**Atharva Rajadnya, Aleksi Knepp, and Yixuan Li**

**ACE 592: SAE**

**Dr. Jared Hutchins**

**6 May 2023**

**ACE 592 Final Project**

**SECTION I: Kuznets Curve for Various Countries**

The provided code imports data from a CSV file named 'co2\_full.csv' and extracts relevant columns including country, year, code, population, GDP, and CO2 emissions per capita. The dataset is cleaned by removing duplicates and missing values. The 'country' column is renamed to 'Country' for consistency. A new column 'gdp\_per\_capita' is created by dividing GDP by population.

The Kuznets curve is a concept in economics that suggests that as an economy develops, income inequality initially increases before decreasing. This theory has been applied to environmental issues, such as carbon emissions, to suggest that as an economy develops, carbon emissions per capita first increase and then decrease after reaching a certain level of development.

The code then creates a Kuznets curve for 20 countries including the United States, China, Brazil, India, Russia, South Africa, Nigeria, Sweden, Japan, Mexico, South Korea, France, Argentina, Germany, United Kingdom, Turkey, Indonesia, Italy, Egypt, and Thailand. For each country, data is filtered for years after 1960, and a scatter plot is created with GDP per capita on the x-axis and CO2 emissions per capita on the y-axis. The resulting plot shows the relationship between economic development and carbon emissions for each country.

The plot reveals some interesting patterns. For example, the United States has a relatively high level of GDP per capita and CO2 emissions per capita compared to other countries, indicating that the country has reached a relatively high level of development. China, on the other hand, has a lower level of GDP per capita but higher CO2 emissions per capita, suggesting that the country is still in the early stages of development in terms of carbon emissions.

Overall, the code provides a useful analysis of the Kuznets curve for 20 countries, highlighting the relationship between economic development and carbon emissions. The plot can be used to inform policy decisions regarding carbon emissions reduction and sustainable development.

**Figure 1: Kuznets Curve (20 Countries)**

A picture containing text, diagram, line, map

Description automatically generated

The Kuznets curve for the 20 countries mentioned suggests that there is a positive relationship between economic development, as measured by GDP per capita, and carbon emissions per capita. However, the relationship is not linear, and there are some variations in the curve's shape for different countries.

For example, for some countries such as the United States, Germany, and Sweden, the curve appears to reach a turning point where the relationship between economic development and carbon emissions becomes negative. This suggests that these countries have reached a certain level of development where they are able to reduce their carbon emissions while continuing to grow economically.

On the other hand, for countries such as China, India, and Indonesia, the curve appears to be still increasing, suggesting that they are still in the early stages of development in terms of carbon emissions, and their emissions may continue to increase as they develop economically.

There are also some countries, such as Brazil and Mexico, where the curve shows a relatively flat relationship between economic development and carbon emissions. This suggests that these countries may have adopted more sustainable development practices that have allowed them to maintain their emissions at a lower level than what is typical for their level of economic development.

**Figure 2: Individual Kuznets Curves**

**(next page due to size constraints)**

The subplots are arranged in a 5x4 grid, with each row representing a different country. The title of each subplot indicates which country it represents. The figure's subtitle indicates that the subplots illustrate the relationship between CO2 emissions per capita and GDP per capita for each country.

Overall, this figure provides a more detailed look at the relationship between economic development and carbon emissions for each of the 20 countries. It allows us to compare the patterns of development for each country and observe any unique features that may not be apparent when looking at the Kuznets curve.

**A picture containing text, diagram, line, map

Description automatically generated**

**Figure 3: Co2 Emissions Over Time**

A picture containing text, diagram, plot, line

Description automatically generated

This plot allows us to visualize how each country's CO2 emissions per capita have changed over time and compare the trends across countries. It can also help identify any global trends in CO2 emissions over time. We can see that some countries like USA have a very high value of average CO2 emissions per capita, while some countries, although rising, have very low values.

Overall, this plot provides a useful visualization of how CO2 emissions per capita have changed over time across multiple countries, which can be valuable for understanding the patterns and trends of carbon emissions globally.

**Figure 4: Individual Kuznets Curves (Time On X Axis)**

The subplots allow us to examine each country's data in detail, including identifying any trends or patterns that may not be visible in the aggregate plot. It provides a useful visualization of how CO2 emissions per capita have changed over time for each of the included countries, which can be valuable for understanding the patterns and trends of carbon emissions across the globe.

**A picture containing text, diagram, line, map

Description automatically generated**

**Figure 5: Kuznets CO2 Maximums Over Time**

A graph of co2 emissions per capita by country

Description automatically generated with low confidence

This plot is identifying the maximum CO2 emissions per capita for each country in the dataset and creating a bubble plot to visualize the data. The plot shows the relationship between the year, CO2 emissions per capita, and GDP (represented by the size of the bubble) for each country. The plot helps us to understand which countries have the highest CO2 emissions per capita, and the size of the bubble indicates their GDP. We can see that countries with higher GDP tend to have larger bubbles, and therefore higher CO2 emissions per capita.

The plot also highlights the trend of increasing CO2 emissions per capita over time for some countries, such as the United States (grey dot which has a peak of more than 20) and China. Overall, this plot emphasizes the need for countries to balance economic growth with reducing their carbon footprint to combat climate change.

The plot suggests that countries that reached their peak earlier in terms of GDP per capita tend to have higher maximum CO2 emissions per capita than countries that are peaking now. This could be due to several reasons:

Technological advancements: Earlier peaking countries had more time to develop and adopt new technologies that would allow them to increase their economic growth and productivity. However, these technologies may have been less environmentally friendly and more reliant on fossil fuels than the technologies currently available.

Environmental awareness: Countries that are peaking now may be more environmentally conscious and have implemented policies and practices that prioritize sustainable development and reduce carbon emissions.

Resource availability: Countries that are peaking now may have access to renewable energy sources that earlier peaking countries did not have. These renewable energy sources may be more cost-effective and environmentally friendly than traditional fossil fuels.

**Figure 6: Kuznets CO2 Maximums by GDP per capita**

A picture containing text, software, screenshot

Description automatically generated

This is a bubble plot that shows the maximum CO2 emissions per capita for each country based on their GDP per capita. The size of each bubble represents the GDP per capita, while the color represents the country. The plot provides an easy-to-understand visualization of the relationship between a country's CO2 emissions, GDP per capita, and the year in which it peaked. The plot can help identify the countries with the highest CO2 emissions per capita and the most significant economic growth.

**Figure 7: Country-wise Total Damages from 1960**

A picture containing screenshot, colorfulness, rectangle, square

Description automatically generated

We calculated the integrated CO2 emissions for each country in the dataset from the year 1960 until the present year. The integrated CO2 emissions are calculated using the trapezoidal rule to estimate the area under the curve of the CO2 emissions over time.

The tree map displays the integrated CO2 emissions for each country, with the size of each rectangle representing the magnitude of the integrated CO2 emissions. The tree map allows us to compare the relative contributions of different countries to global CO2 emissions over time.

The plot shows that China, the United States, and Russia are the top three countries with the highest integrated CO2 emissions, followed by Germany and UK. This suggests that these countries have contributed the most to the overall increase in global CO2 emissions over the past 60 years.

We can also see that some smaller countries, such as South Korea and Turkey, have a relatively high contribution to the total amount of integrated CO2 emissions, due to their high per capita emissions and small populations.

From an economic viewpoint, we can say that the integrated CO2 emissions from 1960 are an indication of the environmental impact caused by each country. The higher the integrated CO2 emissions, the greater the environmental damage caused by a particular country. This can have economic consequences such as increased health care costs, damages to infrastructure and property, and losses in productivity and economic growth. Additionally, countries with high integrated CO2 emissions may face costs associated with adapting to climate change impacts, such as sea level rise, changes in precipitation patterns, and increased frequency and intensity of extreme weather events. Conversely, countries with lower integrated CO2 emissions may benefit economically from their lower environmental impact and may have a comparative advantage in sustainable production and green technologies.

**SECTION II: USA CASE STUDY**

One way to further understand the Kuznets Curve is to perform a case study on one country’s curve. This section will provide a case study of the United States and will analyze the U.S.’s movement along the curve using several different metrics. While Kuznets is traditionally modeled as emissions over GDP, this analysis will measure emissions over time; GDP is almost strictly increasing over time, and having a time metric allows for a more comprehensive analysis of the data. Additionally, the time period for this analysis is 1950-2022 to allow for the most complete data. Below is a simple “adjusted” (emissions over time) Kuznets curve for the U.S.

**Figure 8: Standardized CO2 Emissions in the United States**

A picture containing text, line, screenshot, plot

Description automatically generated

This section seeks to understand the fluctuations above. The next graph will include some significant historical, political, or economic changes in the U.S. over this time period to contextualize the curve. First, the beginning of the JFK administration is marked. JFK’s campaign slogan, “A Time for Greatness,” emphasized his commitment to enhancing production in America (Kennedy). This was enough time after WWII that patriotism had started to wane, so Kennedy aimed to enhance patriotism by this method. He also cut taxes for the middle class, leading to increased demand for goods – goods that were more often produced in the U.S. This led to increased emissions due to increased production and large economic growth throughout the 1960s (Giglio). Next, the signing of the Clean Air Act is marked. This was the first significant policy in the United States of its kind; it is estimated to have saved thousands of lives and to have greatly improved air quality in the U.S. (EPA). After its induction, there is a clear bit of volatility in emissions rather than strictly increasing emissions. The start of the Reagan administration is marked next. Reagan inherited a tumultuous economy, one with strained trade relationships and record inflation. Reagan worked to improve trade relationships and to increase imports rather than production, which is a possible explanation for the slow growth of emissions in this time period (Griswold)[[1]](#footnote-1). The next mark is the dotcom bubble, which was when big tech began to boom in the U.S. In the early days of the internet, a lot of technology was used and produced in the United States. The dotcom boom also connected the world in ways it had not been before (Hayes). The next mark is the Kyoto protocol, which Clinton tried and failed to enter. This had no major impact on U.S. emissions, but it would be remiss not to at least mention the first major international climate accord. Finally, Mass v. EPA is marked, which was a landmark Supreme Court decision that enabled the EPA to regulate GHG emissions. It also marks the decline of U.S. along the Kuznets curve (549 US 497).

**Figure 9: Standardized CO2 Emissions in the United States with Events**

A picture containing text, line, plot, screenshot

Description automatically generated

The next component of analysis is concerned with fuel and energy usage in the U.S. Below, it is clear that renewable usage in the U.S. begins to take off at the same time that CO2 emissions begins to decline. At the end of the period, renewable usage begins to outpace emissions from a standardized perspective (units are not constant; CO2 itself is standardized).

**Figure 10: Standardized CO2 Emissions and Renewables**

A picture containing text, screenshot, diagram, line

Description automatically generated

Below, fossil fuels consumption is included. Standardized fossil fuel consumption is fairly constant across the period, meaning that there is no clear relationship between fossil fuel consumption and emissions or renewable consumption.

**Figure 11: Standardized CO2 Emissions and Renewables/Fossil Fuels**

A picture containing text, line, diagram, plot

Description automatically generated

The next figure looks at year over year growth of CO2 emissions, renewable consumption, and fossil fuel consumption (as a %). There is not much to gain from this visualization, other than a striking matchup between CO2 change and renewable change. Interestingly, fossil fuel consumption growth is extremely variable over the period. This “nonsense” curve is relatively consistent with the previous figure, where there is not much to be learned from the fossil fuel consumption changes.

**Figure 12: Growth of CO2 Emissions and Energy Usage**

A picture containing text, line, font, screenshot

Description automatically generated

Finally, below is a view of all fuel types over the period. One can see the various renewables start to take off in the early 21st century, which is consistent with the overall renewable change. Nuclear energy has also increased rapidly, but appears to be stagnant at this point in time. All fossil fuel types are relatively constant over time, which could potentially mean that they will drop off should all renewable consumption continue to increase. It is important to note that this graph is standardized; all fossil fuels are measured in terawatt hours over 10 (there is no metric prefix for this value). So, renewable consumption will have to continually increase to meet electricity and energy demands in the U.S. However, if consumption trends do continue this way, the U.S. emissions will likely continue to decrease, completing the Kuznets curve.

**Figure 13: Energy Usage of All Sources over Time**

A picture containing text, line, diagram, plot

Description automatically generated

**SECTION III: TEXT ANALYSIS**

To explore the relationship between environmental emissions and political statements, text analysis techniques such as topic modeling, sentiment analysis, and word clouding are employed. The complete texts of State of the Union address from 1790 to 2018 are gathered for this purpose. The State of the Union is an annual speech given by the President of the United States before a joint session of Congress, where a review of the previous year is presented and a legislative agenda for the upcoming year is outlined. The sentiments and vocabulary used often reflect the nation's economic development, political ideology, and various aspects of American life.

It is necessary to differentiate between Presidents with the same last name, such as Theodore Roosevelt and Franklin D. Roosevelt, George H. W. Bush and George W. Bush, Andrew Johnson and Lyndon B. Johnson, John Adams and John Quincy Adams, William Henry Harrison, and Benjamin Harrison. To achieve this, the indexes are renamed, and a new column is created to categorize each President by their political party. Prior to the main analysis, stop words are removed from the text, and all words are transformed into lowercase and tokenized.

**The Valence Aware Dictionary and Sentiment Reasoner (VADER)** is a rule-based sentiment analyzer employed to extract sentiment scores for each State of the Union by year. As shown in Plot XX, the average sentiment score remains stable around 1 from 1790 to 2018. However, notable dips in sentiment are observed in 1790, during the 1940s, and in the first decade of the 21st century. This suggests that in most years, Presidents convey a highly positive attitude and promising statements in their State of the Union addresses. In contrast, during these "slump" years, the sentiment is markedly negative. To verify the accuracy of these observations, the mean sentiment score is calculated for each President. It is discovered that the most significant dips in sentiment mainly occurred during the administrations of Franklin D. Roosevelt, George W. Bush, and George Washington.

**Figure 14: Sentiment of SOTUs since 1800**

Chart

Description automatically generated

Plot xx Sentiment Score by Years and Presidents

The next step involves conducting topic modeling to perform further text analysis and determine whether environment-related words are frequently used in State of the Union addresses. **Latent Dirichlet Allocation (LDA)** is the algorithm employed for topic modeling, so named because it is a generative model with a Dirichlet-distributed prior. LDA is an unsupervised machine learning method. The provided code utilizes LDA to identify the top 10 topics in a corpus of texts represented as a Bag of Words (BOW) corpus and displays the top 10 most probable words for each topic.

The results indicate that the topics generally revolve around themes such as time, war, newness, and greatness. However, without additional context, it is challenging to identify words related to the environment, such as "emission" or "pollution." It is also difficult to ascertain the specific meanings or interpretations of each topic. This suggests that environmental concerns may not be a high-priority issue in State of the Union addresses.

**Table 1: Topic Modeling**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Topic 0 | time | new | us | war | years | work | one | power | also | peace |
| Topic 1 | great | time | us | one | work | every | new | world | present | power |
| Topic 2 | great | present | one | state | new | time | years | war | every | system |
| Topic 3 | great | time | world | every | war | us | new | one | years | peace |
| Topic 4 | new | world | great | every | years | war | make | us | power | peace |
| Topic 5 | new | great | time | one | us | war | world | every | nations | years |
| Topic 6 | great | new | war | world | time | law | one | years | make | service |
| Topic 7 | great | war | time | one | world | new | present | make | every | nations |
| Topic 8 | one | great | us | law | present | state | act | new | world | general |
| Topic 9 | war | new | great | years | state | time | one | shall | every | power |

The final aspect of text analysis is **Word Clouding**, employed to identify representative and frequently used words in the State of the Union addresses. Word clouding is a visualization technique that displays word counts. Plot XX offers a comprehensive overview of the State of the Union, indicating that words such as "time," "law," "great," and "power" are the most common across all Presidents. This implies that the State of the Union primarily focuses on public issues and legislative regulations, while environmental concerns and CO2 emissions are not prominent topics in these statements.

This plot consists of four subplots arranged in a 2x2 grid, with each row representing a different President. Franklin D. Roosevelt, George W. Bush, and George Washington are selected due to their lower sentiment scores compared to other Presidents. The subplots for Franklin D. Roosevelt and George W. Bush reveal that "war" is the most frequent word in their State of the Union addresses. "War" is considered a negative sentiment word in VADER, which significantly lowers the mean sentiment scores of these two Presidents. In the context of their respective eras, Roosevelt's administration was marked by the Great Depression and the onset of World War II, leading to the frequent mention of "war" in his speeches. Similarly, during Bush's administration, military campaigns were initiated in Afghanistan and Iraq in response to terrorist attacks.

In contrast, the words featured in the addresses of Presidents Washington and Obama are more neutral, resulting in higher positive sentiment scores for them.

**Figure 15: Overall Word Cloud**

Text

Description automatically generated

**Figure 16: Word Clouds Examples**

A picture containing text, newspaper

Description automatically generated

In conclusion, the text analysis of State of the Union addresses provides some useful insights, but it does not reveal a significant relationship between environmental emissions and political statements. Environmental concerns may not be consistently prioritized in these speeches, as evidenced by the absence of prevalent environment-related words such as "emission" or "pollution." To better understand and verify the relationship between environmental emissions and political statements, it is crucial to gather a broader range of political and environment-related speeches and statements, as well as consider other types of data and analysis methods. This comprehensive approach would help to paint a clearer picture of how political discourse addresses and influences environmental issues.

**Works Cited**

549 US 497 (2007). “Massachusetts v. EPA.” Oyez, 2005, https://www.oyez.org/cases/2006/05-1120.

EPA. “40th Anniversary of the Clean Air Act.” EPA, Environmental Protection Agency, 2010, <https://www.epa.gov/clean-air-act-overview/40th-anniversary-clean-air-act>.

Giglio, James N. "John F. Kennedy as Domestic Leader: A Perspective on the Literature." in Kennedy: The New Frontier Revisited (Palgrave Macmillan, London, 1998). 222–255.

Griswold, Daniel. “Reagan Embraced Free Trade and Immigration.” CATO, 2004, <https://www.cato.org/commentary/reagan-embraced-free-trade-immigration>.

Hayes, Adam. “What Ever Happened to the Dotcom Bubble?” Investopedia, Investopedia, 18 July 2022, <https://www.investopedia.com/terms/d/dotcom-bubble.asp>.

Kennedy, John F. “THE DEMOCRATIC PARTY: A TIME FOR GREATNESS.” *JFK Presidential Library*, 2023, [https://www.jfklibrary.org/asset- viewer/archives/JFKCAMP1960/1029/JFKCAMP1960-1029-021](https://www.jfklibrary.org/asset-%20viewer/archives/JFKCAMP1960/1029/JFKCAMP1960-1029-021). Accessed 2023.

1. I am not usually in the habit of citing things like the CATO Institute, but they had what I needed. [↑](#footnote-ref-1)