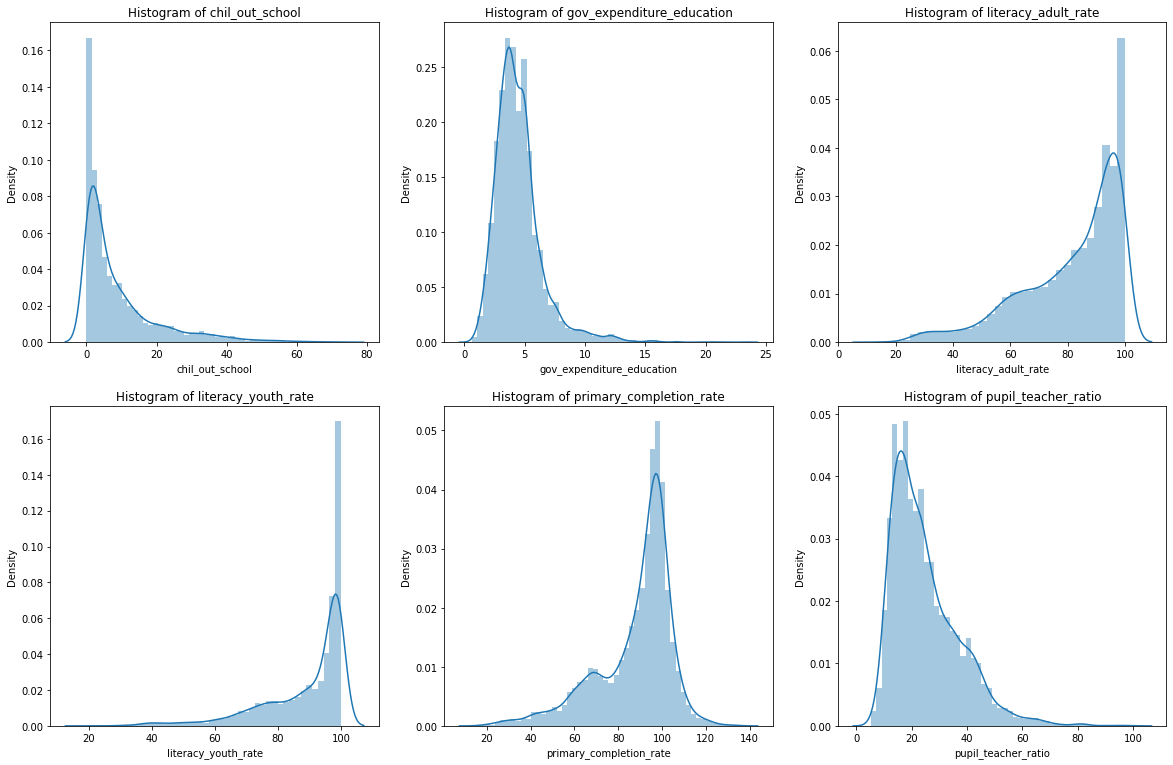
Impact of Economic Growth and Education in Mortality in the 21st Century

Exgended Analysis  
  
Team #6 | Carlos Chávez, Yohana Sandoval, Daniel Durán, Eider Giraldo, Fernán Marsiglia, Juan Álzate, Marta Ordoñez, Mateo Moreno, José López.

Distribution - Independent Variables

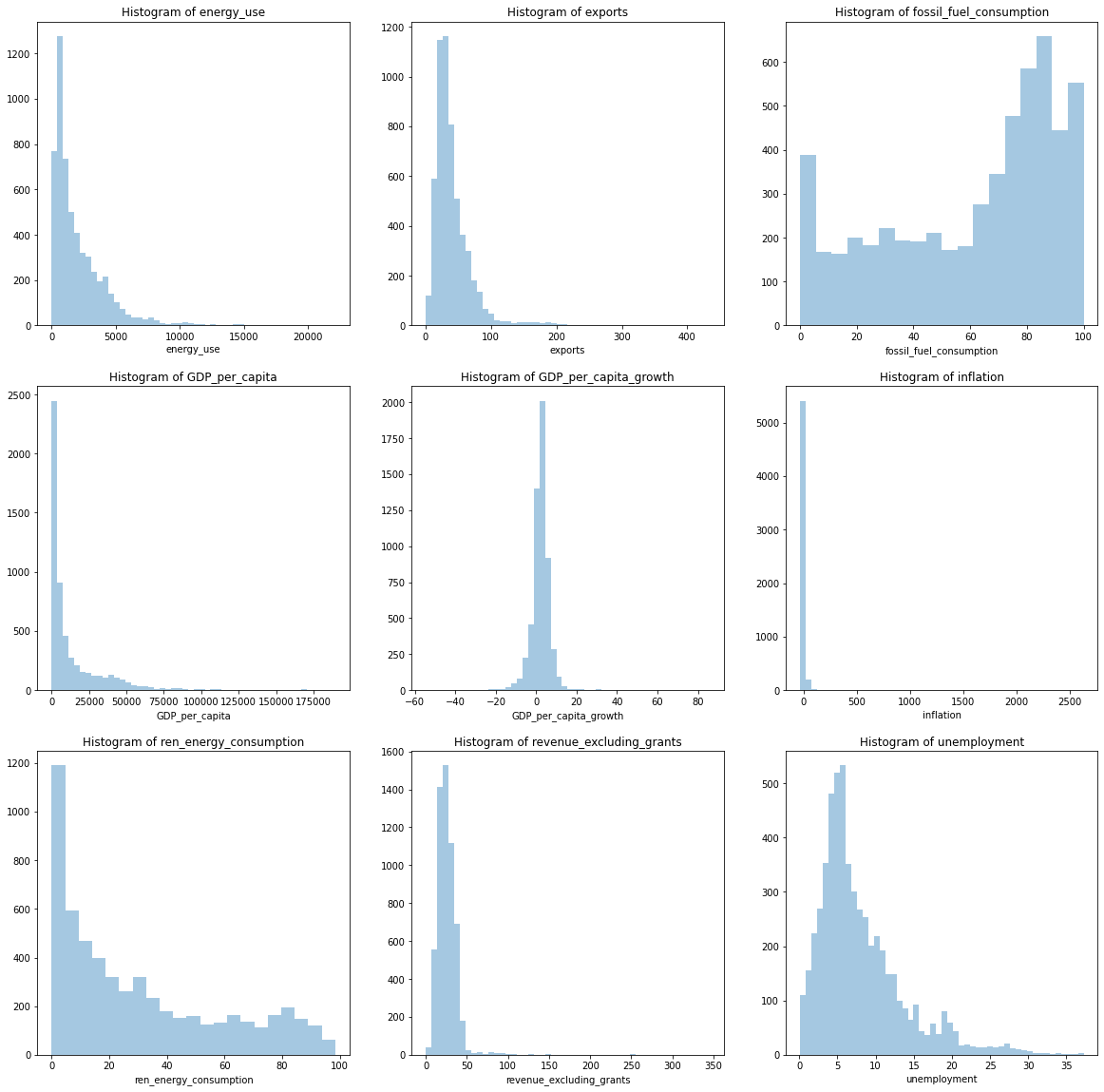
Education



1. The histogram of out-of-school children appears skewed to the right, showing that in most countries the rate of children out of school is between the intervals of 0 to 20. Data showing children out of school exceeding 40 can be evidenced. There is even some atypical data where dropout exceeds 80.
2. The histogram of Public Expenditure on Education is skewed to the right, with a significant concentration of data between intervals 0 and 5.It is also observed that there are countries with high public expenditure that exceed interval 10 and other countries with atypical data that exceed the interval 20.
3. The histogram Literacy rate (i.e. being able to read and write, performing some arithmetic data) of adults is observed skewed to the left with the data concentrated to the right between intervals 100 and 40. This means, there is a high literacy rate in the countries. In addition, it is noted that after the interval 40, there is a lower literacy rate in adults.
4. The histogram Literacy rate (i.e. knowing how to read and write, performing some arithmetic data) of young people shows us that it is skewed to the left with the data concentrated to the right between the intervals 100 and 60. This means that the highest literacy rate for youth in countries is located here. In addition, it is noted that after the interval 60, there is a lower literacy rate in young people. There is data from 2000 to 2021.
5. The elementary-school completion rate histogram is skewed to the left suggesting that the data is concentrated between intervals 120 and 40. But after interval 40 a lower elementary-school completion rate is observed.
6. The student-teacher ratio histogram is skewed to the right, evidencing a concentration of data between intervals 0 and 50. From interval 60 on, there is data observed from some countries where the proportion increases and also there are atypical values that relate to a higher proportion exceeding the interval 100.

In general, the education variables explored show us patterns in the visualization that were expected.

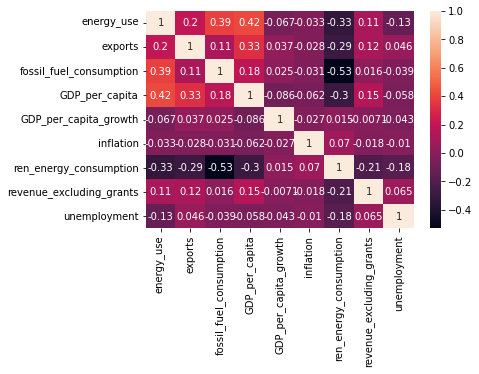
Economy



1. The energy use histogram is skewed to the right with the data concentrated between the intervals 0 to 5000. That means that, most of the countries place their energy use between these intervals. However, after the 5,000 interval and especially the 10,000 interval, there are atypical values ​​ that indicate a greater use of energy by some countries. There is data from the years 2000 to 2015.
2. The exports histogram is skewed to the right with the data concentrated between the 0 and 100 intervals. It also shows atypical data beyond the 100 interval that suggest a higher percentage of exports in some countries. There is data from the years 2000 to 2021.
3. The histogram of fossil fuel consumption indicates an important number of countries that use almost any fossil fuel. However the vast majority uses between 70% to 100% of fossil consumption. There is data from the years 2000 to 2015
4. The histogram of GDP per capita is skewed to the right with concentrated data between the intervals 0 and 50,000. This indicates that most of the countries place their GDP per capita there. In addition, there are atypical values ​​ that exceed the interval of 100,000 with a higher GDP.
5. The per capita GDP Growth histogram has data concentrated between the intervals -20 and 20. This indicates that in most countries the GDP growth is within these ranges. On the other hand, there are atypical values above the interval of 20 where GDP growth has been higher. Additionally, there are other atypical values below -20 that suggest less growth. There is data from the years 2000 to 2021.
6. The inflation histogram is slightly skewed to the right, the data is concentrated near the 0 interval, which indicates that the GDP remains stable in the majority of countries.
7. The Renewable Energy Consumption histogram is skewed to the right, saying that the majority of the couentries uses almost any renewable energy, and there are some few countries that are using a higher percentage of renewable energy.
8. The Income Excluded from Grants histogram is slightly skewed to the right. There is data concentrated between the intervals 0 and 50 and with little variability. It is observed that after the 50 interval there are high incomes and atypical values that are also seen where the income of some countries exceeds the 100 interval. There is data from the years 2000 to 2020.
9. The unemployment histogram is skewed positively or to the right. It is observed that most countries place their unemployment rates between intervals 0 and 20. It is also shown that between interval 20, unemployment is higher in some countries and there is atypical data that even exceeds the interval 35. There is data from the years 2000 to 2020.

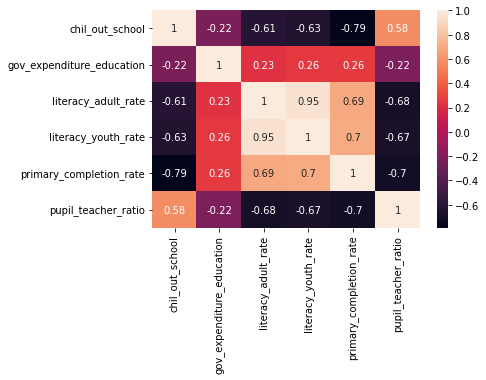
Correlation independent variables

Economy



The heat map of the economy variables did not show significant correlations between them. Except for fossil\_fuel\_consumption and ren\_energy\_consumption with -0.53 indicating a negative correlation.

Education



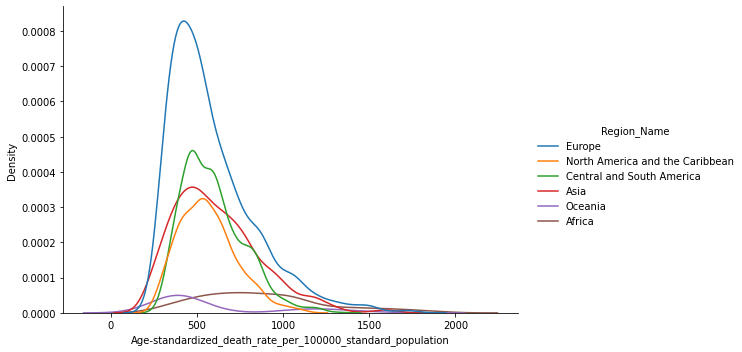
The heat map of the education variables indicates significant correlations between the variables:

* chil\_out school y pupil\_teacher\_ratio con valores de 0.58
* literacy\_adult\_rate y primary\_completion\_rate con valores de 0.69
* literacy\_youth\_rate y primary\_completion\_rate con valores de 0.70
* literacy\_adult\_rate y literacy\_youth\_rate con valores de 0.95

From the aforementioned correlations, the correlation between literacy\_adult\_rate and literacy\_youth\_rate was the most significant with 0.95.

Distribution - Dependent Variable

By region



The frequency diagram that visualizes the Standardized mortality rate by age per 100,000 inhabitants shows the following characteristics for the regions :

Europe is skewed to the right with a leptokurtic data distribution where the mortality rate is concentrated between the intervals 0 and 1000. The mortality rate from the interval 1250 is high and there are atypical values beyond the interval 2000.

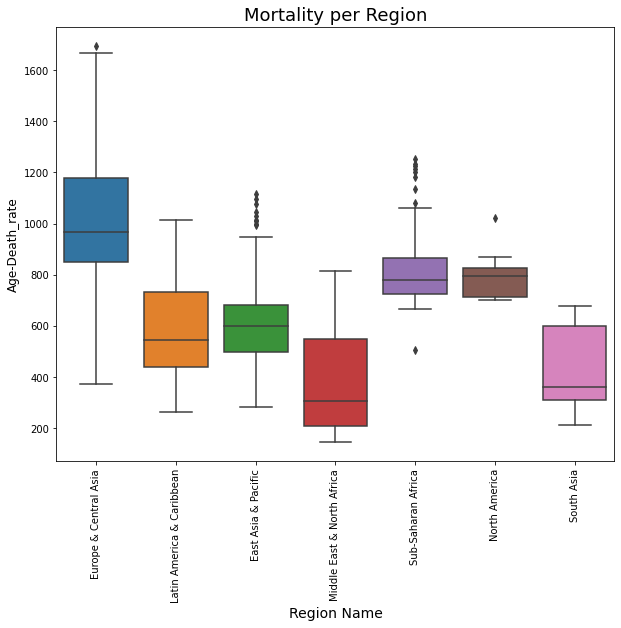
North America and the Caribbean is skewed slightly to the right, the mortality rate of most countries is concentrated between 0 and 1000.

Central and South America are skewed to the right with data concentrated between 0 and 1000, but with atypical values that are close to the 1500 interval.

Asia’s distribution is observed to be skewed to the right, its data on the mortality rate are between the intervals 0 and 1250. In addition, it has atypical values that exceed the interval 2000 and that indicate a high mortality rate.

Oceania data is observed skewed to the right, its distribution is platykurtic and its values are concentrated between 0 and 550, but from this interval atypical data are displayed that exceed the 2000 interval.

Africa is skewed positively or to the right, its data presents a greater dispersion which is expressed in its platykurtic or flattened distribution.



These box or whisker plots express the following characteristics of mortality by region:

Europe and Central Asia have a right-skewed distribution, most countries in this region have age-specific mortality rates between Q1 (interval 830) and Q3 (interval 1180) have a significant dispersion, the mean is close to interval 1000. Shows outliers after interval 1600.

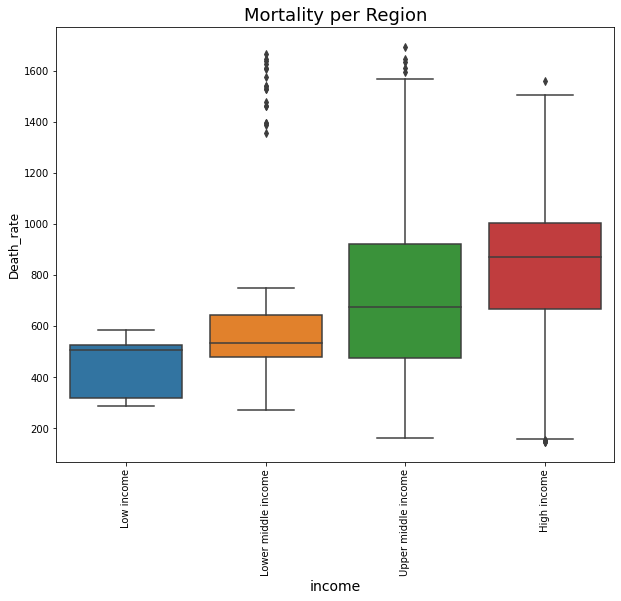
Latin America and the Caribbean whose distribution has a positive bias. The data of the mortality rate by age are concentrated between Q1 (interval 530) and Q3 (interval 780), presenting greater dispersion from Q2. It does not present outliers. East Asia and the Pacific have a slightly left-skewed distribution, their mortality rate is concentrated between Q1 (interval 500) and Q3 (interval 700). Approaching the 1000 interval, outliers are displayed.

The Middle East and North Africa have a positive bias and greater dispersion in the upper whisker. The mortality rate is concentrated between Q1 (interval 200) and Q3 (interval 570). It does not present outliers.

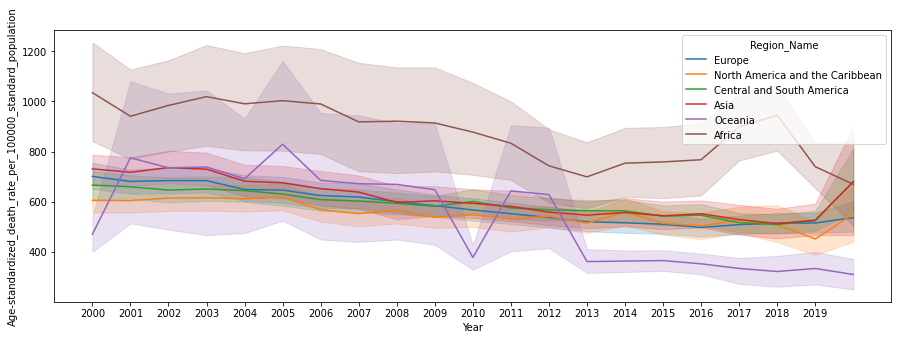
Sub-Saharan Africa is slightly skewed to the right, being more scattered in the tail, and atypical values are observed before the minimum value and after the maximum value.

North America has a negative or left bias, the mean is observed near Q3, it presents atypical values after the maximum value.

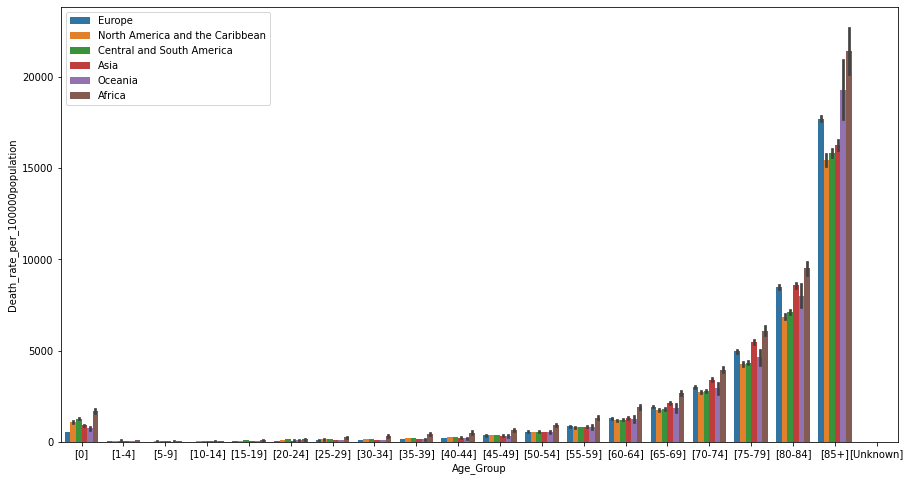
South Africa presents a very pronounced bias to the right, which generates a significant dispersion after the mean. It does not present atypical values.



This box plot shows mortality by region. We can obseve that the mortality rate in low- and lower-middle-income countries is lower than in upper-middle-income and high-income countries.



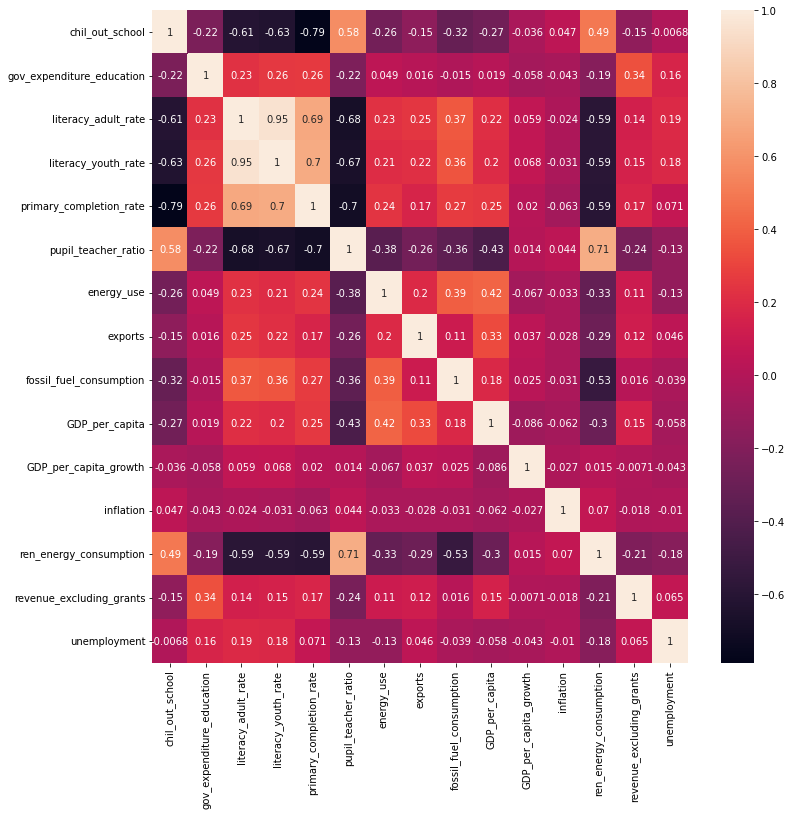
This line graph shows that Europe, North America and the Caribbean, Central and South America, and Asia have a stable standard age-standardized mortality rate per 100,000 population during the years under study (21st century). It is also noted that Africa has the highest rate with respect to the other regions. Finally, Oceania presents significant variations, but it must be taken into account that it only consists of 2 countries, which may mean that its data is smaller compared to the other regions.



This bar graph shows that the mortality rate is higher for ages older than 85 years. But mortality in newborns should also be considered, which compared to ages between 1 and 50 years is high, and Africa is the one with the highest rate of stillbirths. In general, from the age of 50 there is a gradual increase in the mortality rate.

Correlation independent variables

Education and Economy



The heat map of the education and economy variables indicates positive correlations between the following variables:

* chil\_out school and pupil\_teacher\_ratio with values ​​of 0.58
* literacy\_adult\_rate and primary\_completion\_rate with values ​​of 0.69

There are significant positive correlations with the following variables:

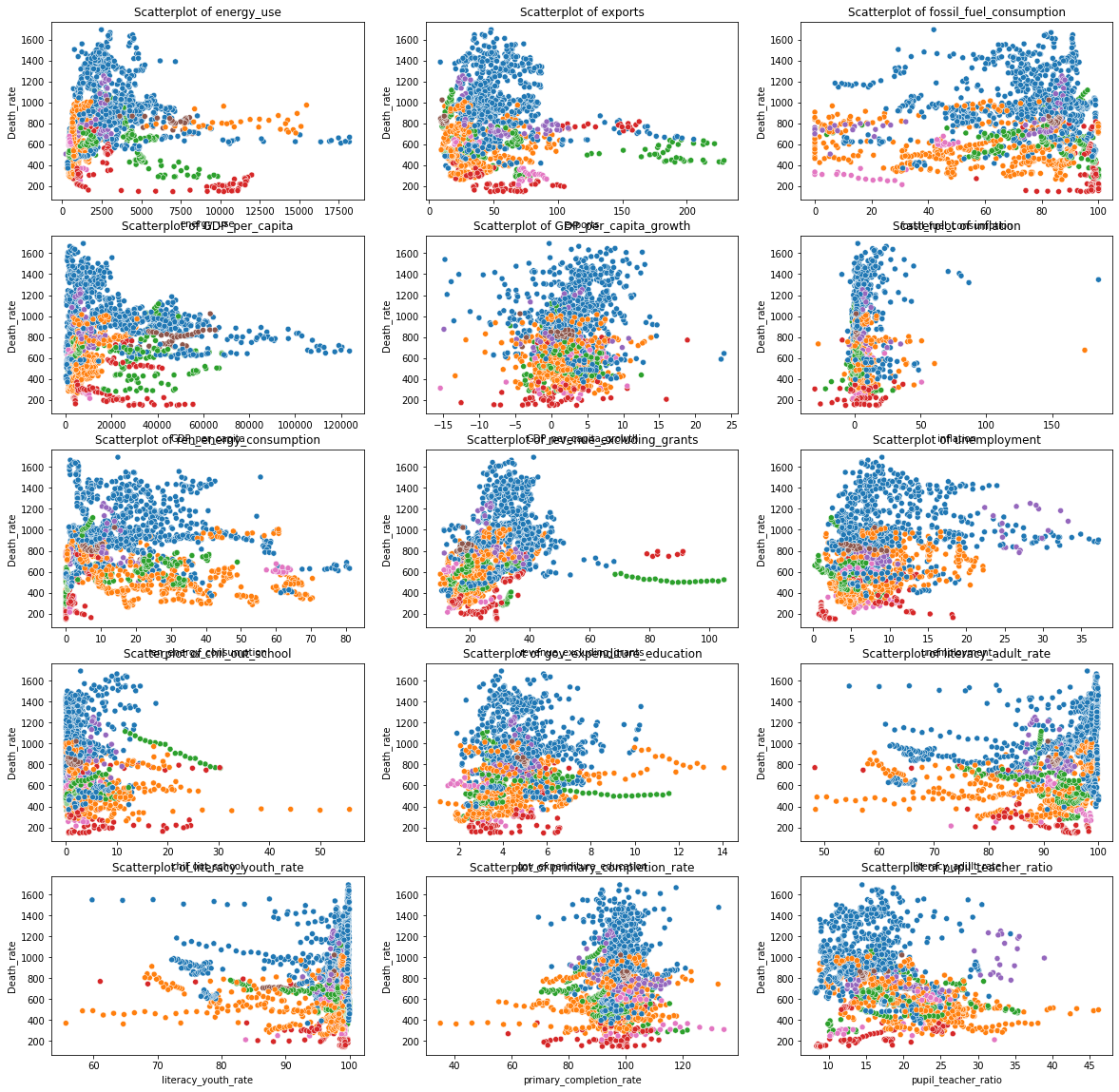
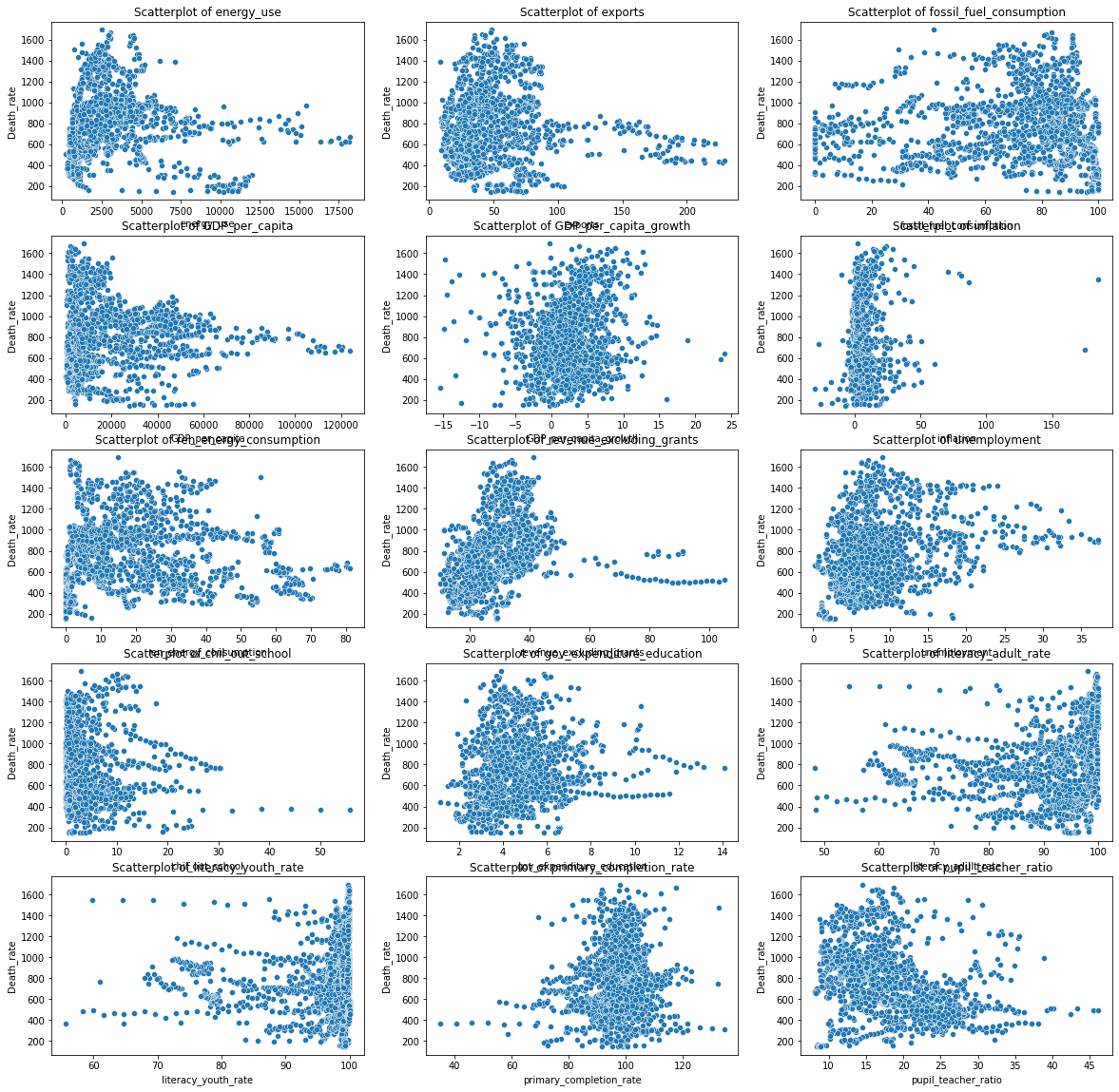
* pupil\_teacher\_ratio and ren\_energy\_consumption with values ​​of 0.71
* literacy\_adult\_rate and literacy\_youth\_rate with values ​​of 0.95

The heat map of the education and economy variables indicates negative correlations between the following variables:

* fossil\_fuel\_consumption and ren\_energy\_consumption with values ​​of -0.53
* literacy\_youth\_rate and ren\_energy\_consumption with values ​​of -0.59
* literacy\_adult\_rate and ren\_energy\_consumption with values ​​of -0.59
* primary\_completion\_rate and ren\_energy\_consumption with values ​​of -0.59
* chil\_out school and literacy\_adult\_rate with values ​​of -0.61
* chil\_out school and literacy\_youth\_rate with values ​​of -0.63
* literacy\_youth\_rate and pupil\_teacher\_ratio with values ​​of -0.67
* literacy\_adult\_rate and pupil\_teacher\_ratio with values ​​of -0.68

There are significant negative correlations with the following variables:

* primary\_completion\_rate and pupil\_teacher\_ratio with values ​​of 0.7
* chil\_out school and primary\_completion\_rate with values ​​of -0.79



In the last two images we can observe the relationships exhibited by the education and economic growth variables with respect to mortality. Here we want to highlight the change that exists between the relationships when we mark the regions with different colours. It is clear that when the data are grouped by region we see clearly marked clusters with different patterns to all the data when they are not grouped.