1. A description of the data

We acquired the data of this project from a website called “eurostat”. It’s an official website which provides extensive categories of European statistics. The three data files that we downloaded are in csv formats, and include data of European unemployment rate of youth, GDP and dropout rates of European countries. In terms of the unemployment and dropout rate, the corresponding csv files contain eight columns of data ranging from time, country to age and value, etc. As for the GDP csv file, it has six columns of data including time, country, value and so on.

Since we’re focusing on the correlations between unemployment rate and the two variables of GPD and dropout rate, the raw data needs to be edited and reformatted in order to efficiently extract the relevant data we’re focusing on, which should be the year, country name and value of the variable in general. The approach that we’ve chosen to process the raw data is to use the csv module provided by Python to write some scripts which read the original data and filter it based on the criteria to precisely select the data that we want (year, country and value), and then write the filtered data into a new csv file, which basically results in the truncated version.

(I think Priya also has acquired some data with regard to shape files for maps, so some specifications about where the data was found and how to use the data should be elaborated as well.)

1. A description of the mapping from data to visual elements

For the third visualization, the scales that I used are linear and ordinal scales. A single rectangle is generated for each of the three variables per country and per year. Since there’s no space between two adjacent rectangles, we chose the ordinal scale to generate the discrete x value for each rectangle. The same approach can also be applied to generate the scale for y values. In order to show the tendency of the changes reflected on one specific variable with the elapse of time, apart from line graph and bar chart, we came with something that is equally intuitive. We used the changes in the shade of color to indicate the changes in numeric values. In other words, we’re trying to map the numeric value of each of the three variables into the shade of the color.

For example, in terms of the unemployment rate, the greater the value becomes, the darker its corresponding color should be. The color scale is chosen as a linear one. It’s important to note that with respect to each factor (unemployment rate, GDP and dropout rate), the color scale should be the same across all the different countries so that it’s intuitive to reflect the numeric differences among these countries.

1. The story

The first visualization gives us the first look of some of the European countries that have relatively high rates of unemployment, such as Greece and Spain, as well as those countries with low rates, like France and Germany. Since the unemployment rate we presented in this map is fixed to a certain year, we would also like to gain a broader view as to how the unemployment rate of youth people changes from year to year so that we can pick the country that is worthwhile to research on. From this plot, the country that we had chosen to continue delving is Greece, Spain, France and Germany.

Talking about showing trends or tendencies, the first form of visualization that comes into our mind would be the line graph. It’s very intuitive and is pretty good at showing the relationship between variables. Therefore, we decided to plot the unemployment data using the line graph. As we can see from the graph, Greece and Spain have undergone a dramatic increase ever since the year of 2008, which is the exact year when the European economic crisis bursted out. In terms of Germany and France, theses two countries do not seem to be severely affected by the crisis, and even surprisingly, Germany’s unemployment rate of youth people decreases as time move on. The distinct performance of the four countries after economic crisis led us step further into finding out the rooted cause of the high unemployment rate, in other words, what are some of the factors that might exert an impact on it.

Since we’re focusing on the unemployment rate of the youth people in Europe, the education level of each individual is definitely something we should take into consideration. We would like to see whether there’s an impact of dropout rate on the unemployment rate of the youth. The assumptions that we made is the relationship between these two variables should be positively correlated, which means that the unemployment rate should increase as the dropout rate goes up. Another factor that we considered is the GDP, which is one of the most important indicators of economic growth of each country. We used the approach described previously and the relationship of multiple factors can be shown in a single graph. It’s quite intuitive to see that GDP and unemployment rate are negatively correlated, which means that the decreasing GDP will result in an increasing rate of unemployment. However, France seems to be the only exception that the bigger value of GDP doesn’t have a prominent impact on the unemployment rate probably due to the fact of the differences of each country’s own policies. However, it’s still quite convincing to say, from the rest countries that are chosen as representatives, GPD does exert an impact on the unemployment rate of youth people in Europe. The lower values of GDP tend to result in higher values of unemployment rate, which in accordance with our original assumption. As for the dropout rate, we’re surprising to find that the change in dropout rate as time moves on has presumably become a trend across Europe since the four countries share the same pattern that no matter how the other two factors vary, the dropout rates always go for a gradual decrease, which naturally makes us wonder if the number of people dropping out of school goes up year by year, then why the unemployment rate of some country still increase. We think this is somewhat unreasonable. We researched into this phenomenon, and we think the reasoning behind this is that there’s a growing mismatch between the skills that young people have and the positions that are offered on the job market, by attaining a certain level of education does not necessarily mean you would be more competitive compared with other people that are less educated than you. Therefore, from a statistical perspective, the dropout rate can’t explain for the increasing in unemployment rate of youth people. These two factors are not related with one another based on the visualization.