Exploring the Correlation Between Global Trends in Agriculture Production and Energy Usage: Local vs. Global Perspectives

Xinyang Gao

ECO225: Big Data Tools for Economists

Prof. Khazra

April.11, 2023

Introduction

How have the average temperatures changes varied over time across different areas and the entire surface of the Earth, including land and ocean? Does the Earth experience warming trends in only certain cities or countries or across the entire surface? Does energy usage and agriculture production correlated with the temperature change? This project aimed to answer these questions. The first part of the project studied the variation in average temperatures, temperature changes, maximum temperature changes, and standard deviation of temperature across different cities and the entire surface of the Earth. The second part focused on analyzing temperature changes and their potential factors that correlated with the temperature like carbon emissions and fossil fuel consumption across different countries and different continents. The third project find more new variables and dataset, such as forest area, agriculture production, energy consumption and generation, etc across different countries of different years, and visualize them to seek the relationship between the global warming and these factors. The final part of the project run the OLS regression and regression tree with dependent variable such as average temperature or agriculture production and independent variables such as forest area, threatened species, different type energy usage, greenhouse gas emissions, and agricultural productivity to examine the correlation between the dependent variables and independent variables.

Description of finding

From first part of the project, it find that average temperature data for entire surface for both land and land plus ocean has increased roughly 1.5 °C during last 150 years. The temperature of entire surface is less scatter meaning the variation between the coldest and hottest month are smaller and smaller during the last 150 years. The second part of the project visualize the temperature trend of every countries for last 150 years and find that the temperature of almost each of them has risen for 1 to 2 °C. These answer that the global warming trend for both global and local area are increasing. Combine the third part and forth part of the project, find that the average temperature is positively correlated with fossil fuel energy consumption per capita and negatively correlated with crops production, electricity consumption per capita, forest area, and renewables energy consumption per capita.

Compare with other paper

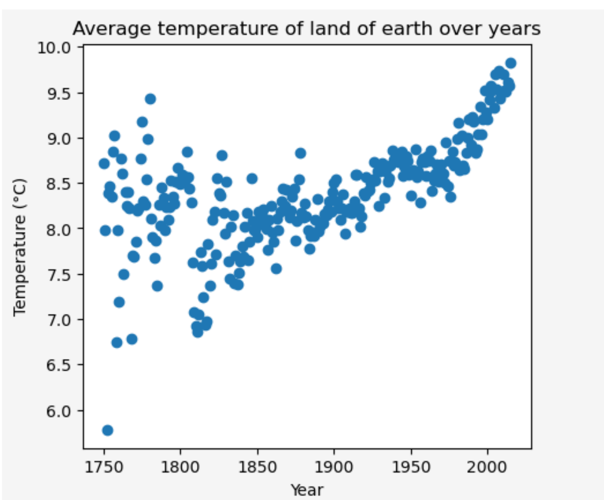
In the report *"Summary for Policymakers — Global Warming of 1.5 ºC"* by Intergovernmental Panel on Climate Change, it has a similar research question with this project. The report also has a consistent finding with ours that the average global surface temperature had risen 1.5 ºC, and they predict how the temperature will change when human reduce the greenhouse gas emission. While theirs focuses on the global picture, this project scales these scales to the national level. Also, their report discuss how the temperature would decrease if just decrease the carbon emission, while this project will provide detailed interpretation about the correlation between temperature change and carbon emission with control other factors. Moreover, this project will provide new insights by running regression tree which is a machine learning technique which separate the global scope to several subgroups which reduce the bias.

Summary for Introduction

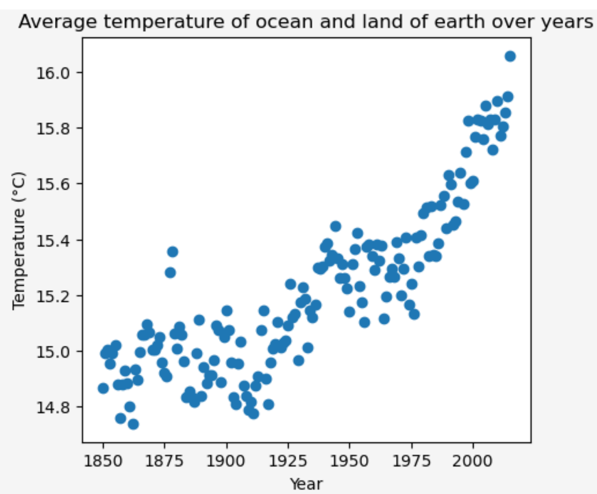
Overall, this project aimed to study the variation in average temperatures, temperature changes, and then examine potential factors that correlated with temperature changes across different areas and the entire surface of the Earth. Let's explore in detail the charts of the specific relevant data mentioned above and our regression results.

Global and Locally Temperature Trend

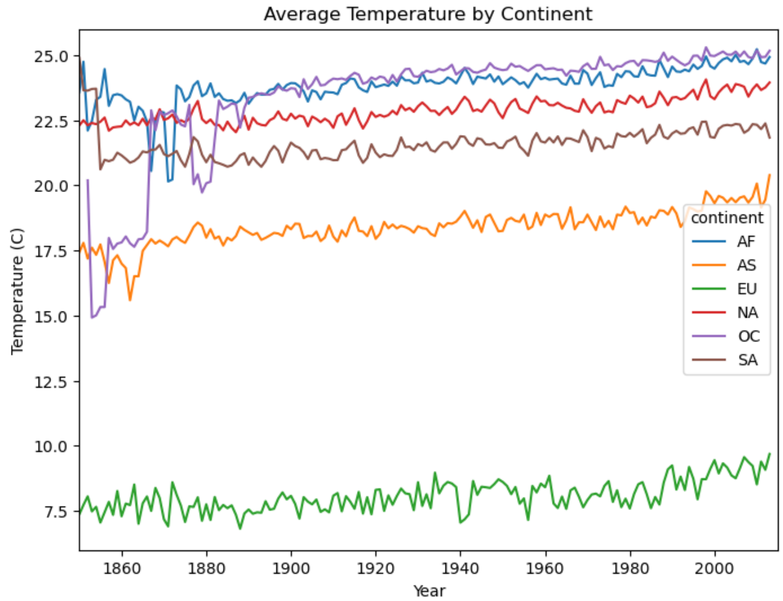
This section visualize the temperature trend for entire surface had group by continents over year. Even the code part provide the temperature trend for all countries, however it will be redundant to present the graph here since it is too much therefore we use temperature data for continents instead.



Graph.1 Average Temperature of Surface of Earth, based on data from Kaggle



Graph.1 continued



Graph.2 Average Temperature of Continents, based on data from Kaggle

The data for Graph.1 and Graph.2 which present the temperature all over the world is from Berkeley Earth. In Graph1, it is observed that the average temperature of both land and ocean has been consistently rising. Over the past 250 years, the average temperature of land has increased by nearly 2°C, while the average temperature of land and ocean combined has increased by approximately 1°C over the last 200 years. Furthermore, the pace of the increase appears to be accelerating. The Graph.2 displays the average temperature for each continent, calculated by averaging the temperature of all countries within the continent for each year. The data indicates a significant and consistent increase in average temperature across all continents including North America, South America, Asia, Africa, Oceania, and Europe over the past 150 years. The increase is relatively steady. Notably, Europe has been the most fluctuate of all the continents. Moreover, we can observe that all continents have experienced an upward trend in temperature since around 1960. Notably, Europe has been the most fluctuate of all the continents. Moreover, we can observe that all continents have experienced an upward trend in temperature since around 1960.

Summary Statistic for Global Surface Temperature

Temperature Difference of Global over years

Year	LandAverageTemperature	LandAndOceanAverageTemperature
1850	13.743	3.950
1851	12.100	3.586
1852	12.136	3.753
2013	11.318	3.386
2014	11.375	3.471
2015	11.170	3.356

Table.1, based on data from Kaggle

Temperature Standard Deviation of Global over years

Year	LandAverageTemperature	LandAndOceanAverageTemperature
1850	4.726851	1.369906
1851	4.652832	1.339856
1852	4.553437	1.386375
2013	4.306305	1.274133
2014	4.396015	1.331519
2015	4.120330	1.229084

Table. 2, based on data from Kaggle

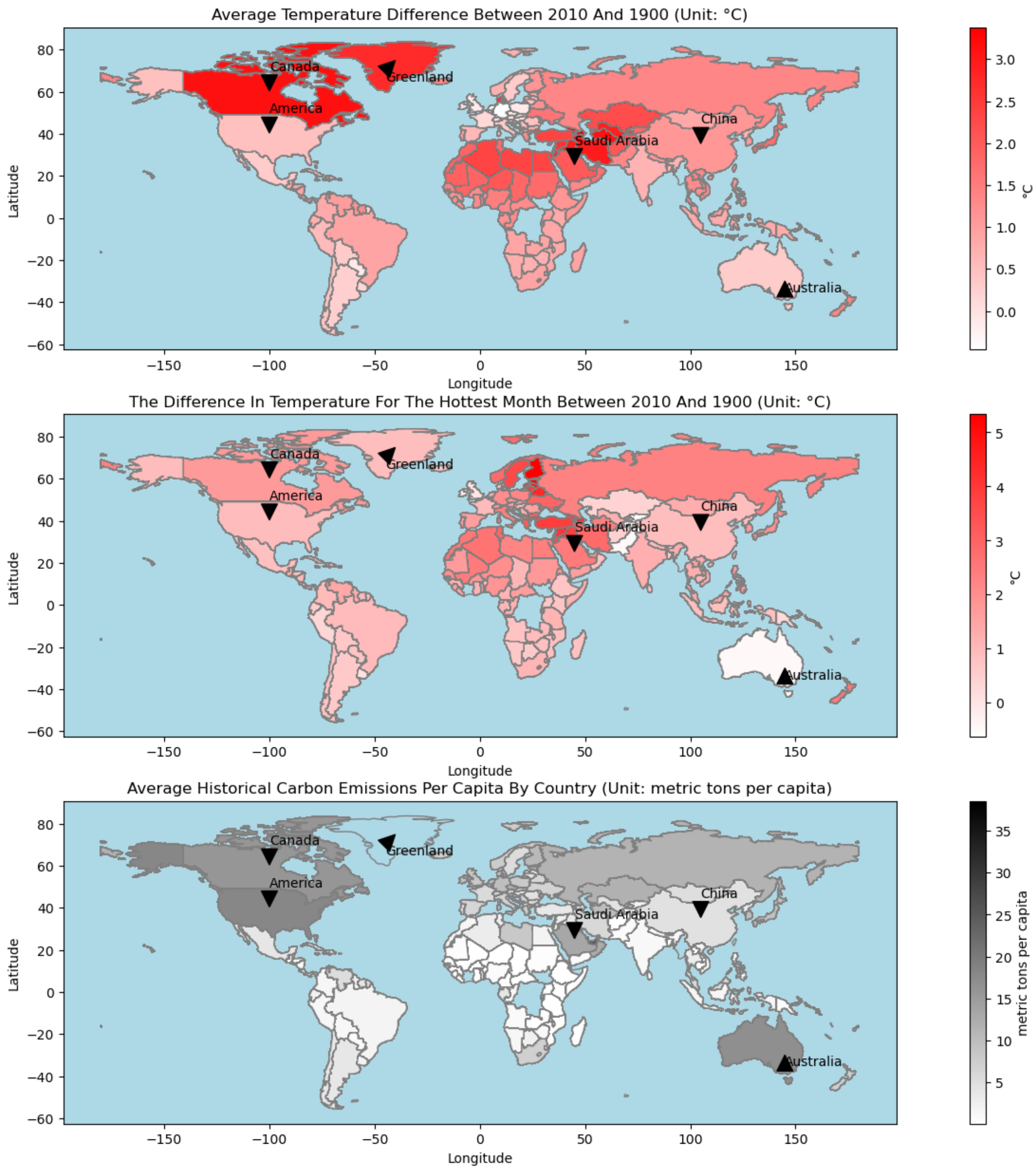
Table.1 present the temperature difference between the coldest and hottest month which is an important indicator of the temperature variability throughout the year. The data shows that the land temperature difference has decreased by 2 degrees Celsius between 1850 and 2015. This indicates a more stable temperature range throughout the year. The decrease in temperature difference for the land plus ocean column is less pronounced but still shows a slight decrease. This could be an indication of a shift towards a more stable climate in recent years. However, it's important to note that these changes are still within the natural variability of the climate and may not necessarily indicate a long-term trend. Further analysis is needed to determine the significance of these changes.

Table.2 presents the standard deviation of temperature measures the variability of temperature within each column over time. In this dataset, it is observed that the standard deviation of temperature for both land and land plus ocean did not change significantly between 1850 and 2015. This means that the spread of temperature values around the mean remained relatively stable over time, suggesting that there was no significant shift in the overall variability of temperature on land or land plus ocean globally.

The analysis of the temperature dataset reveals consistent evidence of global warming over the past century. The increase in average and maximum temperatures, as well as the decrease in temperature difference, suggests a shift towards a more stable and hot climate. However it is still unclear how different region affect by global warming without a map visualization. Next section will show the how different countries were affected by the global warming and the potential factors that correlated it.

Maps for Local Temperature Changes and Carbon Emission per Capita

At this point, it is evident to answer the first question that the global warming happened both locally and globally. However, it remains unclear which regions are getting hotter and which factors are correlated with temperature changes. In this section, some maps are provided to visualize the temperature changes, carbon emission per capita, fossil fuel energy consumption per capita, and electricity generation in different regions. It is important to examine various factors that contribute to climate change. The maps provide crucial insights into the impact of human activities on the environment. The average historical carbon emissions of different countries will give us an idea of the contribution of each country. Finally, exploring the electricity generation and fossil fuel usage in 2018 for different countries will help us to understand the current state of affairs and identify areas where immediate action is needed. By



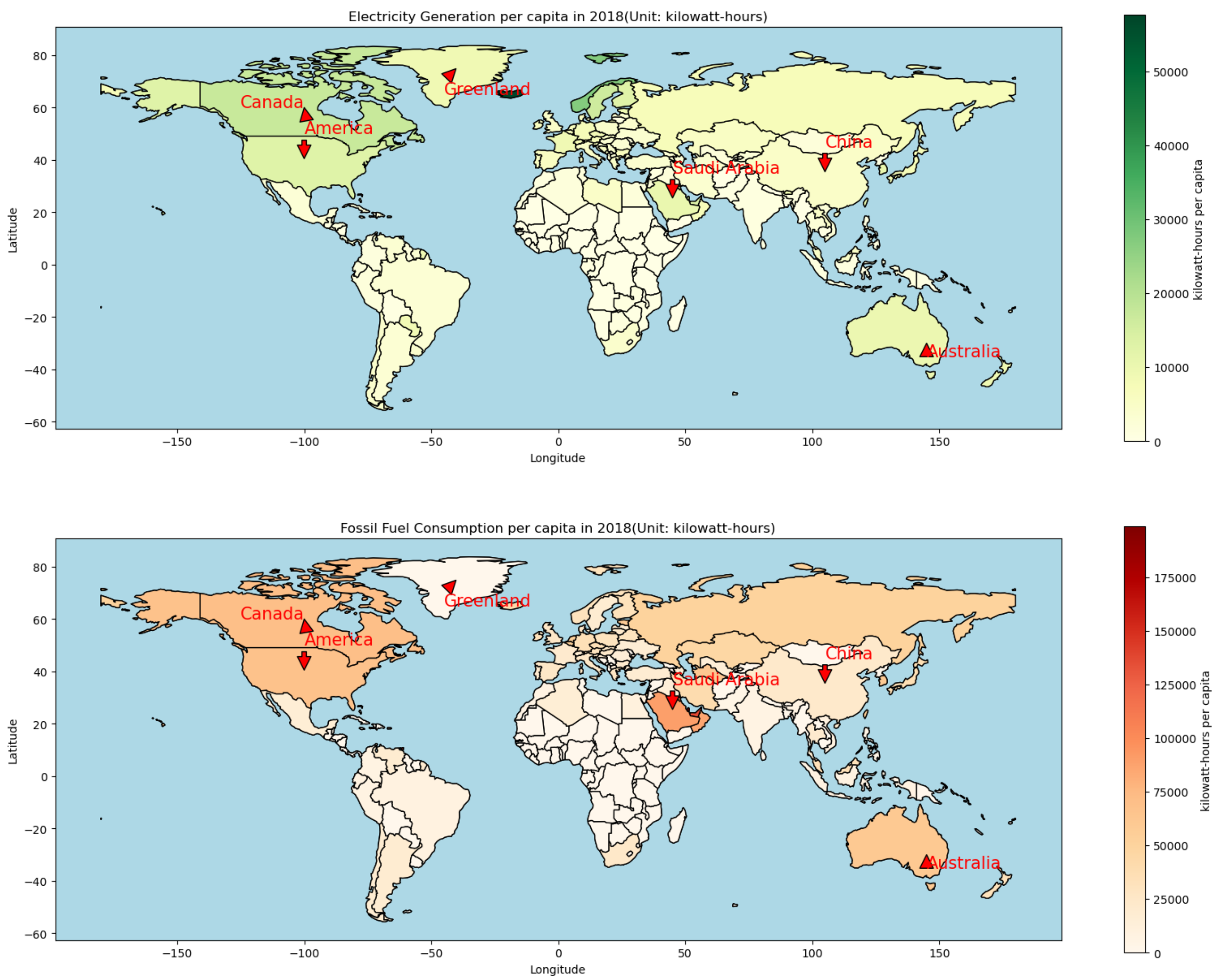
Map.1 Maps for Average Temperature Difference and Hottest Month Temperature Difference between 2010 and 1900, Highest Carbon Emission per Capita, based on data from Kaggle and World Bank

considering these various aspects, we will be able to draw a more comprehensive conclusion about the global warming trends occurring across the globe. The data for CO2 emissions per capita are from Our World in Data

Map.1 provide visual representations of important climate and environmental data for various countries around the world. The first map shows the average temperature difference between 2010 and 1900 in degrees Celsius, highlighting areas where the greatest increases in temperature have occurred over the past century. The second map displays the difference in temperature for the hottest month between 2010 and 1900, providing insights into how climate change has affected seasonal temperatures in different regions. Finally, the third map presents the average historical carbon emissions per capita by country in metric tons, offering a glimpse into the varying contributions to global carbon emissions across different nations. The annotations on each map highlight specific countries of interest, such as China, Australia, Saudi Arabia, Greenland, America, and Canada, and offer insights into how these countries are impacted by climate change or contribute to global carbon emissions. The first map shows the average temperature difference between 2010 and 1900 for various countries. It is observed that countries in North America such as Canada and Greenland have experienced a significant temperature change over the years. Similarly, countries in North Africa and West Asia, which are annotated on the map as being around Saudi Arabia, have also experienced a significant temperature change. However, the second map, which shows the difference in temperature for the hottest month between 2010 and 1900, presents a different trend. It is observed that Europe has been most impacted by the change in temperature, as most areas in the region have experienced a big difference for the hottest month between the two time periods. It present a week correlation with average temperature map. Some countries in Europe are outliers that even they don't have significant change with average temperature, but their hottest month temperature increases significantly. The third map shows the average historical carbon emissions per capita by country, which indicates a strong correlation between carbon emissions and average temperature. Interestingly, Canada in North America and the region around Saudi Arabia, in North Africa and West Asia, have both high carbon emissions per capita and significant temperature changes between 1900 and 2010. This suggests that high carbon emissions are contributing to global warming in these regions. Australia, on the other hand, is an outlier, with high carbon emissions per capita but relatively low temperature change and almost no change in the hottest month between 1900 and 2010. This could be due to a variety of factors, such as differences in climate or energy sources. However, it is clear that efforts to reduce carbon emissions need to be implemented globally to address the issue of global warming.

Maps for Local Fossil Fuel Energy Consumption per Capita and Electricity Generation per Capita

Map.2 depict the electricity generation per capita and fossil fuel consumption per capita of countries worldwide in 2018. The energy consumption data are from Our World in Data. The maps provide valuable insights into the relationship between the two factors and their implications on global warming. One observation from the maps is that countries with higher electricity generation per capita tend to have higher fossil fuel consumption per capita. This finding highlights the high correlation between electricity generation and fossil fuel consumption, which is not surprising given that most countries rely heavily on fossil fuels to generate electricity. Another interesting finding is that China, despite being the highest electricity generator, is not the highest fossil fuel consumer per capita. Other countries such as Australia, Canada, USA, and Saudi Arabia have higher fossil fuel consumption per capita than China. Moreover, the fossil fuel consumption per capita map is highly consistent with the carbon



Map.2 Maps for Electricity Generation per Capita and Fossil Fuel Consumption per Capita in 2018, based on data from Kaggle

emission per capita map, indicating that fossil fuel consumption is the main factor contributing to carbon emissions and global warming. The consistency between the two maps also suggests that reducing fossil fuel consumption is an effective way to combat climate change. In addition, the fossil fuel consumption per capita map correlates highly with the hottest month temperature change map. This correlation provides evidence that fossil fuel consumption is the primary factor driving global warming, as the burning of fossil fuels releases greenhouse gases into the atmosphere, leading to rising temperatures and other climate-related changes. Map.2 provide valuable insights into the relationship between electricity generation, fossil fuel consumption, and global warming. The findings highlight the need to reduce our reliance on fossil fuels and transition to cleaner sources of energy to combat climate change. However to discover the true relationship, it is necessary to run regression and see the correlation between those factors. The regression result is presented in the next section.

Regression Result

The dependent variable selected is average temperature. To ensure that the results are reliable and accurate, the regression controlled for various factors, including GDP, population, and years. With control these factors, it reduce the interference factor and omitted variable bias. The independent variables include forest area, carbon emission per capita, fossil fuel consumption per capita, electricity per capita, renewable energy per capita, and crops production per capita.

When selecting Xs for a regression analysis, it is essential to choose variables that are relevant to the relationship between the independent and dependent variables. According to the theory of climate change, the increase in carbon emissions and the depletion of natural resources have led to a rise in global temperatures, which affects crop yield. Therefore, it is reasonable to include variables related to carbon emissions, renewable energy, and natural resources, such as forest area, in the regression analysis. Furthermore, the agriculture and industry sectors play a crucial role in determining crop yield and average temperature. As such, it is important to include independent variables related to both sectors, such as fossil fuel consumption, electricity consumption, and crops production per capita. These variables will help us understand the contribution of each sector to the dependent variables.

		Dependent variable:AverageTemperature	
	(1)	(2)	(3)
Forest area	-14.434^{***}		-9.680^{***}
	(2.416)		(2.513)
Production per capita (kg)	-0.616^{***}		-0.021
	(0.095)		(0.127)
Year	-0.012	0.123^{***}	-0.001
	(0.030)	(0.031)	(0.043)
carbon emission per capita	-0.670^{***}		-1.140^{***}
	(0.037)		(0.186)
const	43.534	-233.927^{***}	16.145
	(59.318)	(61.151)	(85.795)
fossil_energy_per_capita		94.995^{***}	226.238^{***}
		(11.221)	(27.978)
gdp	-0.827^{***}	-0.491^{***}	-0.312
	(0.187)	(0.187)	(0.204)
per_capita_electricity		-1662.951^{***}	-634.139^{***}
		(127.401)	(195.392)
population	-3.772^{**}	-2.305	-2.470
	(1.672)	(1.698)	(1.769)
renewables_energy_per_capita		319.595^{***}	-310.346^{***}
		(47.063)	(91.246)
Observations	3,079	2,186	1,469
R^2	0.181	0.161	0.189
Adjusted R^2	0.179	0.159	0.184
Residual Std. Error	11.192 (df=3072)	11.720 (df=2179)	11.015 (df=1459)
F Statistic	113.002^{***} (df=6; 3072)	69.774^{***} (df=6; 2179)	37.673^{***} (df=9; 1459)
Note:			$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table.3 Regression Result Table, base on data from Berkeley Earth, World Bank, Our World in Data,and Kaggle

The table.3 shows the results of an OLS regression analysis, where the dependent variable is the AverageTemperature, and the independent variables are forest area, production per capita, carbon emission per capita, fossil energy per capita, renewables energy per capita, per capita electricity, GDP, population, and year. The analysis includes data from all countries, and the variables are measured in different units. The AverageTemperature is in degree celsius. CropsProductionPerCapita is in kilogram per person. CarbonEmissionPerCapita is in metric tons per person. ForestArea is in billion square kilometre. FossilEnergyPerCapita, RenewableEnergyPerCapita, ElectricityPerCapita are in million kilo-watts per person. GDP is in trillion US dollars. Population is in billion. Focus on the specification.3, the R-squared value of the model is 0.189, indicating that the model explains only 18.9% of the variation in the AverageTemperature. This could due to that the percentage change for temperature is too small. The F-statistic of 37.67 indicates that the model is statistically significant, with a p-value of 3.83e-62. The coefficients of the independent variables provide information about their relationship with the dependent variable. The negative coefficient of the forest area (-9.68) indicates that a decrease in forest area is associated with an increase in the average temperature. The negative coefficient of production per capita (-0.021) suggests that an increase in production per capita is associated with a decrease in the average temperature.Suprisingly, the negative coefficient of carbon emission per capita (-1.14) indicates that an increase in carbon emission per capita is associated with an decrease in the average temperature. The positive coefficient of fossil energy per capita (226.24) indicates that an increase in fossil energy per capita is associated with an increase in the average temperature. The negative coefficient of renewables energy per capita (-310.7017) suggests that an increase in renewable energy per capita is associated with a decrease in the average temperature. The negative coefficient of per capita electricity (-634.14) indicates that an increase in per capita electricity is associated with a decrease in the average temperature. The coefficient of GDP (-0.312) is negative but not statistically significant, suggesting that GDP may not be a significant predictor of the average temperature. The negative coefficient of population (-2.47) indicates that an increase in population is associated with an decrease in the average temperature. The coefficient of the year variable is also not statistically significant.

The negative coefficient of forest area suggests that deforestation may be contributing to an increase in the average temperature, which may be happening locally. The positive coefficient of fossil energy per capita suggests that countries relying heavily on fossil fuels may experience higher average temperatures, which may be happening globally. Similarly, the negative coefficient of renewables energy per capita indicates that countries that prioritize the use of renewable energy sources may be able to reduce their average temperature, which may be happening both globally and locally. The negative coefficient of per capita electricity also suggests that reducing energy consumption may lead to a decrease in the average temperature, which may be happening both globally and locally.

Regression Tree

Rerun the regression with the regression tree technique. This allows to divide the data in different subgroups which decrease the prediction error. The regression tree analysis reveals

(Fig.1) interesting insights on the relationship between carbon emissions, forest area, and temperature. The average temperature across all data points is 8.7 degrees Celsius, with the carbon emission per capita serving as the primary threshold that divides the data into two groups. Countries with carbon emission per capita smaller than 2.53 metric tons have an average temperature of 19.6 degrees Celsius, although with only a fifth of the total observations. On the

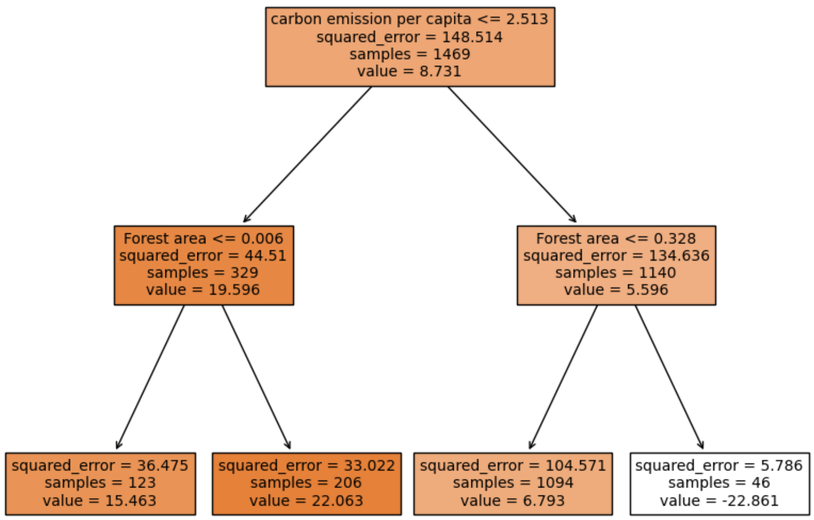


Fig.1, Regression Tree Result, based on data from Berkeley Earth,
World Bank, Our World in Data,and Kaggle

other hand, those with greater carbon emission per capita have an average temperature of 5.6 degrees Celsius, with a much larger number of observations. The second layer of the tree focuses on the right branch, with the most data, revealing that countries with carbon emission per capita greater than 2.53 and forest area smaller than 0.328 billion square kilometres have an average temperature of 6.7 degrees Celsius, with 1094 observations. However, those with forest area greater than that have an average temperature of -22.861 degrees Celsius, albeit with only 46 observations. These finding consistent with the OLS regression that the countries with smaller carbon emission per capita in that year tends to have higher average temperature.

The comparison between OLS regression and Tree regression is an interesting topic, and it highlights some important differences between these two methods. It is evident that the tree regression provides a significant insight that OLS regression fails to do so. The division of data into two groups based on carbon emission per capita is a crucial factor in understanding the relationship between temperature and carbon emission. The coefficient of carbon emission per capita in OLS regression is negative, which is an illusion that more carbon emission "cause" the average temperature lower. However, in regression trees, it provides insight into the relative higher carbon emission per capita area with lower temperature, suggesting that one possible reason is the colder area needing more energy to generate heat and keep warm. This shows that tree regression can differentiate groups based on temperature and carbon emission per capita more effectively than OLS regression. Furthermore, the regression tree further divides the forest area into four subgroups, which are colder area with more forest, colder area with less forest, hot area with more forest, and hotter area with less forest. This is an essential indicator to estimate the area's climate and helps control the factors that affect the average temperature, such as fossil fuel energy, electricity, etc. However, in OLS regression, it is difficult to separate these groups unless we manually add dummy variables, which is not an easy task as selecting the threshold can be tricky.

Data Source

In this project the temperature data comes from Berkeley Earth. CO2 emissions per capita, greenhouse gas emissions, forest area and biodiversity data from world Bank, and energy consumption from Our World in Data, and agriculture data form Kaggle. These data records of the entire surface of the Earth, spanning all countries. By analyzing these records, we can observe the changes in global and local temperatures over time and potential factor related and affected by it. This will give us a better understanding of how average temperature is changing globally and locally and their impact to different areas.

Conclusion

After all analysis we finally can answer the question. How have the average temperatures changes varied over time across different areas and the entire surface of the Earth, including land and ocean? Does the Earth experience warming trends in only certain cities or countries or across the entire surface? Does energy usage and agriculture production correlated with the temperature change?

Even the result from this project may not perfect but the analysis of the temperature dataset provides compelling evidence that global warming is a real phenomenon. The increase in average and maximum temperatures, coupled with a decrease in temperature difference, indicates a shift towards a more stable climate. While the standard deviation of temperature remains relatively stable, the observed increase in frequency of high temperatures suggests a trend towards a warmer climate across the globe. Furthermore, the analysis of average and hottest month temperatures across various continents reveals a clear and consistent upward trend. From

North America to South America, Asia, Africa, Oceania, and Europe, all continents have experienced a steady increase in temperatures over the past 1.5 centuries. This trend is particularly evident after 1960, with Asia and Europe experiencing a significant increase in temperature.

The Analysis of the maps generated from the energy consumption data for 2018 has revealed some interesting findings. The two maps, per capita electricity generation and per capita fossil fuel consumption, are highly correlated, indicating that countries with high electricity generation also tend to have high fossil fuel consumption. The evidence strongly suggests that fossil fuel consumption is the primary factor contributing to global warming, and its effects are not limited to certain areas but are occurring globally. The consistency of the per capita fossil fuel consumption map with the per capita carbon emission and temperature change maps further supports this conclusion.

The regression model Specification3 of Table.3 is an effective tool for analyzing the impact of human activity and climate on agriculture worldwide. The coefficients and their standard errors provide insight into the direction and magnitude of the relationships between the variables and yield. The model includes variables related to temperature, human activity, and economic development, providing evidence for investigating the impact of global warming and human activity on crop yields. However, it is important to note that the coefficients alone cannot answer the question of whether global warming is happening globally or locally, and more evidence needs to be considered. The comparison between OLS regression and Tree regression highlights the benefits of using tree regression to differentiate groups based on temperature and carbon emission per capita more effectively than OLS regression. Overall, Specification3 can be a valuable tool for investigating the relationship between climate, human activity, and agriculture.

The analysis of various datasets provides compelling evidence of the reality and severity of global warming, driven primarily by human activities such as burning of fossil fuels and deforestation. The increase in average and maximum temperatures, loss of forest area, and impact on agriculture are just a few of the many consequences of this phenomenon. The correlation between per capita fossil fuel consumption and per capita electricity generation highlights the need for a transition to cleaner energy sources. The use of regression models such as Specification3 and Tree regression can help investigate the impact of climate change and human activity on agriculture, but further evidence needs to be considered to fully understand the extent of this global issue. It is crucial that we take action to mitigate the effects of global warming and work towards a sustainable future for our planet.

New findings

The project presented is unexpected that the correlation between temperature change and carbon emission is negative while control time, population, and GDP, with interpretation that more carbon emissions don't necessarily cause lower average temperatures. Regression trees reveal that areas with higher per capita carbon emissions may actually have lower temperatures, indicating that colder areas require more energy for heating and staying warm. The regression tree analysis used in this project allows for the data to be divided into different subgroups, providing a more detailed analysis of the relationship between carbon emissions, forest area, and temperature. The findings reveal that countries with carbon emissions per capita lower than 2.53

metric tons had an average temperature of 19.6 degrees Celsius, and the regression tree identified four subgroups based on forest area and temperature, which provides a more nuanced understanding of the factors affecting the average temperature.

Limitation and future work

The coefficients in a statistical model indicate the relationship between independent variables and dependent variables. They can suggest how certain factors may be contributing to changes in temperature or yield over different regions, but they do not necessarily prove the independent variables cause the average temperature increase or the yield decrease. To determine whether those independent variable has cause effect, more various sources of evidence such as temperature records, sea level rise, and changes in precipitation patterns need to be examined and it should take a rigorous experiment design to exclude bias. Thus, while coefficients provide valuable information, they alone cannot be interpreted the cause effect.

To improve the paper in the future, it is suggested to collect more data on agricultural production and yield from different regions, and explore the use of different types of regression models such as other machine learning model to analyze the data. Additionally, the limitations of the analysis such as outdated data and exclusion of certain variables need to be considered. To address these limitations, more recent data and additional variables should be incorporated into the regression models. For example, we should address and classify the country by their climate types, and need exclude some unnecessary outliers. Furthermore, conducting more in-depth analyses of specific regions or crops type could provide more targeted insights into the impact of climate change on agriculture.

Reference:

CO2 emissions (metric tons per capita). World Bank Open Data. (n.d.). Retrieved April 15, 2023, from https://data.worldbank.org/indicator/EN.ATM.CO2E.PC

Earth, B. (2017, May 1). *Climate change: Earth surface temperature data*. Kaggle. Retrieved April 15, 2023, from https://www.kaggle.com/datasets/berkeleyearth/climate-change-earth-surface-temperature-data

Nations, U. (2017, November 16). *Global Food & Agriculture Statistics*. Kaggle. Retrieved April 15, 2023, from https://www.kaggle.com/datasets/unitednations/global-food-agriculture-statistics?resource=download

Summary for policymakers. Global Warming of 1.5 ºC. (n.d.). Retrieved April 15, 2023, from https://www.ipcc.ch/sr15/chapter/spm/

World development indicators: The World Bank. World Development Indicators | The World Bank. (n.d.). Retrieved April 15, 2023, from http://wdi.worldbank.org/table/3.4

World development indicators: The World Bank. World Development Indicators | The World Bank. (n.d.). Retrieved April 15, 2023, from http://wdi.worldbank.org/table/3.9

World Energy Consumption. Kaggle. (n.d.). Retrieved April 15, 2023, from https://www.kaggle.com/datasets/pralabhpoudel/world-energy-consumption