

# research proposal

*by* Wenxiao Hu

---

**Submission date:** 04-Nov-2023 08:26PM (UTC+0800)

**Submission ID:** 2217200151

**File name:** 288459\_Wenxiao\_Hu\_research\_proposal\_3156534\_130920292.pdf (166.86K)

**Word count:** 789

**Character count:** 4428

## Research Proposal

Athena Zhang(3036177823), HU Wenxiao (3036174950)

### Background

In light of the dire climate change situation, it is important to analyze energy policy design and the corresponding environmental performance because it allows for the identification and implementation of effective strategies for reducing greenhouse gas emissions, a primary contributor to climate change. Hopefully, this information will encourage the transition towards cleaner and more sustainable energy sources for policy makers, helping to mitigate the environmental impact of conventional energy production methods in a more concrete way. Additionally, such analysis can reveal practical opportunities for improving energy efficiency, which can lead to significant cost savings as well as environmental benefits.

### Research question

Focus: energy policy and performance

Question: how indicative is energy policy design 1. diction and 2. monetary value on actual performance (reduced impact on the environment)

### Data and Methodology:

Analysis discussion:

1. Scope of policies: The jurisdiction of some policies is not the whole country, but some regions of the corresponding country. In our calculations, we classify them into the country, but their influence may not be the same as that of the country. Policies have different scales and scope. This affects the relevance of policy to performance.
2. Policy implementation effectiveness: First of all, the implementation of policies takes time, and the policies represented in the policy data set were first implemented in 2020. The performance data set represents the environmental effect in 2022. The general environment changes relatively slowly. We cannot ensure that the implemented policies will be fully implemented in 2022, which will affect the correlation between budget and performance to a certain extent.
3. Geographical layout of countries: It is impossible to analyze a country's climate change progress by isolating its own efforts as surrounding states could significantly impact other states' environment indirectly. This should be held into account when analyzing the implications of the results.

### FOR ENERGY POLICY DATA ONLY

#### 1. Source

retrieved from dataset: <https://www.energypolicytracker.org/policies-analysis/>  
global energy policies from 2020-22 period

#### 2. Cleaning

- a. Some policies have been proposed and allocated funds, however, they have not been put into use - data should be cleaned so that only the "active" or employed policies are analyzed and hence, the correlation will be based on normative data
- b. The policies whose "Date of entry into force" are NA in the table should be deleted
- c. We also need to unify the expression of the country's name. (For example, for South Korea, one of the two data sets is called "Republic of Korea" and the other is called "South Korea")
- d. use the `"complete. cases()"` code to remove invalid policies with NA in "date of entry into force". Also delete policies in which the "Value committed, USD" column is 0.

#### 3. Text mining + correlation 1

for policy names ('diction'): to determine if any words were significantly correlated with "Value Committed"('monetary value'): and if so, what

*[would need help from the professor. process would be likely to be similar to foreign aids example]*

find keywords for "name of policy", use `'packages(tm)'` to build a corpus for policy name and find the `"word_count()"` for keywords, and use `"sort()"` to arrange the frequency.

energy	investment
119	110
act	infrastructure
4	67
million	gas
66	58
support	fund
58	53
jobs	renewable
52	37
projects	tax
36	35
electricity	new

{problem: there are still keywords that have little relationship with the policy (e.g. new, since, per). Do we need to delete them by manual screening? Will be better to use the energy type as the "fiction"?}

#### 4. data visualization

for one categorical variable (fiction) and one quantitative variable (monetary value), we plan to draw "colored frequency polygon" or "boxplot" to show their relationships.

#### FOR ENVIRONMENTAL PERFORMANCE DATA ONLY

##### 1. Source

retrieved from index: <https://epi.yale.edu/>  
comprehensive summary of countries' environmental performance

##### 2. Extract

There are many indicators that represent performance but only ones related to 'energy' should be selected - i.e. only use indicators under 'Air Quality' and 'Climate Change Mitigation'

##### 3. Cleaning

some raw data only has data up to 2019 so we should check which datasets have updated or data past 2021 and remove the ones that don't

#### TOGETHER

##### 1. Correlation 2

correlate 'fiction' and 'performance'. hypothesis: no correlation  
two possible ways:

- a. using "fiction" (whether keywords of policy names or energy type) (categorical variable) and "performance" (quantitative variable) to draw color histogram and other data visualization chart.
- b. Add up "policies containing the same keyword (there may be duplications because there are multiple keywords in a policy name) or add up the monetary amounts of policies with the same energy type" (quantitative variable) and "performance" (quantitative variable) ) to draw the regression model.

##### 2. Correlation 3

correlate 'monetary value' and 'performance'. hypothesis: positive correlation

#### Division of Labour

Wenxiao	Athena
Clean + manipulate 'energy policy data'	Clean + manipulate 'performance data'
Correlation 2/3	Correlation 2/3

Great job! You have made it crystal clear what research questions you are asking and how you would like to use data science tools to answer these questions. I especially like that you have already performed some preliminary analysis with the data. Keep up the good work :)

# research proposal

---

GRADEMARK REPORT

---

FINAL GRADE

GENERAL COMMENTS

/100

---

PAGE 1

---

**Text Comment.** A

PAGE 2

---

**Text Comment.** Great job! You have made it crystal clear what research questions you are asking and how you would like to use data science tools to answer these questions. I especially like that you have already performed some preliminary analysis with the data. Keep up the good work :)