Title: *Annual Energy Outlook*

Course Title: **Research Methods**
Course Number: **ETM 565/665**
Instructor: **Dr. Timothy R. Anderson**
Term: **Spring**
Year: **2015**
Author(s): **Pooya Soltantabar**

ETM OFFICE USE ONLY
Report No.:
Type: Student Project
Note:

"The *Annual Energy Outlook* presents a projection and analysis of US energy supply, demand, and prices through 2035. The projections are based on results from the Energy Information Administration's National Energy Modeling System".[1] This report contains a dataset with various information about production, import and export of different types of energy such as oil, gas, nuclear, hydro, etc. In this project I will do some statistical analysis on this dataset such as regression and correlations, clustering, etc. Also I will draw some graphs on the data to visualize the data and their interrelations.

The data has two parts, first is actual data from 2007-2014 and the second part is prediction for 2015-2035; although the prediction part is not real data, the aim of this project is to use research method tools, especially R studio, to practice some statistical techniques. Table 1 is the overview of my dataset.

Supply, Disp	Oil Produc	Natural Ga	Dry Natur	Coal Prod	Nuclear Pc	Hydropow	Biomass P	Other Ren	Total Prod	Oil Import	Liquid Fue	Natural Ga	Total Imp	Petroleum	Natural Gas Exports	Coal Expo	Total Expo
2007	10.75	2.41	19.62	23.49	8.46	2.45	3.15	0.99	72.14	21.91	6.98	4.72	34.6	2.83	0.83	1.51	5.17
2008	10.51	2.57	21.14	23.86	8.46	2.46	3.97	1.17	74.23	21.39	6.38	4.06	32.79	3.71	1.01	2.07	6.8
2009	11.3	2.47	21.18	22.02	8.49	2.57	3.55	1.43	72.99	20.05	5.61	3.87	30.09	3.92	1.03	1.49	6.43
2010	11.39	2.4	20.56	21.54	8.52	2.68	3.88	1.86	73.38	19.41	5.28	3.93	29.41	3.31	1.02	1.72	6.05
2011	11.67	2.32	19.9	22.62	8.54	2.76	3.95	2.21	74.5	19.79	5.61	3.69	29.9	3.45	0.86	1.71	6.02
2012	11.8	2.28	19.59	23.18	8.56	2.81	4.2	2.64	75.67	19.61	5.76	3.53	29.73	3.47	0.91	1.77	6.15
2013	12.3	2.21	19.13	23.49	8.59	2.88	4.4	2.99	76.89	18.9	5.65	3.32	28.72	3.54	0.97	1.7	6.21
2014	12.44	2.21	19.15	23.78	8.7	2.92	4.53	3.03	77.76	18.52	5.46	3.37	28.19	3.55	1.04	1.59	6.19
2015	12.56	2.22	19.39	23.61	8.75	2.96	4.64	3.03	78.36	18.25	5.29	3.46	27.79	3.58	1.12	1.49	6.18
2016	12.58	2.22	19.45	23.75	8.8	2.96	4.96	3.03	78.94	17.98	5.26	3.61	27.63	3.59	1.17	1.44	6.2
2017	12.9	2.23	19.71	23.82	8.91	2.96	5.22	3.03	80.02	17.36	5.16	3.72	27.01	3.61	1.22	1.42	6.25
2018	13.34	2.25	20.18	24.08	9.03	2.96	5.47	3.03	81.67	16.42	5.09	3.71	25.99	3.6	1.28	1.41	6.3
2019	13.77	2.27	20.45	24.35	9.15	2.96	5.76	3.03	83.1	15.7	5.03	3.7	25.27	3.6	1.33	1.39	6.33
2020	14.14	2.24	21.01	24.3	9.26	2.96	6.05	3.04	84.36	15.28	4.94	3.53	24.93	3.64	1.39	1.41	6.45
2021	14.43	2.21	21.82	24.37	9.29	2.96	6.41	3.04	85.94	14.7	4.89	3.38	24.24	3.67	1.44	1.43	6.54
2022	14.65	2.23	22.11	24.97	9.29	2.96	6.76	3.05	87.43	14.23	4.87	3.31	23.62	3.7	1.49	1.4	6.6
2023	14.6	2.24	22.37	25.16	9.29	2.96	7.32	3.07	88.4	13.99	4.85	3.26	23.43	3.72	1.54	1.31	6.57
2024	14.63	2.25	22.76	25.56	9.29	2.96	8.05	3.08	89.98	13.52	4.8	3.25	22.93	3.68	1.59	1.35	6.62
2025	14.67	2.26	22.96	25.74	9.29	2.96	8.68	3.1	91.06	13.21	4.78	3.24	22.58	3.71	1.64	1.19	6.54
2026	14.68	2.27	23.16	26.02	9.29	2.96	9.24	3.1	92.05	12.99	4.74	3.24	22.41	3.72	1.67	1.12	6.51
2027	14.96	2.27	23.52	25.82	9.29	2.96	10.08	3.12	93.36	12.38	4.7	3.22	21.92	3.71	1.7	1.05	6.47
2028	15.2	2.3	23.79	26.11	9.29	2.96	10.49	3.13	94.61	11.89	4.78	3.16	21.5	3.69	1.73	1	6.43
2029	15.1	2.31	24	26.46	9.29	2.96	10.85	3.16	95.44	11.87	4.78	3.13	21.4	3.72	1.76	1	6.48
2030	14.88	2.31	24.49	26.78	9.29	2.97	11.16	3.19	96.33	11.86	4.8	3.12	21.27	3.73	1.79	0.94	6.45
2031	14.82	2.32	24.63	27.05	9.29	2.97	11.52	3.21	96.99	11.89	4.83	3.1	21.27	3.76	1.8	0.92	6.49
2032	14.94	2.33	24.85	27.21	9.29	2.97	11.71	3.24	97.7	11.75	4.84	3.04	21.16	3.77	1.82	0.91	6.5
2033	14.75	2.34	25.32	27.21	9.33	2.99	11.79	3.31	98.17	11.79	4.87	2.97	21.38	3.8	1.83	0.88	6.5
2034	14.65	2.34	25.45	27.33	9.39	2.99	11.94	3.36	98.55	12.01	4.92	2.92	21.64	3.84	1.84	0.84	6.51
2035	14.83	2.35	25.61	27.57	9.44	3.01	12.08	3.4	99.36	11.95	4.96	2.84	21.54	3.86	1.84	0.83	6.53

Table 1 – The dataset

First of all, I am going to write some hypotheses and try to evaluate them with statistical techniques.

- ✓ **Hypothesis 1:** Regarding contemporary concerns, I guessed that the amount of nonrenewable energy production, or to be more accurate exploitation, will reduce in the future.

- ✓ **Hypothesis 2:** The amount of renewable (clean) energy will increase in the future.

- ✓ **Hypothesis 3:** There is a question that if the country is going to increase its gas production, will this surplus export to other countries or is it going to be used inside the country?

- ✓ **Hypothesis 4:** Is there any relationship between increasing coal production and its export amount?

- ✓ **Hypothesis 5:** Is there any relationship between total energy import and producing nonrenewal types of energy?

- ✓ Hypothesis 1: Regarding contemporary concerns, I guessed that the amount of nonrenewable energy production, or to be more accurate exploitation, will reduce in the future.
- ❖

I have plotted the production trend of three kinds of nonrenewable energy: Oil, Gas and Coal. Figure 1 demonstrates this trend.

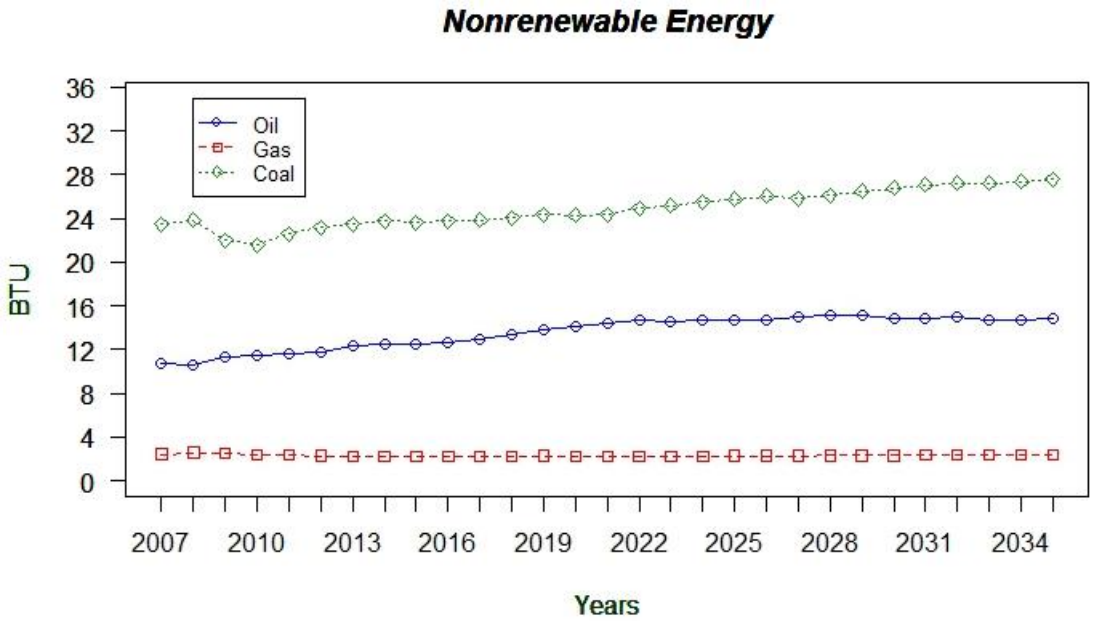

Figure 1- trend of exploitation nonrenewable energy

As the plot represents, the trend seems contradictory to my hypothesis. Surprisingly, it shows that the coal resources would be the main part of nonrenewable energy resource!

- ❖ Hypothesis 2: The amount of renewable (clean) energy will increase in the future.

I was worried about the environment and I decided to plot the clean energies such as biomass production and hydropower production to find out what is the share of these kind of energy production resources in future years. The two scatter plots in figure 2 and figure 3 are the trend of production of these two kind of energy resources.

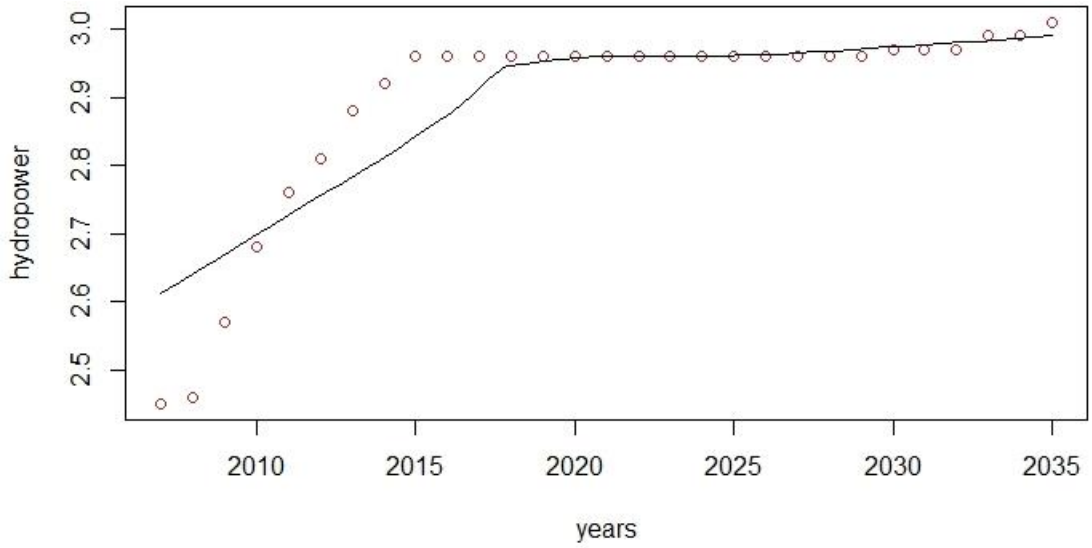

Figure 2- trend of production of hydroelectric power

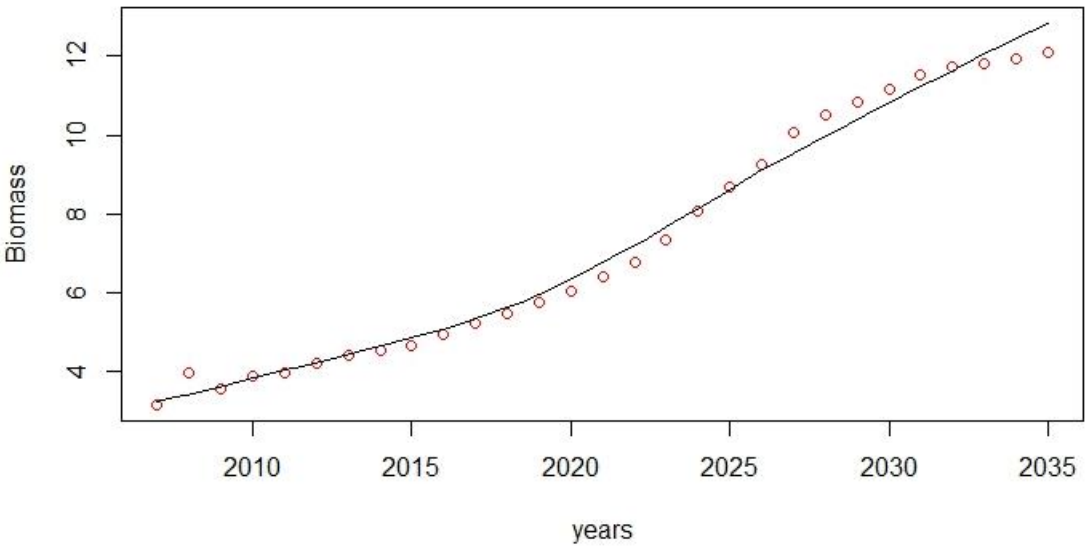

Figure 3- trend of production of biomass energy

Fortunately, these two plots show that the clean energy suppliers are going to increase their products. It is clear that the biomass section curve rises more steeply, while on the other hand the slope of hydropower is milder. This is understandable because of lack of water resources and other environmental concerns which I am going to discuss a little bit more. These hydropower energy concerns are categorized to three different parts:[2]

- Land of Use
- Wildlife Impacts
- Life-cycle Global Warming Emissions

According to the land of use, there is always a risk of flood for the area on which the damn has been built. Flooding land for a hydroelectric reservoir has an extreme environmental impact: it destroys forest, wildlife habitat, agricultural land, and scenic lands. In many instances, such as the Three Gorges Dam in China, entire communities have also had to be relocated to make way for reservoirs.[3]

The second impact would be the wildlife one; hydroelectric power has an impact on aquatic ecosystems. For instance, turbine blades can kill fishes; or because of the stagnation of the reservoir, the amount of sediments and nutrition of reservoir is higher than normal which this can cause unbalanced crowd of aquatic ecosystems. [4]

Global warming emissions are produced during the installation and dismantling of hydroelectric power plants, but recent research suggests that emissions during a facility's operation can also be significant. Such emissions vary greatly depending on the size of the reservoir and the nature of the land that was flooded by the reservoir.[5]

- ✓ Hypothesis 3: There is a question that if the country is going to increase its gas production, will this surplus export to other countries or is it going to be used inside the country?
- ❖

To answer this question I tested the correlation and regression between exporting the gas and consuming it inside the country. We can see the results on figures 4 to 8.

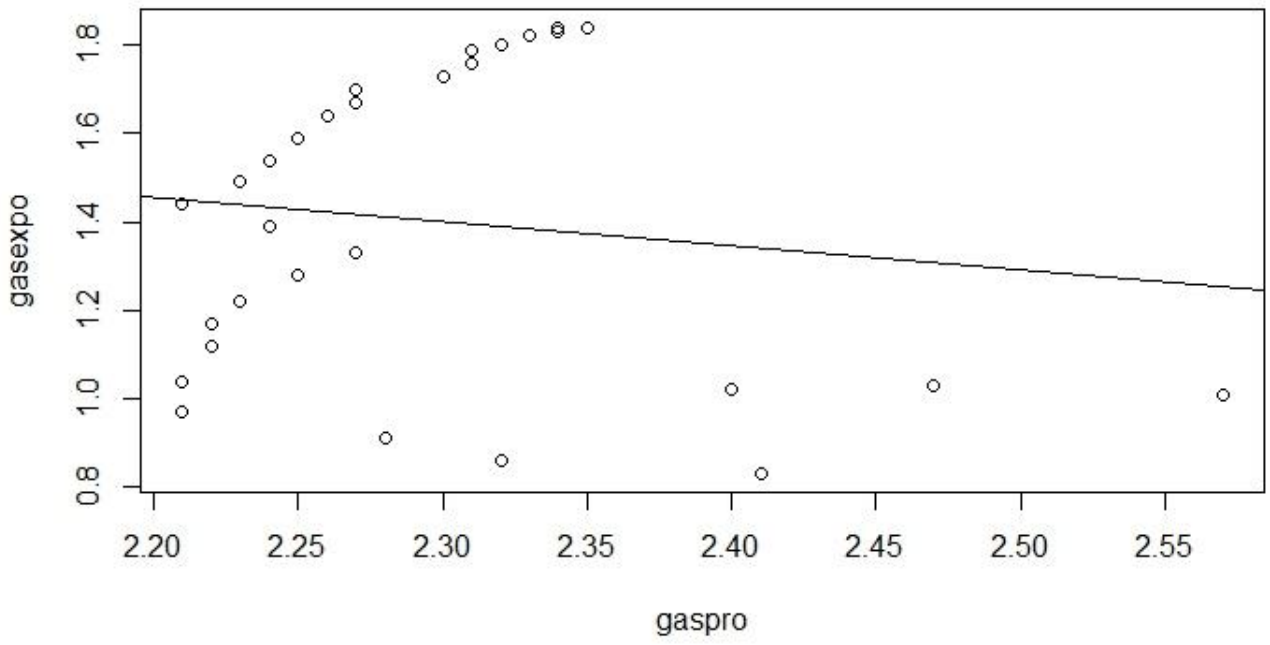

Figure 4- Regression of import and export of gas

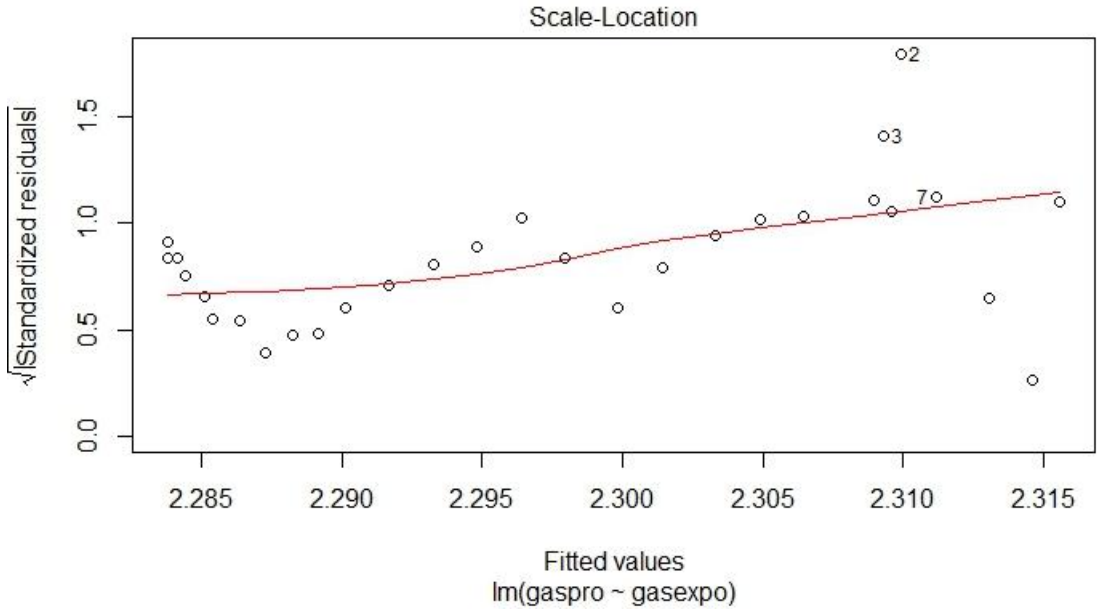

Figure 5- Regression of import and export of gas

This is the summary of the regression:

Residual standard error: 0.08428 on 27 degrees of freedom
Multiple R-squared: 0.01712, Adjusted R-squared: -0.01928
F-statistic: 0.4704 on 1 and 27 DF, p-value: 0.4986
Residual standard error: 0.08428 on 27 degrees of freedom
Multiple R-squared: 0.01712,
Adjusted R-squared: -0.01928
F-statistic: 0.4704 on 1 and 27 DF,
P-value: 0.4986

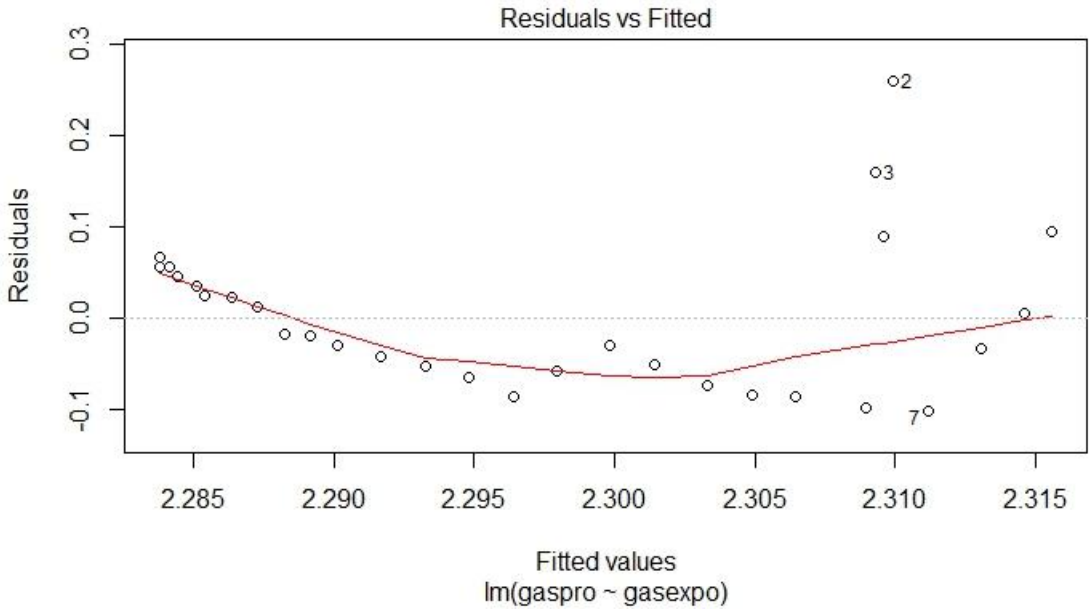

Figure 6- Regression of import and export of gas

At first glance in figure 4, it can be assumed that there is a relationship between producing and exporting the gas; but, when we refer to the p-value of the regression, we can say that it is not reliable and we cannot conclude that more production will lead to more export of this product or vice versa. This model shows that just 1% of observations can be defined by the predicted model! On the other hand, the adjusted R-squared is negative which is strange and odd result in this model. Furthermore, P-value with 0.49

amount reflects the unreliability of the regression. The standardize residuals can be seen in figure 6.

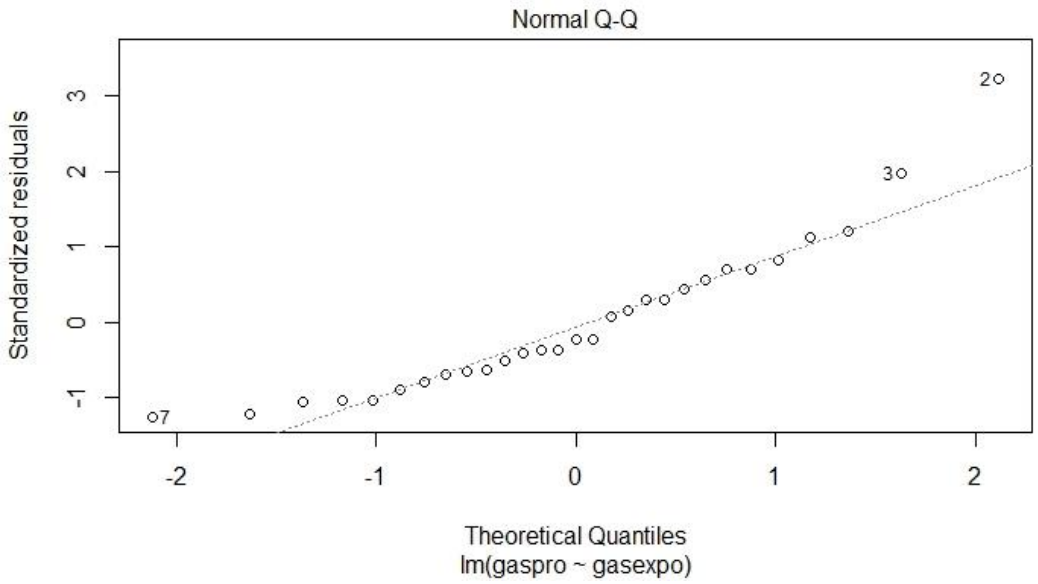

Figure 7- Regression of import and export of gas

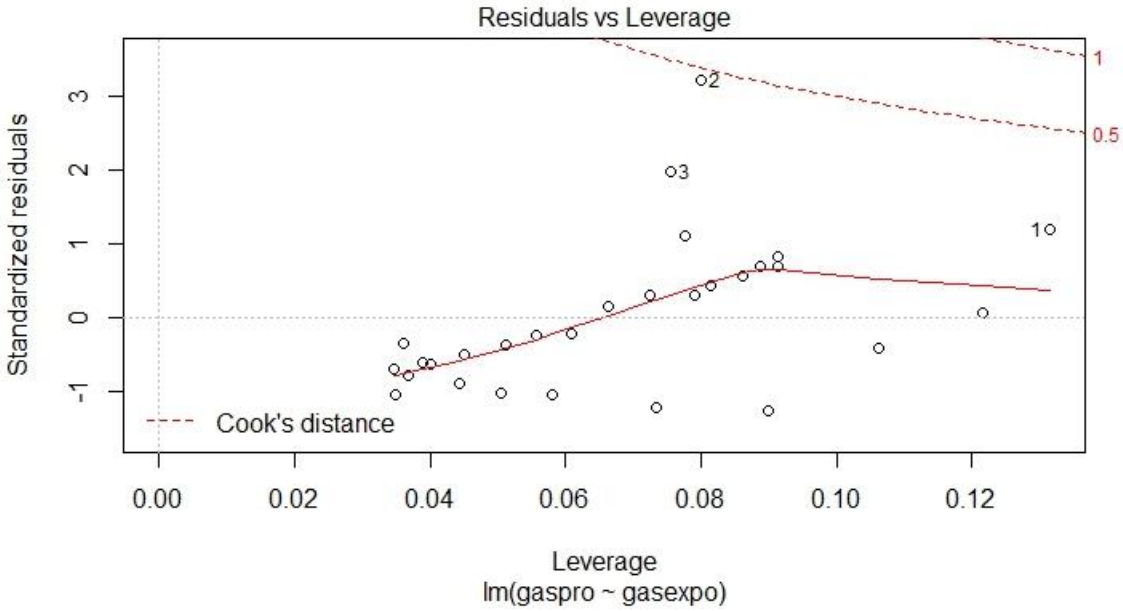

Figure 8- Regression of import and export of gas

- ✓ <u>Hypothesis 4</u>: Is there any relationship between increasing coal production and its export amount?

❖

I did the hypothesis 3 test again; this time for coal production and its export amount. Here is the result of regression:

Residual standard error: 0.7764 on 27 degrees of freedom
Multiple R-squared: 0.7937,
Adjusted R-squared: 0.786
F-statistic: 103.9 on 1 and 27 DF,
P-value: 9.443e-11

This time the P-value shows that the regression is reliable and we can conclude that there is a relationship between exporting and producing coal. In addition, R-squared value indicates that 79% of the observations can be defined by the model. Figures 9 to 13 show the visual results of this regression.

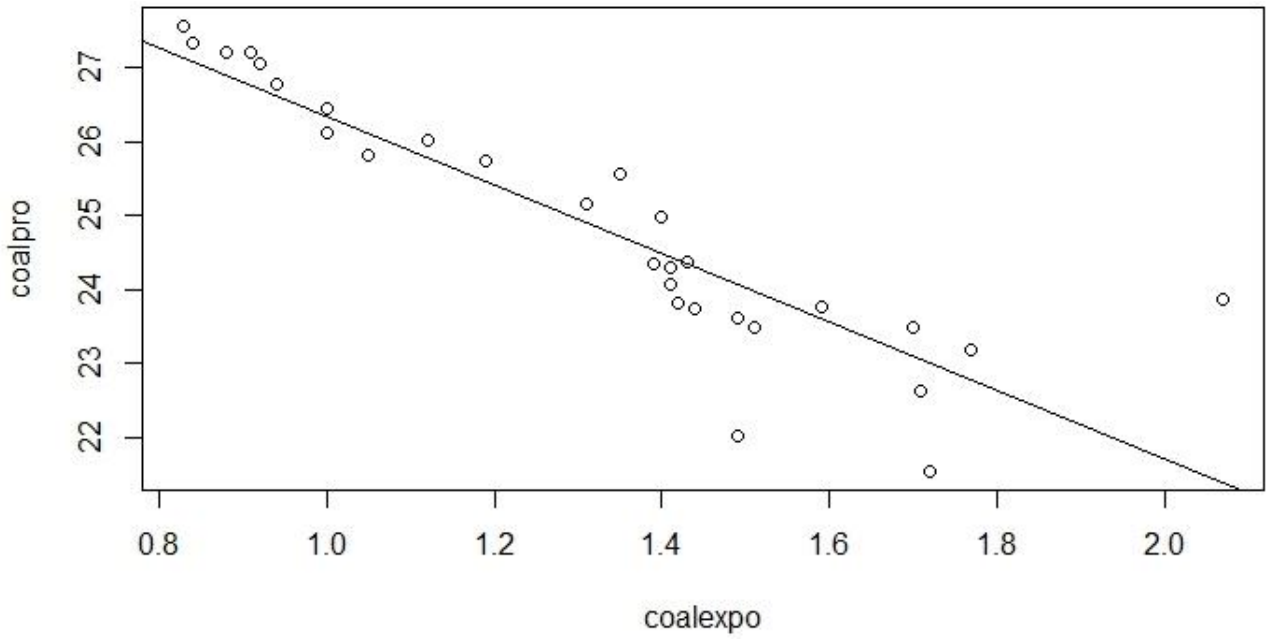

Figure 9- Regression of import and export of coal

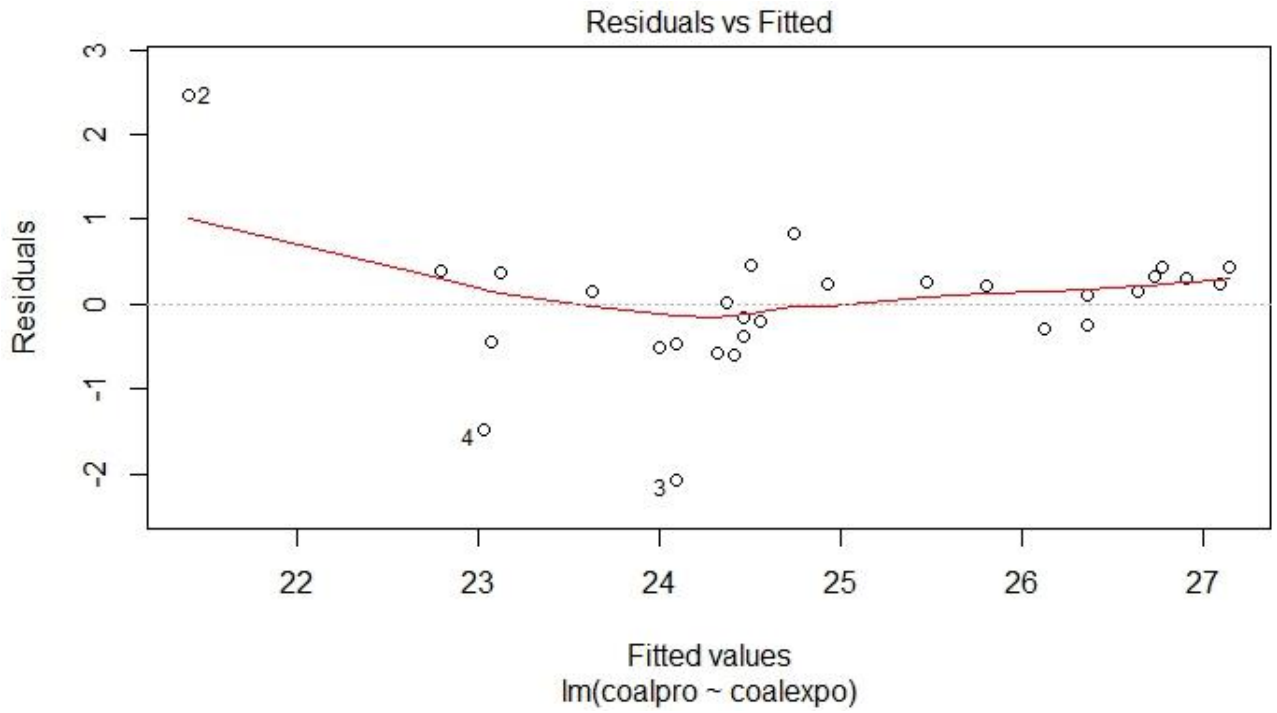

Figure 10- Regression of import and export of coal

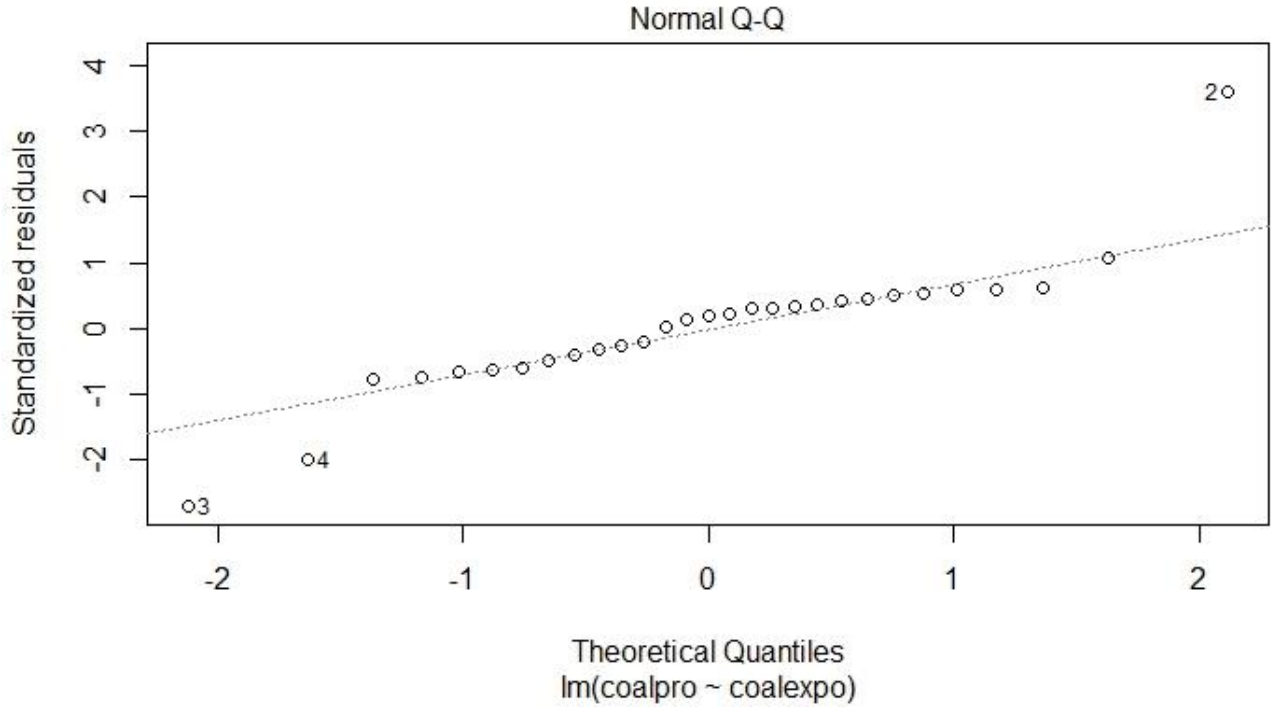

Figure 11- Regression of import and export of coal

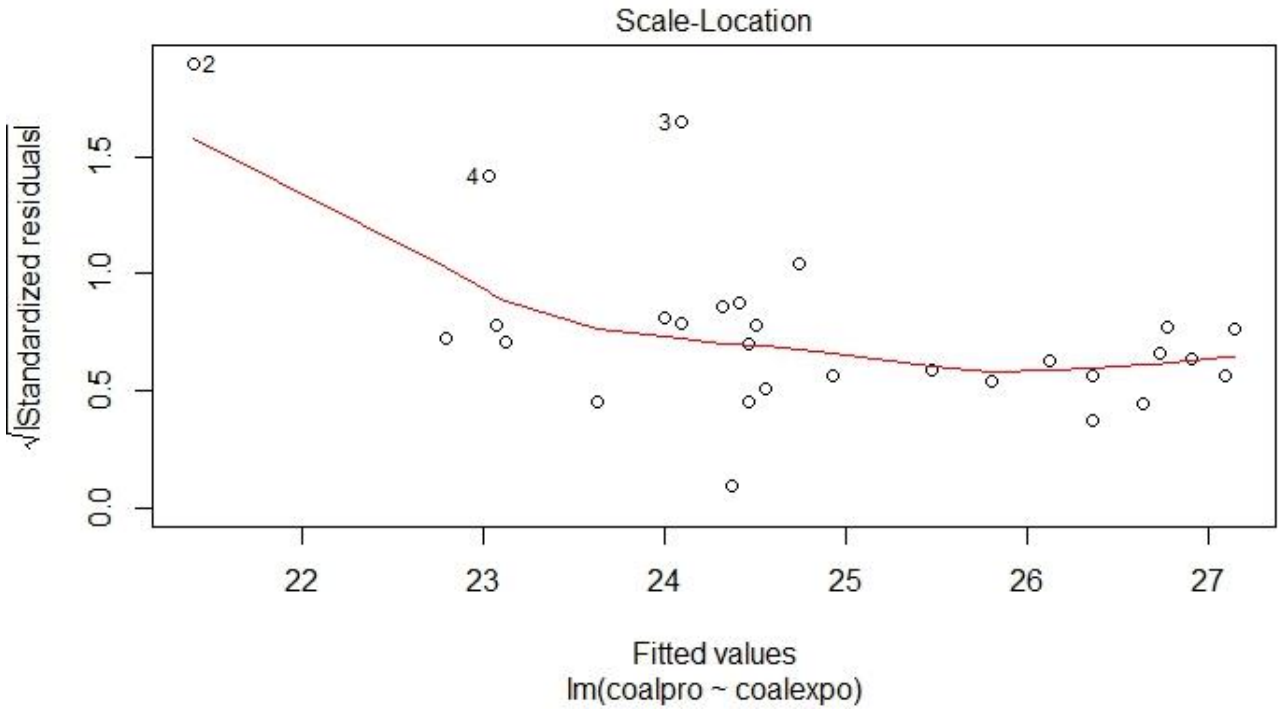

Figure 12- Regression of import and export of coal

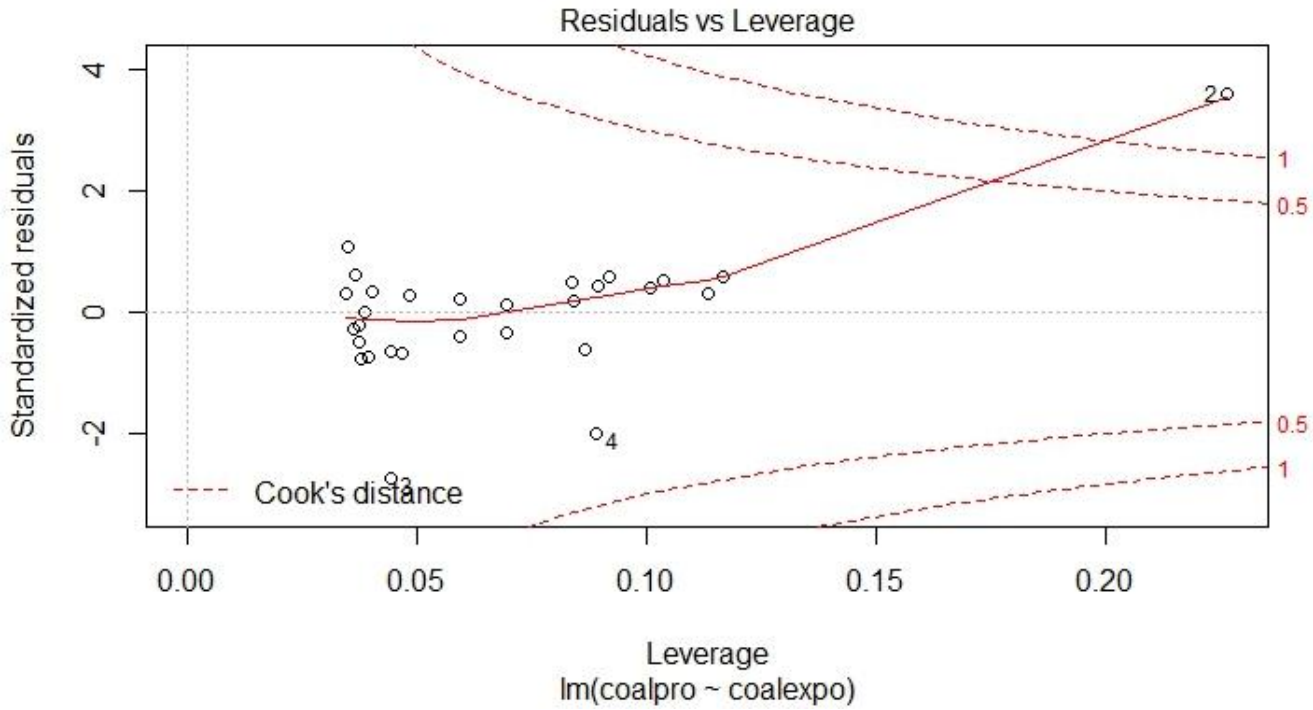

Figure 13- Regression of import and export of coal

- Hypothesis 5: Is there any relationship between total energy import and producing nonrenewal types of energy?

Regarding emission concerns, I want to test the relationship between total amount of nonrenewal energy import and their production. I will use multi-regression between total energy imports and gas, oil and coal production. Here is the summary of the regression:

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 69.79183    5.13160  13.600 4.69e-13 ***
gaspro      -3.50222    2.28917  -1.530    0.139
coalpro     -0.06541    0.19637  -0.333    0.742
oilpro      -2.56610    0.24311 -10.555 1.07e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7343 on 25 degrees of freedom
Multiple R-squared:  0.968,       Adjusted R-squared:  0.9641
F-statistic: 251.8 on 3 and 25 DF,
P-value: < 2.2e-16
```

The negative coefficients of the three different dirty energies show that there is reverse relationship between import and production of these kinds of energy. The big multiple R-squared shows that the line is fitted and about 97% of variations can be explained by the model.The residual standard error confirms that the distance between observed and predicted values is less that 0.73 and this is acceptable. In addition the very small P-value demonstrates that the interpretation is reliable. Figure 14 shows the output of this multi-regression.

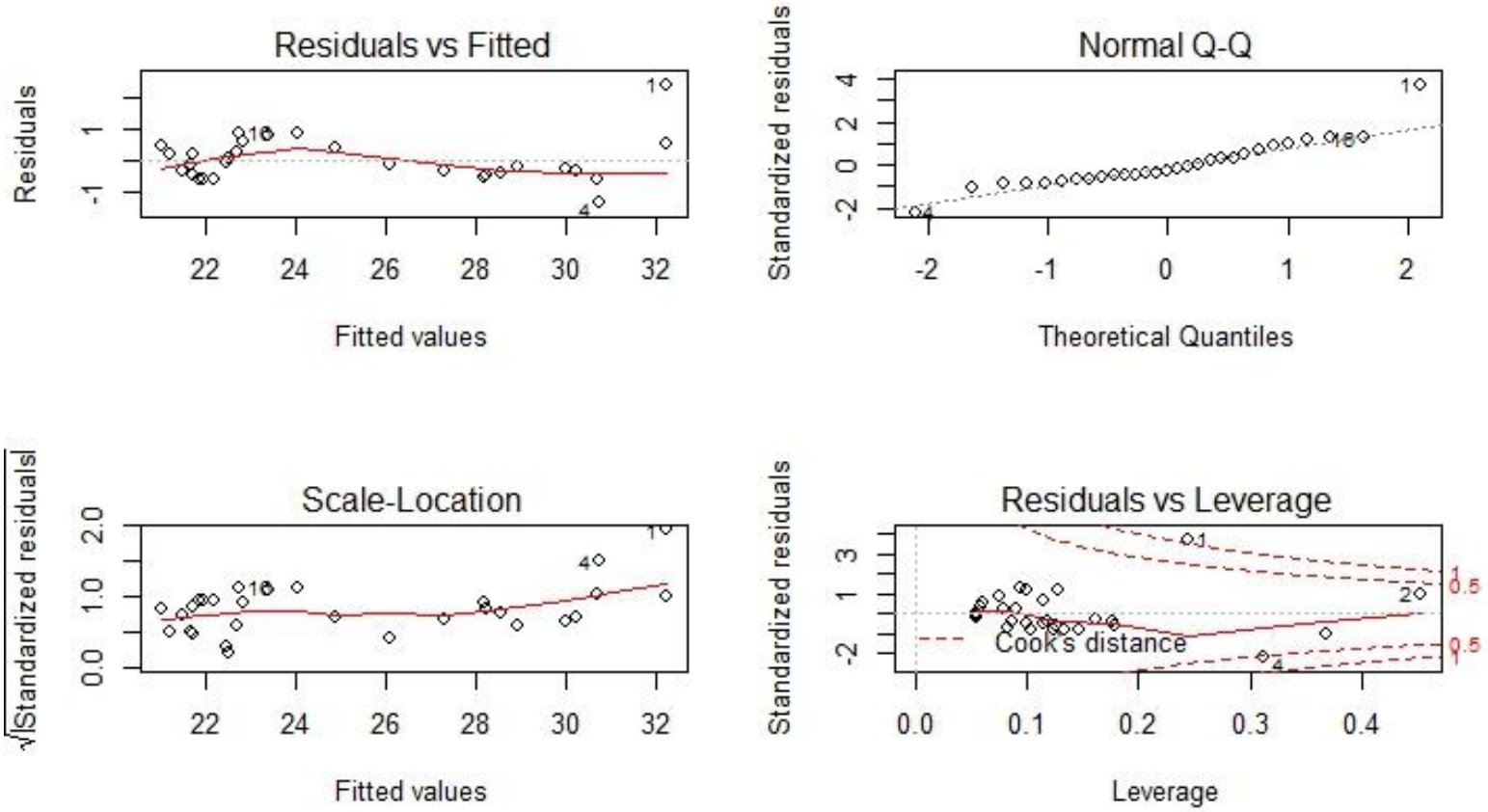

Figure 14- Multi-regression of import and production of nonrenewable energies

Figure 14 shows that when the importing is higher, the producing is reduced. This would be a good hint that the total amount of dirty power is not growing drastically.

For the next step, I want to do some techniques that I have learned from this class. Actually, because of the characteristic of my dataset, some techniques might not make sense and cannot be explained.

Figure 15 is a cluster of production of different types of energy. It shows that at the first level we have two groups. Because of the large amount, coal lies on one group and the

other types of energy can be put on the second group. This can illustrate the larger impact of coal in this dataset.

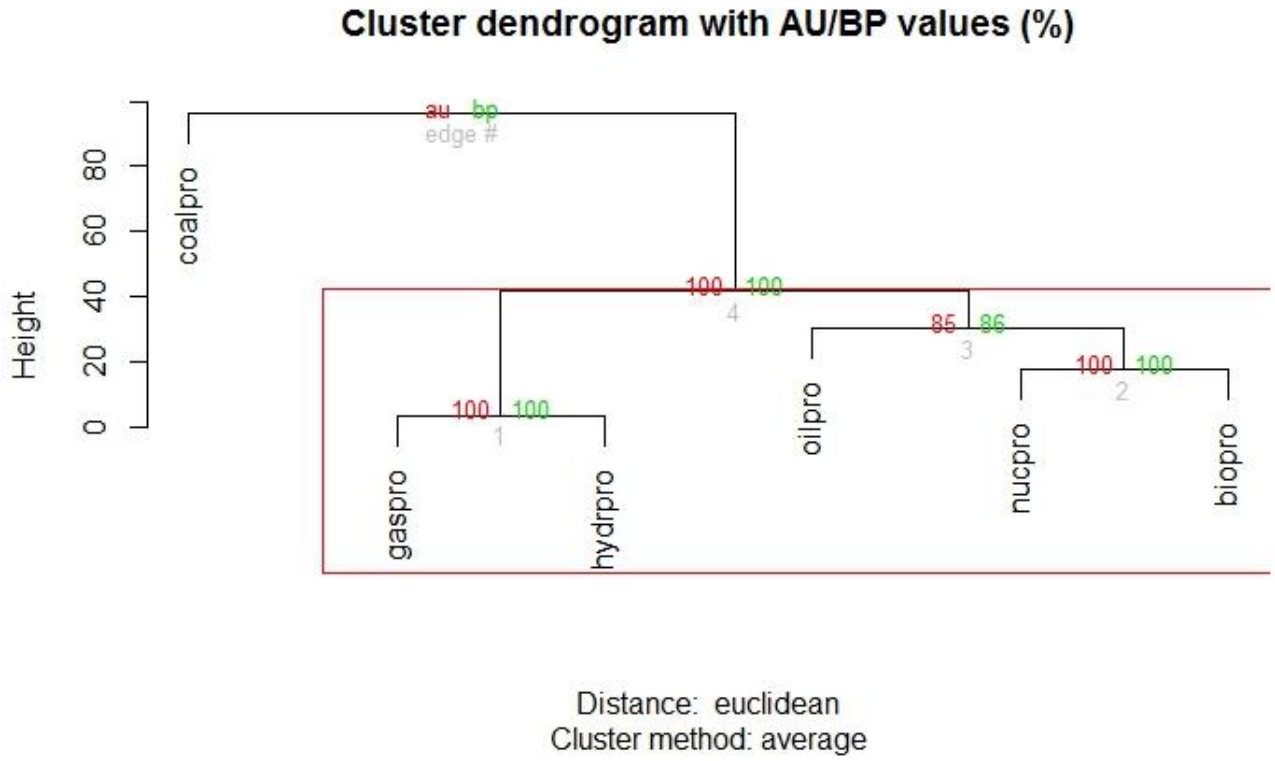

Figure 15- Clustering production of different types of energy

On the other hand we can test the impact of different types of energy by factor analysis. I want to find out that if I can put some types of energy in one group or factor according to their similarities. First I run the PCA to determine the number of factors. Figure 16 shows the histogram of variation of different values in the dataset and the reasonable number of factors which can be defined in this regard. This output tells that I can choose 3 factors but it might not make sense to use all three.

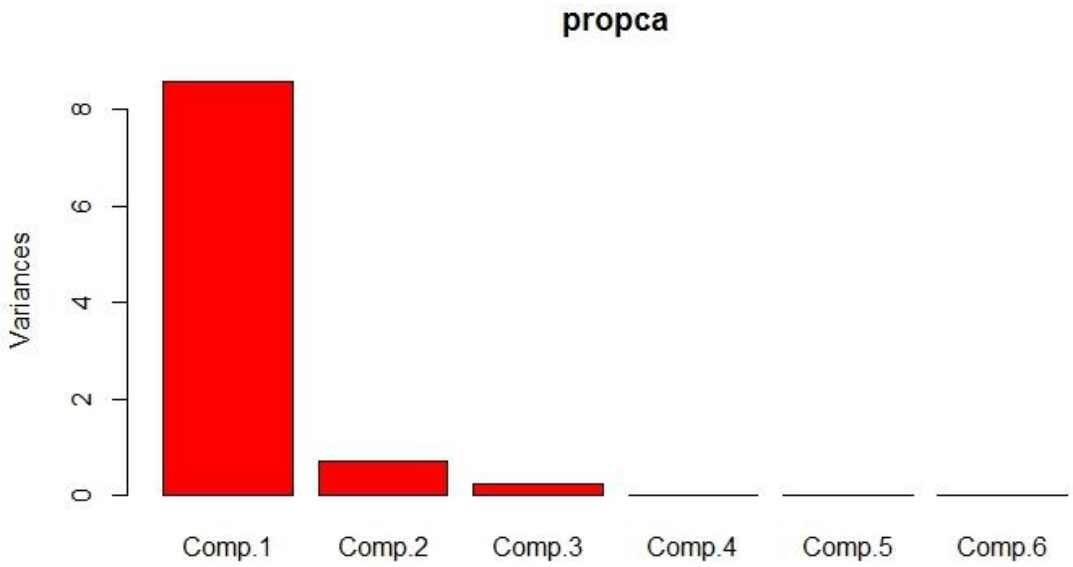

Figure 16- PCA test for the number of factors

I did the factor analysis and this is the summary of that:

Loadings:

	Factor1	Factor2	Factor3
Oil.Production	0.792	0.530	0.259
Natural.Gas.Production	0.151	-0.923	-0.113
Dry.Natural.Gas.Production	0.986		
Coal.Production	0.937	0.225	
Nuclear.Power.Production	0.799	0.476	0.361
Hydropower.Production	0.444	0.893	
Biomass.Production	0.974	0.208	
Other.Renewable.Energy. Production	0.486	0.870	

	Factor1	Factor2	Factor3
SS loadings	4.522	3.008	0.218
Proportion Var	0.565	0.376	0.027
Cumulative Var	0.565	0.941	0.968

Test of the hypothesis that 3 factors are sufficient.
The chi square statistic is 37.7 on 7 degrees of freedom.
The p-value is 3.45e-06

As it can be seen in figure 16, putting these values to different factors is not reasonable and the result of factor analysis confirms this. By comparing the coefficients of three factors we can conclude that the highest coefficient values are in factor one and the rest of the factors do not contain valuable coefficient values. So, here factor analysis cannot help us.

Conclusion

Regarding the characteristics of my dataset I could not get good results in my analysis. One of the main issues was the prediction characteristic of this dataset which leads the data to have an unrealistic distribution or trend. All in all, the aim of the project was struggling with the concepts of statistics and applying R studio to find out the relationships between the data.

References:

[1] "EIA-Annual Energy Outlook 2010."[Online]. Available: http://www.eia.gov/oiaf/archive/aeo10/index.html. [Accessed: 02-Jun-2015].

[2] "Environmental Impacts of Hydroelectric Power | Union of Concerned Scientists." [Online].Available:http://www.ucsusa.org/clean_energy/our-energy-choices/renewable-energy/environmental-impacts-hydroelectric-power.html#.VXNdic9Viko. [Accessed: 06-Jun-2015].

[3] "China - Three Gorges Dam - Impact - New York Times." [Online]. Available: http://www.nytimes.com/2007/11/19/world/asia/19dam.html?pagewanted=all&_r=0. [Accessed: 06-Jun-2015].

[4] "NREL: Energy Analysis - Renewable Electricity Futures Study." [Online]. Available: http://www.nrel.gov/analysis/re_futures/. [Accessed: 06-Jun-2015].

[5] *IPCC, 2011: IPCC Special Report on Renewable Energy Sources and Climate Change Mitigation.* United Kingdom and New York, NY, USA.