6210500111 Tharinrath Jatupattrapiboorn

Data

- Environment data
  - Drinking water : The percentage of people using at least basic water services. This indicator encompasses both people using basic water services as well as those using safely managed water services. Basic drinking water services is defined as drinking water from an improved source, provided collection time is not more than 30 minutes for a round trip. Improved water sources include piped water, boreholes or tubewells, protected dug wells, protected springs, and packaged or delivered water.
  - CO2 emission : Carbon dioxide emissions are those stemming from the burning of fossil fuels and the manufacture of cement. They include carbon dioxide produced during consumption of solid, liquid, and gas fuels and gas flaring.
  - Forest area : Forest area is land under natural or planted stands of trees of at least 5 meters in situ, whether productive or not, and excludes tree stands in agricultural production systems (for example, in fruit plantations and agroforestry systems) and trees in urban parks and gardens.

- Economic Data
  - GDP : is the sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products. It is calculated without making deductions for depreciation of fabricated assets or for depletion and degradation of natural resources. Data are in current U.S. dollars. Dollar figures for GDP are converted from domestic currencies using single year official exchange rates. For a few countries where the official exchange rate does not reflect the rate effectively applied to actual foreign exchange transactions, an alternative conversion factor is used.

- Population : Total population is based on the de facto definition of population, which counts all residents regardless of legal status or citizenship. The values shown are mid year estimates.
- Unemployment : Unemployment refers to the share of the labor force that is without work but available for and seeking employment. Definitions of labor force and unemployment differ by country.

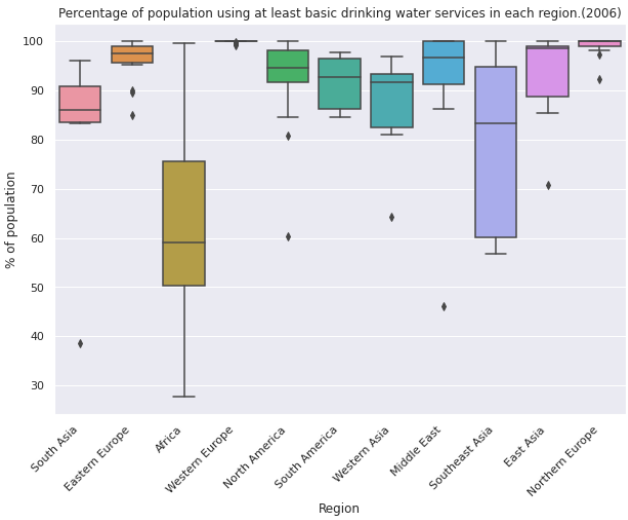
Part 1

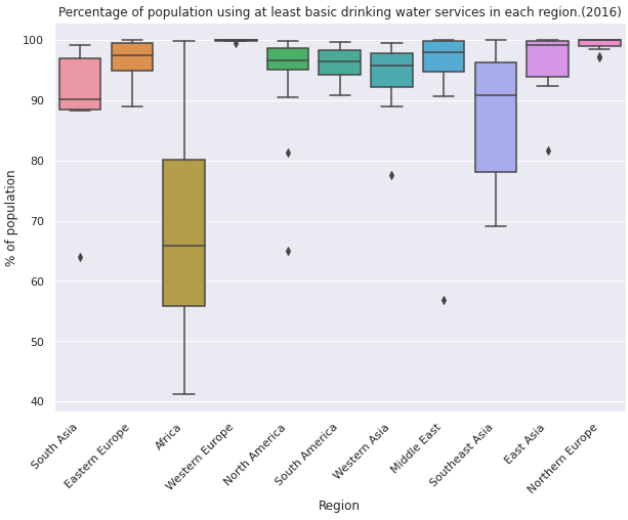
In this part, We will analyze the environment data in all regions except Australia and Oceanea and Thailand.

We already calculate and decide that we'll choose 2006 to represent the past and 2016 to represent the present.

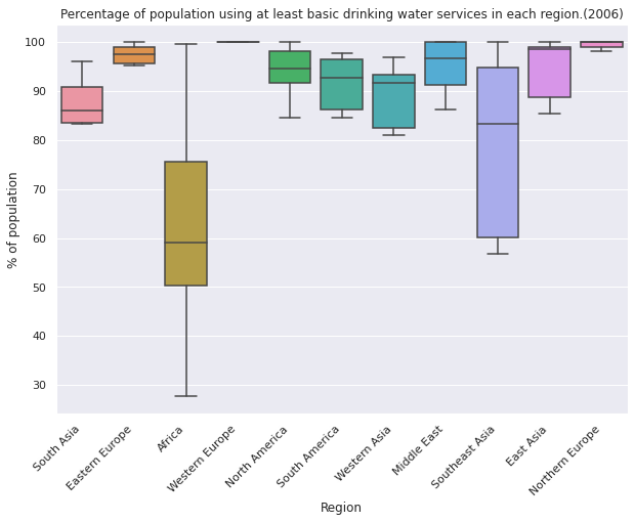
Analysis of drinking water data.

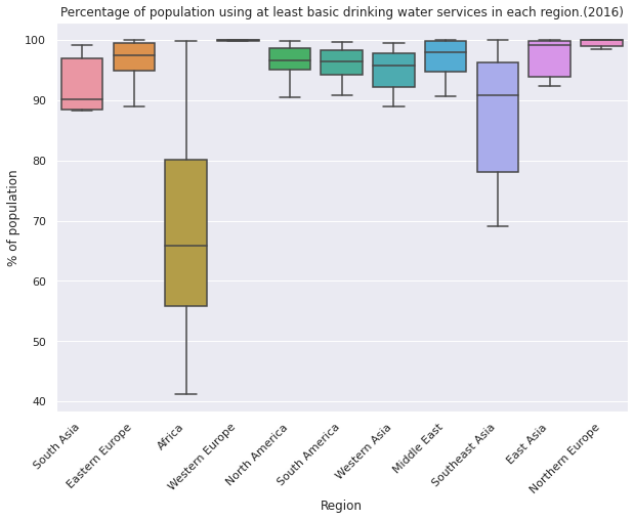
Graph with Outliers





Graph without Outliers





From 2006 graph, we can see that Africa has the lowest percentage of population using at least basic drinking water services, Western Europe has the highest percentage ,and all regions except Africa and Southeast Asia have the percentage of more than 80%.

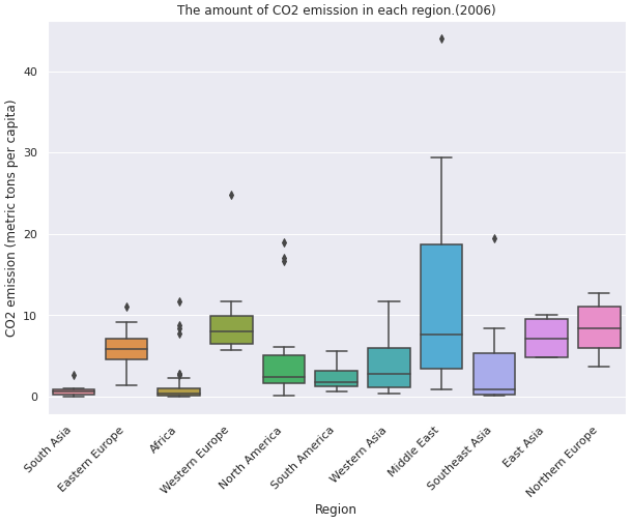
If we take a look further, Africa has the widest box plot. This is due to the fact that the countries in Africa are extremely different in terms of economics, basic utilities, growth and politics.

From 2016 graph, boxplot tends to be the same width but with a higher median. The reason behind this is that all regions develop overtime. As time passes, the governments of every country need to do better to provide basic utilities for everyone in their country.

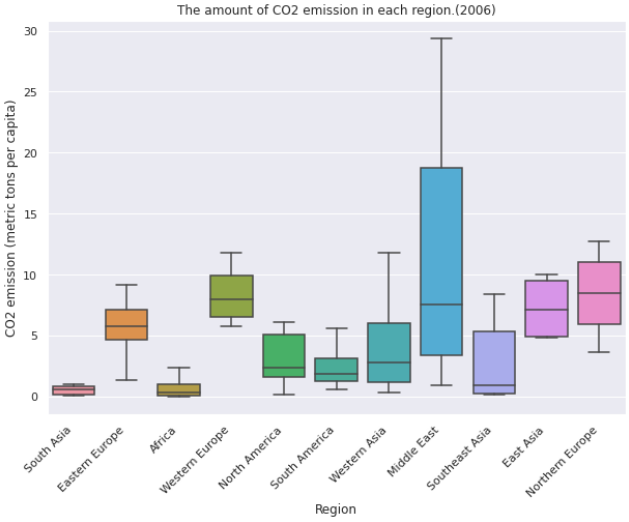
Summary : Percentage of population using at least basic drinking water services are increasing overtime. This can conclude that basic drinking water services are the important part of basic utilities that humans need.

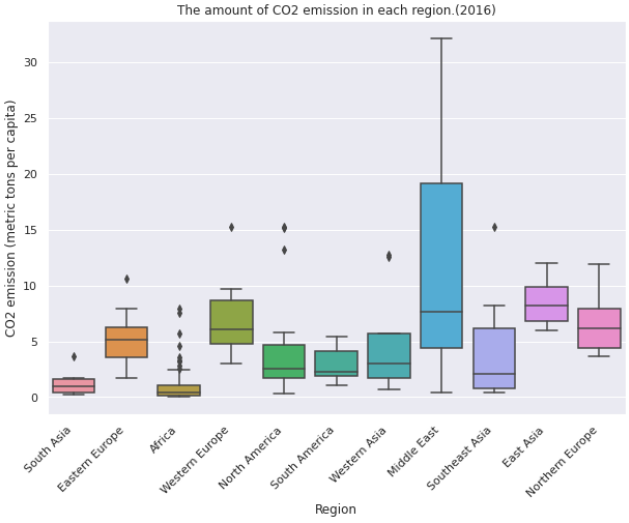
Analysis of CO2 emission data.

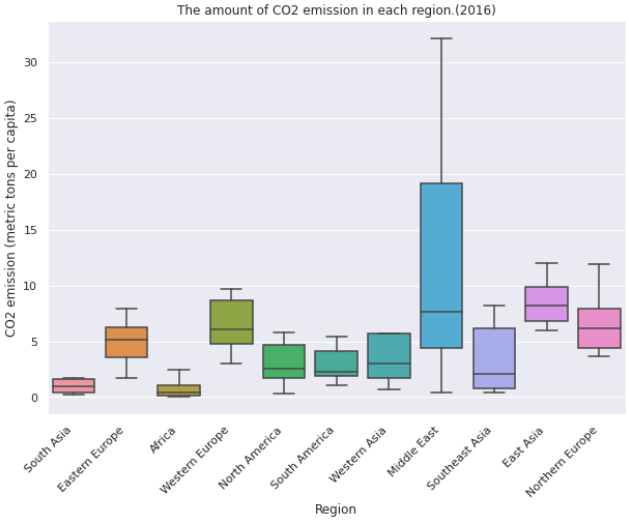
Graph with outliers



Graph without outliers.







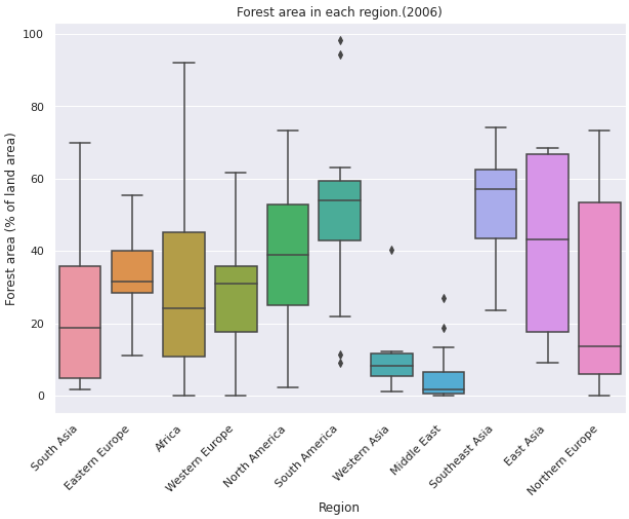
From 2006 graph, we can see that the Middle East has the highest amount of CO_2 emission. The reason behind that is the Middle East is the producer of crude oil and fossil fuels. And the region that has the lowest co2 emission is Africa. This is because the rise of the economy in Africa is still low and there is a low amount of industrial estate.

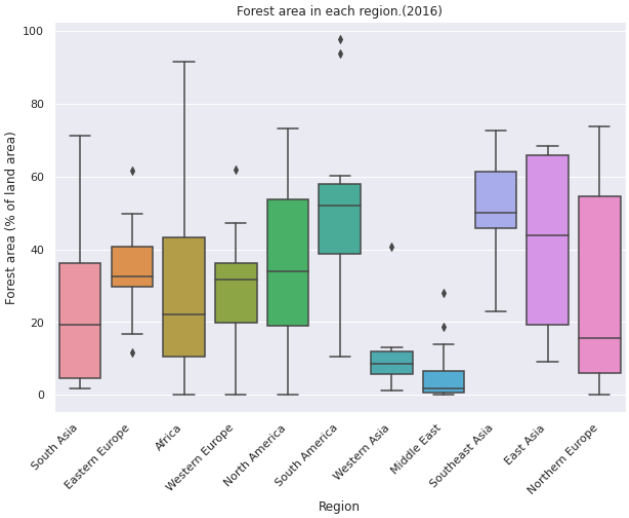
From 2016 graph, The Middle East still has the highest amount of CO_2 emission.But all regions have decreased the amount of CO_2 emission. This is because the world is more concerned and acknowledges the global warming crisis and reducing CO_2 emission is one of the ways to prevent that.

Summary : The amount of CO_2 emission decreases from time to time. Due to the fact that most of the world partakes in the environmental agreements. And people paid more attention to the global warming crisis.

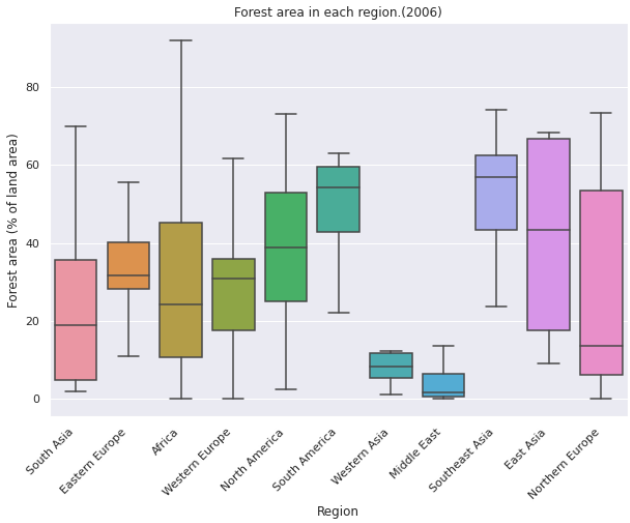
Analysis of forest data.

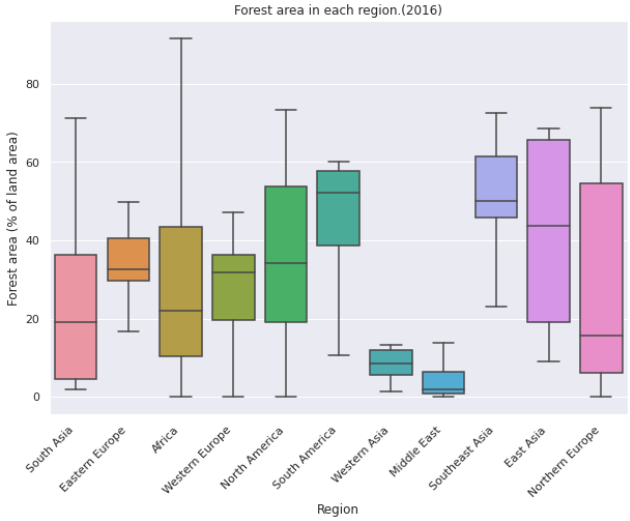
Graph with outliers.





Graph without outliers.





From 2006 graph, we can see that the Middle East has the least amount of forest area.This is because in the Middle East most of the areas are desert. Other regions have a wide range of forest area, because of the difference in terrain, policy and economics.

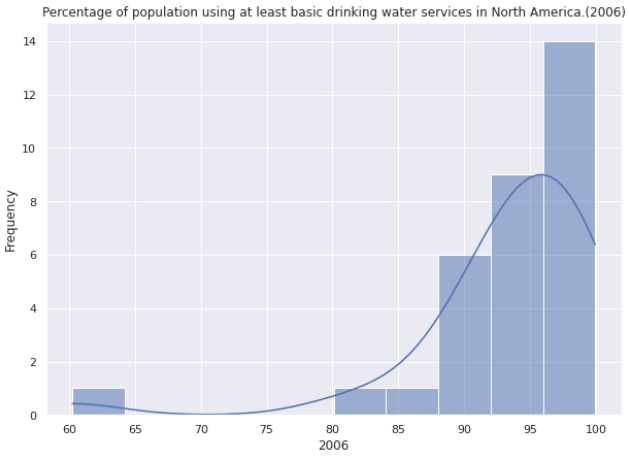
From 2016 graph, The numbers are still nearly the same because people do care about forests and the global warming crisis.But the forest needs time to regrow so the numbers are not rising.

Summary : There is no significant difference between 2006 and 2016 but it’s still a great sign because there is no decrease in forest area.

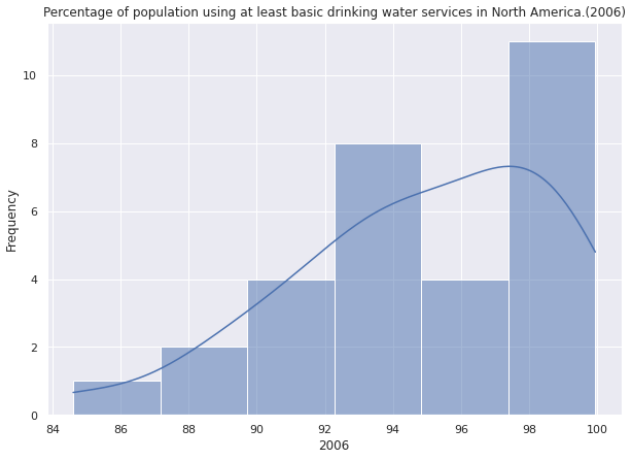
Part 2

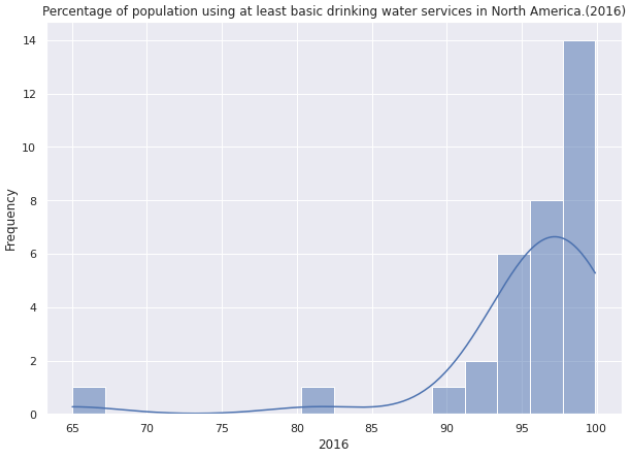
Analysis of drinking water data.

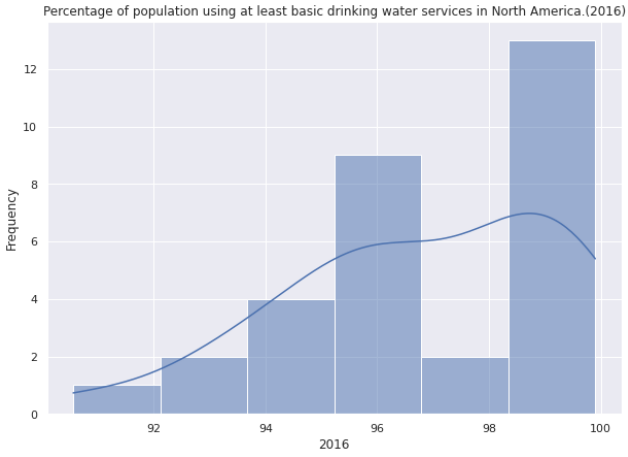
Graph with outliers.

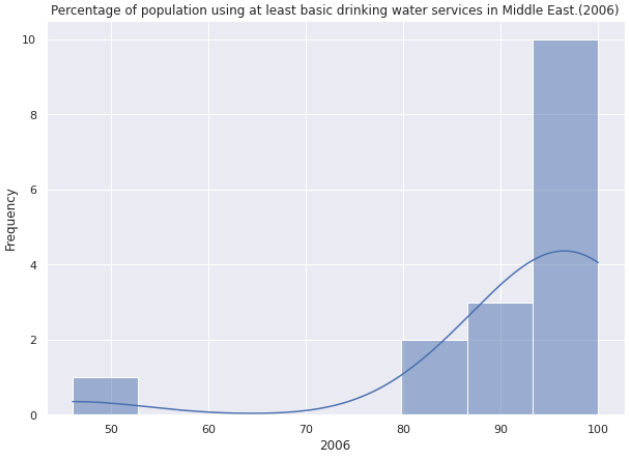


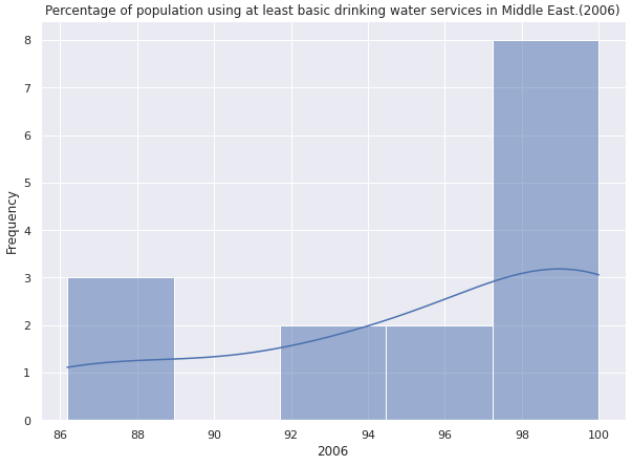
Graph without outliers.

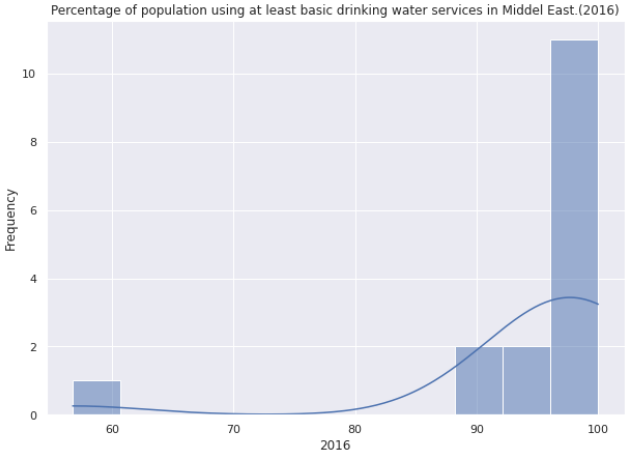


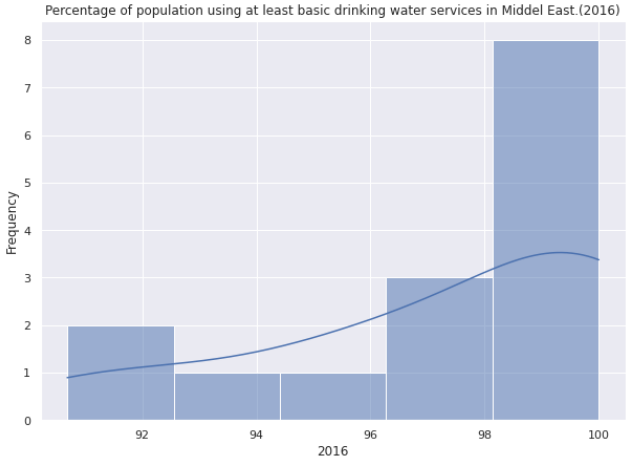


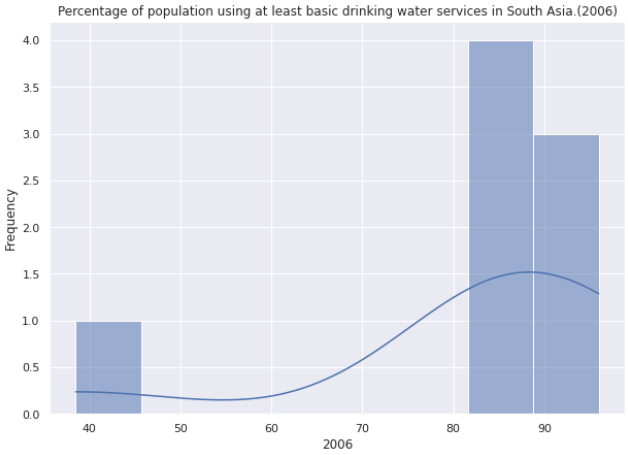


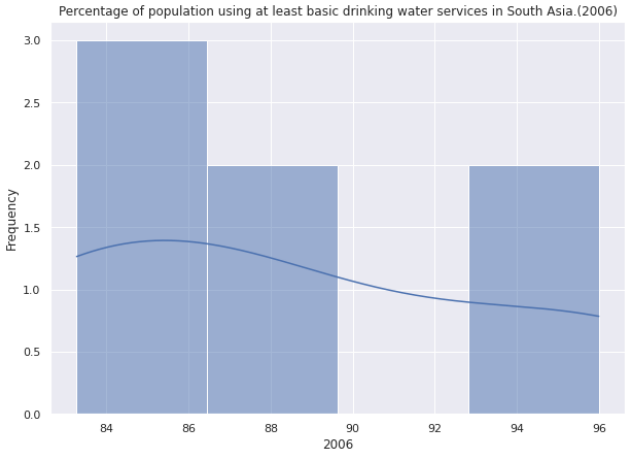


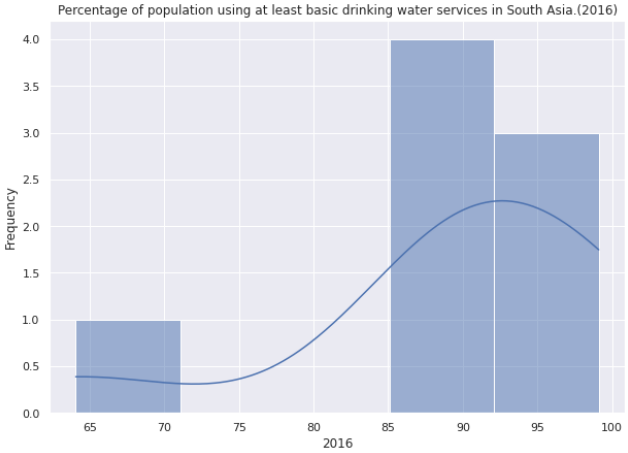


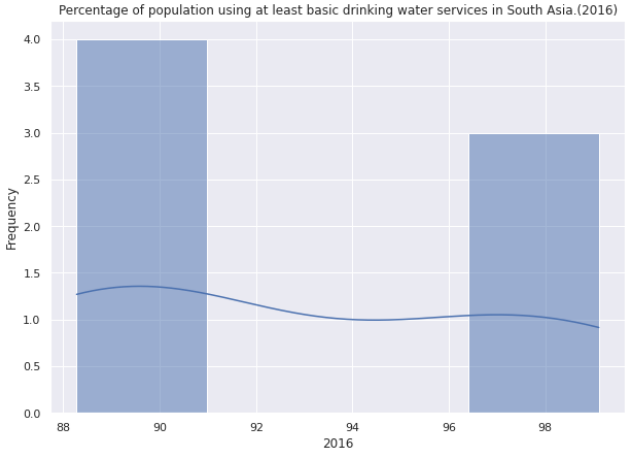

Percentage of population using at least basic drinking water services in Middel East.(2016)
Frequency
2016


Percentage of population using at least basic drinking water services in Middel East.(2016)
Frequency
2016


Percentage of population using at least basic drinking water services in South Asia.(2006)
Frequency
2006


Percentage of population using at least basic drinking water services in South Asia.(2006)
Frequency
2006


Percentage of population using at least basic drinking water services in South Asia.(2016)
Frequency
2016


Percentage of population using at least basic drinking water services in South Asia.(2016)
Frequency
2016

From all graphs, we take a look at the distribution of each region.

- North America : The graphs are left skewed.
- Middle East : The graphs are left skewed.
- South Asia : The graphs are left skewed, but close to normal.

When we compare 2006 and 2016, the percentage increases in every region. This result matches the result from part 1.

When we take a look at each region. We can see that North America has the most percentage of people using at least basic water services.
Next we have the Middle East. And lastly we have South Asia.

2006

With Outliers	Thailand	North America	Middle East	South Asia
Mean	96.279146	93.362284	92.444401	82.295577
Median	-	94.633220	96.726261	86.037700
SD	-	7.498839	13.350310	18.420752
IQR	-	8.289319	91.330006	7.230099

Without Outliers	Thailand	North America	Middle East	South Asia
Mean	96.279146	94.883665	95.535639	88.555593
Median	-	95.336790	98.107350	88.108965
SD	-	3.806806	5.210434	5.488645
IQR	-	92.776978	7.231268	8.636337

2016

With Outliers	Thailand	North America	Middle East	South Asia
Mean	99.44567	95.415305	94.749541	89.316723
Median	-	96.579888	97.971614	90.201556
SD	-	6.525568	10.589868	11.065998
IQR	-	3.573047	5.058989	8.403246

Without Outliers	Thailand	North America	Middle East	South Asia
Mean	99.44567	96.847293	97.280469	92.925557
Median	-	96.739186	98.993880	90.806744
SD	-	2.445515	3.216434	4.616416
IQR	-	3.436696	4.100798	7.926702

From the statistics data table.We can see that the mean increase in every region. This also matches the result from graphs above and the results from part 1.

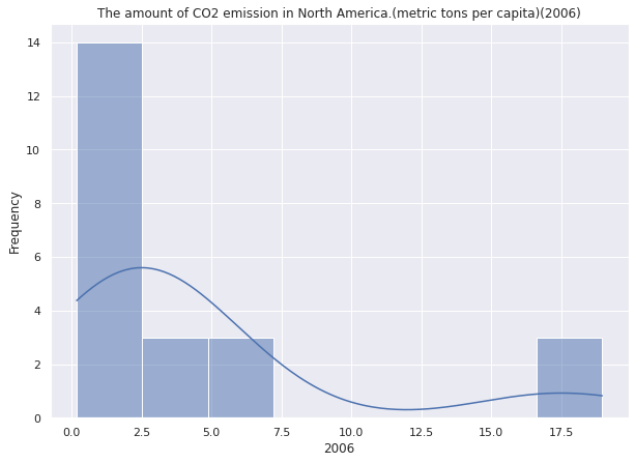
If we look through each region compared to Thailand. We can see that Thailand's average number of percentage of people using at least basic water services is above other regions.

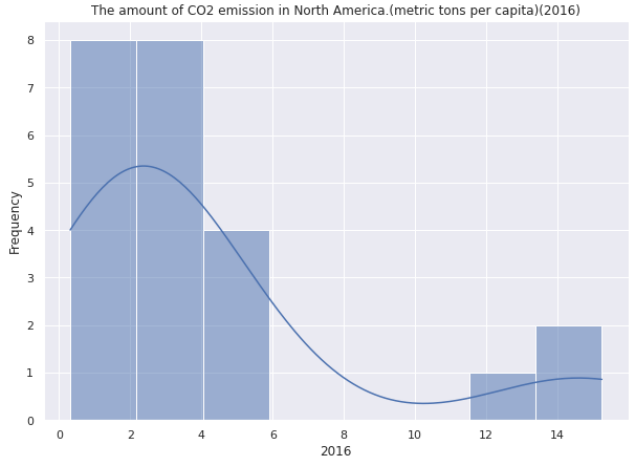
And the Middle East has higher numbers compared to North America.

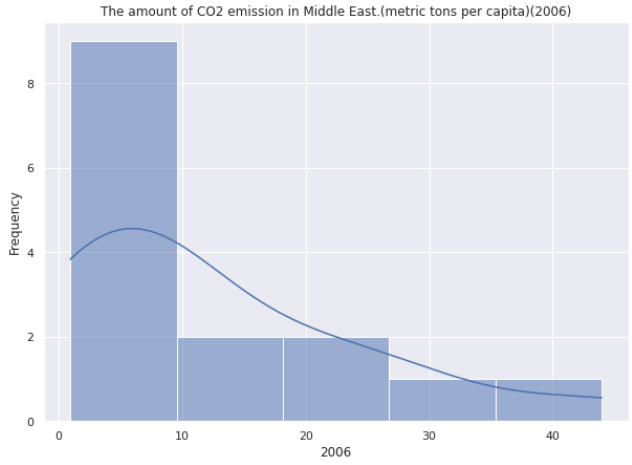
But we need to be aware that we only compute Thailand vs Region, so we can't conclude that Thailand is better than every country in other regions.

Analysis of CO2 emission data.

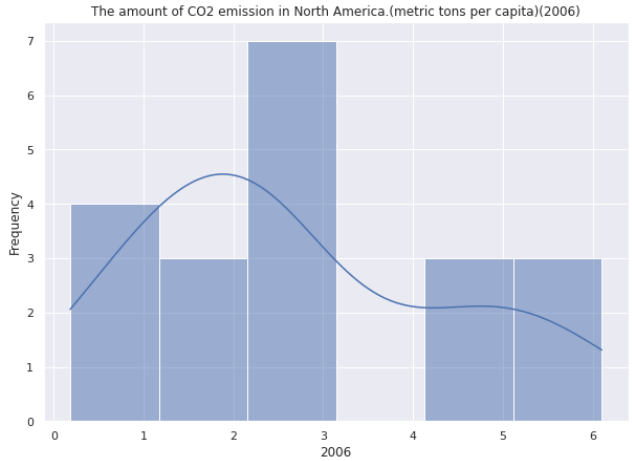
Graph with outliers.

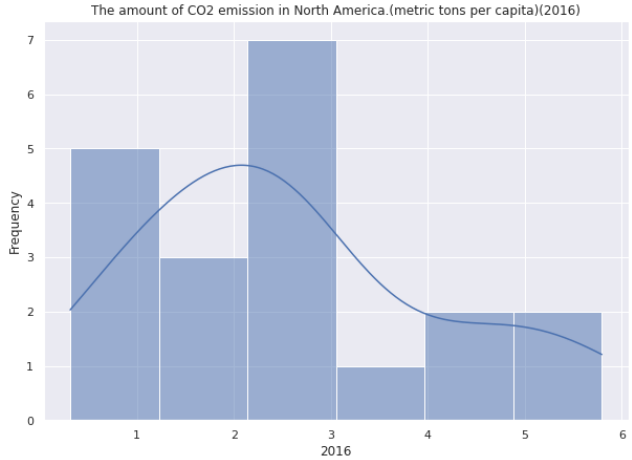


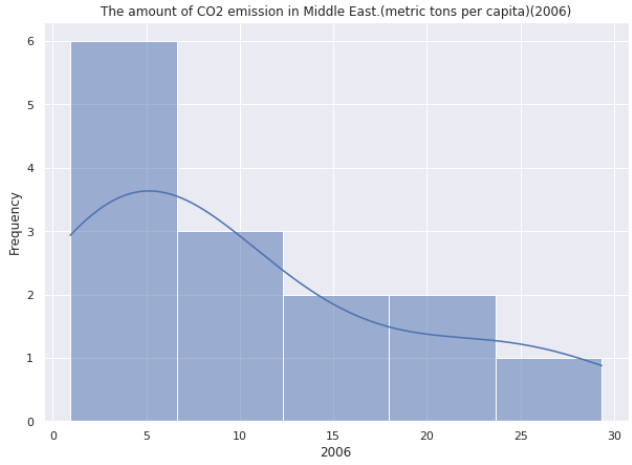


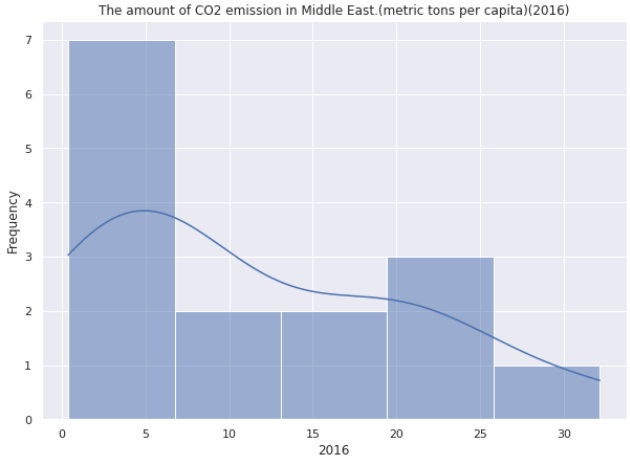


Graph without outliers.

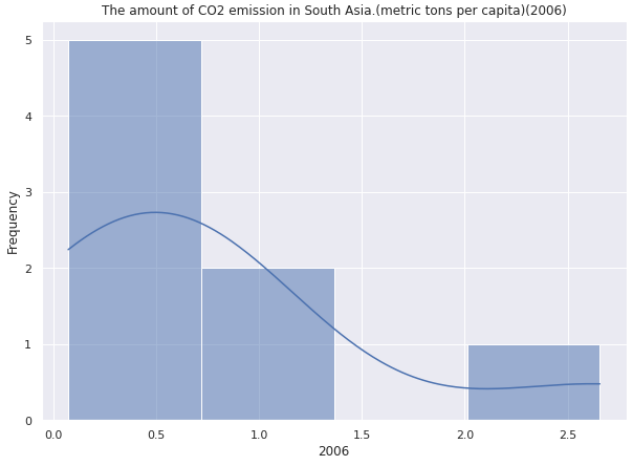


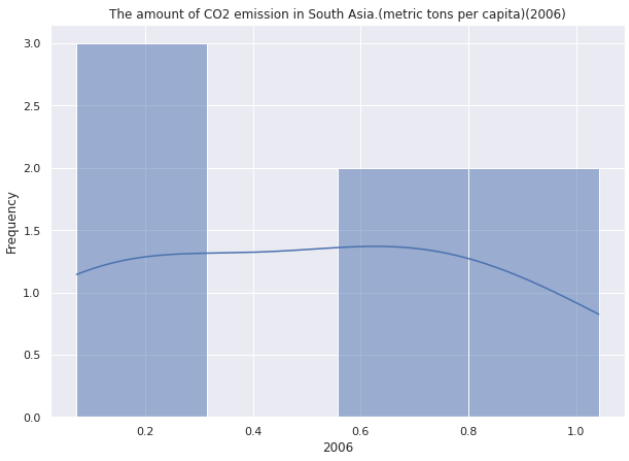


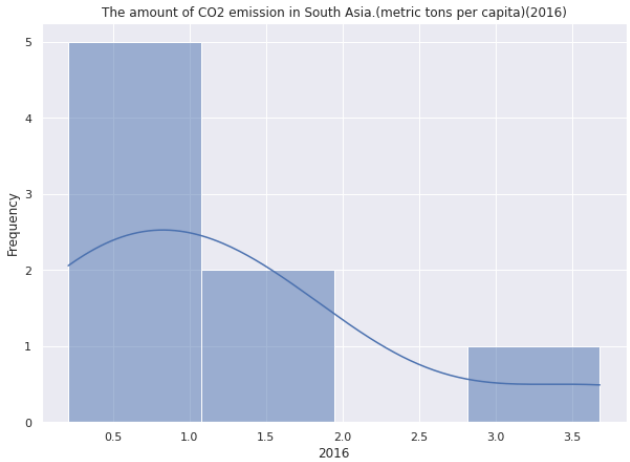


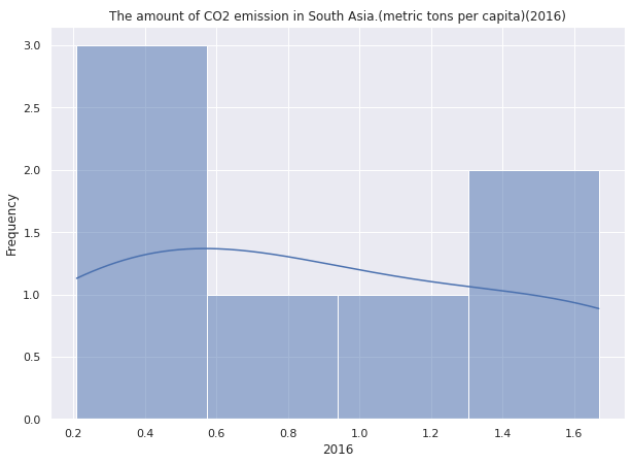

The amount of CO2 emission in Middle East.(metric tons per capita)(2016)
Frequency
2016

The amount of CO2 emission in Middle East.(metric tons per capita)(2016)
Frequency
2016


The amount of CO2 emission in South Asia.(metric tons per capita)(2006)
Frequency
2006


The amount of CO2 emission in South Asia.(metric tons per capita)(2006)
Frequency
2006


The amount of CO2 emission in South Asia.(metric tons per capita)(2016)
Frequency
2016


The amount of CO2 emission in South Asia.(metric tons per capita)(2016)
Frequency
2016

From all graphs, we take a look at the distribution of each region.

- North America : The graphs are right skewed, but close to normal.
- Middle East : The graphs are right skewed.
- South Asia : The graphs are normal.

When we compare 2006 and 2016, the amount tends to be decreasing in North America and the Middle East.But there is a small increase in South Asia.

When we take a look at each region. The Middle East has the most amount of CO2 emission. These results match the result from Part 1.
Next we have North America, and South Asia has the lowest amount of CO2 emission among 3 regions.

2006

With Outliers	Thailand	North America	Middle East	South Asia
Mean	3.313806	4.632674	12.599432	0.771662
Median		2.403914	7.594394	0.619266
SD		5.370264	12.329773	0.834874
IQR		3.448392	15.339192	0.650038

Without Outliers	Thailand	North America	Middle East	South Asia
Mean	3.313806	2.696156	10.357859	0.502389
Median		2.318714	7.275134	0.608454
SD		1.722234	9.085705	0.369378
IQR		2.800822	10.877269	0.541254

2016

With Outliers	Thailand	North America	Middle East	South Asia
Mean	3.780122	4.152318	11.295772	1.242188
Median		2.524296	7.613494	0.960519
SD		4.369211	9.623630	1.130102
IQR		3.006151	14.753565	1.21794

Without Outliers	Thailand	North America	Middle East	South Asia
Mean	3.780122	2.594666	11.295772	0.893887
Median		2.488199	7.613494	0.886749
SD		1.545733	9.623630	0.598099
IQR		1.45337	14.753565	0.936643

From the statistics data table.We can see that the mean decreases in North America, but increases in Middle East and South Asia by small amount.

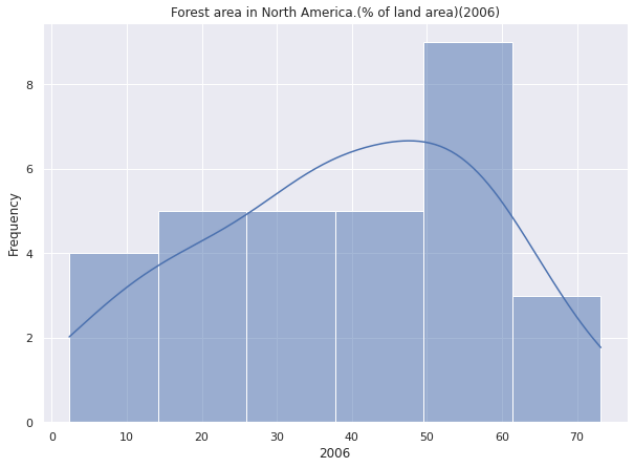
If we look through each region compared to Thailand. We can see that Thailand's average CO2 emission is more than North America and South Asia, but lower compared to the Middle East.

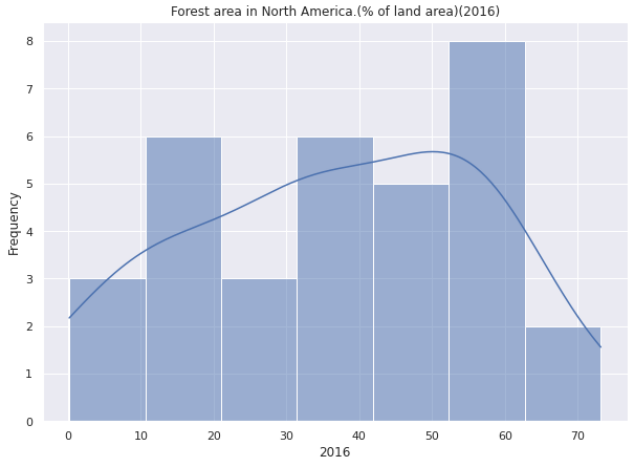
South Asia’s CO2 emission is significantly lower compared to the Middle East, And the Middle East produces the most CO2 emission among three regions.

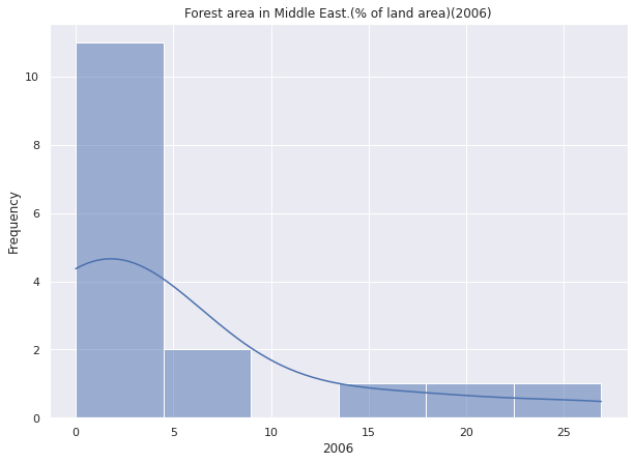
These results match the results from Part 1.

Analysis of forest data.

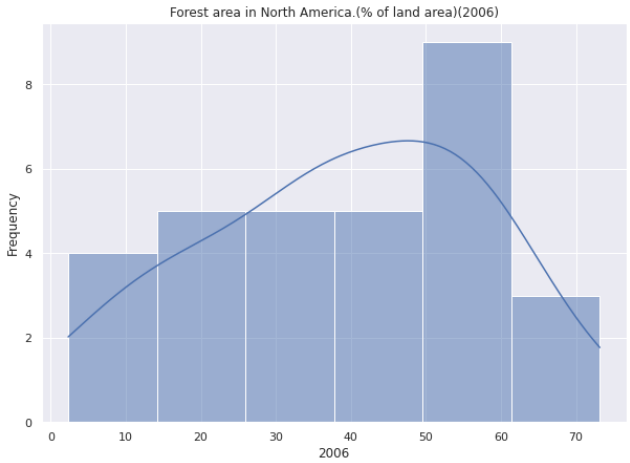
Graph with outliers.

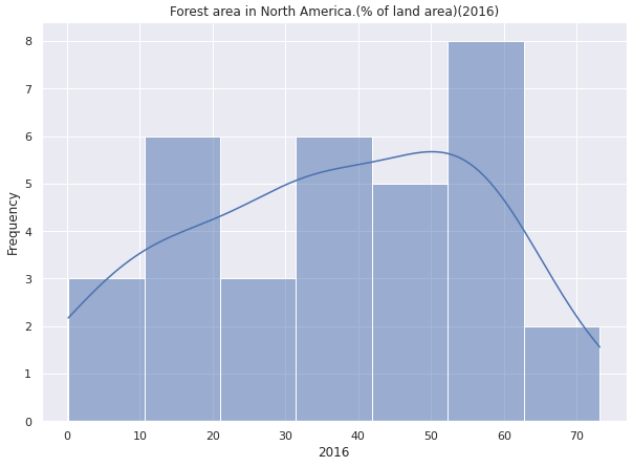


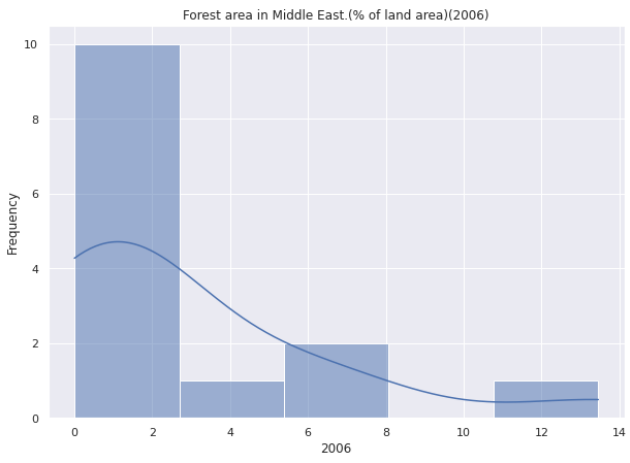


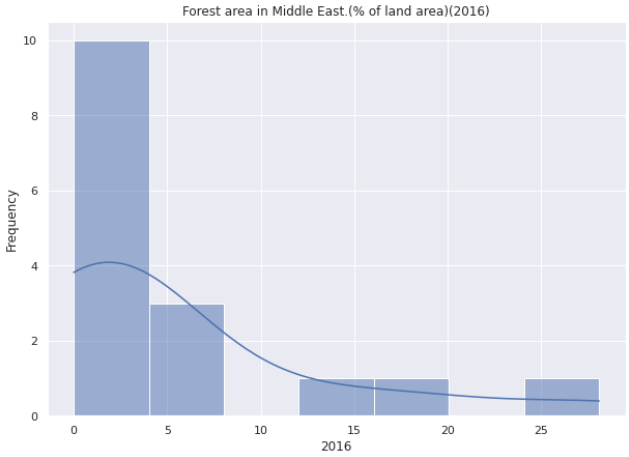


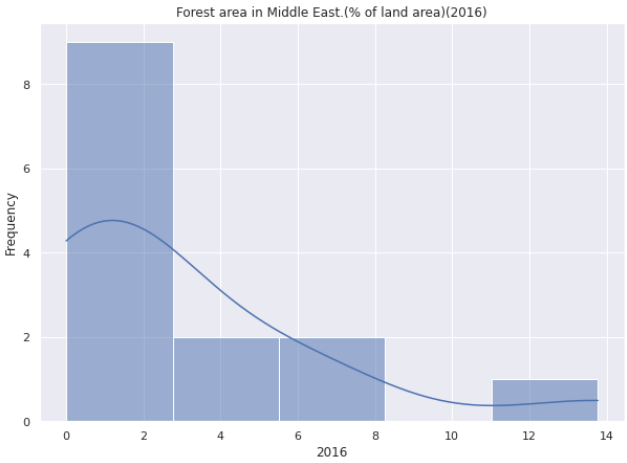
Graph without outliers.

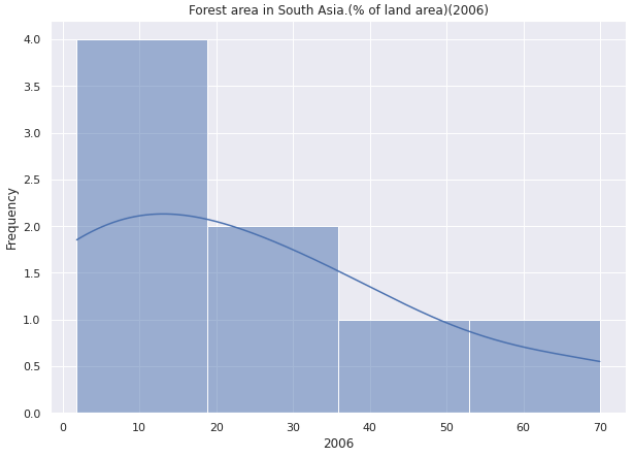


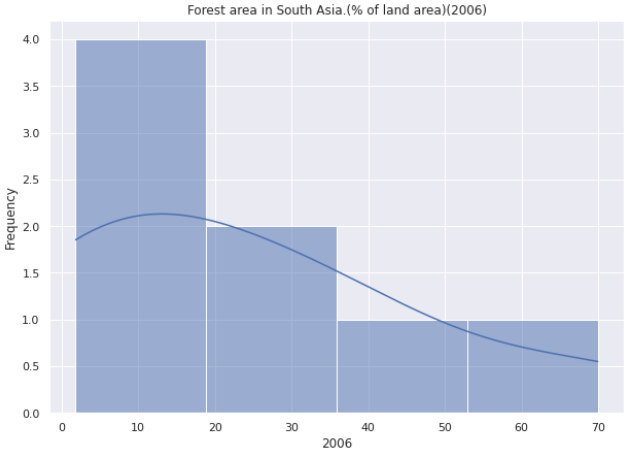


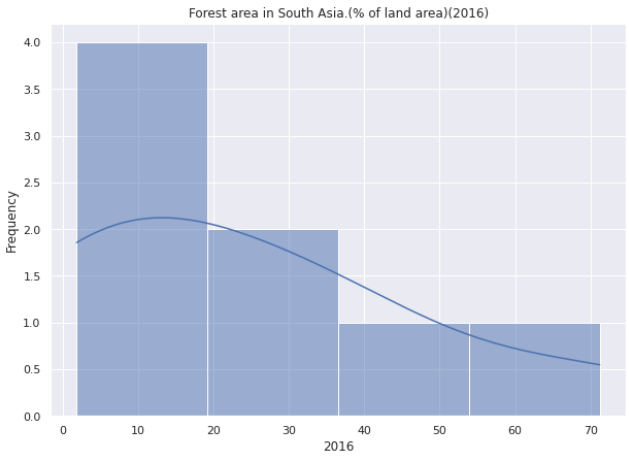


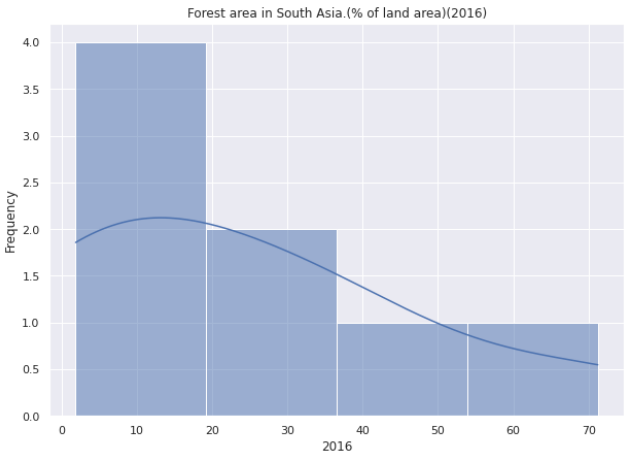

Forest area in Middle East.(% of land area)(2016)
Frequency
2016


Forest area in Middle East.(% of land area)(2016)
Frequency
2016


Forest area in South Asia.(% of land area)(2006)
Frequency
2006


Forest area in South Asia.(% of land area)(2006)
Frequency
2006


Forest area in South Asia.(% of land area)(2016)
Frequency
2016


Forest area in South Asia.(% of land area)(2016)
Frequency
2016

From all graphs, we take a look at the distribution of each region.

- North America : The graphs are normal/
- Middle East : The graphs are right skewed.
- South Asia : The graphs are right skewed.

When we compare 2006 and 2016, the amount tends to be the same. Since the forest takes time to grow.On the other hand we can't see any decrease in the amount of the forest percentage.This is due to the fact that the world takes the forest into account and tries to preserve them as much as possible.

When we take a look at each region. North America has the most percentage of the forest compared to land. And the Middle East has the least percentage. This is because in the Middle East most land are desert.

These results match the results from Part 1.

2006

With Outliers	Thailand	North America	Middle East	South Asia
Mean	38.44859	38.749920	5.391962	24.099558
Median		38.828469	1.701564	18.861479
SD		18.611497	7.797480	23.523741
IQR		27.959498	5.866361	30.901895

Without Outliers	Thailand	North America	Middle East	South Asia
Mean	38.44859	38.749920	2.907931	24.099558
Median		38.828469	1.314098	18.861479
SD		18.611497	3.791154	23.523741
IQR		27.959498	3.458563	30.901895

2016

With Outliers	Thailand	North America	Middle East	South Asia
Mean	39.180646	36.499210	5.515356	24.393998
Median		34.055994	1.792373	19.190298
SD		19.779419	8.019548	24.024655
IQR		34.633215	5.784259	31.708965

Without Outliers	Thailand	North America	Middle East	South Asia
Mean	39.180646	36.499210	2.961440	24.393998
Median		34.055994	1.391303	19.190298
SD		19.779419	3.820713	24.024655
IQR		34.633215	3.521758	31.708965

From the statistics data table.We can see that the mean decreases in North America, but slightly increases in the Middle East ,South Asia and Thailand.

If we look through each region compared to Thailand. We can see that Thailand's average forest area is more than all three regions.

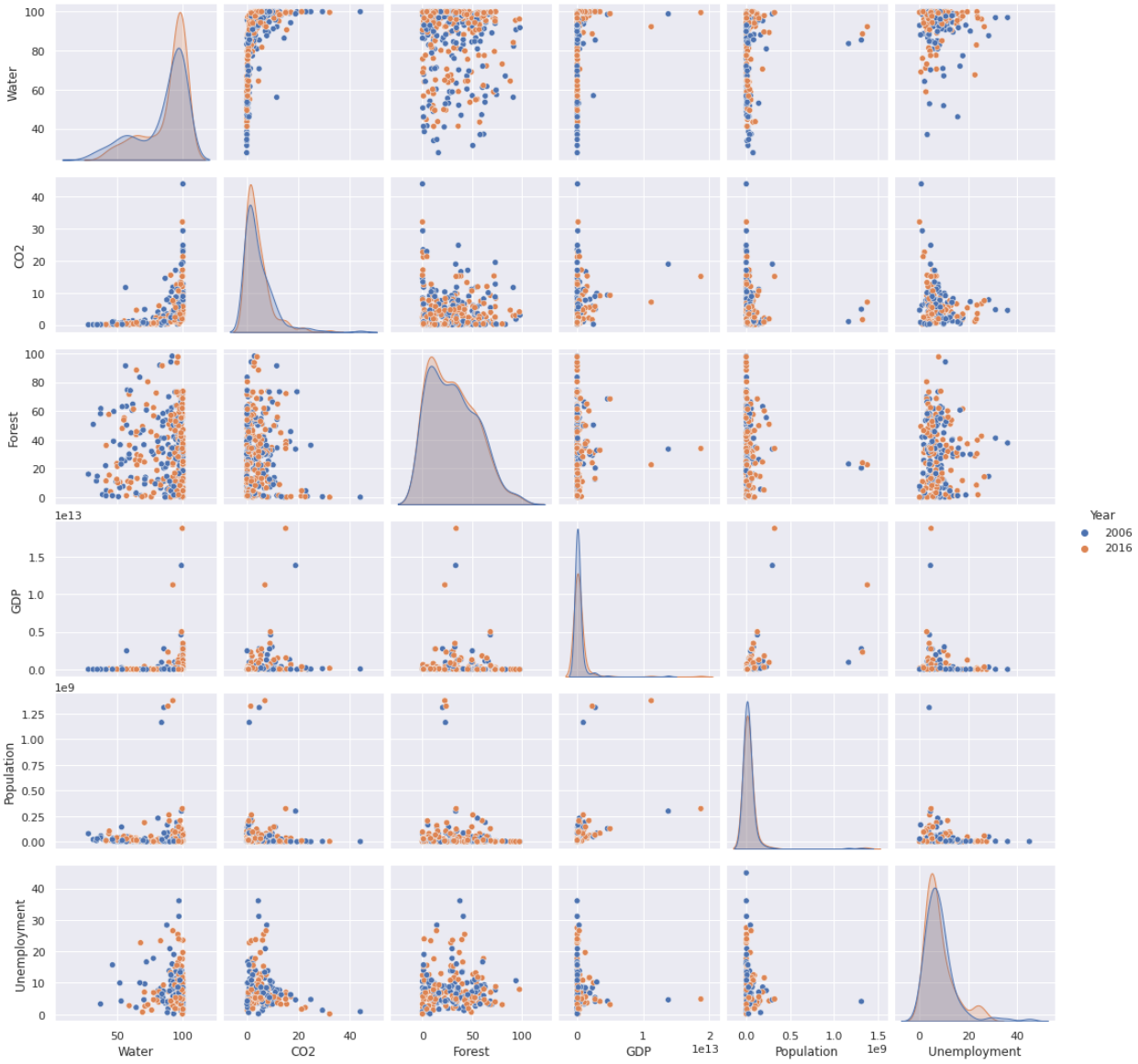
The Middle East has the least amount of forest area, because most of the land is desert.

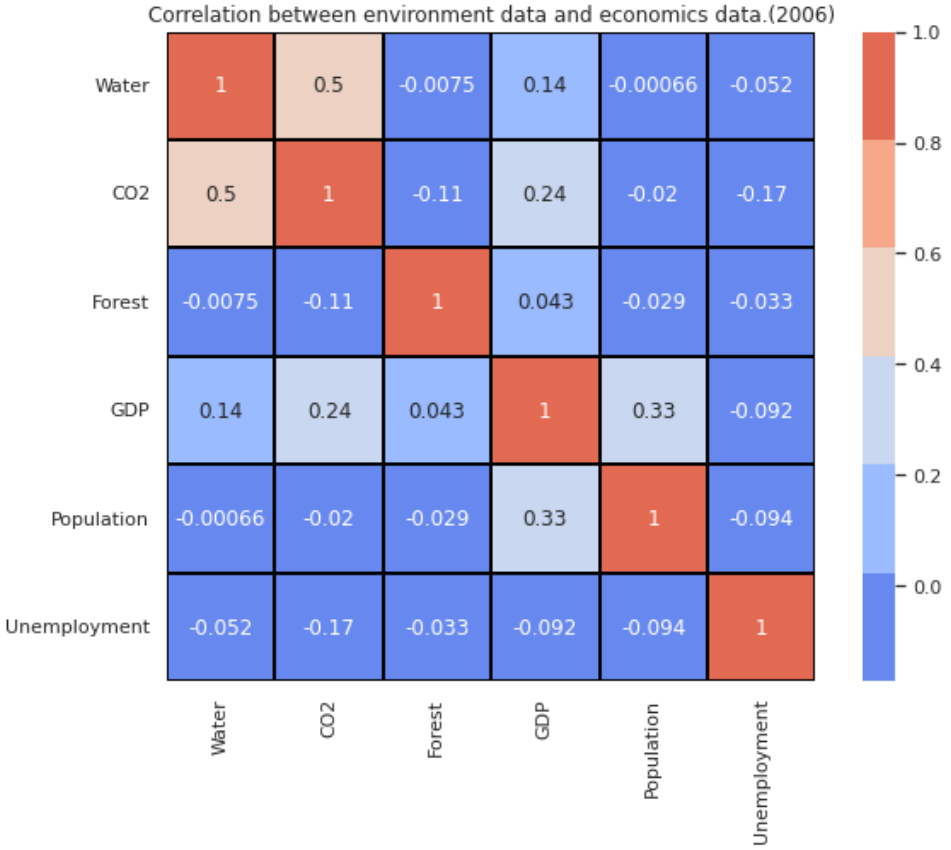
These results match the results from Part 1.

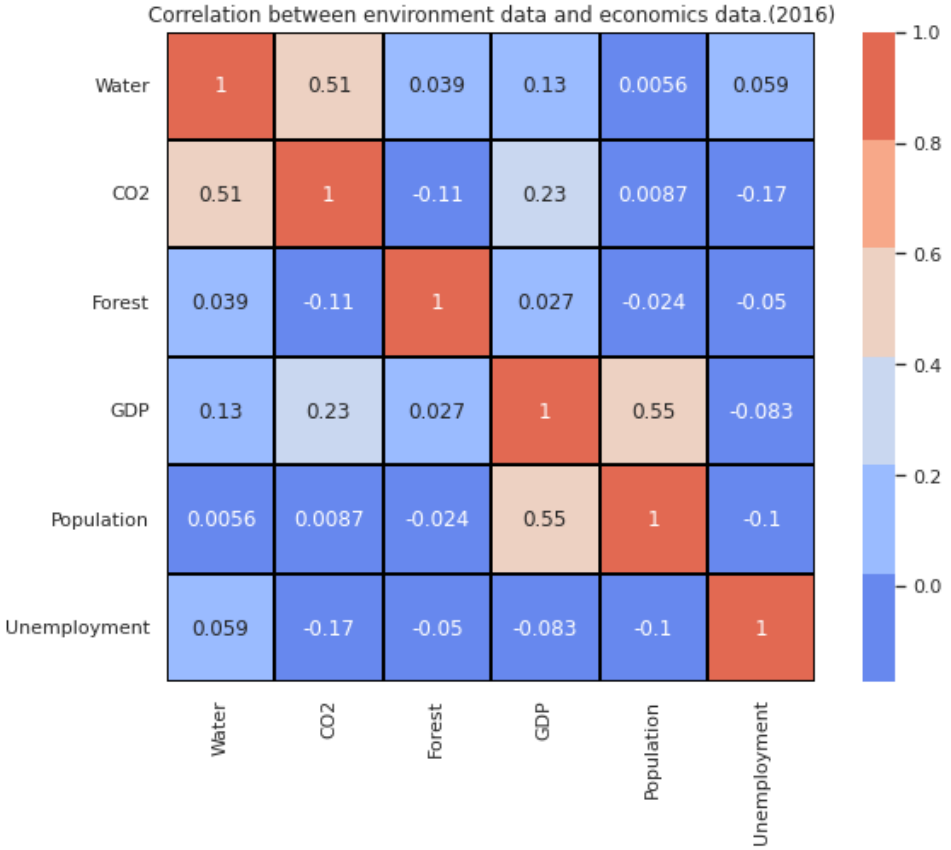
Part 3

Analysis of correlation between environment data and economic data.

Scatter plot of environment and economics data.



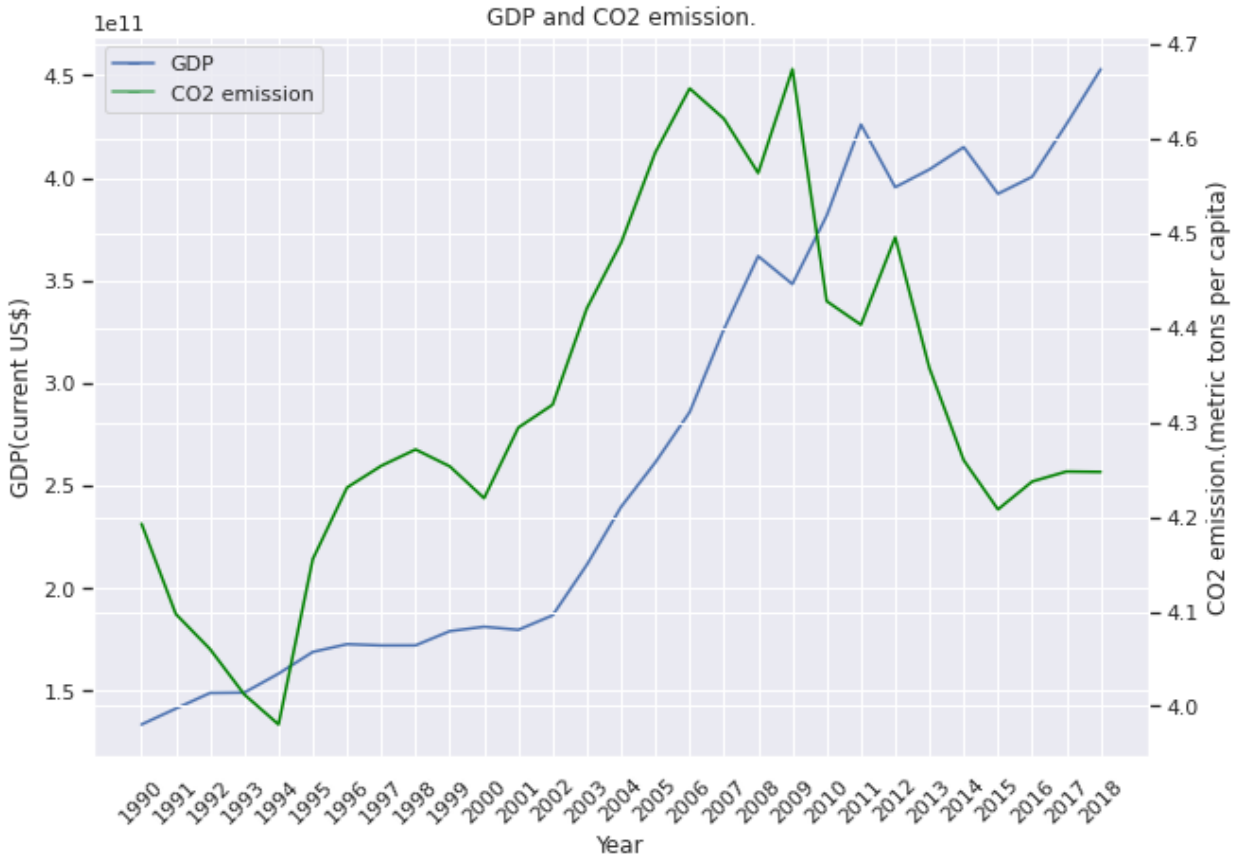

Correlation between environment data and economics data.(2006)
Water
CO2
Forest
GDP
Population
Unemployment
1
0.5
-0.0075
0.14
-0.00066
-0.052
0.5
1
-0.11
0.24
-0.02
-0.17
-0.0075
-0.11
1
0.043
-0.029
-0.033
0.14
0.24
0.043
1
0.33
-0.092
-0.00066
-0.02
-0.029
0.33
1
-0.094
-0.052
-0.17
-0.033
-0.092
-0.094
1
1.0
0.8
0.6
0.4
0.2
0.0


Correlation between environment data and economics data.(2016)
Water
CO2
Forest
GDP
Population
Unemployment
1
0.51
0.039
0.13
0.0056
0.059
0.51
1
-0.11
0.23
0.0087
-0.17
0.039
-0.11
1
0.027
-0.024
-0.05
0.13
0.23
0.027
1
0.55
-0.083
0.0056
0.0087
-0.024
0.55
1
-0.1
0.059
-0.17
-0.05
-0.083
-0.1
1
1.0
0.8
0.6
0.4
0.2
0.0

As we can see from the scatter plot and heatmap of the data,the only data that has enough correlation coefficient to determine the correlation between environment data and economics data is CO_2 and GDP with 0.23 correlation coefficient.

Part 4

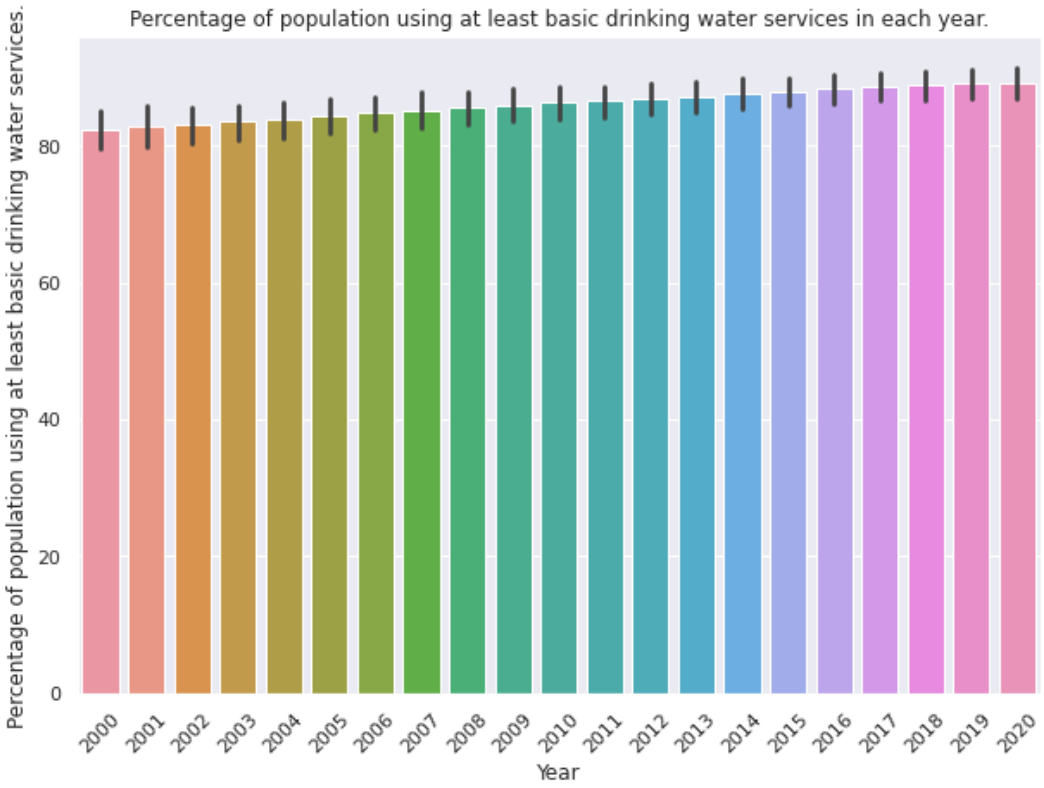
CO2 and GDP.(Highest correlation coefficient between two datasets)

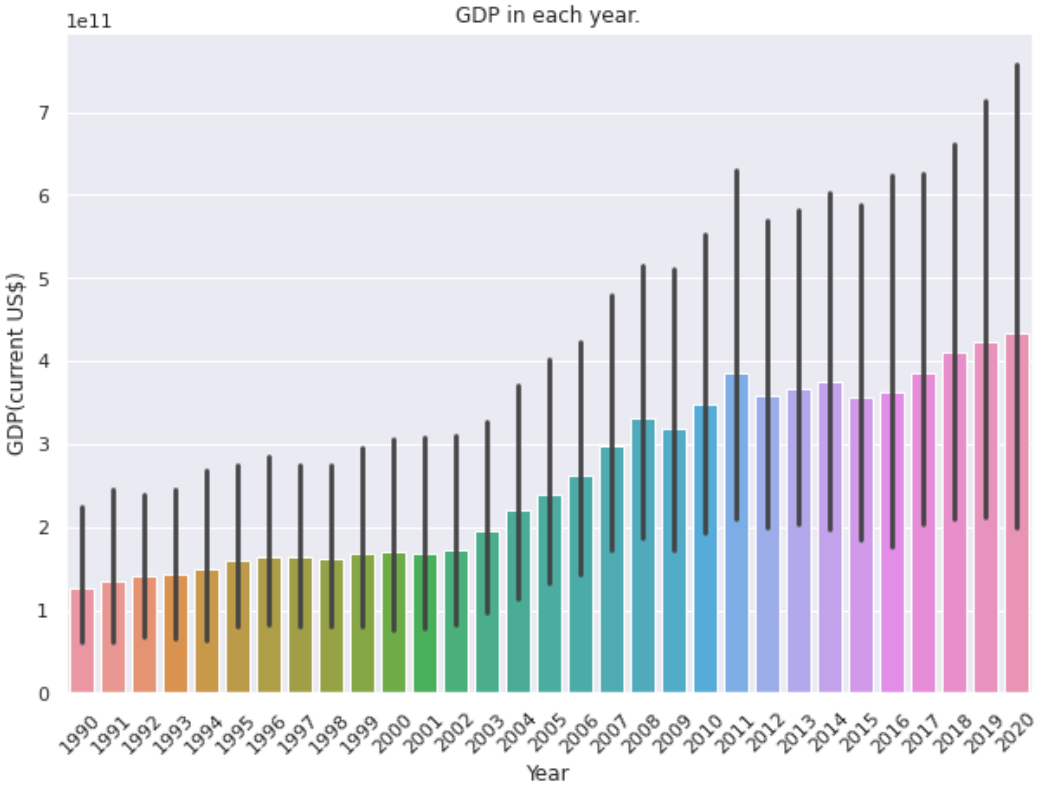


As we can see, before 2009 CO2 and GDP had the same trend.But after that CO2 emission dropped in serious amounts.The reason behind this is the Copenhagen Accord where countries representing over 80% of global emissions have engaged.Other reasons is the change in industrial knowledge and the development of clean energy.

So we can't actually conclude that CO2 and GDP have low correlation, but we need to look for a specific time and other factors to determine the result.

Water and GDP.(No correlation)





Since both of them have no correlation.We will try to see why.

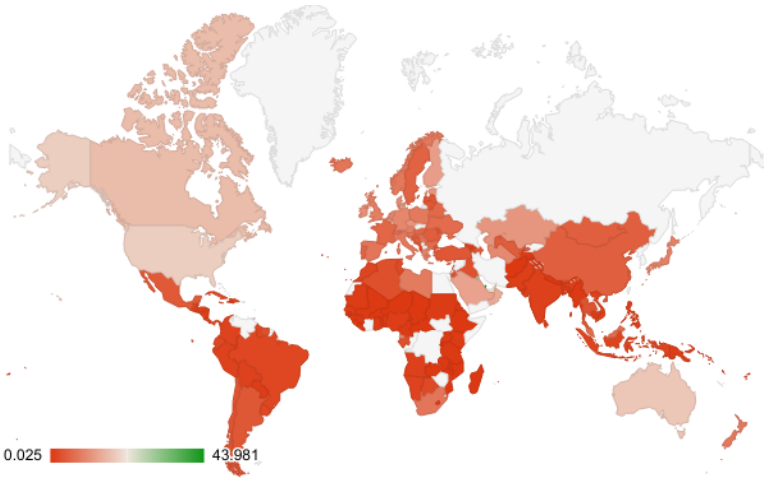
If we look at the water data. We can see that it is quite high since 2000. This is because drinking water is the most basic utility that humans need. And It is increasing because we need to provide every human basic needs.

On the other hand,GDP data.We can see that it’s rising and falling from time to time.But overall it increases.This is because of some financial crisis makes the gdp fall.But the gdp and economics tend to grow overtime.

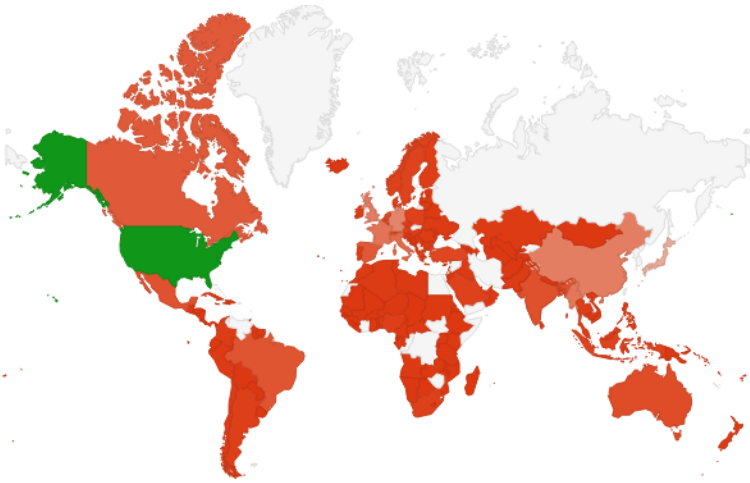
The reason why both of them have no correlation is because even though the economics of the country is falling, every human still needs a good quality water to drink.And it’s a government duty to provide every person in their country with this basic need.

CO2 emission,GDP in each country.(Geovisualization)

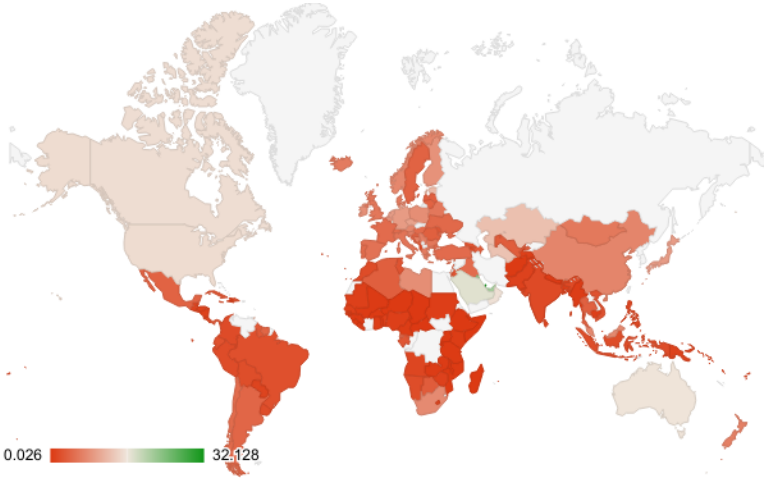
CO2 2006



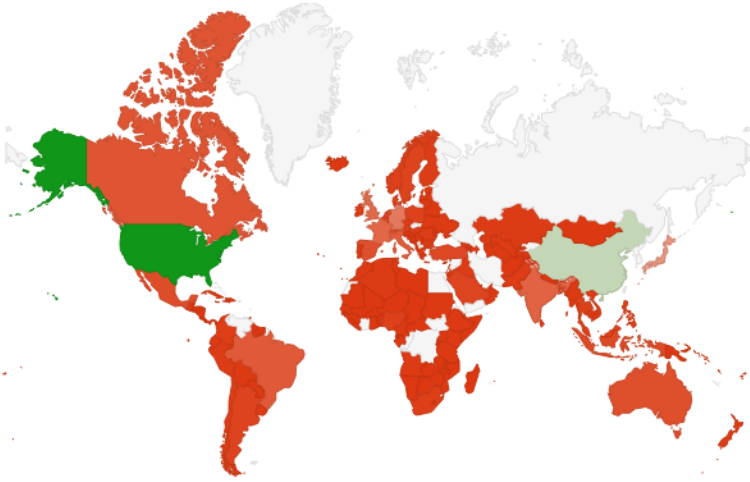
GDP 2006



CO2 2016



GDP 2016



From geovisualization comparing 2006 and 2016. The amount of CO2 emission slightly decreases. The amount of GDP tends to increase. And GDP in China is increasing in significant amounts.