# Analysis of Carbon Emission Trends

Saathvik Suresh and Rahul Choudhary

## Introduction

1. Throughout our lifetimes, carbon emissions have been a growing concern for our planet. These emissions contributes to climate change and environmental degradation. In this project, we try to answer these three questions:

    1. ***Which countries have contributed the most to global emissions?***
    2. ***What's the relationship between country GDP and country carbon emissions?***
    3. ***What are the contribution of different sectors towards CO2 emissions?***

2. Our project starts with showcasing the increase in global emissions over the past decades, identifying the top five countries with the highest emissions. We then see the relationship between the economic growth and emission by comparing GDP and carbon emissions. In conclusion, we see the emissions made by different sectors, using bar graphs and line charts to see the top contributing industries to emission and looking at trends over time.

## Background Information

- Carbon emissions, are in the form of carbon dioxide or CO2, which is one of the major reason for climate change. These emissions are produced by activities such as burning fossil fuels, transportation, construction, and agriculture. Over the last 50 years, global carbon emissions have risen.

    1. ***Which countries have contributed the most to global emissions?***
    We will use the carbon emission data to identify the top five countries with the highest emissions. A line graph will help us see trends, and a table will show a summary. For this question we will be using an emissions per capita dataset that tracks many countries over many years.

2. ***What are the contribution of different sectors towards CO2 emissions?***
   Different sectors have different levels of carbon emissions. We will visualize a emissions by sector bar graph to see which sectors are responsible for the most emission. For this question we will be using an emissions by sector dataset to see which countries have the most CO2 output over the years.

3. ***What's the relationship between country GDP and country carbon emissions?***
   Economic development often leads to increase in industrial activity and energy use. We will use a GDP dataset and track it alongside the emissions by sector dataset to see if there is a correlation between the 2.

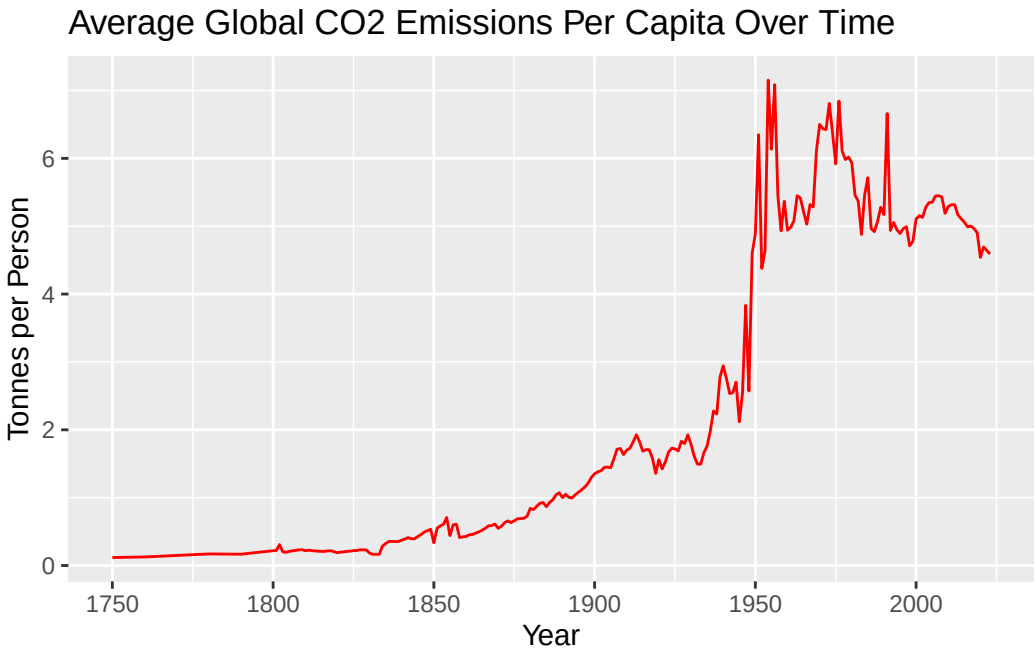
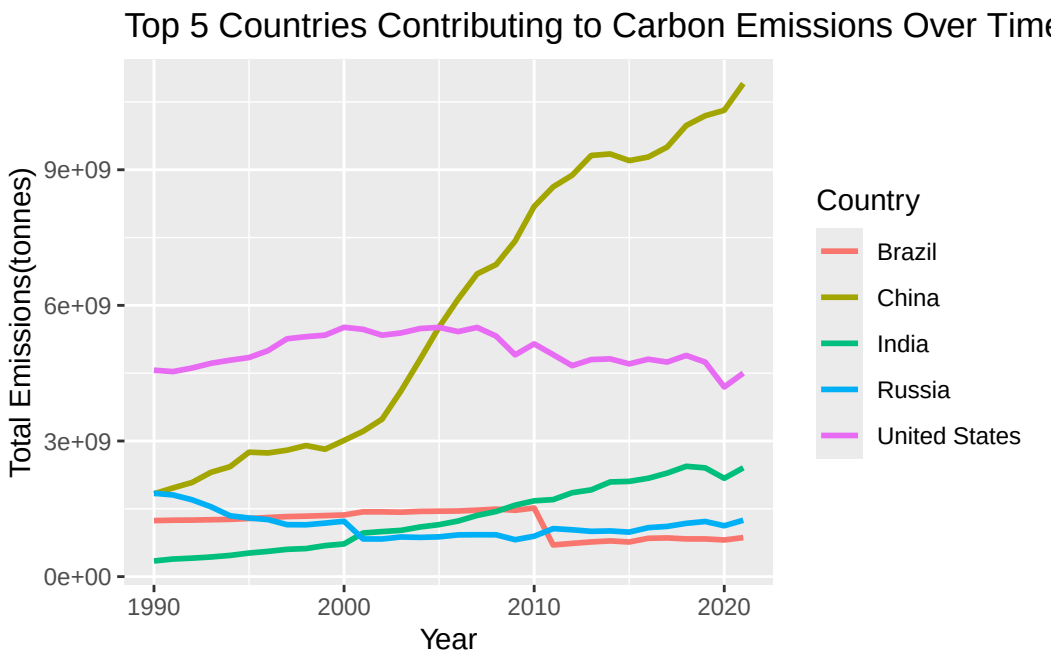



Table 1: Average Total Emissions by Country

| Country | Average Emissions |
|---|---|
| China | 5927070017 |
| United States | 4993443129 |
| India | 1307694061 |
| Brazil | 1175641249 |
| Russia | 1135610309 |

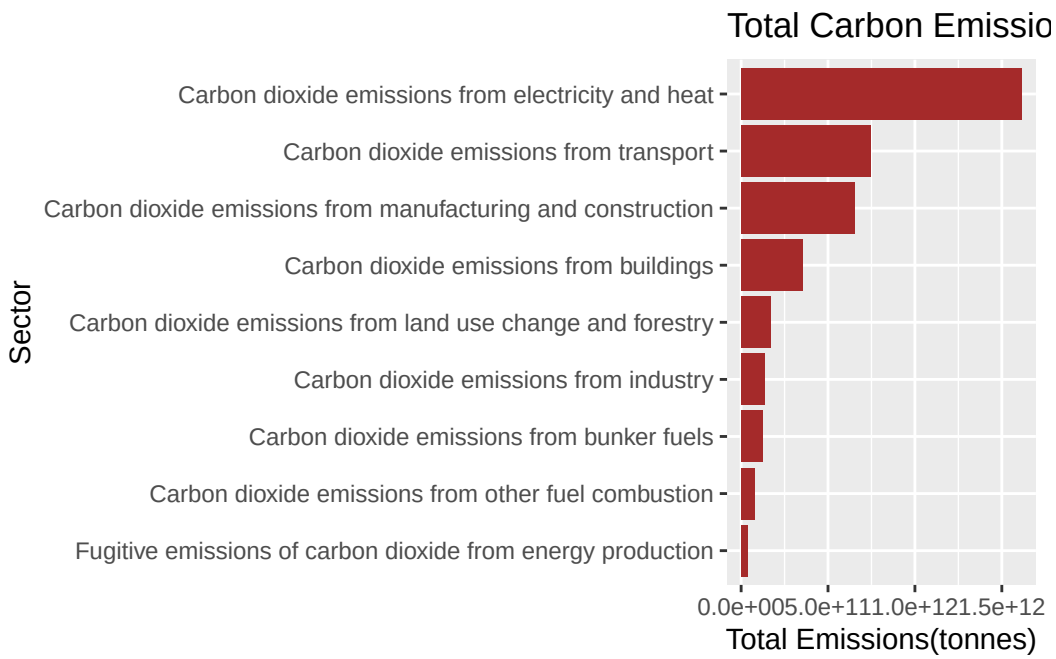
## Top 5 countries

- Here we've established that the top 5 countries in CO2 Emissions are China, United States, Brazil, India, and Russia. Next we will look at the trend of CO2 Emissions over the years in each of those countries to get a better understanding of their respective policies on the matter.



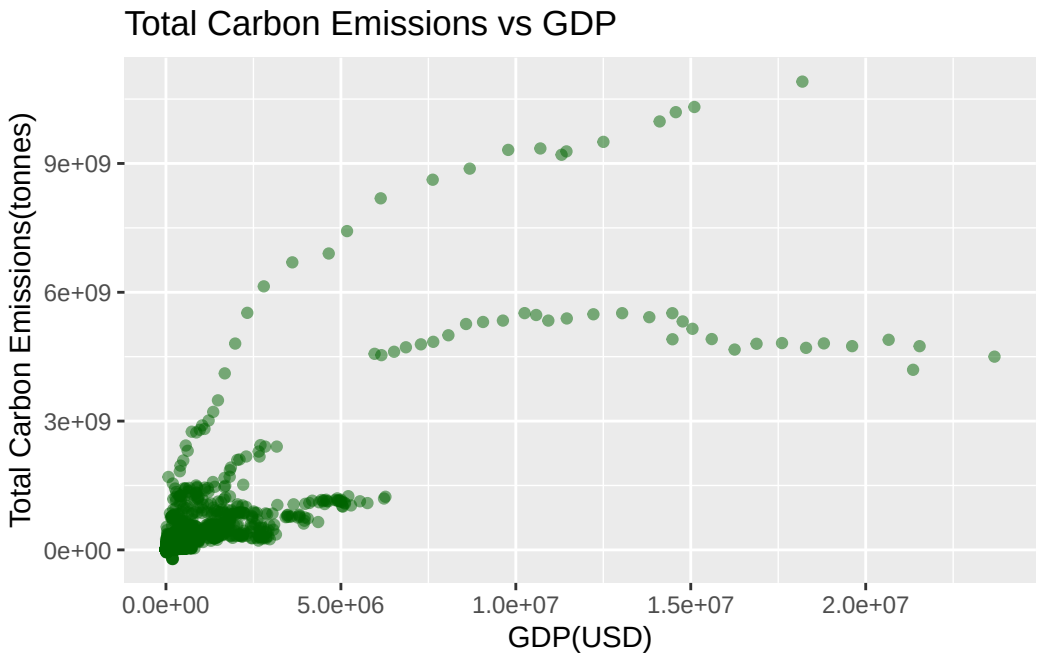

## What does this mean

- We can clearly see that between countries, there are various trends of how the idea Carbon Emissions are being treated. Countries like China that are increasingly unconcerned with how their industrialization affects the rest of the world show a clear steep upward slope of Carbon Emissions. Whereas countries like the US and Russia have grown more environmentally conscious over the years showing their slight downward slope.




## Which Sector is Performing Worst

- We can see that the largest contributor to emissions is consumer heat and electricity. This makes sense as everyone in the world needs electricity and heat. However, something interesting to note is that CO2 from transport is almost half of heat and electricity. Furthermore, manufacturing and construction accounts for nearly the same amount. So there is nearly the same amount of CO2 being spent on transport and manufacturing as there is for the heat and electricity of the 8 billion people who reside on Earth. Also worth noting is that the actual energy being produced doesn't contribute very much to the total amount of emissions which is a common misconception.




## Relationship between GDP and Carbon Emissions

- Intuitively, we can reason that there is a positive relationship between Carbon Emissions and GDP. After all, its reasonable that countries that produce more energy and goods, are also producing more emissions through the use of fossil fuels, electricity, and large machinery. The above scatter plot measuring the relationship between Carbon Emissions and GDP backs this up. As we can see there is a clear positive slope in the plot.
- However, another important detail to realize is that there are countries on the scatter plot that have very high Carbon Emissions with respect to their GDP. What this means is that there are countries that are underproducing, but still contributing a significant amount of emissions. Through this graph we can see that and furthermore, through further analysis we would be able to deduce which countries are responisble for that part of the graph.

## Conclusion

Looking at all the data together, it's clear that carbon emissions have been rising steadily over time, especially in the last few decades. When we looked at the average emissions per person around the world, the trend was almost always going up, showing how widespread and long term the issue has become.

We found that the top five countries contributing to CO2 emissions are China, the United States, India, Brazil, and Russia. China's emissions are still climbing fast, likely due to rapid industrial growth, while countries like the US and Russia have leveled out a bit, probably because of policies and cleaner technology.

When we broke emissions down by sector, electricity and heat were the biggest contributors, which makes sense because almost everyone relies on those. But it was interesting to see that transport and manufacturing weren't that far behind, which shows how much emissions come from how we move and build things. Also, energy production itself wasn't as big of a contributor as many people might assume.

Finally, the GDP vs. emissions graph showed that richer countries tend to emit more, which makes sense. But there were some outliers, countries that emit a lot even though they don't have very high GDPs. That's a sign that not all emissions are just a result of being rich or developed, some countries are just less efficient or more polluting than others.

Overall, the project gave a solid picture of who's contributing to emissions, how it connects to the economy, and which sectors are driving the most damage. It's clear there's no one size fits all fix, but understanding the data helps point to where change is most needed.

## Data Summary

In this project, we have use three datasets:

1. **CO  and Greenhouse Gas Emissions**
   Source: Our World in Data
   Purpose: This dataset was used to get insights on the countries which contributed most towards the global emissions.
   Cases: Each case is a country in a particular year and their annual CO2 emissions per capita.
   Attributes: Country, Years

2. **GDP by Country**
   Source: Wikipedia
   Purpose: This dataset was used to get insights on the relationship between country's GDP v/s country's emissions.
   Cases: Each case is a country and the columns are GDP in particular years.
   Attributes: Country, Years

3. **CO2 Emissions by Sector**
   Source: Our World in Data
   Purpose: This dataset was used to get insights on the sectors which contributed the most towards global emissions.
   Cases: Each case is a country in a particular year and emissions done by various

sectors(1990-2021).
Attributes: Country, Year, Sectors CO2 emissions

### References


- Our World in Data – CO  and Greenhouse Gas Emissions
  Ritchie, H., Roser, M., & Rosado, P. (2020). CO  and greenhouse gas emissions. Our World in Data. https://ourworldindata.org/co2-and-greenhouse-gas-emissions
- Wikipedia contributors. (2025, June 29). List of countries by past and projected GDP (nominal). In *Wikipedia, The Free Encyclopedia*. Retrieved 04:52, July 1, 2025, from https://en.wikipedia.org/w/index.php?title=List_of_countries_by_past_and_projected_GDP_(nominal)&oldid=1297935771
- Our World in Data – CO  Emissions by Sector (Grapher Tool)
  Our World in Data. (n.d.). CO  emissions by sector. https://ourworldindata.org/grapher/co-emissions-by-sector?tab=table


ountry, names_to = “Year”, values_to = “Emissions” ) %>% mutate(Year = as.integer(Year))

## Calculate the global average CO2 emissions per capita for each year

avg_val <- pc_long %>% group_by(Year) %>% summarise(Average = mean(Emissions, na.rm = TRUE))

## Create a line plot of average emissions over time

ggplot(avg_val, aes(x = Year, y = Average)) + geom_line(color = “red”) + labs( title = “Average Global CO2 Emissions Per Capita Over Time”, x = “Year”, y = “Tonnes per Person” )

## Code Appendix

```
# Emissions per capita over time graph code
#_____________________________________
library(readr)
library(dplyr)
library(tidyr)
```

```r
library(ggplot2)

# Load the per capita emissions dataset
per_capita <- read_csv("C:/Users/Saathvik/Desktop/STAT184Project/datasets/per_capita_emissio

# Convert the dataset from wide format to long format
# Each row will now represent a country and year with corresponding emissions
pc_long <- per_capita %>%
  pivot_longer(
    cols = -Country,
    names_to = "Year",
    values_to = "Emissions"
  ) %>%
  mutate(Year = as.integer(Year))

# Calculate the global average CO2 emissions per capita for each year
avg_val <- pc_long %>%
  group_by(Year) %>%
  summarise(Average = mean(Emissions, na.rm = TRUE))

# Create a line plot of average emissions over time
ggplot(avg_val, aes(x = Year, y = Average)) +
  geom_line(color = "red") +
  labs(
    title = "Average Global CO2 Emissions Per Capita Over Time",
    x = "Year",
    y = "Tonnes per Person"
  )

# Top 5 Countries Emission Table
#_______________________________
library(readr)
library(dplyr)
library(knitr)
# Load the emissions by sector dataset
emissions <- read_csv("C:/Users/Saathvik/Desktop/STAT184Project/datasets/emission_by_sector.

# Identify the relevant columns in the dataset
emission_cols <- setdiff(names(emissions), c("Entity", "Code", "Year"))

# Sum across all sector columns
emissions <- emissions %>%
```

```r
  mutate(Total_Emissions = rowSums(select(., all_of(emission_cols)), na.rm = TRUE))

# Compute the average total emissions for each country
country_avg <- emissions %>%
  group_by(Entity) %>%
  summarise(Average_Emissions = mean(Total_Emissions, na.rm = TRUE)) %>%
  arrange(desc(Average_Emissions))

# Identify the target countries which we found when examining the dataset excluding the grou
target_countries <- c("United States", "China", "India", "Brazil", "Russia")

# Rename columns for display and filter for the target countries
filtered_avg <- country_avg %>%
  filter(Entity %in% target_countries) %>%
  rename(Country = Entity, `Average Emissions` = Average_Emissions)

kable(filtered_avg, caption = "Average Total Emissions by Country")

# Emissions Plot of Top 5 Countries code
#-----------------------------------
library(readr)
library(dplyr)
library(ggplot2)

emissions <- read_csv("C:/Users/Saathvik/Desktop/STAT184Project/datasets/emission_by_sector.
                      show_col_types = FALSE)

emission_cols <- setdiff(names(emissions), c("Entity", "Code", "Year"))

# Convert all emission values to numeric
emissions <- emissions %>%
  mutate(across(all_of(emission_cols), ~ readr::parse_number(as.character(.))))

# Calculate total emissions for each row
emissions <- emissions %>%
  rowwise() %>%
  mutate(Total_Emissions = sum(c_across(all_of(emission_cols)), na.rm = TRUE)) %>%
  ungroup()

target_countries <- c("United States", "China", "India", "Brazil", "Russia")
filtered_data <- emissions %>%
  filter(Entity %in% target_countries)
```

```r
# Create a line plot showing total emissions over time for each of the top 5 countries
ggplot(filtered_data, aes(x = Year, y = Total_Emissions, color = Entity)) +
  geom_line(size = 1) +
  labs(
    title = "Top 5 Countries Contributing to Carbon Emissions Over Time ",
    x = "Year",
    y = "Total Emissions(tonnes)",
    color = "Country"
  )

# Highest Contributing Sectors to Emission Bar Graph code
#_____________________________________________________
library(readr)
library(dplyr)
library(ggplot2)

emissions <- read_csv("C:/Users/Saathvik/Desktop/STAT184Project/datasets/emission_by_sector.c


emission_cols <- setdiff(names(emissions), c("Entity", "Code", "Year"))

emissions <- emissions %>%
  mutate(across(all_of(emission_cols), ~ readr::parse_number(as.character(.))))

sector_totals <- emissions %>%
  summarise(across(all_of(emission_cols), ~ sum(., na.rm = TRUE))) %>%
  pivot_longer(cols = everything(), names_to = "Sector", values_to = "Total_Emissions")

# Create a horizontal bar plot showing total emissions by sector
ggplot(sector_totals, aes(x = reorder(Sector, Total_Emissions), y = Total_Emissions)) +
  geom_bar(stat = "identity", fill = "brown") +
  coord_flip() +
  labs(
    title = "Total Carbon Emissions by Sector",
    x = "Sector",
    y = "Total Emissions(tonnes)"
  )

# GDP Relationship to Emissions Scatter Plot code
#_____________________________________________
library(readr)
library(dplyr)
```

```r
library(ggplot2)
library(tidyr)

gdp <- read_csv("C:/Users/Saathvik/Desktop/STAT184Project/datasets/GDP_estimates.csv")
emissions <- read_csv("C:/Users/Saathvik/Desktop/STAT184Project/datasets/emission_by_sector.

# Change the GDP data from wide to long format
gdp_long <- gdp %>%
  pivot_longer(cols = -Country, names_to = "Year", values_to = "GDP") %>%
  mutate(
    Country = tolower(Country),
    Year = as.numeric(Year),
    GDP = as.numeric(gsub(",", "", GDP))
  )

emission_cols <- setdiff(names(emissions), c("Entity", "Code", "Year"))

# Format the dataset so that it can match up with the GDP dataset so we can use it
emissions_clean <- emissions %>%
  mutate(across(all_of(emission_cols), ~ readr::parse_number(as.character(.)))) %>%
  rowwise() %>%
  mutate(Total_Emissions = sum(c_across(all_of(emission_cols)), na.rm = TRUE)) %>%
  ungroup() %>%
  mutate(
    Country = tolower(Entity),
    Year = as.numeric(Year)
  )

# Merge GDP and emissions datasets by country and year
merged <- inner_join(gdp_long, emissions_clean, by = c("Country", "Year")) %>%
  drop_na(GDP, Total_Emissions)

# Create a scatter plot of GDP vs Total Emissions
ggplot(merged, aes(x = GDP, y = Total_Emissions)) +
  geom_point(alpha = 0.5, color = "darkgreen", size = 1.5) +
  labs(
    title = "Total Carbon Emissions vs GDP",
    x = "GDP(USD)",
    y = "Total Carbon Emissions(tonnes)"
  )

# Data Wrangling Code
```

```r
#____________________
#dependencies

library(mosaicData)
library(knitr)
library(rvest)
library(stringr)
library(tidyverse)

#Scraping the gdp table from web (General data wrangling)
page <- "https://en.wikipedia.org/wiki/List_of_countries_by_past_and_projected_GDP_(nominal)"
tableList <- page %>%
  read_html() %>%
  html_nodes(css = "table") %>%
  html_table(fill = TRUE)

#assigning a varible to all the diffrent gdp tables (General data wrangling)
gdp_table1 = tableList[[1]]
gdp_table2 = tableList[[2]]
gdp_table3 = tableList[[3]]
gdp_table4 = tableList[[4]]
gdp_table5 = tableList[[5]]
gdp_table6 = tableList[[6]]

#There was a * after the name of countries in the first table (Regular expression)
gdp_table1 <- gdp_table1 %>%
  mutate(`Country (or area)` = gsub("\\*", "", `Country (or area)`))

#renaming the column (General data wrangling)
joined_table5 <- joined_table5 %>%
  rename(Country = `Country (or dependent territory)`)

gdp_table1 <- gdp_table1 %>%
  rename(Country = `Country (or area)`)

#Joining the tables to make a big table with all the data (Join operation)
joined_table2 <- full_join(gdp_table2, gdp_table3, by = c("Country (or dependent territory)"
joined_table3 <- full_join(joined_table2, gdp_table4, by = c("Country (or dependent territor
joined_table4 <- full_join(joined_table3, gdp_table5, by = c("Country (or dependent territor
joined_table5 <- full_join(joined_table4, gdp_table6, by = c("Country (or dependent territor

#(General data wrangling)
```

```r
filtered_gdp_table2 <- gdp_table1 %>%
  filter(Country %in% joined_table5$Country)

#(Transformation function)
gdp_table1 <- gdp_table1 %>%
  mutate(Country = str_trim(tolower(Country)))

joined_table5 <- joined_table5 %>%
  mutate(Country = str_trim(tolower(Country)))

#(General data wrangling)
common_countries <- intersect(gdp_table1$Country, joined_table5$Country)
common_countries

joined_table5_not_matched <- joined_table5 %>%
  filter(!(Country %in% gdp_table1$Country))

gdp_table1_non_matched <- gdp_table1 %>%
  filter(!(Country %in% joined_table5$Country))

gdp_table1_non_matching
joined_table5_non_matching

gdp_table1[45, "Country"] <- "republic of the congo"
gdp_table1[126, "Country"] <- "federated states of micronesia"

last10 <- tail(names(final_table), 10)
final_table <- final_table %>%
  relocate(all_of(last10), .before = 2)

#(Transformation function)
final_table <- final_table %>%
  mutate(across(everything(), ~na_if(.x, "")))

#(Join operation)
final_table <- left_join(joined_table5, gdp_table1, by = "Country")

# (General data wrangling)
per_capita_emission <- read_csv("per_capita_emission.csv")
gdp_estimates <- read_csv("GDP_estimates.csv")
per_capita_emission <- per_capita_emission %>%
  rename(Country = `Entity`)
```

```r
#(Pivot wider)
per_capita_emission <- per_capita_emission %>%
  pivot_wider(
    names_from = Year,
    values_from = `Annual CO  emissions (per capita)`
  )

# (General data wrangling)
common_countries <- intersect(gdp_estimates$Country, per_capita_emission$Country)
common_countries

per_capita_emission <- per_capita_emission %>%
  mutate(Country = str_trim(tolower(Country)))

non_matched_countries <- per_capita_emission %>%
  filter(!(Country %in% gdp_estimates$Country))

per_capita_emission[131, "Country"] <- "federated states of micronesia"


per_capita_emission <- per_capita_emission[per_capita_emission$Country != "world", ]

non_matched_countries

#turn table into csv (General data wrangling)
write_csv(final_table, "GDP_estimates.csv")
write_csv(per_capita_emission, "per_capita_emission.csv")

#
```