Global Greenhouse Gas Emissions

Amrit Bhat, Corbin Charpentier, Ernesto Cediel, Marie Hoeger

Motivation

The Arctic is melting, and the Siberian tundra is on fire. These are massive problems for many reasons, including the fact that the tundra is one of the biggest carbon dioxide sinks *in the world,* but only when it is frozen. In response, an Intergovernmental Panel on Climate Change (IPCC) convened in 2018 and found that the average global temperature on Earth must be capped at 1.5°C to avoid catastrophic changes that will transform life on Earth as we know it.

This turning point is sneaking up on us, and we must make drastic changes and reduce our greenhouse gas emissions–soon. **Our group's goal is to characterize where global greenhouse gas emissions come from to build a better understanding of what type of policies will be most effective in limiting greenhouse gas emissions**. We would like to understand which are the key factors (economical, trade, growth, agricultural, energy, water, sanitation, demographics, etc.) that are driving countries towards becoming better or worse in limiting their CO2 equivalent (CO2e) emissions.

Questions

Part 1: Characterizing global CO2e trends

- Has there been an increase in CO2e emissions in the past 50 years?
    - Test for global year-over-year increase. Create differences between each subsequent year pairs (t, t-1), of CO2 global emissions for all years of data, calculate **mean and confidence interval** for year differences between 1959-2020. Here we would test if the average of these global emissions are greater than 0. (Sample size =#ofyears-1.)
    - Use linear regression or other functions within a graph to find how it is increasing, exponentially, linearly, logarithmically?
- What are the global trends broken down by each type of CO2 equivalent (carbon dioxide, methane, nitrous oxide, sulfur hexafluoride, perfluorocarbon gasses, hydrofluorocarbon gasses)?
    - Use linear regression or other functions within a graph to find how it is increasing, exponentially, linearly, logarithmically?
- Where are we headed in the next 10, 20, 50 years if nothing changes?
    - We use extrapolation (Ceteris paribus) and graph 3 (good, same, bad) scenarios for the different years.

Part 2: We will explore how the panorama of CO2e looks like in the world by country

All of the analysis questions we will respond below will consider two target variables, C02e per capita and CO2e emissions as a % of the total CO2e emissions.

- **Are there observable patterns and differences amongst different groupings?**
    - Which countries have a specific increase pattern (exponential, linear)? Are there any relationships (features) between countries with this pattern?
        - Linear regression
    - What are the differences between developing and developed countries? What are the growth rates for each and which are the ones in each group that contributes the most and which are becoming a threat?
        - Linear regression, graphs, slopes.
    - Does the country income group affect the C02e emissions?

        - ANOVA test for CO2e emissions for the 4 income groups.
        - Using Radar plot analysis
    - Does being an island affect the CO2e emissions generated?
        - Two sample z-test
- **At a country level, are there any low hanging fruits available so that countries could improve their CO2e emissions in the short-run? (based of the countries we found in the previous question that contributed the most and/or are becoming a threat)**
    - For each country, which features correlate with CO2e emissions?
        - Compare various economic indicators with CO2e, e.g. GDP, logistical quality, transport methods (rail, road, air), gasoline price, active trade (imports and exports), **using regression analysis**.
        - Do conclusions change if we perform the analysis using *CO2e-per-capita* vs *percent-of-total*?

Data Source

Our main data source is the World Bank's suite of [world development indicators](),

The **World Bank World Development Indicators database** is a compilation of internationally comparable statistics about global development. It also contains the CO2e emissions (broken down by specific greenhouse gas) and the population for all countries.

The dataset characteristics are:

- Spans 50 years, though early data is sparser than recent (from 1960 to 2020)
- 263 countries and aggregations (encompassing more than one country)
- Comprised of 7890 rows and 404 columns

Our Plan

- Part 1:
    - Aggregate data for yearly global data
    - Run various analyses outlined in questions (part 1)
- Part 2:
    - Normalize CO2e per country based off of different factors
        - Per capita
        - Percentage of total global emissions
    - Choose year of analysis based on individual factor of analysis and the number of available data points
    - Apply linear regression on normalized CO2e measures vs. different metrics from the world indicators dataset (ex: GDP, amount of coastal area, access to electricity)
    - Look for features highly correlated to CO2e
    - Deep dive into a few features and outlier countries

Challenges and open questions

- If we decided to use linear regression to understand the effect of the different features to CO2e, we will have to check if the dataset complies with the multiple assumptions of Linear Regression (independence, linearity, normality and homoscedasticity), if not we would have to use alternate more robust models.

- Given the huge scope and millions of effects that CO2e can be associated with, we will have to be very cautious while handling analysis, especially, we must be very careful with regards to confounding factors and causation conclusions.
- We don’t yet know the minimum year for which we’ll reliably have CO2e data.