



---

## Information Visualization Project "World Inequality Database" Visualization

---

Gasmi Adel  
Baradza Jonathan  
Mavunga Prosper Julius  
Abdelhak Zabour  
Kryeziu Jehona

Professor: Blanch Renaud

February 7, 2020

# 1 INTRODUCTION

In this project, the aim was to provide a visualization of a subset of World Inequality Database, which is a database containing the latest information in economic inequality across different countries. Having a subset of it, we wanted to find a way to visualize it so that to emphasize the most important information and to answer some of our questions regarding that dataset. To achieve this we used different tools such as the libraries: d3js, TopoJSON, d3-queue etc. As a result, we could do some interactive visualization which was able to reflect our perception of the most relevant content. It is mainly composed of a world map and 3 charts, which are updated to give information according to the country which is selected.

## 2 FULL VIEW OF THE VISUALIZATION

The result is an interactive dataset, the structure of which can be seen in the image below. Dynamicity is provided by the slider which changes the colors on the countries in the map according to the different period of time, by selection of one country in the map which updates the plot in the three charts accordingly to the data corresponding to that country and the possibility to also select the country from a select list.

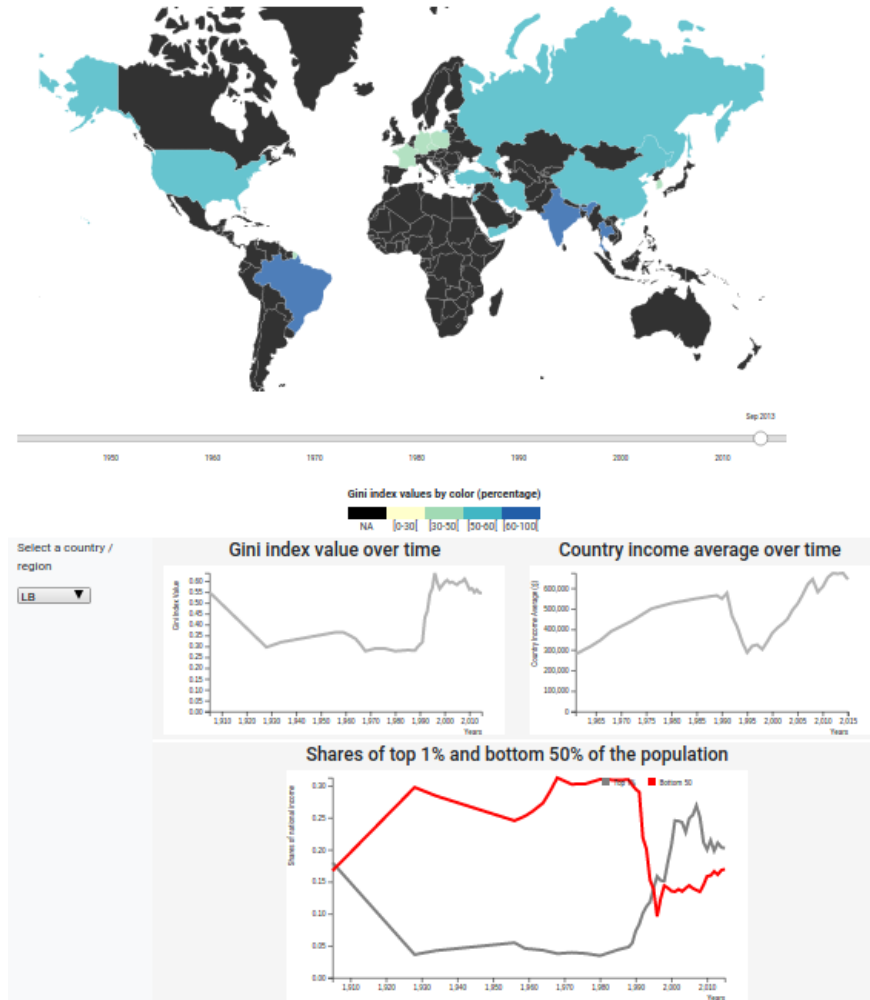


Figure 1: A view of the overall visualization structure

## 3 FOCUS TO THE FIRST PART: MAP

Let's see in more detail the map. Moving the slider, we can see the change of Gini for the countries of the world through months and years, and the value of the color can be interpreted with the help of the legend below it.

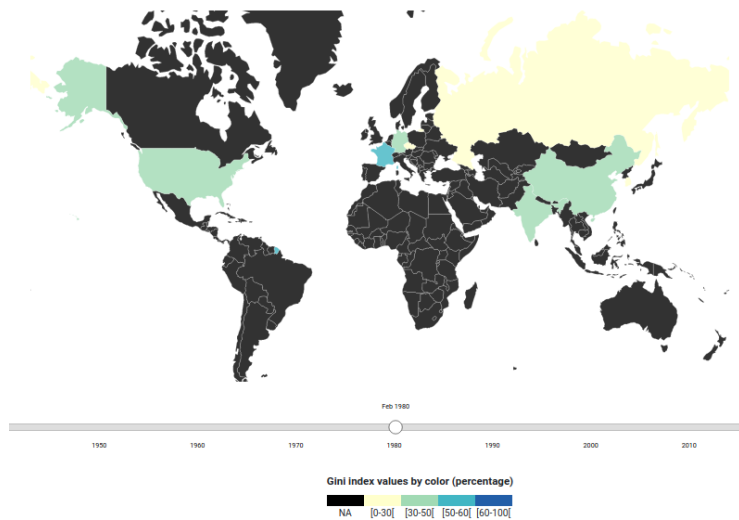


Figure 2: First part of the visualization: Map.

As it can be seen from the image above, we don't have data for most of the countries regarding the Gini index for 1980, but for those countries that we have France has the biggest value of Gini index, while Russia has the lowest. Besides those, when one selects the map of one country, a text pops-up with its name and the number of population.

## 4 FOCUS ON THE SECOND PART: CHARTS

In the second part of the visualization we have three charts which show changes of some variables during the change of time for a specific country chosen either by selecting the map or the select list. Their values all correspond to the same country, and all change when we select a different country.

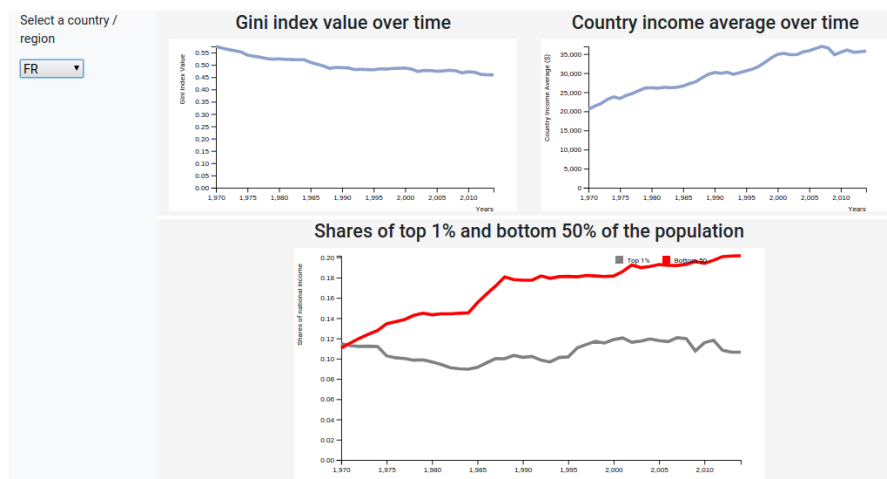


Figure 3: Second part of the visualization: Charts

## 5 CHART 1

In the first chart we have the change of Gini index value over the change of time for a specific selected country. Gini's values vary from 0 to 1.

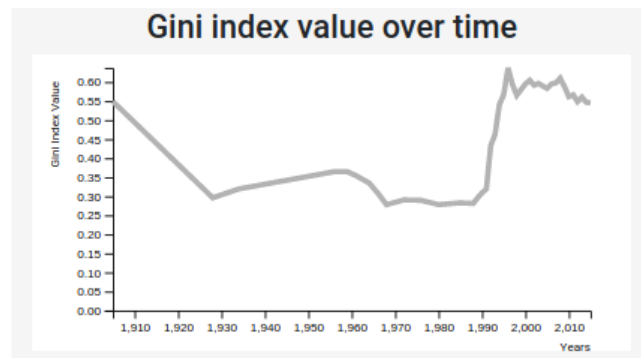


Figure 4: Chart 1 showing time vs. Gini index.

## 6 CHART 2

In the second chart, we can observe the change of country's average income during the years. The income is expressed in dollars \$. In fig. we can see one example of such chart.

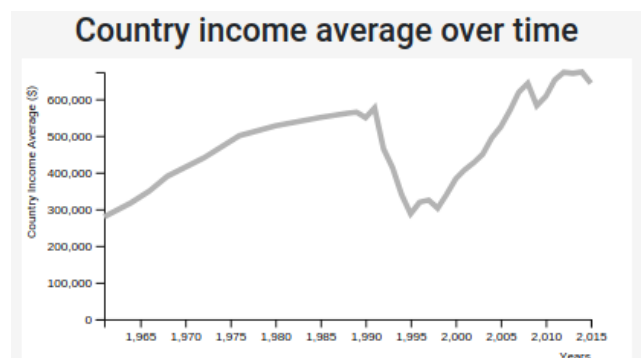


Figure 5: Chart 2 showing Time vs. Average Income

## 7 CHART 3

Chart 3 shows the share of national income among the top 1% (in grey) and bottom 50% (in red) of the population. The share of income is expressed in percentage (%). One example can be seen in Fig.5 below:

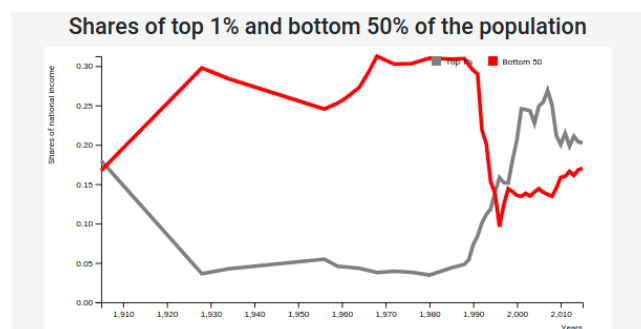


Figure 6: Chart 3 showing shares of top 1% and bottom 50%

## 8 PREPARATION OF DATA FOR CHART 3

The data used for Chart 3 are top 1% and bottom 50% of the population. Considering that they were not available in the dataset, we had to process the data we had in order to derive them. To be more concrete,

we filtered by conditioning the "high" column and got the needed attributes from the data. This step can be observed in Fig.6 below:

```
[ ]: import pandas as pd

df1 = pd.read_csv("D:\incomeall.tsv", sep="\t")
df1.head()

[ ]: filtered_low50 = df1.loc[df1['high']>=0.5]
filtered_high_90 = df1.loc[df1['high']>=0.99]
filtered_top1 = filtered_high_90
filtered_top1['top1_share'] = 1 - filtered_high_90['cumul']
filtered_top_1 = filtered_top1[['country', 'year', 'top1_share']]

filtered_low_50 = filtered_low50[['country', 'year', 'cumul']]
filtered_low_50.rename(columns={'cumul': 'bottom50_share'}, inplace=True)
filtered_low_50

[9]: new_df = pd.merge(filtered_top_1, filtered_low_50)
new_df

[9]:
```

	country	year	top1_share	bottom50_share
0	AE	2009	0.15571	0.10894
1	AE	1998	0.17838	0.12151
2	BH	2015	0.18236	0.14034
3	BH	2005	0.17960	0.14612
4	BH	1995	0.18890	0.14076
5	BR	2015	0.27841	0.12252
6	BR	2014	0.27599	0.12511
7	BR	2013	0.27540	0.12371
8	BR	2012	0.27269	0.12095

Figure 7: Piece of code showing the preparation of data to use in Chart 3.

## 9 QUESTION 1: WHICH WERE THE COUNTRIES WITH THE LARGEST GINI INDEX IN JANUARY 2000?

To answer this question, we set the slider at that date and observe the colors of the countries indicating 4 intervals of its values. As we can see from the Fig. 8, the countries with the darkest green in the map are Russia, USA and Yemen, which means that they had the largest Gini index, being in the range of [0.5,0.6].

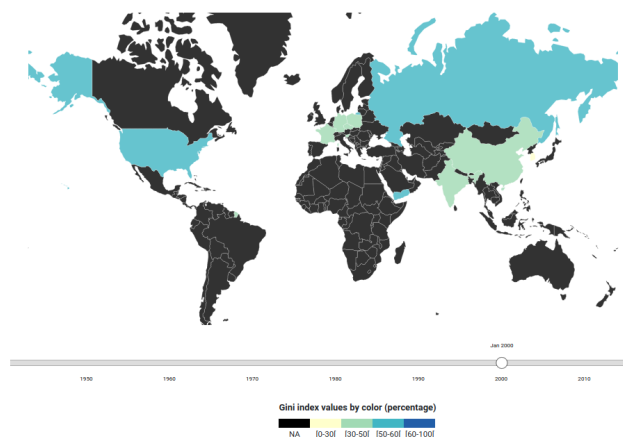


Figure 8: Answer for question 1 has to be inferred from this chart.

## 10 QUESTION 2: WHAT IS THE EVOLUTION OF INCOME AVERAGE FOR RUSSIA OVER THE YEARS?

To answer this question, we can select "Russia" in select list or map and observe the second chart. We see that Russia had a considerable increase after a decrease in 90's.

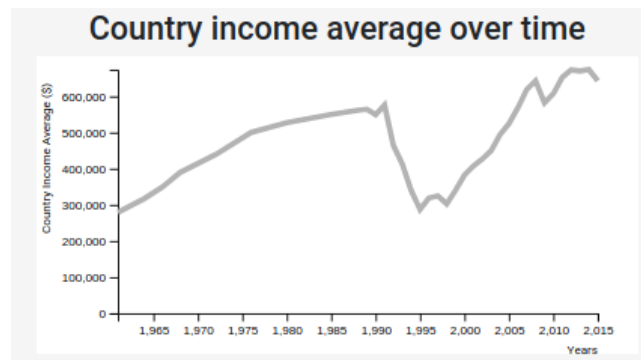


Figure 9: Answer from question 2 can be inferred from this chart.

# 11 QUESTION 3: HOW DO WE COMPARE THE SHARE OF INCOME BETWEEN TOP 1% AND BOTTOM 50% OF POPULATION FOR FRANCE?

We compare them by putting in the same graph and see their corresponding values. For France, it looks like this:

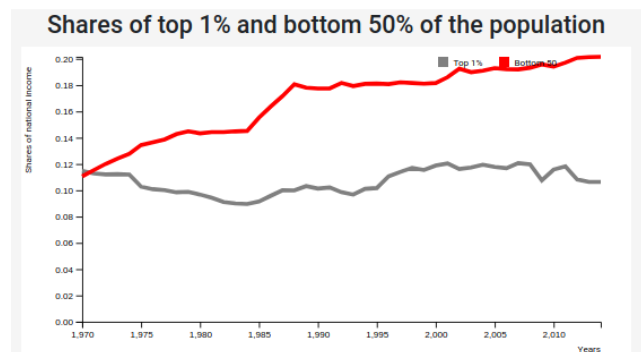


Figure 10: Answer to question 3 can be inferred from this chart.

This means that in France, during the last years the bottom 50% of the population has more income than the top 1%.