Info Visualization project

BRITEL, Malak
BENEVIDES, Gabriel
CHEN, Tyiani
LARAQUI, Omar
IDHBI, Yassir

February 2020

1 Introduction

The world is experiencing inequality whether between different countries or within the same country. The "World Inequality" data set makes it possible to highlight these aspects of the world's socio-economic situation. As part of the Information Visualization course project, we were led to develop visualizations regarding the inequality of the world through the data set that we were provided. A few questions arose when the given the data set was presented. We will proceed to discuss the following questions:

- What is the distribution of income resources for each segment of the population for each country per year?
- How to compare the average income and the Gini index between countries?
- Is there a common trend between annual GDP Growth and Gini Index amongst BRICS countries in the 21st century?
- Is there a relation between Average income and Gini index?

We will try to explain each visualization chart in the next sections.
Finally, we will conclude by mentioning a few of the problems faced during the project.

2 Aggregation of data

Before we start developing the visualizations, we agreed to aggregate the data. First, we aggregate the data of the different countries in a single .csv file by adding the full name of the country, its code, as well as its continent. This would help later on by having a more domain specific structure of the data for our visualizations. We added another file containing the continents and their

countries to also establish this relation in the data set. The purpose of the second aggregation was to group the different quantities into more expressive percentages of the population. It was not easy. Indeed, each country gave its censuses in a distinct way. For example, Korea has only given the cumulative 0 to 90, 90 to 95, 95 to 99 and finally 99 to 100. So we had to adapt our code. All the aggregated files were generated by a custom Python code in which we took into consideration all the differences of each country's data.

Finally, it's interesting to mention that we had to import a new set of Data to develop a response to the third question. This is explained more in detail in the respective section of this report. Suffice it to say that the quantity annual GDP growth wasn't available in the original data set provided for the project and so the group members had to actively find an alternative source to be able to construct this visualization.

3 First graph

In the first graph, we tackled the question: how resources are distributed over the population in each country and each year?
We choose a **Radial Stacked Bar Chart** representation to answer to this question. The percentages in the circles represent the percentages of the total income of the country which on total amount to 100%.
The percentages in the circles represent the percentages of the total income of the country which represent 100%.
Moreover, the colors represent the cumulative income of a percentage of the population. Since we did not have the data aggregated in the same way, we associated a color for each representation. As previously mentioned, we do not have for all countries data for each year, so for a given year we only display the countries for which we have the information available.
The range years displayed is from 1905 to 2016. The graph is also animated. Thanks to the play button we can visualize the evolution of the graph over time. We can also stop or reset the graph to 0.

4 Second graph

In the second graph we needed to compare the average income and the Gini index between the countries given in the dataset. For this task we chose to implement a **bubble chart** where we can represent the countries as bubbles so that we can compare the countries just by the size of the circles representing them.
Two versions of the graph have been done, in which we represented the Gini Index and the Average Income of the countries given in the dataset.
In the Gini index version, we used circles to represent the different countries and the Gini index is represented by the size (radius) of the circle. Respectively, the average income version use the value of the average income to represent

the countries by a circle where the big circles represent the countries with the highest average income.
The actual values of each circle and additional details are represented by a tooltip that appears whenever a cursor is hovering a specific circle.
The values represented are the last Average income and Gini given by each country.

5 Third graph

In the third graph, we tried to add an extra set of data, corresponding to annual GDP growth in the BRICS economic group of countries. The data used for this visualisation was obtained from the official World Bank open Database, which is freely available online.

In addition, we used a Python script, to create an aggregation of Gini and GDP growth data in the same file in order to facilitate the processing using D3.

We were interested in countries belonging to the BRICS economic group, that is, emerging economies in the beginning of the century with potential to become major geopolitical powers by the end of the century. The official definition for BRICS include:

- Brazil
- Russia
- India
- China
- South Africa

However, Gini data was scarcely available for South Africa. Considering it to be one of the smallest countries with least potential to affect world-wide economics, a decision was made to leave it out of the analysis to preserve a consistent data set.

In terms of the visualization we have the two values plotted on the same graphic over the same period of time, from 2000-2015. In the dashed line we can see values corresponding to Gini evolution each year, and in the solid line we can see the evolution of the GDP variation. The user has an option to click on each country to plot/unplot in the graph. Initially, we see all BRIC countries in the same plot. Although simple and straight-forward, this visualization allows us to mention some interesting points. With the shy exception possibly of the Russian Federation, other countries have experienced immense GDP growth in the 21st century. This is due to many different situations, internal investment in social programs to warm the economy and lower central bank interest rates in the case of Brazil. We can also mention the boom in commodity prices in the case of India and Russia. For the Chinese economy, it had been on a enormous growth trend for a few decades before. We can finally note that, despite moments of

excellent GDP increase, virtually none of the countries succeeded in drastically reducing their inequality Gini index. In the case of India for example, it is even worst in 2015 in relation to the year 2000. This goes to show that we cannot, with this data and at this point, simply correlate GDP growth to reduction of inequalities, because these in fact, come from an ensemble of socio-economic factors and public policies which aren't solely tied to the growth of an economy.

6 Fourth graph

To answer to the fourth question, we choose a **Dual Axis Y Line Chart** representation. The first Y axis is for the Average income while the second for Gini index. We can select any country and visualize the two lines over years. The blue one represents the Average income and the red one the Gini index. We notice that most of countries like France and US have a decrease in 2008 in the average income due to the economic crisis of 2008.
We can't have a general conclusion on the graph because it depends on each county.

7 Conclusion

Through our visualizations we answered the questions that we defined beforehand and we highlighted the inequality over countries. We aggregated data and added a new database.
Among the difficulties we encountered during this project, it was the heterogeneity of the data, both in terms of census year and the method of accumulation used by each country.
To conclude, despite the difficulties, we have tried to answer the questions using different visualizations that represent the best the world inequality.