A COMPARATIVE ANALYSIS OF CARBON EMISSIONS FROM COUNTRIES OF VARYING FOSSIL FUEL DEPENDENCE

A Thesis
by
MEGAN K. MACDONALD

Submitted to the Graduate School
Appalachian State University
in partial fulfillment of the requirements for the degree of
MASTER OF SCIENCE IN APPLIED DATA ANALYTICS
(SUSUTAINABILITY CONCENTRATION)

May 2020
Department of Applied Data Analytics,
Walker College of Business

A COMPARATIVE ANALYSIS OF CARBON EMISSIONS FROM COUNTRIES OF VARYING FOSSIL FUEL DEPENDENCE

A Thesis
by
MEGAN K. MACDONALD
May 2020

APPROVED BY:

Tammy Kowalczyk, Ph.D.
Chairperson, Thesis Committee

Lakshmi Iyer, Ph.D.
Member, Thesis Committee

Gregg Marland, Ph.D.
Member, Thesis Committee

Dennis Gilfillan, Ph.D.
Member, Thesis Committee

Timothy Forsyth, Ph.D.
Chairperson, Department of Computer Information Systems

Mike McKenzie, Ph.D.
Dean, Cratis D. Williams School of Graduate Studies

Abstract

A COMPARATIVE ANALYSIS OF CARBON EMISSIONS FROM COUNTRIES OF VARYING FOSSIL FUEL DEPENDENCE

Megan K. MacDonald
B.S., Boston College
M.S., Appalachian State University

Chairperson: Tammy Kowalczyk, Ph.D.

While aggregate global policy exists to combat climate change, there is significantly less understanding about individual countries' unique pathways to reduce emissions of carbon dioxide and limit climate impacts. Research is lacking concerning the ability of small countries, island nations, developing countries, and countries with fossil fuel-based economies to reach global climate targets. While some research exists about individual countries, this study utilizes multiple data analytics techniques to understand their unique emissions trends and what drives those trends. Ten study countries were selected and initially analyzed for trends in their emissions from fossil fuel combustion using quadratic regression. Drivers of emissions were further analyzed using the Kaya Identity and a decomposition analysis using all countries with available data. A cluster analysis was performed on this global set of countries to identify how these Kaya factors -population, wealth, energy intensity, and carbon intensity- could be used to identify similar groups of emitters. The best performing clustering was formed when three clusters were selected; one large cluster of 146 countries, an intermediate cluster of 23 countries mainly driven by growth in wealth (per capita gross domestic product (GDP)), and

four countries mainly driven by decreasing energy intensity (total energy supply per unit of GDP) and growing wealth. This model suggested separate treatment of the 3 “heavy emitters” (China, the United States and India). While the heavy emitters have followed certain pathways to where their emissions are now, this research shows that other countries also have unique drivers and will follow individual emissions pathways. While the cluster modeling showed that some grouping is possible, emissions drivers are still largely specific to a country. As more countries continue to emit more heavily over time, climate targets will need to reflect these differences between countries instead of simple targets currently used by the international community.

Acknowledgments

I would like to express my gratitude to my committee, Dr. Dennis Gilfillan, Dr. Gregg Marland, and Dr Lakshmi Iyer. Without their support and domain knowledge in environmental science, data science, and analytics for good, this effort would not have been possible. They provided me with constant guidance, patience, and insight. I would like to thank my thesis chair, Dr. Tammy Kowalczyk, for her endless support throughout this process. She has inspired me to continue to study sustainability and align my own research with the United Nations Sustainable Development Goals. I would also like to acknowledge the Barnes Student Research Award and the Cratis D. Williams Graduate School for funding a presentation of this research at the Special Interest Group Decision Support and Analytics Symposium in Munich, Germany in 2019.

Table of Contents

Abstract iv

Acknowledgments vi

List of Tables ix

List of Figures x

Chapter 1: Introduction 1

Chapter 2: Review of the Literature 6

2.1 CO_2 and Global Climate Change 6

2.2 Global Climate Change Policy 7

2.3 CO_2 Emissions Inventories 8

2.4 The Kaya Identity 10

2.5 Data Analytics and CO_2 Emissions 11

Chapter 3: Research Methodology 14

3.1 Data Sources 14

3.2 Trend Analysis 15

3.3 Kaya Factors and LMDI Decomposition 19

3.4 Cluster Analysis 20

3.5 Sensitivity Analysis of Clustering 25

Chapter 4: Results and Discussion 27

4.1 Trend Analysis of Total Emissions 27

4.2 Trend Analysis of Fuel Usage 28

4.3 Kaya Factors Breakdown and LMDI Decomposition 32

4.4 Cluster Analysis 36

4.5 Sensitivity Analysis 46

Chapter 5: Conclusions & Future Research 48

References 51

Appendix 57

Vita 62

List of Tables

Table 1: Source information of primary datasets used in this analysis 15

Table 2: Estimates of carbon emissions in from fossil fuel combustion for 10 selected countries in 2005. Note that total emissions also include emissions from cement manufacture and gas flaring. CO_2 emissions are in metric tons of carbon. Per Capita emissions are in metric tons of Carbon per person. 16

Table 3: Descriptions of 10 countries emissions habits for 2015 in thousand metric tons. Note that total emissions also include emissions from cement manufacture and gas flaring. CO_2 emissions are in metric tons of carbon. Per Capita emissions are in MtC. 17

Table 4: Kaya factors broken down for 10 selected countries in 2005. Population is measured in millions of people, Wealth is measured in the 2011 Int. dollar per person, Energy Intensity is in MJ at the 2011 GDP level, and carbon intensity is in thousand metric tons of carbon per unit of energy. 32

Table 5: Kaya factors broken down for 10 selected countries in 2005. Population is measured in millions of people, Wealth is measured in the 2011 Int. dollar per person, Energy Intensity is in MJ at the 2011 GDP level, and carbon intensity is in thousand metric tons of carbon per unit of energy. 33

Table 6: Dunn Index, number of indices that selected k as the best cluster value, Percentage Variance Explained, Average Within cluster Sum of Squares (WSS), Between Cluster Sum of Squares (BSS) and Total Sum of Squares (TSS) for 3, 4 and 5 clusters. The Average WSS is calculated by adding up the WSS values for all clusters and dividing by the number of clusters. WSS, BSS, and TSS are in Euclidean distance. 38

Table 7: Sensitivity Analysis measures are shown such as Dunn Index, membership in Cluster 1, membership in Cluster 2, membership in Cluster 3, Number of negative silhouette widths, Percent variance explained, Between Cluster Sum of Squares (BSS) and Total Sum of Squares (TSS). Percent Variance is calculated by dividing TSS by BSS and multiplying by 100.. 47

List of Figures

Figure 1: Emissions of three selected countries between 2005 and 2015 measured in GtC (Metric Gigatons of Carbon). These are the three largest emitters of CO_2 in 2005 – 2015. The global total is shown as a solid line for perspective. 2

Figure 2: Emissions trends as quadratic functions from 2005 – 2015 emissions data points for 10 select countries based on fuel type emissions and total emissions. Emissions are in GtC for the United States and China, in ktC for Iceland and Marshall Islands, and in MtC for Argentina, Ethiopia, Germany, India, Thailand, and Saudi Arabia. Grey bars indicate the 95% confidence interval for the trend lines. Total emissions are the sum of all other emissions types. 31

Figure 3: LMDI breakdown of Kaya components for 10 selected countries from 2005 - 2015. Panel 1 includes the heavy emitters measured in GtC, Panel 2 includes the moderate emitters in MtC, and Panel 3 includes the low emitters in ktC. The circle represents the change in CO_2 over time for this period. Each Kaya component is represented by how attributable the change in that factor over time is to the change in CO_2 over time. 34

Figure 4: Cluster means are depicted as the black horizontal lines in the boxes describing each cluster. The box structure spans the interquartile range, while the lines indicate the highest and lowest values. Outliers are indicated by dots outside of the box and whisker structure. 41

Figure 5: Clustered countries compared on an axis showing change in wealth for 2005-2015 and an axis of change in carbon intensity over 2005-2015. Cluster 1 is shown in the lightest text, while Cluster 3 is shown in the darkest text. 43

Figure 6: Clustered countries compared on an axis showing change in wealth for 2005-2015 and an axis of change in energy intensity over 2005-2015. Cluster 1 is shown in the lightest text, while Cluster 3 is shown in the darkest text. 44

Figure 7: Clustered countries compared on an axis showing change in wealth for 2005-2015 and an axis of change in population intensity over 2005-2015. Cluster 1 is shown in the lightest text, while Cluster 3 is shown in the darkest text. 46

Chapter 1: Introduction

Emissions of carbon dioxide (CO_2) from the combustion of fossil fuels are related to an increase in the amount of CO_2 in the Earth's atmosphere and a significant shift in the global climate system. It is widely understood that there is a significant and critical need for society to reduce emissions of CO_2, but there has been difficulty in achieving international agreement on how this can be achieved, and how the burden should be distributed. This thesis utilizes data on emissions of CO_2 (represented by the carbon content of the emitted CO_2) at the country level to explore the patterns of emissions, the changes in emissions over time, and the technical as well as economic factors that are driving the changes. While the emissions patterns are based on fuel consumption and reveal how a country is emitting over time, the drivers are based on causes of emissions and reveal why a country is emitting the way they are over time. Carbon emissions have been increasing and shifting over time as nations develop their economies, as populations increase and shift, and energy demand and resources change.

The current leaders in global emissions from fossil fuel combustion and other industrial processes are China, the United States and India. In 1990, the United Sates was the heaviest emitter and the only country emitting over one Gigatons of Carbon (GtC) (Gilfillan et al. 2019). China surpassed the United States as the global leader in emissions in 2005; in 2015 China emitted over 2.5 GtC. Although India has not surpassed the one GtC threshold, this total has grown about 276% since year 1990. For comparison, the global emissions total has grown about 62% since 1990 (Gilfillan et al. 2019, Friedlingstein et al. 2019). In the years after 2015, these three countries accounted for almost 50% of global emissions. This demonstrates that China, the United States, and India are “heavy emitters” globally, and they will be referred to as such in this thesis. **Figure 1** below shows the shift of emissions from 2005-2015, and how the heavy emitters have been accelerating emissions compared to other countries.

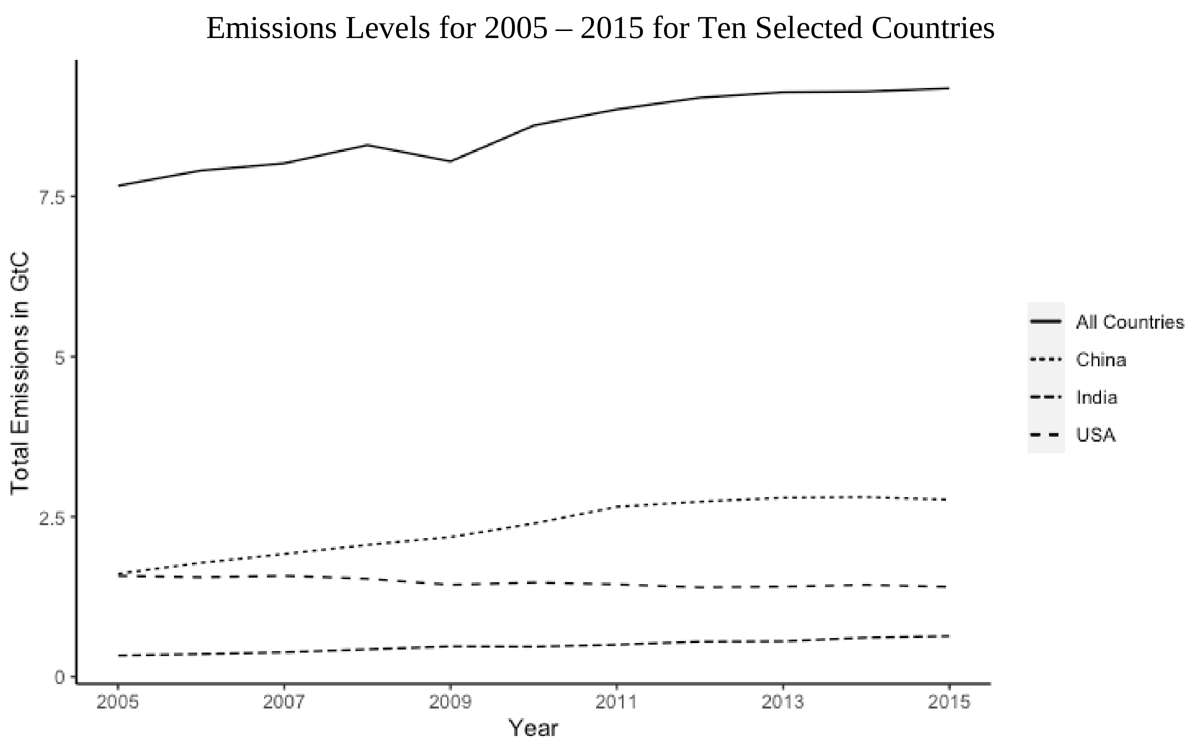


Figure 1: Emissions of three selected countries between 2005 and 2015 measured in GtC (Metric Gigatons of Carbon). These are the three largest emitters of CO_2 in 2005 – 2015. The global total is shown as a solid line for perspective.

Figure 1 shows that the increase in total carbon emissions is connected to the emergence of heavy emitters such as China and India. However, all countries contribute to increasing global CO_2 emissions in the atmosphere, not solely the heavy emitters. Each country has an individual profile of emissions over time, including emissions from different fuel types and different drivers of emissions. These differences reflect the need for different pathways to mitigating emissions globally. Understanding what drives emissions allows for developing effective policies specific to individual nations based on how they can most effectively lower carbon emissions.

Emissions are driven by the types of fuel consumed, and the characteristics of the society. A common way to integrate the demographic, economic, and energy aspects of CO_2 emissions of a society is the Kaya Identity (Kaya, 1990). This is an equation that relates carbon emissions to population, wealth, carbon intensity, and energy intensity. CO_2 emissions can be decomposed from this identity to evaluate how each factor contributes to a country's CO_2 emissions over time.

Data analytics tools are useful for interpreting large datasets that are collected over time, such as carbon emissions data. These tools can derive insights and lead to new conclusions. They are also useful for extremely large datasets. While there are examples in the literature of these analyses being used on single countries, there are far less applications of data analytics on large, global datasets.

This analysis will use data analytics tools to answer the following questions:

1. Beyond the 3 heavy emitters, what are some country-level patterns of emissions over 2005 - 2015
2. What are the common drivers of country wide emissions over 2005 – 2015?
3. Can countries be clustered into groups based on their emissions drivers?

To answer the first question, a regression trend analysis will be used to examine ten specific countries emissions over time. This analysis is done to fit an equation to the curve of total emissions as well as emissions from liquid, solid, and gas fuel consumption. These curves illustrate the patterns of emissions from individual fuels in selected countries. Ten countries were selected to allow for simplicity and a deeper analysis of some individual countries. These 10 countries will be used throughout this research for comparison and discussion of implications of this research.

To answer the second question, the focus is shifted from emissions fuel usage patterns to emissions drivers. This requires the use of the Kaya Identity to break down four drivers of emissions for all countries, and then a log mean Divisia index (LMDI) will be used to break down those factors into index values. These index values show which drivers have changed over time and are most greatly attributable to carbon emissions over time for every country. The LMDI values of the ten study countries will be presented for more in-depth comparison.

Finally, the third question will be answered with a cluster analysis. This is an unsupervised learning analytics tool that will group all of the countries based on their common drivers. Once countries are grouped like this, the collective drivers of emissions can be analyzed, and a global pattern will emerge. While all countries will be used to form the clusters, the ten selected countries as well as some interesting cases will again be examined more in depth for discussion.

Although global approaches to reduce emissions have been researched extensively (Allen et al., 2018), there is less research regarding an individual country's pathway. There is not a universal approach to reducing emissions; a country with high emissions and a large economy will need to approach reducing emissions differently than a developing country, a country with currently declining emissions, or even an island nation with limited natural resources. By studying the different approaches countries can take to reduce emissions, multiple policies can be developed that are tailored to specific types of countries, informing local decision-making to reduce global emissions. As countries are impacted by climate change, providing unique approaches to combating emissions will be the most effective for individual countries.

Chapter 2: Review of the Literature

2.1 CO_2 and Global Climate Change

The Earth's atmosphere is composed of many gases, some of which absorb infrared radiation and cause heat to build up within the atmosphere. These gases are known as greenhouse gases (GHG), and this warming from heat retention in the atmosphere is the greenhouse effect. Notable GHGs include CO_2, water vapor, and methane (Le Treut et al., 2007). Without the natural greenhouse effect, Earth would radiate additional heat to space through longwave radiation, making the surface temperature -19 ^{o}C (Le Treut et al., 2007).

Although the greenhouse effect allows Earth to have a habitable temperature, this greenhouse warming effect has accelerated in the industrial era. Global atmospheric CO_2 levels have risen by 40% since pre-industrial levels (Birdsey et al., 2018). One of the largest causes of this increase is fossil fuel combustion, which emits CO_2 and increases the atmospheric concentration of GHG's (Friedlingstein et al.; 2019, Jackson et al. 2018). The burning of fossil fuels and cement production are responsible for 90% of net CO_2 emissions from anthropogenic actions (Jackson et al., 2018). Increased CO_2 in the atmosphere is stimulating changes in the climate due to increased global surface temperature, potentially leading to: increased frequency and duration of extreme weather events; rising sea level; species extinction and migration; and increased human competition for water and other natural resources (Allen et al., 2018; Birdsey et al., 2018; Le Treut et al., 2007).

2.2 Global Climate Change Policy

Currently, there is a global initiative to address climate change and rising carbon emissions. This global awareness is embodied by the Intergovernmental Panel on Climate Change (IPCC), a group created to collect and assess the scientific information relative to understanding the uncertain future regarding climate change (IPCC, 2019). This assessment body was founded in 1988 by the United Nations Environment Programme as well as the World Meteorological Organization and has produced five assessment reports, along with multiple special reports as requested by the United Nations Framework Convention on Climate Change (UNFCCC) (IPCC, 2013). While the IPCC explores possibilities and scenarios for different levels of responses to climate change, it does not recommend that policymakers take specific actions. However, these assessment reports provide a scientific foundation for countries to utilize when making individual climate policy, and for the UNFCCC to utilize when making global decisions about climate policy at the United Nations Climate Conferences (IPCC, 2013).

The 2018 IPCC special report on Global Warming of 1.5 $_{o}$C concludes that the best climate outcome occurs when surface temperature remains at less than a 1.5 $_{o}$C increase from the pre-industrial global mean temperature (Allen et al., 2018). However, the global mean surface temperature has been increasing at a rate of roughly 0.2 $_{o}$C per decade since 1975; if this trend continues, a projected business-as-usual scenario suggests that the temperature could rise by 2 $_{o}$C to 3 $_{o}$C from the 2000 level by 2100 (Hansen et al. 2006). More current statistical research confirms that the global temperature will most likely rise by 2 $_{o}$C or more by 2100 (Raftery et al. 2017). These temperature increases might seem inconsequential, but they carry large implications for the global climate system. It is estimated that 75% of warm weather extremes and 18% of extreme precipitation events can be attributed to the global temperature rising (Fischer & Knutti, 2015). Due to these potential outcomes, the IPCC recommends the global mean temperature does not deviate more than 2 ^{o}C from the reference mean, and preferably not more than 1.5 ^{o}C (Allen et al., 2018).

Because the mean global temperature change can be observed with a sufficient degree of accuracy and can be easily communicated, the 2 ^{o}C limit on warming has become a global benchmark (Knutti et al., 2016). This goal is simple for policymakers to communicate and the public to understand. However, it is an aggregate number that does not take into consideration the complexity of countries contributing to this goal. It is a much different challenge for a country like China to meet their contribution to this global limit than it is for a country like the Marshall Islands. An aggregate solution such as the IPCC recommends includes all countries but does not take into consideration the very different challenges individual countries will face in order to limit warming by 2 ^{o}C.

2.3 CO_2 Emissions Inventories

This thesis uses the Carbon Dioxide Information Analysis Center (CDIAC) emissions database for analysis. This database contains emissions data by country from 1751 to 2016 and contains information on emissions from solid fuels, liquid fuels, gas fuels, cement manufacture, and gas flaring (Andres et al., 1999; Gilfillan et al., 2019). Andres et al. (1999) estimated CO_2 emissions before 1950 using energy statistics published by Etemad et al. (1991) and Mitchell (1983, 1992, 1993, 1995). Data after 1950 are calculated using statistics published by the United Nations and the methods of Marland and Rotty (1984). These United Nations statistics on production and trade of energy as well as nonenergy use fuels are collected using surveys to member countries and published in the *Energy Statistical Yearbook* (United Nations, 2009). The calculations to estimate carbon emissions are a product of fuel production and consumption data, the fraction of fuel that is oxidized, and the average carbon content of the fuel (Marland & Rotty, 1984). Fossil fuel CO_2 emissions are combined with data from the United States Geological Survey to estimate CO_2 emissions from cement manufacture (Griffin, 1987; US Geological Survey, 2009). Data from the United States Energy Information Administration is used to supplement emissions estimates from gas flaring (Boden et al., 2010).

The CDIAC database is comprised of data from the United Nations Statics Office that reflects reference data guided by the IPCC Guidelines (Andres et al., 2012). Outlined by the IPCC, the reference approach is a top-down method that calculates fossil fuel emissions from a country's energy supply statistics while the sectoral method that is based on energy consumption in the individual sectors (Eggleston et al., 2006)

CDIAC is not the only emissions database. A summary of some of the different carbon emissions databases is described by Andres et al. (2012). These databases include CDIAC, the Emissions Database for Global Atmospheric Research (EDGAR), the International Energy Agency (IEA), the Energy Information Administration of the United States Department of Energy (EIA) and the United Nations Framework Convention on Climate Change (UNFCCC) (Andres et al., 2012). These data sets are based on answers to surveys sent to individual countries, so estimation is required when countries return incomplete answers (Andres et al., 2102). These databases vary in size and focus. A database that considers emissions from only the United States is the Vulcan Project (Gurney et al., 2009). The CDIAC time series of emissions is the base dataset for the global carbon budget and is used to inform IPCC reports for cumulative emissions. It is a comprehensive dataset for the purposes of this research.

2.4 The Kaya Identity

The Kaya Identity is often used to decompose carbons emissions into the factors stimulating changes in those emissions. The Kaya Identity was first described by Professor Yoichi Kaya in 1990 at an IPCC presentation (Kaya, 1990). It can be used to compute what drives changes in CO_2 emissions at a global, national, or even local level (Fan & Lei, 2016; Hatzigerogiou et al., 2008; Pani & Mukhopadhyay, 2010). The Kaya identity is represented by Eq. 1:

$$C = P \times \frac{GDP}{P} \times \frac{E}{GDP} \times \frac{C}{E} \tag{1}$$

Where C is the carbon emissions at a certain time (usually a year), E is the energy supply, GDP is the gross domestic product, and P is population. This identity states that the drivers of change in CO_2 emissions are population (P), wealth $(\frac{GDP}{P})$, energy intensity $(\frac{E}{GDP})$, and carbon intensity $(\frac{C}{E})$.

The Kaya Identity is used extensively in research pertaining to decomposing emissions. An example of this is the decomposition of emissions in Ireland by O' Mahony (2013), which determined that emissions from wealth and population factors were countered by emissions from changes in energy intensity and carbon intensity. A decomposition of the Kaya factors in Ghana by Asumadu-Sarkodie and Owusu (2016) identified that changes in emissions are related most strongly to changes in energy use, then wealth, then population. A decomposition of the Kaya factors in Malaysia by Pui and Othman (2019) also found the need for lower energy intensity over time to counter population and economic growth over time in terms of emissions. Finally, a decomposition of China's emissions from the agricultural sector completed by Li et al. (2014) relates the growth of wealth to increase in

CO_2 emissions using the Kaya Identity. Overall, the Kaya Identity is useful for its ability to serve as a reference point for these factors over time (Albrecht et al., 2002).

2.5 Data Analytics and CO_2 Emissions

The Kaya Identity can be utilized as a scheme, but a decomposition mechanism is needed to understand the drivers of the change over time (Boer & Rodrigues, 2019). There are multiple index decomposition analysis techniques, but a LMDI has been commonly utilized in the literature for breaking down data within the Kaya components (Ang, 2004; Ang & Liu, 2001). This indexing technique decomposes annual data into changes from year to year, which are then summed to an index value that is standardized to the selected base year. Benefits of the LMDI decomposition approach include a complete decomposition, i.e. no residual term like in regression, and consistency in aggregation (O'Mahoney, 2013).

Some successful LMDI analyses applied to a single country are the ones compiled on Greece and Ireland respectively (Hatzigeorgiou et al., 2008; O'Mahoney, 2013). While there is prominent literature of these decomposition techniques being applied to single countries or small groups of countries, there is significantly less research considering a more global scope. One global approach was a large scale LMDI decomposition performed on 114 countries for the years 1992 through 2004 (Pani & Mukhopadhyay, 2010**)**. These papers successfully pair the Kaya Identity as the breakdown scheme and the LMDI as the decomposition analytic technique to compare index values between Kaya components.

The field of statistical learning poses great potential to further characterize complex interactions of drivers of CO_2 emissions. Statistical learning is a framework for analyzing datasets that applies statistics and machine learning. Currently, it has been applied to climate science to reveal climate change impacts in daily weather patterns using ridge regression

(Sippel et al., 2020), and to determine the efficacy of neural networks in forecasting as opposed to traditional physically based process models (Dueben & Bauer, 2018). Regression approaches to trend analysis have been performed on the top 25 emitters to inform projection scenarios (Kone & Hume, 2010).

A statistical learning approach that can be used in understanding CO_2 emissions is cluster analysis. Cluster analysis is an analytics tool that groups items that are more similar into clusters, thus separating items that are dissimilar into different clusters. This approach is useful for grouping countries while still considering the decomposition of individual countries' emissions over time. An example of a global approach using cluster analysis was conducted utilizing K-Means cluster analysis on 87 countries based on six factors to compare carbon emissions and life expectancy (Lamb et al., 2014). A further example of a cluster analysis coupled with an LMDI analysis was used by Liao et al. (2019) to examine the driving factors of terminal electricity generation. Clustering analysis is effective for purposes like these where dividing data in groups can reveal relationships between group members and differences between groups (Yu et al., 2012).

While studies have been completed on the emissions of individual countries (Hatzigeorgiou et al. 2008; O'Mahoney, 2013), research is lacking concerning global drivers of emissions. Given the advancement of data analytics, there are new tools that can be applied to questions like these. One of these is the cluster analysis. Some of the most global research efforts are using LMDI or correlation techniques but are not considering grouping techniques for multiple countries to be compared such as the cluster analysis (Le Quéré et al., 2019; Pani & Mukhopadhyay 2010). The one recent example of a cluster modeling paper applied to all countries (Lamb et al., 2014) does not solely examine carbon emissions. This research will serve as an example of big data techniques applied to a global emissions study to answer whether countries can be clustered into groups based on their emissions drivers.

Ultimately, given the latest IPCC report on curbing the global temperature change to 1.5 $_{o}$C, there is reason to study lowering global emissions to slow temperature change. Mitigating carbon emissions is one of the primary ways to do this and must be done by all countries, not just the heavy emitters. While there is research on how individual countries are emitting, there is significantly less literature considering how all global countries are emitting comparatively. This research will use cluster modeling, a data analytics technique, to address this problem and provide information on how some individual countries as well as groups of countries are emitting.

Chapter 3: Research Methodology

3.1 Data Sources

Data for this analysis were collected from several sources and synthesized. Population data were collected from the United Nations World Population Prospects in 2019 as thousands of individuals (United Nations, 2019). This dataset includes total population (both sexes combined) estimates by country in thousands of people.

GDP data were collected from the World Bank World Development Indicators database (World Bank, 2015). GDP data were collected in constant (adjusted for the price of inflation) PPP (purchasing power parity) in 2011 international dollars. An international dollar is defined by the World Bank as being worth the same amount of goods and services as a United States dollar for that given time (World Bank, 2020).

The energy intensity data were compiled as the ratio of primary energy supply, in Megajoules (MJ) to GDP, and were also collected from the World Bank World Development Indicators database (World Bank, 2015). The carbon emissions data were collected from the CDIAC dataset, which contains emissions estimates in metric tons of carbon for solid fuel consumption, liquid fuel consumption, gas fuel consumption, emissions from cement manufacture, emissions from gas flaring, and per capita emissions at the national and global level for the years 1751 – 2016 (Gilfillan et al., 2019). CO_2 emissions from bunker fuels (used to fuel cargo ships) are also present in this dataset, although were not included in national totals. This information is summarized in the **Table 1** below.

Table 1: Source information of primary datasets used in this analysis

Data	Units	Years	Source
Population	Residents (Thousands)	1950-2100	(United Nations, 2019)
GDP	2011 International Dollars	1960-2019	(World Bank, 2015
Primary Energy	$\frac{\text{MJ}}{\text{2011 International Dollars}}$	1960-2019	(World Bank, 2015)
Carbon Emissions	C (metric tons)	1751 - 2016	(Gilfillan et al. 2019)

Data were collected for all countries from these four sources for the years 2005 to 2015. Countries that did not have sufficient data to complete an analysis for this set of years were removed from the combined dataset. The CDIAC emissions estimates were expressed in metric tons of C, not CO_2, for easier tracking of mass flows in the global carbon cycle (Friedlingstein et al., 2019). To convert from metric tons of carbon to CO_2, the CDIAC estimates can be multiplied by 3.67 to account for the differences in molecular weights (Ryan et al., 2010). For the purpose of this research, these estimates will be left in terms of C, not CO_2.

3.2 Trend Analysis

To address the question of differences in patterns of emissions over time, 10 countries were selected as a study group to identify potential variation in emissions trends. This part of the analysis was conducted to explain how some countries emit. Since these ten countries will be tracked throughout the analysis, understanding the patterns of emissions will be useful for discussion later and deeper understanding about the complete emissions profile of these countries. These 10 countries, called the study countries in this paper, include heavy emitting countries, countries with fossil fuel dependence, nations with high potential burdens from climate change, and other widely divergent natures from different continents. Emissions information about these countries can be found in **Table 2** for year 2005 and **Table 3** for year 2015 below. These countries will be re-examined for discussion and comparison later in this analysis.

Table 2: Estimates of carbon emissions in from fossil fuel combustion for 10 selected countries in 2005. Note that total emissions also include emissions from cement manufacture and gas flaring. CO_2 emissions are in metric tons of carbon. Per Capita emissions are in metric tons of Carbon per person.

Study Country	Total Emissions (2005)	Emissions from solid fuel consumption (2005)	Emissions from liquid fuel consumption (2005)	Emissions from gas fuel consumption (2005)	Per-capita emissions (2005)
Argentina	4.4×10^7	8.92×10^5	1.96×10^7	2.24×10^7	1.13
China	1.61×10^9	1.21×10^9	2.32×10^8	2.32×10^7	1.22
Ethiopia	1.34×10^6	0	1.18×10^6	0	0.02
Germany	2.18×10^8	8.62×10^7	7.86×10^7	4.86×10^7	2.66
Iceland	6.08×10^5	1.03×10^5	4.87×10^5	0	2.06
India	3.33×10^8	2.18×10^8	8.18×10^7	1.38×10^7	0.29
Marshall Islands	3.10×10^4	0	3.10×10^4	0	0.60
Saudi Arabia	1.08×10^8	0	7.01×10^7	3.47×10^7	4.54
Thailand	6.20×10^7	1.12×10^7	3.09×10^7	1.47×10^7	0.95
US	1.58×10^9	5.79×10^8	6.68×10^8	3.17×10^8	5.28

Table 3: Descriptions of 10 countries emissions habits for 2015 in thousand metric tons. Note that total emissions also include emissions from cement manufacture and gas flaring. CO_2 emissions are in metric tons of carbon. Per Capita emissions are in MtC.

Study Country	Total Emissions (2015)	Emissions from solid fuel consumption (2015)	Emissions from liquid fuel consumption (2015)	Emissions from gas fuel consumption (2015)	Per-capita emissions (2015)
Argentina	5.48×10^7	1.43×10^6	2.50×10^7	2.68×10^7	1.26
China	2.77×10^9	1.98×10^9	3.62×10^8	1.00×10^8	1.98
Ethiopia	3.74×10^6	2.98×10^5	2.423×10^6	0	0.04
Germany	1.98×10^8	8.41×10^7	6.92×10^7	6.92×10^7	2.43
Iceland	5.42×10^5	9.60×10^4	4.46×10^5	0	1.64
India	6.35×10^8	4.03×10^8	1.63×10^8	2.70×10^7	0.48
Marshall Islands	3.90×10^4	0	3.90×10^4	0	0.74
Saudi Arabia	2.00×10^8	0	1.29×10^8	6.28×10^7	6.35
Thailand	7.56×10^7	1.76×10^7	2.94×10^7	2.36×10^7	1.10
US	1.40×10^9	3.96×10^8	5.87×10^8	4.05×10^8	4.33

These ten countries were used for trend analysis to further understand current emissions trends and project possible future emissions. Trend analysis was completed using ordinary least squares regression for each country and each type of fuel emissions. This involved using a set of predictors, $\boldsymbol{x_n}$, that related to a linear response in a variable Y (Eq. 2),

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 + \cdots \beta_n x_n \qquad (2)$$

where β_n symbolized the response of Y because of x_n. The beta weights, $\beta_{1,2,\ldots n}$, are assessed by determining the minimal sum of squared distances between the predicted and actual values of Y. The predictors in this equation can characterize different variables or transformations of a single variable; in this analysis, the two predictors were based on a polynomial transformation of time in years. Thus, the trend analysis was reduced to estimating 3 parameters based on the transformed yearly values: $\beta_0, \beta_1,$ and β_2. Eq. 2 could then be expressed as:

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 \qquad (3)$$

Where *x* was the number of years since the reference year, 2005. Using a quadratic fit assumed that the acceleration of emissions is constant, which was consistent with studies of global emissions (Andres et al., 2012). By computing this trend analysis on total, solid fuel, liquid fuel, and gas fuel emissions, specific patterns in emissions by fuel type can be analyzed. Historic solid fuel, liquid fuel, and gas fuel emissions data points for 2005 through 2015 were fit using Eq. 3. To evaluate the fit of the least squares' regression models, the R^2 value was computed for each model. R^2 is a ratio that signifies the proportion of variance of a dependent variable explained by the independent variable in a regression model. It is calculated by finding the sum of the squared difference between the predicted and actual values for each observation and dividing by the variance of the observations. Values of R^2 could range from 0 to 1; a value of 1 would indicate that the regression model perfectly aligns with the observed points. An R^2 value of 0 indicates that the model is not explaining any of the observed variance.

The rates of change of emissions data points over the ten-year period were examined. This measure will show which countries have increased emissions the most over time. This was examined for emissions from solid fuel consumption, emissions from liquid fuel consumption, and emissions from gas fuel consumption.

Once the best fit trends were determined for each of the 10 countries and their emissions patterns, they were extended to the year 2050 to create a business as usual projection. The business as usual pattern was based of the sum of the emissions from solid fuel consumption, emissions from liquid fuel consumption, and emissions from gas fuel consumption. This projection implied that no actions were taken to affect the current trend of carbon emissions, and current conditions continue unhindered into the future. These analyses reveal which countries emissions have the potential to become excessive over time if no corrective action is taken.

3.3 Kaya Factors and LMDI Decomposition

In order to further answer what drives countries to emit, an LMDI analysis was completed. This analysis was used to construct index values from the continuous time series data that was collected on each country from 2005 – 2015. These index values are useful for comparison since they are summaries and condense a time series into a single value. These LMDI values reflect how a certain Kaya component affects, or drives, emissions over time. To assess changes in emissions, C, to a certain Kaya factor, C_x, over time, Eq. 1 was rewritten as follows:

$$C \equiv P * \frac{GDP}{P} * \frac{E}{GDP} * \frac{C}{E} = C_p * C_W * C_{EI} * C_{CI} \qquad (4)$$

where C_p is the emissions from population, C_W is the emissions from wealth, C_{EI} is the emissions from energy intensity, and C_{CI} is the emissions from carbon intensity. Changes in emissions were evaluated between the years 2005 and 2015. The variable units and data sources are summarized in **Table 1**. Since energy supply (E in Eq. 1 and Eq. 4) data were not available from the data source in **Table 1**, the energy intensity data were multiplied by the GDP data to get this value for calculating Carbon Intensity ($\frac{C}{E}$ or C_{CI}).

Generally, changes in emissions, ΔC, from the reference year t_1 to year t_2 were decomposed exactly, i.e. no residual terms, as follows. In this equation, ΔC_x was the change in total emissions while ΔC_P was the change in emissions from change in population, ΔC_W was the change in emissions from change in wealth, ΔC_{EI} was the change in emissions from change in energy intensity, and ΔC_{CI} was the change in emissions from change in carbon intensity:

$$\Delta C = \Delta C_P + \Delta C_W + \Delta C_{EI} + \Delta C_{CI} \qquad (5)$$

Where:

$$\Delta C_x = \frac{C^{t_2} - C^{t_1}}{\ln(C^{t_2}) - \ln(C^{t_1})} \ln \frac{C_x^{t_2}}{C_x^{t_1}} \qquad (6)$$

ΔC_x represented the change in emissions that was attributable to changes in factor x over the time period t_1 (2005) to year t_2 (2015). This decomposition gave one final value per Kaya component. Emissions were decomposed for every country with available data for the years 2005 and 2015. The analysis was focused on the ten selected countries in section 3.1. The resulting LMDI values were then used in a global cluster analysis.

3.4 Cluster Analysis

Cluster analysis was conducted to address the question of whether countries could be grouped meaningfully based on LMDI values. The index values from Eq. 6 were used as inputs for this analysis. Cluster modeling is an unsupervised machine learning approach that involves grouping data points by their similarities into clusters, and separating clusters based on their differences. In this case, unsupervised means a clustering model with unlabeled data present rather than a classification model that uses only labelled data (Jain 2010).

This analysis used k-means clustering, which is a type of centroid clustering. In this analysis, a number of clusters (k) were created based on assigning data points to k number of predefined clusters. This was computed by assigning each data point to a random mean based on the selected number of clusters and a distance metric, e.g. Euclidean, Manhattan, or Pearson.

The distance metric was an important part of the cluster model process because it determined how objects would be considered similar and how the clusters would form. For this analysis, Euclidean distance was used to calculate the shortest distances between objects in a straight line, instead of the Manhattan distance (a measurement of gridded distance) or the Pearson distance (a distance that is based on correlation of objects) (Irani et al., 2016).

The K-means algorithm began by selecting k (the predetermined number of clusters) objects from the dataset and assigning them as centroids: the randomly selected means of the clusters. The other objects from the dataset were then assigned to the closest centroid based on the minimum Euclidean distance between that object and the nearest centroid. The preliminary clusters were then formed. At this point, the mean values were recalculated based on the newly formed clusters, and any objects that were closer to the mean of a different cluster were rearranged accordingly. The cluster means would continually be updated, and objects would continue to be reassigned iteratively until objects remain in their respective clusters. At this point, the clusters have stabilized (Jain, 2010).

The ultimate goal of K-means clustering was to minimize the error, or distance, between the objects and the centroid mean. This distance was known as the within cluster variation. The equation used to determine the within cluster variation for an individual cluster C_K is described as follows in Eq. 7:

$$W(C_K) = \sum_{x_i \in C_K} (x_i - \mu_K)^2 \tag{7}$$

Where $W(C_K)$ is the within cluster variation, x_i is an object in cluster C_K, and μ_K is the cluster mean of C_K. The total within cluster variation could be minimized by Eq. 8, which describes the total within cluster variation for all clusters 1 through K in the model:

$$Total\ \ W(C_K) = \sum_{K=1}^{K} \sum_{x_i \in C_K} (x_i - \mu_K)^2 \qquad (8)$$

Where $Total\, W(C_K)$ is the total within cluster variation. Another name for the within cluster variation is the within cluster sum of squares (WSS). A smaller WSS value indicated a smaller distance between cluster objects and their respective means, and overall more compact clusters. A larger WSS value indicated a less compact cluster. More compact clusters were preferred because they indicated higher goodness of fit, or groups of very similar objects. Less compact clusters indicated groups of less similar objects. Because the WSS value was dependent on the number of objects in a cluster, it was important to compare clusters using variance rather than the WSS. Variance could be defined as the WSS value divided by the number of members in the cluster. A higher variance implied the clusters are more spread out. A lower variance implied the clusters are more compact, and thus more similar.

It was also useful to know the distance between cluster centroids to determine how similar entire clusters are to one another. This measure can be described as the between sum of squares distance (BSS). This value was calculated by first computing the squared Euclidean distance from one cluster centroid to all other cluster centroids, then repeating for K clusters in the model. The total BSS distance was the sum of all of these squared distances. A higher value indicated less similar or clusters that are farther apart, while a lower value indicated more similar clusters or clusters that were closer together. An optimal cluster model would have a large BSS and a small WSS, meaning that the distance within clusters was small and the distance between clusters was high.

It was also important to measure the total sum of squares (TSS) distance of all points from the global mean of a variable. To determine the percentage of variance that is explained by the clusters, the BSS could be divided by the TSS value and multiplied by 100. A higher percentage indicated these clusters are explaining a high degree of variance and are therefore effective. This was a useful tool for determining whether adding an additional cluster improved the model. If the percentage of variance explained by the model increased with the addition of a cluster, it was useful. If the percentage of explained variance did not increase or the change was negligible, the addition of that cluster was not needed, and k clusters was sufficient.

One challenge of k-means clustering was that the number of clusters, k, must be predetermined. Validity indices exist to assist in optimizing number of clusters. Some of these were employed in this analysis. One that was considered was the Dunn Index. This measurement compared the smallest distances between clusters to the largest cluster diameter (Legány et al. 2006). If the clusters were well-defined, then the distance between individual clusters would be large, and the distance between points within a single cluster would be small. Therefore, a larger Dunn index indicated an optimal number of clusters. Conversely, if the distance between clusters was small and the distance between points within a cluster was large, this ratio would be smaller, and the clusters were more ill-defined at that value of k.

Beyond the Dunn Index, there were other indices to assess the optimal number of k-means clusters employed in this analysis. Thirty of these were calculated using the NbClust package, a statistical package in R software. This reported the optimal number of clusters as reported by the majority of indices included (Charrad et al., 2014).

Finally, the silhouette widths were considered for the clusters to determine if objects within a cluster were assigned well. The individual silhouette width described how well each object fit into its cluster. The average silhouette value described how well all objects fit into a cluster, and the quality of that cluster as a whole. The silhouette value was calculated in the following way. First, the dissimilarity, or the average distance between, an object (i) and the members in its cluster was calculated. This value was known as a(i). Then, the lowest dissimilarity between that object (i) and all other clusters was calculated. This was the cluster that the object is most similar to after the one it is currently in. This value was known as b(i). To calculate the silhouette width, a(i) was subtracted from b(i), and the resulting value is divided by the greatest value of either a(i) or b(i) (Lengyel & Botta-Dukat, 2019). This is described below in Eq. 9:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \tag{9}$$

Where $s(i)$ is the silhouette width of an object. The silhouette of an object was the comparison of a data point’s similarity to its own cluster versus its separation from other clusters. Silhouette widths were measured in a range between 1 and -1. A silhouette value close to 1 indicated that an object is strongly related to the assigned cluster. The closer a silhouette width became to 0, the less likely it was classified correctly. If a silhouette value is -1, the data point was clearly mis-clustered. If the average silhouette widths of the data points within a cluster collectively had a silhouette value close to 1, the cluster was well-defined whereas if the data points had a silhouette close to -1, then the cluster as a whole was ill-defined (Kodinariya et al., 2013). Silhouette widths were used to evaluate if any objects were potentially misclustered, and which objects were closely related to clusters.

3.5 Sensitivity Analysis of Clustering

A sensitivity analysis was run to determine how impactful certain variables were on the model as a whole, or how sensitive they were. One simple and effective way to determine a variable's sensitivity was to remove it from the model. This is known as a jack knife (Kott, 2001). This approach was used by removing each of the Kaya components one at a time and rerunning the cluster model. The most sensitive variables would change the percentage of explained variance value ($BSS/TSS \times 100$) the greatest amount. This was because variables that were controlling the explained variance to a higher degree were more critical to the model.

To further examine the effects of removing one variable, the resulting Dunn Index was also compared for each of these jack knife models. A larger change in the Dunn Index indicated higher sensitivity, because removing this variable had a strong enough impact to change the optimal number of clusters. An unchanged or only slightly shifted Dunn Index signified a relatively insensitive variable, or one that did not have a large effect on the optimal number of clusters. In addition, the average silhouette widths of the clusters were also considered in the jack knife models. If the average silhouette widths shifted greatly with a variable removed, it implied that variable was fairly sensitive to the model. Smaller shifts in average silhouette widths for the clusters indicated that the removed variable was not as sensitive to the overall model.

In summary, the cluster analysis consisted of determining the optimal number of clusters with the total percentage of variance explained, the Dunn Index, the silhouette widths, and the Nbclust majority response. The sensitivity analysis consisted of a jackknife operation where one variable was removed at a time and the resulting Dunn Index, cluster membership, number of negative silhouette values, and percentage variance explained were compared to interpret the insights from clusters of similar emitters.

Chapter 4: Results and Discussion

4.1 Trend Analysis of Total Emissions

Trend analysis was performed on the 10 selected countries for the years 2005-2015. The emissions patterns were fit to a quadratic curve according to Eq. 3 for all fuel types. Besides the United States and Germany, all countries total emissions curves exhibit emissions trending upwards. This indicates that they are increasing over time and will not peak naturally unless action is taken.

While all countries were fit with a quadratic term for this analysis, it is important to note that there is one unique case: Iceland. If fit with a linear curve, Iceland's emissions trend downward in projections. If fit with a quadratic curve, Iceland's emissions trend upward in projections due to the shape of the curve. For easier comparison, all of the models were fit with quadratic curves, including Iceland.

The maximum R_2 value is 0.989 for China's gas fuel emissions followed by 0.987 for China's liquid fuel emissions. The minimum R_2 value was 0.054 for Iceland's solid emissions. The average R_2 value is 0.779 for the ten countries, and therefore the quadratic curve is explaining about 78% of the variance. A complete table of the Beta coefficients and R_2 values can be found in the Appendix.

Trend analysis projections revealed that between the years 2015 and 2050 China could increase emissions by 425% and India could increase emissions by roughly 314%. However, there is concern beyond the heavy emitters as well. In this same curve function, Saudi Arabia could increase emissions by almost 328%, implying that focusing on only the heavy emitters in the climate discussion ignores potential future heavy emitters. Argentina's projected emissions in 2050 increase by 99% since 2015, and even the Marshall Islands emissions are projected to increase by 64%. In this analysis, only the United States and Germany displayed a projection of decreasing emissions. Both countries could approach zero emissions before 2050. The 2050 values that were estimated can be found in the Appendix. These scenarios are merely a framework for what individual country emissions could be if no action is taken. However, they highlight some of the countries that have the potential to emit much more in the future such as Saudi Arabia, even if they are not currently considered a heavy emitter.

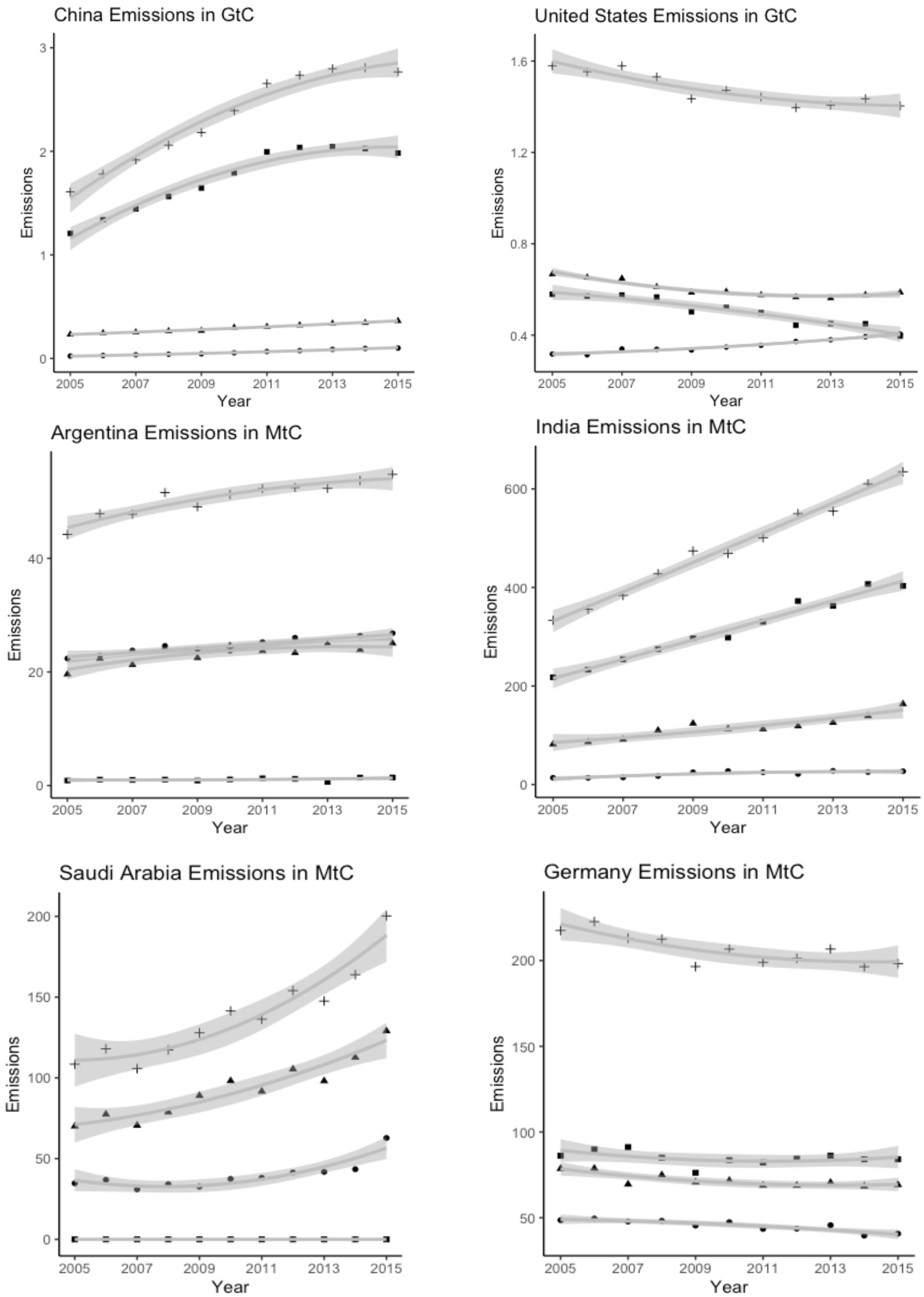
4.2 Trend Analysis of Fuel Usage

Historic emissions were also considered based on the years used in this analysis, 2005 to 2015. **Figure 2** depicts the trend analysis for the 10 selected countries on solid, liquid, and gas fuels, and total emissions. The gray lines indicate the 95% confidence interval for the quadratic function on each set of these data points. Emissions are in GtC, MtC, or ktC depending on the country.

The rate of change in emissions over the 10-year span was highest in China, at roughly 116 MtC a year. This was followed by India, where there was a rate of about 18.5 MtC per year. In both of these countries, the largest rate of change for an individual fuel type comes from solid fuels. The rate of change of solid fuel emissions in China was close to 77.6 MtC a year. The United States and Germany both had emissions trending downward, and therefore negative rates of change. The United States emissions have decreased at a rate of about 17.5 MtC per year over the past 10 years. This is mainly driven by a very large decrease in the rate of solid fuel emissions (almost 18.3 MtC per year) but is counteracted by an increase in the rate of change of gas fuel emissions (about 8.75 MtC per year). Similarly, Germany has also shown decreasing emissions and a negative rate of change in emissions over time. Germany has decreased total emissions at a rate of roughly 1.93 MtC per year. After the heavy emitters, Saudi Arabia has the next largest rate of change in emissions, at about 9.18 MtC per year. Saudi Arabia, along with Ethiopia, primarily emit from liquid fuel usage. Comparatively, the Marshall Islands emissions change at a rate of 0.8 ktC a year, a rate that is almost entirely comprised of liquid fuel emissions.

The use of trend analysis answered the first research question: beyond the three heavy emitters, what are some country-level patterns of emissions over time? While two of the heavy emitters demonstrate the largest emissions rates of change over time, they are mainly powered by solid fuel usage. Other countries in this dataset, including the United States, show a majority of emissions coming from liquid and gas fuel usage. Besides the United States and Germany, the rest of these 10 countries all show positive trending emissions in the quadratic model. Different fuel usage profiles emerge even between these 10 countries, suggesting the need for more varied solutions and climate targets.

Emissions Trends in Select Countries 2005 - 2015



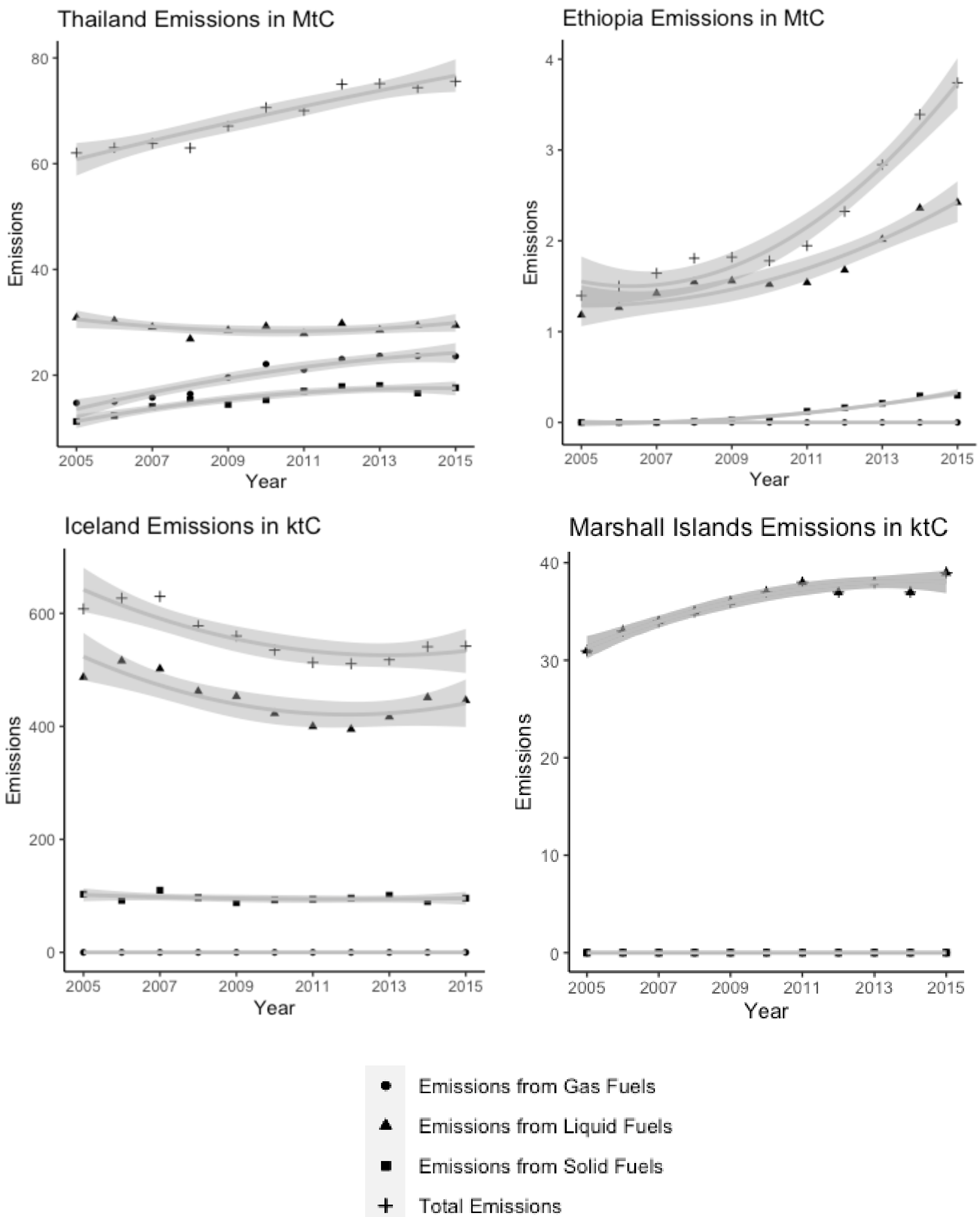


Figure 2: Emissions trends as quadratic functions from 2005 – 2015 emissions data points for 10 select countries based on fuel type emissions and total emissions. Emissions are in GtC for the United States and China, in ktC for Iceland and Marshall Islands, and in MtC for Argentina, Ethiopia, Germany, India, Thailand, and Saudi Arabia. Grey bars indicate the 95% confidence interval for the trend lines. Total emissions are the sum of all other emissions types.

4.3 Kaya Factors Breakdown and LMDI Decomposition

While only the selected countries were used for comparing trends in emissions, changes in emissions were decomposed using a LMDI analysis for all countries in the years 2005 and 2015. Countries with insufficient data were removed and 176 countries remained for analyses. The Kaya components of the 10 selected countries can be found in **Table 4** and **Table 5** for the years 2005 and 2015 respectively. These tables highlight the difference in these 10 countries across the factors.

Table 4: Kaya factors broken down for 10 selected countries in 2005. Population is measured in millions of people, Wealth is measured in the 2011 Int. dollar per person, Energy Intensity is in MJ at the 2011 GDP level, and carbon intensity is in thousand metric tons of carbon per unit of energy.

Study Country	Year	Population (Millions) (* Billions)	Wealth (2011 Int. Dollars/ Person)	Energy Intensity (MJ/2011 Int. Dollar)	Carbon Intensity (MtC/Unit of Energy)
Argentina	2005	38.9	15,600	4.60	1.58×10^{-8}
China	2005	1.33*	5,580	10.3	2.10×10^{-8}
Ethiopia	2005	76.3	736	27.5	9.06×10^{-10}
Germany	2005	81.6	38,000	4.51	1.55×10^{-8}
Iceland	2005	0.295	40,900	11.2	4.47×10^{-9}
India	2005	1.15*	3,410	5.88	1.45×10^{-8}
Marshall Islands	2005	0.055	3,230	6.99	2.49×10^{-8}
Saudi Arabia	2005	23.8	45,800	5.34	1.86×10^{-8}
Thailand	2005	65.4	11,500	5.50	1.49×10^{-8}
United States	2005	295	49,500	6.60	1.63×10^{-8}

Table 5: Kaya factors broken down for 10 selected countries in 2005. Population is measured in millions of people, Wealth is measured in the 2011 Int. dollar per person, Energy Intensity is in MJ at the 2011 GDP level, and carbon intensity is in thousand metric tons of carbon per unit of energy.

Study Country	Year	Population (Millions) (* Billions)	Wealth (2011 Int. Dollar/ Person)	Energy Intensity (MJ/2011 Int. Dollar)	Carbon Intensity (MtC/Unit of Energy)
Argentina	2015	43.1	19,300	4.34	1.52×10^{-8}
China	2015	1.41*	13,200	6.69	2.23×10^{-8}
Ethiopia	2015	101	1,520	13.7	1.79×10^{-9}
Germany	2015	81.8	43,800	3.60	1.54×10^{-8}
Iceland	2015	0.330	44,330	16.6	2.24×10^{-9}
India	2015	1.31*	5,740	4.73	1.79×10^{-8}
Marshall Islands	2015	0.057	3,320	11.4	1.82×10^{-8}
Saudi Arabia	2015	31.7	50,400	5.80	1.90×10^{-8}
Thailand	2015	68.7	15,300	5.41	1.37×10^{-8}
United States	2015	321	53,300	5.41	1.51×10^{-8}

All of these countries experienced a rise in population and wealth over time from 2005 to 2015. The largest positive change in population over this period was in Saudi Arabia. The largest positive change in wealth over this period was in China. Out of these 10 countries, seven of them experienced decreasing energy intensity over time. This suggests that over time, there is a lower cost to produce the same unit of output in these countries. China showed the largest decrease in energy intensity over time. This is likely the result of significant energy policy in China in 2006 aimed at increasing efficiency to meet national targets (Zhou et al., 2010). Of these 10 countries, only Iceland, the Marshall Islands, and Saudi Arabia had an increase in energy intensity over this time period. China, India, Saudi Arabia, and Ethiopia all showed positive change in carbon intensity for this period. A positive change in carbon intensity indicates that more carbon is emitted per unit of energy produced over time.

To capture the change in these values over the 10-year span, the LMDI decomposition was conducted. This analysis was completed on 176 countries, for each of their Kaya components. For continuity of discussion, only results of the same 10 study countries are presented. The results of this decomposition can be seen in **Figure 3** below.

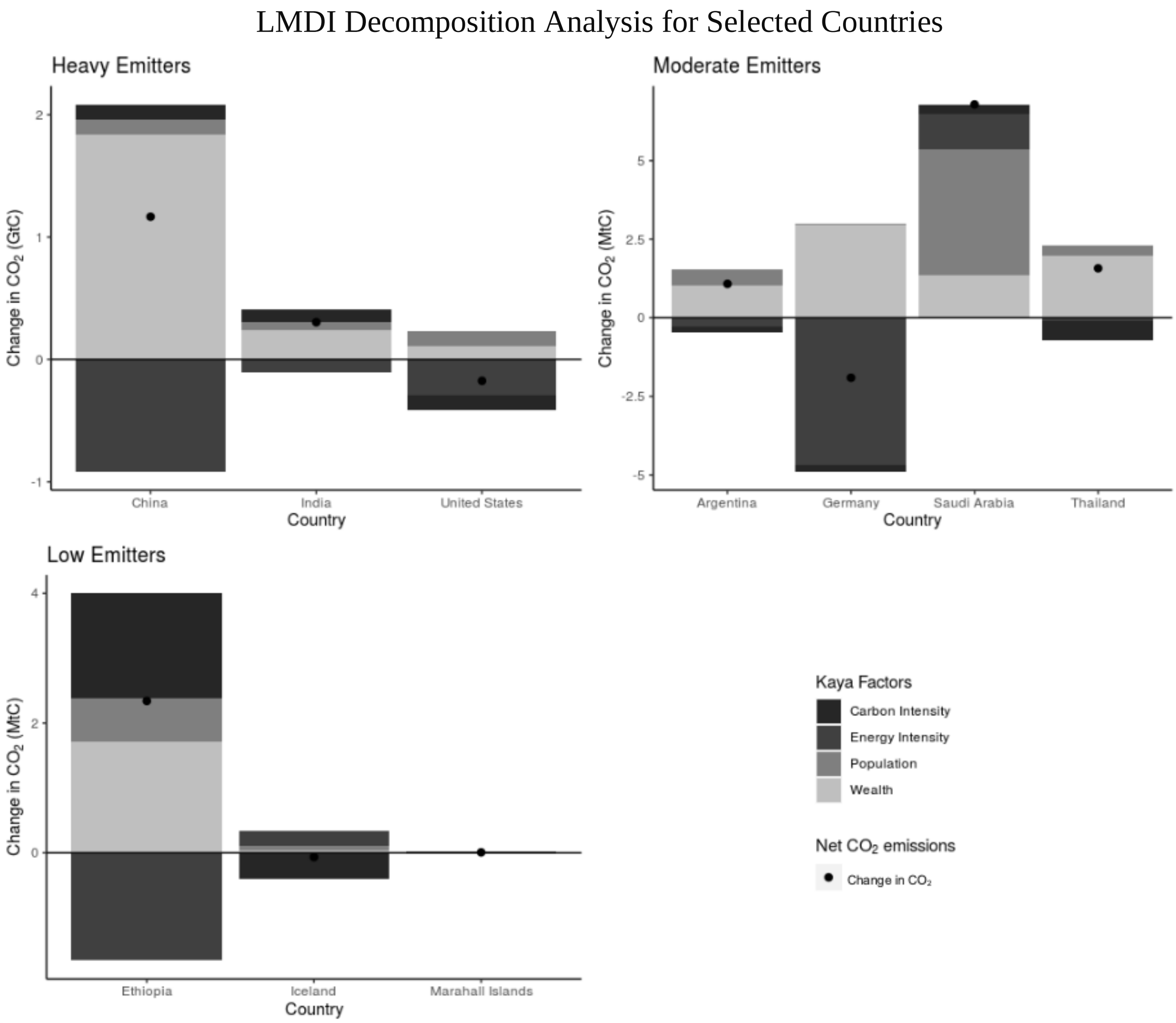


Figure 3: LMDI breakdown of Kaya components for 10 selected countries from 2005 - 2015. Panel 1 includes the heavy emitters measured in GtC, Panel 2 includes the moderate emitters in MtC, and Panel 3 includes the low emitters in ktC. The circle represents the change in CO_2 over time for this period. Each Kaya component is represented by how attributable the change in that factor over time is to the change in CO_2 over time.

This figure separates the 10 study countries into the three heaviest emitters, four moderate emitters, and the three lowest emitters. All three heavy emitters have a deceasing energy intensity over time in common. This indicates that it is taking less energy over time to produce one unit of economic output. All three heavy emitters also have an increasing wealth component over time. Comparatively, population and carbon intensity are not as influential of Kaya components to changing CO_2 emissions for this period. Only the United States has a decreasing carbon intensity value, along with a negative CO_2 value over time. This indicates lower CO_2 emissions per unit of energy over the 10-year study period. The United States is an example of how there are differences in emissions drivers that lead to different emissions pathways even within the heaviest emitters.

For the moderate emitters, an increasing wealth component as well as an increasing population growth is common. Saudi Arabia is distinguished by the largest population growth over time of all 10 countries. Although there is large population growth across the Middle East region, Saudi Arabia has a larger percentage population growth than Bahrain, Jordan, Qatar, the United Arab Emirates, Oman, and Kuwait (Rahman et al., 2017). Furthermore, the highest rate of growth within Saudi Arabia is in urban areas, which correlates with the growth of emissions from transportation and electricity energy sectors (Rahman et al., 2017). Of the four moderate emitters in this group, Argentina and Germany have decreased their carbon intensity over time. This indicates that over time, these countries are able to emit less CO_2 per unit of energy consumed. Germany has a comparatively larger negative energy intensity, similar to some of the heavy emitters.

The low emitters have more unique profiles of emissions drivers. Ethiopia is characterized by a large, positive change in carbon intensity. Egypt and Sudan, Ethiopia’s geographical neighbors, also share this large increase in carbon intensity. Currently, Ethiopia's carbon intensity is largely driven by the agriculture sector (Hamilton & Kelly, 2017). Ethiopia also shows a decreasing energy intensity, similar to the heavy emitters. Only Iceland and Saudi Arabia exhibit a positive change in energy intensity, meaning that more energy is required over time to produce the same unit of output. Finally, the Marshall Islands is barely visible at all, which demonstrates the vast differences in scale of these different countries in terms of emissions.

This analysis answers the second research question which focuses on determining global drivers of emissions. These results demonstrate that while there are some similarities among emitters, there are also large differences in emissions drivers. These drivers can be traced back to specific energy sectors as in Ethiopia, or specific policies such as in China. Whatever the cause these drivers appear to vary greatly by country. Overall, the drivers of emissions are unique at the country level and present different challenges for individual countries to mitigate emissions. These differences could be more effectively met with differentiated climate goals if every country is to be involved with mitigation efforts.

4.4 Cluster Analysis

These LMDI values were ultimately used in a K-means cluster analysis. Exploratory cluster analysis identified China, US, and India as outliers. These countries often clustered individually, which does not allow for comparison with other countries. All three countries exhibit population and carbon intensity LMDI values that are more than three standard deviations outside the mean for all countries. Both China and India's LMDI wealth increase components are roughly two standard deviations away from the mean of all values. India's

LMDI energy intensity value is over three standard deviations from the mean of all values. Only the US and China have emitted over 1 GtC in any years from 1990 to 2015. These extreme differences from other countries make these three outliers, and therefore they will not be included in the clustering models any further.

The optimal number of clusters is decided from a range of one cluster to 20 clusters. The Dunn Index reveals that three clusters are the optimal amount for this remaining dataset. This is supported by 10 of the other 26 methods for choosing optimal clusters in the NbClust package. For three clusters, the Dunn Index is 0.979, where the next closest Dunn Index, 0.660, was indicated for two clusters.

Using three clusters explains 57.6% of the total variance of the LMDI values; the addition of one more cluster would only improve the performance by 5.8%, and two more clusters an additional 14.4%. While more variance is explained in models with more clusters, the cluster membership did not lend itself to grouping and comparison as well. The five-cluster model explains 77.7% of the variance, but it only does so because Japan and Russia separate from the smallest cluster into individual clusters. Because this research is focused on grouping countries, and had already separated out 3 outliers, choosing a model that further isolated countries is not effective for answering the research questions. The highest Dunn Index and the NbClust recommendation methods are utilized to choose the optimal number of clusters. Both of these methods suggest three clusters. **Table 6** below further shows the values for multiple numbers of clusters.

Table 6: Dunn Index, number of indices that selected k as the best cluster value, Percentage Variance Explained, Average Within cluster Sum of Squares (WSS), Between Cluster Sum of Squares (BSS) and Total Sum of Squares (TSS) for 3, 4 and 5 clusters. The Average WSS is calculated by adding up the WSS values for all clusters and dividing by the number of clusters. WSS, BSS, and TSS are in Euclidean distance.

Number of Clusters	Dunn Index	Nbclust Package	%Variance Explained	Average WSS	BSS	TSS
3 Clusters	0.979	10	57.5	8.07×10^9	3.28×10^{10}	5.70×10^{10}
4 Clusters	0.299	2	63.3	5.23×10^9	3.61×10^{10}	5.70×10^{10}
5 Clusters	0.211	4	77.7	2.62×10^9	4.4×10^{10}	5.70×10^{10}

These optimal three clusters contain the following groups: a large cluster with 146 members (Cluster 3), an intermediate cluster with 23 members (Cluster 2), and a small cluster of four members (Cluster 1). Overall, Cluster 3 contains mostly developing or small nations. Cluster 2 contains some Middle Eastern countries, some European countries, Canada, Australia, and Brazil. Cluster 1 contains only four members: Russia, Japan, Indonesia, and Germany. A full breakdown of the clusters and countries they contain can be found in the **Appendix**. The within cluster variance is used to assess compactness. This is calculated by dividing each within cluster variance by the number of countries in each cluster. The within cluster variance is 2.7×10^9 in Cluster 1, 3.7×10^8 for Cluster 2, and 3.2×10^7 for Cluster 3. Cluster 3 is the most similar due to having the smallest WSS, whereas Cluster 1 is the least similar due to having the largest WSS. This is interesting since cluster 3 is the largest cluster. This implies that the 146 countries in cluster 3 emit more similarly than the 4 countries in Cluster 1.

Silhouette widths are used to assess how well individual members fit within the given clusters. A negative silhouette width indicates that a country is misclustered. Germany and Iraq have the largest negative silhouette widths of all countries, with values of -0.383 and -0.234 respectively. Germany's closest neighboring cluster is Cluster 2 and Iraq's closest is Cluster

3. Other notable potentially misplaced countries included Indonesia, Saudi Arabia, Egypt, France, Ukraine, Thailand, Brazil, Spain, Nigeria, Qatar, Turkmenistan, and Vietnam. Indonesia is placed in Cluster 1 while it's neighboring cluster is 2. The rest of these countries are all members of Cluster 2 and the closest neighboring cluster is 3.

To better understand what is driving each cluster, the cluster means are considered. Cluster means can be seen in **Figure 4** below. The mean of largest magnitude for Cluster 1 is the extremely negative energy intensity value and a large, positive wealth value. The average LMDI wealth index value among all countries in this research is 3,934.7. Japan's wealth value is one standard deviation from this mean while Russia's is more than three standard deviations away from this mean. The high wealth component in this cluster could be being driven by Russia's large change in wealth over time. These countries appear to be more similar in terms of energy intensity. They are all on a similar order of magnitude and negatively changing over time. While the mean LMDI energy intensity value for all countries is -4,025.6, these countries' LMDI energy intensity values are three or more standard deviations away. Cluster 1 can be described as having large wealth growth and large decrease in energy intensity over time. However, it is important to note the large variance in Cluster 1. While these means are useful for telling what drives the clusters, this cluster in particular is very spread out making the interpretation more complex.

Cluster 2 shows similar drivers as Cluster 1, just of a lesser magnitude. Cluster 2 shows a moderate wealth increase over time and a moderate decrease in carbon intensity over time. The wealth outlier in Cluster 2, shown as a dot in **Figure 4,** is the Republic of Korea. Perhaps the most striking thing about Cluster 2 is its slightly negative carbon intensity mean. The most negative carbon intensity LMDI values in this cluster are coming from France, Spain, Canada, and the United Kingdom. This cluster also has the largest population mean of all three clusters. Cluster 2 does contain some countries with very high population growth over time, such as Saudi Arabia, Qatar, and Mexico, which could be driving the higher population mean. The population outlier depicted as a dot above cluster 2 in **Figure 4** is Saudi Arabia. However, while 20 countries in this dataset did experience population loss at low levels over this time period, 152 countries experienced positive population growth over time. Therefore, population growth does not have much effect on defining clusters because population growth is a common factor in all of the clusters.

Cluster 3 contains the rest of the countries in the world closely clustered together with the least amount of WSS variance. These countries have lower wealth growth over time, as well as lower energy intensity growth over time. These countries also have a small, positive carbon intensity mean. While many countries in this cluster have negative carbon intensity index values (the most negative of which being Italy), there are some countries with very high carbon intensities present in this cluster such as Libya, Iran, Singapore, and Pakistan. The outliers present in the population mean of **Figure 4** depicted as dots above cluster 3 are the United Arab Emirates, Iran, Kuwait, and Pakistan respectively. The means of the clusters and the variances within each cluster are presented in **Figure 4** below.

Cluster Mean Plots for Kaya Components

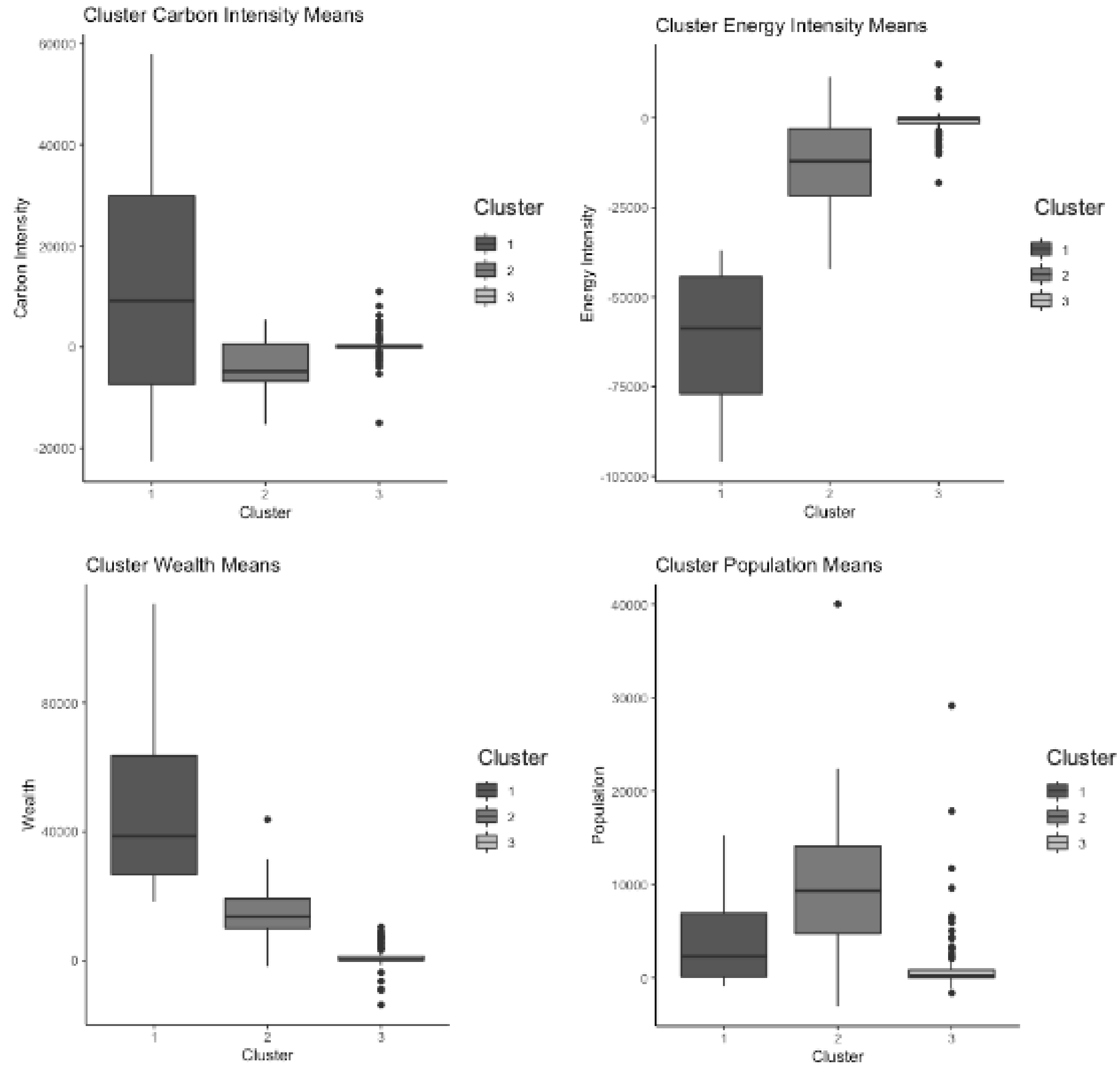


Figure 4: Cluster means are depicted as the black horizontal lines in the boxes describing each cluster. The box structure spans the interquartile range, while the lines indicate the highest and lowest values. Outliers are indicated by dots outside of the box and whisker structure.

The study group members separated into the different clusters, which is useful for comparison and discussion. Of the 10 selected countries, Argentina, Ethiopia, Iceland, and

Marshall Islands are in Cluster 3. Saudi Arabia and Thailand are in Cluster 2. Germany is in Cluster 1.

Some of the 10 study countries emerge more clearly if they are separated by one Kaya component on each axis. In **Figure 5**, all countries are plotted on an axis of change in wealth against an axis of change in carbon intensity. In this image, it is evident how spread apart cluster 1 (consisting of Russia, Japan, Indonesia, and Germany) is. While Russia has a low change in carbon intensity and a high change in wealth, Japan has the opposite condition: a high change in carbon intensity and a low change in wealth. Still, these two countries cluster together in Cluster 1.

Germany fades into the group of countries in Cluster 2 and some of Cluster 1. The low decrease in carbon intensity makes Germany more comparable with Cluster 2 and Cluster 3 as opposed to the larger decrease in carbon intensity seen in Russia that is driving cluster 1. Thailand, another one of the 10 countries, can be visible just below Germany. Thailand shows a similar positive change in wealth and slight decrease in carbon intensity as well.

In this sense, Cluster 1 is a study in just how different drivers of emission can be. Germany, a member of Cluster 1, shows similarities to Thailand, a member of Cluster 2. This shows the potential for some overlap of climate solutions between unlikely countries. However, Japan and Russia are almost opposites, even if they are clustered together. One solution would not be effective for each of these countries. Overall, this shows the need for differentiated climate targets even when there are some similarities.

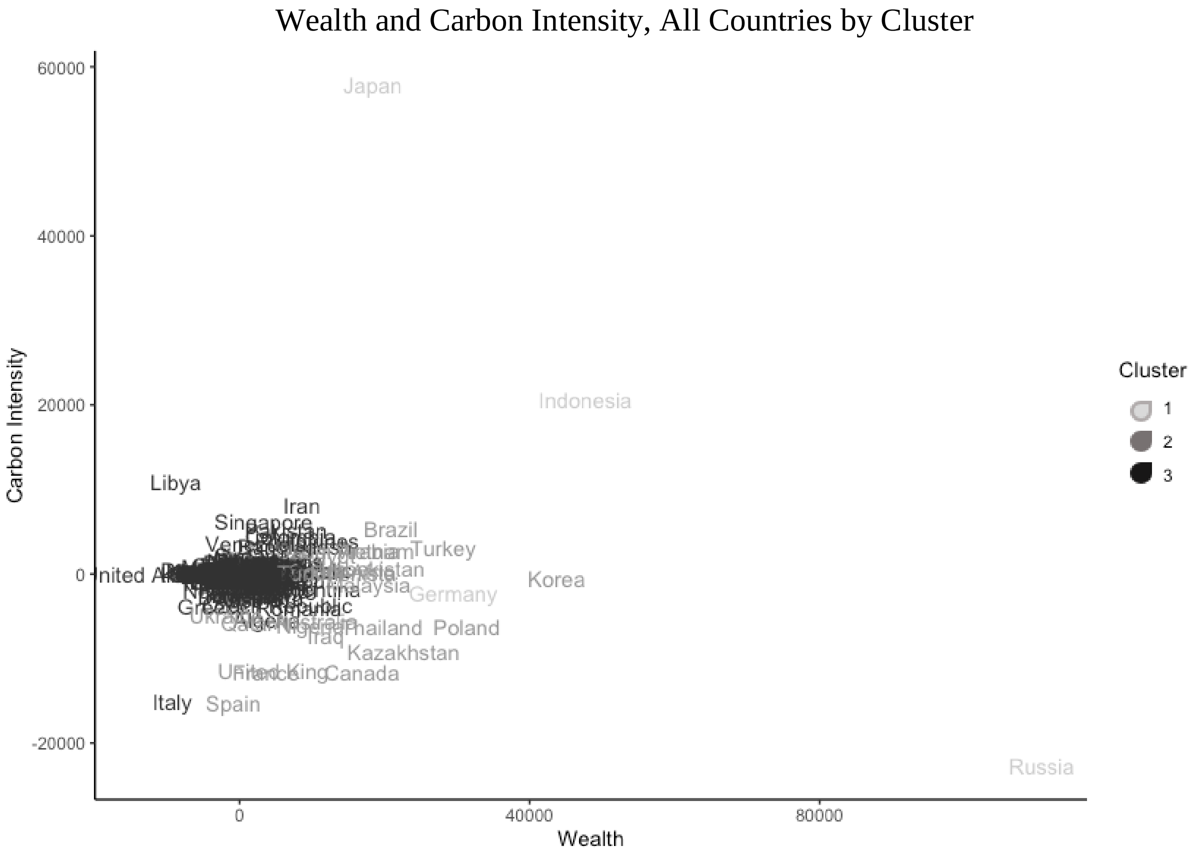


Figure 5: Clustered countries compared on an axis showing change in wealth for 2005-2015 and an axis of change in carbon intensity over 2005-2015. Cluster 1 is shown in the lightest text, while Cluster 3 is shown in the darkest text.

Germany stands out more if it is compared to the other countries on an axis of change in wealth against an axis of change in energy intensity, which is shown in **Figure 6**. Again, the extreme cases of Japan and Russia can be seen when these Kaya components are compared. However, Germany stands apart from Cluster 1 and Cluster 2 when these components are compared due to its greater decrease in negative energy intensity. Past research has linked this decrease of energy intensity in Germany to the shifting of resources out of the manufacturing industry and into the services industry (Koesler et al., 2016). It appears that countries in Custer 2 are beginning to move towards Germany's position between these two Kaya components. Other European countries such as Ukraine, Poland, United Kingdom, and Spain are showing varied levels of decreasing energy intensity over time and increasing wealth over time. Saudi Arabia, one of the study countries, stands out as having one of the highest increases in energy intensity. Other Middle Eastern countries are also separating out due to increasing energy intensity; for examples the United Arab Emirates, Oman, and Iran. This could be due to increased energy demand from a growing population, a growing industrial sector, and a growing standard of living throughout the region in this time period (Nematollahi et al., 2014). Again, the differences in drivers are evident here between the 10 study countries as well as some of the extreme cases in the clusters. A single target goal might not effectively encompass all of these drivers.

Wealth and Energy Intensity, All Countries by Cluster

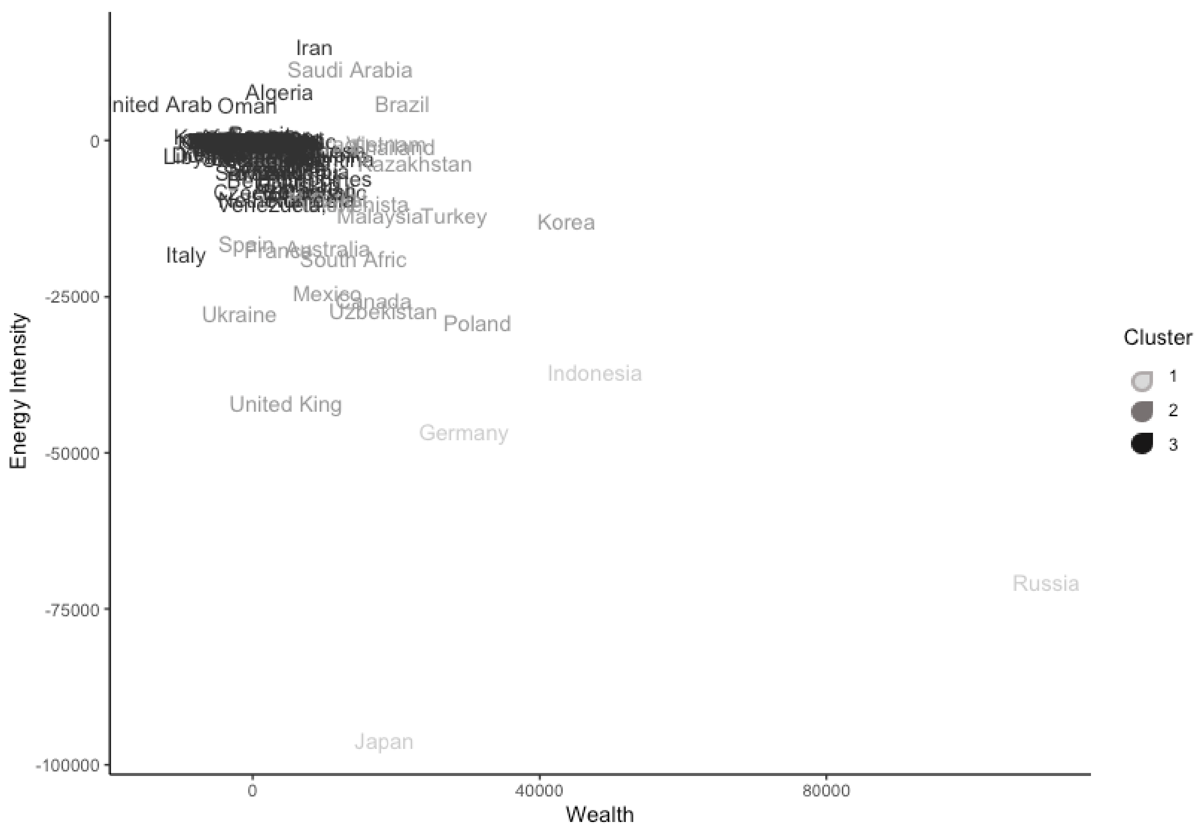


Figure 6: Clustered countries compared on an axis showing change in wealth for 2005-2015 and an axis of change in energy intensity over 2005-2015. Cluster 1 is shown in the lightest text, while Cluster 3 is shown in the darkest text.

Finally, the dispersion of the clusters on the change in wealth over time on one axis and the change in population on the other axis is considered. The Middle Eastern countries such as Saudi Arabia, United Arab Emirates, Qatar, Iran, and Kuwait all show a large, positive change in population, even though they are members of different clusters. Countries in Cluster 1 do not show any large increases in population. Germany can be found on the far-left side of **Figure 7**, showing a small increase in population over these years. Some of the countries that maintained a large growth in wealth despite a lower growth in population are seen in **Figure 7** as well, for examples Korea and Poland. While population is not a large driver for the heavy emitters, it could potentially be a larger driver for countries in the Middle East that are showing large, positive changes in population over time.

Future climate targets will need to incorporate drivers of emissions from all countries, not just drivers that have guided the paths of the heavy emitters. This can be seen even in the few study countries that were followed. While Germany's emissions are currently being driven down by a decrease in energy intensity, Saudi Arabia's are being driven up by an increase in population and Ethiopia's are being driven up by an increase in carbon intensity.

Countries in Cluster 3 are currently being grouped in a small cluster. However, if the drivers of emissions such as wealth and energy intensity continue to rise, these countries could begin contributing more to global emissions totals in even different ways than Cluster 1 and Cluster 2. One climate target might not capture just how different these drivers are as global emissions continue to rise.

Wealth and Population, All Countries by Cluster

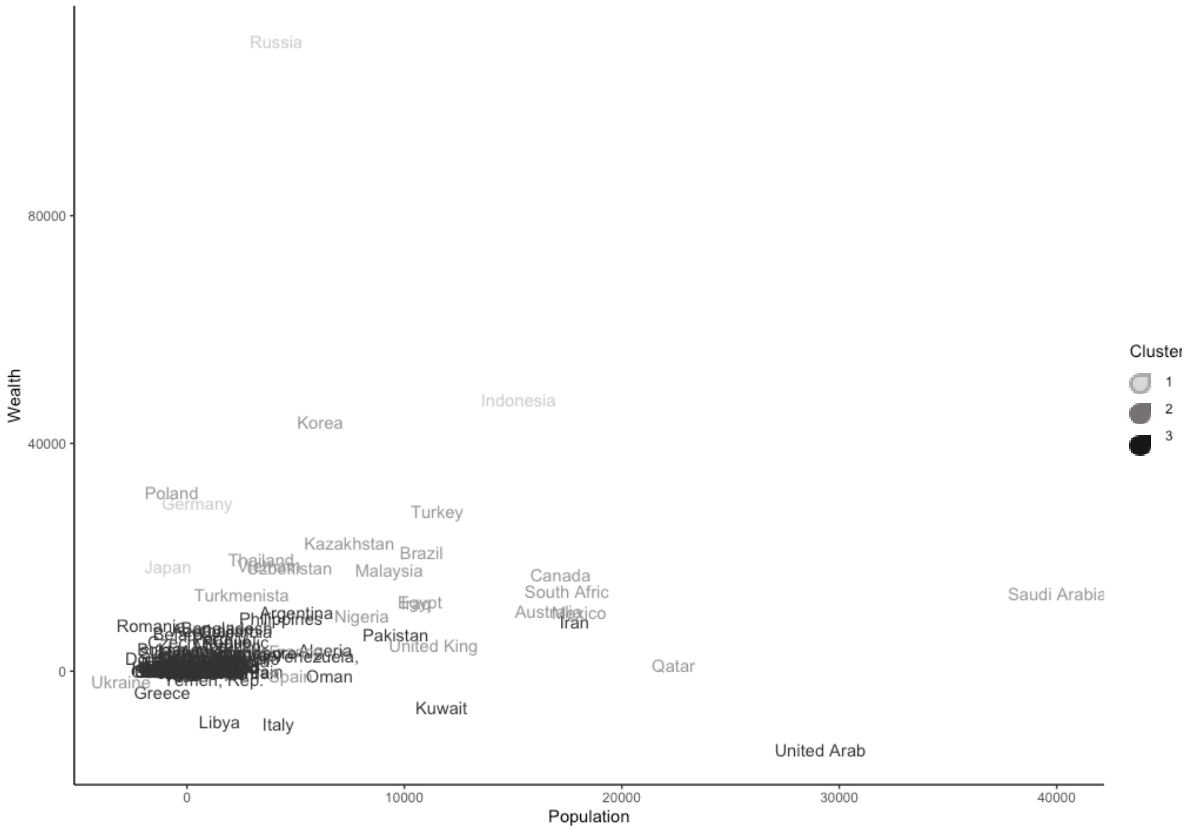


Figure 7: Clustered countries compared on an axis showing change in wealth for 2005-2015 and an axis of change in population intensity over 2005-2015. Cluster 1 is shown in the lightest text, while Cluster 3 is shown in the darkest text.

4.5 Sensitivity Analysis

Sensitivity analysis reveals that clustering is most sensitive to change in wealth. Removing the wealth index decreases the Dunn Index by 54.3%. When wealth is removed, Cluster 1 changes the most in terms of country membership. While the complete cluster model contains Germany, Indonesia, Japan, and Russia in Cluster 1, the model with wealth removed contains more countries in Cluster 1: Iran, Qatar, United Arab Emirates, and Saudi Arabia. Removing the wealth factor completely changes the members of this cluster, demonstrating its sensitivity in this model. Removing wealth causes both the TSS and BSS values to decrease.

This means that the total distance of the model as well as distances between clusters has gone down, implying that wealth is a driver that effectively separates countries and clusters from each other. This low TSS values causes the variance explained value to go up.

After a change in wealth, carbon intensity is the most sensitive Kaya component. Removing carbon intensity decreases the Dunn Index by 24.8%. Eleven countries are potentially misclassified in this model. Cluster 1 contains the same countries as the non-jack knife model while Cluster 2 contains similar countries except for Qatar, which moves from Cluster 2 to Cluster 3. The lowest Dunn Index shift results from removing population, indicating that population as not an overly sensitive factor. The model with population removed has a similar BSS to the model with nothing removed and explains only 5.2% more of the variance. Population growth does not have much effect on defining clusters because population growth is a common factor in all of the clusters. Conversely, wealth is not a common factor among all countries, and that disparity is reflected in the sensitivity of this variable. The results of the sensitivity analysis are shown in **Table 7** below.

Table 7: Sensitivity Analysis measures are shown such as Dunn Index, membership in Cluster 1, membership in Cluster 2, membership in Cluster 3, Number of negative silhouette widths, Percent variance explained, Between Cluster Sum of Squares (BSS) and Total Sum of Squares (TSS). Percent Variance is calculated by dividing TSS by BSS and multiplying by 100.

Factor Removed	Dunn Index	Cluster 1	Cluster 2	Cluster 3	Negative Silhouette Values	Percent Variance	BSS	TSS
None	0.9790	4	23	146	14	57.5	$3.28x10_{10}$	$5.70\ x10_{10}$
Population	0.9726	3	19	151	8	62.7	$3.22x10_{10}$	$5.13x10_{10}$
Wealth	0.4471	4	10	159	5	67.8	$2.34x10_{10}$	$3.45x10_{10}$
Energy Intensity	0.7361	1	21	151	6	63.0	$2.06x10_{10}$	$3.27x10_{10}$
Carbon Intensity	0.6620	4	22	147	11	63.4	$3.19x10_{10}$	$5.03x10_{10}$

Chapter 5: Conclusions & Future Research

One of the 10 focal countries in this analysis is Germany. The pattern of Germany's emissions can be described as decreasing, only two of the 10 countries in this study to do so. This emissions decrease is primarily from a decrease in emissions from gas fuel emissions, as well as a slower decrease in liquid and solid fuel emissions. Germany's decreasing emissions are driven primarily by a decrease in energy intensity. Germany clustered in Cluster 1, which was characterized by an increase in wealth and a decrease in energy intensity and likely related to the shifting of the economy from manufacturing to service based. The decrease in solid and liquid fuel emissions paired with the decrease in energy intensity over time depicts a country that will likely continue to decrease emissions over time.

Comparatively, another one of these 10 focal countries is Saudi Arabia. The pattern of Saudi Arabia's emissions is increasing, with one of the largest changes over time of total emissions in this study besides the heavy emitters. This is largely driven by an increase in emissions from liquid emissions over time. The drivers of emissions for Saudi Arabia are an increase in energy intensity, and an increase in population over time. Saudi Arabia clustered in Cluster 2, where it stands out for a large growth in energy intensity and growth in population from the other members of the cluster. Unlike the case of Germany, Saudi Arabia's positive trending patterns and drivers do not indicate a peak anytime soon in emissions.

Finally, consider the case of Ethiopia. Ethiopia has an emissions pattern that is positively trending over time, due to emissions from primarily liquid fuel usage. The main drivers of emissions are a growth in carbon intensity, primarily from agriculture, a similar growth in population, and a negative energy intensity over time. It appears that Ethiopia's emissions will continue to trend upward over time, given the pattern of emissions from fossil fuels and the drivers increasing emissions outweighing the drivers decreasing them. Ethiopia clustered in Cluster 3, which was made up of similar developing countries. However, Ethiopia's government has been actively involved in global climate plans and initiated the Climate-Resilient Green Economy in 2011. This plan strives to develop Ethiopia into a middle-income economy, withstand the impacts of climate, and install expanded hydroelectric and wind power systems all without raising the net level of carbon emissions throughout this time (Gashaw et al., 2014).

These 3 cuntries (Germany, Saudi Arabia, and Ethiopia) represent different nations that will have to abide by a universal climate goal. A single global climate goal would have to encompass the emissions patterns and drivers of countries like Saudi Arabia, which is experiencing large population growth driving emissions; countries like Germany, which is experiencing a decline in a emissions over time and a decreasing energy intensity over time; as well as countries like Ethiopia, with signs of increasing emissions with development but aggressive policy to counteract such trends (Gashaw et. Al, 2014). The same global target for these three countries will also have to be reachable for the heavy emitters, like China, the United States, and India, as well as countries already starting to feel the effects of climate change, like the Marshall Islands.

The use of data analytics tools makes it possible to consider all countries at a high level while also looking more in depth into certain countries. These tools allowed for considering 10 countries specifically in a trend analysis and all countries in a decomposition and cluster analysis. The results from these analyses suggest that one climate path isn't truly effective for the many different types of countries emitting currently and those who will emit even more in the future.

Future research could consider a different wealth measurement than GDP in this same model to determine a more accurate measure of their economic situation. While GDP is used predominantly in the literature, it could be more insightful to use a different index, such as the Gross National Income, to measure economic progress over time.

Future research could also include altering the time frame used in this research study. In 2015 the Paris Agreement was reached, and the United Nations published the Sustainable Development Goals. This model could look very different with those initiatives in place guiding international and country-level policy and could include a variable that captures policy level. A previous study by Pani and Mukhopadhyay (2010) stated that income was the largest driver of emissions over the years 1993 – 2004. This study reached a similar conclusion, that wealth was the most sensitive driver of emissions in the model. Repeating this study for the years 2005 -2015 could continue this trend or show results that diverge from these.

This model is an example of a data-driven approach to climate solutions. Data analytics tools can derive deeper insights about past trends and build more accurate predictions of the future. This can determine if current policy is truly aligning with the underlying drivers and patterns of emissions at the country level. Policymakers can use these insights to develop more thorough solutions. At the global level, it can be used to determine what more effective climate targets would be, and which paths are most possible for individual countries to follow in the global mitigation effort.

References

Albrecht, J., Francois, D., & Schoors, K. (2002). A shapley decomposition of carbon emissions without residuals. *Energy Policy, 30*(9), 727-736. doi:10.1016/S0301-4215(01)00131-8

Allen, M.R., O.P. Dube, W. Solecki, F. Aragón-Durand, W. Cramer, S. Humphreys, M. Kainuma, J. Kala, N. Mahowald, Y. Mulugetta, R. Perez, M. Wairiu, & K. Zickfeld. (2018). Framing and Context. In *Global Warming of 1.5°C, An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty.* In Press.

Andres, R.J., Fielding, D.J., Marland, G., Boden, T.A., Kumar, N. & Kearney, A.T. (1999). Carbon dioxide emissions from fossil-fuel use, 1751–1950. *Tellus B, 51*, 759-765. doi:10.1034/j.1600-0889.1999.t01-3-00002.x

Ang, B. (2004). Decomposition Analysis for Policymaking in Energy. *Energy Policy, 32*(9), 1131-1139. doi: 10.1016/s0301-4215(03)00076-4

Ang, B. W., & Liu, F. L. (2001). A New Energy Decomposition Method: Perfect in Decomposition and Consistent in Aggregation. *Energy, 26*(6), 537-548. doi: 10.1016/s0360-5442(01)00022-6

Asumadu-sarkodie, S., & Owusu, P. A. (2016). Multivariate co-integration analysis of the kaya factors in ghana. *Environmental Science and Pollution Research International*, *23*(10), 9934-9943. doi: 10.107/s11356-016-6245-9

Birdsey, R., M. A. Mayes, P. Romero-Lankao, R. G. Najjar, S. C. Reed, N. Cavallaro, G. Shrestha, D. J. Hayes, L. Lorenzoni, A. Marsh, K. Tedesco, T. Wirth, & Z. Zhu. (2018). Executive summary. In *Second State of the Carbon Cycle Report (SOCCR2): A Sustained Assessment Report*, Washington, DC, USA, 21-40.

Boden, T.A., G. Marland, and R.J. Andres. (2010). Global, Regional, and National Fossil-Fuel CO2 Emissions. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tenn., U.S.A. doi: 10.3334/CDIAC/00001_V2010

Boer, P.D, & Rodrigues, J. F. D. (2019). Decomposition analysis: when to use which method? *Economic Systems Research, 32*(1), 1-28. doi:10.1080/09535314.2019.1652571

Charrad, M., Ghazzali,., Boiteau, V., & Niknafs, A. (2014). NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. *Journal of Statistical Software, 61*(6), 1-36. doi:10.18637/jss.v061.i06

Dueben, P. D., & Bauer, P. (2018). Challenges and design choices for global weather and climate models based on machine learning. *Geoscientific Model Development, 11*(10), 3999–4009. doi:10.5194/gmd-11-3999-2018

Eggleston, H. S., Buendia, L., Miwa, K., Ngara, T., & Tanabe, K. (2006). 2006 IPCC Guidelines for National Greenhouse Gas Inventories. Japan.

Etemad, B., Luciani, J., Bairoch, P., & Toutain, J.C. (1991). World energy production 1800-1985. Switzerland.

Fan, F., & Lei, Y. (2016). Decomposition analysis of energy-related carbon emissions from the transportation sector in Beijing. *Transportation Research Part D: Transport and Environment, 42,* 135-145. doi:10.1016/j.trd.2015.11.001

Fischer, E. M., & Knutti, R. (2015). Anthropogenic contribution to global occurrence of heavy precipitation and high temperature extremes. *Nature Climate Change, 5*(6), 560-564. doi:10.1038/nclimate2617

Friedlingstein, P., Jones, M. W., OSullivan, M., Andrew, R. M., Hauck, J., Peters, G. P., … Zaehle, S. (2019). Global Carbon Budget 2019. *Earth System Science Data*, *11*(4), 1783–1838. doi:10.5194/essd-11-1783-2019

Gashaw, T., Mebrat, W., Hagos, D., & Nigussie, A. (2014). Climate Change Adaptation and Mitigation Measures in Ethiopia. *Journal of Biology, Agriculture and Healthcare, 4*(15), 148–152. doi: 10.7176/JBAH

Gilfillan, D., Marland, G., Boden, T., Andres, R. (2019). Global, Regional, and National Fossil-Fuel CO2 Emissions. Carbon Dioxide Information Analysis Center at Appalachian State University. Boone, North Carolina.

Griffin, R. C. (1987). CO2 Release from Cement Production 1950-1985. Institute for Energy Analysis, Oak Ridge Associated Universities, Oak Ridge, Tennessee, USA. In G. Marland et al. (1989), Estimates of C02 Emissions from Fossil Fuel Burning and Cement Manufacturing.

Gurney, K. R., Mendoza, D. L., Zhou, Y., Fischer, M. L., Miller, C. C., Geethakumar, S., & de la Rue du Can, S. (2009). High Resolution Fossil Fuel Combustion CO2 Emission Fluxes for the United States. *Environmental Science & Technology, 43*(14), 5535–5541. doi:10.1021/es900806c

Hamilton, T. G. A., & Kelly, S. (2017). Low carbon energy scenarios for sub-Saharan Africa: An input-output analysis on the effects of universal energy access and economic growth. *Energy Policy*, *105*, 303-319. doi:10.1016/j.enpol.2017.02.012

Hansen, J., Sato, M., Ruedy, R., Lo, K., Lea, D., & Medina-Elizade, M. (2006). *Global temperature change. Proceedings of the National Academy of Sciences, 103*(39), 14288-14293.

Hatzigeorgiou, E., Polatidis, H., & Haralambopoulos, D. (2008). CO2 Emissions in Greece for 1990–2002: A Decomposition Analysis and Comparison of Results Using the Arithmetic Mean Divisia Index and Logarithmic Mean Divisia Index Techniques, *Energy 33*(3), 492-499. doi: 10.1016/j.energy.2007.09.014

IPCC. (2013). *IPCC Factsheet: What is the IPCC?* Retrieved from https://www.ipcc.ch/site/assets/uploads/2018/02/FS_what_ipcc.pdf

IPCC. (2019). *History of the IPCC.* Retrieved from https://www.ipcc.ch/about/history/

Irani, J., Pise, N., & Phatak, M. (2016). Clustering techniques and the similarity measures used in clustering: a survey. *International Journal of Computer Applications, 134*(7), 9-14. doi:10.5120/ijca2016907841

Jackson, R. B., Le Quéré, C., Andrew, R. M., Canadell, J. G., Korsbakken, J. I., Liu, Z., Peters, G. P., & Zheng, B. (2018). Global Energy Growth Is Outpacing Decarbonization. *Environmental Research Letters, 13*(12). doi:10.1088/1748-9326/aaf303

Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters, 31*(8), 651-666. doi:10.1016/j.patrec.2009.09.011

Kaya, Y. (1990). Impact of Carbon Dioxide Emission Control on GNP Growth: Interpretation of Proposed Scenarios Paper presented to the IPCC Energy and Industry Subgroup, Response Strategies Working Group, Paris.

Knutti, R., Rogelj, J., Sedláček, J., & Fischer, E. M. (2016). A scientific critique of the two-degree climate change target. *Nature Geoscience, 9*(1), 13. doi:10.1038/ngeo2595

Kodinariya, T. M., & Makwana, P. R. (2013). Review on determining number of Cluster in K-Means Clustering. *International Journal of Advance Research in Computer Science and Management Studies, 1*(6), 90–95. Retrieved from http://www.ijarcsms.com

Koesler, S., Swales, K., & Turner, K. (2016). International spillover and rebound effects from increased energy efficiency in Germany. *Energy Economics, 54*, 444–452. doi:10.1016/j.eneco.2015.12.011

Köne, Aylin Çiğdem, & Tayfun Büke. (2010). Forecasting of CO2 emissions from fuel combustion using trend analysis. *Renewable and Sustainable Energy Reviews 14*(9) 2906-2915.

Kott, P. (2001). The Delete-a-Group Jackknife. *Journal of Official Statistics, 17*(4), 521–526. Retrieved from http://www.jos.nu

Lamb, W. F., Steinberger, J. K., Bows-Larkin, A., Peters, G. P., Roberts, J. T., & Wood, F. R. (2014). Transitions in Pathways of Human Development and Carbon Emissions. *Environmental Research Letters 9*(1). doi: 10.1088/1748-9326/9/1/014011

Liao, C., Wang, S., Zhang, Y., Song, D., & Zhang, C. (2019). Driving forces and clustering analysis of provincial-level CO2 emissions from the power sector in China from 2005 to 2015. *Journal of Cleaner Production, 240.* doi:10.1016/j.jclepro.2019.118026

Le Treut, H., R. Somerville, U. Cubasch, Y. Ding, C. Mauritzen, A. Mokssit, T. Peterson & M. Prather. (2007). Historical Overview of Climate Change. In *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge, United Kingdom: Cambridge University Press.

Legány, C., Juhász, S., & Babos, A. (2006). Cluster validity measurement techniques. *Proceedings of the 5th WSEAS international conference on artificial intelligence, knowledge engineering and data bases* (388-393). World Scientific and Engineering Academy and Society (WSEAS) Stevens Point, Wisconsin, USA.

Lengyel, A., & Botta-Dukát, Z. (2019). Silhouette width using generalized mean-A flexible method for assessing clustering efficiency. *Ecology and Evolution, 9*(23), 13231–13243. doi:10.102/ece3.5774

Li, W., Ou, Q., & Chen, Y. (2014). Decomposition of china's CO2 emissions from agriculture utilizing an improved kaya identity. *Environmental Science and Pollution Research International, 21*(22), 13000-6. doi: 10.107/s11356-014-3250-8

Marland, G., & Rotty, R. M. (1984). Carbon dioxide emissions from fossil fuels: a procedure for estimation and results for 1950-1982. *Tellus B, 36*(4), 232–261. doi: 10.1111/j.1600-0889.1984.tb00245.x

Mitchell, B.R. (1983). *International Historical Statistics: The Americas and Australasia 1750-1988* (pp. 522-525). Gale Research Company, Detroit, United States.

Mitchell, B.R. (1992). *International Historical Statistics: Europe 1750-1988* (pp. 465-485). Stockton Press, New York, United States.

Mitchell, B.R. (1993). *International Historical Statistics: The Americas 1750-1988* (pp. 405-414). Stockton Press, New York, United States.

Mitchell, B.R. (1995). *International Historical Statistics: Africa, Asia and Oceania 1750-1988* (pp. 490-497). Stockton Press, New York, United States.

Nematollahi, O., Hoghooghi, H., Rasti, M., & Sedaghat, A. (2016). Energy demands and renewable energy resources in the Middle East. *Renewable and Sustainable Energy Reviews, 54,* 1172-1181. doi:10.1016/j.rser.2015.10.058

O'Mahony, T. (2013). Decomposition of Ireland's Carbon Emissions from 1990 to 2010: An Extended Kaya Identity. *Energy Policy (59),* 573–581. doi: 10.1016/j.enpol.2013.04.013

Pani, R., & Mukhopadhyay, U. (2010). Identifying the major players behind increasing global carbon dioxide emissions: A decomposition analysis. *Environmentalist, 30*(2), 183-205. doi: 10.107/s10669-010-9256-y

Pui, K. L., & Othman, J. (2019). The influence of economic, technical, and social aspects on energy-associated CO2 emissions in Malaysia: An extended Kaya identity approach. *Energy, 181*, 468–493. doi: 10.1016/j.energy.2019.05.168

Rahman, S. M., Khondaker, A. N., Hasan, M. A., & Reza, I. (2017). Greenhouse gas emissions from road transportation in Saudi Arabia - a challenging frontier. *Renewable and Sustainable Energy Reviews*, *69*, 812–821. https://doi.org/10.1016/j.rser.2016.11.047

Raftery, A. E., Zimmer, A., Frierson, D., Startz, R., & Liu, P. (2017). Less Than 2 °C Warming by 2100 Unlikely. *Nature climate change, 7,* 637–641. https://doi.org/10.1038/nclimate3352

Ryan, M., Harmon, M., Birdsey, R., Giardina, C., Heath, L., Houghton, R., … Skog, K. (2010). A Synthesis of the Science on Forests and Carbon for U.S. Forests. Retrieved from https://www.fs.fed.us/rm/pubs_other/rmrs_2010_ryan_m002.pdf

Sippel, S., Meinshausen, N., Fischer, E. M., Székely, E., & Knutti, R. (2020). Climate change now detectable from any single day of weather at global scale. *Nature Climate Change, 10*(1), 35–41. doi:10.1038/s41558-019-0666-7

United Nations. (2009). 2007 Energy Statistics Yearbook. United Nations Department for Economic and Social Information and Policy Analysis, Statistics Division, New York.

United Nations. (2019). Department of Economic and Social Affairs, Population Division. *World Population Prospects 2019.*

U.S. Geological Survey. (2009). Metals and minerals: U.S. Geological Survey Minerals Yearbook 2007, 1. doi:10.3133/mybvI.

World Bank (2015). World Bank, World Development Indicators. *Energy intensity level of primary energy (MJ/$2011 PPP GDP)*.

World Bank (2015). World Bank, World Development Indicators. *GDP, PPP (constant 2011 international $)*

World Bank (2020). Data Help Desk. What is an "international dollar"? Retrieved from https://datahelpdesk.worldbank.org/knowledgebase/articles/114944-what-is-an-international-dollar

Yu, S., Wei, Y.-M., Fan, J., Zhang, X., & Wang, K. (2012). Exploring the regional characteristics of inter-provincial CO2 emissions in China: An improved fuzzy clustering analysis based on particle swarm optimization. *Applied Energy*, *92*, 552-562. doi:10.1016/j.apenergy.2011.11.068

Zhou, N., Levine, M. D., & Price, L. (2010). Overview of current energy-efficiency policies in China. *Energy Policy*, *38*(11), 6439–6452. doi:10.1016/j.enpol.2009.08.015

Appendix

Table A1. Table of Trend Analysis Results. Beta Coefficients are found in the fitted quadratic function, Adjusted R^2 Value measures explained variance, and 2050 projected value is measured in metric tons of C. *Values were below zero by 2050.

Country	Fuel Type	Beta$_1$	Beta$_2$	Adjusted R^2	2050 Value (C)
Argentina	total	2.57×10^{5}	-6.37×10^{1}	0.86	1.24×10^{8}
	gas	-1.06×10^{3}	3.63×10^{1}	0.80	
	liquid	2.17×10^{5}	-5.38×10^{1}	0.64	
	solid	-1.90×10^{4}	4.74×10^{0}	0.08	
China	total	3.67×10^{7}	-9.09×10^{3}	0.97	1.45×10^{10}
	gas	-1.04×10^{6}	2.59×10^{2}	0.99	
	liquid	-8.96×10^{5}	2.26×10^{2}	0.99	
	solid	3.63×10^{7}	-9.01×10^{3}	0.96	
Ethiopia	total	-1.19×10^{5}	2.97×10^{1}	0.96	2.00×10^{7}
	liquid	-4.65×10^{4}	1.16×10^{1}	0.90	
	solid	-1.52×10^{4}	3.78×10^{0}	0.97	
Germany	total	-1.05×10^{6}	2.60×10^{2}	0.64	0*
	gas	2.08×10^{5}	-5.20×10^{1}	0.76	
	liquid	-5.80×10^{5}	1.44×10^{2}	0.66	
	solid	-6.66×10^{5}	1.66×10^{2}	0.08	
Iceland	total	-7.36×10^{3}	1.83×10^{0}	0.75	2.46×10^{5}
	liquid	-8.58×10^{3}	2.13×10^{0}	0.64	
	solid	-6.57×10^{2}	1.63×10^{1}	0.05	
India	total	-2.07×10^{5}	5.90×101	0.98	2.63×10^{9}
	gas	7.10×10^{5}	-1.76×10^{2}	0.74	
	liquid	-8.61×10^{5}	2.16×10^{2}	0.83	
	solid	-1.19×10^{5}	3.46×10^{1}	0.97	
Marshall Islands	total	3.29×10^{2}	-8.16×10^{2}	0.93	6.46×10^{4}
	gas	3.29×10^{2}	-8.16×10^{2}	0.93	
Saudi Arabia	total	-3.03×10^{6}	7.56×10^{2}	0.89	8.57×10^{8}
	gas	-1.91×10^{6}	4.75×10^{2}	0.80	
	liquid	-1.15×10^{6}	2.87×10^{2}	0.88	
Thailand	total	8.89×10^{4}	-2.17×10^{1}	0.90	8.45×10^{7}
	gas	2.64×10^{5}	-6.54×10^{1}	0.92	
	liquid	-3.07×10^{5}	7.64×10^{1}	0.30	
	solid	2.87×10^{5}	-7.12×10^{1}	0.89	
United States	total	-7.10×10^{6}	1.76×10^{0}	0.82	0*
	gas	-1.84×10^{6}	4.59×10^{2}	0.96	
	liquid	-7.24×10^{6}	1.80×10^{3}	0.93	
	solid	2.58×10^{6}	-6.46×10^{2}	0.91	

Table A2: All Countries used in cluster analysis listed with LMDI values for 4 Kaya components as well as final cluster placement

Country	LMDI (pop.)	LMDI (wealth)	LMDI (EI)	LMDI (CI)	Cluster
Afghanistan	321.96	520.09	725.62	537.39	3
Albania	-79.30	497.30	-474.91	155.91	3
Algeria	6360.35	3769.48	7801.81	-5340.59	3
Angola	2571.10	1856.01	-1685.00	1456.71	3
Antigua and Barbuda	19.70	-8.79	12.66	8.43	3
Argentina	5039.23	10399.75	-2902.98	-1749.37	3
Armenia	-23.22	526.37	-251.77	-132.69	3
Australia	16638.71	10498.74	-17197.19	-5589.72	2
Austria	922.10	1217.01	-2686.21	-2898.24	3
Azerbaijan	1171.43	7242.03	-7862.74	305.88	3
Bahamas, The	73.60	-69.32	152.68	-105.96	3
Bahrain	3057.97	114.97	-966.45	1336.08	3
Bangladesh	1841.00	7627.72	-1474.36	3282.83	3
Barbados	11.32	-8.30	-47.21	22.19	3
Belarus	-207.33	6651.28	-8031.71	1201.65	3
Belgium	1923.40	1978.99	-6142.61	-327.89	3
Belize	32.47	2.23	6.02	2.27	3
Benin	303.09	145.51	197.92	356.58	3
Bhutan	21.21	111.71	-81.85	131.94	3
Bolivia	703.86	1414.84	-109.98	128.05	3
Bosnia	-442.55	1421.35	1191.20	-1518.85	3
Botswana	212.48	376.33	-105.00	-116.47	3
Brazil	10791.51	20853.82	5950.45	5378.19	2
Brunei Daru	209.36	-164.10	120.82	398.53	3
Bulgaria	-826.53	3969.04	-3747.14	-265.53	3
Burkina Faso	166.06	128.33	-72.62	381.04	3
Burundi	22.72	2.31	-47.48	90.45	3
Cambodia	220.19	722.65	66.12	586.73	3
Cameroon	412.77	224.25	-449.30	965.84	3
Canada	17178.79	16864.75	-25615.74	-11534.54	2
Central African Republic	7.66	-13.33	15.99	5.68	3
Chad	80.19	33.01	-64.84	15.65	3
Chile	2055.50	5301.53	-2452.74	101.15	3
Colombia	2252.06	7075.28	-4684.36	4436.01	3
Comoros	10.00	2.13	0.04	-4.17	3
Congo, Dem.	167.61	200.62	-59.49	54.81	3
Congo, Rep.	150.88	81.52	227.92	153.64	3
Costa Rica	241.46	570.05	-396.62	-239.79	3
Cote d'Ivoire	597.26	470.17	-313.47	103.05	3
Croatia	-182.18	278.75	-836.51	-693.22	3
Cyprus	225.54	-138.93	-252.70	-205.92	3
Czech Republic	983.59	5161.86	-8035.98	-3568.97	3
Denmark	502.96	247.71	-2544.71	-2659.56	3
Dominica	0.00	7.12	9.09	-7.21	3
Dominican Republic	710.24	2305.08	-2117.50	588.00	3
Ecuador	1579.44	2273.60	907.10	-1169.70	3
Egypt, Arab	10720.98	12173.83	-9384.42	2271.39	2
El Salvador	79.70	306.93	-378.45	110.75	3
Equatorial	817.27	-247.21	-115.30	-864.76	3
Estonia	-138.63	806.21	-620.63	-166.81	3
Ethiopia	661.84	1720.98	-1660.29	1618.52	3

Fiji	23.76	71.24	110.37	83.63	3
Finland	546.15	144.27	-1440.48	-2478.56	3
France	5019.44	3616.15	-17424.32	-11603.66	2
Gabon	465.89	-21.99	297.64	-639.91	3
Gambia, The	34.26	-8.39	-0.62	28.75	3
Georgia	-88.41	1061.34	-0.88	325.53	3
Germany	467.93	29441.32	-46710.48	-2283.01	1
Ghana	738.84	1234.00	-558.45	1174.22	3
Greece	-1136.62	-3689.89	-613.65	-3807.76	3
Grenada	3.02	3.93	5.23	-0.18	3
Guatemala	824.79	593.12	457.32	-895.98	3
Guinea	138.43	92.41	-56.07	61.80	3
Guinea-Bissau	17.03	4.70	-9.87	6.14	3
Guyana	12.95	188.72	-126.09	82.42	3
Haiti	102.26	50.31	17.31	62.96	3
Honduras	476.71	357.66	-61.04	-101.04	3
Hong Kong	703.45	3277.72	-2802.08	-1507.13	3
Hungary	-430.62	1659.15	-2453.47	-2468.44	3
Iceland	64.40	45.65	222.26	-397.00	3
Indonesia	15249.49	47707.90	-37025.03	20574.04	1
Iran	17829.61	8590.61	15070.33	8089.39	3
Iraq	10535.74	11890.78	-446.89	-7226.76	2
Ireland	1255.49	2528.00	-5061.60	-738.16	3
Israel	3359.54	2852.73	-2724.15	-1008.61	3
Italy	4206.61	-9292.70	-18057.38	-15005.09	3
Jamaica	131.02	-112.40	-629.55	-177.49	3
Japan	-872.16	18367.14	-95981.37	57880.52	1
Jordan	3045.46	-246.10	-1156.35	-248.75	3
Kazakhstan	7474.00	22615.37	-3494.61	-9107.00	2
Kenya	899.43	825.06	-212.00	797.80	3
Kiribati	3.10	0.08	-4.55	2.38	3
Korea, Republic	6132.62	43744.44	-12862.90	-561.08	2
Kuwait	11703.10	-6413.57	620.23	-6.42	3
Kyrgyz Republic	340.18	621.03	-39.75	413.35	3
Lao PDR	175.70	666.65	184.22	1070.40	3
Latvia	-236.74	533.22	-449.80	12.97	3
Lebanon	1802.25	671.98	-182.87	-41.14	3
Lesotho	18.44	239.62	-191.47	24.41	3
Liberia	83.13	67.05	-52.98	56.92	3
Libya	1507.78	-8781.94	-2360.07	10935.21	3
Lithuania	-483.64	1376.06	-1684.65	571.00	3
Luxembourg	602.99	138.30	-1272.01	-92.76	3
Macao SAR	117.55	238.45	-171.65	-113.35	3
Madagascar	195.98	-4.91	49.83	281.16	3
Malawi	79.58	77.56	-195.85	100.70	3
Malaysia	9302.49	17778.57	-11877.88	-1191.64	2
Maldives	87.05	78.29	30.16	-4.50	3
Mali	154.05	42.63	-132.13	549.45	3
Malta	39.87	174.50	-393.68	-114.69	3
Marshall Islands	1.24	0.93	16.86	-11.04	3
Mauritania	172.74	104.69	-34.63	113.20	3
Mauritius	30.39	387.54	-200.78	32.85	3
Mexico	18038.17	10387.56	-24404.23	565.71	2
Micronesia	1.00	-1.49	9.71	-3.22	3
Moldova	-28.69	485.98	-509.90	67.13	3
Mongolia	687.50	2519.62	-1194.88	1990.71	3

Morocco	1875.65	4515.31	-2483.58	322.61	3
Mozambique	277.79	430.34	-279.46	842.73	3
Myanmar	327.87	3596.70	-2669.52	1677.84	3
Namibia	149.00	255.68	-119.86	173.18	3
Nepal	58.87	465.52	-215.21	554.63	3
Netherlands	1656.81	3467.36	-9258.88	1566.52	3
New Zealand	1033.63	893.82	140.34	-1850.65	3
Nicaragua	176.56	347.24	-102.72	-144.55	3
Niger	131.89	58.06	-1.73	168.65	3
Nigeria	8040.79	9755.51	-9004.03	-6081.10	2
North Macedonia	22.61	756.45	-890.20	-1032.46	3
Norway	1412.61	239.17	-219.63	-148.64	3
Oman	6557.12	-700.28	5673.88	-1917.69	3
Pakistan	9599.67	6384.32	-7003.80	5121.86	3
Palau	-6.37	10.57	-11.86	8.66	3
Panama	421.17	1336.58	-908.70	313.01	3
Papua New Guinea	350.61	596.07	-319.68	211.32	3
Paraguay	187.52	434.46	-227.04	278.24	3
Peru	1082.95	5747.76	188.66	-2815.99	3
Philippines	4323.68	9285.28	-6023.87	4047.08	3
Poland	-703.12	31316.74	-29167.09	-6231.07	2
Portugal	-208.10	40.37	-2556.68	-1630.27	3
Qatar	22382.73	1088.09	-2584.20	-5639.96	2
Romania	-1643.79	8155.63	-9358.01	-3925.69	3
Russia	4109.08	110657.09	-70641.95	-22615.72	1
Rwanda	49.60	100.83	-89.48	57.05	3
Samoa	4.00	0.48	14.71	0.81	3
Sao Tome and Principe	6.09	6.96	-5.28	2.23	3
Saudi Arabia	40028.82	13593.05	11427.42	2850.70	2
Senegal	595.37	244.95	31.12	443.22	3
Seychelles	10.40	61.87	-144.45	18.18	3
Sierra Leon	50.75	46.76	-65.16	109.65	3
Singapore	3254.51	3312.57	-4285.68	6251.67	3
Slovak Republic	65.50	3399.28	-4980.77	-735.53	3
Slovenia	144.10	292.78	-855.24	-485.75	3
Solomon Islands	11.93	10.54	-18.95	4.47	3
South Africa	17471.99	14045.04	-18937.91	296.43	2
Spain	4772.34	-875.19	-16599.70	-15252.17	2
Sri Lanka	289.90	2335.29	-1579.51	1155.25	3
St. Kitts	4.81	3.80	-12.82	14.21	3
St. Lucia	9.91	6.06	1.41	-5.39	3
Sudan	933.48	897.36	-1647.41	2224.45	3
Suriname	51.35	87.50	-110.58	6.73	3
Sweden	938.30	1454.30	-3734.57	-2281.58	3
Switzerland	1219.27	796.39	-2778.58	-796.21	3
Tajikistan	213.64	427.84	-500.39	558.79	3
Tanzania	692.22	750.30	-513.49	1104.36	3
Thailand	3419.66	19596.92	-1083.19	-6181.97	2
Togo	138.95	129.11	-16.58	102.77	3
Tonga	0.00	3.18	-2.67	1.50	3
Trinidad	637.10	2027.72	-472.84	25.97	3
Tunisia	722.96	1483.38	-224.79	60.65	3
Turkey	11506.14	28060.90	-12008.72	3137.69	2
Turkmenistan	2522.33	13415.52	-10080.20	192.94	2
Uganda	309.84	314.82	-104.65	344.13	3
Ukraine	-3030.10	-1872.62	-27801.31	-4816.95	2

United Arab Emirates	29130.17	-13724.59	6070.92	-33.86	3
United Kingdom	11341.18	4650.01	-42068.77	-11430.71	2
Uruguay	45.39	728.08	127.56	-642.49	3
Uzbekistan	4715.42	18230.81	-27327.31	621.18	2
Vanuatu	6.41	1.41	5.26	6.92	3
Venezuela	5929.98	2713.98	-10097.59	3629.98	3
Vietnam	3784.54	18642.88	-443.53	2706.07	2
Yemen, Rep.	1231.97	-1399.69	-1970.58	266.64	3
Zambia	261.61	337.22	-308.98	321.09	3
Zimbabwe	422.62	752.81	-537.35	-218.34	3

Vita

Megan MacDonald received her bachelor's degree from Boston College in 2017 in Geological Science. Her undergraduate research studied the effects of induced seismicity in wastewater injection wells in the Midwest and was presented at the 2017 Geological Society of America Meeting. During this time, she pursued geologic field mapping at the South Dakota School of Mines and The University of Western Australia.

After receiving this degree, Megan returned home to the Great Plains to assist with perennial agriculture and sustainable food systems research at the Land Institute in Salina, Kansas.

Megan began completing her master's degree in 2019 in data analytics at Appalachian State University. Her research besides this project involves studying the effects of climate change on honeybee disease populations with the Appalachian State Center for Analytics Research and Education. This research was presented at the 16th Annual Southern Appalachian Honeybee Research Consortium and the 2020 American Beekeeping Federation Conference.

Megan is passionate about using data for good to drive positive change for people, animals, and the environment.