Leaders in Renewable Energy and the Explosion of Production in China

John Yoo

Introduction and Guiding Question

Energy security is one of the biggest concerns for our government today. The availability of energy at affordable prices is essential to our infrastructure and economic engine. It's a critical component of our country's defense strategy as well; the Department of Defense relies on petroleum for approximately 77% of its energy needs.[1]

China is currently setting the standard as the largest power producer in the world today. Since 2005, the production of solar cells in China has expanded 100-fold and they're still growing in their renewable energy production.[2] We're interested in how this growth in China equates to the rate of production in the United States.

The decisions we make to reduce our dependency on fossil fuels will shape the future of our economy. I believe that the awareness of our energy shortcomings will grow and would like to analyze the current state of our renewable energy output in comparison with other countries.

For our analysis, we're interested in a few guiding questions:

- **Who are the largest producers of energy and how much of their power is driven by renewable energy? Does China really own the world in energy production?**
- **Which countries are commited to renewable energy dependence?**
- **What is the gap between the United States and China and how fast is it growing?**

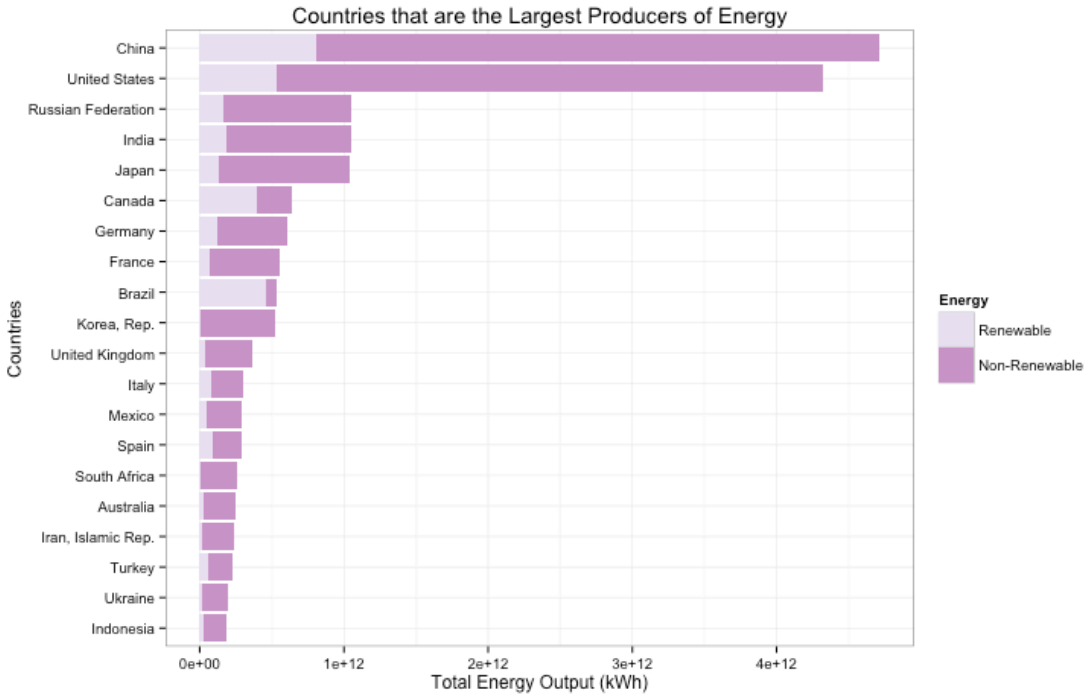
The data that will be used for the analysis of these questions will be the World Bank API, which contains indicators for renewable energy production, as well as total energy production. The data we aren't able to use for our research is renewable energy production through 2013 for

[1] Parthemore, C. (2010). "Fueling the Force: Preparing the Department of Defense for a Post-Petroleum Era". Center for New American Security.

comparisons between US and China as well as for comparisons with all other countries. The reasoning behind this is that the data is unavailable or too sparse to make any meaningful conclusion or inference.

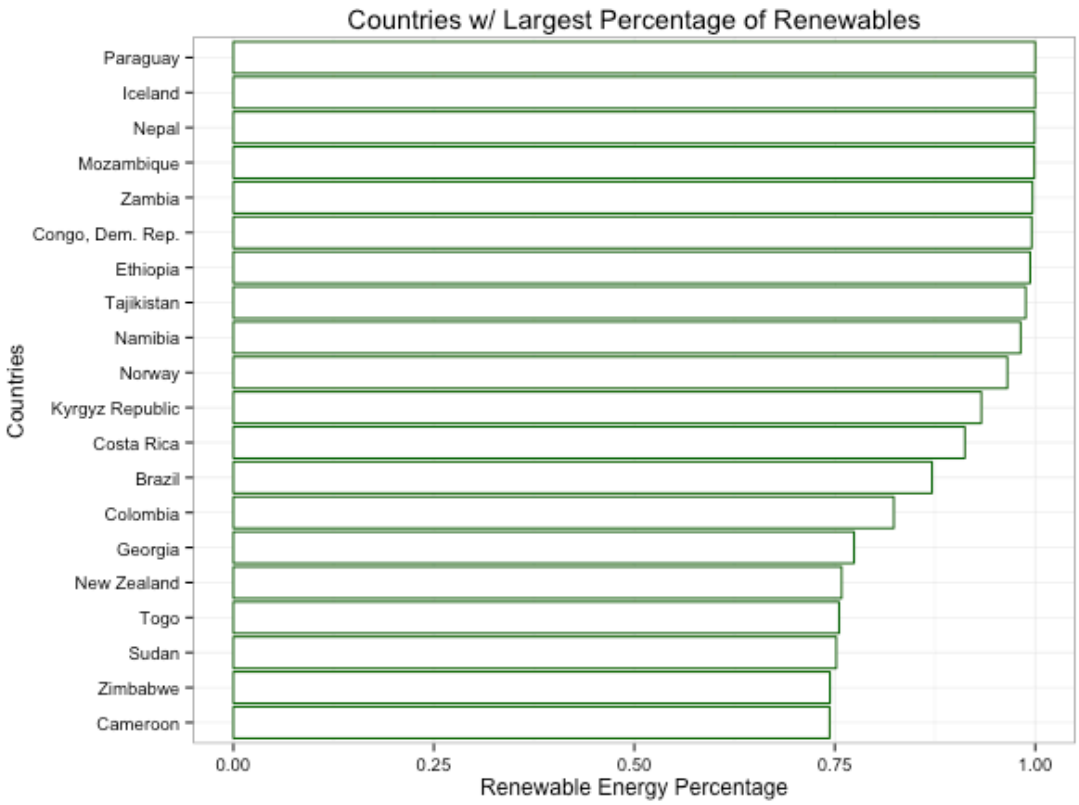
Response

WHO ARE THE LARGEST PRODUCERS OF ENERGY AND HOW MUCH OF THEIR POWER IS DRIVEN BY RENEWABLE ENERGY? DOES CHINA REALLY OWN THE WORLD IN ENERGY PRODUCTION?



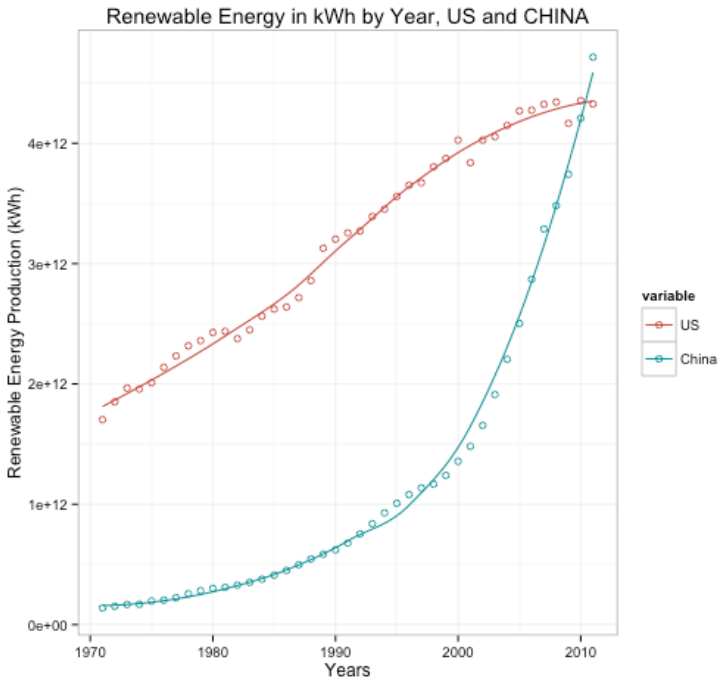
China and the United States together produce more energy than the next twenty highest producers of energy combined. However, production of renewable energy as a percentage of their total production for China and the United Stats is only 17% and 12% respectively. Of the top energy producers, Canada and Brazil rely most heavily on renewables at 63% and 87% respectively.

WHICH COUNTRIES ARE COMMITED TO RENEWABLE ENERGY DEPENDENCE?

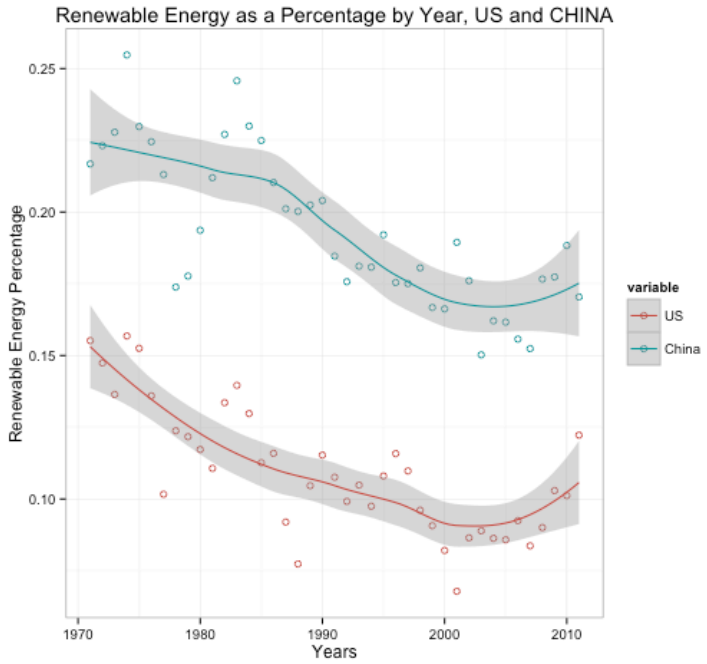


The countries listed in the table above are only the top twenty countries that depend heavily on renewable energy. After performing research to validate the results, we came across some interesting findings. Paraguay's Itaipu Damn provides 90% of its electricity and 19% of Brazil's. Iceland, which is located on a volcano, is supplied by nearly 100%

WHAT IS THE GAP BETWEEN THE UNITED STATES AND CHINA AND HOW FAST IS IT GROWING?



We confirmed speculations that China was growing their production of renewable energy at a rapid pace, while the United States is still heavily dependent on non-renewables. However when, looking at renewable energy production as a production of total energy usage, the countries are very similar in their rate of growth. The current gap of renewable percent is stable at 20%.



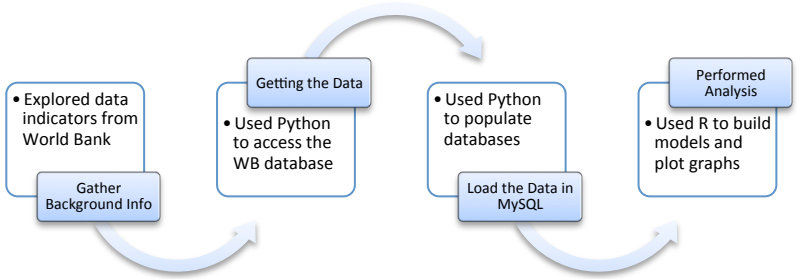
[

Successes and Shortcomings

We were able to provide a response for all of our guiding questions regarding renewable energy production and commitment. We wanted to visualize the proportions of renewable energy production and the countries that were most dependent on renewable energy as a resource. We also wanted to key in on the growth relationship between the US and China. All of these items listed we were able to accomplish.

As stated in the introduction, the analysis would have been more comprehensive if the data for renewable energy production was available through the year 2013. Because of this, we are relying on the trend and predictive power of the linear model for the data from 2011 to make our assumptions. Also, because the scope of this project was limited to the guiding question initially requested, as a follow-on project, we would be interested in tightening our focus on the relationship between US and China while plugging in more indicators to gather more predictive results.

Project Workflow



We began our analysis by searching for the data that we will use for our analysis at the World Bank Data Bank. Once we identified renewable energy production and total electricity production as indicators for our analysis, we used the World Bank API to obtain Pandas data frames for our indicators in 2011 (lines 23 – 30).

Once the data was gathered we formatted it into a single data frame that would be able to manipulate in R (lines 33 – 39). Since we had some NA and 0 values in our data frame, we had to drop them using *".dropna()"* and setting a condition for non-zero values (lines 42 – 43). This allowed us to keep rows that only contained data for each column in the data frame.

While in Python we also needed to gather the data for years 1971 – 2011 for the United States and Canada for their comparative analysis. To do this, we first identified our lists so they would be used with our function (lines 47 – 53). The function uses an *if/else* statement to grab data for either the US or China. We are using the same technique used for gathering our 2011 data on renewable energy for all countries in the world, but we are gathering it in a slightly different way (lines 56 – 66). The *for* loop that calls this for each country in the *countryList* is in lines 68 – 69.

Since we gathered our data in a unique way we have to set up dictionaries in a slightly different method that was used previously (lines 72 -29). Finally, we finish off our work in Python by

sending all the data tables to MySQL while also making CSV files of each data frame (lines 84 – 109).

In R, we start off by importing our libraries and importing our data from MySQL (lines 1 – 15). We then rename columns and add a percentage column, as well as a normal electricity column for a stacking bar graph (lines 20 – 22). After the data is ready for the world data portion of this analysis, we prepare the data frames for graphing (select data, melt data, order data) and graph each individual visual using *ggplot2*.

The same task is done for the comparison between the US and China, but we add a twist by creating a linear regression model for total amount of renewables produced and percentage of renewables used (lines 92 – 93 and 111 – 112).

Separation of Tasks

Python was used mostly for the gathering of data. Its ability to take full advantage of the World Bank API was helpful for understanding the capabilities for the project. The tasks we completed with Python included:

- Gathering the data from the World Bank data bank
- Setting up data frames for output to MySQL
- Connecting to the MySQL database and putting the tables in those databases

R performed a majority of the tasks related to statistics and plotting. It was great for manipulating the data once collected from MySQL and outputting the data in a consumer-friendly format. We spent a lot of time setting up graphs and did a lot of our analysis in R. Tasks include:

- Grabbing the data from the MySQL database
- Manipulating the data frames further for graphing
- Using ggplot2 to visualize the data in several ways (bar plots and scatter plots)
- Creating linear models to calculate relationships between different variables

```
Call:
lm(formula = chnenergy.total.kwh ~ usenergy.total.kwh, data = renewable)

Residuals:
       Min         1Q     Median         3Q        Max
-8.596e+11 -5.615e+11 -8.232e+10  3.550e+11  2.139e+12

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)        -2.616e+12  4.136e+11  -6.325 1.83e-07 ***
usenergy.total.kwh  1.200e+00  1.260e-01   9.523 1.00e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.808e+11 on 39 degrees of freedom
Multiple R-squared:  0.6993,    Adjusted R-squared:  0.6916
F-statistic: 90.69 on 1 and 39 DF,  p-value: 1.003e-11
```

```
Call:
lm(formula = usenergy.renewable.percent ~ chnenergy.renewable.percent,
    data = renewable)

Residuals:
      Min        1Q    Median        3Q       Max
-0.039429 -0.005706 -0.000872  0.007091  0.031087

Coefficients:
                             Estimate Std. Error t value Pr(>|t|)
(Intercept)                 -0.009344   0.017861  -0.523    0.604
chnenergy.renewable.percent  0.615608   0.091499   6.728 5.07e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.01535 on 39 degrees of freedom
Multiple R-squared:  0.5372,    Adjusted R-squared:  0.5253
F-statistic: 45.27 on 1 and 39 DF,  p-value: 5.074e-08
```