a nation is both increasing Real GDP and decreasing income inequality

The findings show that there is little correlation between the real GDP growth rate and the Gini trend. In addition to this, the correlation was calculated to be -0.17, further concluding that there's little correlation. That said, there's still value gained in understanding how many nations have decreased income inequality as well as increased real GDP growth.

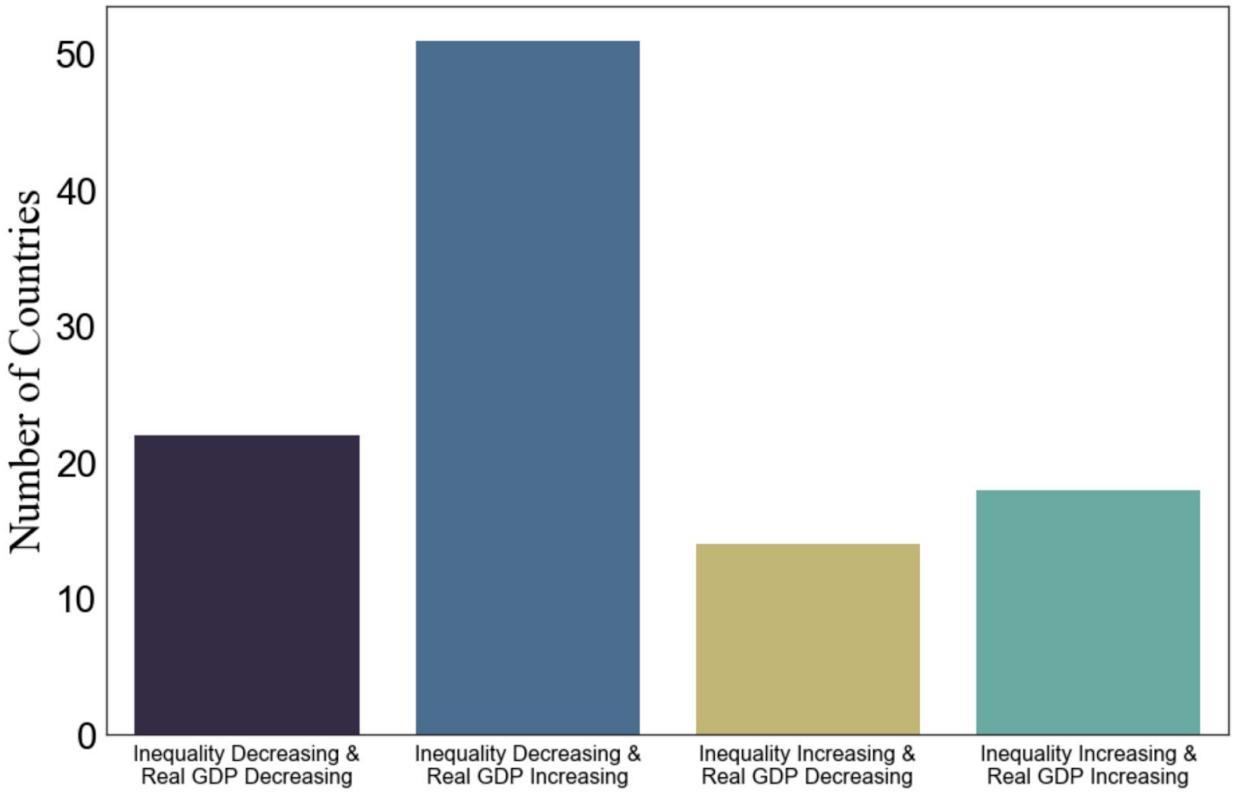


Figure 7. Bar chart identifying the countries that have increased/decreased real GDP and gini

The above figure shows an interesting finding, that the majority of nations across the world are trending in the right direction. They are decreasing income inequality while also increasing real GDP. It's also promising that the smallest bucket are those nations that are increasing inequality while also decreasing real gdp growth.

Income inequality and globalization

The impact that globalization has on equality is a controversial topic of debate in the international community. More evidence that either supports or challenges the theoretical prediction of the positive impact of international trade on income distribution is needed.

In this part of the study, we analyze the relationship between inequality and globalization. In this study, we define globalization as the level of exposure a country's economy has to the global trade of goods and services. The measure we are going to use is the country's total sum of exports and imports divided by the country's GDP. We will refer to this measure as Trade to GDP Ratio.

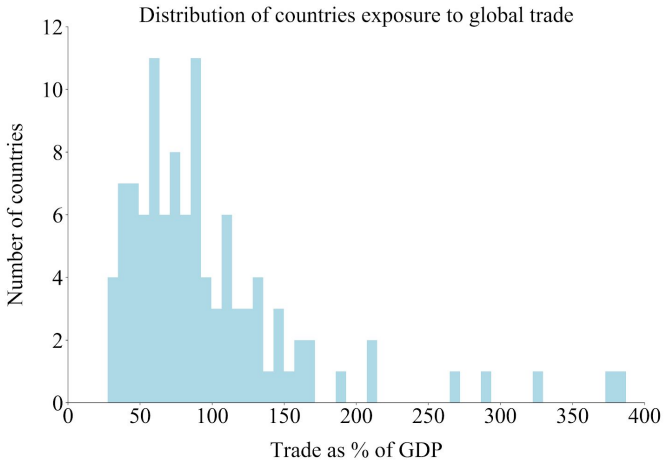


Figure 8. Histogram of countries according to their Trade to GDP ratio for 2018.

Countries have different levels of integration with the global economy. If we take 2018, which is the most recent year in our time series, we see that there is a significant level of dispersion on the Trade to GDP Ratio among countries.

Although the median of the distribution is around 80.7%, the range goes from a minimum of 27.5% to a maximum of 387.1%.

In general, developed economies are more integrated than developing economies, with a Trade to GDP Ratio of 60.4% vs. 49.5%.

Among the regions, there are also notable differences. European and Asian countries are on average the most integrated regions. Africa average is aligned with the world median, while Oceania and Americas are the regions less integrated.

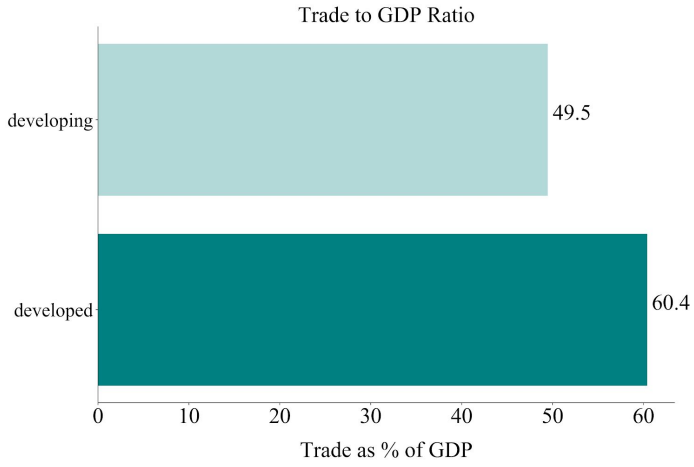


Figure 9. Trade to GDP ratio for developed and developing economies in 2018.

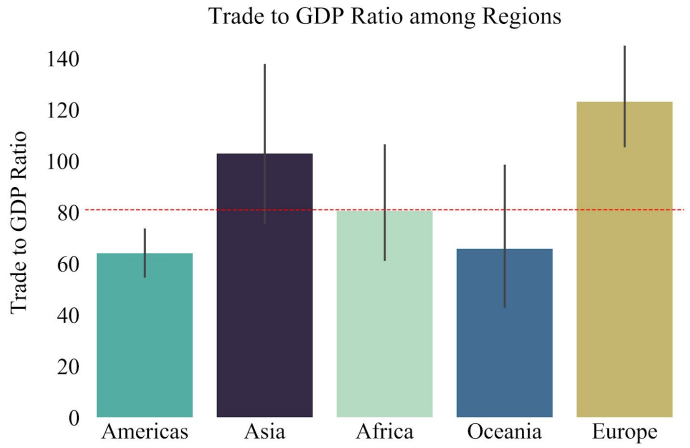


Figure 10. Trade to GDP ratio per region for 2018.

The key question we are going to try to answer is: what is the relationship between exposure to global trade and income inequality? Or in other words, do higher levels of exposure to global trade correlate with lower inequality?

We will try to answer the question through two different angles. The first angle is to simply compare levels of global trade exposure and inequality. For global trade exposure, we will use the Trade to GDP metric for 2018, as described above. For inequality, we will use the Gini Index for 2016, instead of 2018, as it increases the number of countries for which we have data available.

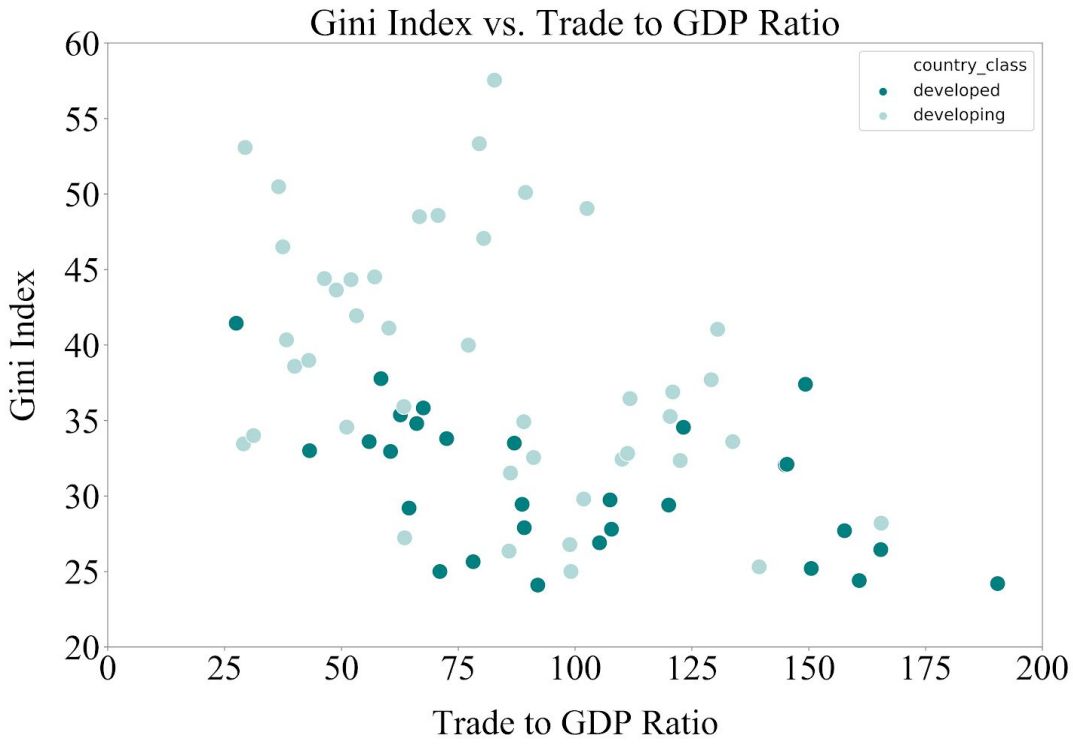


Figure 11. Relationship between Trade to GDP Ratio (2018) and Gini Index (2016) for selected countries.

The second angle compares the change in trade exposure to the changes in inequality, being a more acid test to see if there is indeed a relationship between both variables. For change in trade exposure, we will use the slope of a simple regression of the Trade to GDP metric between 2009 and 2018. Same criteria will be applied for the Gini Index, as we reported on the first part of this study.

On the level comparison, a slightly negative trend can be observed. The Gini Index tends to decrease as a country's Trade to GDP Ratio increases. But the relationship is not as strong as one could expect, with a correlation of only -0.25 between the variables.

However, when analyzing the relationship between the change in both indicators, correlation is equal to zero. In other words, there is no evidence that countries that increased their exposure to global trade in the recent decade have managed to reduce inequality, and vice versa.

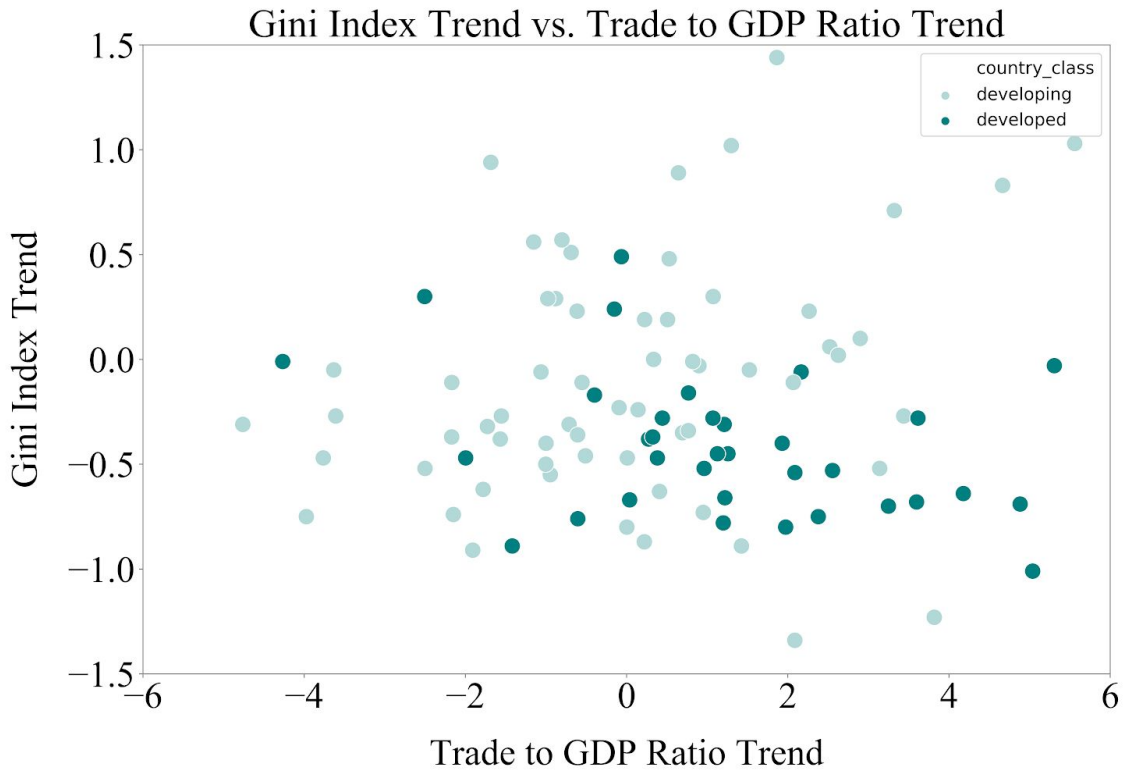


Figure 12. Relationship between the trend of Trade to GDP Ratio between 2009 and 2018, and the trend of Gini Index between 2009 and 2018 for selected countries.

There are likely several other factors affecting a country's inequality level than exposure to global trade, such as domestic market dynamics, labor market conditions, inflow of capital, and government redistributive policies. A more extensive study is required, but for now, we can conclude that more globalization by itself does not automatically lead a country to a more equal society.

Conclusions

- 71% of the 115 analyzed countries decreased their income inequality between 2009 and 2018. In most of these countries the gini index decreased by 1 percentage point or less each year.
- European countries further decreased their already low income inequality. Oceania shows a similar behavior. Asia, America and Africa, with countries that have a wide variety of income inequality levels, exhibit a more heterogeneous behavior in inequality changes.
- This study does not consider more recent changes (2019-2020) in income inequality due to data availability. Effects of phenomena like the COVID-19 pandemic will have a strong effect on income inequality trends around the world.
- Nations with less than $20,000 (USD) in per capita GDP have a large dispersion of income inequality; however, nations with per capita GDPs in excess of $20,000 (USD) tend to have less income inequality overall
- In aggregate, the majority of nations since 2009 have both decreased income inequality as well as increased real GDP. Likewise, the smallest proportion of nations has increased income inequality and decreased real GDP for its citizens
- Countries have different levels of integration with the global economy, with their Trade to GDP Ratio varying from a minimum of 27.5% to a maximum of 387.1%.
- On average, developed economies are more integrated than developing economies, and Europe and Asia are regions more integrated than Oceania and America.
- On a static level comparison, the Gini Index indeed tends to decrease as a country's Trade to GDP Ratio increases, although the correlation is not strong (-0.25).
- When analyzing Trade to GDP Ratio change vs. Gini Index change though, the correlation drops to zero, suggesting little evidence that more exposure to global trade leads automatically to a country's reduction in inequality.
- A more extensive study is needed as there are likely other variables affecting the inequality beyond exposure to globalization, such as domestic market dynamics, labor market conditions, inflow of capital, and government redistributive policies.

Jupyter Notebooks

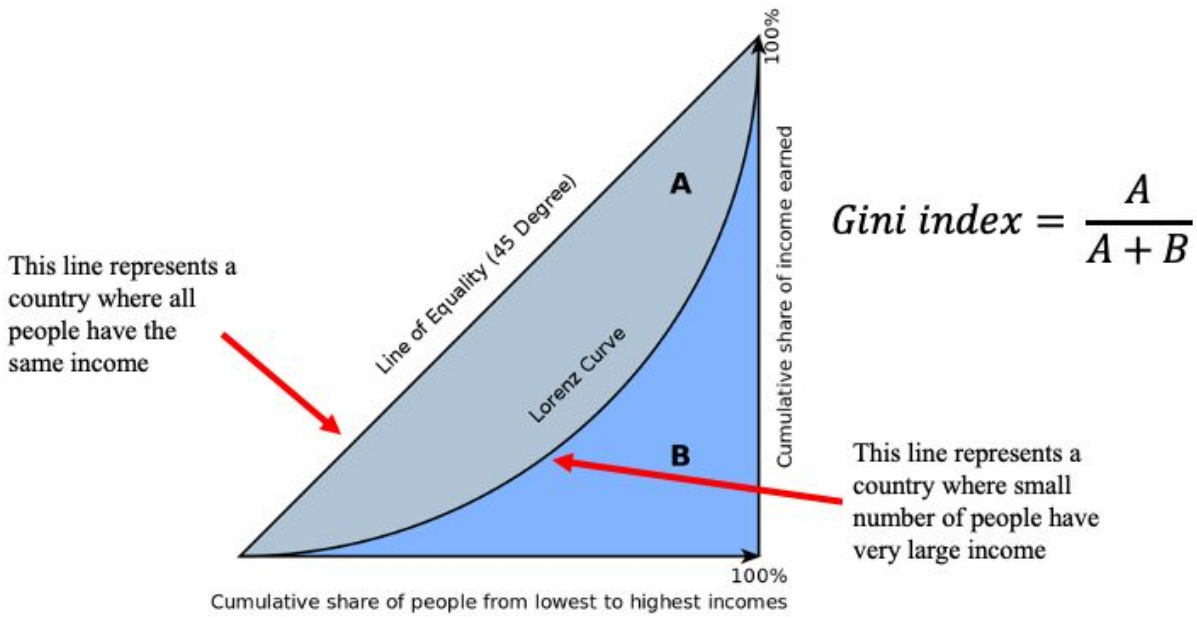
- **Q1-World_Income_Inequality_Trend_FINAL.ipynb**- World Income Inequality dataset manipulation and visualizations related to research question 1.
- **Q2_GDP_Analysis.ipynb** Exploration of the GDP analysis and visualizations answering question 2.
- **Q2_Relationship_GDP_Inequality_dataset_prep.ipynb** Data preparation file calculating and merging Real GDP trend, GDP, and GDP per capita.
- **Q3-Relationship_Trade_Intequality_dataset_prep.ipynb**- Datasets merging and preparation related to research question 3.
- **Q3-Relationship_Trade_Inequality_data_analysis.ipynb-** Data exploration and visualizations related to research question 3.

Appendix

A. Gini Index Definition

The Gini index measures the extent to which the distribution of income (or, in some cases, consumption expenditure) among individuals or households within an economy deviates from a perfectly equal distribution.

A Gini index of zero represents perfect equality and 100, perfect inequality.



$$Gini\ index = \frac{A}{A+B}$$

References

- https://en.wikipedia.org/wiki/Gini_coefficient
- World map shape file: https://www.naturalearthdata.com/downloads/110m-cultural-vectors/