ISEN660 Final Project

Impacts of Human Activities on Climate Changes – Bitcoin Mining

Devarsh Raval, Sandeep Dasari, Tsz Yi Yang

Aggie Code of Honor

For many years Aggies have followed a Code of Honor, which is stated in this very simple verse: An Aggie does not lie, cheat or steal or tolerate those who do. The Aggie Code of Honor is an effort to unify the aims of all Texas A&M men and women toward a high code of ethics and personal dignity. For most, living under this code will be no problem, as it asks nothing of a person that is beyond reason. It only calls for honesty and integrity, characteristics that Aggies have always exemplified. The Aggie Code of Honor functions as a symbol to all Aggies, promoting understanding and loyalty to truth and confidence in each other.

# Executive Summary

# Introduction

## Global warming vs climate change with Human activity

Is it global warming or climate change has been discussed and appeared in internet searches over decades. Public interest shows that people believe the global temperature is becoming warmer rather than just climate changes over time [1]. Yet the average temperature on the earth's surface has shown an increasing trend since 1900 and has constantly grown after the 1970s. The global temperature has increased by 1 ˚C by 2010 and 1.5 ˚C in 2020 compared to the average surface temperature reported between 1950-1980. This increasing trend has not stopped until today [2]. One major hypothesis of global warming is due to the emission of greenhouse gases such as carbon dioxide, methane, etc. which is often the byproducts of human activities. These activities and associated global warming factors can be traced back to the industrial revolution back in 18 century but these records just came into public eyes since 1965 [3]. Fossil oil, natural gases, electricity consumption, agricultural emissions, and more become important criteria for global warming studies.

# Methods

## Environmental Data

The environmental dataset was published by Berkeley Earth. The newest version was last updated in January 2021 at the time this report was written. This dataset contains a variety of environmental measurements retrieved between 1750 to December 2020 from satellite images to ground weather stations readings across nearly every nation with sampling intervals reported monthly and annually in Celsius respectively. In this study, a fraction of the dataset between 1965 to 2020 absolute land temperature and annual temperature anomaly were used to demonstrate the global warming trend for those countries of interest.

## Human Activity Data

Human activity is considered one of the major factors when discussing the causes of the global warming trend. Human activity can be measured in terms of energy transformations which are associated with population growth, energy consumption, carbon dioxide emissions, or other forms of pollutions. The primary dataset was collected from BP Statistical Review and Our World in Data. This dataset contains energy transformations associated with human activities includes total carbon dioxide emission, electricity generation, primary energy consumption per fuel type, natural resource preservation, and consumption for fossil fuel, coal, gas, nuclear energy, and renewable energy. The earliest record of human activity was reported in 1965 up to 2020 when this report was written which has bottlenecked sampling interval no earlier than 1965 to validate the correlation of human activity impacting climate changes.

## Exploratory Data Analysis & Data Visualization

The primary computational analysis was performed in Python 3.8 with numerical libraries including Numpy and Pandas; statistical libraries including Seaborn and Scikit-learn; data visualization library Matplotlib. Analytical tasks were performed in Linux OS with GPU acceleration option enabled. The dataset (data frame, df) used in this project was integrated from the listed resources and was heavily lifted during the data cleaning process. The raw\_df has 6629 data entities and 10 columns variables so the raw\_df.shape() was (6629, 10). However, the preliminary cleaning lifted off a large chunk of raw\_df which contains missing values, yet the target year-interval was clipped only between 1965 to 2020 and the target locations have narrowed down from 123 countries to 10 countries of interest including 1 global sum. The final dataset (from here and forth it is defined as ***df***) has a size of (560,10).

There are 10 columns of variable which contains “country”, “year”, “population”, “geographical region”, “total CO2 emissions”, “generated electricity”, “primary energy consumption”, “annual land temperature anomalies”, “annual land absolute temperature”, and “annual Bitcoin electricity consumption”. The df entity properties are shown in table 1.

## Random Forest Regression Analysis

Random forest analysis (RFA) is a machine learning technique with supervised dataset based on ensemble learning. RFA is an algorithm that regresses K out of N samples per target variable to generate a decision tree and predicts an output prediction. This process will be performed multiple times across all variables and essentially form multiple branches of decision trees, as indicated in the name, forming a random forest. RFA is commonly used in classification problems and regression problems. In this analysis, RFA regression was performed to validate the correlation between temperature changes among all variables.

Table 1. The list of properties of df entities

|  |  |  |
| --- | --- | --- |
| Variable | Description | Data Type |
| Country | Global sum, Canada, China, Germany, Iran, Ireland, Kazakhstan, Malaysia, Russia Federation, and the USA, | Object |
| Year | 1965 - 2020 | Int64 |
| pop | Population per country between 1965 – 2020, reported in millions resolution | Float64 |
| Region | geographical region across continents: Asia Pacific, Africa, CIS, Europe, Middle East, North America, South & Central America | Object |
| co2\_mtco2 | Total CO2 emissions reported in million tonnes | Float64 |
| elect\_twh | Generated electricity per country reported in terawatt per hour (TWh) | Float64 |
| primary\_ej | Primary energy consumptions by all fuel types reported in exajoules (ej) | Float64 |
| annual\_anomaly | Annual land temperature anomalies reported in Celsius (˚C) | Float64 |
| annual\_temp | Absolute annual land temperature reported in Celsius (˚C) | Float64 |
| annual\_consumption | annual Bitcoin electricity consumption reported in terawatt per hour (TWh) *\*There was typo in the code and raw dataset that Bitcoin electricity was reported annually, not monthly* | Float64 |

# Results & Discussion

## Global trend

Human activities and their intensity are positively associated with the population growth in the region. This trend is shown in figure 1 which in the time being between 1965 and 2020, the population has grown from 3.3 billion in 1965 to almost 8 billion in 2020 globally. The human activity measurements showed similar increases nearly double to triple the initial value in 1965. Numerical records showing 1. CO2 emissions were 11,189.7 and 32284.1 million tonnes from 1965 to 2020 respectively; 2. Primary energy consumption was 155.22 ej in 1965 and increased to 556.63 ej in 2020; 3. Electricity consumption was reported at 9880.0 TWh in 1985 and 26823.2 TWh in 2020; 4. Electricity consumption for Bitcoin mining was first estimated in 2010 (negligible small value) and soon bloom to 0.95 TWh in 2013 and 66.91 TWh in 2020; 5. Temperature anomaly is calibrated value based on the differences between measured absolute temperature and baseline temperature averaged from 1951-1980. Temperature anomalies were showing relatively stabled flat line before 1975 and have linearly increased ever since. The global temperature anomaly has increased over 1˚C around 2010 and almost 1.5˚C in 2020 compared to baseline temperature. Note that the primary energy consumption has decreased in 2020 due to pandemic whereas it was reported 581.51 ej in 2019 but decreased to 556.63 ej in 2020. Electricity is hard to be stored in large quantity therefore the generated electricity is considerably equivalent to electricity consumption, yet the transmission loss is not discussed in the raw dataset.

The global trend of a variety of variables along the years is plotted in figure 2 where each variable trend was normalized between 0 and 1 to eliminate biasing of the absolute values. The red dashed line shows the trend of the annual temperature anomaly. All the variable trend shows high linear similarity to annual temperature anomaly along the years which indicates human activities might have high impacts on the global warming issue.

## Breakdown variable trends by country of interest

Due to the time constraints, only 9 individual countries and 1 global sum were picked up for discussion in terms of human activity in particular locations and their average temperature anomalies. The targeting countries were Canada, China, Germany, Iran, Ireland, Kazakhstan, Malaysia, Russia Federation, and the USA which were reported as the top 9 Bitcoin mining contributing countries. The country population trend (figure 3) shows that the population in Mainland China has grown from 724 million to 1.44 billion; the USA has grown from 200 million to 331 million, the population curve is relatively flat for the rest of the countries.

The total CO2 emissions trend (figure 4) shows that most countries had CO2 emission growth from 1965 and peaked out around 1990 then maintained flat or showing decreasing trend. Russian Federation peaked CO2 emissions in the early 1980s then cut significantly in the mid-1980s. China has been the only country showing exponential CO2 emission growth after 2000 and still climbing. China has reportedly weighted the biggest CO2 emission country among all countries at 9900 million tonnes in 2020 following the USA was reported with 4457 million tonnes CO2 emission.

The primary energy consumption (figure 5) and electricity consumption (figure 6) shows a similar trend where China has exponential growth in utilizing primary energy as well as electricity usage after 2000. The USA was long-time being the largest energy-consuming country yet China became the largest single country as an energy consumer around 2010. USA consumed 95.1 ej in 2000 and 87.8 ej in 2020 whereas China consumed 42.5 ej in 2010 and 145.5 ej in 2020 for primary energy consumption. The USA generated 4052 TWh in 2000 and 4287 TWh in 2020 whereas China generated 1356 TWh in 2000 and 7779 TWh in 2020 for electricity consumption.

Annual temperature anomaly is a calibrated temperature indicator against the baseline temperature (1951-1980 average temperature). Negative anomaly means cooler temperature and positive anomaly refers to the warmer temperature. The temperature anomaly trend (figure 7) shows 3 climate change regions where the temperatures were generally cooler (blue color) before 1975 and the temperature started bouncing up and down between 1975 to mid-1980s and the overall temperature trend has then increased significantly (orange color).

## Random forest analysis

In this section, random forest regression analysis (RFRA) was performed using scikit-learn library. The training and test data split ratio was 75% and 25% with 420 training labels and 140 test labels, respectively. Each iteration will take 20 estimators from 45 samples per variable at a time and a smaller decision tree was generated per iteration per variable until the samples and variables were exhausted. The RFRA algorithm repeated adding regression trees and formed 8 layers deep of tree nodes with nearly 256 leaf nodes in this analysis. A small portion of the forest is displayed in figure 8 for demonstration purposes, the full diagram of RFRA is attached as png file separately from this report. The regression accuracies were calculated with mean absolute error (MAE) at 0.01766%, mean squared error (MSE) at 0.00378%, and root mean squared error (RMSE) at 0.06148%.

The top event is the annual temperature anomaly was smaller than 0.6C where the left direction is True and right direction is False. When the node dive down the pivot event will appear in the intermediate node. 40 nodes were associated with the country of the region, 24 nodes were associated with the time in years, 26 nodes were associated with population and CO2 emissions, 11 nodes were associated with primary energy consumption and electricity consumption. At the very bottom node presents the lead node, which shows the regression of the predicted annual temperature anomaly. This preliminary result shows using RFRE for computing probabilities of human activities can affect global warming can be promising. However, due to the bias in country selection, this result may not reflect the whole dynamic worldwide.

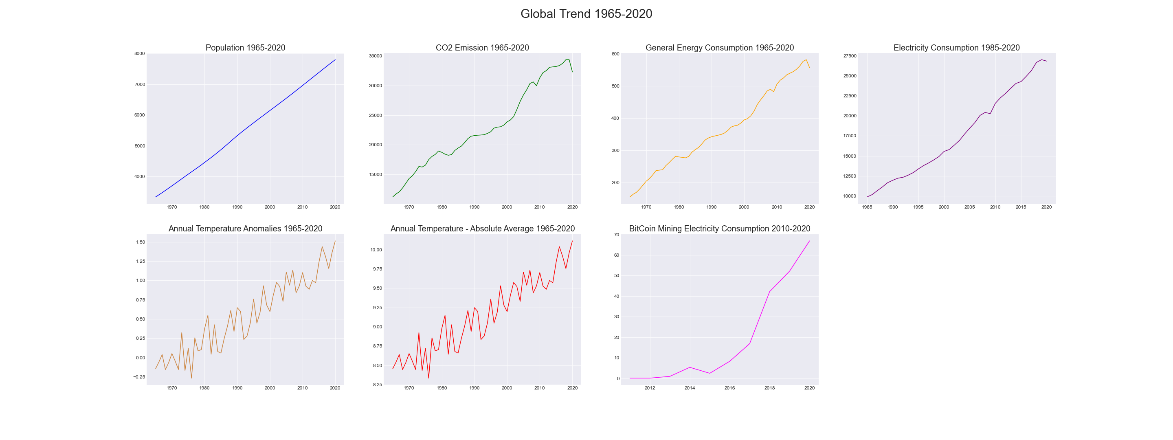


Figure 1. The unscaled global trend of the world population (top left), CO2 emissions, primary energy consumption, electricity consumption, annual temperature anomaly, annual absolute temperature, and Bitcoin electricity consumption (data available from 2019-2020, bottom left) showing linear growth along with the population growth.

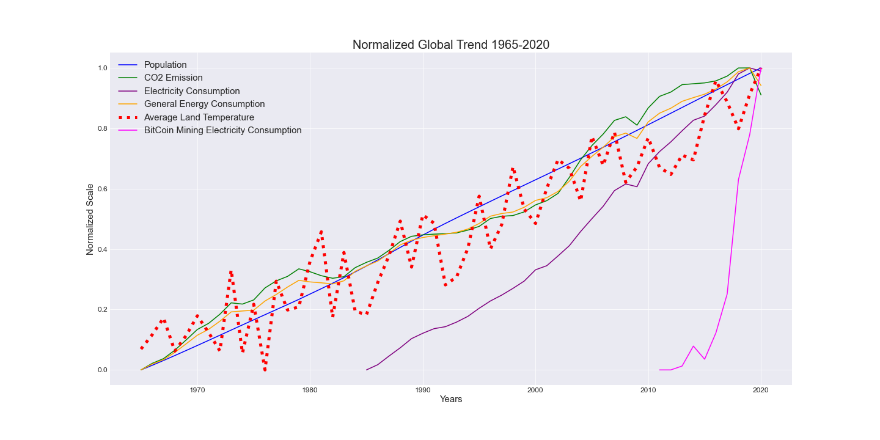


Figure 2. Normalized variable trends against year between 1965 to 2020. The red dashed line shows the trend of the annual temperature anomaly. All the variable trend shows high linear similarity to annual temperature anomaly along the years.

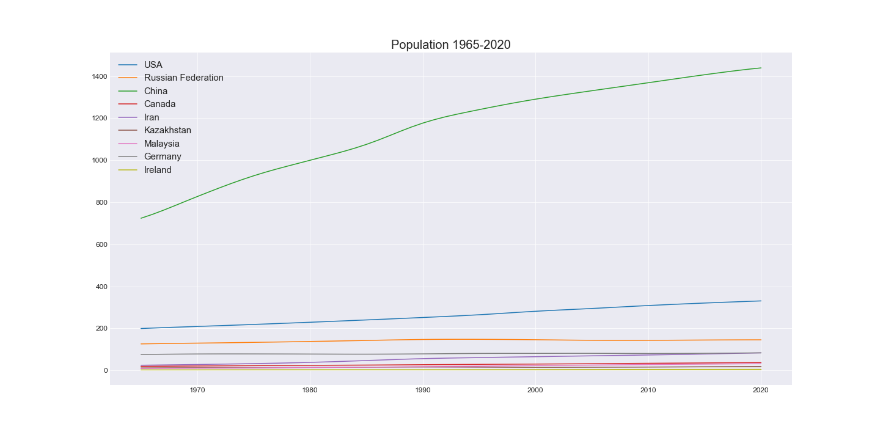


Figure 3. Population trend by country of interest by year.

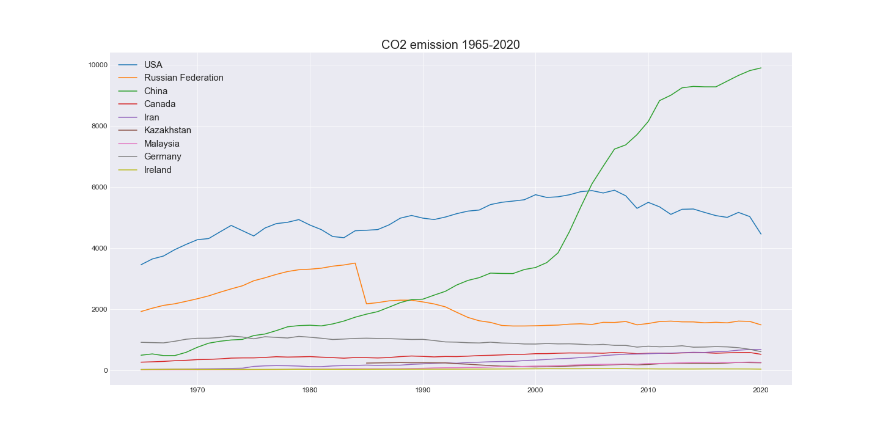


Figure 4. CO2 emission trend by country of interest by year.

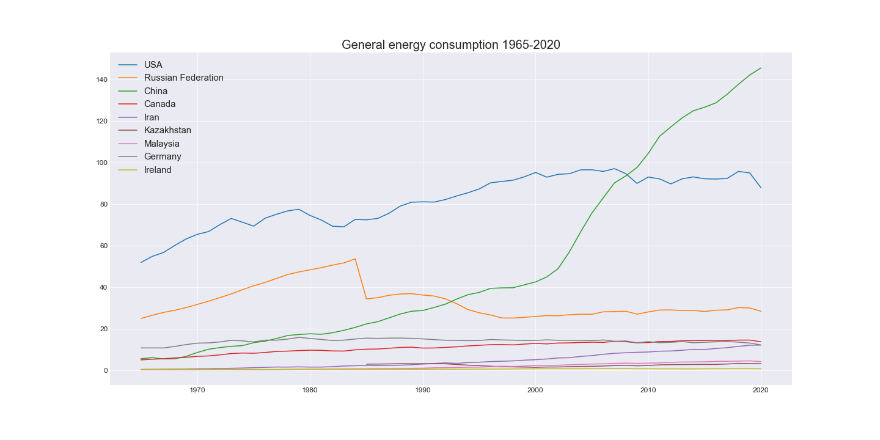


Figure 5. Primary energy consumption trend by country of interest by year.

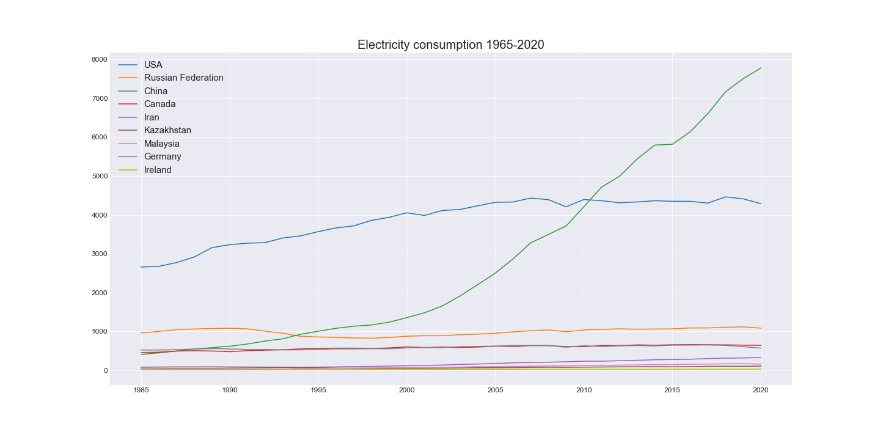


Figure 6. Electricity consumption trend by country of interest by year.

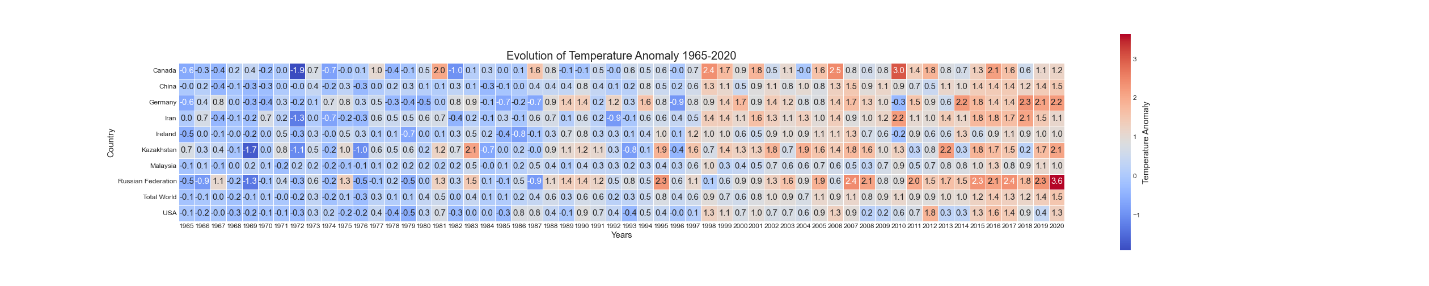


Figure 7. Temperature anomaly trend by country of interest by year. Negative anomaly showing more blue and positive anomaly showing more orange.

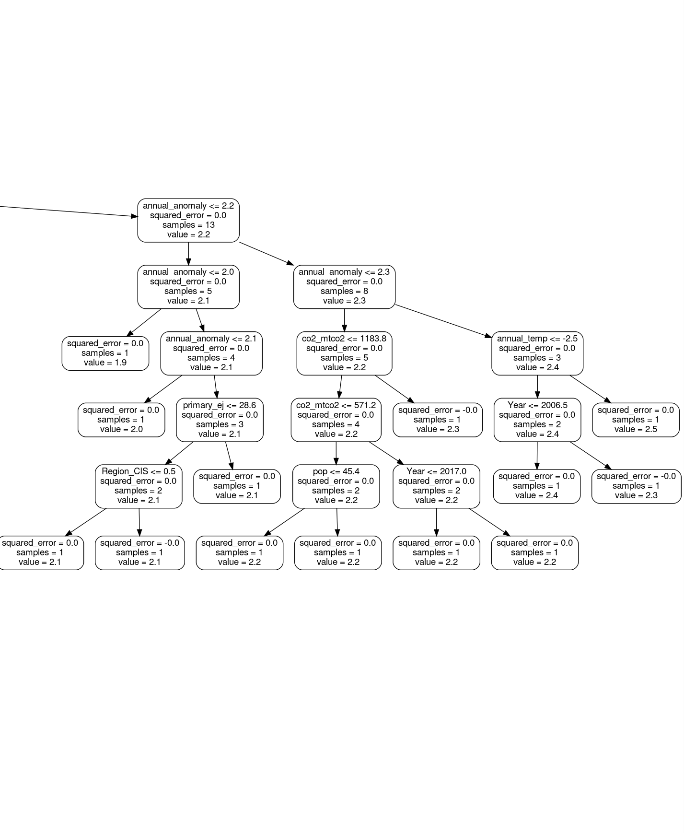


figure 8. This graph demonstrates only a partial branch of the entire random forest.

# Conclusions

The land temperature is showing an increasing trend globally and per country of interest, as well as human activities and their byproducts, also showing an increasing trend. The growth curve of population, CO2 emission, primary energy, and electricity consumption match the increases of temperature trend. RFRA is potentially useful to be used in computing probabilities of global warming events against human activity variables. However, a comprehensive dataset containing more data points from different countries is essential for computing worldwide dynamics for global warming RFRA for future study.

# List of Abbreviations

df: data frame, the cleaned dataset used for data visualizations, and random forest analysis.

˚C: temperature unit in Celsius

Ej: energy consumption unit in exajoules

TWh: electricity consumption unit in terawatt per hour

RFA: random forest analysis

RFRA: random forest regression analysis

MAE: mean absolute error

MSE: mean squared error

RMSE: root mean squared error

# References

|  |  |
| --- | --- |
| [1] | M. Y. D. J. Y. K. a. G.-J. J. Lineman, "Talking about climate change and global warming.," *Plos one,* vol. 10, no. 9, 2015. |
| [2] | R. R. M. R. J. S. P. A. R. J. W. J. C. C. W. a. S. M. Rohde, "Berkeley earth temperature averaging process.," *Geoinformatics & Geostatistics: An Overview 1,* vol. 2, pp. 1-13, 2013. |
| [3] | V. Buyanov, "BP: statistical review of world energy 2011," *Economic Policy,* vol. 4, pp. 38-55, 2011. |

# Appendices