The Data Open Datathon Report

Improving Community Health with Renewable Energy Source

Peiying Yu, Jiahao Huo, Xiaohui Li, Chuchuan Wu

Feb 23, 2019

Introduction

Traditional energy sources have successfully led the human society to a highly-civilized level with industrial revolutions. However, at the meantime, the traditional energy sources are non-renewable and create serious society problems. Toxic gases, which are produced from consumption of non-renewable energy sources (e.g. petroleum, coal, natural gas), can induce various health problems and potentially threaten human being's life.

Hence, we ask a fundamental question: whether or not the advent of renewable energy source can reduce health problems caused by toxic gases emitted from traditional non-renewable sources. To answer this question, we need to find whether toxic gases produced from non-renewable energy will cause respiratory health issues. In this report, we focus on asthma disease, which is common respiratory disease, and we carefully examine the following relationships: (1) causality between renewable energy source consumption and air quality, (2) causality between air quality and asthma rate, (3) causality between renewable energy consumption and reduction in toxic air components, and (4) causality between air components and air quality.

Data Cleaning

To gain insight on how the distribution of energy consumption in renewable and non-renewable energy change air quality and reduce asthma problems, We choose to incorporate 'sed.csv' dataset with asthma and air quality (AQI) data obtained from Centers for Disease Control (CDC) [1] and United States Environmental Protection Agency (EPA) [2] respectively.

We use code 'RETCB' to extract total consumption in renewable energy (in Billion BTU unit) for each state in the United States from 2000-2016, which is the year range of our asthma data. To obtain consumption in non-renewable energy (in Billion BTU unit), we first use code 'TETCB' to extract total consumption in energy (in Billion BTU unit), and subtract the previously extracted consumption in renewable energy quantity from this quantity.

Since states with larger area will consume more energy, we wish to obtain consumption in renewable and non-renewable energy per unit area for each state. With the state area data [3], we calculate the consumption per unit area (BTU/km^2) in both renewable and non-renewable energy sources for each US state from 2000-2016.

Data Exploration

A glimpse into the pair correlation plot presents us with an interesting pattern, there is no seeming correlation between renewable consumption and the air quality index, while an irrational correlation between renewable consumption, which without further proof is highly likely to spurious. However, we regarded it as a potential and wanted to further investigate the relationships between renewable consumption, air quality and the prevalence of asthma.

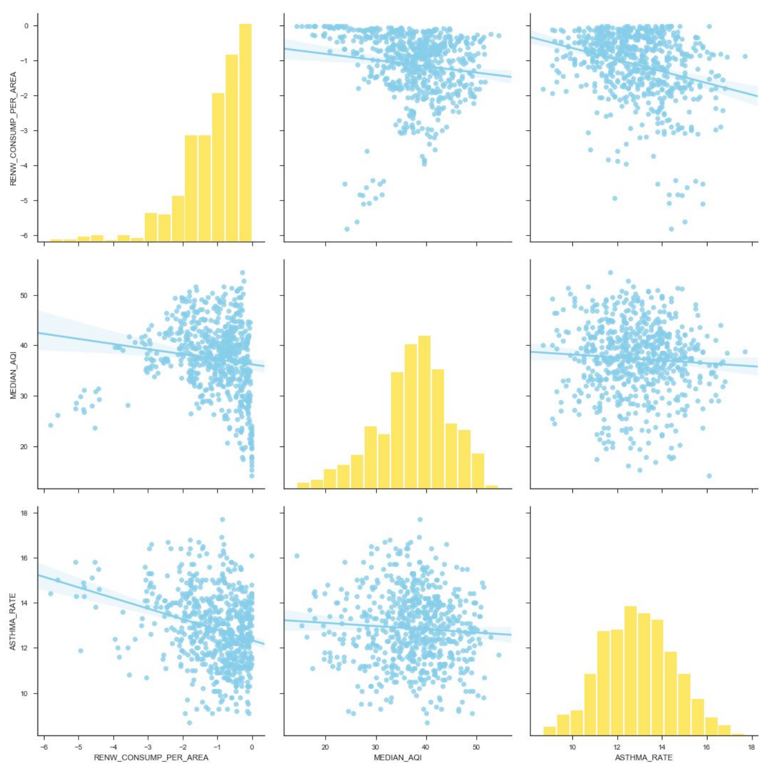

Fig.1.

Then, we wanted to know if asthma rate is correlated with the amount of consumption of non-renewable energy over year. Below is two sets of sample graphs showing the correlation pattern in two states, Wyoming and Nevada, from year 2000 to year 2016.

As can be seen from the graph, the normalized asthma rates and the normalized amounts of consumption of non-renewable energy showed similar pattern in two states.

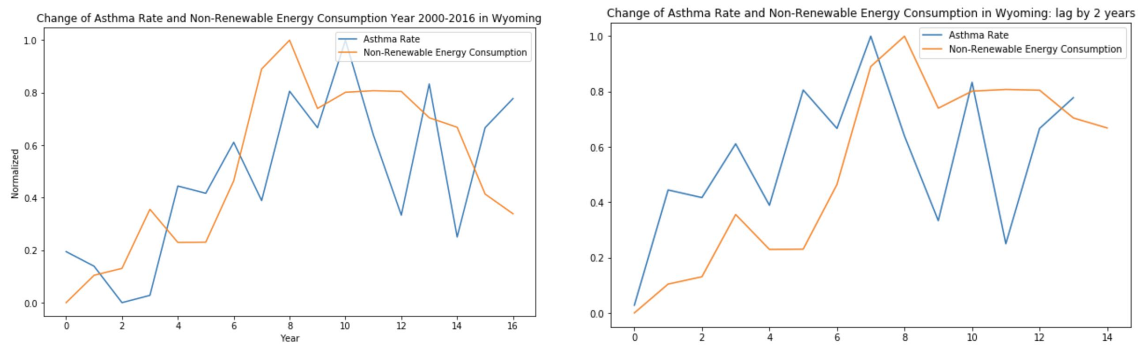

Fig.2.

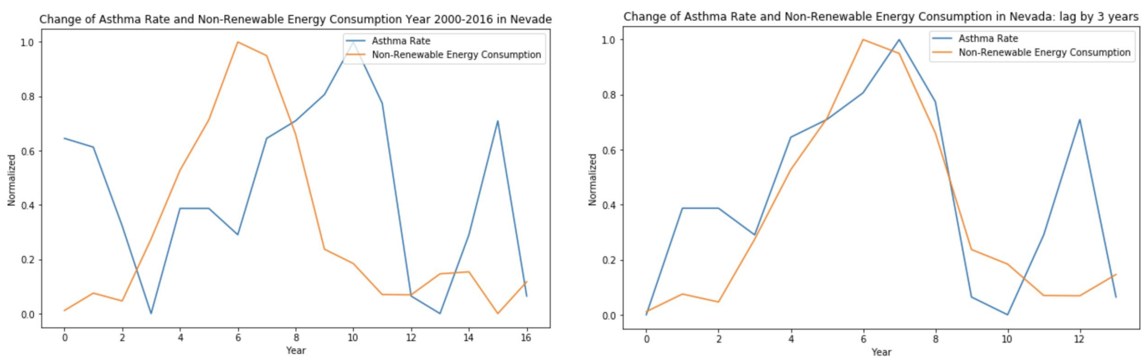

Fig.3.

After we saw these similar patterns, we used Granger Causality Test to explore if there is any causal relationship between asthma rates and amount of consumption of non-renewable energy. The results of the Granger Causality Test in these states are significant. And the lag is usually delayed by 2 to 3 years. In other words, the effect of the amount of non-renewable energy consumption in one state will be seen on its residents in about 2 to 3 years in a form of occurrence of asthma diseases.

During the EDA process, we found that there exist some relationships between asthma rate, amount of renewable/ non-renewable energy consumed, and the air quality, which motivated us to pursue further by starting our modeling part to validate our hypothesis.

Modeling

Methods

Clustering: air quality has influence on air component

To test causality between non-renewable energy consumption and asthma rate in a certain state, we use Granger Causality Test

Causality

- Renewable energy on asthma health
  We implement the panel data in considering the cross-sectional heterogeneity as well as the time dimension.

  Initially, we apply Breusch-Pagan test and find significant sign of heteroskedasticity suggesting panel regression.

Table 1: Table of Breusch-Pagan Test Results

BP_Statistics	14.6064
df	7
p_value	0.0413**

We then implemented the two standard fixed-effect and random-effect panel regression and both found statistically significant negative impacts towards asthma health.

Table 2: Table of Fixed-Effect Regression Results

	ASTHMA_RATE
RENW_CONSUMP_PER_AREA	-1.86***
	(0.14)
TOT_CONSUMP_PER_AREA	0.02
	(0.03)
N	849
R^2	0.20
Adjusted R^2	0.15
F Statistic	99.90*** (df = 2; 797)

*p < .1; **p < .05; ***p < .01

Table 3: Table of Random Effect Regression Results

	ASTHMA_RATE
RENW_CONSUMP_PER_AREA	-1.79***
	(0.17)
TOT_CONSUMP_PER_AREA	0.05
	(0.04)
FLOW_VOLUME	0.06**
	(0.03)
N	649
R^2	0.18
Adjusted R^2	0.12
F Statistic	44.08*** (df = 3; 605)

*p < .1; **p < .05; ***p < .01

Further, we make a comparison of the two models from the Hausman test, where the null hypothesis is that the preferred model is random effects vs. the alternative the fixed effects. It basically tests whether the unique errors are correlated with the regressors, the null hypothesis is they are not. The test result (p-value of 2.2e-16

) shows that fixed effect regression is preferred.

Table 4: Table of Hausman Test Results

Chi_Square_Statistics	12.211
p_value	0.002***

- Air quality on asthma health
  The detailed air quality impact on asthma health is also validated from the regression.

Table 5: Table of Fixed Effect of Renewable Energy towards Asthma Health

	Dependent variable:
	ASTHMA_RATE
K	−13.211**
	(5.270)
WSO2	−0.384***
	(0.033)
NSO4	1.035***
	(0.271)
TNH4	−1.130***
	(0.272)
Observations	649
R^2	0.298
Adjusted R^2	0.246
F Statistic	63.986*** (df = 4; 604)
Note:	*p<0.1; **p<0.05; ***p<0.01

- Air quality on asthma health
  We would like to validate our hypothesis that the air quality has a impact on the asthma health condition. And we also apply the panel regression. The results support our hypothesis with significant gas factors.

Table 6: Table of Fixed Effect of Renewable Energy towards Air Quality

	Dependent variable:
	MAX_AQI
TOT_CONSUMP_PER_AREA	1.076
	(0.957)
RENW_CONSUMP_PER_AREA	−34.402***
	(4.138)
Observations	650
R^2	0.135
Adjusted R^2	0.075
F Statistic	47.189*** (df = 2; 607)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 6: Table of Fixed Effect of Renewable Energy towards Air Quality

	Dependent variable:
	MAX_AQI
TOT_CONSUMP_PER_AREA	1.076
	(0.957)
RENW_CONSUMP_PER_AREA	-34.402^{***}
	(4.138)
Observations	650
R^2	0.135
Adjusted R^2	0.075
F Statistic	47.189^{***} (df = 2; 607)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 5: Table of Fixed Effect of Renewable Energy towards Asthma Health

	Dependent variable:
	ASTHMA_RATE
K	-13.211^{**}
	(5.270)
WSO2	-0.384^{***}
	(0.033)
NSO4	1.035^{***}
	(0.271)
TNH4	-1.130^{***}
	(0.272)
Observations	649
R^2	0.298
Adjusted R^2	0.246
F Statistic	63.986^{***} (df = 4; 604)
Note:	*p<0.1; **p<0.05; ***p<0.01

- Renewable energy on air quality

Table 6: Table of Fixed Effect of Renewable Energy towards Air Quality

	Dependent variable:
	MAX_AQI
TOT_CONSUMP_PER_AREA	1.076
	(0.957)
RENW_CONSUMP_PER_AREA	-34.402^{***}
	(4.138)
Observations	650
R^2	0.135
Adjusted R^2	0.075
F Statistic	47.189^{***} (df = 2; 607)
Note:	*p<0.1; **p<0.05; ***p<0.01

We use LightGBM to select air components features by regressing on medium air quality (AQI) and found out that five air components influence the air quality.

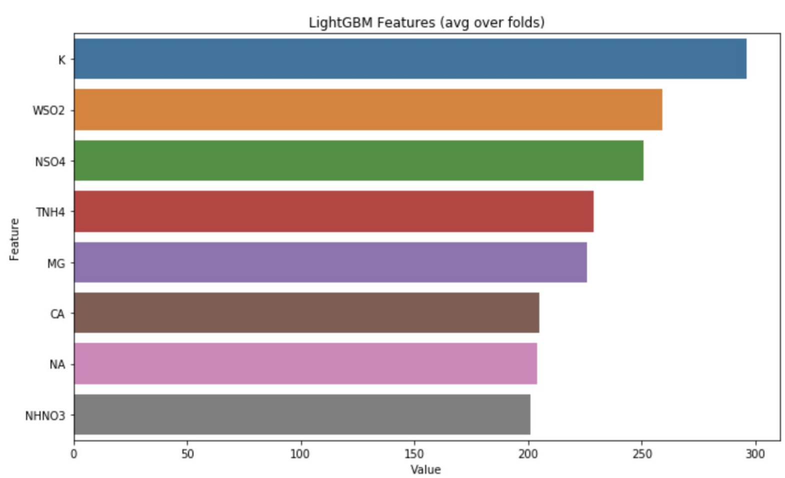

Fig.4.

Then, we use clustering to find what components have major influence on the air quality.

y label: asthma rate, significant feature found via lightGBM. The result is shown below.

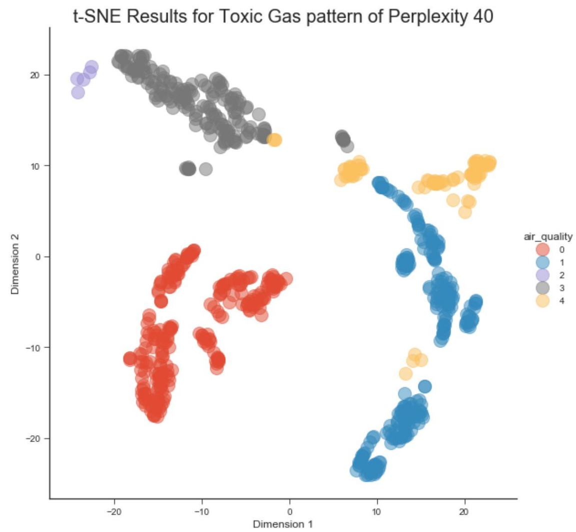

Fig.5.

Then, we produced a radar graph to show the distinction of air components for good air quality and bad air quality

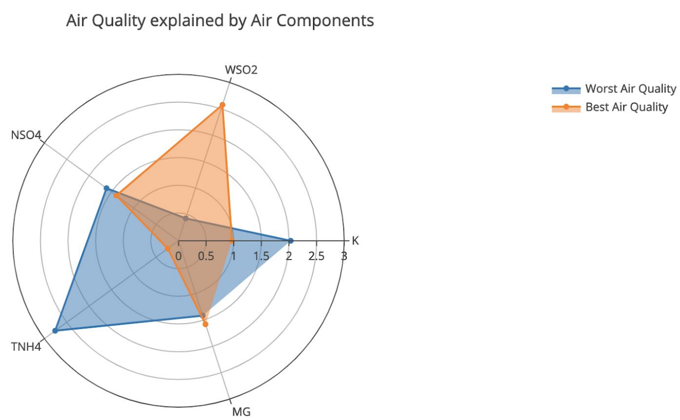

Fig.6.

Conclusions

From the results described above we find that renewable energy source has significant influence on asthma rate. Increase in renewable energy source consumption improves air quality; Reduction in certain air components enhances air quality; Increase in air quality reduces asthma rate; Increase in renewable energy consumption reduces in toxic air components.

Thus, renewable energy source can reduce asthma rate in the United States. Hence, policy makers should focus more on promoting usage of infrastructures or utilities with renewable energy sources, such as wind, solar, and to improve community health and enhance the social welfare.

References:

[1] https://www.cdc.gov/asthma/brfss/2016/brfssdata.htm
[2] https://aqs.epa.gov/aqsweb/airdata/download_files.html
[3]https://en.wikipedia.org/wiki/List_of_U.S._states_and_territories_by_area