GDP and correlation with Social factors

Catarina Filipe
Instituto Superior Técnico
Lisboa, Portugal
catarina.c.filipe@tecnico.ulisboa.pt

Diogo Ramalho
Instituto Superior Técnico
Lisboa, Portugal
diogo.ribeiro.ramalho@tecnico.ulisboa.pt

Lucas Perry
Instituto Superior Técnico
Lisboa, Portugal
lucasperry@tecnico.ulisboa.pt

ABSTRACT
GDP is an economic indicator that helps measuring a country's standard of living. This work investigates the relationship the EU between this indicator, the political party governing the countries and three social factors: unemployment for different ages and education levels, reported criminality in different income areas and emigration. To do this we collected data from the Eurostat for the EU countries and years between 2008 and 2018, and then created a visualization that helps understanding how these indicators are related. Our analysis showed a correlation between higher GDP and more politically right beliefs. It was observed that many countries chose political parties during the peak of the crisis where the GDP decreased, and then many returned to more leftist parties. We also noted that many countries with higher GDP had less impact on these social factors and recovered faster. We conclude that the strong causality we expected between GDP and the social factors was not observed, even though there is more tendency for higher GDP countries to have lower unemployment, emigration and criminality.

Author Keywords
GDP; Unemployment; Criminality; Emigration; Crisis; Politics

ACM Classification Keywords
Human-centered-computing~Visualization~Visualization systems and tools

INTRODUCTION

Domain
In the last decade, the countries of the European Union have faced the repercussions of a big financial crisis that started in 2007-2008 and sparked great recession in national economies. These setbacks reflect on their GDPs variation, with some countries having neutral or even negative years in terms of economic development, and in other social factors.

A country's GDP is calculated by taking into account the monetary worth of a nation's goods and services. GDP per capita helps us understand how the economy of a certain country is growing with its population. The analysis of this indicator helps measuring the general population's standard of living.

Naturally, the decrease of wealth in a country may be associated with loss of jobs, given by the unemployment rate. Consequently, it may also result in a loss of income that was necessary to pay for a house and food. This may lead to an increase in crime, violence or vandalism, explained by the reduction of people's standard of living. Also, when there is a crisis people tend to look for better life opportunities and if their country is not providing them, they will look for it outside. So, we analyzed the emigration rates within the EU. Finally, we looked for an association between the political beliefs (left/right) of parties governing the countries and the variation in GDP.

Relevance and Motivation
According to Eurostat [1], economic security, physical safety and material living conditions are some of the key elements in life quality standards, where people can pursue their self-defined well-being.

Thus, we think this analysis is important because having a better understanding of the impact of the GDP per capita in a country may result in more informed opinions by people when influencing, discussing, voting and protesting. Hopefully, also by our leaders when governing.

From a more personal perspective, in our course we rarely have the opportunity to study a more social domain, such as Economy. As we're people concerned with the future of our country, we were curious to see how some social factors were related to an indicator so commonly overheard in our public discussion, GDP and its growth. Very often we hear that we don't grow enough economically when compared to other countries in the EU, and so we were curious how that could then be related with key elements in people's lives.

Questions
The tools related to our work will be explained in Related Work section. However, none of the tools above show the impact of GDP on the 3 social factors we mentioned: unemployment, criminality and emigration. Moreover, none of the tools show a visualization that joins more than 2 indicators. So, by using many idioms, we expect to gain an overall perspective of the variations through the years for all countries, where different political decisions were made, and consequently different outcomes resulted as well.

These are the questions we proposed in Checkpoint 1:

- What is the evolution of each country's GDP per capita?
- Which countries have a GDP per capita higher than the EU average?
- Has the country's GDP per capita any correlation with its geographical location?
- Do countries with lower GDP per capita have bigger emigration fluxes?
- Do individuals who have a lower income situation report a bigger rise in criminality rates in their area in comparison to higher income areas, when the GDP decreases?
- Is there an age group more affected by unemployment when there is a GDP variation?
- Does the increase in unemployment result in an increase of emigration?
- Does a variation on GDP affect unemployment in the same way for people with different education attainment levels?
- Do the beliefs (left/right) of a political party influence the evolution of the country's GDP?

New questions raised:

- Has the country's GDP per capita any correlation with its geographical location?
- How is the total reported criminality in a country affected by its GDP variation?
- Were countries with higher GDP less affected by the debt crisis in terms of unemployment?

RELATED WORK

It's worth noticing that most of our inspiration came from classes, by learning the proper ways to visualize types of abstract data and then applying these. Even the ones we created were a mix of intuition and knowing the best channels to represent different types of attributes. Still, we checked some work similar to ours in order to guarantee that no one had already implemented our idea and to try finding new insights.

There is a scientific online publication by Oxford University, called Our World In Data [2], that focuses on large global problems such as poverty, hunger and inequality. There is a section for Economic Growth [3] where they compare it with indicators like child mortality (health), expected years of schooling (education), as can be seen in Figure 1, and productivity which are also key elements to a good quality of life according to Eurostat [1], but they don't compare the three social dimensions that we mentioned above. They use similar approaches to ours, using parallel coordinates charts, line charts and choropleth maps to visualize the data.

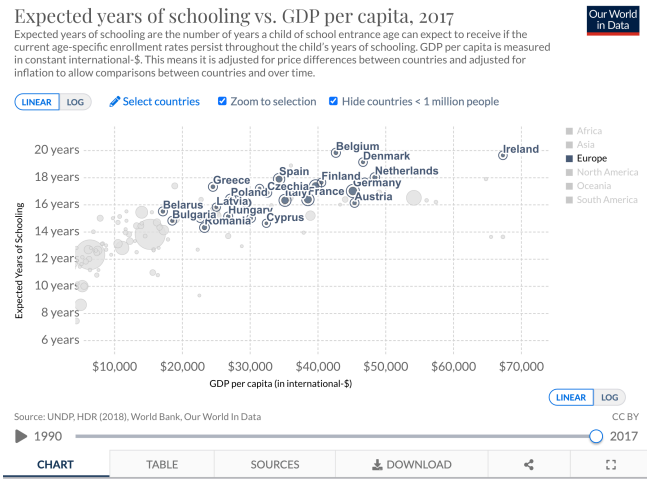

Figure 1. Expected years of schooling vs GDP per capita for EU countries, 2017 (Our World in Data)

They also allow to see an animated evolution through the years and selected countries for the defined indicators.

The Eurostat website has many databases [4] about economy, social conditions, environment, science and others. That's where we ended up getting most of our information. Along with these datasets it provides simple visualizations for each dataset. We can select the countries and the years we wish to visualize. For instance, if we have a dataset with the unemployment rate by sex, age and educational attainment level of the EU countries from 2008 to 2018 [5], they have three options of visualization: bar chart, line chart and choropleth map, as can be seen in Figure 2. We could also have filtered the dataset by sex, age and educational attainment and visualize different data. Still, none of their visualizations tries to find relationships between indicators in the same idiom and they don't show more than one idiom in the same visualization, which are things our work does.

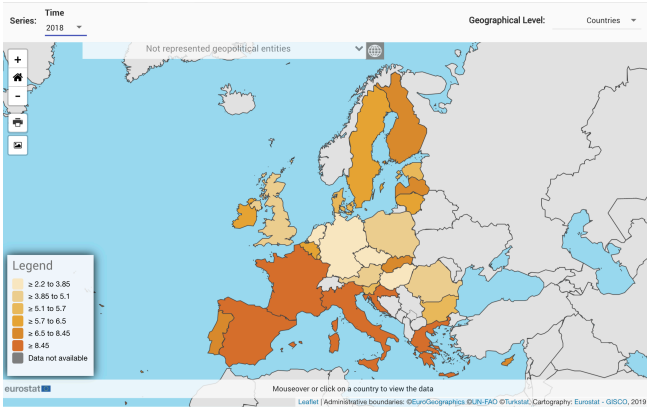

Figure 2. Unemployment rate by sex, age and educational attainment level in EU countries, 2018 (Eurostat)

Figure 3. Layout of the visualization. Year: 2011; Countries: France, Spain, Deutschland, Finland

THE DATA

We needed data about GDP, elections results, total population and 3 social indicators: unemployment, emigration, and criminality. Apart from the elections results [10], we found all our needed datasets in Eurostat's website [6],[7],[8],[9], all of them in a csv format.

We had to make some cleaning in our datasets before joining them, such as deciding the level of attributes' granularity that we deemed necessary, dealing with countries in excess and making sure all years we wanted to cover were there.

The GDP per capita dataset was the least problematic, as we had only the countries, the years and their respective GDP per capita.

The emigration dataset had the number of people emigrated, as a total, per gender and age, we discarded those last two.

In the unemployment dataset, we had gender, age, and education granularity for unemployment rates. We discarded the gender.

The criminality dataset had attributes for the percentage of total criminality, as well as the criminality felt in high- and low-income areas. We decided to keep them all.

From the elections results we kept EU countries and the parties that had the majority in each election. Those last values were encoded from 0 to 10, where we later considered 0 to 5 a leftist government and 5-10 a right one.

We originally wanted to cover the 2000-2020 year interval, but the point where all the datasets could be joined without missing years was from 2008 up to 2018, hence it being our chosen data range. Besides this, we had to filter out countries that did not belong to the European Union and were present in our datasets, as we didn't want to include them in our sample. To deal with missing values, we used a sentinel value (-101), chosen as it wouldn't interfere with the domain of any variable.

With most of the attributes we derived measures: to get the percentage of emigration, we had to correlate the emigration dataset with a total country population dataset. We derived the annual growth for the GDP, emigration, and total unemployment. We derived a difference of the normalized growth between GDP and education levels in unemployment. And finally, a criminality growth difference derived measure between the low-income and high-income areas.

In the prototype, we ended up not using most of these derived measures, as they didn't suit or came as an advantage to the idioms that we formulated. We finished with 15 attributes for 28 countries for each year between 2008-2018.

VISUALIZATION

Overall Description

As can be seen in Figure 3, we implemented 5 idioms that globally represent the following: *choropleth map*, in the upper right, representing the EU countries' GDP per capita; a *parallel coordinate chart*, in the bottom left of the interface, showing the GDP per capita and its correlation with social indicators in all countries; a *line chart*, in the upper center, where we have the highlighted countries' GDP per capita evolution throughout the years; a *double bubble graph*, in the bottom right, that represents the criminality in poorer and richer areas; and a *Cleveland Dot plot*, in the upper left, where we have the unemployment by age group or education level.

We can have up to 4 countries highlighted, each one of those countries has a color associated, that is maintained in every idiom, those are yellow, orange, light green, and lilac. The average of all EU countries, *EU28*, appears permanently in all idioms in blue, dashed when represented by a line. The map idiom contains a slider, where the year that is being observed can be changed. In the upper center of the interface, we have a box with the current selected year and highlighted countries, written in white.

The selection of the countries can be done by either selecting a country area in the choropleth map, searching on the search bar or selecting a line corresponding to a country on the parallel chart. When 4 countries are selected (max selections), selecting a different 5th country removes the firstly selected one. Selecting a country that already exists in the selected ones removes it from the selections.

Rationale

Techniques and visual encodings

Starting by the *choropleth map*, this visual encoding was the obvious one for us to implement. Our visualization is focused on the EU and so it came right to mind that a map would be the perfect way to display information relative to each country and that this map could be used as a logical input channel for the user to select different countries. This map makes use of different saturations of the same color to represent the different values of GDP per country, having four levels of color for each of the < 10K GDP, 10K – 20K GDP, 20K -30K GDP and >= 30K GDP levels. By toggling the Political Party button, this color scheme is replaced by a red and blue scheme, where each country lead by a left type of party is filled with the red color, and the one's lead by a right type of party being filled with the blue color. In addition to this information, by having a geographical representation of each country, the user can perceive different understandings on how the location of each country in the EU can also be related to its GDP and Social Indicators values. A search dropdown is provided, where the user can select of search for a country by name instead of doing so directly on the map. Finally, in this idiom the user makes use of a range selector, where the currently selected year is updated according to the user's input.

The *line chart* was the other contender for the "obvious idiom that came to mind reward". The GDP evolution is best represented with a line and having the possibility to select up to 4 countries, the comparison between different colored lines is made easy on the eye and allows for the user to compare and conclude on the differences existing between the countries GDP values almost instantly. In addition to comparing GDP values between countries, this graph also allows, with a toggle button behavior, to place on the lines circle marks that showcase the type of political party in charge in each year. And so, by toggling the political party information, the user can also understand how one country's GDP evolved in relation to the political party throughout the period it was on charge.

Our data had information regarding three different social indicators (Emigration, Criminality and Unemployment). For the user to see how these three correlated between each other and with the GDP, an idiom that represented four different variables of data was needed. The *parallel chart* originated from this need and it allows the user to compare between the four attributes. In this graph, the user can select each line, updating the selected countries list. The also implemented brushing tool extends the insights one can get from visualizing the graph, by filtering and highlighting the brushed range relative to each of the attributes.

The *Cleveland plot* focusses on the unemployment indicator, which allows the user to deeper analyse the data regarding this indicator. This graph has two toggle buttons that allow the user to compare the levels of unemployment per country to either its population's *education* levels or *age* ranges. We display three different marks per country, and we chose shape and spatial location as our channels as these are the most effective ones for categorical attributes. These attributes are Primary, Secondary and Tertiary education levels when the *education button* is toggled, or 15-24, 25-39 and 40-64 age ranges when the *age button* is toggled. The different shapes are circle, square and triangle.

Finally, the last graph was created by the group and we call it the *double bubble graph*. This graph, focusing on criminality, allows the user to compare the levels of criminality between poorer and richer areas of each country. The data is represented with two overlapped filled circles and so, by analysing which circle is the biggest, one can understand which type of area experiences bigger levels of criminality. The overlapping of circles creates a donut/ring. The bigger the donut/ring, the bigger is the difference of criminality levels between the poor and the rich areas. Even though people are not very good at estimating values encoded by different areas, we know that bubble size perception depends on adjacent bubbles, and so by having them side by side as we do and by observing that

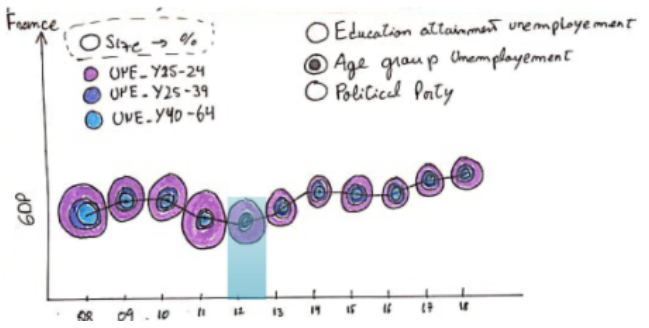

Figure 4. Unemployment with GDP mixed chart

the values for the EU28 average seldom change, we can get an idea of which countries were most affected and the gap between the different income areas.

Evolution of the prototype

Prior to the implementation of this visualization, the conception of its different idioms went through different stages and changes.

The *choropleth map* initially did not represent the political party information and did not provide a search bar. The representation of the political party adds another level of integration between idioms and helps the user to understand the geographical correlation between countries and their political party tendencies. The incorporation of the search bar was a must since most of the users do not fully know the location of each country in the map, facilitating country selection.

The initial idea for the *line graph* had the data of the GDP combined with data related to unemployment and political party, as seen in Figure 4. This was not ideal since two data attributes were constantly hidden. The comparison between two or more countries was made impossible with so much information and bubbles on the screen. Furthermore, the GDP and its evolution were one of the focus points of our visualization, and having this information obstructed as seen in Figure 4 just did not feel right.

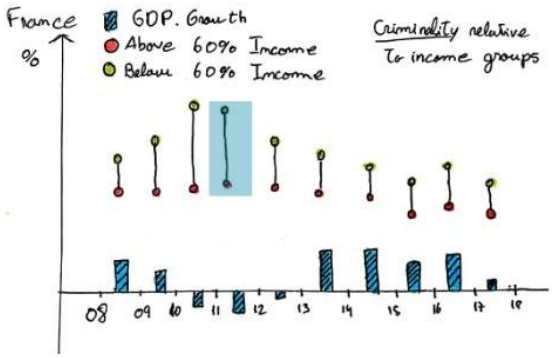

Figure 5. Criminality with GDP growth mixed chart

The graph the represented *criminality* data also changed. The initial idea was to have the criminality being compared with GDP, as seen in Figure 5. The main issue was that having negative and positive values of GDP evolution (bars) could be misleading and confusing, even more so when compared to the information regarding the differences between low- and high-income areas (double dot line).

Most of the initial ideas focused on having the GDP information represented in all the idioms. We did this because we were not seeing the visualization as a whole, where two different idioms could be used together to reach an understanding. With our current implementation, there is a patterned separation between social indicators and GDP, and the grouping of both. Now one can directly compare GDP and social indicators with the parallel chart, or indirectly do so by analysing the different idioms that focus solely on one specific social indicator. Furthermore, the final idioms allow for the user to always be able to analyse data of several selected countries with this data not being clustered or obstructed, making for easily comparable in-graph-data.

Demonstrate the Potential

When answering the question: "Is there an age group more affected by unemployment when there is a GDP variation?" we can do so by analyzing three idioms.

In the initial year of 2008 of the visualization, Portugal is 4 years out from its lowest GDP value in 2012, as can be seen in Figure 6. Here the unemployed percentage of people in the age group of 15 to 24 years old is around 15% and for the other age groups this value is around 8%. When Portugal reached its low in terms of GDP in 2012, the percentage of unemployed population between 15-24 years old increased 25%, skyrocketing to close to 40%, while the other age groups saw a much lower rise of unemployment percentages, reaching around 14% of unemployment which means an increase of around 5- to 10%.

Another question was: "Were countries with higher GDP less affected by the debt crisis in terms of unemployment?" As can be seen in Figure 7, the higher GDP countries are close to the EU unemployment rate average in 2008. In 2012, after the peak of the crisis, the average is significantly higher than most of these countries, because lower GDP countries were more affected.

Another question was: "How is the total reported criminality in a country affected by its GDP variation?" By analyzing the graphs in Figure 8, we can see that despite Latvia (light green) having a practically constant GDP evolution, its criminality levels dropped significantly. On the other hand, Ireland (orange) who saw a significant rise in its GDP, experienced little to no changes regarding its criminality levels. We can conclude that, contrary to common belief, the criminality does not depend directly on the GDP evolution of a country.

Figure 6. Unemployment rate variation, 16-24 age group,

IMPLEMENTATION DETAILS

One of our teammates was already knowledgeable about JavaScript, HTML and CSS and none of us had ever used D3.js. So, in implementation terms the first challenge was to learn the tools we needed to do our work. What sometimes happened to us is that things were working or not working, and we didn't know why. So, we had to first understand the concepts behind what a select and an append were for instance, so that we could then organize our idioms the way we wanted them, and not just copy code. We also faced some problems when designing the dimensions of the idioms to be adapted to different computers. We were using the px CSS units and so it was showing differently on different computers. We later adapted to the viewport height (vh) and width (vw).

To integrate the views, we had to maintain some global variables such as the current year, the selected countries and whether the political party button was selected. Then, every time something changed, we called a general update function that updated every idiom using these global variables. For example, to do brushing on two or more attributes, we had to store the limit values for each attribute, so that when we were drawing on the next update, we know which countries satisfy all the constraints.

Once we understood the logic behind drawing something at a specific location from a certain dataset, using certain scales, we were able to draw all the svg's. We started by drawing the line chart inspired on the code provided by the teachers. It was a good opportunity to experiment and understand what certain commands do. The parallel coordinates [11] and the map were adapted from examples on the web. The others were implemented from scratch.

Using the experience gained on the previous three, we already understood the logic. On the Cleveland Dot Plot we draw circles, squares and triangles. In the Double Bubble chart, we had to draw first the bigger bubble, so that it didn't superimpose the smaller one.

CONCLUSION & FUTURE WORK

We did end up responding to all of our original questions, we created a choropleth map where we can analyze distributions and trends geographically, a parallel coordinate chart where we can respond to all questions about social indicators, in their total percentage. For more detailed insights about the social indicators we have the unemployment's Cleveland plot and the criminality's double bubble graph. The line graph shows us the evolution of GDP per capita throughout the years, that beyond its own value, can help us extract perspectives about the data from other graphs. Besides this, an average of the EU28, was constant in all the idioms.

With time and funds, we would scale our prototype to the world, instead of just the EU, as all indicators are foreseen to have really distinct values between continents, this would allow us to analyze the contrasts and correlations between indicators better. Besides this the number of social indicators could be expanded. Another thing would be perfecting the transitions so that our visualization could run smoother. We could also use it to hire a web designer, so that the look&feel could be even more appealing.

If we were to start over, we would be more knowledgeable about the process, so we would spend less time on collecting data and treating it, and more on thinking about which idioms best answer our questions.

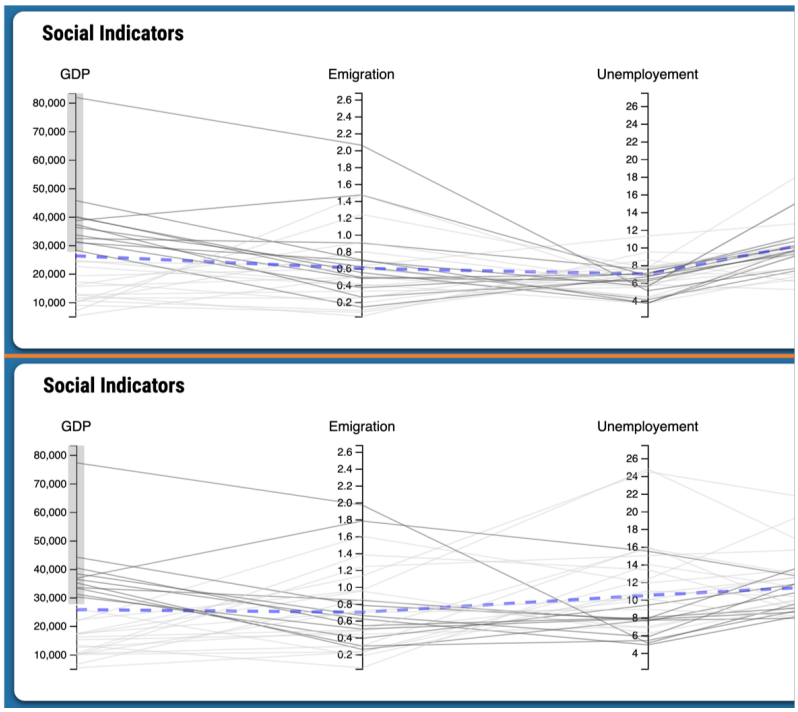

Figure 7. Comparison unemployment for higher GDP countries, 2008 (top) and 2012 (bottom)

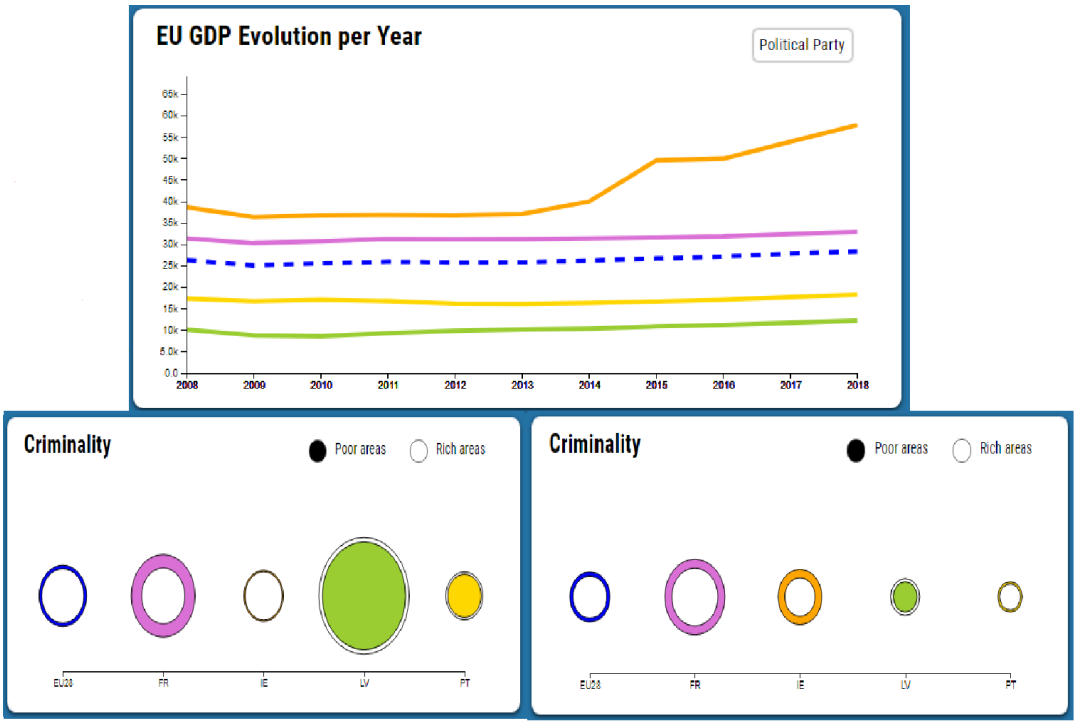

Figure 8. Criminality evolution with GDP variation, 2008 (bottom left) and 2018 (bottom right

REFERENCES

1. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Quality_of_life_indicators
2. https://ourworldindata.org/
3. https://ourworldindata.org/economic-growth
4. https://ec.europa.eu/eurostat/data/database
5. https://ec.europa.eu/eurostat/databrowser/view/lfsa_urgaed/default/map?lang=en
6. https://ec.europa.eu/eurdostat/databrowser/view/sdg_08_10/default/table?lang=en
7. https://ec.europa.eu/eurostat/databrowser/view/ILC_MDDW03__custom_52281/default/table?lang=en
8. https://ec.europa.eu/eurostat/databrowser/view/tps00177/default/table?lang=en
9. https://ec.europa.eu/eurostat/databrowser/view/lfsa_urgaed/default/table?lang=en
10. http://www.parlgov.org/data/table/view_election/
11. https://www.d3-graph-gallery.com/graph/parallel_basic.html