Greenhouse Gas Emissions and Renewable Energy from the UN: Dataset Analysis and Visualization in R

Matthew Hileman
University of Colorado at Colorado Springs
mhileman@uccs.edu

Abstract

World greenhouse gas emissions have spiked dramatically in the last century. In order to reduce these emissions in order to preserve the global environment, The United Nations are attempting to reduce the production of these harmful gasses. One such method is to switch the world use more electrical based systems that are powered by renewable energies. Discussion such, this report is divided into the following categories: (1) background and layout, (2) cleaning and manipulation, (3) individual analysis for greenhouse gas emissions, (4) individual analysis for solar and wind energy production, (5) combined mixed model analysis.

Background and Layout

Greenhouse Gas Emissions (GHGs)

The first dataset comes from the United Nations Framework Convention on Climate Change (UNFCCC), which are concerned on global climate change and measuring the effect thereof. Data is collected using scientific guidelines defined in their National Greenhouse Gas Inventories and Collections guidelines. The data itself is a collection of yearly estimates, while it is derived from individual sensory and air sampling.

Containing over 25 unique countries in the United Nations, the data is divided into 7 major greenhouse gas types - CH4, C02, HFCs, N20, NF3, PFCs, and SF6. Each gas is contained in its own separate csv file, with the year, country, and volumetric amount which is simply labeled "Value." This value is the emissions annual estimate, measured in kiloton C02 equivalent. Because each gas has different effects, densities, and physics, they have to be converted to a standard measurement, which in this case is C02 (the reason behind this choice will be seen later). There is also a total for all the gasses combined, which is labeled GHGs.

The data first starts at 1990 and spans until 2018. Some countries did not have any readings in the 1990's, and instead began recording in the early 2000's (mostly small, and insignificant in total emissions statistics). The data is mostly clean, the only issue will be in insuring the years that new countries began inserting new readings does not effect the data analysis with any significance.

Solar and Wind Energy Productions

The greenhouse gasses are interesting on their own, but cannot really be compared with each other using mixed models - at least not in an interesting or beneficial way. So, to have a nice dataset to use mixed model analysis, I chose to import another dataset that records the average annual energy production from the UN. The data was taken from the same source as the Greenhouse Gas Emissions, from a different sector known as Energy Statistics Database in the United Nations Statistics Division. Much like the greenhouse emissions data, this dataset follows similar scientific collection protocols and guidelines relevant to itself.

The solar and wind energy productions are separated into their own files as well, with a slightly different structure. There are footnotes at the bottom of the csv files which will need to be removed. The new measurement per year and country is under "Quantity," which is described in its own column as being: Kilowatt-hours, million. That is, the quantity is the Kilowatt hours in the millions.

Like the greenhouse gas datasets, there are countries whose data does not start until later. We will need to make sure this does not effect our analysis. There is also some random data from 2019 from a few small countries, while no other countries have any data on 2019.

Cleaning and Manipulation

Greenhouse Gas Emissions (GHGs)

Since the data is contained in several different files, each one had to be imported separately. I decided to combine them into one single data frame after importing, adding another variable named "gas" to classify between the different gasses. I also included the total greenhouse gas emissions in the combined result (GHGs). There were whitespaces in some of the values so I passed the import with strip white.

If there had been any significant errors, I would like to have used one of the imported cleaning tools available in R from the tidyverse library. Thankfully, the data was not too erroneous - but there were different years present for each country. I chose to leave them in due to two reasons: (1) their emissions were very small in comparison to larger countries, and (2) the sum between countries could be taken in yearly increments. Because of these, the data would not be misrepresented if included in the visualizations.

Solar and Wind Energy Productions

The solar and wind data files had some erroneous rows at the bottom of the dataset, as well as a few missing years. These missing years were much more reasonable, as some countries did not develop or implement solar/wind technology until recent - so their values are low for the ones that are missing. By taking the total of the energy productions for any year, this would eliminate any missing year issues.

Various other manipulations were done for both the greenhouse gas datasets and the energy datasets as the visuals presented more of a story and what was needed next was then decided. Such manipulations will be discussed when they surface in the visualization discussion.

Greenhouse Gas Emissions - Visualizations

Line Plots by Gas Type

The first illustration compares gas types with each other, to see which are most present in the greenhouse gasses emissions *(figure 1)*. The overwhelming majority of gas consists of C02, with roughly 450,000 kilotons in emissions starting in 1990. The next closest is CH4 at about 75,000 kilotons in 1990. This is a nice plot for C02, but all the other gasses are drowned out because of the large scale. Because of this, it would be beneficial to also plot the gasses without C02. As for the C02, it can be seen rising until 2007 until it drops dramatically and begins to continuously fall after. Although not a dramatic decline, realize the plot is from 0 - so the quantities are quite significant.

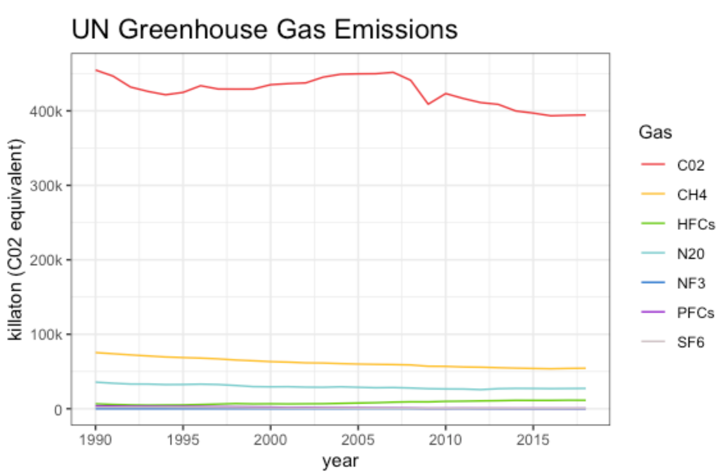

Figure 1: *United Nations Greenhouse Gas Emissions by type. Notice how C02 is the dominant gas.*

After removing C02 and plotting the next level, we are able to more clearly see the pattern of CH4, N20, and HFCs (HFCs are Hydrofluorocarbons. See *figure 2*). The other three gasses were still so small in emissions that they could not be seen, so another level of plotting would be helpful.

CH4 declined significantly - from 75,000 to 55,000 in a 28 year span. N20 slowly fades, but HFCs are seen as being produced more! I had originally predicted that **all** gas emissions would decline, but such is not the case. HFCs are the primary gas emission caused by cooling systems, air-conditioning, insulation, and aerosols. As population increases and computers are produced in higher amounts, it only makes since HFCs would increase as well - an unavoidable gas increase with the expansion of technology.

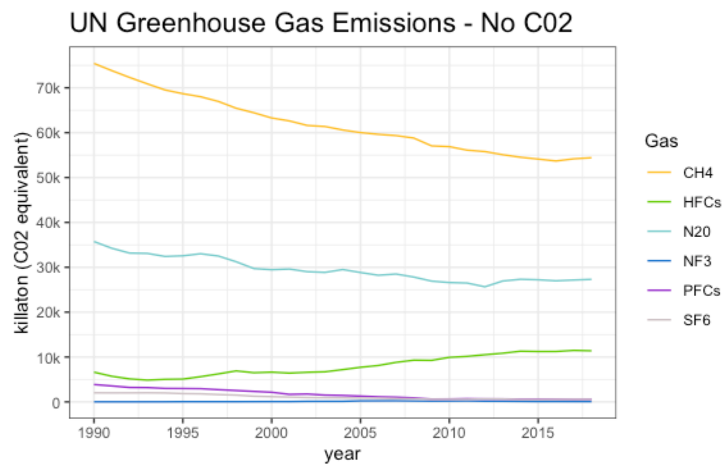

Figure 2: *United Nations Greenhouse Gas Emissions by type, without C02. Notice how HFCs are increasing?*

The final zoom level of the gasses shows emissions of very low impact gasses - SF6, PFCs, and NF3. The only one not declining is that of NF3, which has a bit of a bump in the early 2000's before decreasing again by 2018. By 2018, the gas emissions from these three gasses is less than 1000 kilotons.

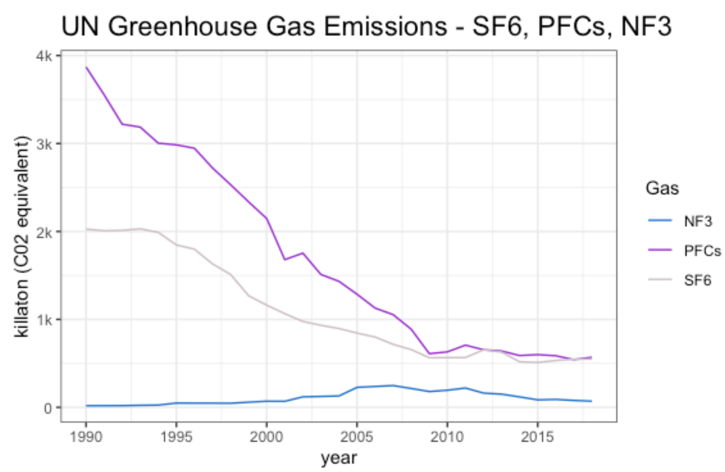

Figure 3: *United Nations Greenhouse Gas Emissions by type, three lowest value gasses. No wonder they couldn't be seen in comparison to the 450,000 kiloton peak of C02.*

After having plotted each gas individually, it would also be nice to have the overall, total greenhouse gas emissions plotted *(figure 4)*. The pattern looks very similar to C02 - and because C02 dominants the majority of emissions, this result is no surprise. Instead of basing the plot from 0, I chose to blow it up and have the bottom start higher so we can see the change more clearly. It seems the UN's efforts to reduce emissions is working.

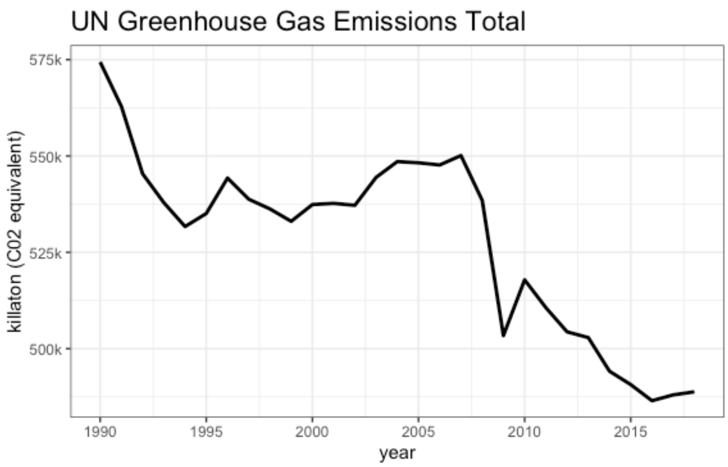

Figure 4: *United Nations Greenhouse Gas Emissions Total. Notice the scale. Although not too drastic, every kiloton of gas is significant to the climate!*

Donut Plots by Gas Type

Instead of just the above charts to illustrate the difference in gas types, I wanted to stress how much C02 there was in comparison to the other gas types. For this, I created three different donut charts - one for 1990, 2005, and 2018 (*figure 5*). Although the change is very slight, this visualization still shows the gas proportions effectively. I was hoping to see C02 emissions drop more than they do here, and it might even misrepresent the data in that regard - since donut charts are percentage based. To that effect, I also added the numeric kiloton totals underneath a percentage label so there is no confusion that the C02 levels did in fact drop. This type of visualization is only useful to show the percentages of gas types.

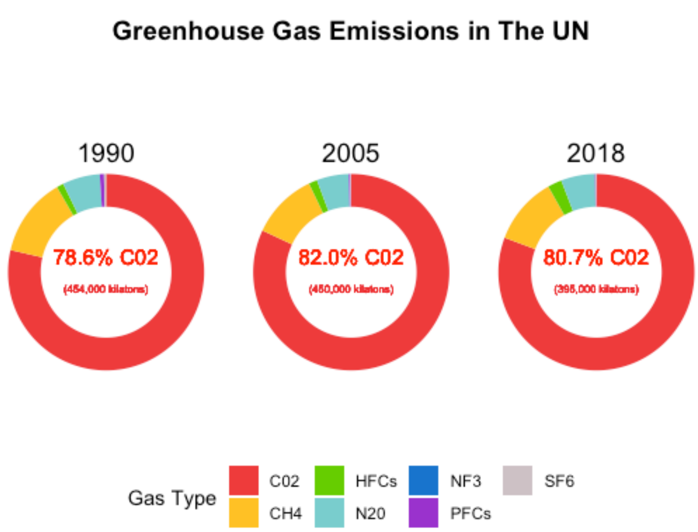

Figure 5: *United Nations Greenhouse Gas Emissions by gas percentages. C02, CH4, N02, and HFCs are significant.*

Bubble Plot by Country

How about countries? Which countries are producing the most greenhouse gasses in the UN? Since there are so many countries, it would be difficult to plot a conventional plot and show this information without it feeling cluttered. There is no China in the dataset, so keep in mind, this dataset is not for the world, but rather the United Nations. I'm unsure why China is not included - and it might be that they do not take measurements of gas emissions. There are ways to estimate China's emissions, which I could have included for a more complete picture - but instead decided to stick to only one source. So, to visualize the remaining countries, a bubble visualization was produced, where the area attributes to the total gas emission for that country (*figure 6*).

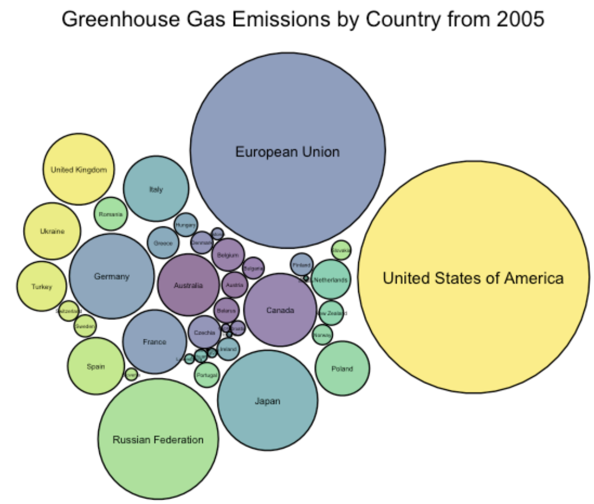

Figure 6: *Greenhouse gas emissions by country or area. The European Union (as a whole), Japan, Russia, and United States are the largest producers. See Bubble.png for a larger graphic.*

Wind and Solar Production - Visualizations

Now that we have a nice representation for greenhouse gas emissions, we should consider why gas emissions are declining. In particular, what produces C02 and CH4? The answer is that power plants and gas fuel produce these gasses. The UN has taken the initiative to try and transition to using mostly renewable energies, such as wind and solar power. As the world transitions to more electric based power and less coal/gas, the more these renewable energies can be utilized.

So, perhaps the decline of greenhouse gas emissions in the recent 8 years is due to more energy coming from solar and wind production. I plotted the results of both solar and wind production in a singular line plot to illustrate just this (*figure 7*). This is not the main dataset, so I won't be plotting anything further, just these to confirm their production has in fact gone upward. The amount of energy produced by these methods was surprising - almost exponential in behavior at first before becoming more linear.

While also interesting on their own, the solar and wind data can help us link gas emissions to their increase. I'm sure that this is not the only reason why greenhouse gasses have gone down, but their is a relationship. It is important not to mistake correlation to causation.

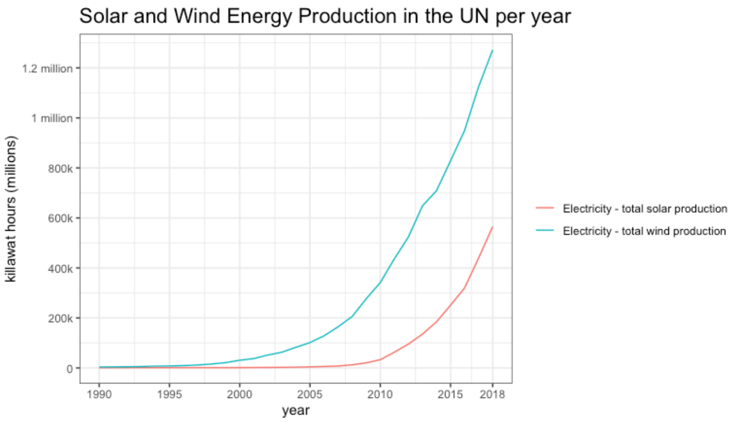

Figure 7: *Solar and Wind Energy Production in the UN. Wind (blue) is about twice as much energy as solar (red). Only in recent years has their production significantly increased (correlating to the recent decline in gas emissions).*

Mix Model Analysis

And now for what we have been leading up to. How do solar and wind energy productions effect greenhouse gas emissions? I predict that they do in fact have a correlation and that as renewable energy increases, greenhouse gas emissions decrease. It's fairly simple to prove this in R. We can use the lm() function with an equation to fit our mixed model to produce a fit. Then, we can use summary() with that fit to see the raw result.

I was unsure the relationship exactly. By comparing the gas emissions plots to the renewable energy increase plots, we can see that the energies are somewhat exponentially increasing, while the gasses decrease slightly. It's probably a linear to poly/exponential relationship. But, let's stark with a simple comparison.

We want to read the change in GHGs (greenhouse gasses), so that becomes our dependent variable and goes first in the equation. The wind and solar energies are the independent variables. I first decided to create a mixed model with each one individually, so:

GHGs ~ solar
and
GHGs ~ wind.

After confirming that there was in fact a relationship, I decided to add them both to the same mixed model with:

GHGs ~ solar + wind.

For the sake of keeping this report a reasonable length, I will only explain the fit of this model's summary(). After running summary on the combined, we are given the information seen in *figure 8*. The residuals are quite high, but that's to be expected with numbers that are high. The import part is that the residuals balance out - that is, adding them together should be closer to 0 scaled with the highest amount. Adding all the residuals gives roughly 170,000, and divided by the max (1,191,680) equates to about 1.44. This is the uncertainty percentage range, which is low.

```
> summary(linear_combined)

Call:
lm(formula = GHGs ~ solar + wind, data = data)

Residuals:
    Min      1Q  Median      3Q     Max
-816143 -289182  -82220  167055 1191680

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.352e+07  1.229e+05 191.425  < 2e-16 ***
solar        4.727e+00  2.028e+00   2.331   0.0278 *
wind        -4.171e+00  7.829e-01  -5.328 1.42e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 467900 on 26 degrees of freedom
Multiple R-squared:  0.8106,    Adjusted R-squared:  0.796
F-statistic: 55.64 on 2 and 26 DF,  p-value: 4.036e-10
```

Figure 8: *Summary of the combined mixed model, as outputted in R. Notice Multiple R-squared and p-value for relational significance.*

The next important output is the Coefficients. The intercept of wind and solar are very important, they tell us how much gas productions are estimated to reduce or increase based on their inputs. Ignoring the topmost because its still without the complete variable input, for every million kilowatt hours produced by solar and wind energy, roughly 4.171 kilotons C02 greenhouse gasses is reduced! That's quite significant.

Now to make sure the data is accurate. Notice multiple R-Squared is 0.8106, which is the variance number. So, variance is 1 - 0.81, or 0.19 (under 0.5, so the data is sound). Also observe that p-value is a very small number (4 x 10^-10). This number indicates that there is a high relationship between GHG emissions and solar/wind energy production.

Finally, let's take a look at the plot for our fitted model by using plot() to the fit in R *(figure 9)*. First, the normal Q-Q makes sure residuals are distributed evenly. They should all be along the line. There is some issues at the tail end of our normal, but that can be explained by the fact that energy productions were at 0 for the early years that GHGs were recorded. The first part of the graph seems to be fine.

Next, the Residuals vs Fitted plot shows the pattern of the residuals (to see if it's a linear fit or not). Ours is most certainly curbed at the end, but mostly follows a horizontal pattern (which makes the linear comparison alright). Our plot isn't quite linear anyway, we have two variables influencing one (ex z ~ yx), so not having a completely stable line for this one is acceptable.

The Residuals vs Leverage plot is to identify heavy influentials. So long as nothing crosses the Cook's Distance line, the fit is acceptable. Just make sure that no residuals are on the other side, which would mean there is an outlier influences the data too heavily.

And the scale-location measures equal variance as well. The more horizontally fit this line, the more sound the variance. Ours is mostly fit except for a slight bump in the mid-end.

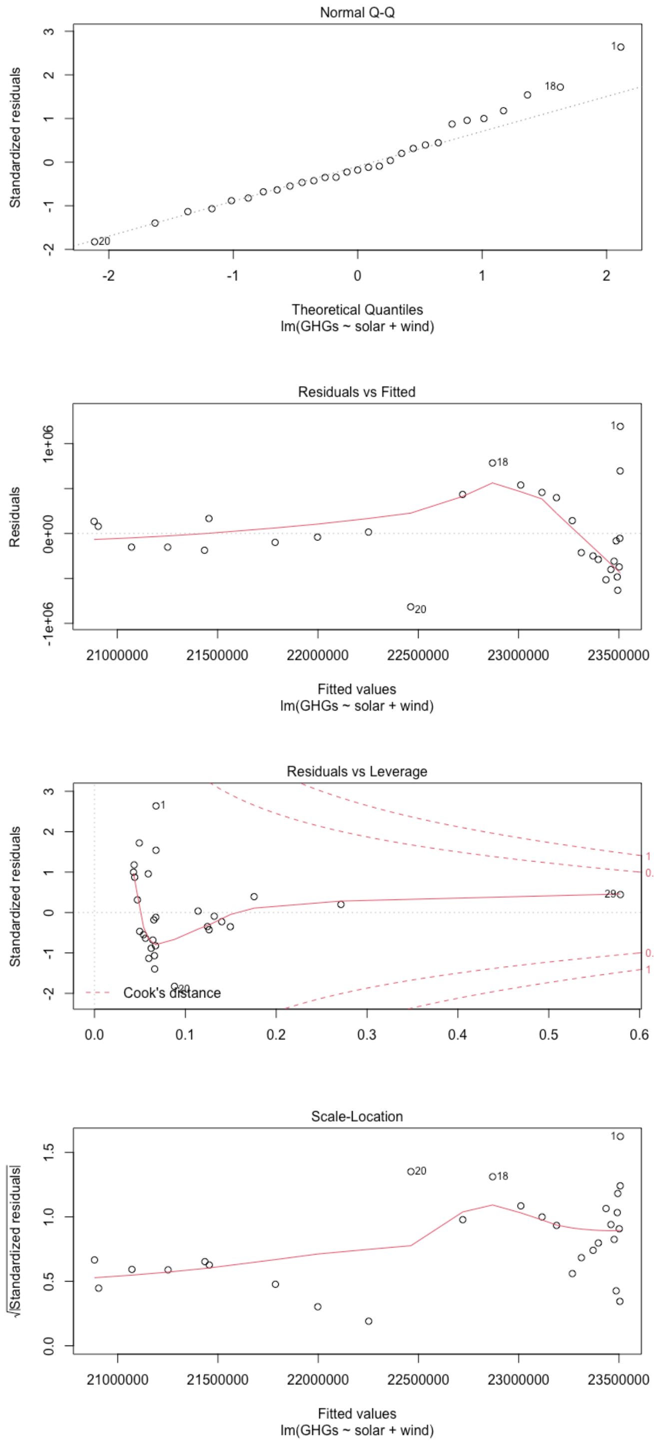

Figure 9: *Result of running plot() on the linear mixed model.*

After having compared the linear mixed model, I decided to add a polynomial mixed analysis as well. Like the linear, I first started with each energy source (wind and solar) individually before combined. Likewise, a summary and interpretation was done on this mixed model. The following equation is the final:

GHGs ~ solar^4 + wind^4.

Now that we have our mixed models, let's plot them against gas emissions and energy production. I produced a total of 3 graphics for this section, for each of the energy sources and then the combined final *(figure 10, 11, 12)*.

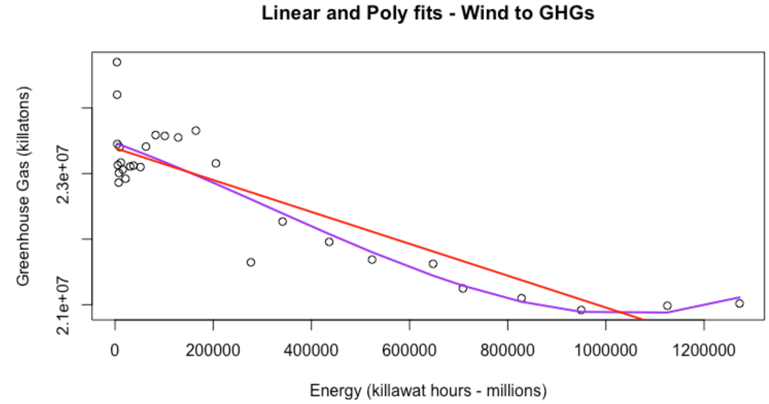

Figure 10: *Result of running wind's linear (red) and poly (purple) fits. The linear doesn't fit as well as the poly equation.*

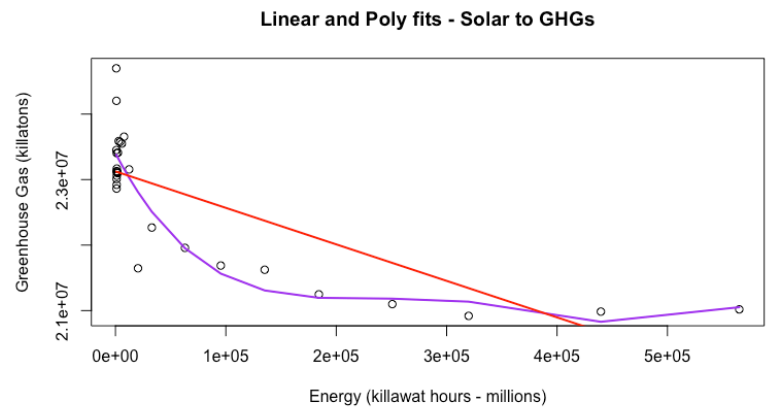

Figure 11: *Result of running solar's linear (red) and poly (purple) fits. As with he wind, the poly fits better.*

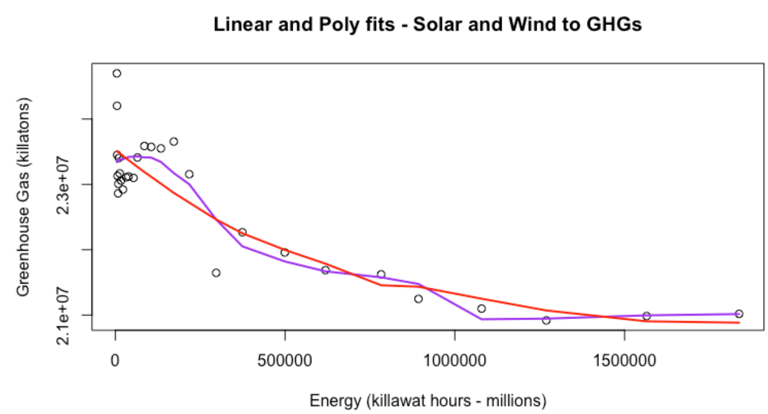

Figure 12: *Result of running the combined linear (red) and poly (purple) fits. Both fit the plot well, but I prefer the linear combined (red) as it fits the general idea better.*

The plot tells us that the greenhouse gas emissions diminishes slower as wind and solar energy increases. I did not expect this result, and though the amount would be more significant at the end. The relationship is not 1 to 1 however, and we did not consider that other variables may be effecting the greenhouse gas emissions. At the very least, the production of the solar and wind energies has made the emissions come to a stand still.

Conclusions

Solar and wind energy production in the UN has impacted greenhouse gas production, or so our mixed model's indicate. We did not account for population growth or increased power usage, which may increase just how impactful these renewable energies really are. In any case, renewable energies have reversed or slowed the annual production of greenhouse gasses in the UN.

References

UN Data, "A world of information" databases, 20221. Available at: http://data.un.org/Explorer.aspx
(Datasets also included in zip under Final_Part_2/data)