# "Can Innovation solve the Global Climate Crisis?"

## CO2 Emissions and Innovation metrics dataset

# PROJECT 1

## 1.1 Introduction

Climate change, driven by rising carbon dioxide ($CO_2$) emissions, is a global challenge requiring urgent solutions (IPCC, 2021).While industrial growth boosts economies, it also increases environmental degradation (World Bank, 2020). Many believe that innovation—through new technologies and better policies—can help reduce emissions while maintaining economic growth (Schumpeter, 1942). Clean energy, energy efficiency, and industrial improvements are key to making this shift (OECD, 2019).

This project explores whether innovation alone can solve the global climate crisis. Using data from the Global Innovation Index and the Global Carbon Budget (Friedlingstein et al., 2022), along with income classifications from the World Bank (World Bank, 2020), this study examines how different countries approach sustainability. Wealthy nations tend to lead in green technology but still have high emissions due to industrial activity and consumption (IEA, 2021). In contrast, poorer nations rely on older energy sources, limiting their ability to switch to cleaner alternatives (Sachs et al., 2021).

Investing in research and development (R&D), clean technology, and better infrastructure can help reduce emissions over time (Popp, 2012). Countries that prioritize innovation often find long-term solutions to environmental issues (Acemoglu et al., 2014). However, technology alone may not be enough—carbon pricing, subsidies for green energy, and strong international policies also play a major role (Nordhaus, 2019).

This study aims to answer whether innovation can truly solve climate change or if broader economic and policy changes are needed. The findings could help policymakers balance economic growth with environmental protection, ensuring a more sustainable future (Stern, 2007).

### 1.1.1 Variables

For this project, the chosen X variables selected are:

1. Global Innovation Index (GII): This measures a country's overall innovation performance and reflects how technological advancements may influence emissions.

2. R&D Spending: This is critical for fostering technological innovations that can improve energy efficiency and reduce emissions.

3. High-tech Manufacturing: This variable represents the extent of a country's manufacturing sector dedicated to high-tech production. It is included to capture the role of advanced industries in influencing emissions through both production processes and clean technology innovations.

4. GDP per Unit of Energy Use: This measures energy efficiency, quantifying the economic output generated per unit of energy consumed. Countries with higher values are expected to have more efficient energy use, potentially leading to lower emissions per capita.

5. High-tech Imports: This variable represents a country's reliance on importing advanced technology. It is included because it reflects the ability of countries to access and implement innovative solutions, which can either increase or decrease emissions depending on the type of imported technologies.

6. ISO 14001 Environmental Certificates: This metric captures the extent to which organizations in a country adhere to international environmental management standards.

7. Innovation Efficiency Ratio: This measures how effectively a country converts its innovation inputs into tangible outputs. A higher efficiency ratio suggests better utilization of resources for innovation, which may translate into sustainable practices and reduced emissions.

The chosen Y variable (outcome) is Total Per Capita $CO_2$ Emissions, which represents the average amount of $CO_2$ emitted per individual in a country. This variable directly reflects the environmental impact of a country relative to its population and is essential for understanding the relationship between innovation, economic development, and environmental outcomes.

Controlling for income groups (high, upper middle, lower middle, and low) is essential as income levels influence innovation capacity, technology adoption, and energy efficiency. High-income countries invest more in clean technologies and energy efficiency, while lower-income nations rely on traditional energy sources. This ensures income disparities do not confound the relationship between innovation and $CO_2$ emissions.

These variables capture the interplay between innovation, governance, energy efficiency, and income levels, providing a comprehensive framework for analyzing their combined impact on emissions and addressing the research question effectively.

## 1.2 Data Cleaning

### 1.2.1 CO2 Emissions Data Set

In [107…

```
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np
import seaborn as sns
import geopandas as gpd
from IPython.display import display, HTML
from IPython.display import display, Markdown

!pip install stargazer
import statsmodels.api as sm
from stargazer.stargazer import Stargazer
from IPython.display import display, HTML

import undetected_chromedriver as uc
from selenium.webdriver.common.by import By
import time

from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeRegressor, plot_tree
from sklearn.metrics import mean_squared_error

from sklearn.ensemble import RandomForestRegressor

import warnings
warnings.filterwarnings('ignore', category=FutureWarning)
```

```
Requirement already satisfied: stargazer in /opt/anaconda3/lib/python3.11/
site-packages (0.0.7)
```

In [2]:

```
#import data set
data = pd.read_csv("GCB2022v27_MtCO2_flat-Copy1.csv")
```

In [3]:

```
#Drop Other column
data.drop(columns = ['Other'], inplace = True) #drop rows which have NaN
data.dropna(inplace=True)
data.set_index('Country', inplace=True)
data = data[data.index != 'Global']
```

```
filtered_data_year = data[(data['Year'] >= 2000) & (data['Year'] <= 2020)
display(Markdown("**Table 1: Cleaned Primary Data Set ; CO2 Emissions fro
filtered_data_year.head()
```

**Table 1: Cleaned Primary Data Set ; CO2 Emissions from sources for Years 2000–2020**

Out[3]:

| | ISO 3166-1 alpha-3 | Year | Total | Coal | Oil | Gas | Cement | I |
|---|---|---|---|---|---|---|---|---|
| **Country** | | | | | | | | |
| **Afghanistan** | AFG | 2000 | 1.047128 | 0.003664 | 0.787760 | 0.223504 | 0.010216 | 0.0 |
| **Afghanistan** | AFG | 2001 | 1.069098 | 0.069616 | 0.762112 | 0.208848 | 0.006538 | 0.0 |
| **Afghanistan** | AFG | 2002 | 1.340995 | 0.055109 | 0.727438 | 0.547416 | 0.011033 | 0.0 |
| **Afghanistan** | AFG | 2003 | 1.559602 | 0.091813 | 0.991575 | 0.466408 | 0.009807 | 0.0 |
| **Afghanistan** | AFG | 2004 | 1.237247 | 0.091600 | 0.908672 | 0.227168 | 0.009807 | 0.0 |

The values represent the annual emissions in metric tons, reflecting the contribution of each source to overall emissions. These statistics highlight variations in energy use and industrial activity across time and regions, offering insights into emission trends and their potential drivers, which are crucial for identifying areas where mitigation efforts can have the greatest impact.

In [4]:
```
filtered_data_year_2020 = filtered_data_year[filtered_data_year["Year"] =
filtered_data_year_2020 = filtered_data_year_2020.reset_index()

filtered_data_year_2011 = filtered_data_year[filtered_data_year["Year"] =
filtered_data_year_2011 = filtered_data_year_2011.reset_index()
```

## 1.2.2 Global Innovation index Dataset

In [5]:
```
innovation = pd.read_csv("Global_Innovation_index.csv") # Loading dataset
innovation.rename(columns={' ': 'Indicators'}, inplace=True) # Setting co
```

In [6]:
```
filtered_GII = innovation.iloc[1:3,135:266]
filtered_GII_new = innovation.iloc[1:5,135:266]
filtered_GII_new['Indicator'] = innovation.iloc[1:5, 0].values

# Move the 'Indicator' column to the first position
columns = ['Indicator'] + [col for col in filtered_GII_new.columns if col
filtered_GII_new = filtered_GII_new[columns]
```

```
filtered_GII_new = filtered_GII_new.set_index('Indicator').T # Flip rows
filtered_GII_new = filtered_GII_new.rename_axis('Country') # Rename the i
filtered_GII_new.index = filtered_GII_new.index.str.rstrip('.1')
filtered_GII_new

# Display the table
display(Markdown("**Table 2 (a): Cleaned Data Set ; Key Innovation Perfor
filtered_GII_new
```

**Table 2 (a): Cleaned Data Set ; Key Innovation Performance Metrics by Country**

Out[6]:

| Indicator<br>Country | Global Innovation Index | Innovation Efficiency Ratio | Innovation Input Sub-index | Innovation Output Sub-index |
|---|---|---|---|---|
| **Albania** | 27.1 | 0.4 | 39.6 | 14.6 |
| **Algeria** | 19.5 | 0.3 | 30.5 | 8.5 |
| **Argentina** | 28.3 | 0.5 | 38.3 | 18.4 |
| **Armenia** | 32.6 | 0.7 | 38.1 | 27.1 |
| **Australia** | 48.4 | 0.5 | 62.9 | 33.8 |
| **...** | ... | ... | ... | ... |
| **Uzbekistan** | 24.5 | 0.3 | 38.2 | 10.8 |
| **Viet Nam** | 37.1 | 0.8 | 42.1 | 32.2 |
| **Yemen** | 13.6 | 0.4 | 19.9 | 7.3 |
| **Zambia** | 19.4 | 0.3 | 30.7 | 8.0 |
| **Zimbabwe** | 20.0 | 0.4 | 28.0 | 11.9 |

131 rows × 4 columns

These metrics indicate each country's capacity and effectiveness in fostering innovation, with higher values reflecting stronger performance. For example, Australia scores a GII of 48.4, with notable efficiency and high input and output sub-indices, highlighting its advanced innovation ecosystem. Such insights are valuable for comparing countries' innovation environments and identifying areas for policy improvements or resource allocation to enhance innovation.

In [7]:

```
# Extract a single row by index
row_GdpPerEnergyUse = innovation.iloc[44]
row_ISO = innovation.iloc[46]

filtered_infrastructure = innovation.iloc[44:47,135:266]
filtered_infrastructure['Indicator'] = innovation.iloc[44:47, 0].values
```

```
filtered_infrastructure_first_column = innovation.iloc[44:47, 0]

#moving indicator at the front:
columns_infrastructure = ['Indicator'] + [col for col in filtered_infrast
filtered_infrastructure = filtered_infrastructure[columns_infrastructure]

filtered_infrastructure_reversed_df = filtered_infrastructure.set_index('

# Rename the index to "Country" for row headings
filtered_infrastructure_reversed_df = filtered_infrastructure_reversed_df

filtered_infrastructure_reversed_df.index = filtered_infrastructure_rever
filtered_infrastructure_reversed_df
filtered_infrastructure_reversed_df.columns = ['GDP per unit of energy us
                                               'ISO 14001 environmental c
```

In [8]:
```
# Extract a single row by index
row_GdpPerEnergyUse = innovation.iloc[44,135:266].T
row_ISO = innovation.iloc[46,135:266].T
row_RD = innovation.iloc[28,135:266].T
row_Hightech = innovation.iloc[75,135:266].T
row_manufacturing = innovation.iloc[91,135:266].T

variables = [row_GdpPerEnergyUse, row_ISO, row_RD, row_Hightech, row_manu
for i in variables:
    i.index.name = 'Country'

# Merging all columns
merged_indicators = pd.concat([row_GdpPerEnergyUse,row_ISO, row_RD, row_H

merged_indicators.columns = ['Gdp Per Unit of Energy Use','ISO 14001 envi
                             'Research and development (R&D)', 'High-tech
                             'High-tech and medium high-tech manufacturin

merged_indicators.index = merged_indicators.index.str.rstrip('.1')
# Display the table
display(Markdown("**Table 2 (b): Cleaned Data Set ; Other Key Metrics by
merged_indicators
```

**Table 2 (b): Cleaned Data Set ; Other Key Metrics by Country**

Out[8]:

| Country | Gdp Per Unit of Energy Use | ISO 14001 environmental certificates | Research and development (R&D) | High-tech imports | High-tech and medium high-tech manufacturing, % |
|---|---|---|---|---|---|
| Albania | 44.1 | 29.9 | 0.0 | 1.7 | 2.8 |
| Algeria | 29.7 | 1.2 | 5.1 | 28.1 | 4.6 |
| Argentina | 27.8 | 11.9 | 28.1 | 29.0 | 0.0 |
| Armenia | 21.4 | 0.6 | 1.2 | 19.5 | 4.2 |
| Australia | 26.2 | 16.0 | 59.4 | 34.0 | 33.7 |
| ... | ... | ... | ... | ... | ... |
| Uzbekistan | 13.6 | 1.0 | 2.0 | 23.9 | 28.2 |
| Viet Nam | 19.4 | 16.1 | 7.0 | 96.6 | 50.7 |
| Yemen | 67.4 | 0.8 | 0.0 | 18.2 | 0.0 |
| Zambia | 10.7 | 1.6 | 0.0 | 18.1 | 11.7 |
| Zimbabwe | 3.3 | 8.4 | 0.3 | 16.3 | 26.9 |

131 rows × 5 columns

Table 2 presents additional metrics related to countries' economic and technological performance. They provide a multidimensional perspective on sustainability, technological progress, and innovation intensity across countries. For instance, Australia's high GDP per unit of energy use and R&D investment (59.4%) reflect efficient resource use and innovation-driven growth.

### 1.2.3 Controlling for Income Groups

In [9]:

```
# Load datasets
country_data = pd.read_csv("GCB2022v27_MtCO2_flat-Copy1.csv")
incomeGroup = pd.read_csv("Income_group.csv")

country_data.rename(columns={"ISO 3166-1 alpha-3": "Country Code"}, inpla
combined_data = country_data.merge(incomeGroup, on="Country Code", how="l

# Extract relevant columns
result = combined_data[["Country", "Country Code", "IncomeGroup", ]]
result = result.dropna()
result = result.drop_duplicates()

# Grouping countries on 4 categories
```

```
grouped = result.groupby("IncomeGroup")["Country"]
high_income_countries = result[result["IncomeGroup"] == "High income"].re
low_income_countries = result[result["IncomeGroup"] == "Low income"].rese
Upper_middle_income_countries = result[result["IncomeGroup"] == "Upper mi
Lower_middle_income_countries = result[result["IncomeGroup"] == "Lower mi

# Create a summary table for income groups
income_group_summary = result['IncomeGroup'].value_counts().reset_index()
income_group_summary.columns = ['Income Group', 'Number of Countries']

# Caption for the table
display(Markdown("**Table 3: Distribution of Countries by Income Group**"
income_group_summary
```

**Table 3: Distribution of Countries by Income Group**

Out[9]:

| | Income Group | Number of Countries |
|---|---|---|
| **0** | High income | 75 |
| **1** | Upper middle income | 53 |
| **2** | Lower middle income | 51 |
| **3** | Low income | 26 |

Table 3 presents the distribution of countries by income group.This breakdown is important for analyzing how income levels influence various indicators such as innovation, $CO_2$ emissions, and development metrics, allowing for targeted comparisons and policy recommendations.

### 1.2.4 Innovation index by year

In [10]:
```
innovation_by_year = pd.read_excel("innovation_by_year.xlsx")
```

In [11]:
```
# Rename the column "Economies" to "Country"
innovation_by_year = innovation_by_year.rename(columns={"Economies": "Cou
innovation_by_year = innovation_by_year.set_index("Country")

# Display the updated DataFrame to confirm
display(Markdown("**Table 4: Innovation Indicators by Country (2011-2022)
innovation_by_year.head()
```

**Table 4: Innovation Indicators by Country (2011-2022)**

Out[11]:

| | Year | ID | Institutions | Human capital and research | Infrastructure | Market sophistication | Busin sophisticat |
|---|---|---|---|---|---|---|---|
| **Country** | | | | | | | |
| **Albania** | 2011 | 2 | 65.2 | 32.7 | 26.4 | 47.5 | 1 |
| **Albania** | 2012 | 2 | 55.0 | 26.2 | 33.6 | 49.7 | 2 |
| **Albania** | 2013 | 2 | 58.9 | 27.1 | 31.1 | 56.8 | 2 |
| **Albania** | 2014 | 2 | 58.8 | 22.8 | 34.1 | 61.9 | 2 |
| **Albania** | 2015 | 2 | 60.1 | 21.8 | 39.0 | 59.1 | 2 |

This table presents country-wise innovation indicators from 2011 to 2022, covering metrics such as human capital, infrastructure, market sophistication, business sophistication, technology outputs, and the Global Innovation Index, along with GDP and income levels.

In [12]:

```
# Extract only rows where Year is 2020
innovation_2020 = innovation_by_year[innovation_by_year["Year"] == 2020]
innovation_2020 = innovation_2020.reset_index()
# Extract only rows where Year is 2011
innovation_2011 = innovation_by_year[innovation_by_year["Year"] == 2011]
innovation_2011 = innovation_2011.reset_index()
```

In [13]:

```
reg_merged = pd.merge(innovation_by_year, filtered_data_year, on=["Countr
reg_merged = reg_merged.dropna()  # Remove rows with NaN values
reg_merged = reg_merged.rename(columns={"Per Capita": "Per Capita CO2 Emi
```

## 1.3 Summary Statistics Tables

In [14]:

```
# Filter data for each decade
decade_1 = filtered_data_year[(filtered_data_year['Year'] >= 2000) & (fil
decade_2 = filtered_data_year[(filtered_data_year['Year'] >= 2010) & (fil

# Group by country and sum the total emissions for each decade
decade_1_emissions = decade_1.groupby('Country')['Total'].sum().reset_ind
decade_2_emissions = decade_2.groupby('Country')['Total'].sum().reset_ind

# Rename columns for clarity
decade_1_emissions.rename(columns={'Total': 'Total Emissions 2000-2009'},
decade_2_emissions.rename(columns={'Total': 'Total Emissions 2010-2020'},
```

In [15]:

```
total_per_capita_emissions = filtered_data_year.groupby('Country')['Per C
```

```
# Convert the Series to a DataFrame
total_per_capita_emissions = total_per_capita_emissions.reset_index()
total_per_capita_emissions.columns = ['Country', 'Total Per Capita Emissi
```

In [16]:

```
country_emissions = filtered_data_year.groupby('Country')['Total'].sum()
mean_emissions_per_country = filtered_data_year.groupby('Country')['Total
country_emissions = country_emissions.reset_index()
country_emissions.rename(columns={'Total': 'Total Emissions'}, inplace=Tr
country_emissions['Mean Emissions over 20 years'] = mean_emissions_per_co

country_emissions_copy = country_emissions.copy()

# Merge Datasets
country_emissions.merge(total_per_capita_emissions, on = 'Country', how =
country_emissions = country_emissions.merge(decade_1_emissions, on='Count
country_emissions = country_emissions.merge(decade_2_emissions, on='Count
country_emissions = country_emissions.merge(total_per_capita_emissions, o
columns_order = ['Country', 'Total Emissions', 'Total Per Capita Emission
                 'Total Emissions 2000-2009', 'Total Emissions 2010-2020'

country_emissions = country_emissions[columns_order]
# Display the result
display(Markdown("**Table 5: Summary of Total, Per Capita, and Decadal CO
country_emissions.head(5)
```

**Table 5: Summary of Total, Per Capita, and Decadal $CO_2$ Emissions by Country (2000–2020)**

Out[16]:

| | Country | Total Emissions | Total Per Capita Emissions | Mean Emissions over 20 years | Total Emissions 2000-2009 | Total Emissions 2010-2020 |
|---|---|---|---|---|---|---|
| **0** | Afghanistan | 134.716516 | 4.279085 | 6.415072 | 23.749169 | 110.967347 |
| **1** | Albania | 95.202807 | 32.295674 | 4.533467 | 39.398730 | 55.804077 |
| **2** | Algeria | 2655.136465 | 71.876520 | 126.435070 | 978.604375 | 1676.532090 |
| **3** | Andorra | 10.645796 | 144.730713 | 0.506943 | 5.389744 | 5.256052 |
| **4** | Angola | 455.129260 | 19.056985 | 21.672822 | 178.610861 | 276.518399 |

Table 5 summarizes the total, per capita, and decadal $CO_2$ emissions by country from 2000 to 2020. It provides insights into total emissions (in metric tons), emissions per person, and emissions trends across two decades (2000–2009 and 2010–2020). For instance, Algeria's total emissions stand out at 2,656 metric tons, with a significant increase in emissions between the decades, indicating industrial growth and increased energy consumption. This data is critical for understanding both absolute and relative contributions to global emissions, highlighting the need for tailored mitigation strategies for high emitters and regions experiencing rapid increases in emissions.

In [17]:
```
country_emissions_sorted = country_emissions.sort_values(by='Total Emissi
country_emissions_sorted.reset_index(drop = True, inplace = True)
```

In [18]:
```
# Filter data for the years 2000–2009
filtered_2000_2009 = filtered_data_year[(filtered_data_year['Year'] >= 20
emissions_by_source_2000_2009 = filtered_2000_2009.groupby('Country')[['C
emissions_by_source_2000_2009.columns = [
    'Total CO2 Emissions from Coal (2000–2009)',
    'Total CO2 Emissions from Gas (2000–2009)',
    'Total CO2 Emissions from Oil (2000–2009)',
    'Total CO2 Emissions from Cement (2000–2009)',
    'Total CO2 Emissions from Flaring (2000–2009)',
    'Total CO2 Emissions Per Capita (2000–2009)'
]

# Filter data for the years 2010–2020
filtered_2010_2020 = filtered_data_year[(filtered_data_year['Year'] >= 20
emissions_by_source_2010_2020 = filtered_2010_2020.groupby('Country')[['C
emissions_by_source_2010_2020.columns = [
    'Total CO2 Emissions from Coal (2010–2020)',
    'Total CO2 Emissions from Gas (2010–2020)',
    'Total CO2 Emissions from Oil (2010–2020)',
    'Total CO2 Emissions from Cement (2010–2020)',
    'Total CO2 Emissions from Flaring (2010–2020)',
    'Total CO2 Emissions Per Capita (2010–2020)'
]

#Table 2c
# Group by 'Country' and sum the 'Per Capita' emissions
total_per_capita_emissions = filtered_data_year.groupby('Country')['Per C
total_per_capita_emissions = total_per_capita_emissions.reset_index() # C
total_per_capita_emissions.columns = ['Country', 'Total Per Capita Emissi
display(Markdown("**Table 6(a): Distribution of CO2 Emissions by Country
emissions_by_source_2000_2009.head(5)
```

**Table 6(a): Distribution of CO2 Emissions by Country and Source (2000–2009)**

Out[18]:

| Country | Total CO2 Emissions from Coal (2000–2009) | Total CO2 Emissions from Gas (2000–2009) | Total CO2 Emissions from Oil (2000–2009) | Total CO2 Emissions from Cement (2000–2009) | Total CO2 Emissions from Flaring (2000–2009) | Total CO2 Emissions Per Capita (2000–2009) |
|---|---|---|---|---|---|---|
| **Afghanistan** | 3.918703 | 3.208415 | 16.472577 | 0.105508 | 0.043968 | 0.957895 |
| **Albania** | 1.113567 | 0.230733 | 35.289161 | 2.765268 | 0.000000 | 12.954547 |
| **Algeria** | 24.054262 | 477.490561 | 336.404097 | 55.547741 | 85.107714 | 29.712802 |
| **Andorra** | 0.000000 | 0.000000 | 5.389744 | 0.000000 | 0.000000 | 72.655515 |
| **Angola** | 0.000000 | 12.596832 | 68.667024 | 4.101869 | 93.245136 | 9.030205 |

In [19]:

```
display(Markdown("**Table 6(b): Distribution of CO2 Emissions by Country
emissions_by_source_2010_2020.head()
```

**Table 6(b): Distribution of CO2 Emissions by Country and Source (2010–2020)**

Out[19]:

| Country | Total CO2 Emissions from Coal (2010–2020) | Total CO2 Emissions from Gas (2010–2020) | Total CO2 Emissions from Oil (2010–2020) | Total CO2 Emissions from Cement (2010–2020) | Total CO2 Emissions from Flaring (2010–2020) | Total CO2 Emissions Per Capita (2010–2020) |
|---|---|---|---|---|---|---|
| **Afghanistan** | 37.706606 | 3.145864 | 69.674635 | 0.440242 | 0.000000 | 3.321190 |
| **Albania** | 5.328273 | 0.981071 | 38.371974 | 10.986403 | 0.136357 | 19.341127 |
| **Algeria** | 9.393741 | 821.727024 | 579.772818 | 102.380332 | 163.258173 | 42.163718 |
| **Andorra** | 0.000000 | 0.000000 | 5.256052 | 0.000000 | 0.000000 | 72.075198 |
| **Angola** | 0.000000 | 16.623900 | 169.579512 | 13.843195 | 76.471791 | 10.026780 |

In [20]:

```
display(Markdown("**Table 6(c): Total Per Capita CO2 Emissions by Country
total_per_capita_emissions.head(5)
```

**Table 6(c): Total Per Capita CO2 Emissions by Country over 20 years**

Out[20]:

| | Country | Total Per Capita Emissions |
|---|---|---|
| 0 | Afghanistan | 4.279085 |
| 1 | Albania | 32.295674 |
| 2 | Algeria | 71.876520 |
| 3 | Andorra | 144.730713 |
| 4 | Angola | 19.056985 |

This set of tables (6a, 6b, and 6c) provides a detailed breakdown of $CO_2$ emissions by source (coal, gas, oil, cement, and flaring) and per capita for two distinct decades: 2000–2009 and 2010–2020. Table 5(a) focuses on the earlier decade, while Table 5(b) examines the later decade, and Table 5(c) highlights total per capita emissions over 20 years.

In [21]:

```
# Filtering data to only get the main countries with respective index val
filtered_GII = innovation.iloc[1:3,135:266]
filtered_GII['Indicator'] = innovation.iloc[1:3, 0].values

# Move the 'Indicator' column to the first position
columns = ['Indicator'] + [col for col in filtered_GII.columns if col !=
filtered_GII = filtered_GII[columns]

filtered_GII = filtered_GII.set_index('Indicator').T # Flip rows and colu
filtered_GII = filtered_GII.rename_axis('Country') # Rename the index to
filtered_GII.index = filtered_GII.index.str.rstrip('.1')

# merging all the filtered data:
all_indicators = filtered_GII.merge(merged_indicators, on = 'Country', ho
```

In [22]:

```
display(Markdown("**Table 7: Global Innovation Metrics and Environmental
all_indicators.head(5)
```

**Table 7: Global Innovation Metrics and Environmental and Technological Indicators by Country**

Out[22]:

| Country | Global Innovation Index | Innovation Efficiency Ratio | Gdp Per Unit of Energy Use | ISO 14001 environmental certificates | Research and development (R&D) | High-tech imports | n |
|---|---|---|---|---|---|---|---|
| **Albania** | 27.1 | 0.4 | 44.1 | 29.9 | 0.0 | 1.7 | |
| **Algeria** | 19.5 | 0.3 | 29.7 | 1.2 | 5.1 | 28.1 | |
| **Argentina** | 28.3 | 0.5 | 27.8 | 11.9 | 28.1 | 29.0 | |
| **Armenia** | 32.6 | 0.7 | 21.4 | 0.6 | 1.2 | 19.5 | |
| **Australia** | 48.4 | 0.5 | 26.2 | 16.0 | 59.4 | 34.0 | |

This table provides a preview of the top 5 rows (Alphabetically) for key innovation metrics and environmental indicators by country that will be used for analysis. It displays a merged dataset combining key global innovation metrics with environmental and technological indicators by country. It integrates data from previous tables. By consolidating these diverse metrics, the table provides a comprehensive view of each country's performance in innovation, sustainability, and technological advancement, enabling more holistic analysis and cross-country comparisons.

In [23]:
```
# Merging Aggregate Per Capita emissions with innovation standards and In
merged_data_perCapita = filtered_GII.merge(total_per_capita_emissions,
                                          left_index=True, right_on='Coun

# Move the "Country" column to the start
columns = ['Country'] + [col for col in merged_data_perCapita.columns if
merged_data_perCapita = merged_data_perCapita[columns]

merged_data_perCapita = merged_data_perCapita.merge(result, on  = 'Countr

display(Markdown("**Table 8: Aggregate Per Capita Emissions with Innovati
all_indicators.head(5)
```

**Table 8: Aggregate Per Capita Emissions with Innovation Standards categorised by Income Groups**

Out[23]:

| Country | Global Innovation Index | Innovation Efficiency Ratio | Gdp Per Unit of Energy Use | ISO 14001 environmental certificates | Research and development (R&D) | High-tech imports | m |
|---|---|---|---|---|---|---|---|
| Albania | 27.1 | 0.4 | 44.1 | 29.9 | 0.0 | 1.7 | |
| Algeria | 19.5 | 0.3 | 29.7 | 1.2 | 5.1 | 28.1 | |
| Argentina | 28.3 | 0.5 | 27.8 | 11.9 | 28.1 | 29.0 | |
| Armenia | 32.6 | 0.7 | 21.4 | 0.6 | 1.2 | 19.5 | |
| Australia | 48.4 | 0.5 | 26.2 | 16.0 | 59.4 | 34.0 | |

Table 8 integrates total per capita $CO_2$ emissions with innovation indicators and income group classifications, presenting a comprehensive dataset for the first five countries. By linking environmental performance with innovation standards and economic classifications, this table enables a multidimensional analysis to explore correlations and disparities across countries, highlighting how income and innovation influence environmental sustainability.

In [24]:
```
# Select specific columns to create a new DataFrame
selected_columns = [
    ' Global Innovation Index',
    'Research and development (R&D)',
    'High-tech imports',
    'High-tech and medium high-tech manufacturing, %'
]

new_dataframe = all_indicators[selected_columns]

# Display the new DataFrame
new_dataframe = new_dataframe.merge(result, on= 'Country', how = 'inner')
new_dataframe = new_dataframe.drop(columns = ['Country Code'])
columns_to_convert = ['Research and development (R&D)', 'High-tech import
                      'High-tech and medium high-tech manufacturing, %']

for column in columns_to_convert:
    new_dataframe[column] = pd.to_numeric(new_dataframe[column], errors='
```

In [25]:
```
# Filter out non-numeric columns
numeric_columns = new_dataframe.select_dtypes(include='number').columns

# Group by 'IncomeGroup' and calculate the mean for numeric columns only
averages_by_income_group = new_dataframe.groupby('IncomeGroup')[numeric_c

# Optional: Rename columns for clarity
averages_by_income_group.rename(columns={
```

```
    'Global Innovation Index': 'Average Global Innovation Index',
    'Research and development (R&D)': 'Average R&D',
    'High-tech imports': 'Average High-tech Imports',
    'High-tech and medium high-tech manufacturing, %': 'Average High-tech
}, inplace=True)

display(Markdown("**Table 9: Average Innovation Metrics by Income Group**
averages_by_income_group.head(5)
```

**Table 9: Average Innovation Metrics by Income Group**

Out[25]:

| | IncomeGroup | Global Innovation Index | Average R&D | Average High-tech Imports | Average High-tech Manufacturing |
|---|---|---|---|---|---|
| **0** | High income | 43.809091 | 35.856818 | 24.895455 | 38.159091 |
| **1** | Low income | 19.550000 | 1.420000 | 25.320000 | 2.590000 |
| **2** | Lower middle income | 24.235714 | 4.528571 | 24.628571 | 14.635714 |
| **3** | Upper middle income | 30.912121 | 11.521212 | 29.651515 | 20.912121 |

This table displays the average values of key innovation indices, including the Global Innovation Index, R&D spending, high-tech imports, and high-tech manufacturing, categorized by income group

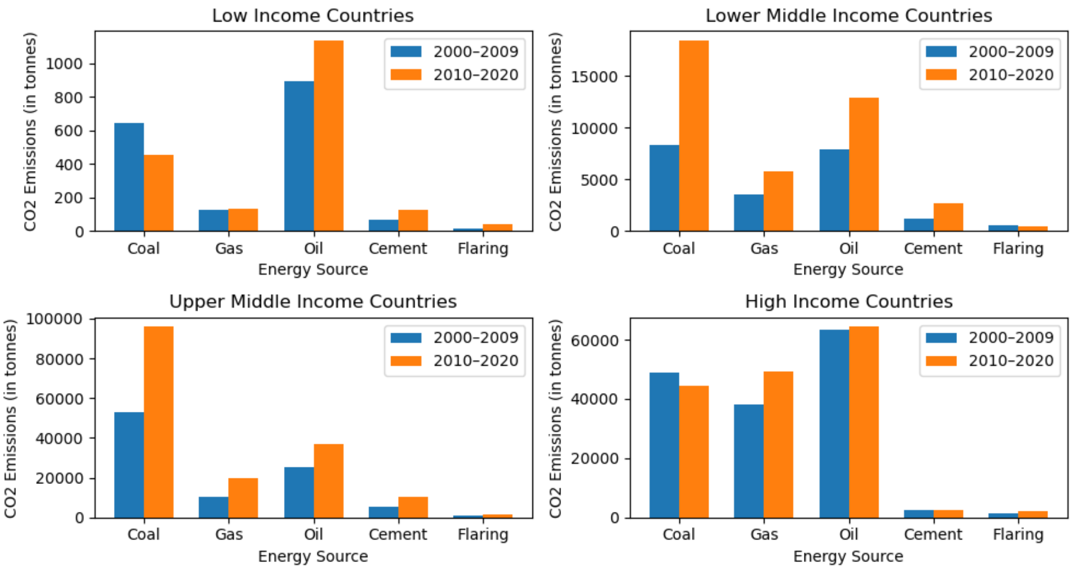
## 1.4 Plots, Figures and Histograms

In [26]:

```
country_emissions_source1 = emissions_by_source_2000_2009.merge(result, o
country_emissions_source2 = emissions_by_source_2010_2020.merge(result, o
# Example data filtering
low_income = country_emissions_source1[country_emissions_source1['IncomeG
lower_middle_income = country_emissions_source1[country_emissions_source1
upper_middle_income = country_emissions_source1[country_emissions_source1
high_income = country_emissions_source1[country_emissions_source1['Income

# Helper function to plot bar charts
def plot_income_group_bar_chart(group_data1, group_data2, title):
    energy_sources = ['Coal', 'Gas', 'Oil', 'Cement', 'Flaring']
    bar_width = 0.35
    x = np.arange(len(energy_sources))

    # Total emissions for the group in the first and second decade
    emissions_2000_2009 = group_data1.iloc[:, 1:6].sum().values
    emissions_2010_2020 = group_data2.iloc[:, 1:6].sum().values

    plt.bar(x - bar_width / 2, emissions_2000_2009, bar_width, label='200
    plt.bar(x + bar_width / 2, emissions_2010_2020, bar_width, label='201
```

```
    plt.xticks(x, energy_sources)
    plt.xlabel('Energy Source')
    plt.ylabel('CO2 Emissions (in tonnes)')
    plt.title(title)
    plt.legend()

# Plot for Low Income
plt.figure(figsize=(10, 6))
plt.subplot(2, 2, 1)
plot_income_group_bar_chart(
    low_income,
    country_emissions_source2[country_emissions_source2['IncomeGroup'] ==
    'Low Income Countries'
)

# Plot for Lower Middle Income
plt.subplot(2, 2, 2)
plot_income_group_bar_chart(
    lower_middle_income,
    country_emissions_source2[country_emissions_source2['IncomeGroup'] ==
    'Lower Middle Income Countries'
)

# Plot for Upper Middle Income
plt.subplot(2, 2, 3)
plot_income_group_bar_chart(
    upper_middle_income,
    country_emissions_source2[country_emissions_source2['IncomeGroup'] ==
    'Upper Middle Income Countries'
)

# Plot for High Income
plt.subplot(2, 2, 4)
plot_income_group_bar_chart(
    high_income,
    country_emissions_source2[country_emissions_source2['IncomeGroup'] ==
    'High Income Countries'
)

# Add a title for the entire figure
plt.suptitle(' Figure 1: CO2 Emissions by Income Group and Energy Source

# Adjust layout to make space for the title
plt.tight_layout(rect=[0, 0, 1, 0.95])

# Show the plot
plt.show()
```

**Figure 1: CO2 Emissions by Income Group and Energy Source for 2 decades (2000-2009 vs 2010-2020)**




High-income countries, with higher innovation scores, show greater reliance on cleaner energy sources like gas and oil, reflecting their capacity to adopt sustainable technologies. In contrast, lower-income countries exhibit increased emissions from coal and oil, highlighting their dependency on traditional energy sources due to limited innovation capacity. The temporal trends further emphasize that innovation plays a critical role in driving sustainable energy transitions and reducing $CO_2$ emissions globally.

In [27]:
```
# Scatter plot
plt.figure(figsize=(10, 6))
plt.scatter(
    merged_data_perCapita[' Global Innovation Index'],
    merged_data_perCapita['Total Per Capita Emissions'],
    alpha=0.7, edgecolors='k'
)

# Add labels and title
plt.title('Figure 2: Innovation Index Scores and CO2 Emissions Per Capita
plt.xlabel(' Global Innovation Index', fontsize=12)
plt.ylabel('Total Per Capita Emissions', fontsize=12)
plt.grid(True, linestyle='--', alpha=0.6)

# Calculate line of best fit
x = merged_data_perCapita[' Global Innovation Index']
y = merged_data_perCapita['Total Per Capita Emissions']
coefficients = np.polyfit(x, y, 1)  # Fit a 1st-degree polynomial (line)
trendline = np.polyval(coefficients, x)  # Evaluate the polynomial at x

# Plot the trendline
plt.plot(x, trendline, color='red', linestyle='--', label='Line of Best F

# Add legend
```

```
plt.legend()

# Show plot
plt.tight_layout()
plt.show()
```

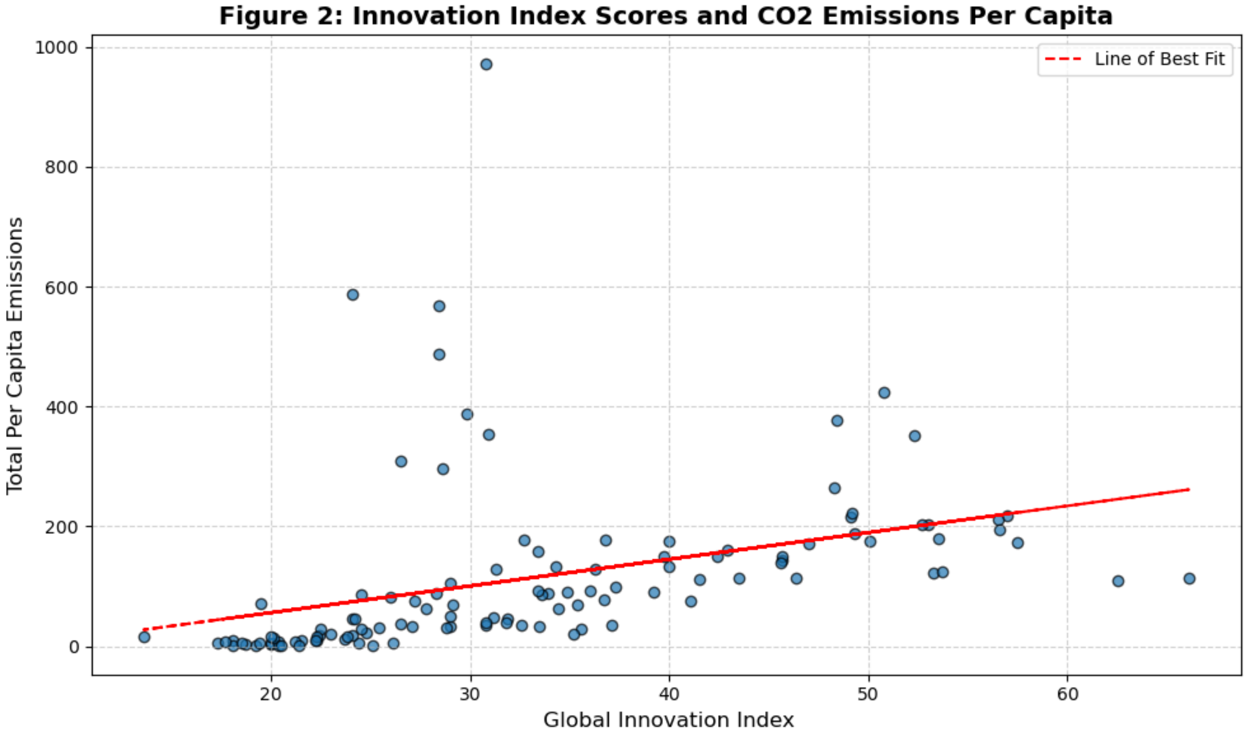



Figure 2 shows the relationship between Global Innovation Index (GII) scores and total per capita $CO_2$ emissions, without grouping by income, as per capita emissions adjust for population differences. The scatterplot with a line of best fit suggests a positive correlation, where higher GII scores are linked to increased per capita emissions due to industrialization and energy use in innovation-driven economies. However, it also highlights innovation's dual role in enabling sustainable technologies to reduce emissions. This visualization addresses the research question by connecting innovation, income, and $CO_2$ emissions.

In [28]:
```
merged_df = merged_indicators.merge(merged_data_perCapita, on='Country',

# Merge the tables on the common column (e.g., 'Country')
merged_df = merged_data_perCapita.merge(merged_indicators, on='Country',

plt.figure(figsize=(10, 6))
sns.scatterplot(
    data=merged_df,
    x='Gdp Per Unit of Energy Use',
    y='Total Per Capita Emissions',
    hue='IncomeGroup',  # Correct column name
    style='IncomeGroup',  # Optional, for different markers
    palette='viridis',
    alpha=0.8,
    s=100
```

```
)
plt.title('Figure 3: GDP Per Unit of Energy Use vs Total Per Capita Emiss
plt.xlabel('GDP Per Unit of Energy Use')
plt.ylabel('Total Per Capita Emissions')
plt.legend(title='Income Group')
plt.grid(True)
plt.show()
```

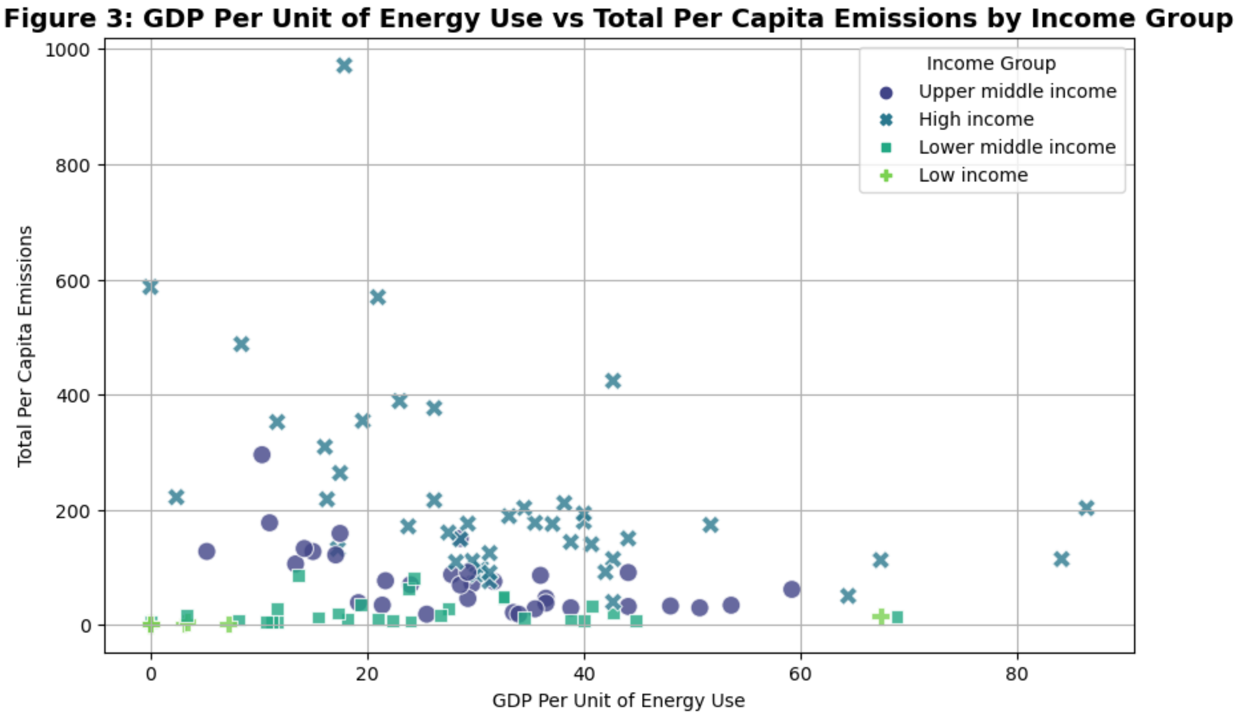
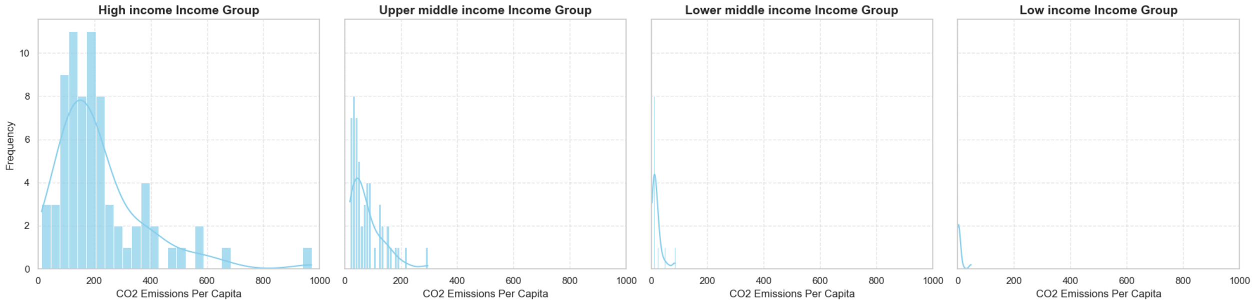



Figure 3 visualizes the relationship between GDP per unit of energy use and total per capita $CO_2$ emissions, categorized by income groups. The scatterplot reveals that high-income countries exhibit higher GDP per unit of energy use, reflecting greater energy efficiency, while also maintaining varied per capita emissions. In contrast, low and lower-middle-income countries cluster around lower GDP efficiency, with correspondingly lower emissions per capita. This graph highlights how income levels influence both economic efficiency and environmental impact, supporting the research question by demonstrating the interplay between economic innovation, energy use, and emissions across different income groups.

In [29]:
```
histogram = total_per_capita_emissions.merge(result, on= 'Country', how =
# Define the custom order for IncomeGroup

income_order = ['High income', 'Upper middle income', 'Lower middle incom
histogram['IncomeGroup'] = pd.Categorical(
    histogram['IncomeGroup'],
    categories=income_order,
    ordered=True
)

sns.set_theme(style="whitegrid")
fig, axes = plt.subplots(1, 4, figsize=(20, 5), sharey=True)
```

```
x_min, x_max = 0, 1000  # Adjust as per data range

# Plot histograms for each income group
for i, group in enumerate(income_order):
    data = histogram[histogram['IncomeGroup'] == group]
    sns.histplot(
        data=data,
        x='Total Per Capita Emissions',
        kde=True,
        ax=axes[i],
        bins=30,
        color='skyblue',  # Adjust color to match
        alpha=0.7  # Match transparency
    )
    axes[i].set_title(f'{group} Income Group', fontsize=14, fontweight='b
    axes[i].set_xlabel('CO2 Emissions Per Capita', fontsize=12)  # Add x-
    axes[i].set_ylabel('Frequency', fontsize=12) if i == 0 else axes[i].s
    axes[i].grid(True, linestyle='--', alpha=0.5)  # Add gridlines
    axes[i].set_xlim(x_min, x_max)

# Adjust layout for better appearance
plt.tight_layout()
plt.show()
```




The histograms display the distribution of $CO_2$ emissions per capita across income groups. Each panel illustrates the frequency of different per capita emission levels within each group, offering insights into their variability and typical ranges.

High-income countries show a wider range and higher frequency of emissions per capita, reflecting greater industrialization and energy use. In contrast, lower-income groups display lower emissions, with distributions concentrated at the lower end. This visualization highlights the stark contrasts between income groups in terms of $CO_2$ emissions, emphasizing the role of economic capacity and innovation in shaping environmental outcomes. These histograms directly connect to the research question by visualizing how emissions patterns differ across income levels in the context of global innovation.

In [30]:
```
averages_by_income_group.set_index('IncomeGroup', inplace = True)

# Create a horizontal bar plot
fig, ax = plt.subplots(figsize=(12, 6))
```

```
# Define metrics and y-axis locations
metrics = averages_by_income_group.columns
income_groups = averages_by_income_group.index
y_pos = np.arange(len(income_groups))

# Set bar width and spacing
bar_width = 0.2

# Define custom colors for the metrics
custom_colors = ['#1f77b4', '#ff7f0e', '#2ca02c', '#d62728']  # Example c

# Plot each metric as a bar for each income group
for i, (metric, color) in enumerate(zip(metrics, custom_colors)):
    ax.barh(y_pos + i * bar_width, averages_by_income_group[metric],
            height=bar_width, label=metric, color=color)

# Add labels, title, and legend
ax.set_yticks(y_pos + bar_width * (len(metrics) - 1) / 2)
ax.set_yticklabels(income_groups)
ax.set_xlabel('Average Value')
ax.set_title('Figure 5: Average Innovation Metrics for 20 years by Income
ax.legend(title='Innovation Metrics', bbox_to_anchor=(1.05, 1), loc='uppe

# Adjust layout for better visibility
plt.tight_layout()
plt.show()
```

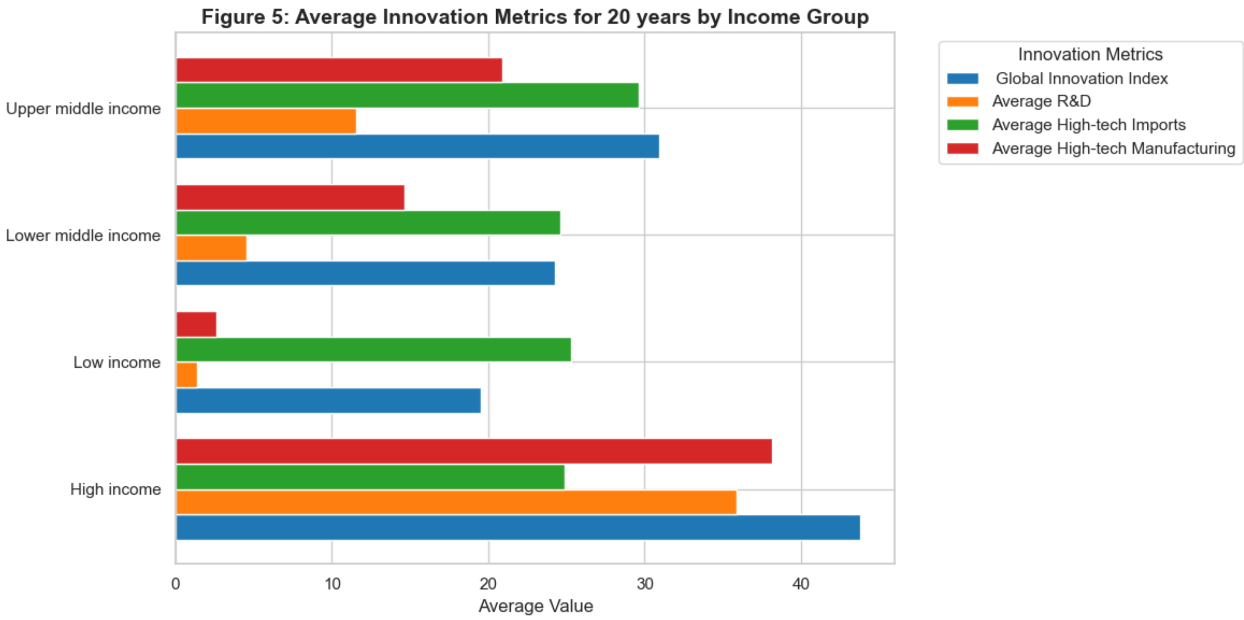



Figure 5 compares average innovation metrics over 20 years across income groups. High-income countries lead in all metrics, with the highest GII and R&D investments, reflecting their capacity for innovation-driven growth. Upper-middle-income countries show moderate innovation levels, while lower-middle and low-income countries lag, especially in R&D and high-tech manufacturing. This visualization highlights the link between income levels and innovation capacity, showing how resources and infrastructure drive disparities in technological advancement and environmental outcomes.

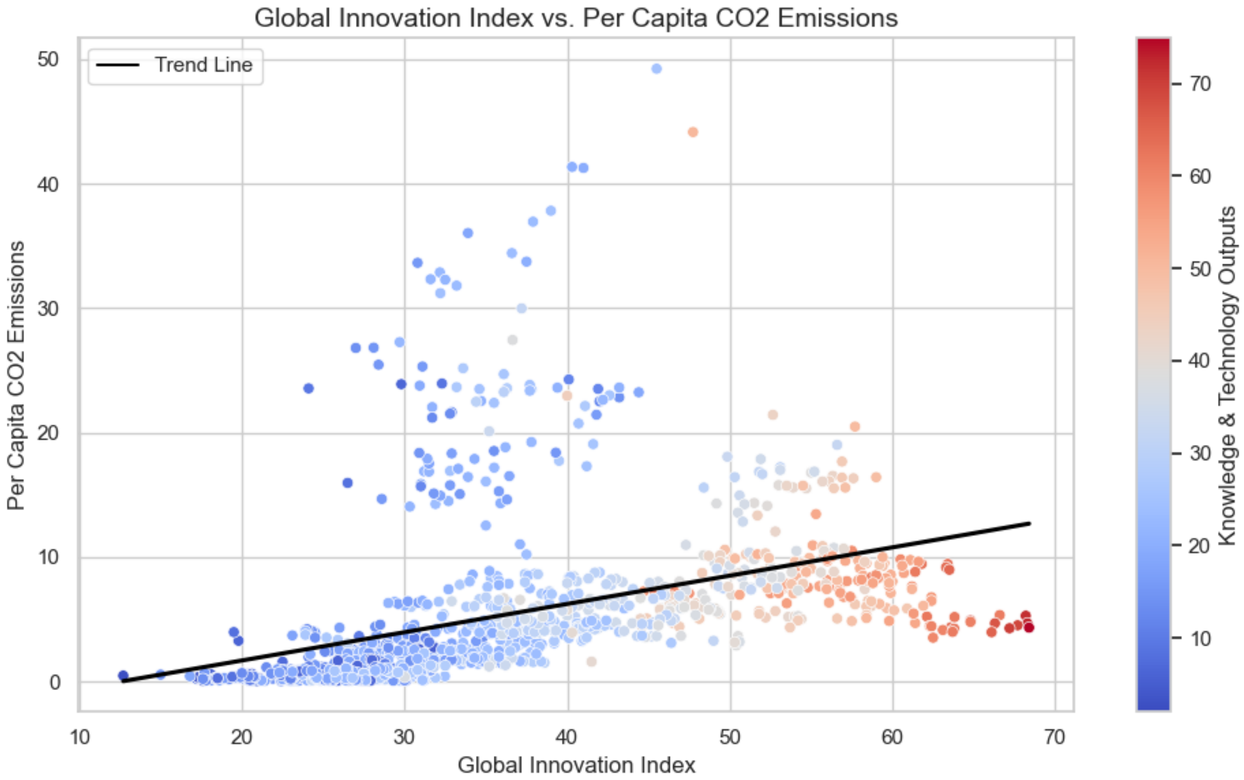
# PROJECT 2

## 2.1 Message

Innovation alone does not necessarily reduce $CO_2$ emissions; its relationship is influenced by knowledge & technology outputs, income levels and institutional support.

In [31]:
```
import matplotlib.pyplot as plt
import seaborn as sns
import matplotlib as mpl
import matplotlib.lines as mlines

# Set up the figure
fig, ax = plt.subplots(figsize=(10, 6))

# Scatter plot with color only (NO bubble size)
scatter = sns.scatterplot(
    data=reg_merged,
    x="Global Innovation Index",
    y="Per Capita CO2 Emissions",
    hue="Knowledge and technology outputs",
    palette="coolwarm",
    legend=False,
    ax=ax
)

# Add trend line
sns.regplot(
    data=reg_merged,
    x="Global Innovation Index",
    y="Per Capita CO2 Emissions",
    scatter=False,
    color='black',
    ci=None,
    ax=ax
)

# Add colorbar manually to explain color gradient
norm = plt.Normalize(
    vmin=reg_merged["Knowledge and technology outputs"].min(),
    vmax=reg_merged["Knowledge and technology outputs"].max()
)
sm = plt.cm.ScalarMappable(cmap="coolwarm", norm=norm)
sm.set_array([])
```

```
cbar = fig.colorbar(sm, ax=ax)
cbar.set_label("Knowledge & Technology Outputs", fontsize=12)

# Labels and title
ax.set_title("Global Innovation Index vs. Per Capita CO2 Emissions", font
ax.set_xlabel("Global Innovation Index", fontsize=12)
ax.set_ylabel("Per Capita CO2 Emissions", fontsize=12)

# Create manual legend for the black trend line
trend_legend = mlines.Line2D([], [], color='black', label='Trend Line')

# Add it to the top-left
ax.legend(handles=[trend_legend], loc='upper left', frameon=True)

# Show plot
plt.tight_layout()
plt.show()
```



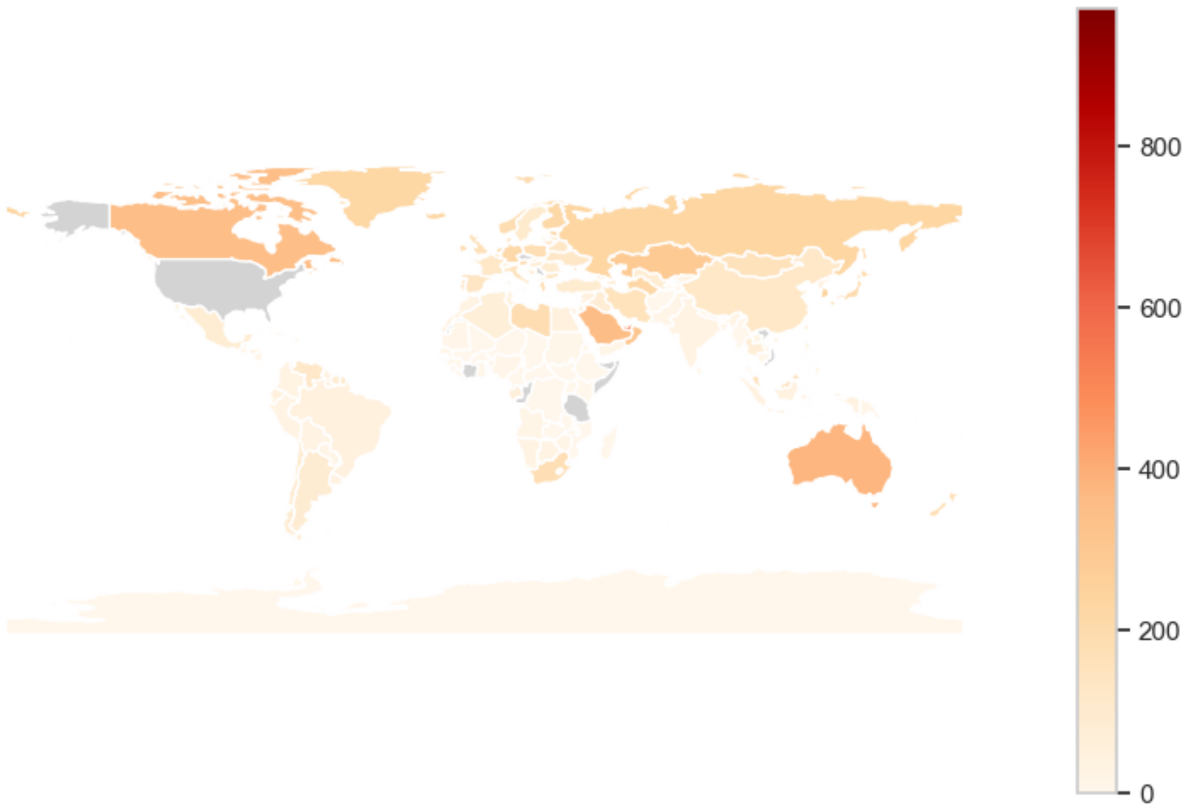

This scatter plot shows that higher innovation levels are generally associated with higher per capita $CO_2$ emissions. However, countries with stronger Knowledge & Technology Outputs (darker red) sometimes have lower emissions, suggesting that innovation alone is not enough — it must be paired with the right knowledge systems or policies.

## 2.2 Maps and Intepretation

In [32]:
```
world = gpd.read_file("ne_110m_admin_0_countries")
# Standardize country names for merging
```

```
world["ADMIN"] = world["ADMIN"].str.strip().str.lower()
total_per_capita_emissions["Country"] = total_per_capita_emissions["Count

# Merge the emissions data with the world GeoDataFrame
world_merged = world.merge(total_per_capita_emissions, left_on="ADMIN", r

# Plot the map
fig, ax = plt.subplots(1, 1, figsize=(10, 6))
world_merged.plot(
    column="Total Per Capita Emissions",  # Column to visualize
    cmap="OrRd",  # Color scheme
    legend=True,  # Add a legend
    ax=ax,
    missing_kwds={"color": "lightgrey", "label": "No Data"}  # Handle mis
)
display(Markdown("### **Map 1: Global CO2 Emissions Per Capita (2010-2020
ax.set_axis_off()

plt.show()
```

### **Map 1: Global CO2 Emissions Per Capita (2010-2020)**

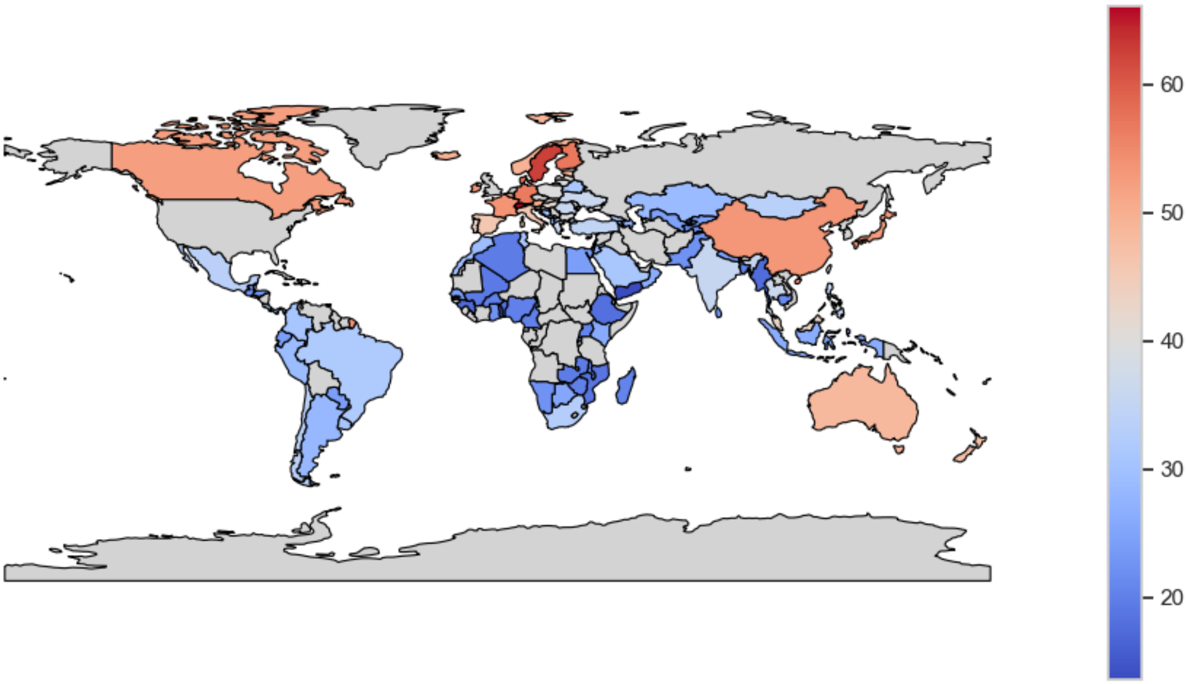



This choropleth map visualizes global $CO_2$ emissions per capita from 2010 to 2020. The color scale ranges from light beige (low emissions) to dark red (high emissions), representing differences in per capita emissions across countries.

In [33]:
```
world["ADMIN"] = world["ADMIN"].str.strip().str.lower()
filtered_GII.index = filtered_GII.index.str.strip().str.lower()
```

```
# Convert filtered_GII from index-based to a DataFrame with a country col
filtered_GII_df = filtered_GII.reset_index()

# Merge the innovation index data with the world map
world_merged = world.merge(filtered_GII_df, left_on="ADMIN", right_on="Co

# Plot the heatmap
fig, ax = plt.subplots(1, 1, figsize=(12, 6))
world_merged.plot(
    column=" Global Innovation Index",  # Column to visualize
    cmap="coolwarm",  # Color scheme
    linewidth=0.8,
    edgecolor="black",
    legend=True,
    ax=ax,
    missing_kwds={"color": "lightgrey", "label": "No Data"}  # Handle mis
)

display(Markdown("### **Map 2: Average Global Innovation Index by Country
ax.set_axis_off()

plt.show()
```

### Map 2: Average Global Innovation Index by Country (2010-2020)

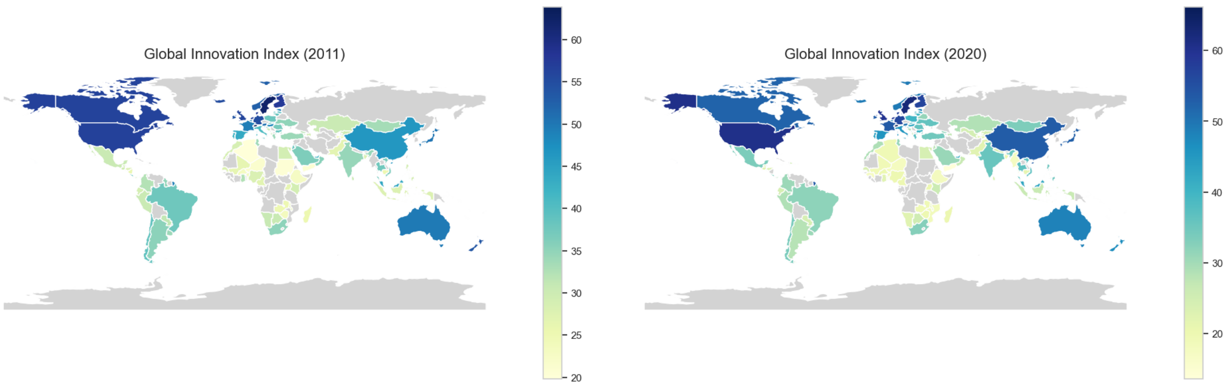



This heatmap visualizes the Global Innovation Index (GII) per country from 2010 to 2020. The color scale ranges from blue (low innovation) to red (high innovation), representing differences in innovation levels across nations.

In [34]:
```
world["ADMIN"] = world["ADMIN"].str.strip().str.lower()
innovation_2011["Country"] = innovation_2011["Country"].str.strip().str.l
```

```
innovation_2020["Country"] = innovation_2020["Country"].str.strip().str.l

world_merged_2011 = world.merge(innovation_2011, left_on="ADMIN", right_o
world_merged_2020 = world.merge(innovation_2020, left_on="ADMIN", right_o

fig, axes = plt.subplots(1, 2, figsize=(20, 6))

world_merged_2011.plot(
    column="Global Innovation Index",
    cmap="YlGnBu",
    legend=True,
    ax=axes[0],
    missing_kwds={"color": "lightgrey", "label": "No Data"}
)
axes[0].set_title("Global Innovation Index (2011)", fontsize=16)
axes[0].set_axis_off()

world_merged_2020.plot(
    column="Global Innovation Index",
    cmap="YlGnBu",
    legend=True,
    ax=axes[1],
    missing_kwds={"color": "lightgrey", "label": "No Data"}
)
axes[1].set_title("Global Innovation Index (2020)", fontsize=16)
axes[1].set_axis_off()

plt.tight_layout()
display(Markdown("### **Map 3: Global Innovation Index for two years ;201
plt.show()
```

### Map 3: Global Innovation Index for two years ;2010 and 2020




These side-by-side heatmaps illustrate the Global Innovation Index (GII) by country in 2011 and 2020, highlighting how innovation levels have evolved over the decade. The color scale ranges from light yellow (low innovation) to dark blue (high innovation).Innovation has remained concentrated in high-income regions, with North America and Western Europe continuing to lead. China's innovation growth is noticeable, reflecting its advancements in technology and R&D. Developing nations still face innovation gaps, but some regions are showing gradual progress.

# 2.3 Regression

## 2.3.1 Economic justifications for variables and relationship between X and Y

Global Innovation Index (GII)

- Represents a country's technological progress and ability to develop cleaner energy solutions.
- OECD (2019) findings suggest higher innovation levels correlate with greater energy efficiency, potentially reducing $CO_2$ emissions.

Human Capital and Research

- Investment in education and research fosters the development of green technologies (endogenous growth theory, Romer, 1990).
- NCBI (2023) shows that R&D intensity has a causal impact on lowering emissions, as countries with stronger research ecosystems develop energy-efficient solutions.

Infrastructure

- The quality of a country's infrastructure influences how energy is consumed and produced. Better public transport, renewable energy grids, and smart cities reduce emissions (World Bank, 2020).
- The urbanization-emissions hypothesis suggests that as infrastructure improves, economies transition to cleaner energy sources.

Knowledge and Technology Outputs

- Measures how much innovation translates into practical applications (e.g., patents, clean technology).
- The Porter Hypothesis (1991) suggests that technological innovation is driven by environmental policies, leading to more efficient production and lower emissions.

The relationship between $CO_2$ emissions per capita and innovation-related variables is likely non-linear, following patterns observed in economic theories. The Environmental Kuznets Curve (EKC) suggests emissions rise with industrialization before declining as clean technology adoption increases. Similarly, technology diffusion follows an S-curve, where innovation initially leads to higher emissions before reducing them as adoption scales. Your regression results show mixed coefficients, indicating a threshold effect—where Global Innovation Index and

Knowledge & Technology Outputs impact emissions differently at various stages. Infrastructure and R&D investments also exhibit diminishing returns, meaning their effectiveness in reducing emissions slows over time. This suggests a non-linear, threshold-dependent relationship, where innovation must reach a critical level before significantly lowering emissions.

## 2.3.2. Regression tables

In [35]:
```
# Define dependent variable
import statsmodels.api as sm
y = reg_merged['Per Capita CO2 Emissions']

X1 = sm.add_constant(reg_merged[['Global Innovation Index']])
X2 = sm.add_constant(reg_merged[['Global Innovation Index', 'Human capita
X3 = sm.add_constant(reg_merged[['Global Innovation Index', 'Human capita
X4 = sm.add_constant(reg_merged[['Global Innovation Index', 'Human capita

model1 = sm.OLS(y, X1).fit()
model2 = sm.OLS(y, X2).fit()
model3 = sm.OLS(y, X3).fit()
model4 = sm.OLS(y, X4).fit()
display(Markdown("**Regression 1: Determinants of Per Capita CO₂ Emission
stargazer = Stargazer([model1, model2, model3, model4])
display(HTML(stargazer.render_html()))
```

**Regression 1: Determinants of Per Capita $CO_2$ Emissions**

| | *Dependent variable: Per Capita CO2 Emissions* | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Global Innovation Index | 0.227*** | 0.026 | 0.298*** | 0.154*** |
| | (0.015) | (0.032) | (0.045) | (0.054) |
| Human capital and research | | 0.175*** | 0.144*** | 0.144*** |
| | | (0.024) | (0.024) | (0.024) |
| Infrastructure | | | | 0.104*** |
| | | | | (0.022) |
| Knowledge and technology outputs | | | -0.231*** | -0.186*** |
| | | | (0.028) | (0.030) |
| const | -2.852*** | -1.313** | -3.995*** | -4.070*** |
| | (0.572) | (0.600) | (0.669) | (0.663) |
| Observations | 1192 | 1192 | 1192 | 1192 |
| $R^2$ | 0.164 | 0.199 | 0.242 | 0.256 |
| Adjusted $R^2$ | 0.163 | 0.198 | 0.240 | 0.254 |
| Residual Std. Error | 5.942 (df=1190) | 5.819 (df=1189) | 5.663 (df=1188) | 5.612 (df=1187) |
| F Statistic | 233.634*** (df=1; 1190) | 147.819*** (df=2; 1189) | 126.421*** (df=3; 1188) | 102.320*** (df=4; 1187) |
| Note: | | | | *p<0.1; **p<0.05; ***p<0.01 |

The regression model estimates Per Capita $CO_2$ Emissions as a function of innovation and related economic factors. The Global Innovation Index (0.154, $p<0.01$) is positively correlated with emissions. A 1-point increase in GII increases $CO_2$ emissions by 0.154 metric tons per capita. This suggests that innovation, in its current form, is contributing to emissions, likely because industrialization and economic growth come with fossil fuel use.

A 1-unit increase in technology outputs reduces emissions by 0.186 metric tons per capita. This implies that specific types of innovation, such as clean energy or sustainable technologies, can counteract overall emissions growth.

A 1-point increase in education and R&D spending leads to 0.144 more metric tons of $CO_2$ emissions per capita. Better infrastructure is linked to 0.104 additional metric tons of $CO_2$ per capita, likely due to increased energy and industrial use.

The $R^2$ is 0.256 which means that the model explains 25.6% of $CO_2$ emissions variations. The F-statistic is 102.32 ($p < 0.01$). It shows that the model is statistically significant, meaning the variables collectively explain a significant portion of emissions changes.

In [36]:
```
# Ensure income levels are categorical
reg_merged["Income_Level"] = reg_merged["Income level (GNI Thresholds)"].

income_dummies = pd.get_dummies(reg_merged["Income_Level"], prefix="Incom

reg_merged = pd.concat([reg_merged, income_dummies], axis=1)
```

In [37]:
```
y = reg_merged['Per Capita CO2 Emissions']

X = sm.add_constant(reg_merged[['Global Innovation Index',
                                'Income_2.0', 'Income_3.0']])  # Assuming

reg_merged['Income_2.0'] = reg_merged['Income_2.0'].astype(int)
reg_merged['Income_3.0'] = reg_merged['Income_3.0'].astype(int)

reg_merged["Innovation_Income2"] = reg_merged["Global Innovation Index"]
reg_merged["Innovation_Income3"] = reg_merged["Global Innovation Index"]

X_interaction = sm.add_constant(reg_merged[['Global Innovation Index',
                                            'Income_2.0', 'Income_3.0',
                                            'Innovation_Income2', 'Innova

model_interaction = sm.OLS(y, X_interaction).fit()

display(Markdown("**Regression 2: Income level analysis**"))
stargazer_interaction = Stargazer([model_interaction])
display(HTML(stargazer_interaction.render_html()))
```

**Regression 2: Income level analysis**

| | *Dependent variable: Per Capita CO2 Emissions* |
|---|---|
| | (1) |
| Global Innovation Index | $0.137^{***}$ |
| | (0.018) |
| Income_2.0 | $-6.616^{***}$ |
| | (2.207) |
| Income_3.0 | $-4.191^{*}$ |
| | (2.151) |
| Innovation_Income2 | 0.079 |
| | (0.075) |
| Innovation_Income3 | 0.040 |
| | (0.061) |
| const | $2.230^{***}$ |
| | (0.808) |
| Observations | 1192 |
| $R^2$ | 0.221 |
| Adjusted $R^2$ | 0.218 |
| Residual Std. Error | 5.745 (df=1186) |
| F Statistic | $67.413^{***}$ (df=5; 1186) |
| Note: | $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$ |

The regression output presents the impact of innovation, income levels, and their interactions on per capita $CO_2$ emissions. The Global Innovation Index (0.137, $p<0.01$) has a statistically significant positive effect, indicating that higher innovation levels are associated with increased $CO_2$ emissions. The model explains about 22.1% of the variation in $CO_2$ emissions ($R^2 = 0.221$).

The highest income group (Income_4.0) is omitted for comparison, which means that coefficients for Income_2.0 and Income_3.0 reflect differences relative to the highest income group. The income dummy variables (Income_2.0 = -6.616, $p<0.01$ and Income_3.0 = -4.191, $p<0.1$) suggest that lower-income and middle-income countries have significantly lower per capita emissions compared to the highest-income group. This aligns with global trends where high-income countries tend to have more industrialization and consumption, leading to higher per capita emissions.

The interaction terms (Innovation_Income2 = 0.079, Innovation_Income3 = 0.040) are not statistically significant, implying that the relationship between innovation and emissions does not vary significantly across income levels.This suggests that innovation alone is not a strong determinant of $CO_2$ emissions when income levels are considered.

In [38]:
```
# Create interaction terms
reg_merged["Innovation_GDP"] = reg_merged["Global Innovation Index"] * re
reg_merged["Innovation_Fossil"] = reg_merged["Global Innovation Index"] *
reg_merged["Innovation_Knowledge"] = reg_merged["Global Innovation Index"
reg_merged["Innovation_Policy"] = reg_merged["Global Innovation Index"] *

y = reg_merged["Per Capita CO2 Emissions"]

X1 = sm.add_constant(reg_merged[["Global Innovation Index"]])  # Baseline
X2 = sm.add_constant(reg_merged[["Global Innovation Index", "Innovation_G
X3 = sm.add_constant(reg_merged[["Global Innovation Index", "Innovation_F
X4 = sm.add_constant(reg_merged[["Global Innovation Index", "Innovation_K

model1 = sm.OLS(y, X1).fit()
model2 = sm.OLS(y, X2).fit()
model3 = sm.OLS(y, X3).fit()
model4 = sm.OLS(y, X4).fit()

display(Markdown("**Regression 3: Influence of Economic Factors on relati
display(Markdown("We estimate four regression models, each adding interac
display(HTML(stargazer.render_html()))
```

**Regression 3: Influence of Economic Factors on relation between Innovation and CO2 Emissions**

We estimate four regression models, each adding interaction terms to explore how different economic factors influence the relationship between innovation and $CO_2$ emissions.

| | *Dependent variable: Per Capita CO2 Emissions* | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Global Innovation Index | 0.227*** | 0.026 | 0.298*** | 0.154*** |
| | (0.015) | (0.032) | (0.045) | (0.054) |
| Human capital and research | | 0.175*** | 0.144*** | 0.144*** |
| | | (0.024) | (0.024) | (0.024) |
| Infrastructure | | | | 0.104*** |
| | | | | (0.022) |
| Knowledge and technology outputs | | | -0.231*** | -0.186*** |
| | | | (0.028) | (0.030) |
| const | -2.852*** | -1.313** | -3.995*** | -4.070*** |
| | (0.572) | (0.600) | (0.669) | (0.663) |
| Observations | 1192 | 1192 | 1192 | 1192 |
| $R^2$ | 0.164 | 0.199 | 0.242 | 0.256 |
| Adjusted $R^2$ | 0.163 | 0.198 | 0.240 | 0.254 |
| Residual Std. Error | 5.942 (df=1190) | 5.819 (df=1189) | 5.663 (df=1188) | 5.612 (df=1187) |
| F Statistic | 233.634*** (df=1; 1190) | 147.819*** (df=2; 1189) | 126.421*** (df=3; 1188) | 102.320*** (df=4; 1187) |
| Note: | | | | *p<0.1; **p<0.05; ***p<0.01 |

Model 1 establishes a basic relationship between innovation and $CO_2$ emissions. Model 2 tests if the effect of innovation on $CO_2$ emissions depends on income levels. Model 3 determines if innovation's impact on emissions changes based on fossil fuel dependency. Model 4 tests if innovation is more effective at reducing $CO_2$ in knowledge-driven or policy-regulated economies.

According to the regression table we see that innovation alone increases $CO_2$ emissions, likely due to industrial expansion. It reduces emissions in wealthier countries, indicating they invest more in clean technologies. Innovation is not significantly reducing emissions in fossil-fuel-heavy nations, meaning clean energy transitions are slow. Knowledge-based economies benefit from innovation, suggesting R&D and technological output are crucial for climate action. Innovation works best in knowledge-intensive economies but requires further policy support.

### 2.3.3 : Best specification and Evaluation regressions

This analysis helps identify which economic conditions allow innovation to reduce emissions effectively. The Knowledge & Policy Model (Model 3) is the most relevant because it highlights that innovation works best when combined with strong institutions and technological advancements.

Adjusted $R^2$:Measures how well the model explains variation in $CO_2$ emissions. Higher Adjusted $R^2$ suggests the model is capturing more relevant factors.

Statistical Significance (p-values): Coefficients should be statistically significant ($p < 0.05$ or $p < 0.01$) to confirm their effect is not random.

Sign and Magnitude of Coefficients:A negative coefficient on Innovation × Knowledge suggests that knowledge-intensive economies use innovation effectively to reduce emissions. A positive coefficient on Innovation × Fossil Fuels suggests that innovation is not mitigating emissions in fossil-heavy economies.

# 3. Final Project

## 3.1 Potential Data to Scrape

Data Scraped from: https://www.lens.org The green patent data enhances my paper by providing a focused view of sustainability-driven innovation. While the Global Innovation Index reflects overall innovation capacity, green patent counts highlight environmental innovation efforts.

I scraped country-level green patent data from the search results on Lens.org using filters for green technologies based on the IPC Green Inventory. After standardizing country names, I merged this with my existing dataset to compare general innovation with green innovation. This helps assess whether highly innovative countries are also leaders in climate-focused innovation, directly supporting my research on innovation's impact on $CO_2$ emissions.

Although I didn't merge the datasets to avoid losing country-level data, I compared them side by side to examine whether the most innovative countries are also leading in green innovation, helping address my research question on innovation's role in reducing $CO_2$ emissions.

## 3.2 Potential Challenges

Scraping data from Lens.org can be tricky because the website loads content using JavaScript, so I couldn't use basic scraping tools like requests or BeautifulSoup. Instead, I had to use Selenium, which opens a browser and is slower to run. The site also has limits on how much data you can access at once and may require logging in, which makes it hard to scrape a lot of data or do it regularly.

For this project, I didn't need to run the scraper over time—I just collected the green patent counts for the top countries once. So while I couldn't scrape everything, I was able to get enough data to compare and visualize the top 15–25 countries for my analysis.

However, since patent data changes over time, I would need to scrape it again each year to keep the data up to date and track changes over time.

## 3.3 Data Scraping from Website

In [94]:
```
!pip install -q undetected-chromedriver # I have to import these here as
```

In [95]:
```
# I have to import these here as if i do it above, the code crashes
import undetected_chromedriver as uc
from selenium.webdriver.common.by import By
import time
```

In [96]:
```
# open the browswer
chrome_options = uc.ChromeOptions()
driver = uc.Chrome(options=chrome_options)
```

In [97]:
```
# get the url for website
url = "https://www.lens.org/lens/search/patent/list?p=0&n=10&s=_score&d=%
driver.get(url)
```

In [98]:
```
# I have to import these here as if i do it above, the code crashes
from selenium.webdriver.support.ui import WebDriverWait
from selenium.webdriver.support import expected_conditions as EC
from selenium.webdriver.common.by import By

page_source = driver.page_source
# Wait for the Jurisdiction <h3> to be clickable and then click it
jurisdiction_header = WebDriverWait(driver, 10).until(
    EC.element_to_be_clickable((
        By.XPATH,
        # Look for an h3 that has a child <span> exactly matching "Jurisd
```

```
        "//h3[span[@class='list-icons--label ng-binding' and text()='Juri
    ))
)
jurisdiction_header.click()
```

In [100…

```
load_more_button = WebDriverWait(driver, 10).until(
    EC.element_to_be_clickable((
        By.CSS_SELECTOR,
        "li.facets-more--link.link.ng-scope > a.ng-binding"  # or any uni
    ))
)
load_more_button.click()
```

In [101…

```
# Extract the country & count for each label
labels = driver.find_elements(
    By.CSS_SELECTOR,
    "label.checkbox.tri-state-checkbox.ng-binding.ng-scope"
)
```

In [102…

```
results = []
for lbl in labels:
    # e.g., lbl.text might be "(239,044)\nChina"
    text = lbl.text.strip()
    lines = text.split("\n")
    if len(lines) == 2:
        raw_count = lines[0].strip()   # "(239,044)"
        country = lines[1].strip()    # "China"

        # If you want just digits, remove parentheses/commas
        count_str = raw_count.strip("()").replace(",", "")

        results.append({"country": country, "count": count_str})
df = pd.DataFrame(results)
display(Markdown("#### **Table: Top 5 Countries by Green Patent Counts (S
df.head(5)
```

#### **Table: Top 5 Countries by Green Patent Counts (Scraped Data)**

Out[102…

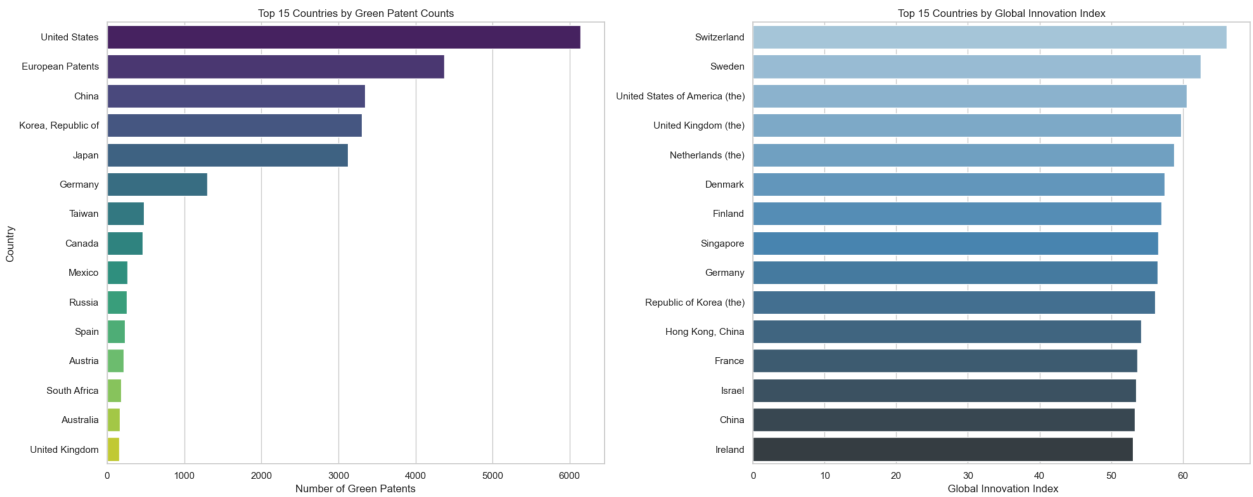
| | country | count |
|---|---|---|
| **0** | United States | 6143 |
| **1** | European Patents | 4371 |
| **2** | China | 3345 |
| **3** | Korea, Republic of | 3303 |
| **4** | Japan | 3123 |

## 3.4 Visualising the Scraped Dataset

In [103…
```
df['count'] = df['count'].astype(int)

top15 = df.sort_values(by='count', ascending=False).head(15)

filtered_GII_new_reset = filtered_GII_new.reset_index()

# Ensure 'Global Innovation Index' is numeric
filtered_GII_new_reset[' Global Innovation Index'] = pd.to_numeric(filter

top15_gii = filtered_GII_new_reset.sort_values(by=' Global Innovation Ind
```

In [104…
```
sns.set(style="whitegrid")
fig, axes = plt.subplots(nrows=1, ncols=2, figsize=(20, 8))

# Plot 1: Green Patent Counts
sns.barplot(data=top15, x='count', y='country', palette='viridis', ax=axe
axes[0].set_title('Top 15 Countries by Green Patent Counts')
axes[0].set_xlabel('Number of Green Patents')
axes[0].set_ylabel('Country')

# Plot 2: Global Innovation Index
sns.barplot(data=top15_gii, x=' Global Innovation Index', y='Country', pa
axes[1].set_title('Top 15 Countries by Global Innovation Index')
axes[1].set_xlabel(' Global Innovation Index')
axes[1].set_ylabel('')

# Adjust layout
plt.tight_layout()
display(Markdown("### **Comparing Innovation Index and Focus on Green Pat
plt.show()
```

### **Comparing Innovation Index and Focus on Green Patents by Country**




The two charts highlight the top 15 countries by Global Innovation Index and Green Patent Counts. While countries like the United States, Germany, and South Korea rank high in both, others show a stronger presence in one metric. Switzerland and

Sweden lead in overall innovation, whereas the U.S. and China dominate in green patents, reflecting a focus on sustainable technology. This comparison shows that while there is overlap, countries vary in their emphasis on general innovation versus environmentally focused innovation.

# 3.5 Regression Tree

## Regression Tree Objective Function

The objective function of a regression tree is to minimize the **total squared error** across all leaf nodes:

$$\min_T \sum_{j=1}^{J} \sum_{i \in R_j} \left( y_i - \bar{y}_{R_j} \right)^2$$

Where:

- ( y_i ): actual $CO_2$ emissions per capita for country-year ( i )
- ( \overline{y}_{R_j} ): mean $CO_2$ emissions per capita for the leaf node (region) ( R_j )
- ( R_j ): set of observations in the ( j^{th} ) region (leaf)
- ( J ): total number of leaves in the tree
- ( T ): the structure of the tree (how it splits on variables like `Global Innovation Index`, `R&D`, etc.)

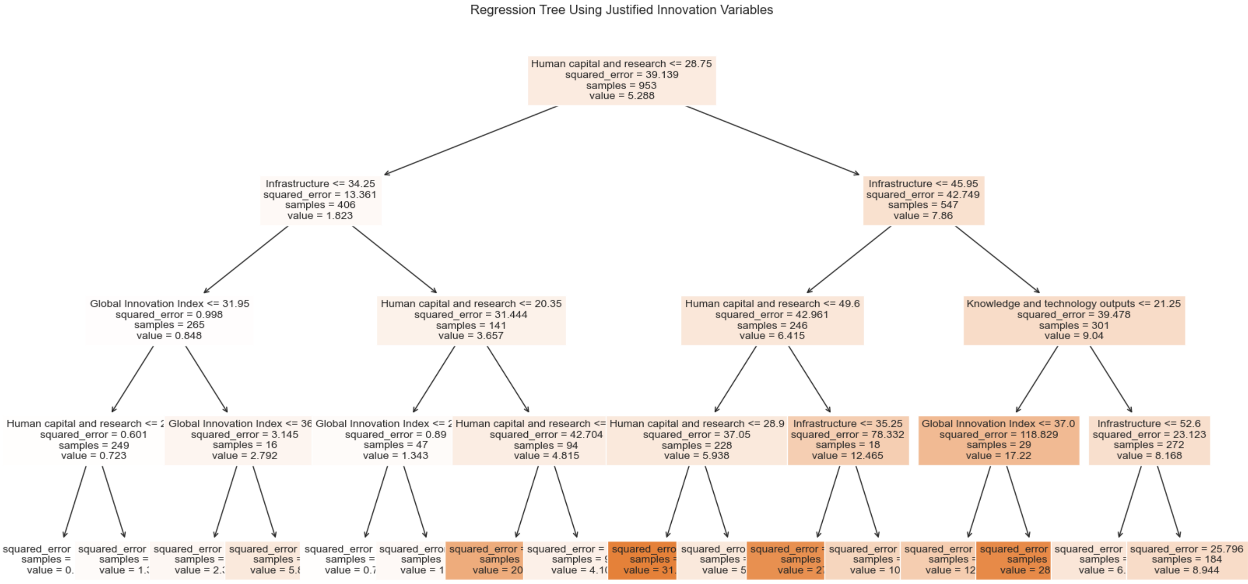
This objective ensures that each split reduces the variation in emissions within the resulting groups.

In this project, we use a regression tree to study how innovation factors like the Global Innovation Index, R&D spending, and technology output affect a country's per capita $CO_2$ emissions. The tree works by splitting countries into groups based on these indicators, so that countries in each group have similar $CO_2$ emissions. Its goal is to minimize the total difference between actual emissions and the average emissions in each group. This helps us find non-linear patterns and interactions between innovation variables that a regular regression might miss.

## Preferred Model Regression Tree (Justified Xs Only)

In [93]:
```
selected_Xs = [
    'Global Innovation Index',
    'Human capital and research',
    'Infrastructure',
    'Knowledge and technology outputs'
```

```
]

# Combine selected X variables and y into one DataFrame
data = reg_merged[selected_Xs + ['Per Capita CO2 Emissions']].dropna()

# Split cleanly into X and y
X_selected = data[selected_Xs]
y_selected = data['Per Capita CO2 Emissions']

X_train, X_test, y_train, y_test = train_test_split(X_selected, y_selecte

tree_model = DecisionTreeRegressor(max_depth=4, random_state=42)
tree_model.fit(X_train, y_train)

plt.figure(figsize=(20, 10))
plot_tree(tree_model, feature_names=X_selected.columns, filled=True, font
plt.title("Regression Tree Using Justified Innovation Variables")
plt.show()

# Step 4: Evaluate model performance
y_pred = tree_model.predict(X_test)
mse = mean_squared_error(y_test, y_pred)
print(f"Test MSE (Preferred Variables Tree): {mse:.2f}")
```




```
Test MSE (Preferred Variables Tree): 50.11
```

Interpretation:

This model uses only the core innovation indicators: Global Innovation Index, Human capital and research, Infrastructure, and Knowledge and technology outputs. The root node still splits on human capital, confirming its importance in explaining differences in emissions. The tree shows that countries with low scores in innovation and infrastructure tend to have lower $CO_2$ emissions, while countries that invest more in these areas tend to have higher emissions per capita.

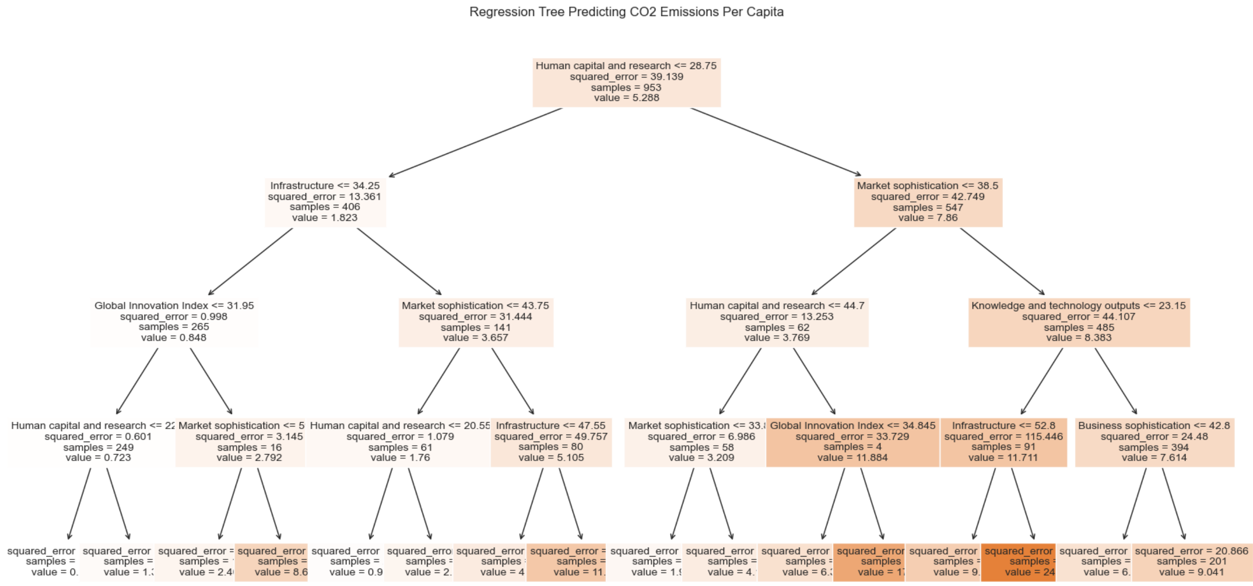
This result may reflect the development-emissions trade-off: countries that are more innovative and infrastructure-rich also tend to be more industrialized, which leads to higher emissions. However, knowledge and technology outputs also appear in later splits, which could indicate that innovation eventually helps moderate emissions in more developed economies.

Error of Prediction:

Test MSE: 50.11 This model has a higher error compared to the full model, suggesting that it doesn't capture as much variation in emissions. However, it is more interpretable and theoretically sound, making it a better fit for policy-oriented analysis. It's also less likely to overfit.

## Full Model Regression Tree (All Xs)

In [69]:
```
# Define predictors (Xs) from reg_merged
predictors = [
    'Global Innovation Index',
    'Human capital and research',
    'Infrastructure',
    'Knowledge and technology outputs',
    'Market sophistication',
    'Business sophistication',
    'Creative outputs'
]

# Drop missing values for those predictors + Y
reg_tree_data = reg_merged.dropna(subset=predictors + ['Per Capita CO2 Em

X = reg_tree_data[predictors]
y = reg_tree_data['Per Capita CO2 Emissions']

# Split data
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,

# Fit regression tree
tree = DecisionTreeRegressor(max_depth=4, random_state=42)
tree.fit(X_train, y_train)

# Visualize the tree
plt.figure(figsize=(20, 10))
plot_tree(tree, feature_names=predictors, filled=True, fontsize=10)
plt.title("Regression Tree Predicting CO2 Emissions Per Capita")
plt.show()

# Evaluate performance
y_pred = tree.predict(X_test)
mse = mean_squared_error(y_test, y_pred)
print(f"Test MSE: {mse:.2f}")
```




```
Test MSE: 36.59
```

#### Interpretation

The regression tree shows that Human Capital and Research is the most important factor driving per capita $CO_2$ emissions, as it's the first split in the tree. This suggests that countries with low investment in education and R&D (below ~28.75) tend to have significantly lower emissions, possibly due to lower industrial development. For countries with higher levels of human capital, the next important split is often on Market Sophistication and Infrastructure; More sophisticated markets and stronger infrastructure likely enable higher levels of industrial activity, which can increase emissions.

The Global Innovation Index appears deeper in the tree — suggesting that while overall innovation is relevant, its effect on emissions is more nuanced and depends on how it's combined with other factors like infrastructure and human capital. The tree also shows that Knowledge and Technology Outputs and Business Sophistication matter for high-innovation countries, possibly because tech-intensive industries contribute to emissions in middle- to high-income economies.

#### Error of prediction

Test MSE: 36.59 This relatively low error suggests that the full model captures a substantial amount of the variance in emissions across countries. However, including all variables increases the risk of overfitting, meaning some splits might reflect noise rather than meaningful economic relationships.

#### Comparing the two models:

The full model performs better in terms of prediction accuracy, but the preferred model aligns more closely with economic theory and is easier to explain. This trade-off between interpretability and accuracy is important in policy analysis and model design.

## 3.6 Random Forest

In [108…

```
# Train Random Forest on your selected variables
rf_model = RandomForestRegressor(n_estimators=100, random_state=42)
rf_model.fit(X_train, y_train)

# Predict and evaluate
y_pred_rf = rf_model.predict(X_test)
mse_rf = mean_squared_error(y_test, y_pred_rf)
print(f"Random Forest Test MSE: {mse_rf:.2f}")

# Get feature importances
importances = rf_model.feature_importances_
feature_names = X_selected.columns

# Create a sorted DataFrame for easy interpretation
importance_df = pd.DataFrame({
    'Feature': feature_names,
    'Importance': importances
}).sort_values(by='Importance', ascending=False)

display(Markdown("**Random Forest Regression**"))
display(importance_df)
```

```
Random Forest Test MSE: 37.62
```

**Random Forest Regression**

| | Feature | Importance |
|---|---|---|
| **1** | Human capital and research | 0.368014 |
| **0** | Global Innovation Index | 0.252963 |
| **2** | Infrastructure | 0.192417 |
| **3** | Knowledge and technology outputs | 0.186605 |

The Random Forest model helps us better understand which innovation factors are most important in predicting a country's $CO_2$ emissions per capita. According to the model, "Human capital and research" was the most important variable. This supports what we saw in the regression trees — countries that invest more in education and research tend to have different emission levels. "Infrastructure" also ranked high in importance, which makes sense because better infrastructure often comes with more industrial activity and energy use, which can increase emissions.

On the other hand, "Global Innovation Index" had a lower importance in the Random Forest, which is interesting. It suggests that while this index gives a broad overview of innovation, it's not as useful when we already include more specific variables like infrastructure and education. One surprise was that "Knowledge and technology outputs" ranked a bit higher here than in the regression tree. This might mean that even countries producing a lot of new technology still have high emissions, maybe because of how energy-intensive innovation can be.

# 3.7 OLS VS ML

The OLS regression provides an overall understanding of how each innovation-related factor is associated with $CO_2$ emissions per capita. It assumes a linear relationship — meaning the effect of each variable is consistent across all countries. The results showed that variables like Human capital and research, Infrastructure, and the Global Innovation Index are positively associated with emissions, while Knowledge and technology outputs have a negative relationship. This suggests that while development and innovation tend to increase emissions, the effective application of innovation (such as green technology) can help reduce them.

However, the OLS model does not account for nonlinear relationships or interactions between variables. It treats all countries the same, regardless of their development stage or context. This is where the regression tree model adds value. Instead of assuming uniform effects, the tree model splits the data based on variable thresholds — such as when human capital is below or above a specific value. This allows the model to better capture complex behaviors and turning points in the data.

From an economic standpoint, the regression tree provides more context-specific insights. It shows that countries with similar levels of a particular variable (e.g., infrastructure) can have very different emissions depending on other factors. These conditional patterns are especially useful for policymakers, as they highlight when and where certain variables start to have a stronger or weaker impact.

Therefore, while OLS gives a broad and interpretable view of the average relationships, the regression tree reveals deeper nonlinear structures and threshold effects. Together, they offer a more complete picture of how innovation relates to environmental outcomes.

# 3.8 CONCLUSION

This project examined whether innovation can help reduce global $CO_2$ emissions. Using combined datasets—including innovation scores, emissions data, income levels, and tech indicators—we explored how innovation-related factors influence per capita emissions across countries.

Linear regressions showed that higher innovation (measured by the Global Innovation Index) often correlates with higher emissions, likely due to industrialization and energy use. However, specific types of innovation—like “Knowledge and Technology Outputs”—were linked to lower emissions, suggesting that the nature of innovation matters more than its quantity.

Regression trees and random forests added more depth. These models highlighted “Human Capital and Research” and “Infrastructure” as stronger predictors of emissions than broad innovation scores. Trees also revealed thresholds—points where emissions shift sharply based on variable levels—offering insights not captured by linear models. The importance matrix from the random forest confirmed that targeted investments in education and infrastructure are key to lowering emissions.

Future research could use panel data to track changes over time, and include policy variables like carbon taxes or renewable energy targets to improve predictions. Moreover, more advanced models could help uncover complex interactions missed by basic trees.

Comparing countries with and without strong climate policies may reveal what innovation strategies work best along with adding richer variables—like energy types or environmental laws—could further strengthen the analysis.

Overall, innovation can help reduce emissions, but only when focused on clean, efficient, and socially beneficial outcomes. Policymakers should prioritize education, sustainable infrastructure, and environmentally focused innovation—such as green technology development—for more effective climate strategies.

Word Count :

# 4. CITATION

• Acemoglu, D., Aghion, P., Bursztyn, L., & Hemous, D. (2014). The environment and directed technical change. American Economic Review, 102(1), 131-166. https://doi.org/10.1257/aer.102.1.131

• Friedlingstein, P., Jones, M. W., O’Sullivan, M., Andrew, R. M., et al. (2022). Global carbon budget 2022. Earth System Science Data, 14(8), 4653-4723. https://doi.org/10.5194/essd-14-4653-2022

• Frontiers in Environmental Science. (2022). The impact of innovation on $CO_2$ emissions: The threshold effect of financial development. Frontiers in Environmental Science. Retrieved from https://www.frontiersin.org/journals/environmental-science/articles/10.3389/fenvs.2022.980267/full

• Free Policy Briefs. (2024). The impact of technological innovations and economic growth on $CO_2$ emissions in different countries. Free Policy Briefs. Retrieved from https://freepolicybriefs.org/2024/04/08/economic-growth-co2-emissions/

• International Energy Agency (IEA). (2021). Energy efficiency 2021: A global analysis of energy efficiency trends and policies. IEA Publications. Retrieved from https://www.iea.org/reports/energy-efficiency-2021

• Intergovernmental Panel on Climate Change (IPCC). (2021). Climate change 2021: The physical science basis.Cambridge University Press. Retrieved from https://www.ipcc.ch/report/ar6/wg1/

• National Center for Biotechnology Information (NCBI). (2021). A comparative analysis of the relationship between innovation and carbon emissions in the transportation sector. NCBI. Retrieved from https://pmc.ncbi.nlm.nih.gov/articles/PMC8055480/

• National Center for Biotechnology Information (NCBI). (2023). Revisiting R&D intensity and $CO_2$ emissions link in the USA using time-varying Granger causality. NCBI. Retrieved from https://pmc.ncbi.nlm.nih.gov/articles/PMC10520821/

• Nordhaus, W. D. (2019). Climate change: The ultimate challenge for economics. American Economic Review, 109(6), 1991-2014. https://doi.org/10.1257/aer.109.6.1991

• Organisation for Economic Co-operation and Development (OECD). (2019). Innovation and environmental sustainability: Policy insights. OECD Publishing. https://doi.org/10.1787/9789264307438-en

• Organisation for Economic Co-operation and Development (OECD). (2019). Measuring innovation in energy technologies. OECD Publishing. https://doi.org/10.1787/9789264311992-en

• Popp, D. (2012). The role of technological change in green growth. Energy Economics, 34(1), S63-S74. https://doi.org/10.1016/j.eneco.2012.08.052

• Sachs, J. D., Schmidt-Traub, G., Kroll, C., Lafortune, G., & Fuller, G. (2021). The decade of action for the Sustainable Development Goals: Sustainable Development Report 2021. Cambridge University Press. https://doi.org/10.1017/9781009106559

• Schumpeter, J. A. (1942). Capitalism, socialism, and democracy. Harper & Brothers.

• Stern, N. (2007). The economics of climate change: The Stern Review. Cambridge University Press. https://doi.org/10.1017/CBO9780511817434

• Taylor & Francis Online. (2020). Income inequality, innovation, and carbon emission. Taylor & Francis Online. Retrieved from https://www.tandfonline.com/doi/full/10.1080/1331677X.2020.1734855

• Taylor & Francis Online. (2024). The interrelationships between carbon dioxide emissions and technological innovation. Taylor & Francis Online. Retrieved from https://www.tandfonline.com/doi/full/10.1080/14786451.2024.2352515

• World Bank. (2020). World development indicators 2020. World Bank Group. Retrieved from https://databank.worldbank.org/source/world-development-indicators

```
In [ ]:
```