Jay Brothers: jbrothers@clarku.edu
Will Setlik: wsetlik@clarku.edu
Ben Seidenberg: bseidenberg@clarku.edu
Christopher Siems: csiems@clarku.edu

Green Energy Exploration

Introduction

Goal seven of the United Nations sustainable development goals is to 'Ensure access to affordable, reliable, and sustainable energy for all.' This is a lofty and important goal. Access to reliable energy is one of the key features of a high standard of living. Aswell, the future of the whole planet depends on our ability as a species to combat rapid climate change, which is primarily driven by energy generation through polluting means. This is the specific aspect of United Nations sustainable development goal seven that we chose to focus on. Specifically, we wanted to examine discussion of clean energy within the United Nations and the relationship between this discussion and actual action in the realm of the switch to clean energy.

Goal seven of the United Nations sustainable development goals relates to one of the most pressing, contentious, and difficult issues of our time. Climate change is an urgent matter with real world ramifications being felt all over the globe, even today. Potentially the most important aspect of solving climate change is making the switch from polluting, fossil fuel based, energy generation to green/clean/sustainable energy generation. Analysis of the dynamics of discussion of these energy types and analysis of the actual data about their usage will be important for informing the public about how discussion among policymakers leads to actual action, if it does. This is why we selected this topic; we care about the future of this planet and we want to make sure that the people in positions of power that we entrust to tackle these issues are having an effect. There is a lot to goal seven of the United Nations sustainable development goals, but understanding the progress that has already been made towards this goal is extremely informative.

Collection

We used two datasets for our exploration. The dataset primarily used to build this project was a comma separated values (CSV) file holding the text of every speech given during the annual General Debate at the United Nations General Assembly from 1970 to 2016. Each row of the dataset represents a speech and is labeled with the year the speech was given and International Organization for Standardization alpha 3 country code, the standard three letter code representing each country. Each row also included a transcript of the speech. The other dataset was another CSV holding comprehensive data about the energy production and usage per year of a set of countries. Each row of the data was labeled with a country's name, a year, a type of energy, and a unit that the energy was being measured in. The data contained within each row was the amount of that energy either generated or used.

From the first dataset, the country and year attributes were used to group the data. The data examined from this dataset was exclusively the speech transcripts themselves. From the second dataset, groupings were done using the country and year labels and examination was done of the type and amount attributes.

Processing

The debate dataset was enormous to begin with. The first step in cleaning it was restricting the time period we were examining. The other dataset only had data from 1990 to 2013, so we felt that restricting the debate data to within that period was a reasonable way to make the processing more manageable. Some other cleaning tactics included dropping the column giving the session of the speech, that information can be gleamed from the year data, and preprocessing the tokens from speeches themselves so that work would be easier down the line. Using these tactics we produced two cleaned datasets, one with every country and one with only the countries included in the other dataset.

year	country	text	tokens	keywords
2013	DMA	I would like to begin \nby congratulating you,...	would like begin congratul...	47

2013	TUV	It is a great honour and \npleasure to partici...	great honour pleasur particip...	43
2010	TJK	I would like to \ncongratulate Mr. Joseph Deis...	would like congratul mr....	42
2013	FSM	I am pleased to join previous \nspeakers in co...	pleas join previou speaker...	42
2010	GUY	I bring to all delegates at this \nsixty-fifth...	bring deleg sixty-fifth...	41
2012	DMA	On behalf of the\nGovernment and people of...	on behalf govern peopl...	41
2013	BRB	It is my distinct honour \nand pleasure to con...	distinct honour pleasur...	38
2013	TTO	This \nevening, as the Chairperson of the Con...	thi even chairperson confer...	37
2013	NRU	May I first congratulate Mr. John \nAshe on hi...	may first congratul mr. john...	36
2012	GRD	I am honoured to address\nthe General...	honour address gener...	35

The energy statistics data set is a collection of yearly country energy use. The data has over 1 million lines, with a thousand categories of energy data types. These categories cover distinct types of energy, their production, consumption, supply, and demand. Many of these categories had insufficient data or were too focused. To fix this, we made a DataFrame holding each energy type and how many observations were in the data for that type. Observations ranged from as little as one, to as much as 1600. For our purposes we only need the overall production of each energy type per country. To do this, we filtered first by country and date so that we only had comparable data. Then we flipped the columns on the data frame so that each energy statistic was a column along with the country and year. This made the data more readable and condensed the information into fewer lines. The final data set consisted of time series data of each country from 1990 – 2013, dates consistent with the other data set. We saved fifteen broad statistics of several types of energy as well as total energy so that we can gauge the importance of each source to countries. In our original plans, we decided to track greenhouse gas emissions too, which required merging of another data set. Upon further consideration, we decided to narrow our focus to strictly clean energy.

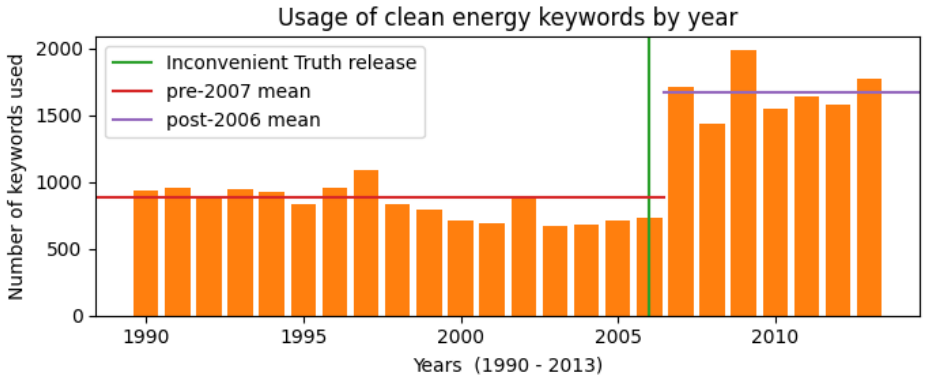
Analysis

Our analysis began with statistical representation of the data we collected on green energy keyword use and green energy generation. With this data we drew charts several types of charts (histograms, bar graphs, line graphs, and scatter plots), we generated and overserved distributions (normal and uniform). We then used this statistical data as the background for cursory natural language processing and machine learning.

The key metrics we used in our examination were a counts of green energy keywords from each of the speeches, energy usage and production data generated and retrieved from our energy dataset, and the sentiment of speech sections devoted to discussion of green energy. These metrics and our statistical methods enabled this exploration.

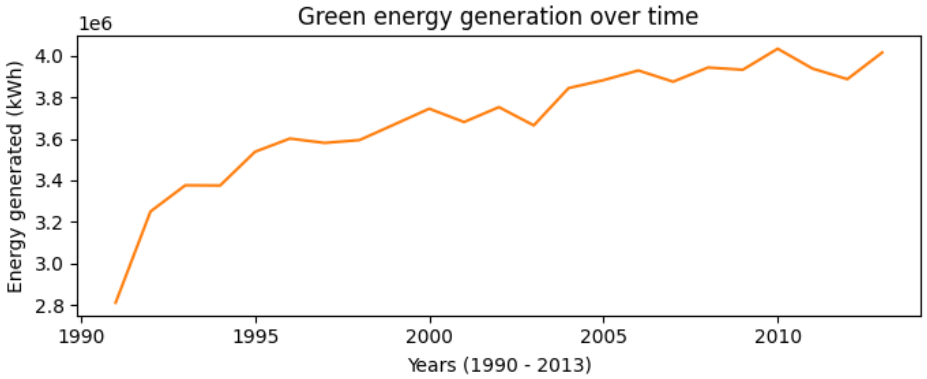
Communication

The primary question we wanted to explore with this report was: what is the relationship between green energy advocacy in the United Nations and the real-world transition to green energy? Our first step in exploring this question was a cursory look at green energy advocacy in the United Nations general debates.



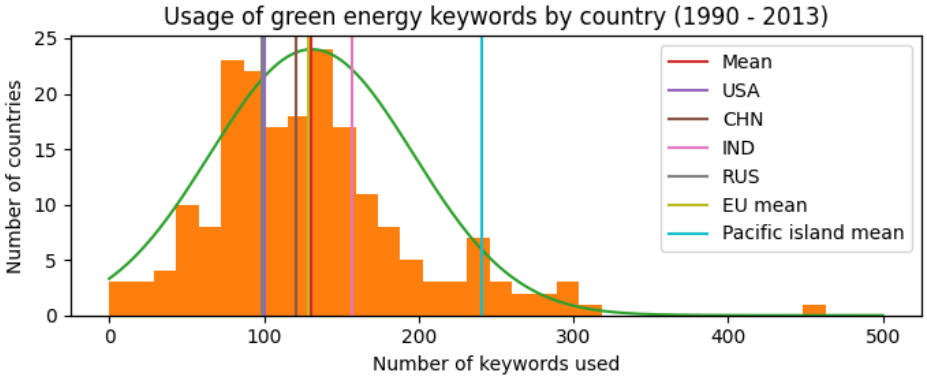
The above graph shows usage of green energy keywords at sessions of the United Nations General Assembly over time. Our expectation was that when graphed we would discover that the data showed an upward trend, this is the case, but it materialized differently than we expected. Instead of a linear or exponential upward trend, there appear to be two uniform distributions, one before 2007 and one after 2006. When investigating possible explanations for this we developed a hypothesis to explain this phenomenon. We propose that this is the effect of *An Inconvenient Truth*'s 2006 release. *An Inconvenient Truth* is often considered the first piece of mainstream media to bring the public's attention to climate change, and it is our hypothesis that this documentary led to heightened awareness of the topic at the United Nations.

Somewhat un-shockingly, in this same time period, green energy generation increased by a not insignificant amount.



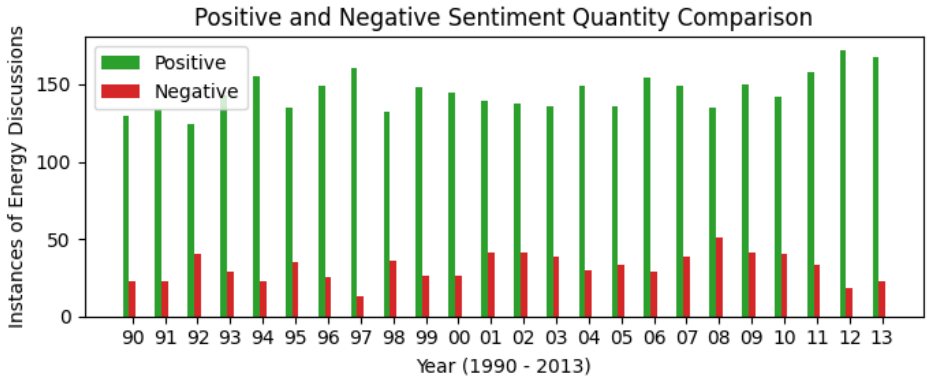
This data only considers nuclear, hydro, wind, and solar energy generation from a subset of UN countries: Australia, Austria, Belarus, Belgium, Canada, Croatia, Cyprus, Czechia, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Japan, Latvia, Lithuania, Malta, Netherlands, New Zealand, Norway, Poland, Portugal, Romania, Russia, Slovakia, Slovenia, Sweden, Switzerland, Turkey, Ukraine, and the United States. We only examined these countries due to a lack of data for other countries.

The next piece of this topic that we chose to examine was the distribution of keyword usage by country. We expected to find that when charted, the number green energy keywords used in each country's speeches during the time period would form a normal distribution. We were about correct.



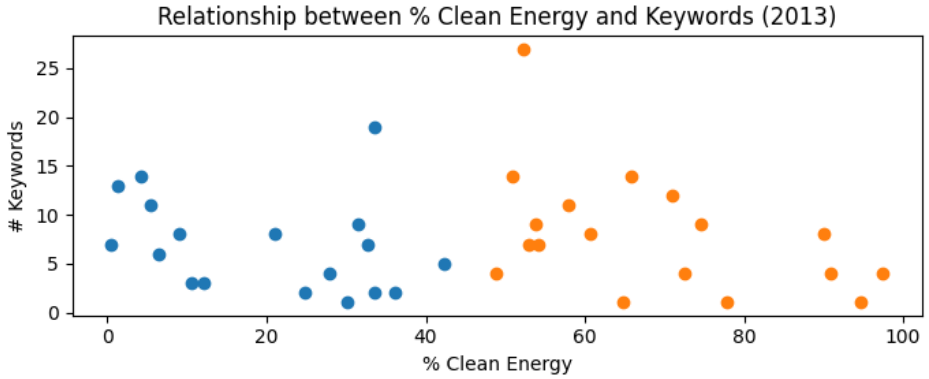
The data did seem to show a rough normal distribution. It seems that most countries have used about 100 to 200 green energy keywords from 1990 to 2013. We tried to find a pattern in the countries that use more and fewer green energy keywords, but mostly failed. The exception was when we examined the Pacific Island nations. Intuitively, the Pacific Island nations that stand at risk of their whole countries going underwater seem to use more keywords in their addresses to the United Nations. Though we suspect that this is less due to their own usage of these technologies, but rather they are appealing to larger nations to make the switch. The single nation that used about 450 keywords in this period is in fact the Federated States of Micronesia.

Another aspect of our work was using sentiment analysis to inspect the general sentiment of green energy discussions. The work exposed that general discussion at the United Nations seems to be less pessimistic than one might expect. Instead, diplomats seem more hopeful than doubtful.



This graph compares the sentiment of parts of speeches about clean energy by year.

Moving onto exploring the relationship between green energy keyword usage and actual green energy usage. Our plan was to chart the smaller set of countries in our data set in a scatter plot where the x-coordinate of a country's point on the chart would be the percent of their electricity generated from green sources and the y-coordinate would be the number of keywords they used. From here we wanted to color the data using the k-means algorithm to show the distinct groupings that we expected to form.



There is seemingly no good relationship between a country's usage of green energy keywords and the makeup of their energy generation. Thus, this part of our exploration was inconclusive.

From our exploration of the data, it has become clear that a nation's willingness to discuss the issue of green energy at the United Nations General Debates does not have a noticeable impact on their domestic usage of these energy sources. In general, most countries seem to agree that green energy must be pursued, but it seems that the actual pursuit of this future falls to many other actors than the governments who send these diplomats to the United Nations.