

Stabilize CO2 Emissions with Renewable Energy

Logan McDavid

2025-07-28

Project Overview

In this project, I explored historical CO2 emissions and renewable energy worldwide using the OWID datasets for energy and CO2. I aimed to answer the question:

How much do renewables need to grow to stabilize CO2 emissions over the next 20 years?

1. Data Cleaning and Preparation

Datasets:

- `owid-energy-data.csv`
 - Data on Energy by Our World in Data
- `owid-co2-data.csv`
 - Data on CO2 and Greenhouse Gas Emissions by Our World in Data

Libraries included in preparation and exploratory analysis:

```
library(readr) # Reading in .csv files
library(dplyr) # For filtering and left join
library(ggplot2) # Plotting in exploratory analysis
library(tidyr) # For reshaping data: pivot_longer()
```

Filter both OWID datasets to contain data from 2003 - 2023.

```
energy_master_data <- read.csv("owid-energy-data.csv") |>
  filter(year >= 2003)

co2_master_data <- read.csv("owid-co2-data.csv") |>
  filter(year >= 2003)
```

In the cleaning process, I grabbed unique countries from both datasets and also divided variables into countries, regions, organizations, and groupings. In my research, I will focus primarily on country-level data. Code here is not included as it is very repetitive.

Merge both datasets into master dataset over `country` and `year`.

```
# Master merged dataset  
merged_master_data <- left_join(co2_master_data, energy_master_data, by = c("country", "year"))
```

2. Exploratory Data Analysis

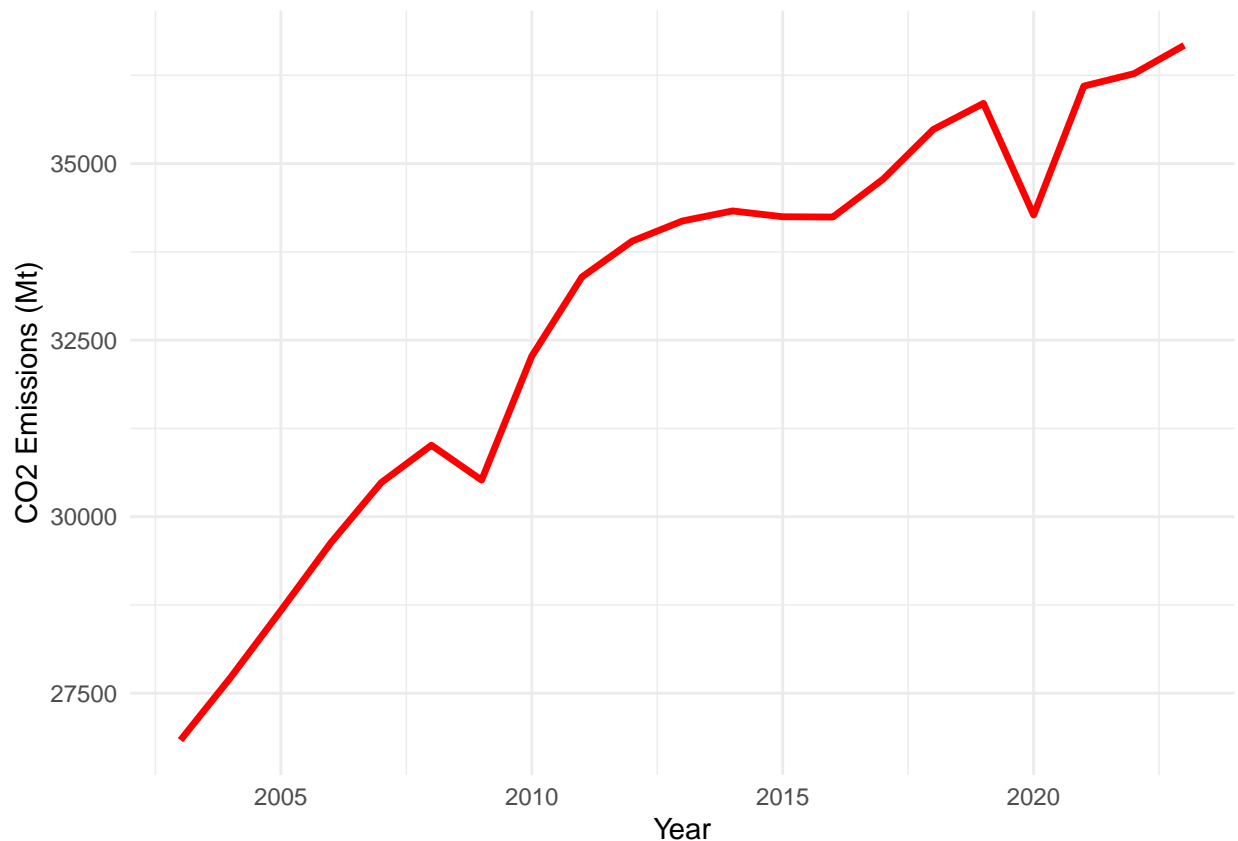
In order to understand the raw data I am working in, here are 4 different visualization to help represent CO2 emissions and renewable energy consumption.

1. Global CO2 Emissions Overtime

To begin my analysis, I simply calculated the **total global CO2 emissions per year** from 2003 - 2023 using country-level CO2 data. This dataset allows us to observe global trends in datasets over time. By visualizing these trends, I aimed to better understand the overall trajectory of global emissions and establish a baseline for evaluating the potential impact of renewable energy adoption in future scenarios.

```
# Grab total emissions per year in dataset
global_co2 <- co2_countries |>
  group_by(year) |>
  summarize(total_co2 = sum(co2, na.rm = TRUE))

# Plot emissions per year
ggplot(global_co2, aes(x = year, y = total_co2)) +
  geom_line(color = "red", size = 1.2) +
  labs(
    y = "CO2 Emissions (Mt)",
    x = "Year"
  ) +
  theme_minimal()
```



```
# Basic summary of plot
summary(global_co2)
```

```
##      year      total_co2
## Min.   :2003   Min.     :26836
## 1st Qu.:2008   1st Qu.:30518
## Median :2013   Median   :34187
## Mean   :2013   Mean     :32899
## 3rd Qu.:2018   3rd Qu.:34779
## Max.   :2023   Max.     :36674
```

According to the summary statistics:

- The **minimum** recorded global CO2 emissions during this period was **26,836 Mt**, while the maximum reached* **36,674 Mt**, suggesting an overall increase of nearly **10,000 Mt** over two decades.
- The **mean** emission value was **32,899 Mt**, with the median at **34,187 Mt**, indicating that recent years are generally above the average, consistent with a rising trend.

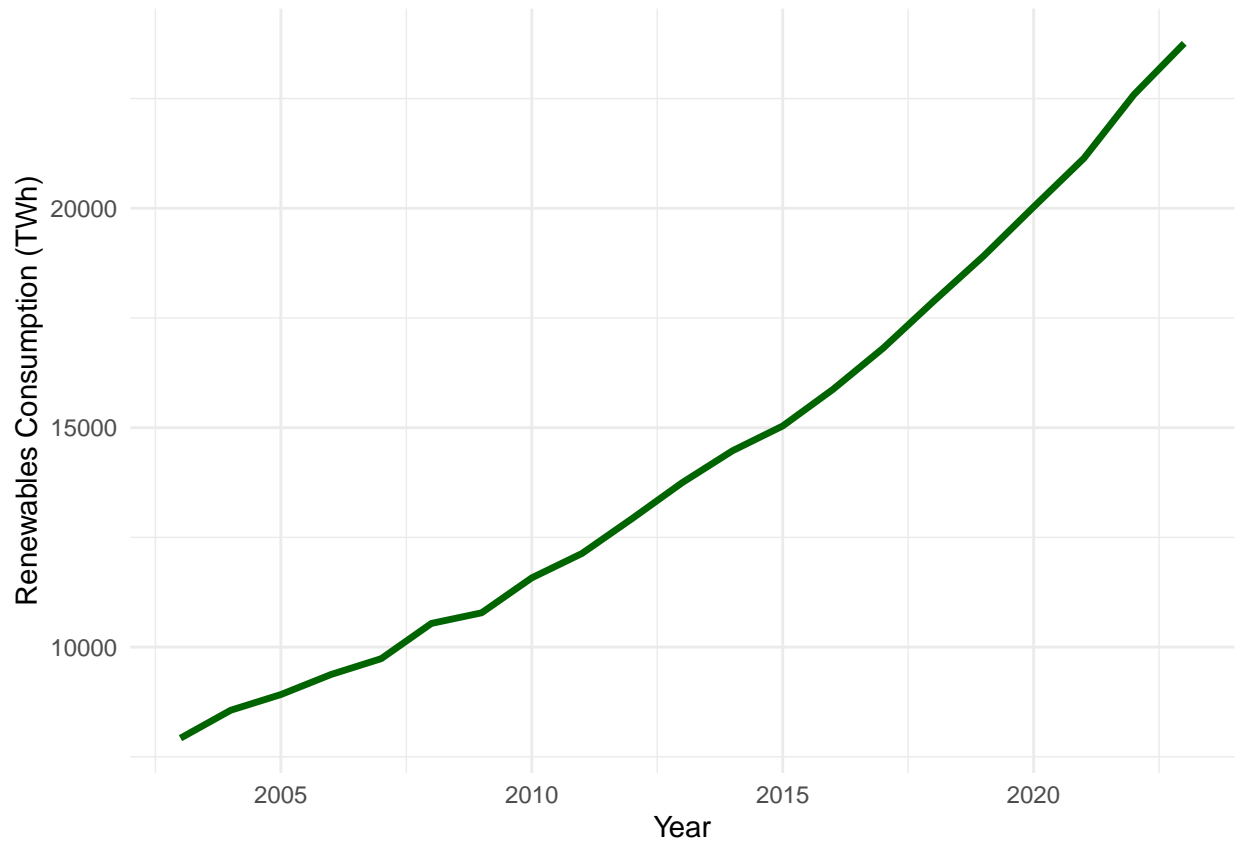
This trend reinforces the urgency of our research question: *How much do renewables need to grow to stabilize CO2 emissions over the next 20 years?*

2. Global Renewable Consumption Overtime

Next, I examined **global renewable energy consumption** from *2003 to 2023*. By aggregating the total renewable energy consumption across all countries, I aimed to identify trends in the global adoption of renewables. This visualization will help us understand whether the increase in renewable energy has kept pace with the rising CO2 emissions observed earlier — and whether current growth trajectories are sufficient to make a meaningful impact on emissions stabilization.

```
# Grab total renewables per year in dataset
global_renewables <- energy_countries |>
  group_by(year) |>
  summarize(total_renewables = sum(renewables_consumption, na.rm = TRUE))

# Plot renewables per year
ggplot(global_renewables, aes(x = year, y = total_renewables)) +
  geom_line(color = "darkgreen", size = 1.2) +
  labs(
    y = "Renewables Consumption (TWh)",
    x = "Year"
  ) +
  theme_minimal()
```



```
# Basic summary of plot
summary(global_renewables)
```

```
##      year      total_renewables
##  Min.   :2003    Min.      : 7924
##  1st Qu.:2008    1st Qu.   :10540
##  Median :2013    Median    :13748
##  Mean   :2013    Mean      :14415
##  3rd Qu.:2018    3rd Qu.   :17873
##  Max.   :2023    Max.      :23758
```

The plot illustrates an upward, exponential trend in global renewable energy consumption from 2003 to 2023. According to the summary statistics:

- In **2003**, total renewable consumption was around **7,924 TWh**, and by **2023**, it had increased to **23,758 TWh** — a **threefold** increase over two decades.
- There was never a year where renewable energy consumption **decreased**, setting the exponential trend to be promising toward renewables.

This rapid growth is promising; however, when compared with the CO₂ emissions summary — which also showed a steady increase — it suggests that renewables are growing **significantly but not yet fast enough to cause a downturn in emissions**.

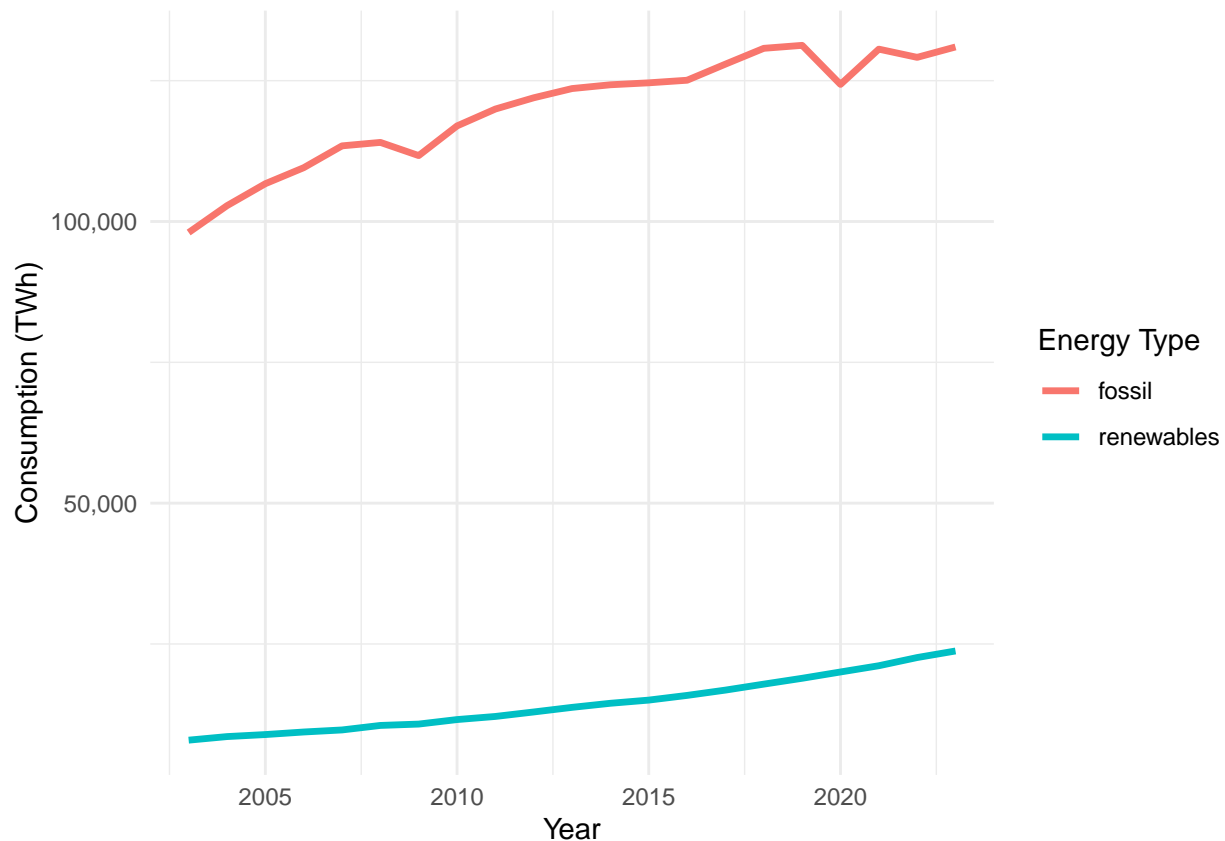
These findings reinforce the need for policy and infrastructure that can **amplify the growth rate of renewables even further**, ensuring that their expansion not only supplements global energy demand but also begins to **displace fossil fuel consumption at scale**.

3. Relationship Between Fossil Fuel Consumption and Renewable Consumption

This plot explores the relationship between **global fossil fuel and renewable energy consumption over time**. The dataset is grouped by year and energy type, summing total fossil fuel and renewable consumption worldwide. The data is reshaped into a long format to enable a direct visual comparison between the two sources of energy.

```
# Group fossil and renewable consumption over years
global_fossil <- energy_countries |>
  group_by(year) |>
  summarize(
    fossil = sum(fossil_fuel_consumption, na.rm = TRUE),
    renewables = sum(renewables_consumption, na.rm = TRUE)
  ) |>
  pivot_longer(cols = c("fossil", "renewables"), names_to = "Type", values_to = "Consumption")

# Plot energy types consumptions over years
ggplot(global_fossil, aes(x = year, y = Consumption, color = Type)) +
  geom_line(size = 1.2) +
  scale_y_continuous(labels = function(x) format(x, big.mark = ",", scientific = FALSE)) +
  labs(
    y = "Consumption (TWh)",
    x = "Year",
    color = "Energy Type"
  ) +
  theme_minimal()
```



The resulting line plot shows that fossil fuel consumption has remained consistently higher than renewable energy consumption throughout the 2003–2023 period. While fossil fuels dominate in total consumption, the gap appears to narrow slightly in more recent years, suggesting a global shift toward renewable energy.

Still, renewables remain a much smaller portion of the total energy mix, indicating the world continues to rely heavily on fossil fuels.

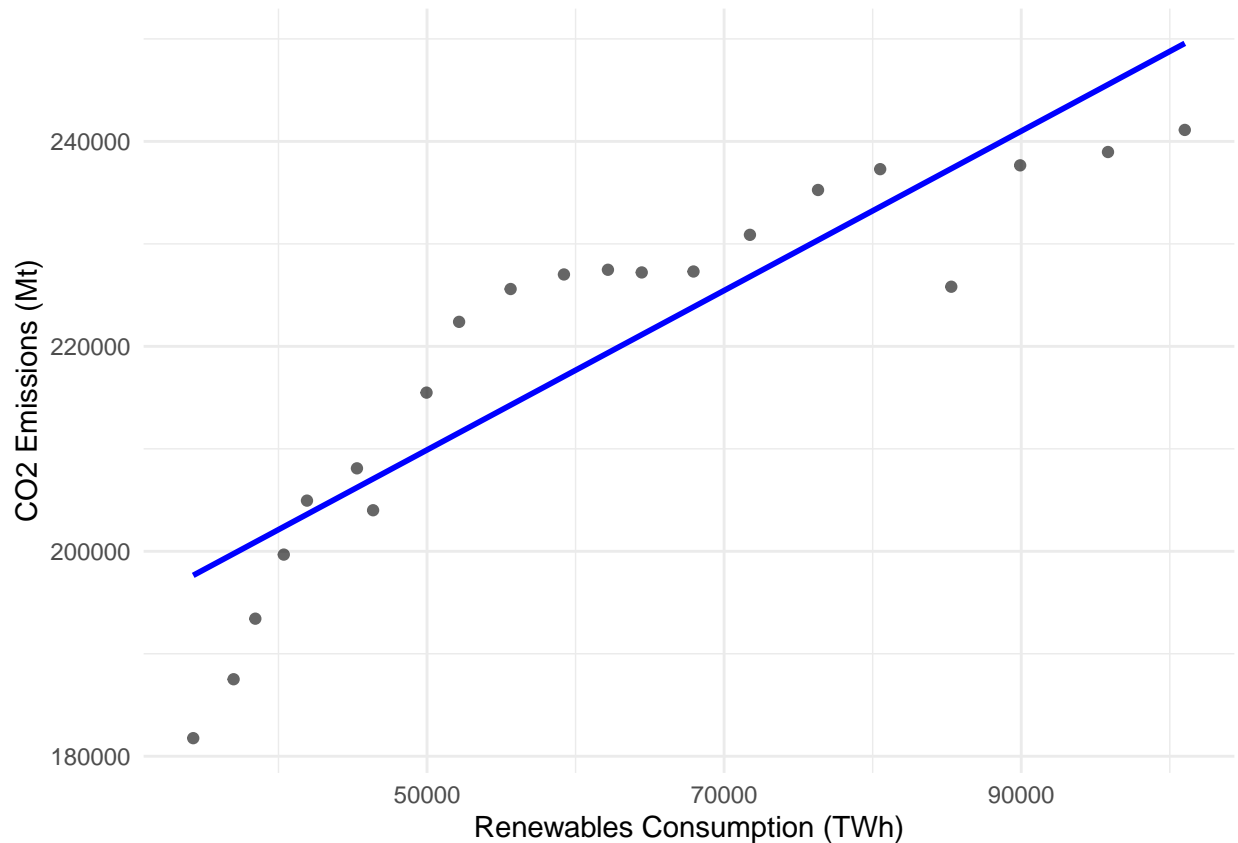
4. CO2 vs. Renewable Consumption

This plot examines the relationship between **global CO2 emissions and renewable energy consumption from 2003 to 2023**. Each *point* represents a *single year*, plotting **total renewable energy consumption** (in TWh) on the **x-axis** and **total CO2 emissions** (in Mt) on the **y-axis**. A **linear regression line** is included to highlight the **overall trend** between the two variables.

```
# Group co2 and renewables over years
merged_world <- merged_master_data |>
  group_by(year) |>
  summarize(
    co2 = sum(co2, na.rm = TRUE),
    renewables = sum(renewables_consumption, na.rm = TRUE)
  )

# Plot linear model
ggplot(merged_world, aes(x = renewables, y = co2)) +
  geom_point(alpha = 0.6) +
  geom_smooth(method = "lm", se = FALSE, color = "blue") +
  labs(
    x = "Renewables Consumption (TWh)",
    y = "CO2 Emissions (Mt)"
  ) +
  theme_minimal()
```

```
## 'geom_smooth()' using formula = 'y ~ x'
```



```
# Print a basic summary
summary(merged_world)
```

```
##      year      co2      renewables
##  Min.   :2003   Min.   :181765   Min.    : 34262
## 1st Qu.:2008   1st Qu.:204943   1st Qu.: 45284
## Median :2013   Median :225816   Median : 59218
## Mean   :2013   Mean   :218994   Mean    : 61704
## 3rd Qu.:2018   3rd Qu.:230882   3rd Qu.: 76321
## Max.   :2023   Max.   :241117   Max.    :101011
```

The scatter plot reveals a **slight positive linear relationship**, suggesting that as renewable consumption has **increased**, CO2 emissions have also **increased**. This may seem counterintuitive, but it likely reflects the fact that *global energy demand has been rising overall*, with **renewable energy supplementing** — **rather than replacing** — **fossil fuels** in many cases. As a result, even with growing renewables, CO2 emissions have continued to rise.

The summary statistics show that **renewable energy consumption** ranged from **34,262 TWh** to over **101,000 TWh**, with a **median** of around **59,218 TWh**. **CO2 emissions** ranged from **181,765 Mt** to **241,117 Mt**, with a **median** of **225,816 Mt**. These figures reinforce the notion that both renewables and emissions have trended upward over the last two decades.

This analysis highlights the complex challenge of transitioning to low-carbon energy: increasing renewable energy alone may not be sufficient to reduce emissions unless it leads to a substantial displacement of fossil fuel use.

3. Methods

In my study, I sought out to use two different modeling approaches. These models worked with data from the **United States of America** as well as the **World**. After testing many different models, I found that these two models were the most telling:

1. Counterfactual Model

To estimate the potential impact of renewable energy on CO2 emissions, a **counterfactual linear regression model** was developed. This model simulated what CO2 emissions would have looked like had renewable energy consumption remained constant at its 2003 level.

A multiple linear regression was fitted using U.S. and World data from a cleaned and merged dataset. The response variable was total **CO2 emissions co2**, and the predictor variables included:

- **Fossil fuel consumption:** fossil_fuel_consumption
- **Population:** population.x
- **Renewable energy consumption:** renewables_consumption

GDP was excluded in the final U.S. model due to multicollinearity concerns or limited marginal predictive value in this context.

After fitting the model:

1. A **copy of the U.S. dataset** was created.
2. The **renewable energy consumption** variable was fixed at its **2003 value** across all years.
3. Using the trained linear model, **counterfactual CO2 emissions** were predicted under the assumption that renewable energy consumption did not grow after 2003.

The difference between the actual and counterfactual CO2 emissions reflects an **estimate of emissions avoided** due to renewable energy growth. A plot was generated to compare actual emissions with the simulated counterfactual scenario across time.

2. Multiple Linear Regression Model

To project future U.S. CO2 emissions under varying renewable energy growth scenarios, a **multiple linear regression model** was developed using historical data from the cleaned and merged dataset.

The model was trained to predict annual **CO2 emissions co2** based on the following key predictors:

- **Fossil fuel consumption:** fossil_fuel_consumption
- **Population:** population.x
- **Renewable energy consumption:** renewables_consumption

Although GDP was initially included, it was excluded in the final model due to multicollinearity concerns and limited marginal explanatory power in the U.S.-specific context.

To simulate future emissions, the following steps were taken:

1. Trend extrapolation for predictors

- Linear trend models were fitted for each of the three predictor variables using historical U.S. data. These models projected values for **fossil fuel consumption**, **population**, and **renewables consumption** from **2024 to 2043**.

2. Simulating renewable energy growth

- A custom function was implemented to simulate various **renewable energy growth rates**, ranging from **0% to 10% annually**. For each growth rate, renewable energy consumption was projected exponentially, while fossil fuel use and population followed their respective linear trends.

3. CO2 predictions

- The fitted multiple linear regression model `usa_lm` was used to predict **future CO2 emissions** under each renewable growth scenario. These predictions allowed for the comparison of CO2 trajectories depending on how rapidly renewable energy is adopted in the future.

4. Output and visualization

- Predicted emissions across all growth scenarios were plotted over time (2024–2043), with the **2023 CO2 level** shown as a dashed baseline for reference. A summary table reported projected emissions in the final forecast year (2043) for each growth rate scenario.

This forecasting model enabled an exploration of how different renewable energy adoption paths may influence future CO2 emissions in the U.S., highlighting the potential long-term climate benefits of accelerated renewable energy investment.

4. Tools

In order to run these models, you will need to install these packages:

1. Base R

No explicit loading is required for base R functions, which powered the core of both modeling approaches:

- `lm()` — used to fit linear regression models for both the counterfactual and forecasting models.
- `predict()` — used to generate predicted CO2 values based on the fitted models.
- `data.frame()` — used extensively to build and manipulate datasets.
- `tail()` — used to extract the final observed CO2 value.
- `seq()`, `paste0()`, and other control functions — used for simulation setup and labeling.

2. ggplot2

- Used to create all visualizations, including:
 - Actual vs. counterfactual CO2 trends.
 - Forecasted CO2 emissions under varying renewable growth rates.
 - Custom legends, color coding, themes, and plot labels.
- Key functions: `ggplot()`, `geom_line()`, `geom_hline()`, `labs()`, and `theme_minimal()`.

3. dplyr

- Functions like `filter()`, `select()`, `arrange()`, and the pipe operator `|>` are part of the `dplyr` package. These were used for:
 - Creating summary tables.
 - Filtering by year.
 - Sorting predicted values for final comparison.

5. Results

I began by revisiting our central question: **How much do renewables need to grow to stabilize CO2 emissions over the next 20 years?** To investigate this, I employed two key methods—**Counterfactual Simulation** and **Multiple Linear Regression Modeling**—to analyze both American and global data. For each method, I'll present results and findings specific to the United States and then for the world as a whole.

1. Counterfactual Simulation

The counterfactual simulation allows us to model hypothetical scenarios of renewable energy growth and estimate the corresponding effects on CO2 emissions. This method helps quantify how different rates of renewable expansion could alter emission trajectories.

I. United States of America

```
# This estimates CO2 as a function of known drivers, including renewables
model <- lm(co2 ~ fossil_fuel_consumption + gdp.x + population.x + renewables_consumption, data = merged_master_data)

# Subset dataset for United States
usa_model <- merged_master_data[merged_master_data$country == "United States", ]

# Fit linear model predicting CO2 from major drivers
usa_lm <- lm(co2 ~ fossil_fuel_consumption + population.x + renewables_consumption, data = usa_model)

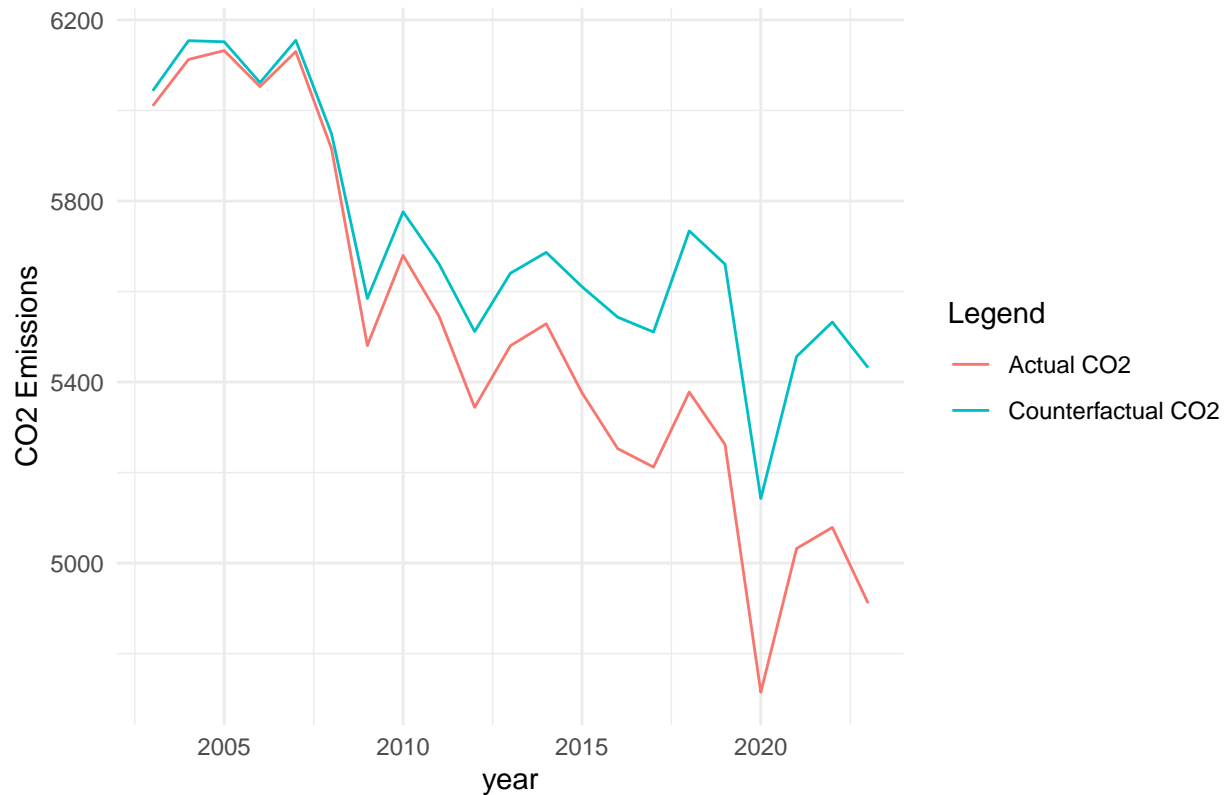
# Copy dataset and fix renewables to 2003 level
counterfactual_data <- usa_model
renewable_2003 <- counterfactual_data$renewables_consumption[counterfactual_data$year == 2003]
counterfactual_data$renewables_consumption <- renewable_2003

# Predict counterfactual CO2 emissions
counterfactual_data$counterfactual_co2 <- predict(usa_lm, newdata = counterfactual_data)

# Prepare comparison dataframe
comparison_df <- data.frame(
  year = usa_model$year,
  actual_co2 = usa_model$co2,
  counterfactual_co2 = counterfactual_data$counterfactual_co2
)

# Plot actual vs counterfactual
library(ggplot2)
ggplot(comparison_df, aes(x = year)) +
  geom_line(aes(y = actual_co2, color = "Actual CO2")) +
  geom_line(aes(y = counterfactual_co2, color = "Counterfactual CO2")) +
  labs(title = "Counterfactual Simulation: U.S. CO2 Emissions",
       y = "CO2 Emissions",
       color = "Legend") +
  theme_minimal()
```

Counterfactual Simulation: U.S. CO2 Emissions



```
# Total CO2 emissions avoided
sum(comparison_df$counterfactual_co2 - comparison_df$actual_co2)
```

```
## [1] 4365.408
```

The simulation shows that if the U.S. had not increased renewable energy consumption since 2003, CO2 emissions would have been substantially higher over the study period. The gap between actual and counterfactual CO2 emissions widens over time, indicating meaningful emissions avoided due to renewable growth. This suggests that renewable expansion in the U.S. has had a tangible mitigating effect on CO2 emissions.

The total CO2 emissions avoided in the United States is estimated to be **~4365 Mt**. This is equivalent to eliminating nearly **9 years'** worth of current U.S. aviation emissions, or removing all passenger vehicles in the U.S. for **over 2 years**. While renewables have not yet reversed overall emissions trends, this counterfactual underscores their growing and measurable impact in reducing what emissions could have been.

II. World

```
# Subset dataset for the world
world_model <- merged_master_data[merged_master_data$country == "World", ]

# Fit linear model predicting CO2 from major drivers
world_lm <- lm(co2 ~ fossil_fuel_consumption + population.x + renewables_consumption, data = world_model)

# Copy dataset and fix renewables to 2003 level
counterfactual_data <- world_model
renewable_2003 <- counterfactual_data$renewables_consumption[counterfactual_data$year == 2003]
```

```

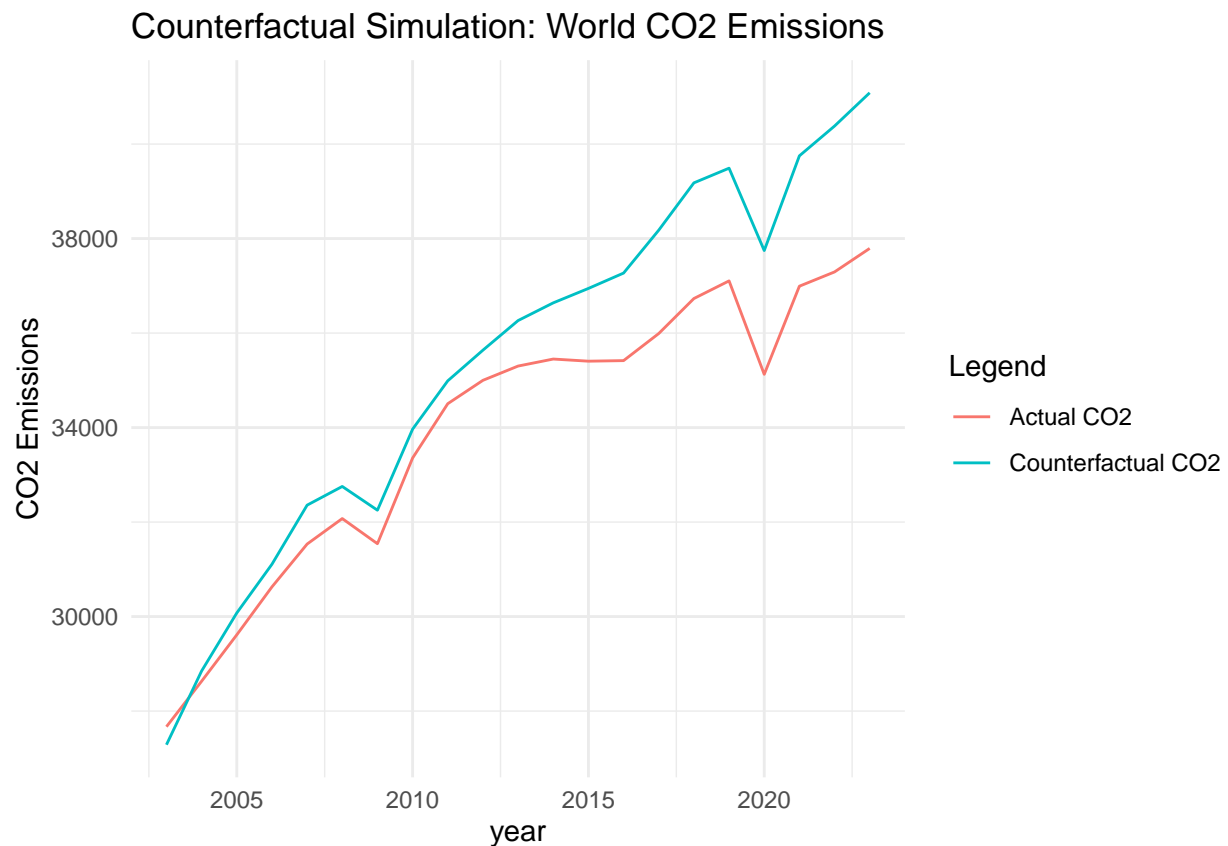
counterfactual_data$renewables_consumption <- renewable_2003

# Predict counterfactual CO2 emissions
counterfactual_data$counterfactual_co2 <- predict(world_lm, newdata = counterfactual_data)

# Prepare comparison dataframe
comparison_df <- data.frame(
  year = world_model$year,
  actual_co2 = world_model$co2,
  counterfactual_co2 = counterfactual_data$counterfactual_co2
)

# Plot actual vs. counterfactual CO2
library(ggplot2)
ggplot(comparison_df, aes(x = year)) +
  geom_line(aes(y = actual_co2, color = "Actual CO2")) +
  geom_line(aes(y = counterfactual_co2, color = "Counterfactual CO2")) +
  labs(title = "Counterfactual Simulation: World CO2 Emissions",
       y = "CO2 Emissions",
       color = "Legend") +
  theme_minimal()

```



```

# Total CO2 emissions avoided
sum(comparison_df$counterfactual_co2 - comparison_df$actual_co2)

```

```
## [1] 29046.4
```

Similarly, at the global level, maintaining renewable energy consumption at 2003 levels would have led to significantly higher CO2 emissions. The divergence between actual and counterfactual emissions grows steadily, illustrating the critical role of renewable energy expansion worldwide in offsetting emissions growth. However, the scale of emissions avoided, while notable, also highlights that renewable growth alone has not yet reversed the upward trend in global emissions.

The total CO2 emissions avoided in the World is estimated to be **~29046.4 Mt**. To put this into perspective, this is approximately **three times** the annual CO2 emissions of the European Union, or nearly **one year's worth** of global road transport emissions. This finding reinforces that while renewables have not yet decoupled emissions from economic and energy growth, their expansion has already significantly mitigated what could have been a much steeper emissions curve.

Final Summary of Conuterfactual Simulation:

The counterfactual simulations for both the U.S. and the world underscore the importance of renewable energy growth in reducing CO2 emissions compared to a no-growth baseline. However, the continued rise in actual emissions suggests that renewable energy expansion must accelerate further to stabilize and ultimately reduce emissions over the coming decades.

2. Multiple Linear Regression Model

Building off of the counterfactual simulation, I asked myself: **How much do we ultimately need to accelerate renewables to stabilize and ultimately reduce emissions over the coming decades?**

To better understand the relationship between carbon emissions and potential driving factors, I used a multiple linear regression model. This model helps quantify how much variables like fossil fuel consumption and renewable energy use contribute to CO2 emissions. By analyzing the model for both the United States and the world, we can gain insight into different regional trends and the influence of energy consumption on emissions.

I. United States of America

```
# Forecast frame: 2024-2043
future_years <- data.frame(year = 2024:2043)

# Fit linear trend models for predictors
fossil_model <- lm(fossil_fuel_consumption ~ year, data = usa_model)
population_model <- lm(population.x ~ year, data = usa_model)
renewables_model <- lm(renewables_consumption ~ year, data = usa_model)

# Generate predicted values for each predictor
future_years$fossil_fuel_consumption <- predict(fossil_model, newdata = future_years)
future_years$population.x <- predict(population_model, newdata = future_years)
future_years$renewables_consumption <- predict(renewables_model, newdata = future_years)

# Store baseline values
baseline_fossil <- future_years$fossil_fuel_consumption
baseline_population <- future_years$population.x

# Last observed CO2 level (2023)
last_actual_co2 <- tail(usa_model$co2, 1)

# FUNCTION: simulate_renewables_growth
# PURPOSE: To simulate future CO2 under varying renewable growth rates
```

```

simulate_renewables_growth <- function(growth_rate) {
  renewables_start <- future_years$renewables_consumption[1]
  n_years <- nrow(future_years)

  # Project renewables consumption forward at given growth rate
  renewables_projection <- renewables_start * (1 + growth_rate)^(0:(n_years - 1))

  # Create input data frame for CO2 prediction
  prediction_input <- data.frame(
    year = future_years$year,
    fossil_fuel_consumption = baseline_fossil,
    population.x = baseline_population,
    renewables_consumption = renewables_projection
  )

  # Predict CO2 emissions using regression model
  predicted_co2 <- predict(usa_lm, newdata = prediction_input)

  # Return results
  data.frame(
    year = future_years$year,
    renewables = renewables_projection,
    predicted_co2 = predicted_co2
  )
}

# Run simulations across a range of renewable growth rates (0% to 10%)
growth_rates <- seq(0, 0.10, by = 0.01)
results_list <- lapply(growth_rates, simulate_renewables_growth)
names(results_list) <- paste0("Growth_", growth_rates * 100, "%")

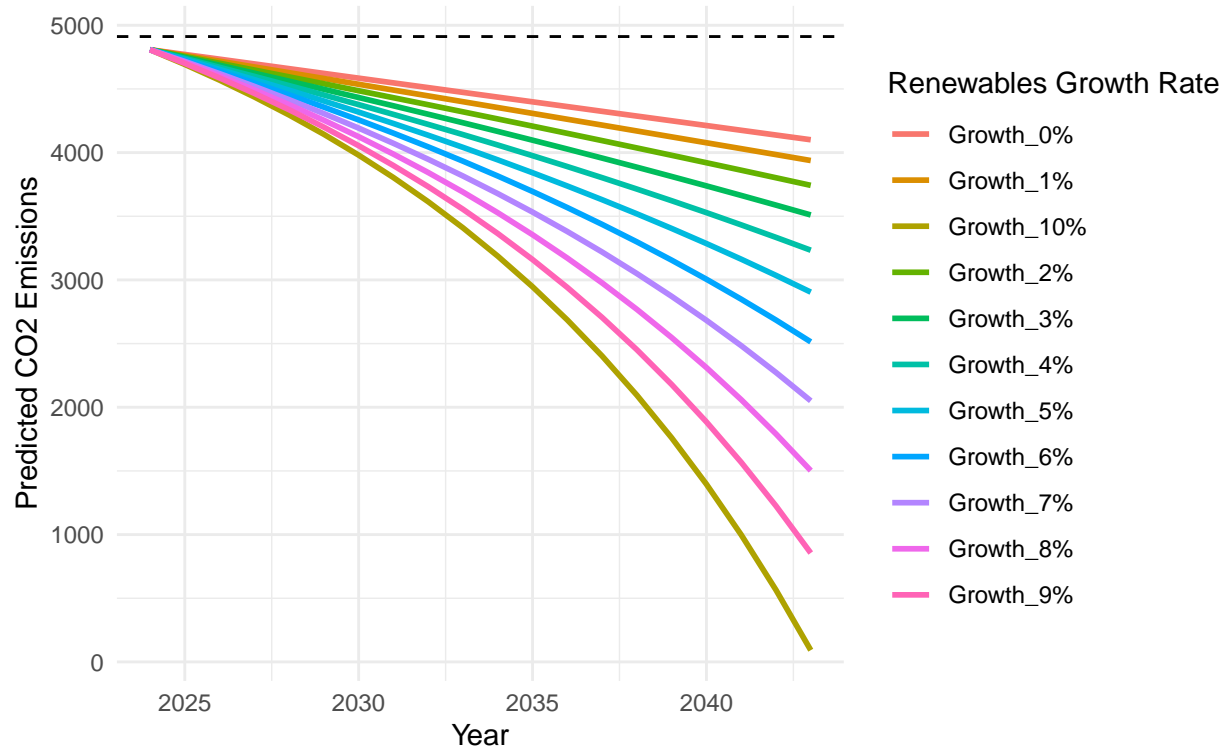
# Combine simulation results into one data frame
all_results <- bind_rows(
  lapply(seq_along(results_list), function(i) {
    df <- results_list[[i]]
    df$GrowthRate <- names(results_list)[i]
    df
  })
)

# # Plot CO2 emission trajectories
ggplot(all_results, aes(x = year, y = predicted_co2, color = GrowthRate)) +
  geom_line(size = 1) +
  geom_hline(yintercept = last_actual_co2, linetype = "dashed", color = "black") +
  labs(
    title = "U.S. CO2 Emissions under Different Renewables Growth Rates (0-10%)",
    subtitle = "Dashed line = last observed CO2 emissions",
    y = "Predicted CO2 Emissions",
    x = "Year",
    color = "Renewables Growth Rate"
  ) +
  theme_minimal()

```


U.S. CO2 Emissions under Different Renewables Growth Rates (0–10%)

Dashed line = last observed CO2 emissions



```
# Final-year emissions for each scenario
summary_table <- all_results |>
  filter(year == max(year)) |>
  select(GrowthRate, predicted_co2) |>
  arrange(predicted_co2)

# Print summary table
print(summary_table)
```

##		GrowthRate	predicted_co2
## 20...1		Growth_10%	91.91582
## 20...2		Growth_9%	855.34190
## 20...3		Growth_8%	1502.56897
## 20...4		Growth_7%	2050.44530
## 20...5		Growth_6%	2513.49985
## 20...6		Growth_5%	2904.24426
## 20...7		Growth_4%	3233.43780
## 20...8		Growth_3%	3510.31947
## 20...9		Growth_2%	3742.81125
## 20...10		Growth_1%	3937.69572
## 20...11		Growth_0%	4100.77107

The U.S. model suggests that CO2 emissions are highly sensitive to renewable energy growth. In the absence of additional growth (0%), emissions are projected to rise from approximately **3,943 Mt in 2023** to **4,101 Mt** by 2043. However, under an aggressive 10% annual growth in renewable energy, emissions could decline

significantly to just ~92 Mt, representing a **near-complete decarbonization** of the energy system by the 2040s.

Most of the emission reductions are nonlinear, with increasingly steep declines occurring at higher growth rates. For example:

- A **5% annual growth** reduces emissions to ~2,904 Mt.
- A **7% growth rate** brings emissions down to ~2,050 Mt.
- A **10% growth rate** drops emissions to under 100 Mt.

These results highlight that **incremental increases in renewable growth yield disproportionately large reductions in emissions**, especially above the 6–7% annual range.

II. World

```
# Forecast frame: 2024-2043
future_years_world <- data.frame(year = 2024:2043)

# Fit linear trend models for predictors
fossil_model_world <- lm(fossil_fuel_consumption ~ year, data = world_model)
population_model_world <- lm(population.x ~ year, data = world_model)
renewables_model_world <- lm(renewables_consumption ~ year, data = world_model)

# Predict global trends
future_years_world$fossil_fuel_consumption <- predict(fossil_model_world, newdata = future_years_world)
future_years_world$population.x <- predict(population_model_world, newdata = future_years_world)
future_years_world$renewables_consumption <- predict(renewables_model_world, newdata = future_years_world)

# Baseline predictors for simulation
baseline_fossil_world <- future_years_world$fossil_fuel_consumption
baseline_population_world <- future_years_world$population.x

# Last observed CO2 level (2023)
last_actual_co2_world <- tail(world_model$co2, 1)

# FUNCTION: simulate_renewables_growth_world
# PURPOSE: To simulate future CO2 under varying renewable growth rates for the world
simulate_renewables_growth_world <- function(growth_rate) {
  renewables_start <- future_years_world$renewables_consumption[1]
  n_years <- nrow(future_years_world)

  # Project renewables forward at specified growth rate
  renewables_projection <- renewables_start * (1 + growth_rate)^(0:(n_years - 1))

  # Assemble prediction input
  prediction_input <- data.frame(
    year = future_years_world$year,
    fossil_fuel_consumption = baseline_fossil_world,
    population.x = baseline_population_world,
    renewables_consumption = renewables_projection
  )

  # Predict emissions
```

```

predicted_co2 <- predict(world_lm, newdata = prediction_input)

# Return results
data.frame(
  year = future_years_world$year,
  renewables = renewables_projection,
  predicted_co2 = predicted_co2
)
}

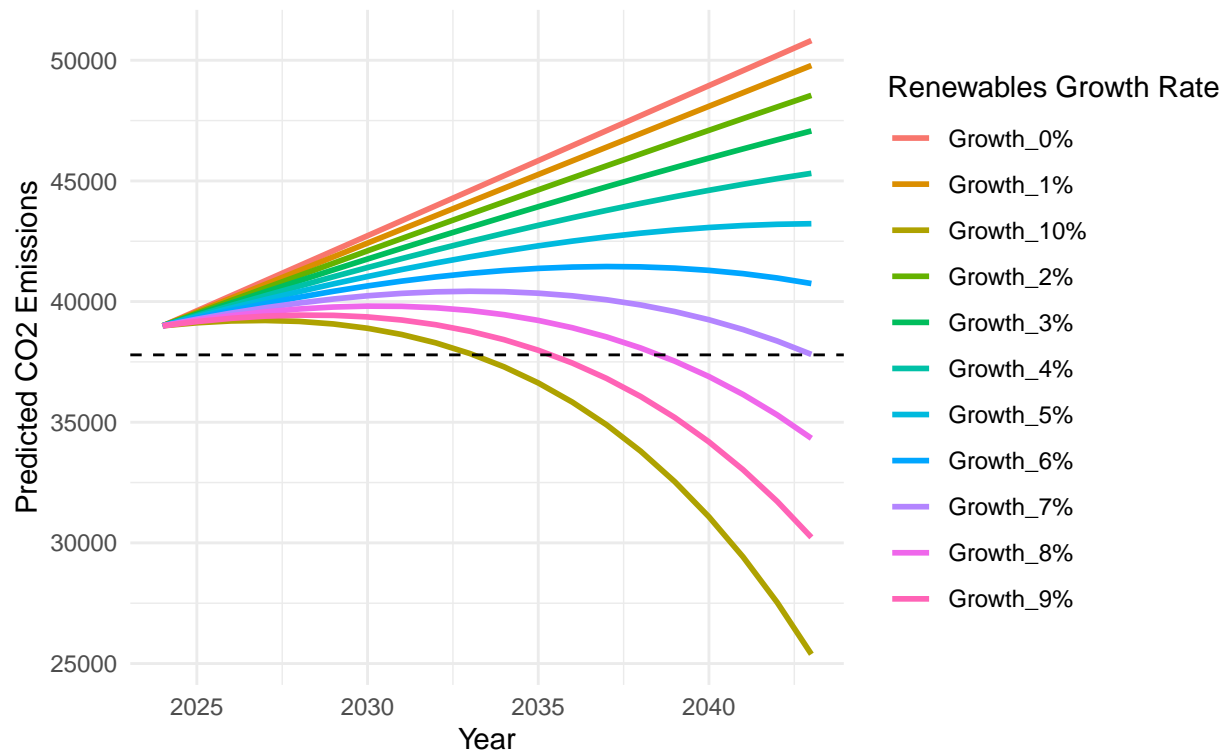
# Run simulations globally
growth_rates <- seq(0, 0.10, by = 0.01)
results_list_world <- lapply(growth_rates, simulate_renewables_growth_world)
names(results_list_world) <- paste0("Growth_", growth_rates * 100, "%")

# Combine results
all_results_world <- bind_rows(
  lapply(seq_along(results_list_world), function(i) {
    df <- results_list_world[[i]]
    df$GrowthRate <- names(results_list_world)[i]
    df
  })
)

# Plot global results
ggplot(all_results_world, aes(x = year, y = predicted_co2, color = GrowthRate)) +
  geom_line(size = 1) +
  geom_hline(yintercept = last_actual_co2_world, linetype = "dashed", color = "black") +
  labs(
    title = "Global CO2 Emissions Under Different Renewables Growth Rates (0-10%)",
    subtitle = "Dashed line = last observed CO2 emissions",
    y = "Predicted CO2 Emissions",
    x = "Year",
    color = "Renewables Growth Rate"
  ) +
  theme_minimal()

```

Global CO2 Emissions Under Different Renewables Growth Rates (0–10%
Dashed line = last observed CO2 emissions



```
# Final-year emissions for each scenario
summary_table_world <- all_results_world |>
  filter(year == max(year)) |>
  select(GrowthRate, predicted_co2) |>
  arrange(predicted_co2)

# Print summary table
print(summary_table_world)
```

##		GrowthRate	predicted_co2
## 20...1		Growth_10%	25386.38
## 20...2		Growth_9%	30229.25
## 20...3		Growth_8%	34335.01
## 20...4		Growth_7%	37810.52
## 20...5		Growth_6%	40747.96
## 20...6		Growth_5%	43226.69
## 20...7		Growth_4%	45314.97
## 20...8		Growth_3%	47071.40
## 20...9		Growth_2%	48546.23
## 20...10		Growth_1%	49782.51
## 20...11		Growth_0%	50816.99

Globally, the model tells a similar story—though on a much larger scale and with slower relative declines. With no acceleration in renewables, global CO2 emissions are projected to rise from around **49,782 Mt** (2023) to **50,817 Mt** by 2043.

At higher growth rates:

- A **5% increase** brings emissions down to **~43,227 Mt.**
- A **7% growth** leads to **~37,811 Mt.**
- A **10% increase** results in a dramatic fall to **~25,386 Mt.**, or nearly **50% below baseline projections.**

While the effect is less dramatic than in the U.S., the global model demonstrates that **substantial emissions reductions are possible only through steep and sustained increases in renewables adoption.**

Final Summary of Multiple Linear Regression model

The multiple linear regression model offers a quantitative window into the role renewable energy growth can play in emissions reduction. For both the United States and the world, a consistent pattern emerges: **only ambitious growth rates in renewables—typically 7–10% annually—are sufficient to offset fossil fuel trends and population growth to produce real reductions in CO2 emissions.**

The U.S. shows greater responsiveness due to a smaller energy footprint and more mature infrastructure, while the global system requires more aggressive and coordinated action. Still, the direction is clear: **rapid acceleration of renewable energy is not just beneficial, it is necessary** if emissions are to be stabilized and then meaningfully reduced over the next two decades.

6. Summary & Conclusions

In my project, I wanted to **examine the relationship between renewable energy consumption and CO2 emissions in the U.S. and globally, using two different modeling approaches**, asking myself:

How much do renewables need to grow to stabilize CO2 emissions over the next 20 years?

My Multiple Linear Regression Model gives us the most telling results, acting as an advisement as to how quickly can we realistically achieve this goal.

To summarize, renergy reduces CO2 emissions significantly both in the U.S. and globally. However, the effect is stronger and more statistically significant in America., based on the regression results. We can further validate this by looking at the R-squared values of the following linear models:

```
summary(usa_lm)$r.squared
```

```
## [1] 0.9948748
```

```
summary(world_lm)$r.squared
```

```
## [1] 0.9933304
```

When interpreted:

- The United States has **~99.5%** of the variation in CO2 emissions explained by **fossil fuel consumption, population, and renewable energy use**. This is extremely high predictive accuracy.
- The World has **~99.3%** of global CO2 emissions variance explained by the same predictors. Also excellent, but slightly less explanatory power than the U.S. model.

The data that the model is portraying also clearly describes that there is a *strong negative correlation* between renewables and CO2, as increasing renewables significantly reduces CO2 emissions.

From a practical standpoint, if the United States wants to leave a strong environmental footprint for the world, changes need to be made toward environmental practices. Some ideas for this could be:

- Continue or expand subsidies/tax incentives for renewable energy.
- Invest in infrastructure (like wind/solar grids, battery storage).
- Discourage fossil fuels (carbon taxes, stricter regulations).
- Establish deadlines that are not just 20 years out, but showing progress throughout the years of shifting toward renewable energy consumption for companies.

Now for the global community:

- Focus on emerging economies—help them skip fossil-heavy development paths.
- Promote international renewable energy partnerships (e.g., tech sharing).
- Align global climate goals (e.g., via COP agreements) with data-based emission strategies.

I believe that this data shows *real potential—now its about commitment*. If investments in renewables continue rising while fossil fuel dependence falls, our models suggest a real path to reducing carbon emissions world-wide, further destructing our planet. But this path depends on political will, equitable global strategies, and sustained innovation. Thank you.

7. Sources

Github: OWID

- <https://github.com/owid/energy-data>
- <https://github.com/owid/co2-data>

Github: SpencerPao / Data_Science

- https://github.com/SpencerPao/Data_Science/tree/main

Youtube - Counterfactual Predictions for Decision-Making

- https://www.youtube.com/watch?v=Ubt_sH2qRMg

Youtube Playlist: Stanford CS229 Machine Learning Course

- <https://www.youtube.com/playlist?list=PLoROMvodv4rMiGQp3WXShMGgzqpfVfbU>

Youtube - Multiple regression for beginners

- https://www.youtube.com/watch?v=Ew__uyYjow1w

ChatGPT

- <https://chatgpt.com/>

Youtube - Making your R Markdown Pretty

- <https://www.youtube.com/watch?v=v048ru5v0BE>