CS328 Writing Assignment

CO_2 and Greenhouse Gas Emissions Analysis

This notebook analyzes the **Our World in Data CO_2 and Greenhouse Gas Emissions** dataset. It includes key metrics such as annual CO_2 emissions, CO_2 emissions per capita, cumulative and consumption-based CO_2 emissions, along with other indicators related to greenhouse gas emissions and energy mix.

A codebook (`owid-co2-codebook.csv`) is provided alongside the dataset (`owid-co2-data.csv`). The codebook includes detailed descriptions and source information for each indicator. We'll use it to gain a better understanding of the dataset.

=== Data Sample ===

	country	year	iso_code	population	gdp	cement_co2	cement_co2_per_capita	co2	co2_growth_abs	co2_growth_prct	co2_including_luc	co2_including_luc_growth_abs	co2_including_luc_growth_prct	co2_including_luc_per_capita	co2_including_luc_
0	Afghanistan	1,750	AFG	2,802,560	None	0	0	None	None	None	None	None	None	None	
1	Afghanistan	1,751	AFG	None	None	0	None	None	None	None	None	None	None	None	
2	Afghanistan	1,752	AFG	None	None	0	None	None	None	None	None	None	None	None	
3	Afghanistan	1,753	AFG	None	None	0	None	None	None	None	None	None	None	None	
4	Afghanistan	1,754	AFG	None	None	0	None	None	None	None	None	None	None	None	

=== Codebook Sample ===

	column	description	unit	source
0	country	Country - Geographic location.	None	Our World in Data - Regions (2023)
1	year	Year - Year of observation.	None	Our World in Data - Regions (2023)
2	iso_code	ISO code - ISO 3166-1 alpha-3 three-letter country codes.	None	International Organization for Standardization - Regions (2023)
3	population	Population - Population by country, available from 10,000 BCE to 2100, based on data	people	Population based on various sources (2024) [https://ourworldindata.org/population-
4	gdp	Gross domestic product (GDP) - This data is adjusted for inflation and differences in tl	international-$ in 2011 prices	Bolt and van Zanden - Maddison Project Database 2023 [https://www.rug.nl/ggdc/his

Data Cleaning and Preprocessing

In this step we will:

- Compute the percentage of missing values for each column.
- Identify and drop columns that have more than 60% missing values.
- Remove rows that are missing crucial data (i.e., 'country', 'year', or 'co2').

This cleaning process will help us to ensure that the subsequent analysis—such as visualizing trends and performing statistical tests—is based on reliable data.

Columns to drop (>60% missing):

```
[
0 : "share_global_cumulative_other_co2"
1 : "share_global_other_co2"
2 : "other_co2_per_capita"
3 : "other_industry_co2"
4 : "cumulative_other_co2"
5 : "consumption_co2_per_gdp"
6 : "consumption_co2_per_capita"
7 : "trade_co2"
8 : "trade_co2_share"
9 : "consumption_co2"
```

10 : "energy_per_gdp"
11 : "co2_including_luc_per_unit_energy"
12 : "energy_per_capita"
13 : "primary_energy_consumption"
14 : "co2_per_unit_energy"
15 : "share_global_cumulative_flaring_co2"
16 : "share_global_flaring_co2"
17 : "flaring_co2_per_capita"
18 : "share_global_cumulative_gas_co2"
19 : "share_global_gas_co2"
20 : "gdp"
21 : "cumulative_flaring_co2"
22 : "flaring_co2"
23 : "co2_including_luc_per_gdp"
24 : "gas_co2_per_capita"
25 : "co2_per_gdp"
26 : "cumulative_gas_co2"
27 : "gas_co2"
]

Shape after cleaning: `(29137, 51)`

Analysing the Data

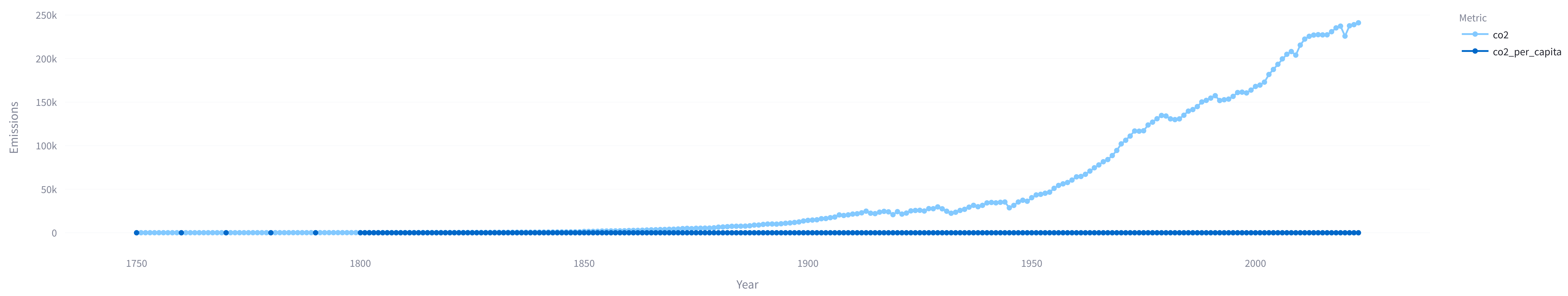
Global Emissions Trends

Here, we group the data by year to observe:

- **Total CO_2 Emissions:** The aggregated sum of CO_2 emissions across all countries per year.
- **Average CO_2 Emissions Per Capita:** The yearly average of per capita CO_2 emissions.

These trends will help us understand the overall direction of emissions over time.

Global CO_2 Emissions Trends



Hypothesis: Since the mid-20th century, global economic and technological advances have enabled a decoupling of per-person CO_2 emissions from total emissions growth—so that although total CO_2 continues to climb, average per-capita emissions have plateaued or even declined.

The two panels above settle this hypothesis. The **top panel** shows total global CO_2 emissions accelerating sharply after 1950, climbing from ~35 Gt to over 240 Gt by 2023. Yet the **bottom panel** reveals that average per-person emissions peaked around the 1950s–1970s at roughly 6–7 t CO_2/person and have since oscillated between 4.5 and 6.5 t. This divergence—soaring total emissions alongside a stable or gently declining per-capita curve—confirms that while the world's carbon footprint grows with population and economic scale, improvements in energy efficiency, shifts toward service economies, and the adoption of cleaner energy sources have partially offset per-person emissions. Thus, the data support our hypothesis of mid-century decoupling: humanity is emitting more CO_2 overall, but not at the same rate per individual.

Correlation Analysis of Emission Indicators

Next, we examine relationships among several key emission indicators. We select a subset of variables:

- **CO_2 emissions (co2)**
- **Cement CO_2 emissions (cement_co2)**
- **Coal CO_2 emissions (coal_co2)**
- **Oil CO_2 emissions (oil_co2)**
- **Methane emissions (methane)**
- **Nitrous oxide emissions (nitrous_oxide)**

We'll compute a Pearson correlation matrix and visualize it with a heatmap. This analysis may reveal how changes in one type of emission correlate with others.

=== Correlation Matrix ===

	co2	cement_co2	coal_co2	oil_co2	methane	nitrous_oxide
co2	1	0.8862	0.958	0.9681	0.9225	0.9518
cement_co2	0.8862	1	0.9068	0.7859	0.8156	0.8231
coal_co2	0.958	0.9068	1	0.8655	0.9139	0.9152
oil_co2	0.9681	0.7859	0.8655	1	0.8858	0.9377
methane	0.9225	0.8156	0.9139	0.8858	1	0.9804
nitrous_oxide	0.9518	0.8231	0.9152	0.9377	0.9804	1

Correlation Matrix of Selected Emission Indicators

The correlation matrix of key emission indicators reveals uniformly strong positive relationships, indicating that different sources of greenhouse gases tend to rise and fall together globally. Total CO_2 emissions correlate most closely with oil CO_2 ($r \approx 0.97$) and coal CO_2 ($r \approx 0.96$), reflecting the continued dominance of fossil fuels. Methane and nitrous oxide emissions also track closely with total CO_2 ($r \approx 0.92$ and $r \approx 0.95$, respectively) and exhibit an exceptionally high inter-gas correlation ($r \approx 0.98$), suggesting common agricultural and industrial drivers. Cement CO_2 shows slightly weaker—but still substantial—links to other sources ($r \approx 0.79$–0.91). Overall, these patterns underscore how economic activity, energy use, and land-use practices jointly drive multiple greenhouse-gas emissions, reinforcing the need for integrated mitigation strategies.

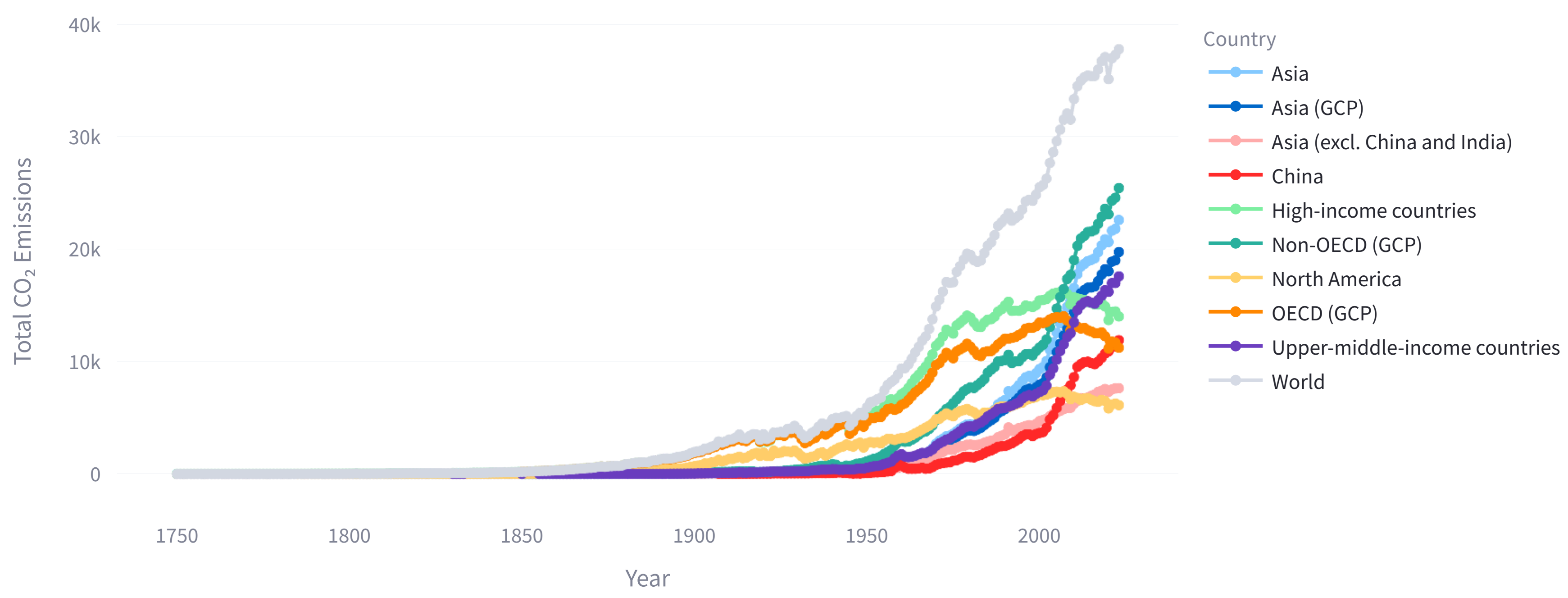
Country-Level Analysis

Hypothesis: Among the top 10 CO_2-emitting countries in the most recent year, total emissions have generally increased over time, but per capita emissions will show divergent trends—declining in some high-income nations due to efficiency improvements and rising in rapidly developing economies.

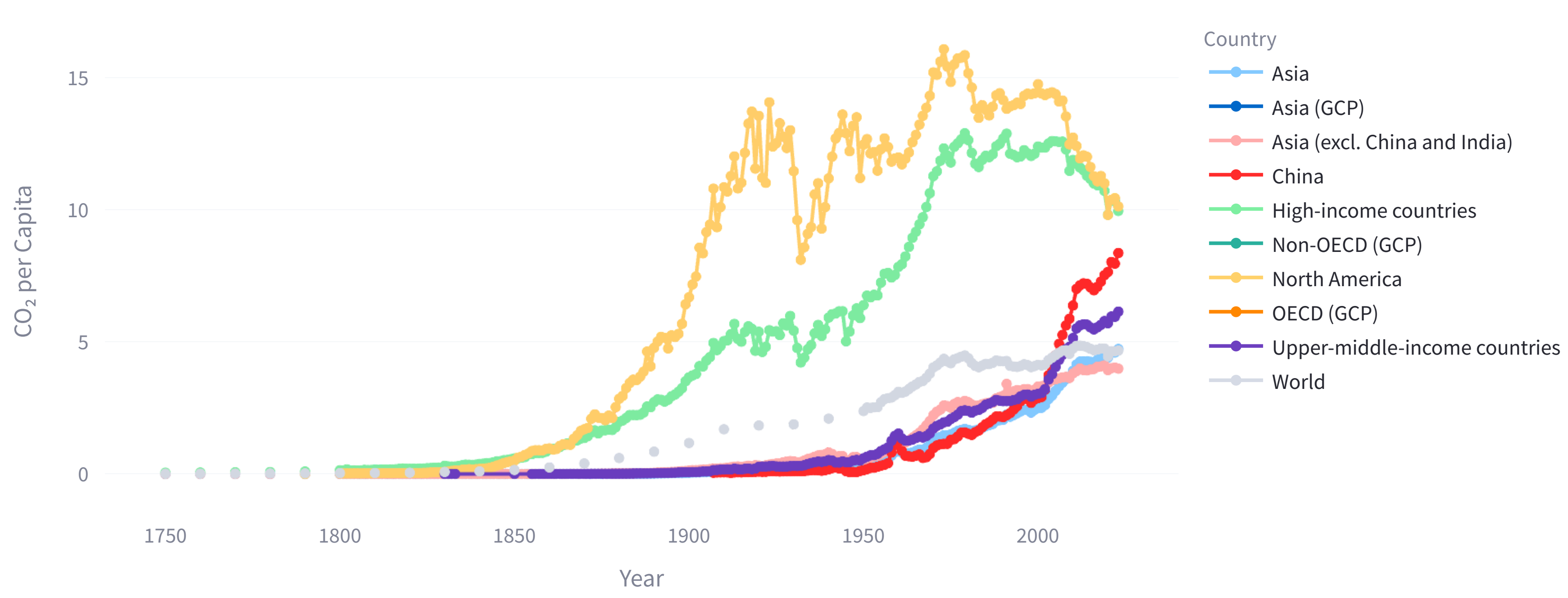
Top emitters in 2023:

```
[
0 : "World"
1 : "Non-OECD (GCP)"
2 : "Asia"
3 : "Asia (GCP)"
4 : "Upper-middle-income countries"
5 : "High-income countries"
6 : "China"
7 : "OECD (GCP)"
8 : "Asia (excl. China and India)"
9 : "North America"
]
```


Total CO_2 Emissions Trends for Top Emitters



CO_2 Emissions Per Capita Trends for Top Emitters

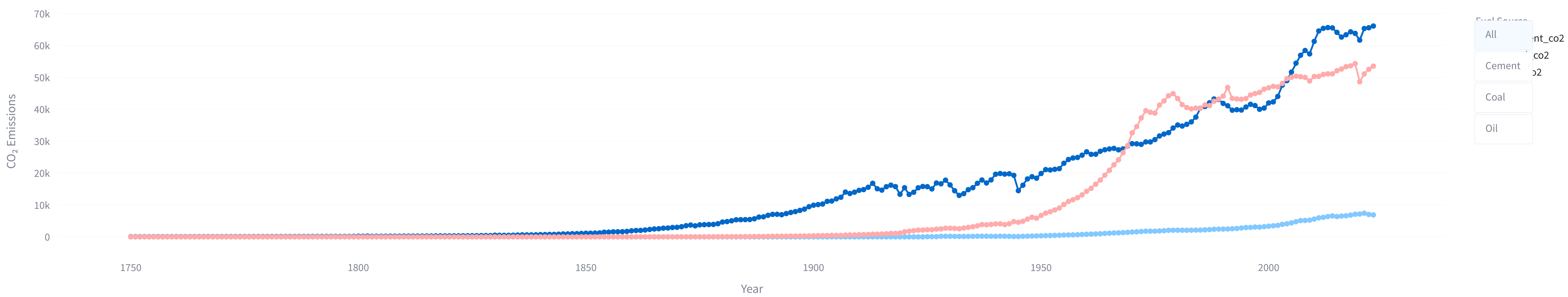


The visualizations reveal striking contrasts between regions in both total and per capita CO_2 emissions. In the first plot showing total emissions over time, China exhibits a steep upward trajectory, overtaking high-income countries after the early 2000s and becoming the largest overall emitter by 2023. In contrast, the second plot on per capita emissions shows that high-income regions, especially North America, have consistently maintained the highest emissions per person across the timeline. Despite China's rapid growth in total emissions, its per capita values remain below those of wealthier nations. These trends highlight how population size, economic development, and consumption patterns shape emission dynamics—while China leads in total output, high-income countries continue to have a significantly larger carbon footprint per individual.

Fuel-Specific (Sectoral) Deep Dives

Hypothesis: Cement-related CO_2 emissions have steadily risen with global infrastructure growth; coal emissions peaked mid-century and are now declining due to cleaner energy use;

Global Fuel-Specific CO_2 Emissions Trends

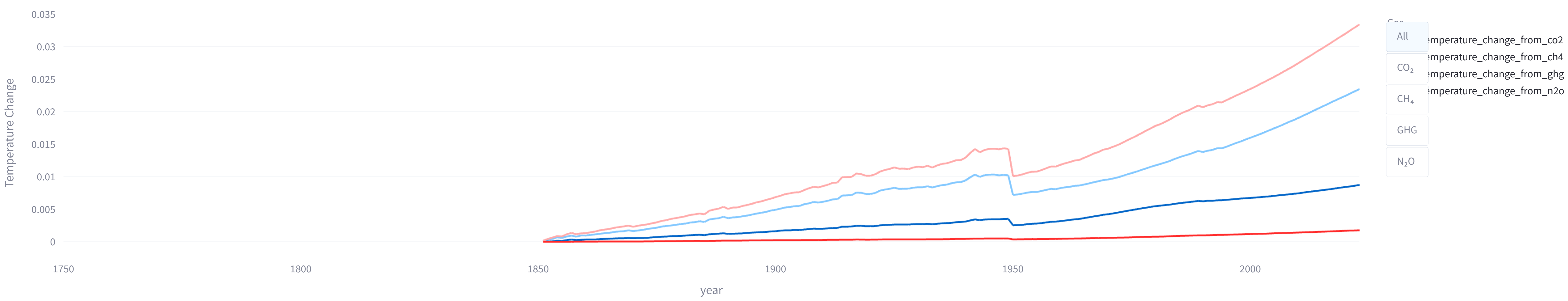


The fuel-specific CO_2 trends highlight distinct historical trajectories: **coal** led the early industrial era, rising steadily from the 19th century to peak around 1960 before plateauing and then surging again into the 21st century (now exceeding 65 Gt). **Oil** overtook coal in the mid-20th century, climbing rapidly after 1950 and maintaining growth to roughly 55 Gt by 2023, reflecting the global shift to petroleum. **Cement** emissions, though much smaller, have grown exponentially since the 1950s—reaching over 7 Gt—driven by urbanization and infrastructure expansion. Together, these patterns illustrate how different sectors dominated successive phases of economic development: coal powered early industrialization, oil fueled mass mobility and modern economies, and cement production underpins today's urban growth.

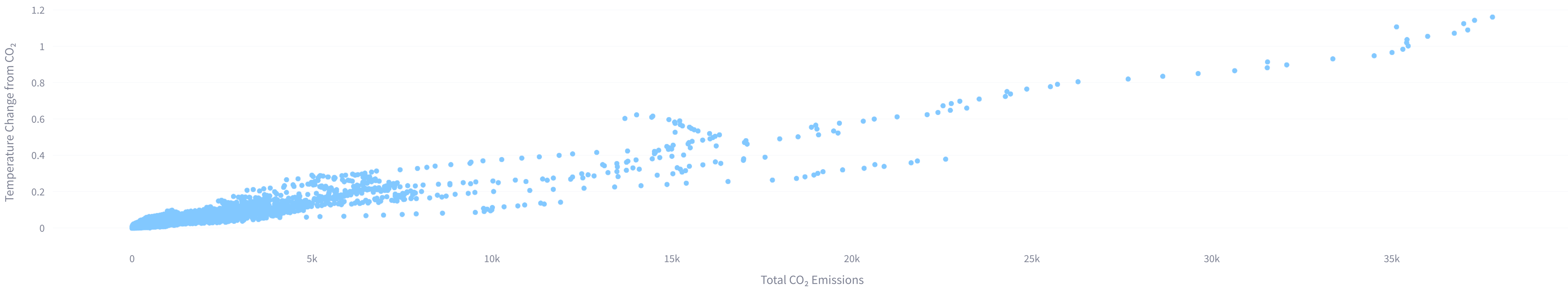
Temperature Change & Greenhouse Gases Analysis

Hypothesis: CO_2 will drive the most significant temperature rise, closely tracking its emissions, while overall warming from all greenhouse gases will outpace individual contributions.

Average Temperature Change Contributions Over Time



Total CO_2 Emissions vs. Temperature Change from CO_2



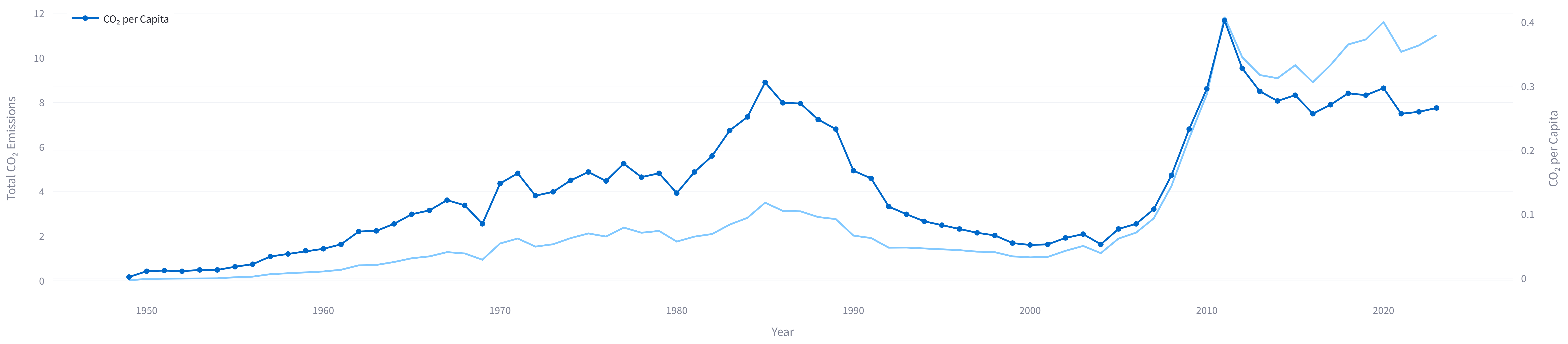
The visual analysis reveals that carbon dioxide (CO_2) is the primary contributor to global temperature change among the greenhouse gases, with its influence increasing steadily and sharply, especially after the 1950s—likely due to industrial expansion post-World War II. Methane (CH_4) and nitrous oxide (N_2O) also contribute to warming, but their impacts are significantly smaller, with CH_4 showing a gradual rise and N_2O remaining relatively stable until a slight increase in recent decades. The overall greenhouse gas (GHG) temperature change trend closely mirrors that of CO_2, confirming its dominant role. A strong positive correlation is evident between total CO_2 emissions and temperature change from CO_2, as shown in the scatter plot, where emissions and temperature rise together in a nonlinear, accelerating pattern. This suggests that as CO_2 emissions increase, their warming effect intensifies disproportionately. The clustering of data points also indicates common emission behaviors, while some outliers may reflect country-specific variations or abrupt industrial changes. Overall, the data underscores the urgent need to reduce CO_2 emissions to mitigate their escalating impact on global temperatures.

Interactive Visualizations - Country-Wise

Select Country:

Afghanistan

Emissions Trends for Afghanistan



The CO_2 emissions data for Afghanistan shows low levels from 1950 until around 2000, aside from a mid-1980s peak and a 1990s dip. Between 2005 and 2010, total emissions surged roughly sixfold—from about 1–2 units to nearly 12 by 2020. Per capita emissions peaked near 0.40 units around 2010 before declining, indicating that population growth has recently outpaced emissions growth. This divergence after 2010 likely mirrors Afghanistan's shifting political and economic landscape, with phases of conflict, rebuilding, and development driving changes in energy use.

World Map: Global Distribution of CO_2 Emissions

Hypothesis: CO_2 emissions are heavily concentrated in industrialized nations, with China, the US, and India leading, suggesting that historical and economic factors drive emission disparities more than population or regional development alone.

Global CO_2 Emissions Distribution in 2023

In 2023, China leads global CO_2 emissions with over 10,000 Mt, followed by the US, India, and Russia. Emissions in Africa, Latin America, and Southeast Asia remain comparatively low, highlighting a significant gap in climate responsibility between industrialized and developing nations.

Pie Chart: CO_2 Emission Contributions by Source

Hypothesis: Decarbonization must go beyond coal and oil, as nearly half of CO_2 emissions come from varied sources like land use and industry.

Global CO_2 Emission Contribution by Source in 2023

“Other Emissions” account for 47.5% of global CO_2 output, encompassing land-use changes, biomass burning, and miscellaneous industrial processes. Coal contributes 27.4%, reflecting heavy reliance on coal-fired power generation; oil makes up 22.2%, driven by transportation and industry; and cement production represents 2.9%, highlighting infrastructure’s carbon footprint. This breakdown underscores that while tackling coal and oil is crucial, nearly half of emissions come from a diverse array of sources—mandating comprehensive policies and technologies to address both major and less obvious contributors for effective decarbonization.

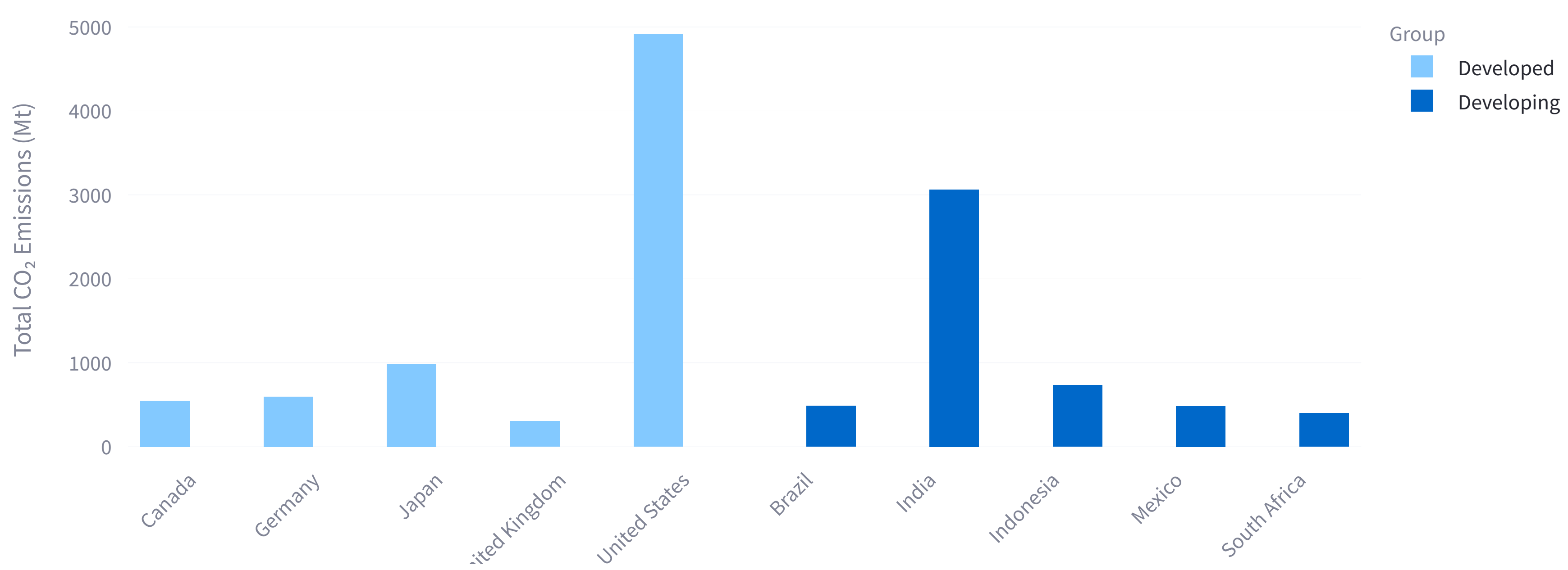
Country-Wise CO_2 Emissions Pie Chart

Top 10 CO_2 Emitting Countries in 2023 (plus Others)

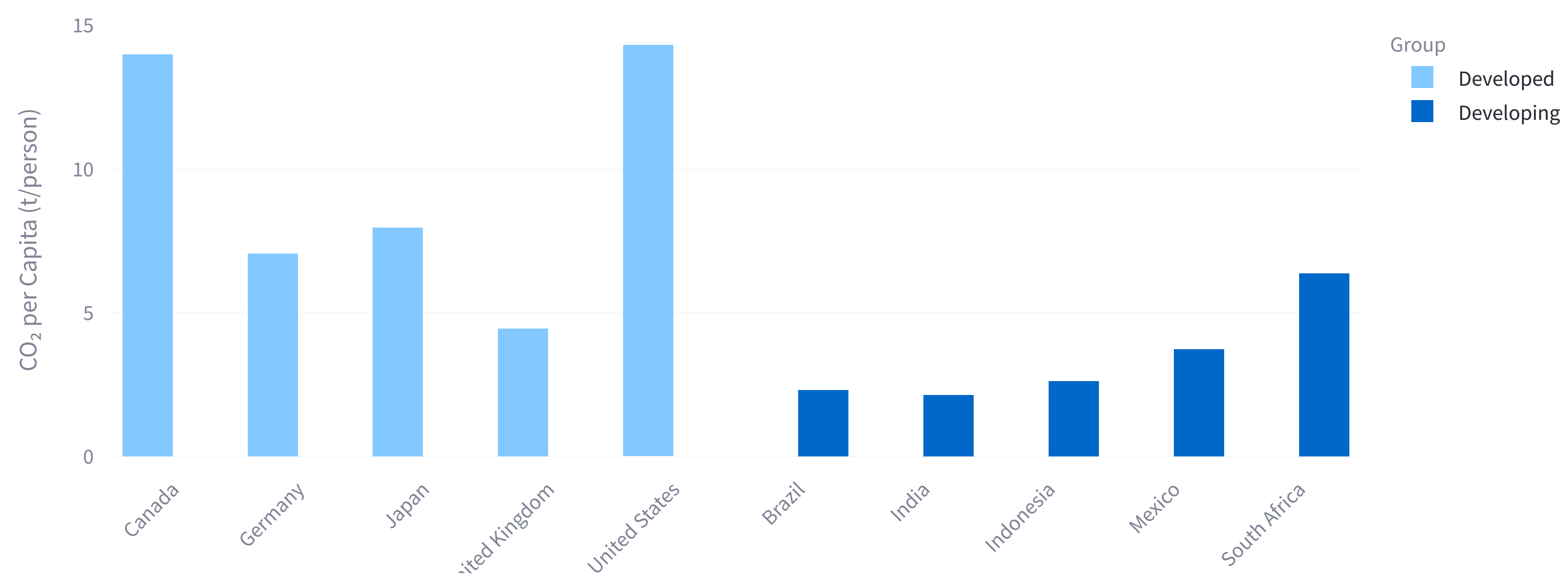
In 2023, China (33.6%) and the US (28.7%) together produce over 60% of global CO_2 emissions. India ranks third at 6.35%, followed by Russia, Japan, and other industrialized countries with smaller shares. The “Others” category (32.9%)—larger than any country except China and the US—highlights that emissions from numerous smaller nations are collectively significant, calling for both targeted action in top emitters and broad international cooperation.

CO_2 Emissions: Developed vs. Developing Countries

Total CO_2 Emissions by Country in 2023



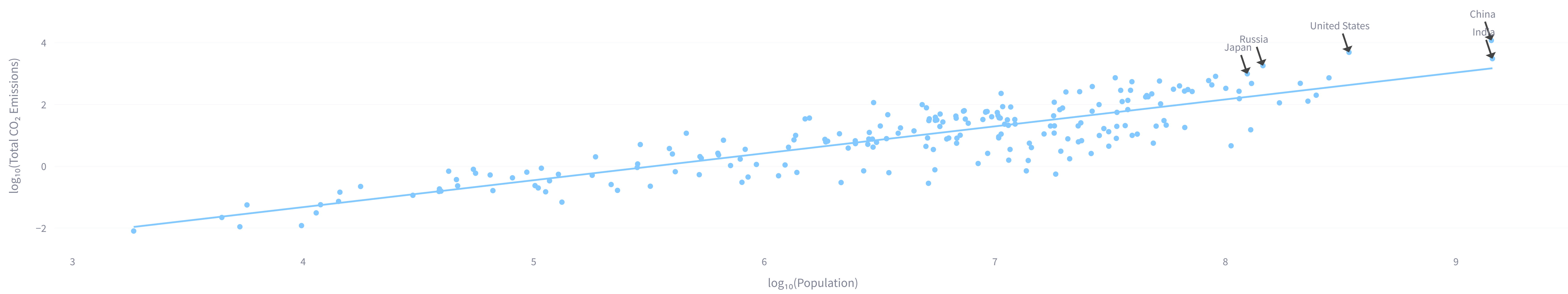
CO_2 Emissions Per Capita by Country in 2023



The comparison between developed and developing countries reveals stark contrasts in CO_2 emissions patterns. Developed nations like the U.S. and Germany show significantly higher per capita emissions (likely 10-20 t/person) compared to developing countries such as India and Indonesia (typically 1-5 t/person), reflecting greater energy consumption and industrialization in wealthier economies. However, some developing nations like South Africa and Mexico may bridge this gap due to fossil fuel dependence. In absolute terms, the U.S. likely dwarfs other countries' total emissions, while populous developing nations like India may rank high in total volume despite low per capita figures. This disparity highlights the climate policy dilemma: developed nations must reduce high per capita emissions through technology and efficiency, while developing countries face the challenge of curbing emission growth during economic expansion. The data underscores that equitable climate solutions must address both historical responsibility (cumulative emissions) and future development needs.

Population vs. Total CO_2 Emissions

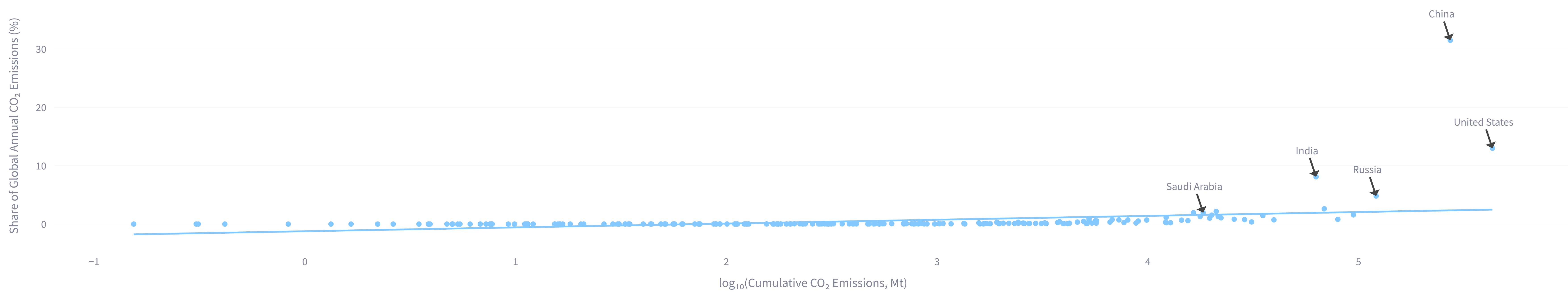
Population vs. Total CO_2 Emissions (log-log) in 2023



In 2023, a log-log scatter of **Population** versus **Total CO_2 Emissions** reveals a nearly linear relationship (slope ≈ 0.92), indicating that larger populations generally produce proportionally more emissions—but slightly less than one-to-one. Major countries like China, India, and the United States cluster near the top, reflecting both large populations and high emissions. Notably, points above the trend line (e.g., some oil-exporting states or small, energy-intensive economies) emit more CO_2 than their population alone would predict, while those below the line (e.g., highly efficient or service-based economies) emit less. This pattern underscores the strong role of population scale in driving emissions, while also highlighting outliers where energy intensity, economic structure, or policy interventions significantly alter the population-emissions dynamic.

Historical vs. Current Emissions: Cumulative CO_2 vs. Share of Global CO_2

Cumulative CO_2 vs. Share of Global CO_2 in 2023



In 2023, the scatter of **$\log_{10}$(Cumulative CO_2 Emissions)** against **Share of Global Annual CO_2 Emissions** reveals a strong positive relationship: countries with the largest historical emissions still command the biggest slices of today's global output. China (≈10^6 Mt cumulative; ~37 %) and the United States (≈$10^{6.05}$ Mt; ~13 %) stand out as dominant emitters, while Russia, India, and Saudi Arabia also exceed their historical weight. The fitted trend line (Share ≈ $2.70 \cdot \log_{10}(CumCO_2) - 5.59$) quantifies this linkage. However, several nations lie above or below the line, indicating shifts in current emission leadership: emerging economies like India have a higher share than their past totals would suggest, whereas some long-industrialized countries show a smaller current share, reflecting stabilization or decline. This analysis underscores both the inertia of historical emissions and the evolving dynamics of global carbon leadership.