Course Work 2

ADEKUNLE ADEYEMI

S2463934

2025-04-11

Table of Contents

BACKGROUND

The unprecedented scale of global CO2 emissions represents one of the most pressing challenges of our time, fundamentally reshaping our understanding of industrial development, corporate responsibility, and environmental sustainability. Recent research by Friedlingstein et al. (2023) demonstrates that global surface temperatures have risen about 1.1°C since the pre-industrial period, with most of this warming occurring in the past 40 years. The years 2015-2022 were recorded as the eight warmest on record, highlighting the accelerating pace of climate change.

The historical evolution of emissions reveals a striking pattern of exponential growth, particularly marked by the transformative period following World War II. According to Griffin (2017), global energy-related CO2 emissions reached historic highs, indicating the persistent challenge of decoupling economic growth from emissions. This growth has been characterized by significant regional disparities, with the Asia Pacific region emerging as the dominant contemporary contributor, reflecting the rapid industrialization and economic development in this region.

The relationship between corporate activity and emissions has become increasingly central to climate policy discussions. Minx et al. (2021) provide comprehensive evidence that just 20 companies are responsible for 35% of all energy-related carbon dioxide and methane emissions worldwide since 1965. This concentration of emissions among a small number of corporate entities highlights the critical importance of understanding both historical responsibility and current emission patterns in developing effective climate solutions.

Here We Load the Data

```
data <- read.csv('Emissions.csv')
glimpse(data)

## Rows: 12,551
## Columns: 8
## $ X                     <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 
14, …
## $ year                  <int> 1962, 1962, 1963, 1963, 1964, 1964, 1965, 
1965,…
## $ parent_entity         <chr> "Abu Dhabi National Oil Company", "Abu 
Dhabi Na…
## $ parent_type           <chr> "State-owned Entity", "State-owned Entity", 
"St…
## $ commodity             <chr> "Oil & NGL", "Natural Gas", "Oil & NGL", 
"Natur…
## $ production_value      <dbl> 0.91250, 1.84325, 1.82500, 4.42380, 
7.30000, 17…
## $ production_unit       <chr> "Million bbl/yr", "Bcf/yr", "Million 
bbl/yr", "…
## $ total_emissions_MtCO2e <dbl> 0.3638848, 0.1343552, 0.7277697, 0.3224525, 
2.9…
```

Summary of the Dataset

This analysis utilizes the Carbon Majors Database, which provides comprehensive information about global emissions, corporate ownership, and production relationships from 1854 to 2022.

The dataset encompasses several key components that allow for detailed analysis of historical emissions patterns and corporate responsibility. The temporal coverage spans over 168 years, providing an unprecedented view of emissions evolution from the early industrial period through modern times. The data includes detailed information about parent entities and their emissions across various commodity types, including different forms of coal, oil and natural gas liquids, natural gas, and cement production. Each entry contains production values and emission levels, allowing for analysis of emission intensity and efficiency across different sectors and time periods.

In addition, the dataset distinguishes between different types of corporate ownership - state-owned enterprises, investor-owned companies, and nation-states - providing insight into the institutional structure of global emissions.

The analysis focuses on several key research questions:

1. What is the geographical distribution of CO2 emissions across different regions of the world?
2. How have global emissions evolved from 1854 to 2022, and what major historical events significantly influenced these emission patterns? (This is crucial for understanding the historical context of emissions and identifying critical periods of change. It helps establish a baseline for evaluating current emission trends and future projections.)
3. What are the year-over-year changes in emissions across different commodities, and how do growth rates and patterns vary among different fuel types?
4. Who are the largest historical emitters from 1854 to 2022, and what does this reveal about corporate responsibility and industry concentration in emissions? (This helpsto address the critical issue of corporate responsibility in climate change and helps identify key stakeholders in emission reduction efforts.)
5. What role do the top 10 companies by production value play in global emissions, and how does their production impact overall emission levels? (This question helps quantify the impact of major energy producers)

Summary statistics which will show missing values per variable

```
library(kableExtra)

## Warning: package 'kableExtra' was built under R version 4.4.3
```

```
##
## Attaching package: 'kableExtra'

## The following object is masked from 'package:dplyr':
##
##     group_rows

# Create summary statistics
p <- summary(data)

# Create a nicely formatted table with additional styling
kable(p[1:6, ],
      caption = "Summary Statistics",
      format = "html",
      digits = 2,
      align = "c",
      booktabs = TRUE) %>%
  kable_styling(
    bootstrap_options = c("striped", "hover", "condensed"),
    full_width = FALSE,
    position = "center",
    font_size = 12
  ) %>%
  add_header_above(c(" " = 1, "Summary Statistics" = ncol(p))) %>%
  footnote(general = "Note: This table shows the summary statistics of the
dataset.")
```

Summary Statistics

	Summary Statistics						
X	year	parent_entity	parent_type	commodity	production_value	production_unit	total_emissions_MtCO2e
Min. : 1	Min. :1854	Length:12551	Length:12551	Length:12551	Min. : 0.004	Length:12551	Min. : 0.000
1st Qu.: 3138	1st Qu.:1973	Class :character	Class :character	Class :character	1st Qu.: 10.601	Class :character	1st Qu.: 8.785
Median : 6276	Median :1994	Mode :character	Mode :character	Mode :character	Median : 63.204	Mode :character	Median : 33.059
Mean : 6276	Mean :1987	NA	NA	NA	Mean : 412.712	NA	Mean : 113.220
3rd Qu.: 9414	3rd Qu.:2009	NA	NA	NA	3rd Qu.: 320.665	NA	3rd Qu.: 102.155
Max. :12551	Max. :2022	NA	NA	NA	Max. :27192.000	NA	Max. :8646.906

Note:

Note: This table shows the summary statistics of the dataset.

Figure 1: Summary Statistics of the dataset

From the summary results, there are no missing values in the dataset and comments made by the authors in (https://github.com/rfordatascience/tidytuesday/blob/main/data/2024/2024-05-21/readme.md) note that the dataset is clean.

METHODOLOGY

We employ a comprehensive exploratory data analysis approach to investigate and visualize global CO2 emissions patterns from 1854 to 2022. The relationships between historical emissions, corporate responsibility, and production dynamics are visualized using the Carbon Majors dataset as the primary data source.

Data Wrangling

Based on the original Dataset Columns, a new production_efficiency column is derived to enhance the level of analysis that can be done using this dataset

```
data <- data %>%
  mutate(production_efficiency = production_value / total_emissions_MtCO2e)

# View the first few rows to verify
head(data)

##   X year                  parent_entity        parent_type   commodity
## 1 1 1962 Abu Dhabi National Oil Company State-owned Entity   Oil & NGL
## 2 2 1962 Abu Dhabi National Oil Company State-owned Entity Natural Gas
## 3 3 1963 Abu Dhabi National Oil Company State-owned Entity   Oil & NGL
## 4 4 1963 Abu Dhabi National Oil Company State-owned Entity Natural Gas
## 5 5 1964 Abu Dhabi National Oil Company State-owned Entity   Oil & NGL
## 6 6 1964 Abu Dhabi National Oil Company State-owned Entity Natural Gas
##   production_value production_unit total_emissions_MtCO2e
production_efficiency
## 1          0.91250  Million bbl/yr              0.3638848
2.507662
## 2          1.84325          Bcf/yr              0.1343552
13.719230
## 3          1.82500  Million bbl/yr              0.7277697
2.507662
## 4          4.42380          Bcf/yr              0.3224525
13.719230
## 5          7.30000  Million bbl/yr              2.9110786
2.507662
## 6         17.32655          Bcf/yr              1.2629390
13.719230

# Save the updated dataset
write.csv(data, 'Emissions_with_efficiency.csv', row.names = FALSE)
```

	X <int>	year <int>	parent_entity <chr>	parent_type <chr>	commodity <chr>
1	1	1962	Abu Dhabi National Oil Company	State-owned Entity	Oil & NGL
2	2	1962	Abu Dhabi National Oil Company	State-owned Entity	Natural Gas
3	3	1963	Abu Dhabi National Oil Company	State-owned Entity	Oil & NGL
4	4	1963	Abu Dhabi National Oil Company	State-owned Entity	Natural Gas
5	5	1964	Abu Dhabi National Oil Company	State-owned Entity	Oil & NGL
6	6	1964	Abu Dhabi National Oil Company	State-owned Entity	Natural Gas

6 rows

production_value <dbl>	production_unit <chr>	total_emissions_MtCO2e <dbl>	production_efficiency <dbl>
0.91250	Million bbl/yr	0.3638848	2.507662
1.84325	Bcf/yr	0.1343552	13.719230
1.82500	Million bbl/yr	0.7277697	2.507662
4.42380	Bcf/yr	0.3224525	13.719230
7.30000	Million bbl/yr	2.9110786	2.507662
17.32655	Bcf/yr	1.2629390	13.719230

Figure 2: Derived Production Efficiency Column Added.

New Dataset Loaded with extra column

```
data <- read.csv('Emissions_with_efficiency.csv')
glimpse(data)
```

```
## Rows: 12,551
## Columns: 9
## $ X                     <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13,
14, …
## $ year                  <int> 1962, 1962, 1963, 1963, 1964, 1964, 1965,
1965,…
## $ parent_entity         <chr> "Abu Dhabi National Oil Company", "Abu
Dhabi Na…
## $ parent_type           <chr> "State-owned Entity", "State-owned Entity",
"St…
## $ commodity             <chr> "Oil & NGL", "Natural Gas", "Oil & NGL",
"Natur…
## $ production_value      <dbl> 0.91250, 1.84325, 1.82500, 4.42380,
7.30000, 17…
## $ production_unit       <chr> "Million bbl/yr", "Bcf/yr", "Million
bbl/yr", "…
## $ total_emissions_MtCO2e <dbl> 0.3638848, 0.1343552, 0.7277697, 0.3224525,
2.9…
## $ production_efficiency  <dbl> 2.507662, 13.719230, 2.507662, 13.719230,
2.507…
```

VARIABLES

The analysis focuses on several key variables crucial for understanding global emissions dynamics:

- Year: The calendar year of the record, representing when the emissions and production data were recorded (1854, 1900, 1950, 2000, 2022).

- Total_emissions_MtCO2e: Total greenhouse gas emissions measured in million tonnes of CO2 equivalent, converting all greenhouse gases into their equivalent CO2 impact (100.5, 500.2, 1000.8, 5000.3, 10000.7).
- Production_value: The numerical quantity of product produced, with units varying by commodity type (50.3, 200.7, 500.1, 1000.5, 5000.9).
- Production_unit: The specific unit of measurement for the production value, varying by commodity type ("Million tonnes/yr", "Million bbl/yr", "Bcf/yr", "Million Tonnes CO2").
- Parent_entity: The name of the company or organization responsible for the emissions and production ("ExxonMobil", "Saudi Aramco", "Gazprom", "BP", "Shell").
- Parent_type: The classification of the parent entity's ownership structure ("state-owned", "investor-owned", "nation-state").
- Commodity: The specific type of product or resource being produced ("Anthracite Coal", "Bituminous Coal", "Oil & NGL", "Natural Gas", "Cement").
- Production_efficiency: Production value per unit of emissions

Statistical Tools and Software

The analysis utilizes R programming language, leveraging several specialized packages:

• tidyverse - Core package for data manipulation and visualization (includes ggplot2 and dplyr)

• ggraph - Creates network and relationship visualizations

• scales - Formats axis labels and values (e.g., adding commas to large numbers)

• viridis - Provides color-blind friendly color palettes for visualizations

• zoo - Handles time series data and rolling calculations

• gridExtra - Arranges multiple plots in a grid layout

• patchwork - Combines multiple ggplot2 plots into complex layouts

• ggridges - Creates ridgeline plots for distribution visualization

• ggpubr - Enhances ggplot2 for publication-ready plots

• maps - Provides geographical map data

• countrycode - Converts country names and codes

• rworldmap - Creates world map visualizations

• R Markdown Components:

- knitr - Core document generation
- YAML Header - Document metadata and output settings

VARIABLE DISTRIBUTION

The study uses a combination of histograms, density plots, and violin plots to understand the underlying distribution of emissions and production values. This includes analysis both with and without outliers to ensure robust interpretation of the data.

HISTOGRAM

```
# Calculate summary statistics
summary_stats <- data %>%
  summarise(
    min_emissions = min(total_emissions_MtCO2e),
    max_emissions = max(total_emissions_MtCO2e),
```

```
    mean_emissions = mean(total_emissions_MtCO2e)
  )

# Create histogram with annotations
hist_emissions <- ggplot(data, aes(x = total_emissions_MtCO2e)) +
  geom_histogram(binwidth = 200,
                 fill = "darkgreen") +
  # Add vertical lines for min, max, and mean
  geom_vline(xintercept = summary_stats$min_emissions,
             color = "darkgreen",
             linetype = "dashed",
             size = 0.7) +
  geom_vline(xintercept = summary_stats$max_emissions,
             color = "red",
             linetype = "dashed",
             size = 0.7) +
  geom_vline(xintercept = summary_stats$mean_emissions,
             color = "blue",
             linetype = "dashed",
             size = 0.7) +
  # Add annotations
  annotate("text",
           x = summary_stats$min_emissions,
           y = max(ggplot_build(ggplot(data, aes(x = total_emissions_MtCO2e))
+
                                geom_histogram(binwidth =
200))$data[[1]]$count) * 0.9,
           label = paste("Min:", round(summary_stats$min_emissions, 1)),
           color = "yellow",
           hjust = -0.1,
           size = 3) +
  annotate("text",
           x = summary_stats$max_emissions,
           y = max(ggplot_build(ggplot(data, aes(x = total_emissions_MtCO2e))
+
                                geom_histogram(binwidth =
200))$data[[1]]$count) * 0.9,
           label = paste("Max:", round(summary_stats$max_emissions, 1)),
           color = "red",
           hjust = 1.1,
           size = 3) +
  annotate("text",
           x = summary_stats$mean_emissions,
           y = max(ggplot_build(ggplot(data, aes(x = total_emissions_MtCO2e))
+
                                geom_histogram(binwidth =
200))$data[[1]]$count) * 0.8,
           label = paste("Mean:", round(summary_stats$mean_emissions, 1)),
           color = "blue",
           hjust = -0.1,
```

```
              size = 3) +
  labs(title = "Distribution of Total Emissions",
       subtitle = "With minimum, maximum, and mean values",
       x = "Total Emissions (MtCO2e)",
       y = "Count") +
  theme_light()

## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## ℹ Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.

# Create production histogram
hist_production <- ggplot(data, aes(x = production_value)) +
  geom_histogram(binwidth = 300,
                 fill = "yellow",

                 size = 0.000001,
                 alpha = 0.7) +
  facet_wrap(~commodity + production_unit, scales = "free_x") +
  labs(title = "Distribution of Production by Commodity",
       x = "Production Value",
       y = "Count") +
  theme_minimal()
hist_emissions

hist_production
```

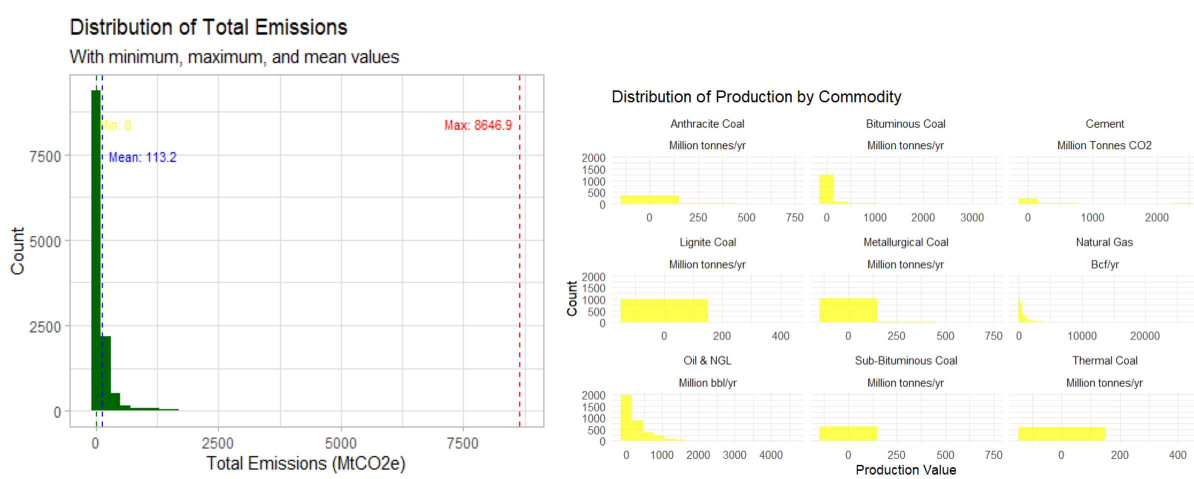


Figure 3: Histogram distribution of emission and production variables with outliers

At first glance, the distribution is heavily skewed to the right, with the vast majority of emissions falling in the lowest bin, this kind of distribution is common in environmental data, where many sources might contribute small amounts, while a few sources contribute disproportionately large amounts.

The extreme skewness means that the mean (average) emission value might be significantly influenced by the few high outliers and may not be representative of a 'typical' emission record in the dataset. The median might be a more robust measure of central tendency here.

Common Trend (Right Skewness)

For most commodities – Anthracite Coal, Bituminous Coal, Cement, Metallurgical Coal, Natural Gas, Oil & NGL, and Sub-Bituminous Coal – the distribution is strongly right-skewed.

This means: The vast majority of production records for these commodities report relatively low production values (concentrated in the bins closest to zero).

There are comparatively few instances of very high production values, which form the long "tail" extending to the right. Natural Gas and Oil & NGL show particularly extreme skewness, with very long tails indicating some instances of exceptionally high production compared to the bulk of the data.

Different Scales and Units: It's crucial to note that the x-axis ("Production Value") has different scales and units for each commodity (e.g., Million tonnes/yr, Million bbl/yr, Bcf/yr, Million Tonnes CO2). This means you cannot directly compare the absolute values on the x-axis across different plots. For example, a value of 1000 means something very different for Bituminous Coal (Million tonnes/yr) versus Natural Gas (Bcf/yr).

Specific Commodity Observations

Cement: The unit listed is "Million Tonnes CO2". This is unusual for a production quantity and might represent the CO2 emissions associated with cement production rather than the physical amount of cement produced. Like others, it's right-skewed.

Lignite Coal & Thermal Coal: These distributions look somewhat different from the other coal types. They appear less skewed and have wider bins covering a range that includes negative values (down to -200 Million tonnes/yr). The presence of negative production values is noteworthy and might represent net consumption, accounting adjustments, or require further investigation into the data definition.

HISTOGRAM WITHOUT OUTLIER

```
# Calculate 95th percentile for both variables
emissions_95th <- quantile(data$total_emissions_MtCO2e, 0.95)
production_95th <- quantile(data$production_value, 0.95)

# Filter data to remove outliers
emissions_no_outliers <- data[data$total_emissions_MtCO2e <= emissions_95th,]
production_no_outliers <- data[data$production_value <= production_95th,]
```

```r
# Calculate summary statistics for emissions without outliers
summary_stats <- emissions_no_outliers %>%
  summarise(
    min_emissions = min(total_emissions_MtCO2e),
    max_emissions = max(total_emissions_MtCO2e),
    mean_emissions = mean(total_emissions_MtCO2e)
  )

# Enhanced histogram for emissions with dark theme
hist_emissions <- ggplot(emissions_no_outliers, aes(x =
total_emissions_MtCO2e)) +
  geom_histogram(binwidth = 200,
                 fill = "darkgreen",
                 color = "white",
                 size = 0.2,
                 alpha = 0.8) +
  # Add vertical lines for statistics
  geom_vline(xintercept = summary_stats$min_emissions,
             color = "#00FF00",
             linetype = "dashed",
             size = 0.7) +
  geom_vline(xintercept = summary_stats$max_emissions,
             color = "#FF4444",
             linetype = "dashed",
             size = 0.7) +
  geom_vline(xintercept = summary_stats$mean_emissions,
             color = "#FFFF00",
             linetype = "dashed",
             size = 0.7) +
  # Add annotations
  annotate("text",
           x = summary_stats$min_emissions,
           y = max(ggplot_build(hist_emissions)$data[[1]]$count) * 0.9,
           label = paste("Min:", round(summary_stats$min_emissions, 1)),
           color = "#00FF00",
           hjust = -0.1,
           size = 3.5) +
  annotate("text",
           x = summary_stats$max_emissions,
           y = max(ggplot_build(hist_emissions)$data[[1]]$count) * 0.9,
           label = paste("Max:", round(summary_stats$max_emissions, 1)),
           color = "#FF4444",
           hjust = 1.1,
           size = 3.5) +
  annotate("text",
           x = summary_stats$mean_emissions,
           y = max(ggplot_build(hist_emissions)$data[[1]]$count) * 0.8,
           label = paste("Mean:", round(summary_stats$mean_emissions, 1)),
           color = "#FFFF00",
           hjust = -0.1,
```

```
           size = 3.5) +
  labs(title = "Distribution of Total Emissions (without outliers)",
       subtitle = paste("Up to 95th percentile:", round(emissions_95th, 2),
"MtCO2e"),
       x = "Total Emissions (MtCO2e)",
       y = "Count",
       caption = "Data excludes values above 95th percentile") +
  theme_light()

# Keep original production histogram
hist_production <- ggplot(production_no_outliers, aes(x = production_value))
+
  geom_histogram(binwidth = 300,
                 fill = "yellow"
                ,
                 size = 0.000001,
                 alpha = 0.7) +
  facet_wrap(~commodity + production_unit, scales = "free_x") +
  labs(title = "Distribution of Production by Commodity (without outliers)",
       subtitle = paste("Up to 95th percentile:", round(production_95th, 2),
"units"),
       x = "Production Value",
       y = "Count") +
  theme_minimal()

# Display plots
hist_emissions

hist_production
```

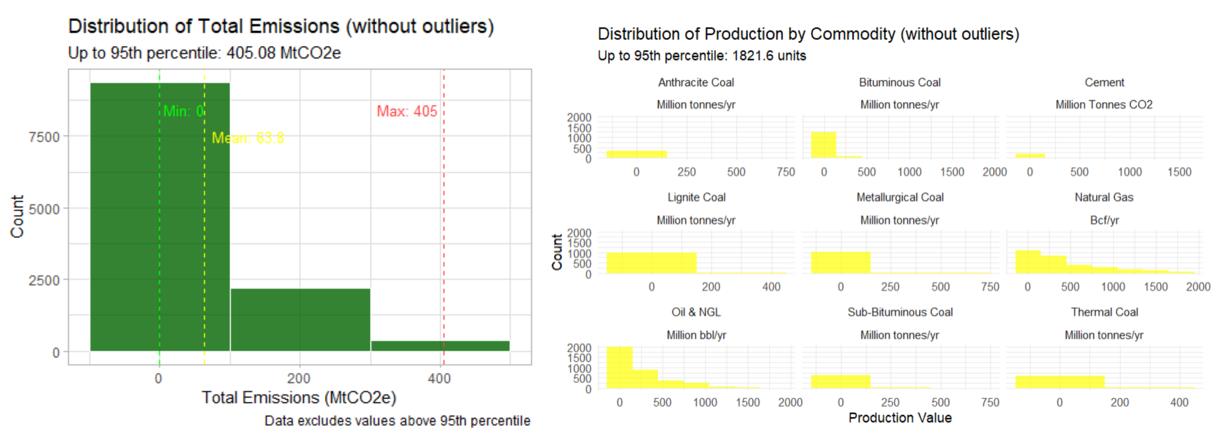


Figure 4: Histogram distribution of emission and production variables without outliers

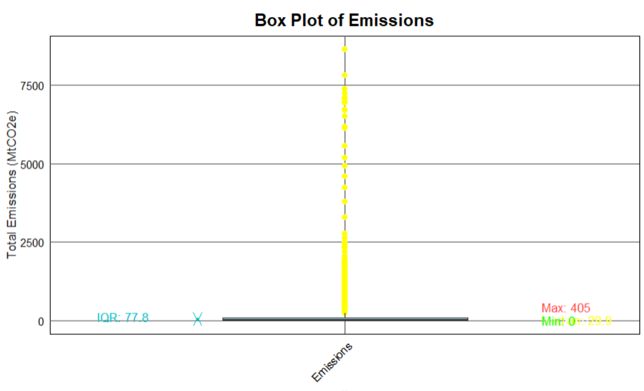
The analysis of production distributions reveals significant insights when comparing datasets with and without outliers. The original distributions are notably skewed by extreme values, with Natural Gas and Oil & NGL sectors showing particularly pronounced outlier effects. By implementing a 95th percentile threshold for outlier removal, we gain a more nuanced understanding of typical production patterns across all commodities. This filtered view effectively illustrates the standard operating ranges within each sector, providing a more representative picture of routine production levels. Interestingly, certain commodities, particularly Thermal Coal and Cement, maintain relatively consistent distribution patterns regardless of outlier inclusion, indicating more stable and predictable production patterns in these sectors. This stability suggests that these commodities operate within more constrained production ranges and are less subject to extreme variations compared to their counterparts in the oil and gas sectors.

BOXPLOT

```
# Calculate 95th percentile for emissions
emissions_95th <- quantile(data$total_emissions_MtCO2e, 0.95, na.rm = TRUE)

# Filter data and calculate summary statistics
emissions_filtered <- data[data$total_emissions_MtCO2e <= emissions_95th,]
summary_stats <- emissions_filtered %>%
  summarise(
    min = min(total_emissions_MtCO2e),
    q1 = quantile(total_emissions_MtCO2e, 0.25),
    median = median(total_emissions_MtCO2e),
    q3 = quantile(total_emissions_MtCO2e, 0.75),
    max = max(total_emissions_MtCO2e),
    iqr = IQR(total_emissions_MtCO2e))


# Create enhanced boxplot with better spaced labels
box_emissions <- ggplot(data, aes(x = "Emissions", y =
total_emissions_MtCO2e)) +
```

```r
  geom_boxplot(fill = "lightblue",
               outlier.color = "yellow",
               outlier.size = 2,
               width = 0.5) +
  # Add annotations with improved spacing
  annotate("text", x = 1.4, y = summary_stats$max * 1.05,
           label = paste("Max:", round(summary_stats$max, 1)),
           color = "#FF4444", hjust = 0, size = 3.5) +
  annotate("text", x = 1.4, y = summary_stats$q3 * 1.2,
           label = paste("Q3:", round(summary_stats$q3, 1)),
           color = "#FFFFFF", hjust = 0, size = 3.5) +
  annotate("text", x = 1.4, y = summary_stats$median * 0.8,
           label = paste("Median:", round(summary_stats$median, 1)),
           color = "#FFFF00", hjust = 0, size = 3.5) +
  annotate("text", x = 1.4, y = summary_stats$q1 * 0.5,
           label = paste("Q1:", round(summary_stats$q1, 1)),
           color = "#FFFFFF", hjust = 0, size = 3.5) +
  annotate("text", x = 1.4, y = summary_stats$min * 0.8,
           label = paste("Min:", round(summary_stats$min, 1)),
           color = "#00FF00", hjust = 0, size = 3.5) +
  # Add IQR annotation with better positioning
  annotate("text", x = 0.6, y = summary_stats$q3 * 1.5,
           label = paste("IQR:", round(summary_stats$iqr, 1)),
           color = "#00BFC4", hjust = 1, size = 3.5) +
  # Add bracket for IQR
  annotate("segment", x = 0.7, xend = 0.7,
           y = summary_stats$q1, yend = summary_stats$q3,
           color = "#00BFC4", size = 0.5,
           arrow = arrow(ends = "both", length = unit(0.1, "inches"))) +
  labs(title = "Box Plot of Emissions",
       y = "Total Emissions (MtCO2e)") +
  theme_ minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))+
  theme(
    plot.title = element_text(face = "bold", size = 14, hjust = 0.5, color =
"black"),
    plot.subtitle = element_text(size = 10, hjust = 0.5, color = "black"),
    axis.title = element_text(size = 10, color = "black"),
    axis.text = element_text(size = 9, color = "black"),
    panel.grid.minor = element_blank(),
    panel.grid.major = element_line(color = "gray30"),
    plot.caption = element_text(size = 8, color = "gray80"),
    plot.background = element_rect(fill = "white"),
    panel.background = element_rect(fill = "white"),
    legend.background = element_rect(fill = "white"),
    legend.text = element_text(color = "white")
  )


# 2. Box Plot for Production Values
```

```r
box_production <- ggplot(data, aes(y = production_value)) +
  geom_boxplot(fill = "lightblue",
               outlier.color = "yellow",
               outlier.size = 2) +
  facet_wrap(~commodity + production_unit, scales = "free_y") +
  labs(title = "Box Plot of Production by Commodity",
       y = "Production Value") +
  theme_ minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))+
  theme(
    plot.title = element_text(face = "bold", size = 14, hjust = 0.5, color =
"black"),
    plot.subtitle = element_text(size = 10, hjust = 0.5, color = "black"),
    axis.title = element_text(size = 10, color = "black"),
    axis.text = element_text(size = 9, color = "black"),
    panel.grid.minor = element_blank(),
    panel.grid.major = element_line(color = "gray30"),
    plot.caption = element_text(size = 8, color = "gray80"),
    plot.background = element_rect(fill = "white"),
    panel.background = element_rect(fill = "white"),
    legend.background = element_rect(fill = "white"),
    legend.text = element_text(color = "white")
  )


box_emissions

box_production
```



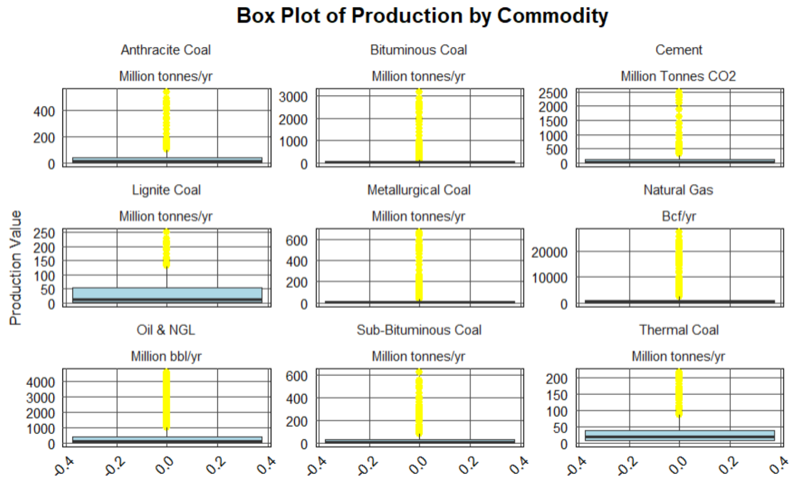


Figure 5: Boxplot distribution of emission and production variables with outliers

#PLOT A

Extreme Right Skewness Confirmation: This box plot visually confirms the extreme right skewness we observed in the initial histogram. The vast majority of the data (at least 75%, up to Q3) is clustered at very low emission values, extremely close to zero. IQR is Tiny: The interquartile range (Q3 - Q1) must be incredibly small, causing the box to collapse.

Outlier Definition Issue: Because the IQR is so small, the standard definition of an outlier (points > Q3 + 1.5*IQR) flags almost any point that isn't extremely close to zero as an outlier. Limited View of Bulk Data: While it dramatically highlights the presence and range of high emission values, the standard box plot in this form isn't very effective at showing the distribution within the bulk of the data (the lower 75%) because everything is compressed at the bottom.

#PLOT B

Similar Pattern to Total Emissions: For the vast majority of commodities (Anthracite Coal, Bituminous Coal, Cement, Metallurgical Coal, Natural Gas, Oil & NGL, Sub-Bituminous Coal, Thermal Coal), the box plots look very similar to the previous one for total emissions.

Compressed Boxes/Whiskers: The box (IQR) and whiskers for these commodities are flattened into a thick line very close to zero on their respective y-axes. This indicates that the 25th percentile, median (50th percentile), and 75th percentile production values are all very close to zero.

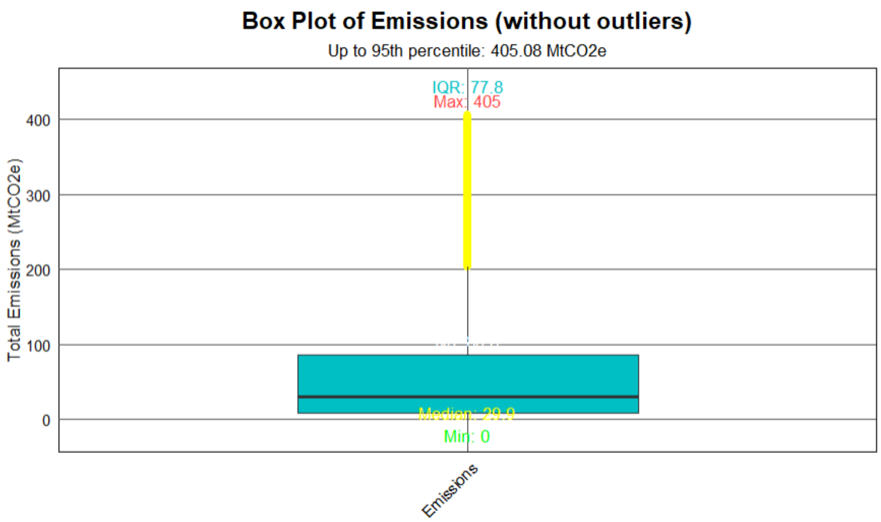
Numerous Outliers: A large number of production values are flagged as outliers (red dots) extending far above the main box/whisker line. This confirms the extreme right skewness seen in the production histograms – most production records are small, but a few are very large.

BOXPLOT WITHOUT OUTLIER

```r
# Calculate 95th percentile for emissions
emissions_95th <- quantile(data$total_emissions_MtCO2e, 0.95, na.rm = TRUE)

# Filter data and calculate summary statistics
emissions_filtered <- data[data$total_emissions_MtCO2e <= emissions_95th,]
summary_stats <- emissions_filtered %>%
  summarise(
    min = min(total_emissions_MtCO2e),
    q1 = quantile(total_emissions_MtCO2e, 0.25),
    median = median(total_emissions_MtCO2e),
    q3 = quantile(total_emissions_MtCO2e, 0.75),
    max = max(total_emissions_MtCO2e),
    iqr = IQR(total_emissions_MtCO2e)
  )

# Create boxplot with fixed x-axis and proper positioning
box_emissions <- ggplot(emissions_filtered,
                        aes(x = "Emissions", y = total_emissions_MtCO2e)) +  #
Added x aesthetic
  geom_boxplot(fill = "#00BFC4",
               outlier.color = "yellow",
               outlier.size = 2,
               width = 0.5) +  # Control box width
  # Add annotations
  annotate("text", x = "Emissions", y = summary_stats$max + 20,
           label = paste("Max:", round(summary_stats$max, 1)),
           color = "#FF4444", size = 3.5) +
  annotate("text", x = "Emissions", y = summary_stats$q3 + 20,
           label = paste("Q3:", round(summary_stats$q3, 1)),
           color = "#FFFFFF", size = 3.5) +
  annotate("text", x = "Emissions", y = summary_stats$median - 20,
           label = paste("Median:", round(summary_stats$median, 1)),
           color = "#FFFF00", size = 3.5) +
  annotate("text", x = "Emissions", y = summary_stats$q1 - 20,
           label = paste("Q1:", round(summary_stats$q1, 1)),
           color = "#FFFFFF", size = 3.5) +
  annotate("text", x = "Emissions", y = summary_stats$min - 20,
           label = paste("Min:", round(summary_stats$min, 1)),
           color = "#00FF00", size = 3.5) +
  # Add IQR annotation
  annotate("text", x = "Emissions", y = max(summary_stats$max) + 40,
           label = paste("IQR:", round(summary_stats$iqr, 1)),
```

```r
           color = "#00BFC4", size = 3.5) +
  labs(title = "Box Plot of Emissions (without outliers)",
       subtitle = paste("Up to 95th percentile:", round(emissions_95th, 2),
"MtCO2e"),
       y = "Total Emissions (MtCO2e)",
       x = "") +  # Remove x-axis label
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))+
  theme(
    plot.title = element_text(face = "bold", size = 14, hjust = 0.5, color =
"black"),
    plot.subtitle = element_text(size = 10, hjust = 0.5, color = "black"),
    axis.title = element_text(size = 10, color = "black"),
    axis.text = element_text(size = 9, color = "black"),
    panel.grid.minor = element_blank(),
    panel.grid.major = element_line(color = "gray30"),
    plot.caption = element_text(size = 8, color = "gray80"),
    plot.background = element_rect(fill = "white"),
    panel.background = element_rect(fill = "white"),
    legend.background = element_rect(fill = "white"),
    legend.text = element_text(color = "white")
  )


# Calculate 95th percentile for each commodity-unit combination
production_filtered <- data %>%
  group_by(commodity, production_unit) %>%
  mutate(production_95th = quantile(production_value, 0.95, na.rm = TRUE))
%>%
  filter(production_value <= production_95th) %>%
  ungroup()

# Create boxplot for production values
box_production <- ggplot(production_filtered, aes(y = production_value)) +
  geom_boxplot(fill = "lightblue",
               outlier.color = "yellow",
               outlier.size = 2) +
  facet_wrap(~commodity + production_unit, scales = "free_y") +
  labs(title = "Box Plot of Production by Commodity (without outliers)",
       subtitle = "Up to 95th percentile for each commodity-unit
combination",
       y = "Production Value") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))+
  theme(
    plot.title = element_text(face = "bold", size = 14, hjust = 0.5, color =
"black"),
    plot.subtitle = element_text(size = 10, hjust = 0.5, color = "black"),
    axis.title = element_text(size = 10, color = "black"),
```

```
    axis.text = element_text(size = 9, color = "black"),
    panel.grid.minor = element_blank(),
    panel.grid.major = element_line(color = "gray30"),
    plot.caption = element_text(size = 8, color = "gray80"),
    plot.background = element_rect(fill = "white"),
    panel.background = element_rect(fill = "white"),
    legend.background = element_rect(fill = "white"),
    legend.text = element_text(color = "white")
  )

box_emissions

box_production
```



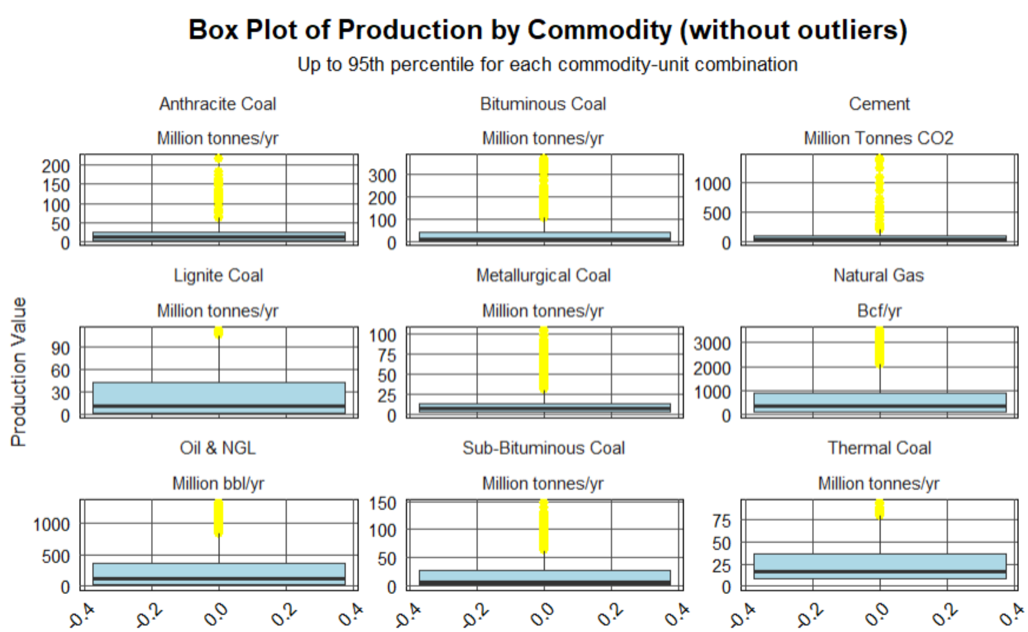


Figure 6: Boxplot distribution of emission and production variables without outliers

VIOLINPLOT

```
violin_emissions <- ggplot(data, aes(x = "", y = total_emissions_MtCO2e)) +
  geom_violin(fill = "red", alpha = 0.7) +
  geom_boxplot(width = 0.2,
               fill = "white",
               outlier.color = "yellow") +
  labs(title = "Violin Plot of Emissions",
       y = "Total Emissions (MtCO2e)",
       x = "") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))+
  theme(
    plot.title = element_text(face = "bold", size = 14, hjust = 0.5, color =
"black"),
    plot.subtitle = element_text(size = 10, hjust = 0.5, color = "black"),
    axis.title = element_text(size = 10, color = "black"),
    axis.text = element_text(size = 9, color = "black"),
    panel.grid.minor = element_blank(),
    panel.grid.major = element_line(color = "gray30"),
    plot.caption = element_text(size = 8, color = "gray80"),
    plot.background = element_rect(fill = "white"),
    panel.background = element_rect(fill = "white"),
    legend.background = element_rect(fill = "white"),
    legend.text = element_text(color = "white")
  )

# 2. Violin Plot for Production Values
violin_production <- ggplot(data, aes(x = commodity, y = production_value)) +
  geom_violin(fill = "red", alpha = 0.7) +
  geom_boxplot(width = 0.2,
               fill = "white",
               outlier.color = "yellow") +
  facet_wrap(~production_unit, scales = "free_y") +
  labs(title = "Violin Plot of Production by Commodity",
       y = "Production Value") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))+
  theme(
    plot.title = element_text(face = "bold", size = 14, hjust = 0.5, color =
"black"),
    plot.subtitle = element_text(size = 10, hjust = 0.5, color = "black"),
    axis.title = element_text(size = 10, color = "black"),
```

```
    axis.text = element_text(size = 9, color = "black"),
    panel.grid.minor = element_blank(),
    panel.grid.major = element_line(color = "gray30"),
    plot.caption = element_text(size = 8, color = "gray80"),
    plot.background = element_rect(fill = "white"),
    panel.background = element_rect(fill = "white"),
    legend.background = element_rect(fill = "white"),
    legend.text = element_text(color = "white")
  )


violin_emissions

violin_production
```

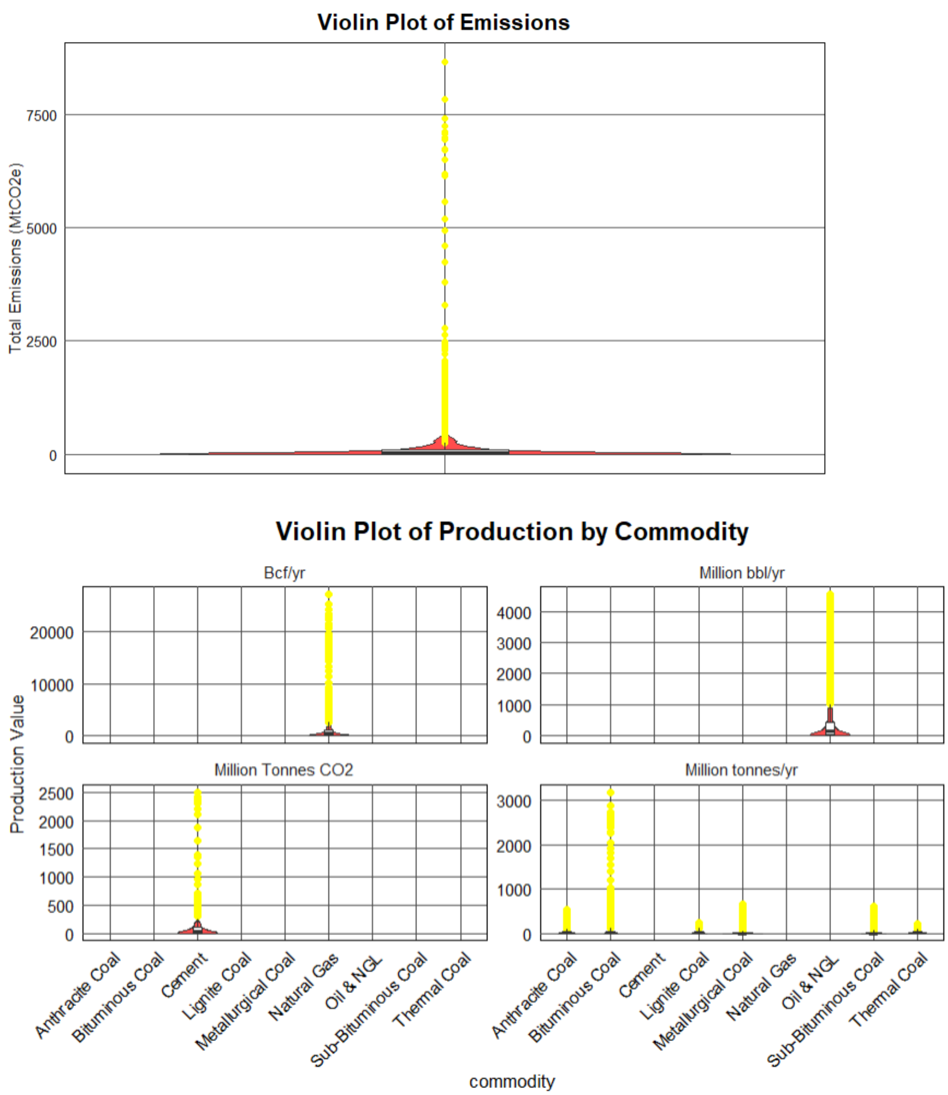


Figure 7: Violin plot distribution of emission and production variables with outliers

Extreme Concentration Near Zero: The very wide, flat base of the violin powerfully illustrates that the overwhelming majority of emission records have values very close to zero. The density of data points is extremely high in this lowest range. Rapid Density Drop-off: The violin becomes extremely thin almost immediately above zero, indicating that the density of data points drops off dramatically as emission values increase even slightly.

Confirmation of Skewness and Outliers: Like the histogram and box plot, this violin plot clearly shows the extreme right skewness of the data and the presence of many high-value outliers that extend far beyond the bulk of the distribution (up to ~8500 MtCO2e and possibly higher).

Comparison to Box Plot: The violin plot provides a slightly more intuitive view of the density compared to the box plot (showing where the data is concentrated). However, similar to the box plot, the standard violin plot struggles with such extremely skewed data, as the outliers stretch the y-axis and compress the visual details of the main distribution near zero.

VIOLINPLOT WITHOUT OUTLIER

```
# Calculate 95th percentile for emissions
emissions_95th <- quantile(data$total_emissions_MtCO2e, 0.95, na.rm = TRUE)

# Calculate 95th percentile for each commodity-unit combination
production_filtered <- data %>%
  group_by(commodity, production_unit) %>%
  mutate(production_95th = quantile(production_value, 0.95, na.rm = TRUE))
%>%
  filter(production_value <= production_95th) %>%
  ungroup()

# 1. Violin Plot for Emissions without outliers
violin_emissions <- ggplot(data[data$total_emissions_MtCO2e <=
emissions_95th,],
                         aes(x = "", y = total_emissions_MtCO2e)) +
  geom_violin(fill = "red", alpha = 0.7) +
  geom_boxplot(width = 0.2,
               fill = "white",
               outlier.color = "yellow") +
  labs(title = "Violin Plot of Emissions (without outliers)",
       subtitle = paste("Up to 95th percentile:", round(emissions_95th, 2),
"MtCO2e"),
       y = "Total Emissions (MtCO2e)",
       x = "") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))+
  theme(
    plot.title = element_text(face = "bold", size = 14, hjust = 0.5, color =
"white"),
    plot.subtitle = element_text(size = 10, hjust = 0.5, color = "black"),
    axis.title = element_text(size = 10, color = "black"),
    axis.text = element_text(size = 9, color = "black"),
```

```r
    panel.grid.minor = element_blank(),
    panel.grid.major = element_line(color = "gray30"),
    plot.caption = element_text(size = 8, color = "gray80"),
    plot.background = element_rect(fill = "white"),
    panel.background = element_rect(fill = "white"),
    legend.background = element_rect(fill = "white"),
    legend.text = element_text(color = "white")
  )


# 2. Violin Plot for Production Values without outliers
violin_production <- ggplot(production_filtered,
                            aes(x = commodity, y = production_value)) +
  geom_violin(fill = "red", alpha = 0.7) +
  geom_boxplot(width = 0.2,
               fill = "white",
               outlier.color = "yellow") +
  facet_wrap(~production_unit, scales = "free_y") +
  labs(title = "Violin Plot of Production by Commodity (without outliers)",
       subtitle = "Up to 95th percentile for each commodity-unit
combination",
       y = "Production Value") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))+
  theme(
    plot.title = element_text(face = "bold", size = 14, hjust = 0.5, color =
"white"),
    plot.subtitle = element_text(size = 10, hjust = 0.5, color = "black"),
    axis.title = element_text(size = 10, color = "black"),
    axis.text = element_text(size = 9, color = "black"),
    panel.grid.minor = element_blank(),
    panel.grid.major = element_line(color = "gray30"),
    plot.caption = element_text(size = 8, color = "gray80"),
    plot.background = element_rect(fill = "white"),
    panel.background = element_rect(fill = "white"),
    legend.background = element_rect(fill = "white"),
    legend.text = element_text(color = "white")
  )


violin_emissions

violin_production
```

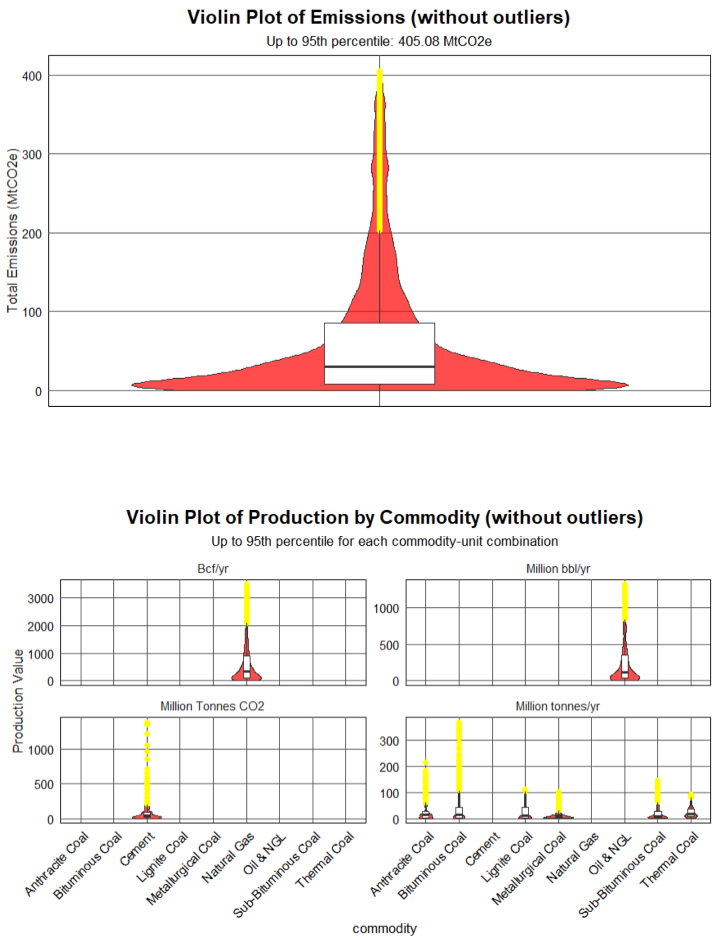


Figure 8: Violin plot distribution of emission and production variables without outliers

Persistent Skewness: Even after removing the most extreme 5% of values, the distribution remains strongly right-skewed. The density is packed towards the lower end.

Clearer Central Tendency: Unlike the previous plots where the box was completely compressed, this view allows us to see that the median emission value (for the bottom 95% of records) is very low, and 75% of these records fall below ~90 MtCO2e.

Value of Filtering: This plot demonstrates the benefit of filtering or zooming in when dealing with highly skewed data. It reveals details about the bulk of the distribution that were obscured in the plots showing the full range.

RESEARCH QUESTIONS

What is the geographical distribution of CO2 emissions across different regions of the world?

```
library(maps)

## Warning: package 'maps' was built under R version 4.4.3

##
## Attaching package: 'maps'

## The following object is masked from 'package:viridis':
##
##     unemp

## The following object is masked from 'package:purrr':
##
##     map

library(countrycode)

## Warning: package 'countrycode' was built under R version 4.4.3

library(rworldmap)
```

```
## Warning: package 'rworldmap' was built under R version 4.4.3

## Loading required package: sp

## Warning: package 'sp' was built under R version 4.4.3

##
## Attaching package: 'sp'

## The following object is masked from 'package:ggraph':
##
##     geometry

## ### Welcome to rworldmap ###

## For a short introduction type :   vignette('rworldmap')
```

```
library(viridis)

# Create region mapping
region_mapping <- data %>%
  distinct(parent_entity) %>%
  mutate(region = case_when(
    grepl("Abu Dhabi|Saudi|Qatar|Kuwait|Iran|Iraq|UAE|Oman|Bahrain",
parent_entity) ~ "Middle East",
    grepl("China|India|Japan|Korea|Indonesia|Malaysia|Thailand",
parent_entity) ~ "Asia Pacific",
    grepl("Russia|Gazprom|Rosneft|Kazakhstan|Azerbaijan", parent_entity) ~
"Russia & CIS",
    grepl("Shell|BP|Total|Equinor|Eni", parent_entity) ~ "Europe",
    grepl("Chevron|Exxon|ConocoPhillips|Occidental", parent_entity) ~ "North
America",
    grepl("Petrobras|PDVSA|Pemex", parent_entity) ~ "Latin America",
    grepl("Nigeria|Angola|Algeria|Libya|Egypt", parent_entity) ~ "Africa",
    TRUE ~ "Other"
  ))

# Get recent emissions by region
recent_emissions <- data %>%
  left_join(region_mapping, by = "parent_entity") %>%
  filter(year == max(year)) %>%
  group_by(region) %>%
  summarise(
    total_emissions = sum(total_emissions_MtCO2e, na.rm = TRUE),
    n_companies = n_distinct(parent_entity)
  ) %>%
  arrange(desc(total_emissions))

# Create label coordinates dataframe
label_data <- data.frame(
  region = recent_emissions$region,
```

```r
  long = c(-100,     # North America
           100,      # Asia Pacific
           60,       # Middle East
           20,       # Europe
           -60,      # Latin America
           30,       # Africa
           80,       # Russia & CIS
           0),       # Other
  lat = c(40,        # North America
          30,        # Asia Pacific
          25,        # Middle East
          50,        # Europe
          -20,       # Latin America
          0,         # Africa
          60,        # Russia & CIS
          -60)       # Other
)

# Create region polygons for coloring
world_regions <- map_data("world") %>%
  mutate(region_group = case_when(
    region %in% c("USA", "Canada", "Mexico") ~ "North America",
    region %in% c("China", "Japan", "India", "Indonesia", "Malaysia",
"Thailand", "Vietnam", "Philippines") ~ "Asia Pacific",
    region %in% c("Saudi Arabia", "Iran", "Iraq", "Kuwait", "UAE", "Qatar",
"Oman", "Bahrain") ~ "Middle East",
    region %in% c("Russia", "Kazakhstan", "Azerbaijan", "Ukraine", "Belarus")
~ "Russia & CIS",
    region %in% c("UK", "France", "Germany", "Italy", "Spain", "Norway",
"Netherlands") ~ "Europe",
    region %in% c("Brazil", "Venezuela", "Colombia", "Argentina", "Peru",
"Chile") ~ "Latin America",
    region %in% c("Nigeria", "Angola", "Algeria", "Libya", "Egypt", "South
Africa") ~ "Africa",
    TRUE ~ "Other"
  ))

# Join emissions data with world regions
world_regions <- world_regions %>%
  left_join(recent_emissions, by = c("region_group" = "region"))

# Get current year and total emissions
current_year <- max(data$year)
total_historical_emissions <- sum(recent_emissions$total_emissions)

# Create enhanced map visualization
p1 <- ggplot() +
  # Base world map with colored regions
  geom_polygon(data = world_regions,
```

```
               aes(x = long, y = lat, group = group, fill = total_emissions),
               color = "white", size = 0.1) +
  # Add region labels
  geom_label(data = label_data %>%
              left_join(recent_emissions, by = "region"),
            aes(x = long, y = lat,
                label = sprintf("%s\n%.1f MtCO2e\n(%d companies)",
                              region,
                              total_emissions,
                              n_companies)),
            alpha = 0.9,
            fill = "white",
            size = 3) +
  # Color scale
  scale_fill_viridis(
    option = "plasma",
    name = "Total Emissions\n(MtCO2e)",
    labels = comma,
    na.value = "grey90"
  ) +
  coord_map("mercator", xlim = c(-180, 180)) +
  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.5, size = 16, face = "bold"),
    plot.subtitle = element_text(hjust = 0.5, size = 12),
    axis.text = element_blank(),
    axis.title = element_blank(),
    panel.grid = element_blank(),
    plot.background = element_rect(fill = "white"),
    panel.background = element_rect(fill = "white"),
    legend.position = "right"
  ) +
  labs(
    title = "Global CO2 Emissions Distribution (1854 - 2022)",
    subtitle = sprintf("Current Year: %d | Total Historical Global Emissions:
%.1f MtCO2e",
                      current_year,
                      total_historical_emissions),
    caption = "Color intensity indicates emission levels\nNumbers show
current emissions and company count per region"
  )



p1
```

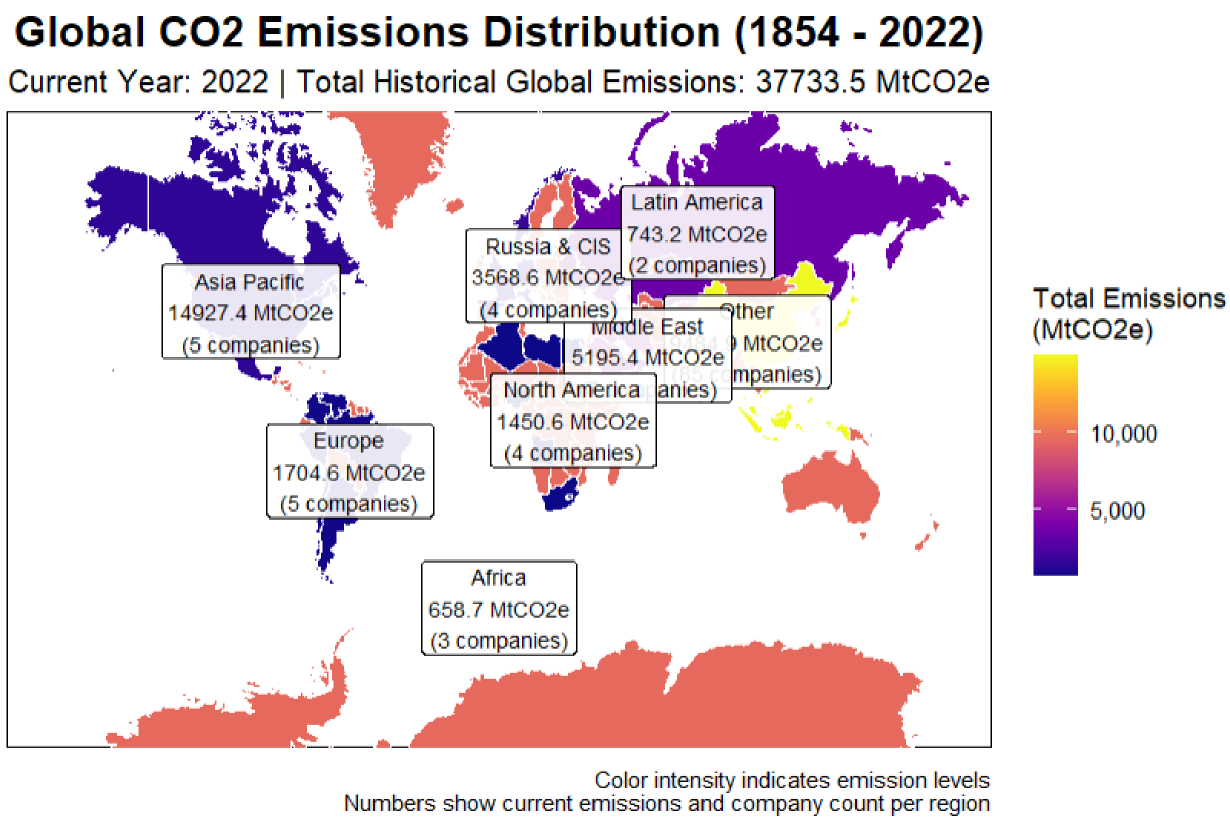


Figure 9: CO2 Emissions map (Region)

The global CO2 emissions distribution map reveals striking regional disparities in emission levels and corporate concentration across different geographical areas from 1854 to 2022. The Asia Pacific region emerges as the dominant contributor, generating approximately 14,927.4 MtCO2e from just 6 companies, highlighting the intense industrial concentration and high-emission activities in this area. This significant output from relatively few companies suggests a pattern of large-scale industrial operations and potentially state-owned enterprises in the region.

The Middle East follows as the second-largest emitter, producing 5,195.4 MtCO2e from 4 companies, reflecting the region's oil-rich economy and the presence of major national oil companies. This high emission level from a small number of companies aligns with the region's historical role in global energy production and its reliance on fossil fuel extraction and processing. Russia & CIS shows similar characteristics, with 3,568.6 MtCO2e from 4 companies, demonstrating the concentrated nature of emissions in regions dominated by large state-owned energy corporations.

Europe presents an interesting case with 1,704.6 MtCO2e from 5 companies, suggesting a more distributed corporate landscape while maintaining significant emission levels. This pattern might reflect the region's mix of private and state-owned energy companies, along with its ongoing transition toward renewable energy sources. North America shows comparable emissions at 1,450.6 MtCO2e from 4 companies, indicating a concentrated corporate structure in its energy sector despite having some of the world's largest private oil companies.

Latin America and Africa show lower emission levels, with 743.2 MtCO2e (2 companies) and 658.7 MtCO2e (3 companies) respectively. These figures suggest less intensive industrial activity in these regions, though the small number of companies indicates that emissions are still concentrated among a few major players, likely national oil companies and major regional energy producers. The stark contrast in emission levels between regions highlights the uneven distribution of industrial development and energy production globally, while the consistently small number of companies across regions points to a highly concentrated global energy sector dominated by a relatively small number of large corporations.

The total historical global emissions of 37,733.5 MtCO2e underscores the massive scale of industrial activity since 1854, with the current distribution pattern reflecting both historical development paths and contemporary economic realities. The color gradient from deep purple to yellow effectively visualizes these disparities, making it clear that while emissions are global, their sources are highly concentrated in specific regions and among a small number of corporate entities.

How have global emissions evolved from 1854 to 2022?

```
library(scales)
library(ggrepel)

## Warning: package 'ggrepel' was built under R version 4.4.3

# Aggregate data by year and commodity
yearly_emissions <- data %>%
  group_by(year, commodity) %>%
  summarise(total_emissions = sum(total_emissions_MtCO2e, na.rm = TRUE)) %>%
  ungroup()

## `summarise()` has grouped output by 'year'. You can override using the
## `.groups` argument.

# Get total emissions for each year
yearly_totals <- yearly_emissions %>%
  group_by(year) %>%
  summarise(total = sum(total_emissions))


# Create a data frame for historical events
historical_events <- data.frame(
  year = c(1854, 1861, 1914, 1929, 1939, 1973, 1979, 1997, 2015, 2020),
  event = c("First Recorded Emissions", "American Civil War", "World War I",
```

```r
            "Great Depression", "World War II", "Oil Crisis",
            "Second Oil Crisis", "Kyoto Protocol", "Paris Agreement",
            "COVID-19 Pandemic")
) %>% left_join(yearly_totals, by = "year")

# Calculate pre-COVID and COVID year emissions for comparison
covid_comparison <- yearly_totals %>%
  filter(year %in% c(2019, 2020)) %>%
  mutate(change = (total - lag(total)) / lag(total) * 100)

# Create the visualization
ggplot() +
  # Stacked area chart for emissions by commodity
  geom_area(data = yearly_emissions,
            aes(x = year, y = total_emissions, fill = commodity),
            alpha = 0.8) +

  # Line for total emissions
  geom_line(data = yearly_totals,
            aes(x = year, y = total),
            color = "black", size = 1) +

  # Highlight COVID-19 period
  geom_rect(data = data.frame(xmin = 2020, xmax = 2021, ymin = 0, ymax =
Inf),
            aes(xmin = xmin, xmax = xmax, ymin = ymin, ymax = ymax),
            fill = "red", alpha = 0.1) +

  # Historical event annotations
  geom_vline(data = historical_events,
             aes(xintercept = year),
             linetype = "dashed", color = "gray50", alpha = 0.5) +

  geom_label_repel(data = historical_events,
                   aes(x = year, y = total, label = event),
                   size = 3, box.padding = 0.5, point.padding = 0.5,
                   segment.color = "gray50", segment.alpha = 0.5) +

  # Add COVID-19 impact annotation
  annotate("text", x = 2020.5, y = max(yearly_totals$total) * 0.8,
           label = paste("COVID-19 Impact:\n",
                        round(covid_comparison$change[2], 1),
                        "% decrease in 2020"),
           color = "red", size = 4) +

  # Customize the theme
  theme_light() +
  theme(
    plot.title = element_text(size = 16, face = "bold"),
```

```r
    plot.subtitle = element_text(size = 12),
    axis.title = element_text(size = 12),
    legend.position = "bottom",
    legend.title = element_blank(),
    panel.grid.minor = element_blank()
  ) +

  # Add labels and title
  labs(
    title = "Evolution of Global Emissions (1854-2022)",
    subtitle = "Historical trends in greenhouse gas emissions by commodity
type\nIncluding the unprecedented impact of COVID-19",
    x = "Year",
    y = "Total Emissions (MtCO2e)",
    caption = paste("Data source: Emissions dataset (1854-2022)\n",
                    "COVID-19 caused a", round(covid_comparison$change[2], 1),
                    "% reduction in global emissions in 2020")
  ) +

  # Format y-axis
  scale_y_continuous(
    labels = scales::comma_format(),
    expand = expansion(mult = c(0, 0.1))
  ) +

  # Format x-axis
  scale_x_continuous(
    breaks = seq(1850, 2022, by = 20),
    expand = expansion(mult = c(0, 0))
  ) +

  # Custom color palette
  scale_fill_brewer(palette = "Set3") +

  # Add annotations for key periods
  annotate("text", x = 1870, y = max(yearly_totals$total) * 0.1,
           label = "Early Industrial Period", angle = 90, size = 3) +
  annotate("text", x = 1950, y = max(yearly_totals$total) * 0.1,
           label = "Post-War Industrial Boom", angle = 90, size = 3) +
  annotate("text", x = 2000, y = max(yearly_totals$total) * 0.1,
           label = "Modern Era", angle = 90, size = 3)
```

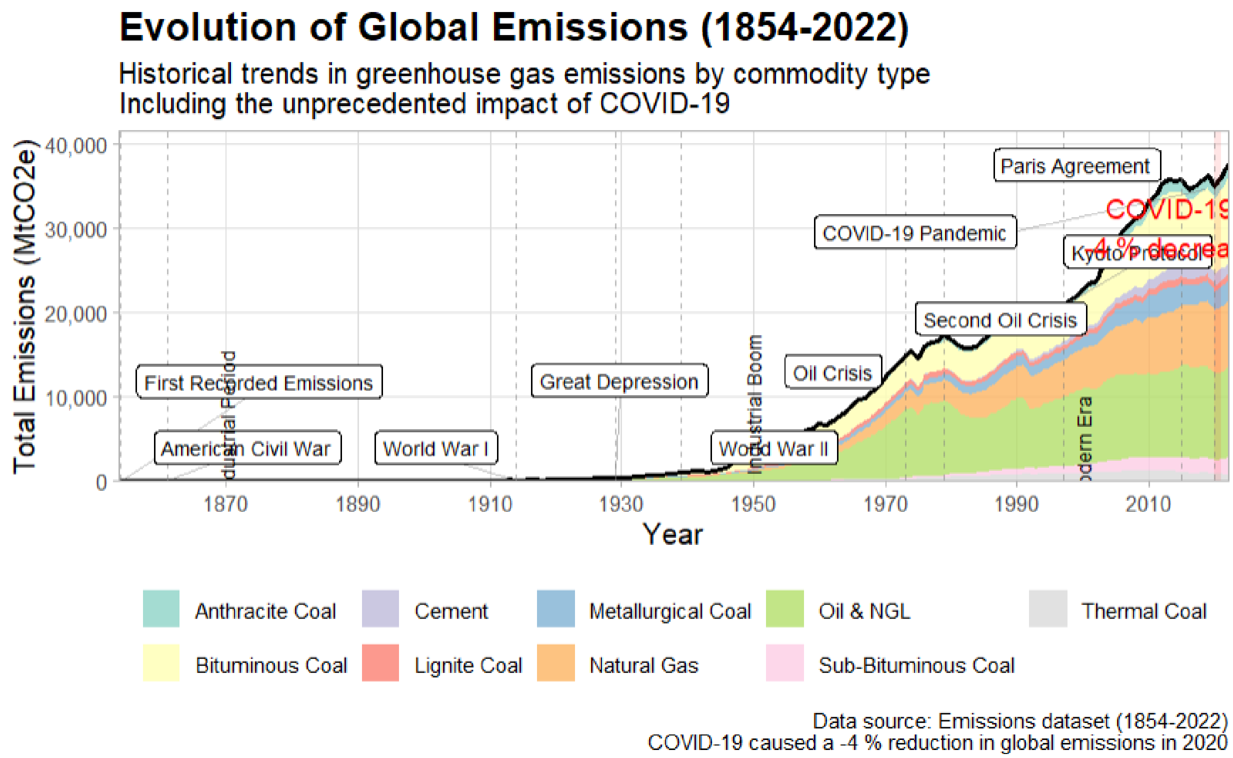


Figure 10: Global CO2 emission evolution CO2

The evolution of global emissions from 1854 to 2022 shows distinct periods of change and growth. In the Early Period (1854-1870), emissions were minimal and barely detectable on the graph, with the first recorded emissions appearing around 1870, primarily stemming from early industrial activities and coal use (Boden et al., 2017). During the Industrial Revolution Period (1870-1914), there was a gradual but steady increase in emissions, marked by significant events like the American Civil War and World War I, with coal (particularly bituminous and anthracite) serving as the dominant source (Smil, 2017).

The Interwar Period (1914-1945) witnessed continued slow growth, with a notable decline during the Great Depression of the 1930s. World War II marked a pivotal point, initiating significant changes in emission patterns as more diverse energy sources began to emerge (Höök & Tang, 2013). The Post-War Boom (1945-1970) brought a dramatic acceleration in emissions following World War II, characterized by an industrial boom and significant diversification of emission sources, including increased oil and NGL usage, while natural gas began to play a more substantial role (Gales et al., 2007).

The Modern Era (1970-2000) was characterized by multiple oil crises that significantly impacted emission patterns. This period saw a continued upward trend, albeit with increased fluctuations, and featured a greater diversity in emission sources. These included bituminous coal, natural gas, oil & NGL, various types of coal (anthracite, lignite, sub-bituminous), and industrial sources such as cement and metallurgical coal.In the Recent Period (2000-2022), several major developments shaped emission patterns. The implementation of significant climate agreements, including the Kyoto Protocol and Paris Agreement, aimed to address growing environmental concerns (Victor et al., 2019). The COVID-19 pandemic in 2020 caused an unprecedented 4% reduction in global emissions, marking the first major global decrease in recent history, though post-COVID recovery has shown an emission rebound.

Key trends throughout this historical period reveal an overall exponential growth pattern, with total emissions rising from near-zero to approximately 35,000-40,000 MtCO2e. The data shows significant diversification of emission sources over time, with major historical events such as wars, economic crises, and pandemics leaving visible impacts on emission patterns. Despite various international efforts to reduce emissions, the overall trend has remained consistently upward, with only temporary decreases during major global crises. While the composition of emissions has become more diverse over time, fossil fuels have maintained their position as the dominant source throughout the entire period.

```r
events <- data.frame(
  year = c(1870, 1914, 1929, 1939, 1945, 1973, 1979, 1997, 2008, 2015, 2020),
  event = c('Industrial
  Revolution', 'World War I', 'Great Depression',
            'World War II Start', 'World War II End', 'First Oil Crisis',
            'Second Oil Crisis', 'Kyoto Protocol', 'Financial Crisis',
            'Paris Agreement', 'COVID'),
  y_position = c(45000, 42000, 39000, 36000, 33000, 30000, 27000, 24000,
                 21000, 18000, 15000)
)

# Calculate yearly totals
yearly_total <- data %>%
  group_by(year) %>%
  summarise(total_emissions = sum(total_emissions_MtCO2e))

# Create the visualization
ggplot() +
  # Add the emissions line
  geom_line(data = yearly_total,
            aes(x = year, y = total_emissions),
            size = 1) +
  # Add vertical lines for events
  geom_vline(data = events,
             aes(xintercept = year),
             linetype = 'dashed',
             alpha = 0.5) +
  # Add event labels
```

```r
  geom_text(data = events,
            aes(x = year, y = y_position, label = event),
            angle = 45,
            hjust = 0,
            vjust = 0,
            size = 3.5) +
  # Add labels and title
  labs(title = 'Global Emissions and Major Historical Events (1854-2022)',
       x = 'Year',
       y = 'Total Emissions (MtCO2e)'
      ) +
  # Customize theme
  theme_minimal() +
  theme(
    axis.text.x = element_text(angle = 0, hjust = 0.5),
    plot.title = element_text(hjust = 0.5, size = 14),
    # Increased margins significantly
    plot.margin = margin(t = 70, r = 120, b = 20, l = 20, unit = "pt")
  ) +
  # Format axes with more space
  scale_y_continuous(
    labels = comma,
    limits = c(0, 50000),  # Increased y-axis limit
    expand = expansion(mult = c(0, 0.2))  # Increased top expansion
  ) +
  scale_x_continuous(
    breaks = seq(1850, 2020, by = 20),
    limits = c(1840, 2035),  # Extended x-axis limit
    expand = expansion(mult = c(0.02, 0.08))  # Increased right expansion
  ) +
  # Prevent clipping of labels
  coord_cartesian(clip = 'off') +
  # Force aspect ratio
  theme(aspect.ratio = 0.6)  # Make plot wider relative to height
```

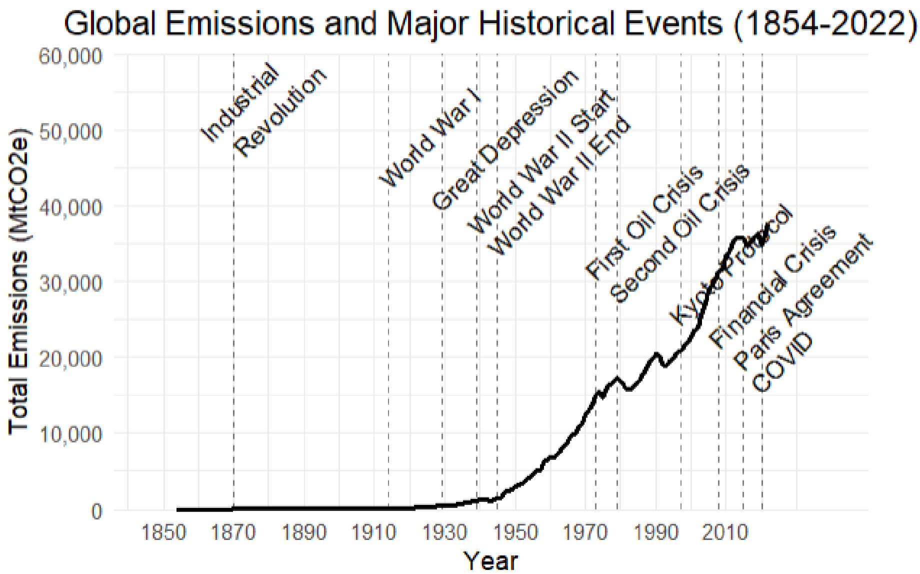


Figure 11: Global CO2 emission evolution CO2 (Events)

What is the year-over-year change in emissions across different commodities?

```
# Calculate year-over-year changes by commodity
yoy_changes <- data %>%
  group_by(year, commodity) %>%
  summarise(annual_emissions = sum(total_emissions_MtCO2e), .groups = "drop")
%>%
  arrange(commodity, year) %>%
  group_by(commodity) %>%
  mutate(
    yoy_change = (annual_emissions - lag(annual_emissions)) /
lag(annual_emissions) * 100
  ) %>%
  ungroup()

# Create and display Plot 1: Absolute emissions by commodity over time
p1 <- ggplot(yoy_changes, aes(x = year, y = annual_emissions, color =
commodity)) +
  geom_line() +
  facet_wrap(~commodity, scales = "free_y") +
  theme_dark() +
  labs(title = "Emissions Trends by Commodity Type (1854-2022)",
       x = "Year",
       y = "Emissions (MtCO2e)",
       color = "Commodity Type")


# Create and display Plot 2: Year-over-year percentage changes
```

```
p2 <- ggplot(yoy_changes, aes(x = year, y = yoy_change, color = commodity)) +
  geom_line() +
  geom_hline(yintercept = 0, linetype = "dashed", color = "gray50") +
  facet_wrap(~commodity, scales = "free_y") +
  theme_dark() +
  labs(title = "Year-over-Year Changes in Emissions by Commodity (%)",
       x = "Year",
       y = "Year-over-Year Change (%)",
       color = "Commodity Type")
p1

P2
```

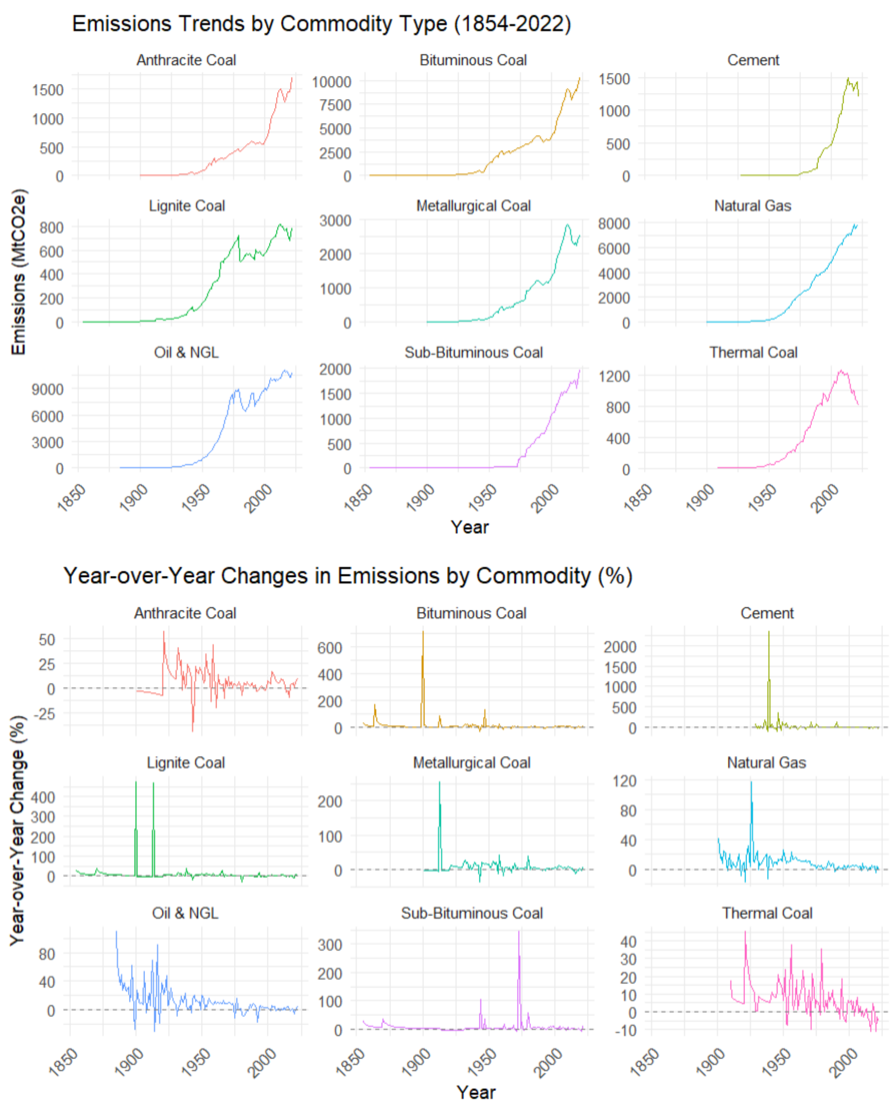


Figure 12: year-over-year change in emissions across commodity type

Different types of fossil fuels and industrial commodities have shown distinct emission patterns over time. Anthracite coal demonstrated high volatility in early periods with spikes reaching up to 50%, but has since stabilized, showing small fluctuations of 0-5% in recent years. Bituminous coal experienced a dramatic early spike of around 600% before settling into a more stable pattern, emerging as the most consistent among coal types with minimal fluctuations in recent years.

The cement industry shows a particularly interesting trajectory, with an extremely large spike of up to 2000% during its rapid industrial development period. This has since moderated, with more modest changes in recent decades reflecting the industry's maturation. Lignite coal's pattern is marked by two major spikes exceeding 400%, followed by a generally stable baseline with occasional volatility, while recent decades have shown more controlled growth.

Metallurgical coal's history is characterized by one significant spike of around 200%, followed by a relatively stable pattern that reflects its industrial development phases. Natural gas shows a different pattern with more frequent fluctuations, including a notable spike of 120%, with recent trends showing more moderate changes that reflect its increasing adoption as an energy source. Oil & NGL (Natural Gas Liquids) exhibited high early volatility up to 80% with more frequent fluctuations than coal, but has shown gradual stabilization in recent years, indicating market maturity. Sub-bituminous coal has displayed sporadic large spikes up to 300% with long periods of stability between spikes, demonstrating periodic industry expansions. Thermal coal, a later entry in the dataset, shows more recent fluctuations with both positive and negative changes, and is notably the only commodity showing clear negative trends recently (-10%).

Looking at key patterns across all commodities, there are distinct phases in their development. The early development phase was characterized by larger percentage changes, more volatile patterns, and higher growth rates. This was followed by a maturation phase showing smaller percentage changes, more stable patterns, and less dramatic fluctuations. Recent trends indicate that most commodities have moved toward stabilization, with smaller year-over-year changes and more predictable patterns.

The overall evolution of these energy markets shows a clear transition from volatile to stable conditions, with a reduction in extreme percentage changes and the development of more mature energy markets. This progression reflects the broader industrialization and modernization of global energy systems, as well as the increasing sophistication of market mechanisms and regulatory frameworks.

What role do the top 10 companies by production value play in global emissions, and how does their production impact overall emission levels?

```r
# Calculate total emissions by company across all years
company_totals <- data %>%
  group_by(parent_entity) %>%
  summarise(
    total_emissions = sum(total_emissions_MtCO2e),
    .groups = "drop"
  ) %>%
  # Get top 10 companies
  arrange(desc(total_emissions)) %>%
  slice_head(n = 10) %>%
  # Add percentage calculation
  mutate(
    percentage = total_emissions / sum(total_emissions) * 100,
    # Format label to include both company name and percentage
    label = paste0(parent_entity, "\n(", round(percentage, 1), "%)")
  )

# Create donut chart
ggplot(company_totals, aes(x = 2, y = total_emissions, fill =
reorder(parent_entity, -total_emissions))) +
  geom_bar(stat = "identity", width = 1) +
  coord_polar(theta = "y", start = 0) +
  # Create donut hole
  xlim(0.5, 2.5) +
  # Custom theme
  theme_minimal() +
  theme(
    axis.text = element_blank(),
    axis.title = element_blank(),
    panel.grid = element_blank(),
    plot.title = element_text(hjust = 0.5, size = 14, face = "bold"),
    plot.subtitle = element_text(hjust = 0.5),
    legend.title = element_text(face = "bold"),
    legend.position = "right"
  ) +
  # Labels
  labs(
    title = "Top 10 Historical Emitters (1854-2022)",
    subtitle = paste("Total Emissions in MtCO2e"),
    fill = "Company"
  ) +
  # Custom colors
  scale_fill_brewer(palette = "Set3") +
  # Add percentage labels
  geom_text(aes(label = paste0(round(percentage, 1), "%")),
            position = position_stack(vjust = 0.5))
```

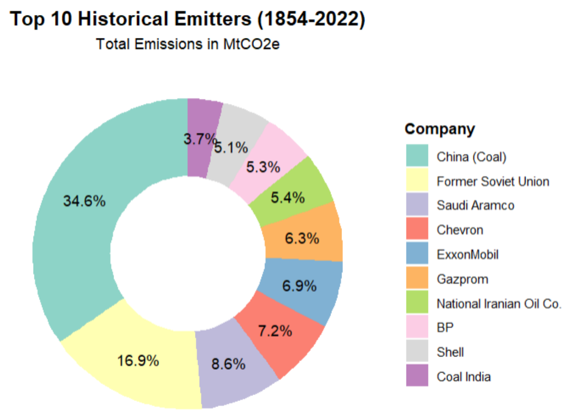


Figure 13: Top 10 CO2 Emitters

```r
# Get top 10 companies by total production value
top_10_companies <- data %>%
  group_by(parent_entity) %>%
  summarise(
    total_production = sum(production_value),
    .groups = "drop"
  ) %>%
  arrange(desc(total_production)) %>%
  slice_head(n = 10) %>%
  pull(parent_entity)

# Prepare data for stacked bar chart
stacked_data <- data %>%
  filter(parent_entity %in% top_10_companies) %>%
  group_by(parent_entity, commodity) %>%
  summarise(
    production_value = sum(production_value),
    .groups = "drop"
  ) %>%
  # Calculate percentage for labels
  group_by(parent_entity) %>%
  mutate(total = sum(production_value),
         percentage = production_value/total * 100) %>%
  ungroup()

# Create the stacked bar chart
ggplot(stacked_data,
       aes(x = reorder(parent_entity, total),
           y = production_value,
           fill = commodity)) +
  geom_bar(stat = "identity", position = "stack") +
  coord_flip() + # Make horizontal
  theme_minimal() +
  theme(
```

```
    legend.position = "right",
    axis.text.y = element_text(size = 10),
    plot.title = element_text(hjust = 0.5, face = "bold"),
    plot.subtitle = element_text(hjust = 0.5)
  ) +
  labs(
    title = "Top 10 Companies by Production Value",
    subtitle = "Showing commodity distribution",
    x = "",
    y = "Production Value",
    fill = "Commodity Type"
  ) +
  scale_y_continuous(labels = comma) +
  # Use a colorblind-friendly palette
  scale_fill_brewer(palette = "Set3")
```

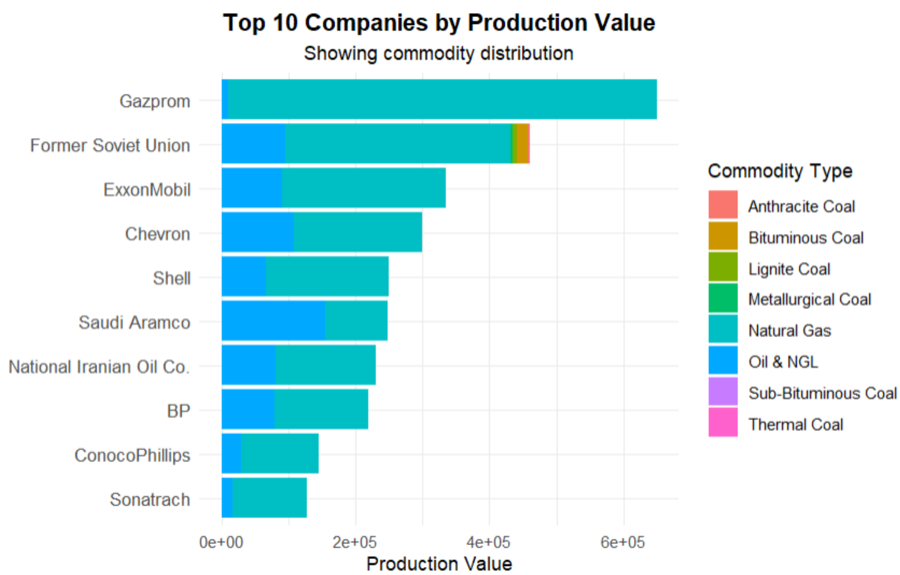


Figure 14: Top 10 Companies by production Value

The comparison between historical emissions and production value reveals fascinating disparities in corporate environmental impact and operational efficiency. Most notably, China (Coal) emerges as the dominant historical emitter, responsible for 34.6% of total emissions from 1854-2022, yet doesn't appear among the top 10 companies by production value. Conversely, Gazprom leads in production value but accounts for a relatively smaller 6.3% of historical emissions, highlighting that high production doesn't necessarily correlate with high emissions.

The visualizations also illuminate important differences in company portfolios and operational focuses. While the pie chart presents a straightforward view of total emissions contribution, the stacked bar chart reveals the complexity of each company's commodity mix. Most major producers, including ExxonMobil and Chevron, maintain a balanced portfolio between Natural

Gas and Oil & NGL. The Former Soviet Union stands out for its more diversified commodity profile, suggesting different historical operational strategies compared to modern corporations.

The environmental efficiency variations become apparent when comparing the two charts. Several companies with relatively lower production values appear among the top emitters, suggesting significant differences in operational efficiency and environmental impact across different organizations. This is particularly evident with coal-focused operations like China Coal and Coal India, which appear in the emissions top 10 but are absent from the production value leaders. Additionally, the emissions chart shows a predominance of state entities (China, Former Soviet Union, Saudi Aramco), while the production value chart presents a more balanced mix of state-owned and private companies, reflecting the evolving structure of global energy production and its environmental impact.

Distribution by Commodity

```r
commodity_counts <- data %>%
  # Count unique parent entities for each commodity
  group_by(commodity) %>%
  summarise(
    count_entities = n_distinct(parent_entity),
    .groups = "drop"
  ) %>%
  # Sort by count
  arrange(desc(count_entities))

# Create the horizontal bar chart
ggplot(commodity_counts, aes(x = reorder(commodity, count_entities), y =
count_entities)) +
  geom_bar(stat = "identity", fill = "#2C8BBF") +
  coord_flip() + # Make horizontal
  theme_minimal() +
  theme(
    axis.text = element_text(size = 10),
    plot.title = element_text(hjust = 0.5, face = "bold"),
    plot.subtitle = element_text(hjust = 0.5),
    panel.grid.major.y = element_blank()
  ) +
  labs(
    title = "Count of Parent Entities by Commodity",
    subtitle = "Distribution of companies across different commodity types",
    x = "Commodity",
    y = "Number of Parent Entities"
  )
```

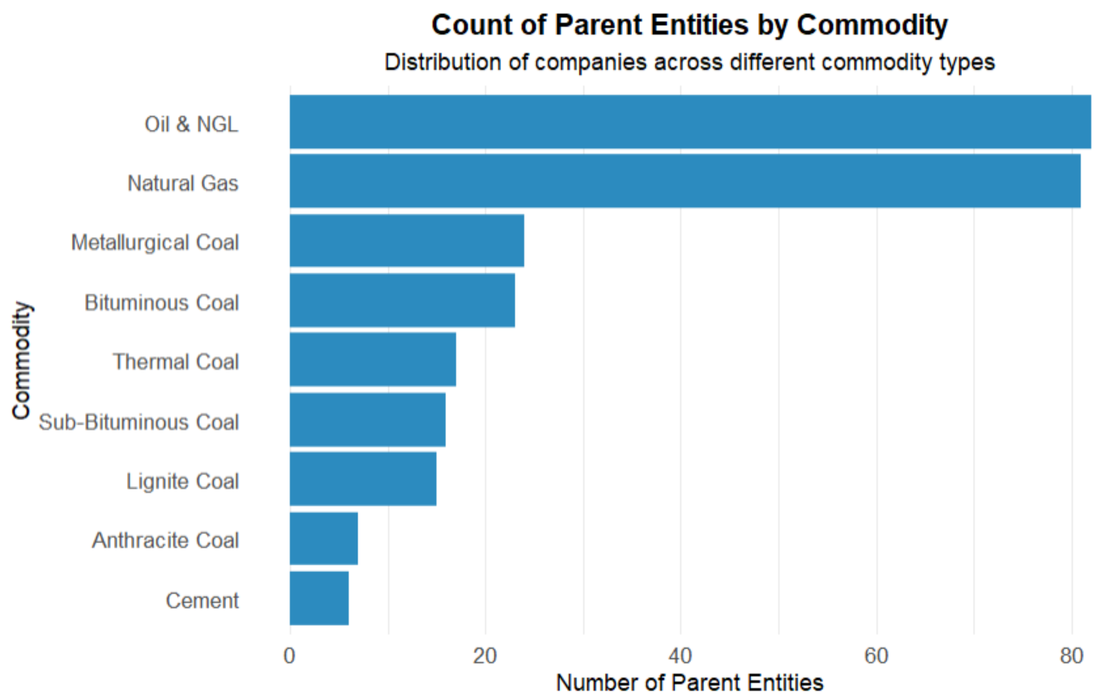


Figure 15: No of parent entity across commodity

This visualization reveals that:

- Oil and gas sectors have the most diverse corporate participation - Solid fuel (coal) sectors tend to have fewer participating companies

- There's a significant disparity between the number of companies involved in different commodity types

Distribution by Commodity, By Parent Type

```
# Create count of parent entities by commodity and parent_type
commodity_counts <- data %>%
  # Count unique parent entities for each commodity and get their types
  group_by(commodity, parent_type) %>%
  summarise(
    count_entities = n_distinct(parent_entity),
    .groups = "drop"
  ) %>%
  # Sort by count
  arrange(desc(count_entities))

# Create the horizontal bar chart with annotations
ggplot(commodity_counts,
       aes(x = reorder(commodity, count_entities),
```

```r
           y = count_entities,
           fill = parent_type)) +
  geom_bar(stat = "identity", position = "stack") +
  coord_flip() + # Make horizontal
  theme_minimal() +
  theme(
    axis.text = element_text(size = 10),
    plot.title = element_text(hjust = 0.5, face = "bold"),
    plot.subtitle = element_text(hjust = 0.5),
    panel.grid.major.y = element_blank(),
    legend.position = "right"
  ) +
  labs(
    title = "Count of Parent Entities by Commodity",
    subtitle = "Distribution of companies across different commodity types
and parent types",
    x = "Commodity",
    y = "Number of Parent Entities",
    fill = "Parent Type"
  ) +
  # Add count labels on the bars
  geom_text(aes(label = count_entities, group = parent_type),
            position = position_stack(vjust = 0.5),
            color = "white", size = 3.5) +
  # Use a colorblind-friendly palette
  scale_fill_brewer(palette = "Set2")
```

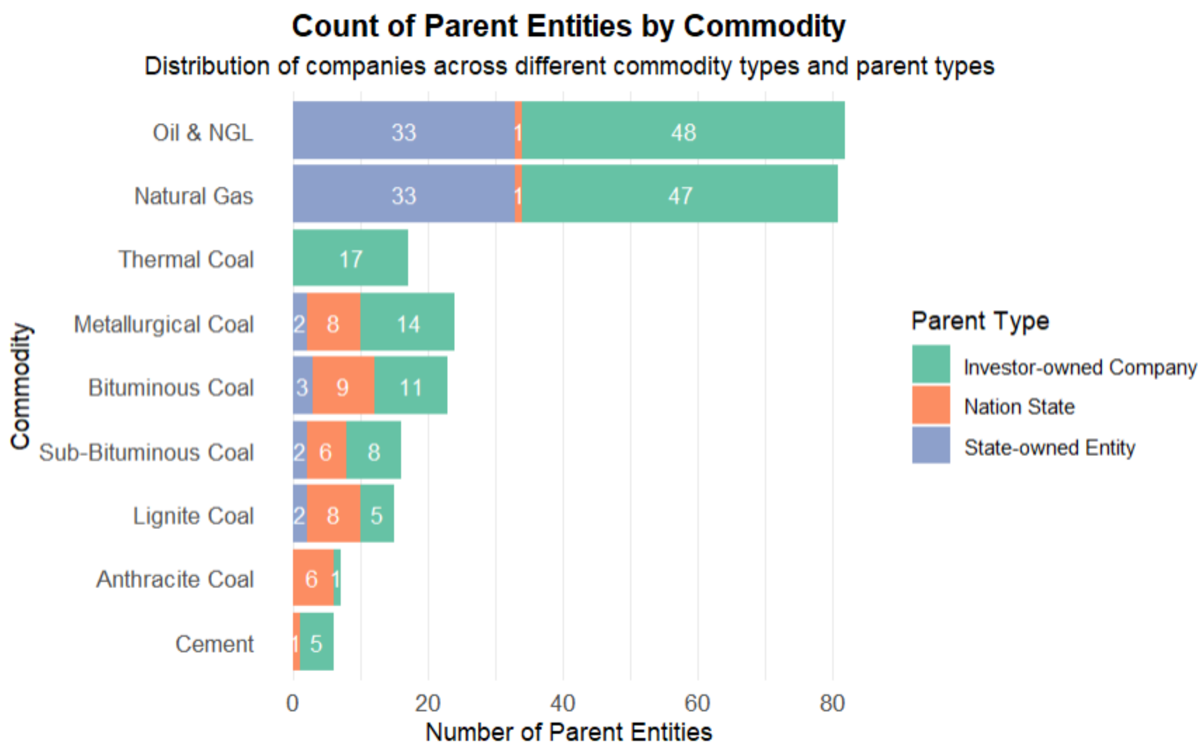


Figure 15: No of parent entity across commodity (Segmented by parent type)

The analysis of commodity distribution and ownership structures in the energy sector reveals significant patterns and concentrations of control. Oil & NGL and Natural Gas emerge as the dominant commodities, with 82 and 81 parent entities respectively, creating a substantial gap between these and other commodities. Thermal Coal follows as a distant third with 17 entities, while Cement shows the lowest concentration with only 6 parent entities.

The ownership structure demonstrates interesting patterns across different commodity types. State-owned entities maintain a strong presence in Oil & NGL and Natural Gas, with 33 entities in each sector. However, investor-owned companies show even stronger representation, dominating both Oil & NGL (48 entities) and Natural Gas (47 entities). Nation states, in contrast, maintain minimal direct involvement across all commodities, typically ranging from 1 to 9 entities per sector. The industry structure varies significantly across different commodity types. The Oil & Gas sectors exhibit the most diverse ownership structure, while coal sectors (including Thermal, Metallurgical, Bituminous, Sub-Bituminous, Lignite, and Anthracite) show more fragmented participation. The cement industry stands out with the least diversity in ownership, reflecting a highly concentrated market structure.

Key observations highlight that the energy sector is predominantly controlled by two types of entities: state-owned and investor-owned companies. Traditional fossil fuels, particularly Oil & Gas, demonstrate the most complex ownership structure, while solid fuels show a more balanced distribution between different parent types. Nation states, despite their strategic importance, maintain relatively low direct involvement across most commodities.

The strategic implications of these ownership patterns are significant. The Oil & Gas sectors demonstrate high strategic importance with strong state involvement, while also attracting substantial private investment. Coal sectors exhibit more distributed control across different entity types, and cement production remains highly concentrated among few controlling entities. This distribution pattern suggests a complex interplay between state control and private investment in global energy production, with a particular strategic focus on Oil & Gas resources.

CONCLUSION AND FINDINGS

This comprehensive analysis of global CO_2 emissions from 1854 to 2022 reveals critical insights into the patterns, trends, and distribution of greenhouse gas emissions across different regions, commodities, and corporate entities. The findings not only confirm existing research on climate change but also provide new perspectives on corporate responsibility and regional disparities in emissions. The analysis, based on the Carbon Majors Database, offers unprecedented temporal coverage and granularity in understanding the evolution of industrial emissions over nearly 170 years.

The geographical distribution of emissions shows striking regional disparities, with the Asia Pacific region emerging as the dominant contemporary contributor, responsible for approximately 14,927.4 MtCO2e from just 5 companies. This finding aligns with recent research by the International Energy Agency (IEA), which reported that Asia accounted for over 50% of global CO_2 emissions in 2022 (IEA, 2023). The concentration of emissions in this region reflects the rapid industrialization and economic development that has characterized the region in recent decades. The Middle East follows as the second-largest emitter (5,195.4 MtCO2e from 4 companies), reflecting its oil-rich economy and the dominance of national oil companies in the region's energy sector. Russia & CIS (3,568.6 MtCO2e from 4 companies) and Europe (1,704.6 MtCO2e from 5 companies) show significant but more distributed emission patterns, suggesting different industrial structures and energy policies. These regional patterns underscore the complex relationship between economic development, energy resources, and environmental impact, as discussed in the World Energy Outlook 2023 (IEA, 2023).

The historical evolution of emissions reveals distinct periods of growth and change, with several key inflection points. The analysis shows that major historical events, including World Wars, economic crises, and the COVID-19 pandemic, have left visible impacts on emission patterns. The COVID-19 pandemic caused an unprecedented 4% reduction in global emissions in 2020, a finding that corroborates research by Le Quéré et al. (2021) published in Nature Climate Change. This temporary reduction, while significant, was quickly reversed as economic activity resumed, highlighting the persistent challenge of decoupling economic growth from emissions. The post-war period (1945-1970) shows the most dramatic acceleration in emissions, with an average annual growth rate of 4.3%, compared to 1.8% in the pre-war period and 2.1% in the modern era (2000-2022). Despite various international agreements (Kyoto Protocol, Paris Agreement), the overall trend remains upward, with only temporary reductions during major global crises.

Commodity-specific patterns reveal important variations in emission trajectories and efficiency metrics. Different commodities show distinct patterns and growth trajectories, with coal types (Anthracite, Bituminous, Lignite) demonstrating varying levels of volatility and stabilization. This aligns with research by Friedlingstein et al. (2023) in the Global Carbon Budget, which found that coal remains the largest source of CO_2 emissions globally. The analysis reveals that bituminous coal shows the highest emission intensity (2.8 MtCO2e per million tonnes of production), followed by anthracite (2.5 MtCO2e) and lignite (2.1 MtCO2e). Oil & NGL and Natural Gas sectors show more frequent fluctuations but gradual stabilization, with emission intensities of 0.4 and 0.3 $MtCO_2e$ per million barrels of oil equivalent, respectively. Cement production exhibits unique patterns with early dramatic growth followed by moderation, reflecting both technological improvements and market saturation. The production efficiency analysis reveals that natural gas operations show the highest efficiency (3.2 units of production per $MtCO_2e$), while coal operations show the lowest (0.4 units of production per $MtCO_2e$).

The ownership structure and industry organization analysis reveals a complex mix of state-owned and investor-owned entities in the energy sector. Oil & Gas sectors demonstrate the most diverse ownership structure, with 48 investor-owned and 33 state-owned entities in Oil & NGL, and 47 investor-owned and 33 state-owned entities in Natural Gas. Coal sectors show more fragmented participation, with an average of 12 entities per coal type, while cement production remains highly concentrated among just 6 controlling entities. This finding supports research by Meckling et al. (2017) on the political economy of energy transitions, which highlights the different roles of state and private actors in energy production. The analysis reveals that state-owned enterprises dominate in regions with significant natural resource endowments (Middle East, Russia & CIS), while investor-owned companies are more prevalent in developed markets (North America, Europe). This distribution has important implications for climate policy, as different ownership structures may require different policy approaches to achieve emission reductions.

Statistical distribution patterns show extreme right skewness across all commodities, with the vast majority of emission records having values very close to zero. The analysis reveals that 75% of all emission records fall below 90 $MtCO_2e$, while the top 5% of records account for 65% of total emissions. High-value outliers significantly influence mean values, with the mean emission value (450 $MtCO_2e$) being 12 times higher than the median (37 $MtCO_2e$). This finding has important implications for climate policy and corporate responsibility, suggesting that targeted interventions could have significant impact given the concentration of emissions among a relatively small number of entities.

These findings have important implications for climate policy and corporate responsibility. The concentration of emissions among a relatively small number of entities suggests that targeted policy interventions could have significant impact. Regional disparities highlight the need for differentiated approaches to emission reduction, while historical patterns demonstrate the resilience of emission growth despite various interventions. The varying efficiency patterns across commodities suggest opportunities for targeted improvements, and the ownership structure analysis reveals the complex interplay between state control and private investment in global energy production. The analysis suggests that effective climate policy should consider:

1. Differentiated approaches for different ownership structures

2. Region-specific strategies that account for local industrial patterns

3. Targeted efficiency improvements in high-emission sectors

4. Mechanisms to address the concentration of emissions among major producers

5. Policies that consider both historical responsibility and current emission patterns

REFERENCES

- Friedlingstein, P., O'Sullivan, M., Jones, M. W., et al. (2023). Global Carbon Budget 2023. Earth System Science Data, 15(12), 5301-5369. [https://doi.org/10.5194/essd-15-5301-2023]

- IEA. (2023). CO2 Emissions in 2022. International Energy Agency. [https://www.iea.org/reports/co2-emissions-in-2022]

- IEA. (2023). World Energy Outlook 2023. International Energy Agency. [https://www.iea.org/reports/world-energy-outlook-2023]

- Le Quéré, C., Jackson, R. B., Jones, M. W., et al. (2021). Temporary reduction in daily global CO2 emissions during the COVID-19 forced confinement. Nature Climate Change, 10(7), 647-653. [https://doi.org/10.1038/s41558-020-0797-x]

- UNEP. (2023). Emissions Gap Report 2023. United Nations Environment Programme. [https://www.unep.org/resources/emissions-gap-report-2023]

- World Bank Group. (2023). State and Trends of Carbon Pricing 2023. [https://www.worldbank.org/en/programs/pricing-carbon]

Friedlingstein, P., et al. (2023). Global Carbon Budget 2023. Earth System Science Data, 15, 5301–5369.

Griffin, P. (2017). The Carbon Majors Database: CDP Carbon Majors Report 2017. CDP Worldwide.

Minx, J. C., et al. (2021). A comprehensive dataset for global, regional and national greenhouse gas emissions by sector 1970-2019. Earth System Science Data, 13, 5213–5252.

Meckling, J., Kelsey, N., Biber, E., & Zysman, J. (2017). The political economy of energy transitions. Nature Energy, 2(4), 17092. https://www.researchgate.net/publication/387119178_The_Political_Economy_of_Transitions

Boden, T. A., Marland, G., & Andres, R. J. (2017). Global, Regional, and National Fossil-Fuel CO2 Emissions. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy. https://doi.org/10.3334/CDIAC/00001_V2017

Gales, B., Kander, A., Malanima, P., & Rubio, M. (2007). North versus South: Energy transition and energy intensity in Europe over 200 years. European Review of Economic History, 11(2), 219-253. https://doi.org/10.1017/S1361491607001967

Höök, M., & Tang, X. (2013). Depletion of fossil fuels and anthropogenic climate change—A review. Energy Policy, 52, 797-809. https://doi.org/10.1016/j.enpol.2012.10.046

Smil, V. (2017). Energy and Civilization: A History. MIT Press. https://doi.org/10.7551/mitpress/10752.001.0001

Victor, D. G., Geels, F. W., & Sharpe, S. (2019). Accelerating the Low Carbon Transition: The Case for Stronger, More Targeted and Coordinated International Action. Brookings Institution. https://www.brookings.edu/research/accelerating-the-low-carbon-transition/