GDPvsGHG-NZ-R

Data Preparation

````
#### Installing required libraries and preparing the environment

```{r}

options(repos = "https://cran.rstudio.com/")
install.packages("tidyverse")
install.packages("dyplr")
install.packages("ggplot2")
install.packages("mice")
install.packages("knitr")
install.packages("corrplot")
install.packages("reshape2")
install.packages("janitor")
install.packages("imputeTS")
install.packages("skimr")
library(skimr)
library(imputeTS)
library(janitor)
library(tidyverse)
library(ggplot2)
library(mice)
library(knitr)
library(corrplot)
library(reshape2)
library(forecast)
library(tseries)
```
````

```
#set working directory
setwd("~/Documents/Data Intelligence")
```

````
#### Importing the datasets

```{r}

gdp ← read_csv("Total-GDP-NZ.csv", show_col_types = FALSE)
total_co2 ← read_csv("Total-CO2.csv", show_col_types = FALSE)
ann_co2 ← read_csv("Ann-CO2.csv", show_col_types = FALSE)
ann_ghg ← read_csv("Ann-GHG-NZ.csv", show_col_types = FALSE)
co2_sector ← read_csv("CO2-sector.csv", show_col_types = FALSE)
gdp_pc ← read_csv("GDP-PC-NZ.csv", show_col_types = FALSE)
ghg_gas ← read_csv("GHG-Gas.csv", show_col_types = FALSE)
ghg_sector ← read_csv("GHG-sector.csv", show_col_types = FALSE)
ghg_pc ← read_csv("GHG_per_capita.csv", show_col_types = FALSE)

```
````

````
#### Descriptive Statistics of all the datasets

```{r}
#Annual GDP

head(gdp)
tail(gdp)
summary(gdp)

#Cumulative annual CO2

head(total_co2)
tail(total_co2)
summary(total_co2)

#Annual CO2

head(ann_co2)
tail(ann_co2)
summary(ann_co2)
````
````

```
#Annual GHG

head(ann_ghg)
tail(ann_ghg)
summary(ann_ghg)

#Cumulative CO2 emissions from multiple sources NZ

head(co2_sector)
tail(co2_sector)
summary(co2_sector)

#GDP per capital NZ

head(gdp_pc)
tail(gdp_pc)
summary(gdp_pc$`GDP per capita`)

#GHG emissions from various gases

head(ghg_gas)
tail(ghg_gas)
summary(ghg_gas)

#GHG emissions from various sectors in NZ
head(ghg_sector)
tail(ghg_sector)
summary(ghg_sector)
```

#filtering countries by 'New Zealand'

````
```{r}
#GHG per capita

ghg_pc <- ghg_pc %>%
  filter(Entity == "New Zealand")
````
````

````
head(ghg_pc)
tail(ghg_pc)
summary(ghg_pc$`Per-capita greenhouse gas emissions in CO₂ equivalents`)

```
````

````
#### Data Preparation for Analysis before joining

#Since most of the data is available from 1990 to 2022, I filtered all the datase

```{r}

ann_co2 <- ann_co2 %>%
  filter(ann_co2$Year >= 1990)

```


```{r}
ann_ghg <- ann_ghg %>%
  filter(Year >= 1990)
```


```{r}
ghg_gas <- ghg_gas %>%
  filter(Year >= 1990)
```


```{r}
total_co2 <- total_co2 %>%
  filter(Year >= 1990)
```
````

````
#### Checking for any missing values in all the datasets

```{r}
sum(is.na(ann_co2))
sum(is.na(ann_ghg))
````

```
sum(is.na(co2_sector))
sum(is.na(gdp))
sum(is.na(gdp_pc))
sum(is.na(ghg_gas))
sum(is.na(ghg_sector))
sum(is.na(total_co2))
```
```

## *Analysis 1 - Descriptive statistics, correlation, and regression analysis of Annual GHG, Methane, Nitrous Oxide, and CO2 emissions in NZ.*

Greenhouse gases are majorly driven by human activities (Ritchie et al., 2023). Carbon dioxide, methane and nitrous oxide are produced by burning fossil fuels and producing goods like steel, cement, plastic and food. New Zealand is one of the highest producers of GHG per capita (Kliejunas et al., 2023).

The first objective is to explore which gas amongst CO2, methane and nitrous oxide contributes significantly to the overall GHG in New Zealand.

````
Merging the Annual GHG and Sector Wise GHG dataset for predictive ar

```{r}

#Preparing the data
ghg1 <- left_join(ghg_gas, ann_ghg, by = "Year") #combining nitroux oxide, me
ghg1 <- clean_names(ghg1)

ghg1$Annual_GHG <- ghg1$annual_greenhouse_gas_emissions_in_co2_equiva
ghg1$annual_greenhouse_gas_emissions_in_co2_equivalents <- NULL

ghg1 <- ghg1 %>%
  select(year, Annual_GHG, everything()) #rearranging columns by year, Annua
````

```
colnames(ghg1) <- c("Year", "Annual_GHG", "Nitrous_Oxide", "Methane", "CO2

#custom summary table
summary_table <- data.frame(
  Variable = colnames(ghg1)[-1],  # Exclude 'Year'
  N = nrow(ghg1),
  Mean = sapply(ghg1[-1], mean, na.rm = TRUE),
  Median = sapply(ghg1[-1], median, na.rm = TRUE),
  SD = sapply(ghg1[-1], sd, na.rm = TRUE),
  Min = sapply(ghg1[-1], min, na.rm = TRUE),
  Max = sapply(ghg1[-1], max, na.rm = TRUE)
)

# Print the summary
print(summary_table)
```

	Variable <chr>	N <int>	Mean <dbl>	Median <dbl>	SD <dbl>	Min <dbl>	Max <dbl>
Annual_GHG	Annual_GHG	33	65856343	65421948	8199717.1	54412040	93721336
Nitrous_Oxide	Nitrous_Oxide	33	13764474	13727416	284867.8	13245013	14379261
Methane	Methane	33	36954772	36972900	859970.6	35421950	38433660
CO2	CO2	33	15137097	14738757	8803954.8	2113234	43169410

4 rows

Summary of Annual GHG, nitrous oxide, methane and CO2 emissions in NZ

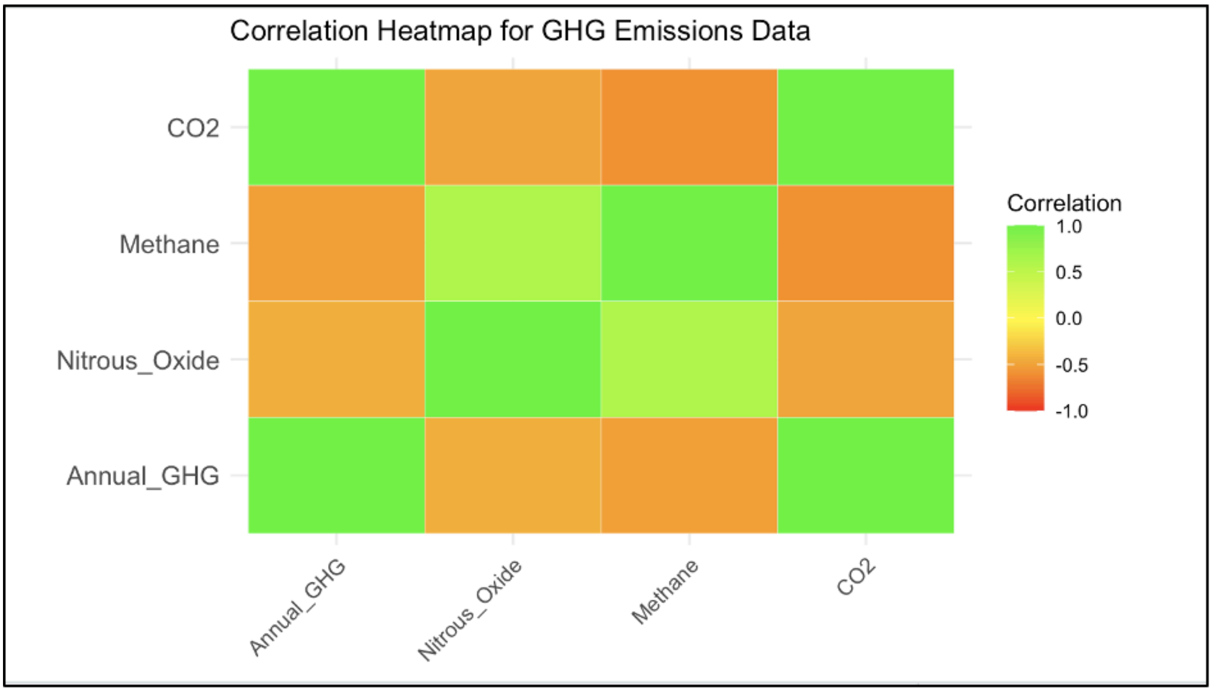
Key findings:

- The dataset contains no missing values.
- The average GHG emissions stood at 65.8 million CO2 equivalent from 1990 to 2022.
- The mean and median are close, suggesting a symmetrical distribution.
- The Standard Deviation of Annual GHG emissions is 8199717.1, slightly higher than the mean 65756343, suggesting the data points are more dispersed.

Further, a correlation analysis was conducted to see which gases highly correlate with Annual GHG emissions in New Zealand.

```
#### Correlation Analysis to check which gas highly correlates with Annual G

```{r}
colnames(ghg1) #checking column names to make sure all the columns are ali

#correlation matrix to check which gases have high correlation with Annual GI

ghg_corr <- ghg1 %>%
  select(Annual_GHG, Nitrous_Oxide,
         Methane, CO2)

correlation_matrix <- cor(ghg_corr, use = "complete.obs")

kable(correlation_matrix, digits = 2, caption = "Correlation Matrix for GHG Emi

```
```

Correlation Matrix for GHG Emissions Data

	Annual_GHG	Nitrous_Oxide	Methane	CO2
Annual_GHG	1.00	-0.42	-0.51	0.99
Nitrous_Oxide	-0.42	1.00	0.62	-0.48
Methane	-0.51	0.62	1.00	-0.59
CO2	0.99	-0.48	-0.59	1.00

Correlation matrix of Annual GHG, Nitrous Oxide, Methane and CO2

```
#correlation heatmap of Annual GHG emissions and other gases
```

```
correlation_data <- melt(correlation_matrix)

ggplot(data = correlation_data, aes(x = Var1, y = Var2, fill = value)) +
  geom_tile(color = "white") +
  scale_fill_gradient2(low = "red", high = "green", mid = "yellow",
                       midpoint = 0, limit = c(-1, 1), space = "Lab",
                       name = "Correlation") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, vjust = 1, size = 10, hjust = 1),
        axis.text.y = element_text(size = 12)) +
  labs(title = "Correlation Heatmap for GHG Emissions Data", x = "", y = "")
```



Correlation heatmap of Annual GHG, Nitrous Oxide, Methane and CO2

Key findings

- Annual GHG and CO2 emissions have a strong positive correlation of 0.99, suggesting that if one increases, the other increases and vice versa.

- Annual GHG emissions have a negative correlation (-0.51) with methane emissions, suggesting the relationship between them is not strong.

Next, multiple regression analysis explores how other gases contribute to overall GHG emissions in NZ.

Here, Annual GHG is the dependent variable, and the rest of the gases are independent variables.

```
```{r}
#regression analysis to check how other gases influence Annual GHG in NZ

colnames(ghg1)
ghg_reg <- lm(Annual_GHG ~  + Annual_GHG + Nitrous_Oxide + Methane + CO
summary(ghg_reg)

kable(summary(ghg_reg)$coefficients, caption = "Regression Coefficients")

#extracting values adjusted R, multiple R and f-statistics from the model
adjusted_r_squared <- summary(ghg_reg)$adj.r.squared
multiple_r <- summary(ghg_reg)$r.squared
f_statistic <- summary(ghg_reg)$fstatistic[1]

cat("Adjusted R-squared:", adjusted_r_squared, "\n")
cat("Multiple R:", multiple_r, "\n")
cat("F-statistic:", f_statistic, "\n")



#plotting regression

ghg1$predicted_Annual_GHG <- predict(ghg_reg, newdata = ghg1)

ggplot(ghg1, aes(x = Annual_GHG, y = predicted_Annual_GHG)) +
  geom_point() +
  geom_abline(intercept = 0, slope = 1, color = "red") +
  labs(title = "Actual vs. Predicted Annual GHG",
       x = "Actual Annual GHG",
```
```

```
    y = "Predicted Annual GHG")
```

```
lm(formula = Annual_GHG ~ +Annual_GHG + Nitrous_Oxide + Methane +
    CO2, data = ghg1)

Residuals:
    Min      1Q  Median      3Q     Max
-3.1272 -1.2077  0.0434  0.8718  6.0536

Coefficients:
               Estimate Std. Error   t value Pr(>|t|)
(Intercept)   4.405e+01  2.299e+01 1.916e+00   0.0653 .
Nitrous_Oxide 1.000e+00  1.615e-06 6.194e+05   <2e-16 ***
Methane       1.000e+00  5.819e-07 1.719e+06   <2e-16 ***
CO2           1.000e+00  5.104e-08 1.959e+07   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.012 on 29 degrees of freedom
Multiple R-squared:      1,     Adjusted R-squared:      1
F-statistic: 1.771e+14 on 3 and 29 DF,  p-value: < 2.2e-16

Adjusted R-squared: 1
Multiple R: 1
F-statistic: 1.771243e+14
```

Regression Analysis of Annual GHG, Nitrous Oxide, Methane and CO2

Key findings:

- The adjusted R square value is 1, suggesting that 100% variation in Annual GHG emissions can be explained by other variables.
- The P-value of all the independent variables is **0.000**, less than 5%, suggesting the other gases are statistically significant to Annual GHG emissions.
- The p-value of the f-statistic is less than 5%, suggesting the model is the best fit.

- Each unit increase in other independent variables increases Annual GHG by **1.000**

Analysis 1: Conclusion and limitations

- Amongst all three gases, CO2 emissions are highly correlate with annual GHG in New Zealand.
- CO2, methane, and nitrous oxide emissions are statistically significant in annual GHG emissions.

CO2 gases majorly contribute to GHG emissions in NZ. Government can define policies around CO2 emission reduction to reduce the overall GHG emissions. While CO2, Methane, and Nitrous Oxide emissions are statistically significant to Annual GHG in New Zealand, it does not consider other possible gases that may contribute to GHG emissions in New Zealand.

Analysis 2 - Descriptive statistics, correlation, and regression analysis of GDP per capita, Annual GHG, Annual CO2 and GHG per capita

Climate change negatively impacts the economy (Economics Observatory, 2024). Moreover, for New Zealand businesses, GHG emissions and climate change remain top problems (EECA, 2022). GHG and CO2 emissions majorly drive Climate Change (United Nations, n.d.). Therefore, it is worth analyzing if they impact GDP per capita in NZ.

This analysis compares the Annual GHG, CO2, GHG per capita and GDP per capita datasets.

```
#Analyzing GDP, Annual GHG, and Annual CO2 emissions to see which vari
able has a high correlation with GDP in New Zealand
```

```
#merging gdp, ghg per capita, annual ghg and annual CO2 datasets

gdp1 <- left_join(gdp_pc, ann_ghg, by = "Year")
gdp1

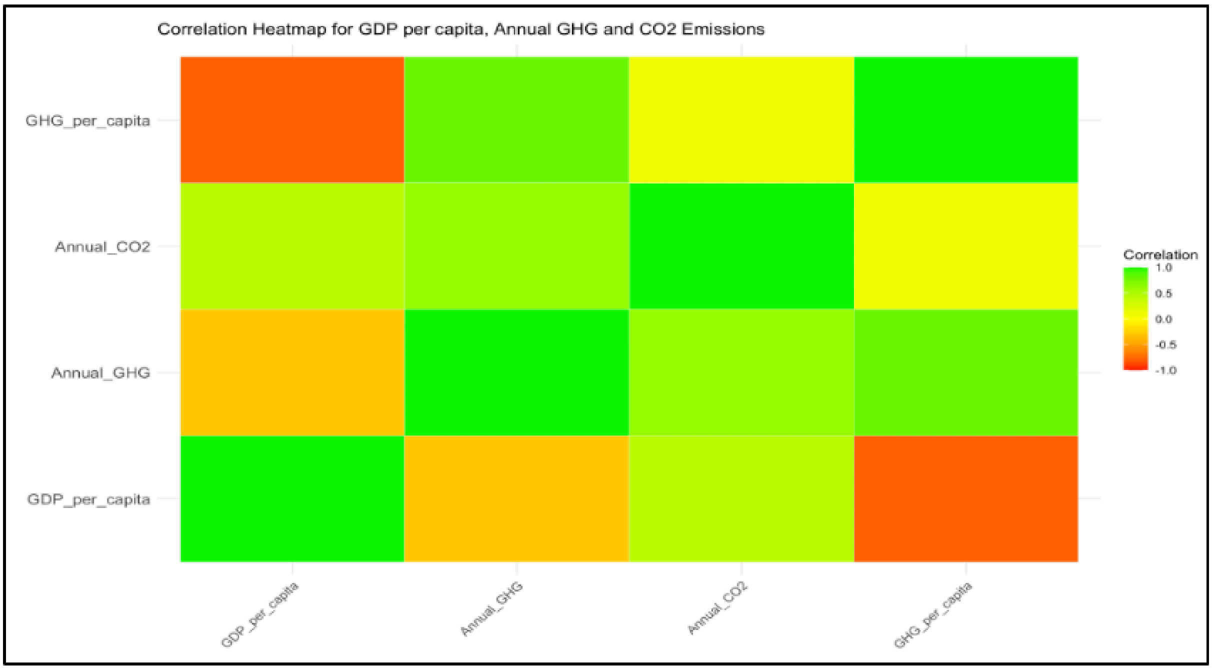
colnames(gdp1)
gdp1 <- gdp1 %>%
  select(Year, `GDP per capita`, everything())

gdp1 <- left_join(gdp1, ann_co2, by = "Year")
gdp1

#filtering ghg per capita data from the year 1990 to 2022
ghg_pc <- ghg_pc %>%
  filter(Year >= 1990)

gdp1 <- left_join(gdp1, ghg_pc, by = "Year")
gdp1$Entity <- NULL
gdp1$Code <- NULL
gdp1$predicted_GDP_per_capita <- NULL


colnames(gdp1) <- c("Year", "GDP_per_capita", "Annual_GHG", "Annual_CO2", "GHG_per_capita")

print(gdp1)
```

````
```{r}
#Descriptive statistics

summary(gdp1)

summary_gdp1 <- data.frame(
  Variable = colnames(gdp1)[-1],  # Exclude 'Year'
  N = nrow(gdp1),
  Mean = sapply(gdp1[-1], mean, na.rm = TRUE),
  Median = sapply(gdp1[-1], median, na.rm = TRUE),
  SD = sapply(gdp1[-1], sd, na.rm = TRUE),
````
````

```
  Min = sapply(gdp1[-1], min, na.rm = TRUE),
  Max = sapply(gdp1[-1], max, na.rm = TRUE)
)

print(summary_gdp1)
```
```

```
      Year       GDP_per_capita    Annual_GHG         Annual_CO2        GHG_per_capita
 Min.   :1990   Min.   :26383   Min.   :54412040   Min.   : 7019598   Min.   :12.62
 1st Qu.:1998   1st Qu.:30539   1st Qu.:60409556   1st Qu.:21422318   1st Qu.:13.55
 Median :2006   Median :37429   Median :65421948   Median :28498340   Median :14.41
 Mean   :2006   Mean   :36033   Mean   :65856343   Mean   :27779043   Mean   :15.95
 3rd Qu.:2014   3rd Qu.:40351   3rd Qu.:69211896   3rd Qu.:34934280   3rd Qu.:18.18
 Max.   :2022   Max.   :45185   Max.   :93721336   Max.   :47972070   Max.   :24.45
```

Descriptive summary

A correlation analysis explores how GDP per capita relates to Annual GHG emissions, Annual CO2 emissions and GHG per capita emissions.

```
#### Correlation Analysis of GDP per capita, Annual GHG and CO2 emissions

```{r}
#correlatoion analysis for GDP and Annual GHG Emissions in NZ

colnames(gdp1)
gdp_corr <- gdp1 %>%
  select(GDP_per_capita,Annual_GHG, Annual_CO2, GHG_per_capita)

gdp_corr_matrix <- cor(gdp_corr, use = "complete.obs")

kable(gdp_corr_matrix, digits = 2, caption = "Correlation Matrix for GDP per ca


#Correlation heatmap

correlation_data1 <- melt(gdp_corr_matrix)
```

```
ggplot(data = correlation_data1, aes(x = Var1, y = Var2, fill = value)) +
  geom_tile(color = "white") +
  scale_fill_gradient2(low = "red", high = "green", mid = "yellow",
                       midpoint = 0, limit = c(-1, 1), space = "Lab",
                       name = "Correlation") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, vjust = 1, size = 10, hjust = 1),
        axis.text.y = element_text(size = 12)) +
  labs(title = "Correlation Heatmap for GDP per capita, Annual GHG and CO2 E
```

Correlation Matrix for GDP per capita, GHG, CO2 emissions and GHG per capita Data

	GDP_per_capita	Annual_GHG	Annual_CO2	GHG_per_capita
GDP_per_capita	1.00	-0.28	0.45	-0.79
Annual_GHG	-0.28	1.00	0.61	0.80
Annual_CO2	0.45	0.61	1.00	0.10
GHG_per_capita	-0.79	0.80	0.10	1.00

Correlation Analysis between GDP per capita, Annual GHG, CO2 emissions and GHG per capita



Correlation Analysis between GDP per capita, Annual GHG, CO2 emissions and GHG per capita

Key findings

- GDP per capita (0.45) positively correlates with Annual CO2 emissions, suggesting one factor influences the other.
- GDP per capita has a strong negative correlation with GHG per capita (-0.79), suggesting an increase in GDP per capita tends to decrease GHG per capita.
- However, GDP per capita negatively correlates (-0.28) with Annual GHG emissions.

Furthermore, a regression analysis is conducted to explore if Annual GHG and CO2 emissions impact GDP per capita in New Zealand.

```
#### regression analysis for gdp, ghg, c02

```{r}

#regression analysis for gdp, ghg, c02 to show how GHG and Annual CO2 imp

colnames(gdp1)

gdp_co_ghg_reg <- lm(GDP_per_capita ~ Annual_GHG +
                     Annual_CO2 + GHG_per_capita, data = gdp1)
summary(gdp_co_ghg_reg)

kable(summary(gdp_co_ghg_reg)$coefficients, caption = "Regression Coeffic

#extracting values adjusted R, multiple R and f-statistics from the model
adjusted_r_squared <- summary(gdp_co_ghg_reg)$adj.r.squared
multiple_r <- summary(gdp_co_ghg_reg)$r.squared

cat("Adjusted R-squared:", adjusted_r_squared, "\n")
cat("Multiple R:", multiple_r, "\n")
```
```

````
```
````

Regression Coefficients				
	Estimate	**Std. Error**	**t value**	**Pr(>|t|)**
(Intercept)	37340.0435806	2356.7238580	15.8440470	0.0000000
Annual_GHG	0.0006152	0.0000919	6.6940416	0.0000002
Annual_CO2	0.0000250	0.0000470	0.5306807	0.5996809
GHG_per_capita	-2665.2791611	189.6716890	-14.0520664	0.0000000

Regression analysis of GDP per capita, Annual GHG and CO2 emissions - 1

```
lm(formula = GDP_per_capita ~ Annual_GHG + Annual_CO2 + GHG_per_capita,
    data = gdp1)

Residuals:
    Min      1Q  Median      3Q     Max
-2272.2  -822.7   285.7   653.6  2873.2

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)     3.734e+04  2.357e+03  15.844 8.12e-16 ***
Annual_GHG      6.152e-04  9.190e-05   6.694 2.42e-07 ***
Annual_CO2      2.496e-05  4.703e-05   0.531      0.6
GHG_per_capita -2.665e+03  1.897e+02 -14.052 1.79e-14 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1162 on 29 degrees of freedom
Multiple R-squared:   0.96,	Adjusted R-squared:  0.9559
F-statistic: 232.1 on 3 and 29 DF,  p-value: < 2.2e-16

Adjusted R-squared: 0.9558778
Multiple R: 0.9600143
```

Regression analysis of GDP per capita, Annual GHG and CO2 emissions - 2

Key findings
```

- Adjusted R-square (0.9559) suggests that 95% of the variation in GDP per capita can be explained by Annual GHG, CO_2, and GHG per capita emissions in NZ.
- F-statistic has a p-value of 2.2e-16, which is less than the significance level of 5%, suggesting the model is the best fit.
- The Annual GHG and GHG per capita are statistically significant to GDP per capita at 5% significance level.
- Each unit increase in Annual GHG reduces GDP by approximately 0.000597 units, suggesting that although the magnitude is small, an increase in GHG emissions potentially reduces GDP per capita.
- The coefficient of GHG per capita is -2665.27, suggesting that each 1-ton increase in GHG per capita tends to decrease GDP per capita by $2665.27.
- Contrastingly, for each unit increase in Annual CO_2, GDP per capita increases by approximately 0.0005665 units. This suggests that higher CO_2 emissions are positively associated with increased GDP per capita.

Analysis 2: Conclusion and limitations

The annual GHG emissions tend to reduce GDP per capita in NZ. Although GDP per capita has a high positive correlation with Annual GHG emissions, it has a negative correlation with Annual CO_2 emissions in New Zealand. Moreover, 95% variation in GDP per capita (dependent variable) can be explained by Annual GHG, Annual CO_2 and GHG per capita emissions (independent variables).

While all three variables are statistically significant to GDP per capita in New Zealand, they can only explain 95% of the variation in GDP per capita. Adding more vital factors may provide a more comprehensive overview of how each affects GDP per capita.

Analysis 3 - Descriptive statistics, correlation, and regression analysis of GDP per capita, Annual GHG and GHG emissions from various

Since the Analysis 2 results suggest that Annual GHG has a significant relation with GDP per capita, the report further conducts a correlation and regression analysis to check GHGs from which sectors in New Zealand highly associate with GDP per capita.

```
#### Exploring GDP and Sector wise GHG in NZ to see which sector significar

```{r}

#merging the GDP and Sector wise GHG emission data

gdp_sec_ghg <- left_join(gdp_pc, ghg_sector, by = "Year")
gdp_sec_ghg

colnames(ann_ghg) <- c("Year", "Annual_GHG")

gdp_sec_ghg <- left_join(ann_ghg, gdp_sec_ghg, by = "Year")
gdp_sec_ghg

gdp_sec_ghg$Country <- NULL  #removed country column as it is not needed

```
```

```
```{r}
#rearranging columns in gdp_sec_ghg

gdp_sec_ghg <- gdp_sec_ghg %>%
  select(Year, `GDP per capita`, everything())

```
```

```
#checking if there are any missing values

sum(is.na(gdp_sec_ghg))

```
```

```
```{r}

#handling missing values

colnames(gdp_sec_ghg)
gdp_sec_ghg <- clean_names(gdp_sec_ghg)

gdp_sec_ghg <- gdp_sec_ghg[1:31, ] #dropped year 2021 and 2022 as no impu
```
```

```
```{r}
#descriptive statistics



summary_gdp_sec_ghg <- data.frame(
  Variable = colnames(gdp_sec_ghg)[-1],  # Exclude 'Year'
  N = nrow(gdp_sec_ghg),
  Mean = sapply(gdp_sec_ghg[-1], mean, na.rm = TRUE),
  Median = sapply(gdp_sec_ghg[-1], median, na.rm = TRUE),
  SD = sapply(gdp_sec_ghg[-1], sd, na.rm = TRUE),
  Min = sapply(gdp_sec_ghg[-1], min, na.rm = TRUE),
  Max = sapply(gdp_sec_ghg[-1], max, na.rm = TRUE)
)
print(summary_gdp_sec_ghg)
```
```
```

Variable <chr>	N <int>	Mean <dbl>	Median <dbl>	SD <dbl>	Min <dbl>	Max <dbl>
gdp_per_capita	31	35479	37261.15	5234.066	26383.26	43272.57
annual_ghg	31	65800166	65324324.00	8458744.192	54412040.00	93721336.00
greenhouse_gas_emissions_from_agriculture	31	42295484	42260000.00	736029.160	40810000.00	43520000.00
greenhouse_gas_emissions_from_land_use_change_and_forestry	31	-19652903	-21520000.00	5570928.225	-24620000.00	-9290000.00
greenhouse_gas_emissions_from_waste	31	4126774	4080000.00	210782.461	3840000.00	4470000.00
greenhouse_gas_emissions_from_buildings	31	1494516	1480000.00	133462.572	1220000.00	1860000.00
greenhouse_gas_emissions_from_industry	31	1592581	1510000.00	507536.322	850000.00	2420000.00
greenhouse_gas_emissions_from_manufacturing_and_construction	31	6037419	6080000.00	562452.770	5080000.00	7160000.00
greenhouse_gas_emissions_from_transport	31	13007742	13890000.00	2101714.664	8910000.00	16590000.00
greenhouse_gas_emissions_from_electricity_and_heat	31	7790645	7470000.00	2085049.850	4390000.00	11840000.00
fugitive_emissions_of_greenhouse_gases_from_energy_production	31	1221613	1230000.00	117078.233	1000000.00	1420000.00
greenhouse_gas_emissions_from_other_fuel_combustion	31	1533871	1540000.00	191845.032	1100000.00	1810000.00
greenhouse_gas_emissions_from_bunker_fuels	31	3186774	3020000.00	747990.139	2160000.00	4980000.00
predicted_gdp_per_capita	31	35479	36994.68	5218.390	27004.07	43390.71

14 rows | 2-8 of 7 columns

Descriptive statistics

```
```{r}

#### Correlation analysis ####

 gdp_sec_ghg_corr ← gdp_sec_ghg %>%
  select(gdp_per_capita, annual_ghg, greenhouse_gas_emissions_from_agricu

gdp_sec_ghg_corr_matrix ← cor(gdp_sec_ghg_corr, use = "complete.obs")

kable(gdp_sec_ghg_corr_matrix, digits = 2, caption = "Correlation Matrix for G



### Correlation heatmap

```{r}

correlation_gdp_ghg_sec ← melt(gdp_sec_ghg_corr_matrix)

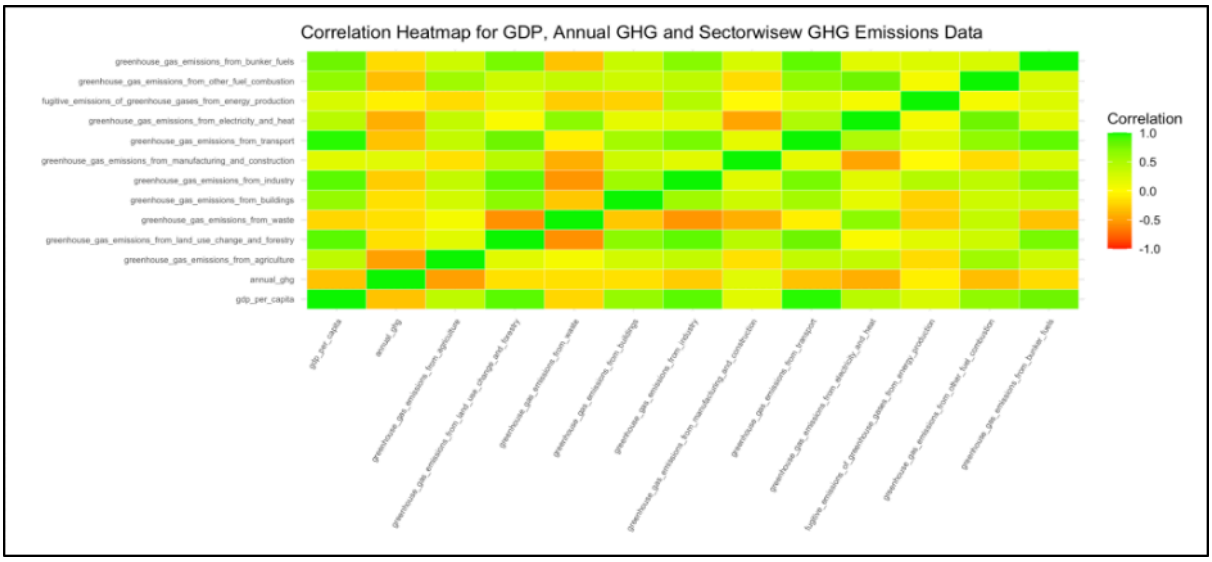
ggplot(data = correlation_gdp_ghg_sec, aes(x = Var1, y = Var2, fill = value)) +
  geom_tile(color = "white") +
  scale_fill_gradient2(low = "red", high = "green", mid = "yellow",
                      midpoint = 0, limit = c(-1, 1), space = "Lab",
                      name = "Correlation") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 60, vjust = 1, size = 6, hjust = 1),
      axis.text.y = element_text(size = 6)) +
```

~~~
  labs(title = "Correlation Heatmap for GDP, Annual GHG and Sectorwisew GH

```
~~~

```
                                                           gdp_per_capita  annual_ghg
gdp_per_capita                                                  1.0000000 -0.31797634
annual_ghg                                                     -0.3179763  1.00000000
greenhouse_gas_emissions_from_agriculture                       0.3991138 -0.49199281
greenhouse_gas_emissions_from_land_use_change_and_forestry      0.8637712 -0.13945116
greenhouse_gas_emissions_from_waste                            -0.2055321 -0.16504513
greenhouse_gas_emissions_from_buildings                         0.6164685 -0.13557907
greenhouse_gas_emissions_from_industry                          0.8592288 -0.26832868
greenhouse_gas_emissions_from_manufacturing_and_construction    0.1780590  0.19206598
greenhouse_gas_emissions_from_transport                         0.9717801 -0.31757007
greenhouse_gas_emissions_from_electricity_and_heat              0.4386638 -0.42780412
fugitive_emissions_of_greenhouse_gases_from_energy_production   0.2427508 -0.07440069
greenhouse_gas_emissions_from_other_fuel_combustion             0.6326006 -0.33910459
greenhouse_gas_emissions_from_bunker_fuels                      0.7887540 -0.18343857
```

Correlation analysis for GDP per capita, Annual GHG and GHG emissions from different sectors of NZ



Correlation matrix for GDP per capita, Annual GHG and GHG emissions from different sectors of NZ

Key findings

- Annual GHG emissions have a high mean (65800166).

- GDP per capita highly correlates with GHG emissions from the transport sector 0.9717801, suggesting the emissions from the transport sector increase GDP per capita in New Zealand.
- Annual GHG (-0.3180) negatively correlates with GDP per capita, implying an overall increase in GHG emissions tends to decrease GDP per capita.
- There is a positive correlation between GHG emissions from industry and GDP per capita (0.8592), suggesting that industrial activities contribute to GDP per capita alongside the creation of greenhouse emissions.
- GHH emissions from land use change and forestry (0.8637712) strongly correlate with GDP per capita, suggesting deforestation or land development is strongly linked to economic growth with a significant increase in GHG emissions.
- GHG emissions from waste (-0.2055321) negatively correlated with GDP per capita, indicating that waste emissions tend to decrease GDP per capita.

The regression analysis further explores the relationship between emissions from each sector and GDP per capita.

```
#Regression analysis for GDP, Annual GHG and Sectorwise emissions in NZ

colnames(gdp_sec_ghg)

gdp_ghg_sec_reg <- lm(gdp_per_capita ~ annual_ghg +
  greenhouse_gas_emissions_from_agriculture +
  greenhouse_gas_emissions_from_land_use_change_and_forestry +
  greenhouse_gas_emissions_from_waste +
  greenhouse_gas_emissions_from_buildings +
  greenhouse_gas_emissions_from_industry +
  greenhouse_gas_emissions_from_manufacturing_and_construction +
  greenhouse_gas_emissions_from_transport +
  greenhouse_gas_emissions_from_electricity_and_heat +
  fugitive_emissions_of_greenhouse_gases_from_energy_production +
  greenhouse_gas_emissions_from_other_fuel_combustion +
  greenhouse_gas_emissions_from_bunker_fuels,
```

```
  data = gdp_sec_ghg)

summary(gdp_ghg_sec_reg)

kable(summary(gdp_ghg_sec_reg)$coefficients, caption = "GDP, Annual GHG

#extracting values adjusted R, multiple R and f-statistics from the model
adjusted_r_squared <- summary(gdp_ghg_sec_reg)$adj.r.squared
multiple_r <- summary(gdp_ghg_sec_reg)$r.squared
f_statistic <- summary(gdp_ghg_sec_reg)$fstatistic[1]
df1 <- summary(gdp_ghg_sec_reg)$fstatistic[2]
df2 <- summary(gdp_ghg_sec_reg)$fstatistic[3]
f_p_value <- pf(f_statistic, df1, df2, lower.tail = FALSE)

cat("Adjusted R-squared:", adjusted_r_squared, "\n")
cat("Multiple R:", multiple_r, "\n")
cat("F-statistic:", f_statistic, "\n")
cat("F-statistic p-value:", f_p_value, "\n")

```
```

GDP, Annual GHG and Sectorwise GHG Regression Coefficients

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.8019e+04	1.312779e+04	1.3725848	0.1867433
annual_ghg	-1.3000e-06	1.600000e-05	-0.0842484	0.9337889
greenhouse_gas_emissions_from_agriculture	-1.8000e-05	3.123000e-04	-0.0575649	0.9547294
greenhouse_gas_emissions_from_land_use_change_and_forestry	8.9300e-05	7.490000e-05	1.1925498	0.2485301
greenhouse_gas_emissions_from_waste	-1.6232e-03	2.068800e-03	-0.7846359	0.4428718
greenhouse_gas_emissions_from_buildings	-7.4030e-04	1.601400e-03	-0.4622910	0.6494117
greenhouse_gas_emissions_from_industry	2.4355e-03	8.973000e-04	2.7143788	0.0142107
greenhouse_gas_emissions_from_manufacturing_and_construction	9.7700e-05	2.589000e-04	0.3771660	0.7104602
greenhouse_gas_emissions_from_transport	2.0529e-03	2.794000e-04	7.3465043	0.0000008
greenhouse_gas_emissions_from_electricity_and_heat	5.6500e-05	1.800000e-04	0.3138405	0.7572502
fugitive_emissions_of_greenhouse_gases_from_energy_production	-1.8390e-03	1.624500e-03	-1.1320551	0.2724675
greenhouse_gas_emissions_from_other_fuel_combustion	1.3638e-03	1.690300e-03	0.8068846	0.4302662
greenhouse_gas_emissions_from_bunker_fuels	-1.1259e-03	5.374000e-04	-2.0951171	0.0505717

Regression analysis of GDP per capita, Annual GHG and Sectorwise GHG emissions - 1
```

```
(Intercept)
annual_ghg
greenhouse_gas_emissions_from_agriculture
greenhouse_gas_emissions_from_land_use_change_and_forestry
greenhouse_gas_emissions_from_waste
greenhouse_gas_emissions_from_buildings
greenhouse_gas_emissions_from_industry                        *
greenhouse_gas_emissions_from_manufacturing_and_construction
greenhouse_gas_emissions_from_transport                       ***
greenhouse_gas_emissions_from_electricity_and_heat
fugitive_emissions_of_greenhouse_gases_from_energy_production
greenhouse_gas_emissions_from_other_fuel_combustion
greenhouse_gas_emissions_from_bunker_fuels                    .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 522.6 on 18 degrees of freedom
Multiple R-squared:  0.994,     Adjusted R-squared:   0.99
F-statistic: 249.3 on 12 and 18 DF,  p-value: < 2.2e-16

Adjusted R-squared: 0.9900316
Multiple R: 0.9940189
F-statistic: 249.2915
F-statistic p-value: 1.908761e-17
```

Regression analysis of GDP per capita, Annual GHG and Sector wise GHG emissions - 2

Key findings

- Adjusted R square value is 0.99, implying that independent variables Annual GHG and GHG emissions from other sectors in NZ can explain a 99% variation in GDP per capita.
- Multiple R is 0.994, showing a strong relationship between dependent and independent variables.
- The f-statistic p-value is 1.908761e-17, suggesting the model is the best fit.
- The P-value of GHG emissions from the transport sector (0.0000008) and industry (0.0142107) is less than the significance level of 5%, suggesting

transport and industry emissions are statistically significant to GDP per capita in NZ.

- Most independent variables demonstrate no statistical significance in GDP per capita, suggesting no firm evidence of whether the variables influence GDP per capita in NZ.

Analysis 3: Conclusion and limitations

The correlation analysis shows that GDP per capita positively correlates with emissions from transport and land use change and forestry. However, Regression analysis demonstrates that although the adjusted R square explains 99% of the variation in GDP per capita caused by Annual GHG emissions and GHG emissions from various sectors, only emissions from transport and industry are statistically significant. Therefore, more variables would be needed to refine the model further and study their impact on GDP per capita in NZ.

Analysis 4 - Descriptive statistics, correlation and regression analysis of GDP per capita, Annual CO2, and CO2 emissions from various sources in NZ.

Since Analysis 1 concluded that CO2 emissions are statistically significant to Annual GHG emissions in NZ, the paper further explores the relationship between Annual CO2 emissions and top CO2 emission sources and studies how they influence GDP per capita in NZ.

The data preparation steps are mentioned in the R file attached to this report.

````
#### Comparing GDP per capita, Annual CO2 emissions and sector

```{r}
````
````

```
#preparing the data

#joining the datasets and rearranging the columns

gdp_sec_co2 <- left_join(gdp_pc, co2_sector, by = "Year")
gdp_sec_co2

gdp_sec_co2 <- left_join(ann_co2, gdp_sec_co2, by = "Year")
gdp_sec_co2 <- clean_names(gdp_sec_co2)

```

```
```{r}
#renaming columns for convention

colnames(gdp_sec_co2) <- c("year", "annual_co2", "gdp_per_capita", "industr

gdp_sec_co2 <- gdp_sec_co2 %>%
  select(year, gdp_per_capita, everything())
```
```

```
```{r}

#checking if there are any missing values

sum(is.na(gdp_sec_co2))

```
```

```
```{r}
#descriptive statistics

summary_gdp_sec_co2 <- data.frame(
  Variable = colnames(gdp_sec_co2)[-1],  # Exclude 'Year'
  N = nrow(gdp_sec_co2),
```
```

````
  Mean = sapply(gdp_sec_co2[-1], mean, na.rm = TRUE),
  Median = sapply(gdp_sec_co2[-1], median, na.rm = TRUE),
  SD = sapply(gdp_sec_co2[-1], sd, na.rm = TRUE),
  Min = sapply(gdp_sec_co2[-1], min, na.rm = TRUE),
  Max = sapply(gdp_sec_co2[-1], max, na.rm = TRUE)
)
print(summary_gdp_sec_co2)
```
````

Variable <chr>	N <int>	Mean <dbl>	Median <dbl>	SD <dbl>	Min <dbl>	Max <dbl>
gdp_per_capita	33	36032.61	37428.85	5.531968e+03	26383.26	45185.31
annual_co2	33	27779042.85	28498340.00	9.635191e+06	7019598.00	47972070.00
industry	33	10756587.21	10818884.00	6.643783e+06	410901.00	21308300.00
cement	33	26081060.55	25925962.00	5.231340e+06	17663076.00	34196964.00
flaring	33	14566765.58	12480285.00	9.254172e+06	1994904.00	30344344.00
gas	33	209826785.03	211462370.00	8.001149e+07	77063010.00	338069500.00
oil	33	588639817.58	580474050.00	1.724005e+08	327052600.00	888133000.00
coal	33	543600172.73	544770750.00	5.849778e+07	453833500.00	637118900.00

Descriptive statistics

```
#Correlation Analysis for GDP per capita, Annual CO2 emissions and cumul
ative CO2 emissions from varios sources in NZ

colnames(gdp_sec_co2)

gdp_sec_co2 ← clean_names(gdp_sec_co2)

 gdp_sec_co2_corr ← gdp_sec_co2 %>%
  select(gdp_per_capita, annual_co2, industry, cement, flaring, gas, oil, coa
l)

gdp_sec_co2_corr_matrix ← cor(gdp_sec_co2_corr, use = "complete.obs")
print(gdp_sec_co2_corr_matrix)

kable(gdp_sec_co2_corr_matrix, digits = 2, caption = "Correlation Matrix fo
r GDP, Annnual CO2, and Cumulative CO2 emissions from multiple sectors i
n NZ")
```
```

```
#correlation analysis

correlation_gdp_ann_sec <- melt(gdp_sec_co2_corr_matrix)

ggplot(data = correlation_gdp_ann_sec, aes(x = Var1, y = Var2, fill = value))
+
  geom_tile(color = "white") +
  scale_fill_gradient2(low = "red", high = "green", mid = "yellow",
                       midpoint = 0, limit = c(-1, 1), space = "Lab",
                       name = "Correlation") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 60, vjust = 1, size = 6, hjust = 1),
        axis.text.y = element_text(size = 6)) +
  labs(title = "Correlation Heatmap for GDP, Annual CO2 and Cumulative CO
2 Emissions from various sources", x = "", y = "")
```

Correlation Matrix for GDP, Annnual CO2, and Cumulative CO2 emissions from multiple sectors in NZ

	gdp_per_capita	annual_co2	industry	cement	flaring	gas	oil	coal
gdp_per_capita	1.00	0.45	0.98	0.98	0.96	0.99	0.98	0.98
annual_co2	0.45	1.00	0.50	0.51	0.55	0.50	0.52	0.51
industry	0.98	0.50	1.00	1.00	0.99	1.00	1.00	1.00
cement	0.98	0.51	1.00	1.00	0.99	1.00	1.00	1.00
flaring	0.96	0.55	0.99	0.99	1.00	0.99	0.99	0.99
gas	0.99	0.50	1.00	1.00	0.99	1.00	1.00	1.00
oil	0.98	0.52	1.00	1.00	0.99	1.00	1.00	1.00
coal	0.98	0.51	1.00	1.00	0.99	1.00	1.00	1.00

Correlation analysis of GDP per capita, Annual CO2 and cumulative CO2 emissions from various sources in NZ

Correlation matrix of GDP per capita, Annual CO2 and cumulative CO2 emissions from various sources in NZ

Key findings

- GDP per capita highly correlates with cumulative CO2 emissions from various sectors (all near 1). This suggests a strong positive relationship that indicates that if GDP per capita increases, the CO2 emissions from these sources also tend to increase.
- Annual CO2 emissions (0.45) have a weak positive correlation with GDP per capita.

Next, a regression analysis is conducted to analyze how independent variables impact GDP per capita in NZ.

~~~~
#### Regression Analysis

```{r}
~~~~

```
colnames(gdp_sec_co2)

gdp_ann_sec_reg <- lm(gdp_per_capita ~ annual_co2 +
  industry + cement + flaring + gas + oil + coal,
  data = gdp_sec_co2)

summary(gdp_ann_sec_reg)

kable(summary(gdp_ann_sec_reg)$coefficients, caption = "GDP, Annual CO2

#extracting values adjusted R, multiple R and f-statistics from the model
adjusted_r_squared <- summary(gdp_ann_sec_reg)$adj.r.squared
multiple_r <- summary(gdp_ann_sec_reg)$r.squared

cat("Adjusted R-squared:", adjusted_r_squared, "\n")
cat("Multiple R:", multiple_r, "\n")

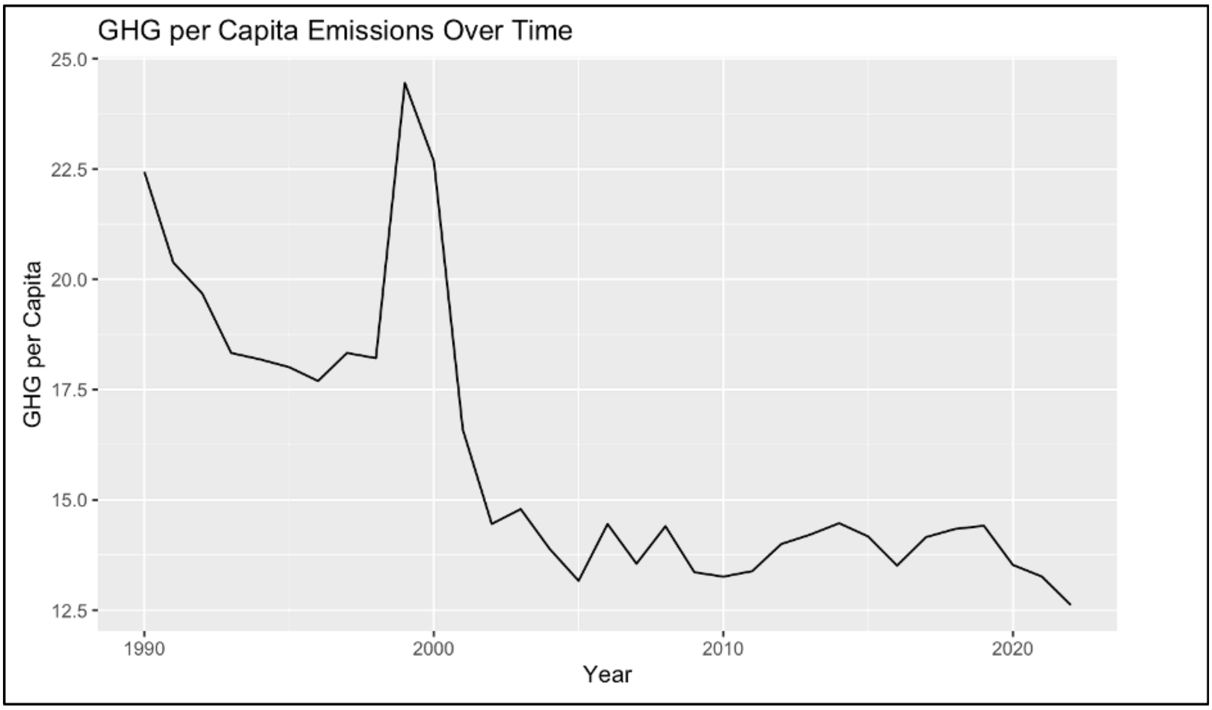
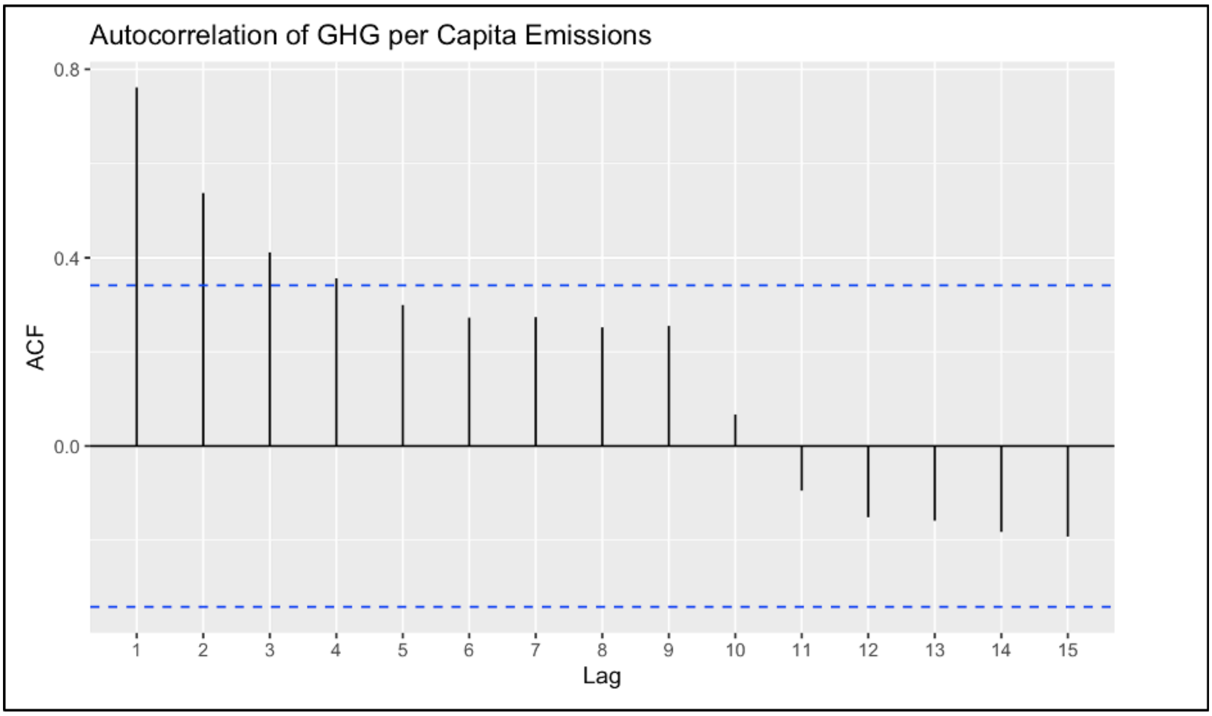
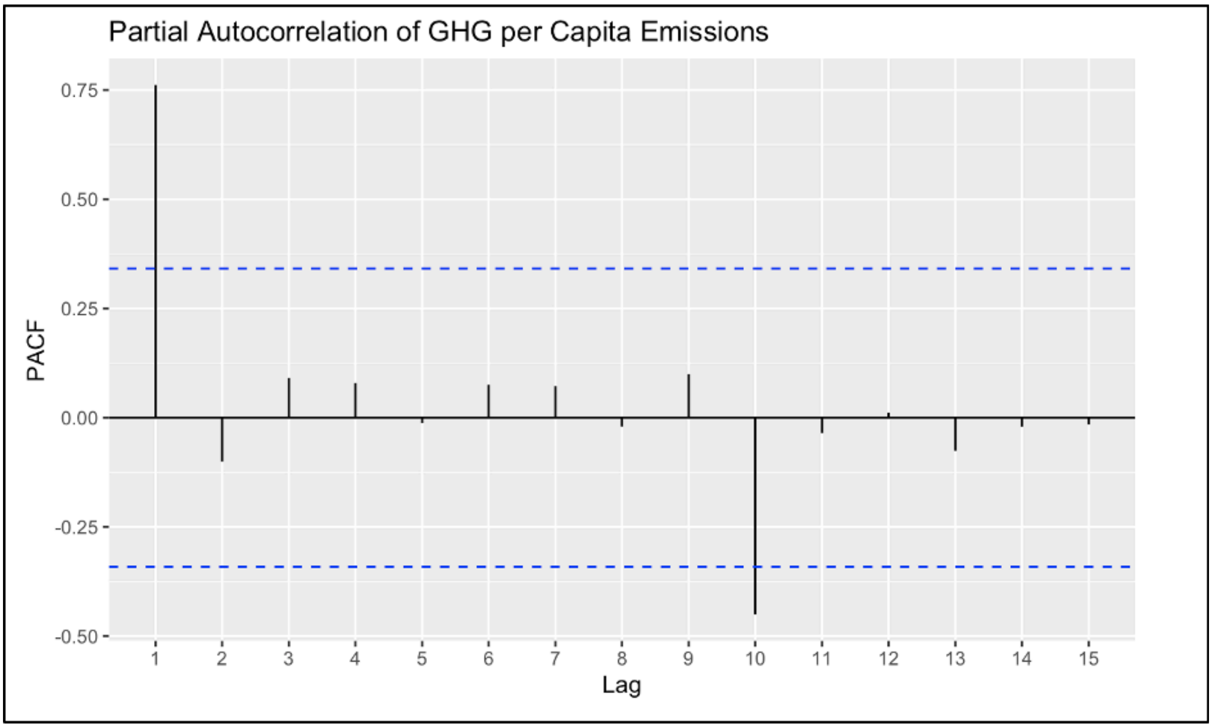
```

GDP, Annual CO2 and Cumulative CO2 Emissions from various sources

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	84970.6439346	3.131451e+04	2.7134592	0.0118819
annual_co2	-0.0000014	1.730000e-05	-0.0788679	0.9377656
industry	0.0035448	1.236000e-03	2.8679446	0.0082733
cement	0.0038452	2.009100e-03	1.9138745	0.0671512
flaring	-0.0016310	4.339000e-04	-3.7593575	0.0009167
gas	-0.0003775	1.580000e-04	-2.3886381	0.0247793
oil	0.0001469	5.120000e-05	2.8693698	0.0082454
coal	-0.0003142	1.019000e-04	-3.0843505	0.0049260

Regression analysis for GDP per capita, Annual CO2 and Cumulative CO2 emissions from various sources - 1

```
```{r}
```
```

```
gdp_sec_co2$predicted_gdp_per_capita <- predict(gdp_ann_sec_reg, newdata

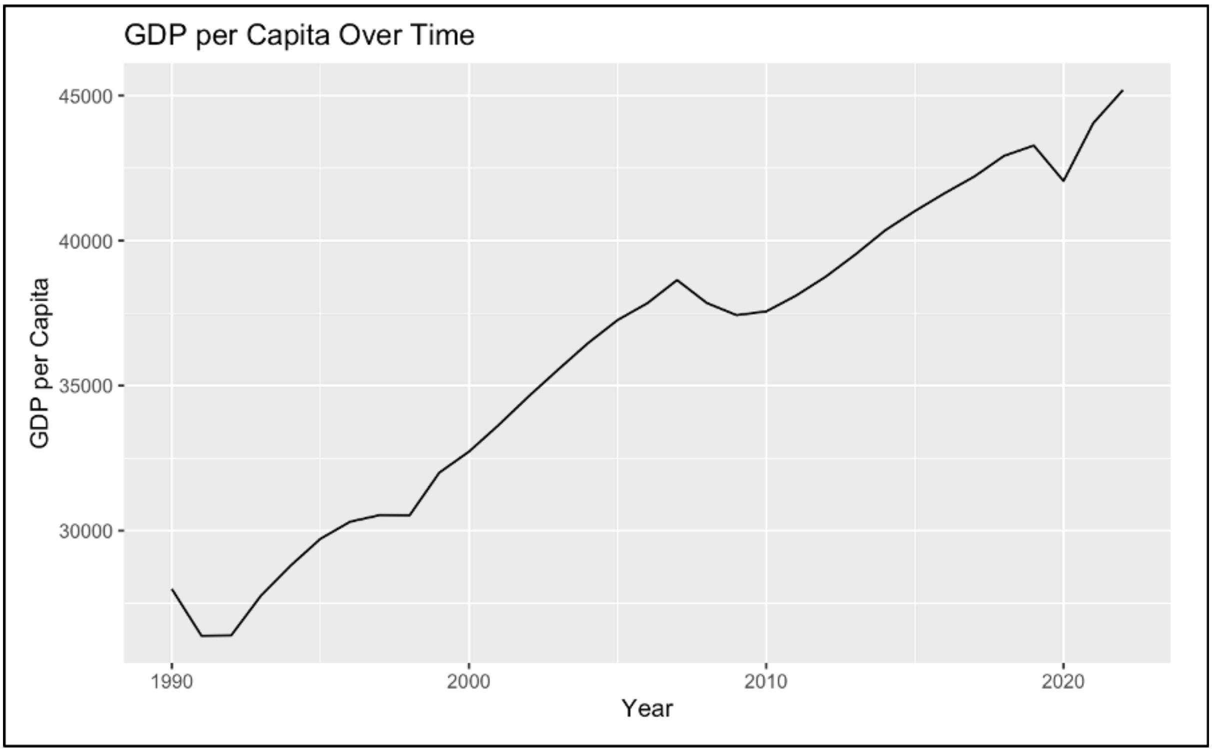
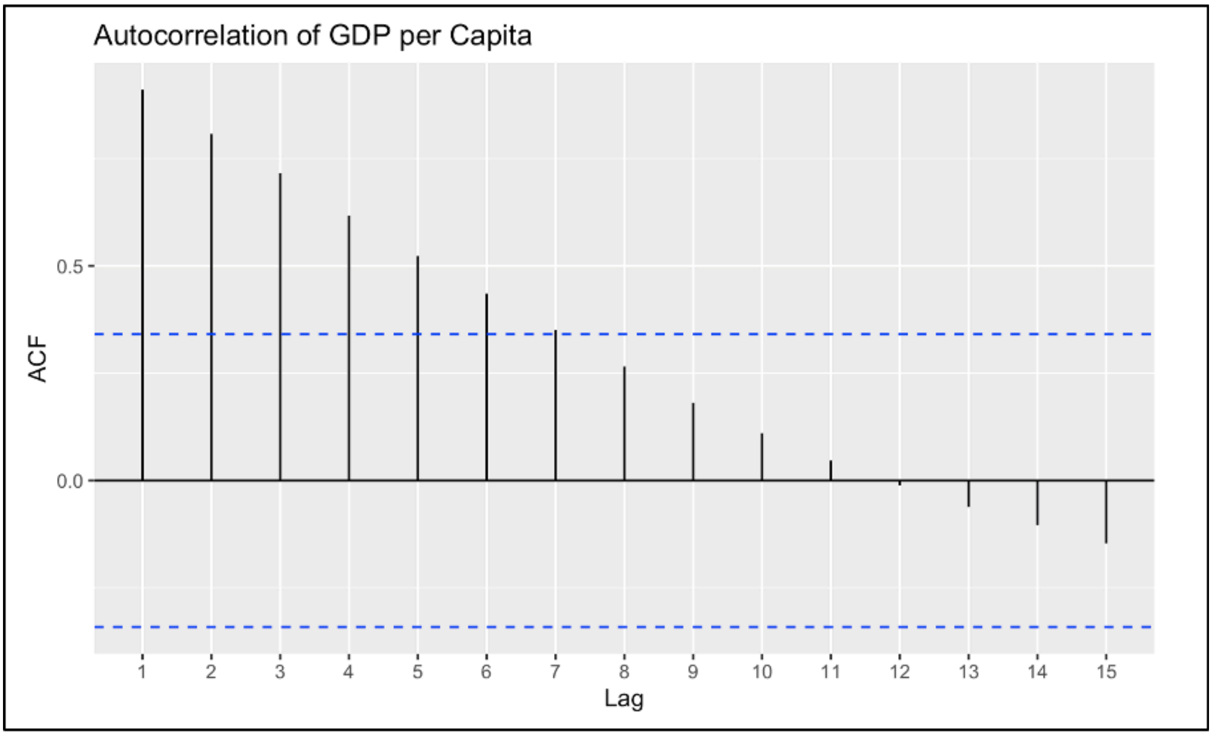
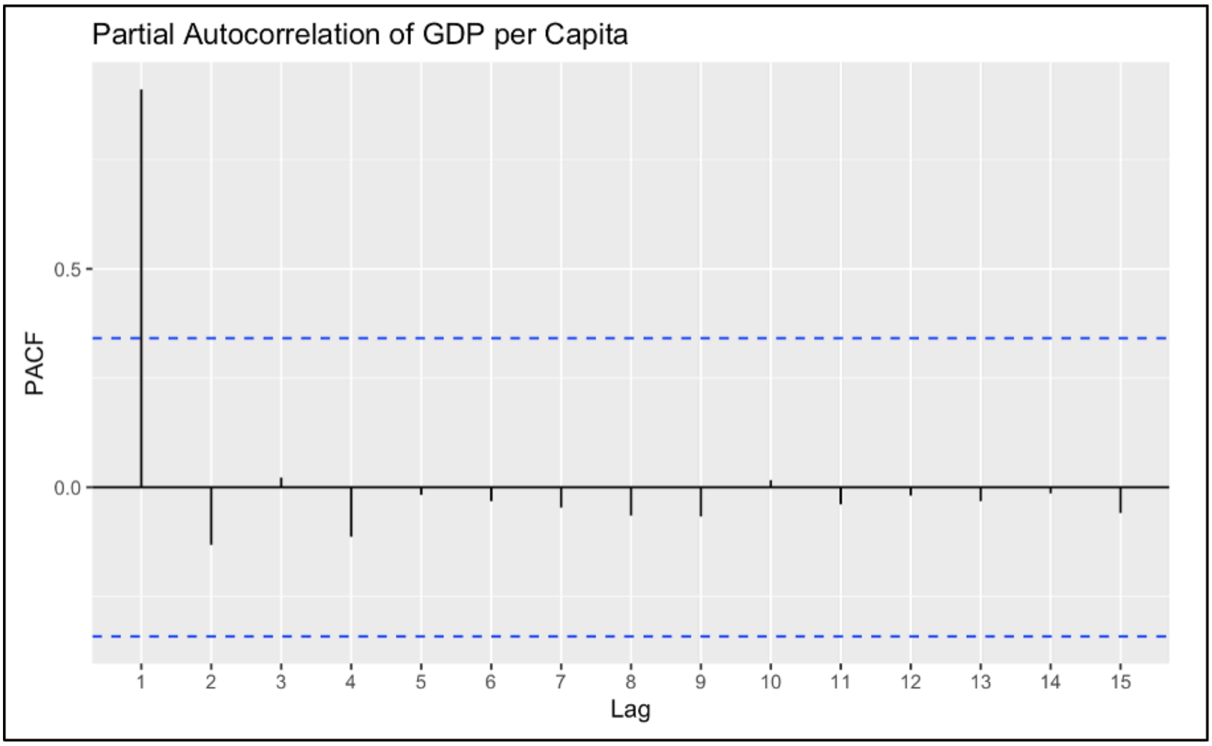
# Plotting the actual values against the predicted values
ggplot(gdp_sec_co2, aes(x = gdp_per_capita, y = predicted_gdp_per_capita))
  geom_point() +
  geom_abline(intercept = 0, slope = 1, color = "blue") +
  labs(title = "Actual vs. Predicted GDP per capita",
       x = "Actual GDP per capita",
       y = "Predicted GDP per capita")
```
```

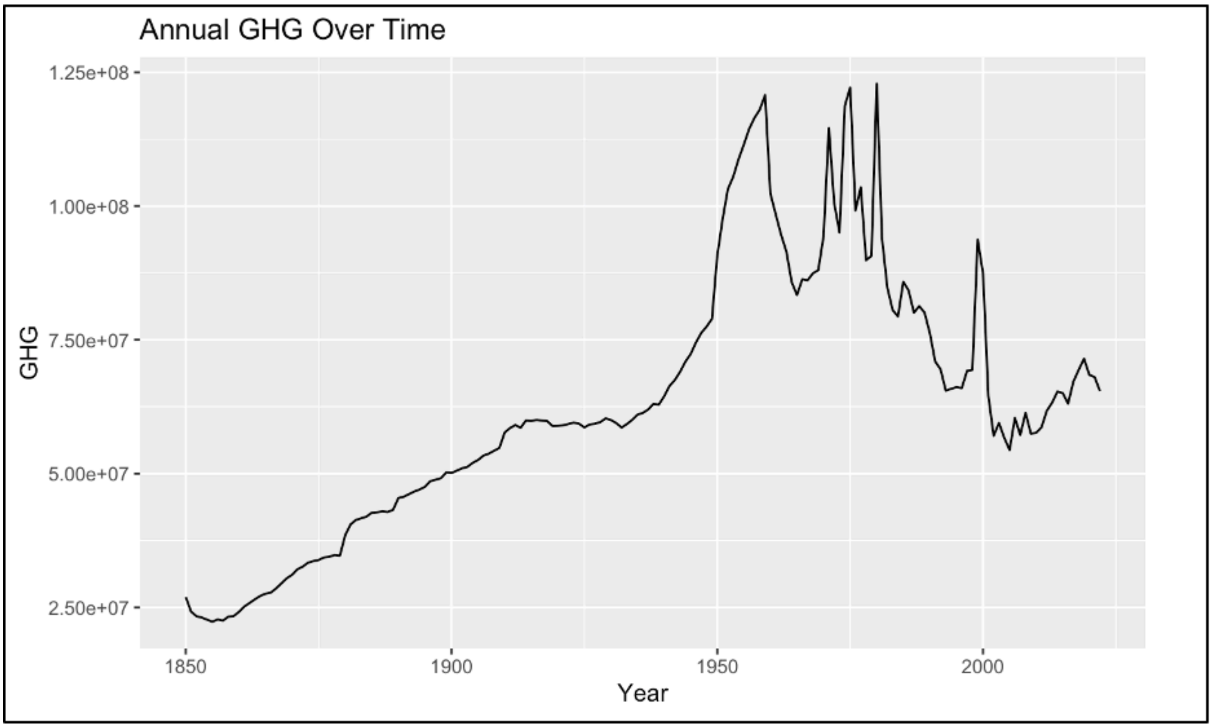
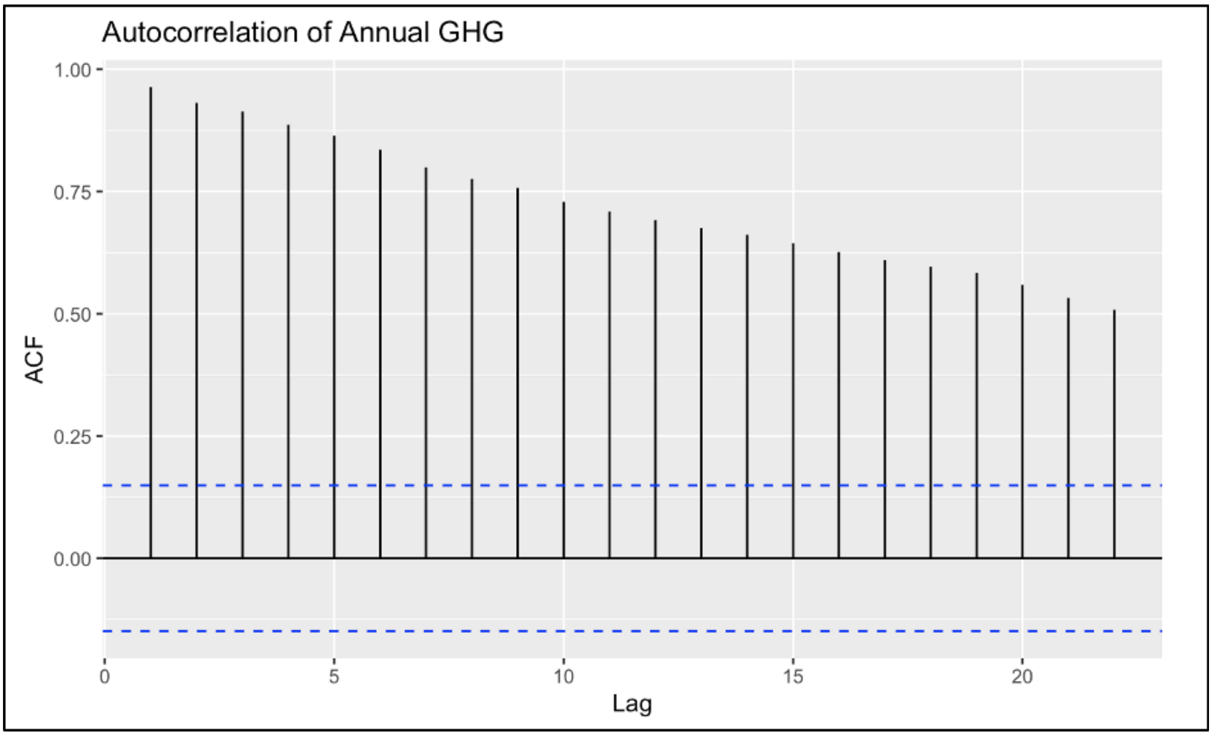
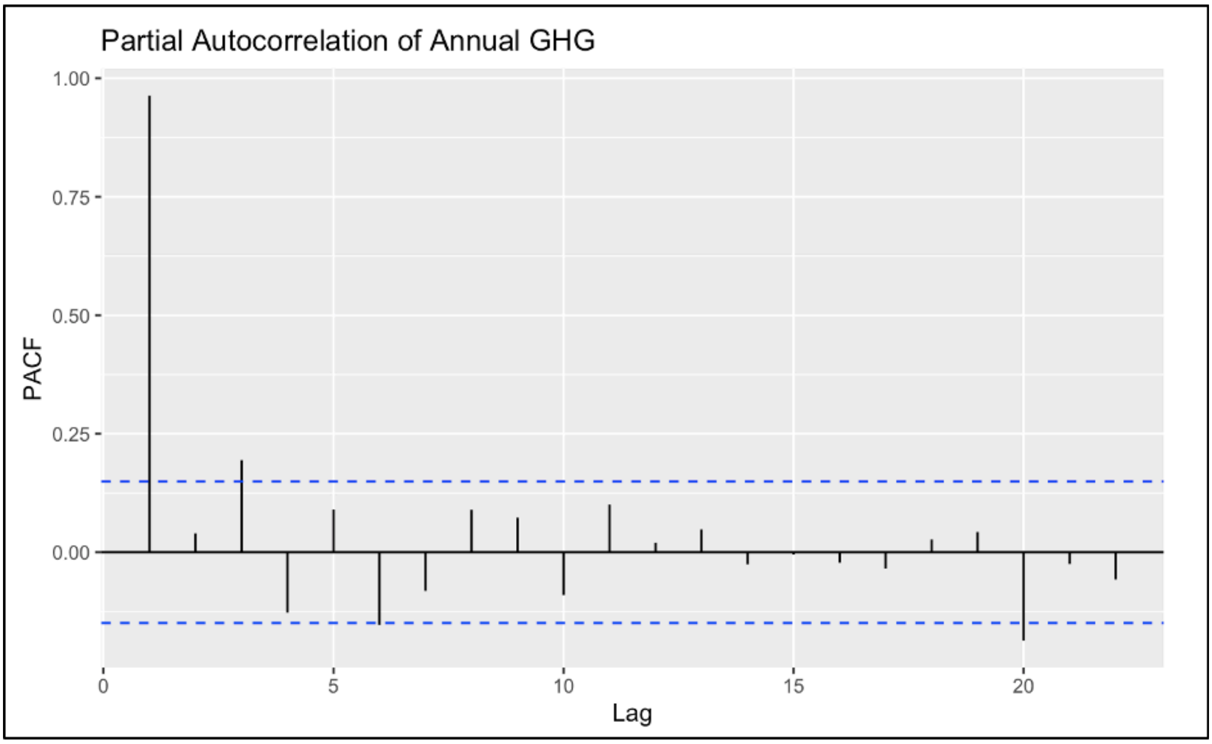
```
Residuals:
    Min      1Q  Median      3Q     Max
-1246.0  -367.9   190.7   375.3   959.5

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  8.497e+04  3.131e+04   2.713 0.011882 *
annual_co2  -1.368e-06  1.735e-05  -0.079 0.937766
industry     3.545e-03  1.236e-03   2.868 0.008273 **
cement       3.845e-03  2.009e-03   1.914 0.067151 .
flaring     -1.631e-03  4.339e-04  -3.759 0.000917 ***
gas         -3.775e-04  1.580e-04  -2.389 0.024779 *
oil          1.469e-04  5.118e-05   2.869 0.008245 **
coal        -3.142e-04  1.019e-04  -3.084 0.004926 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 635.2 on 25 degrees of freedom
Multiple R-squared:  0.9897,    Adjusted R-squared:  0.9868
F-statistic: 343.1 on 7 and 25 DF,  p-value: < 2.2e-16

Adjusted R-squared: 0.9868148
Multiple R: 0.989699
F-statistic: 249.2915
```

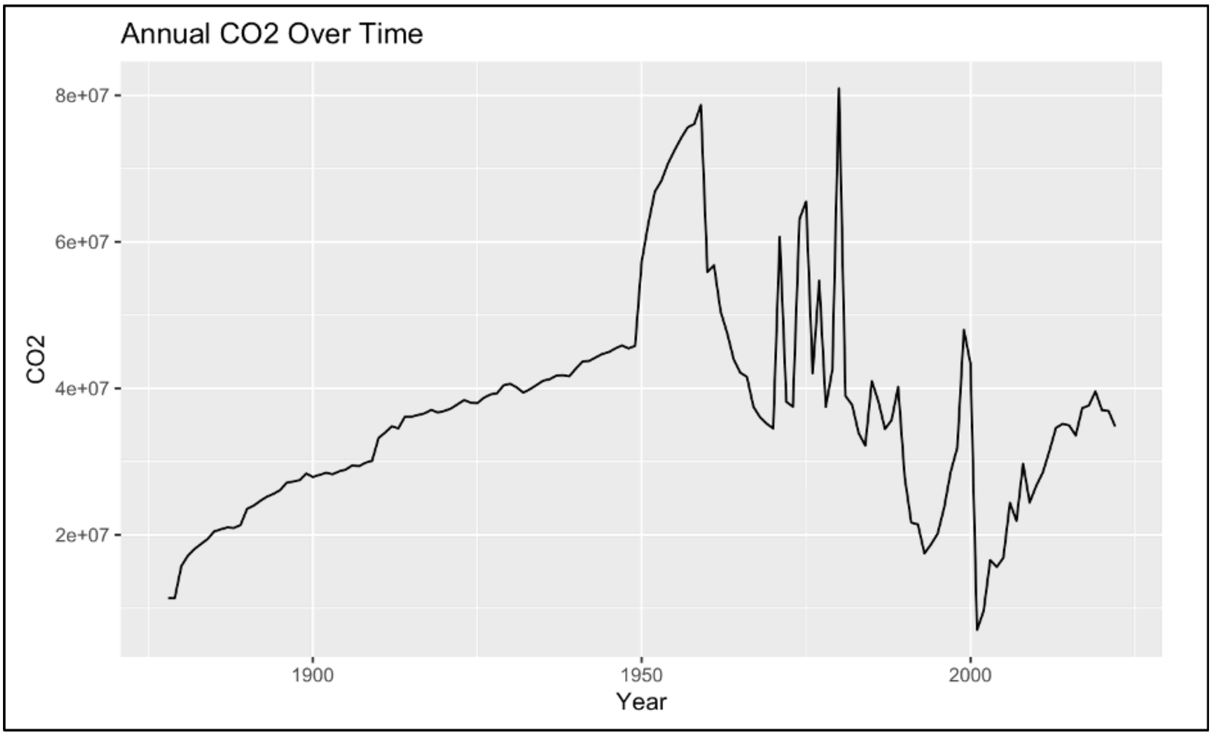
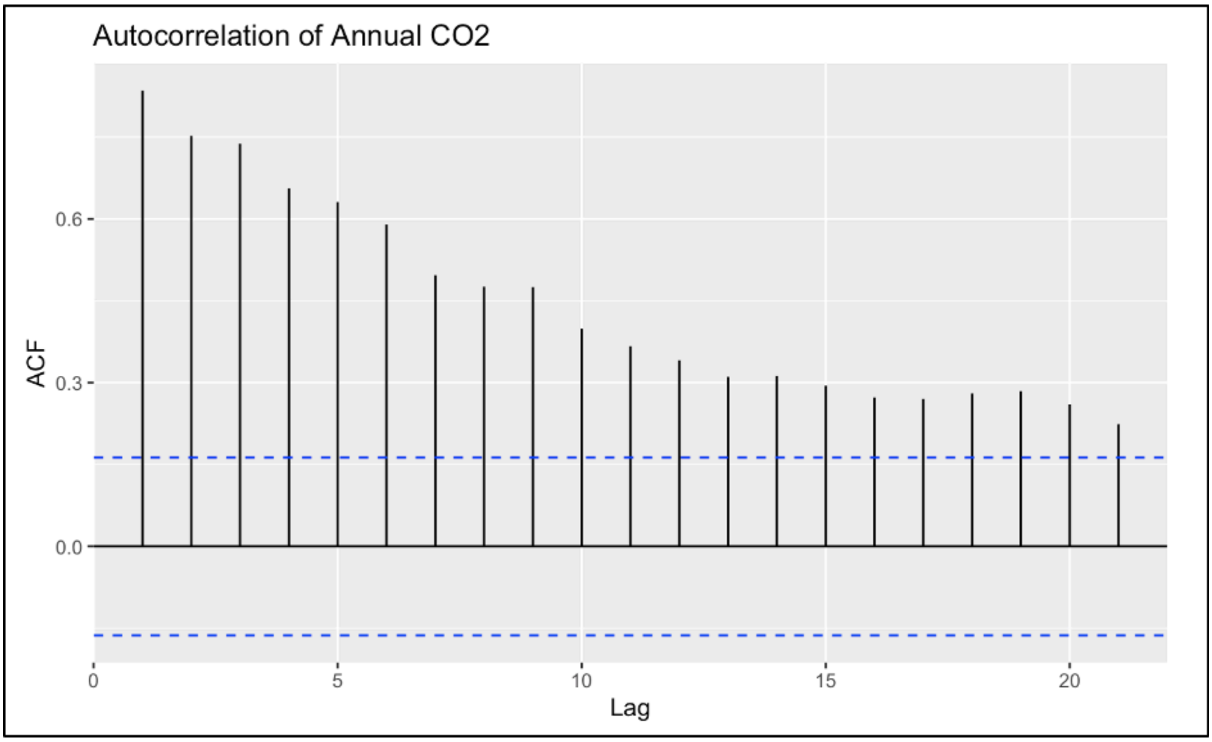
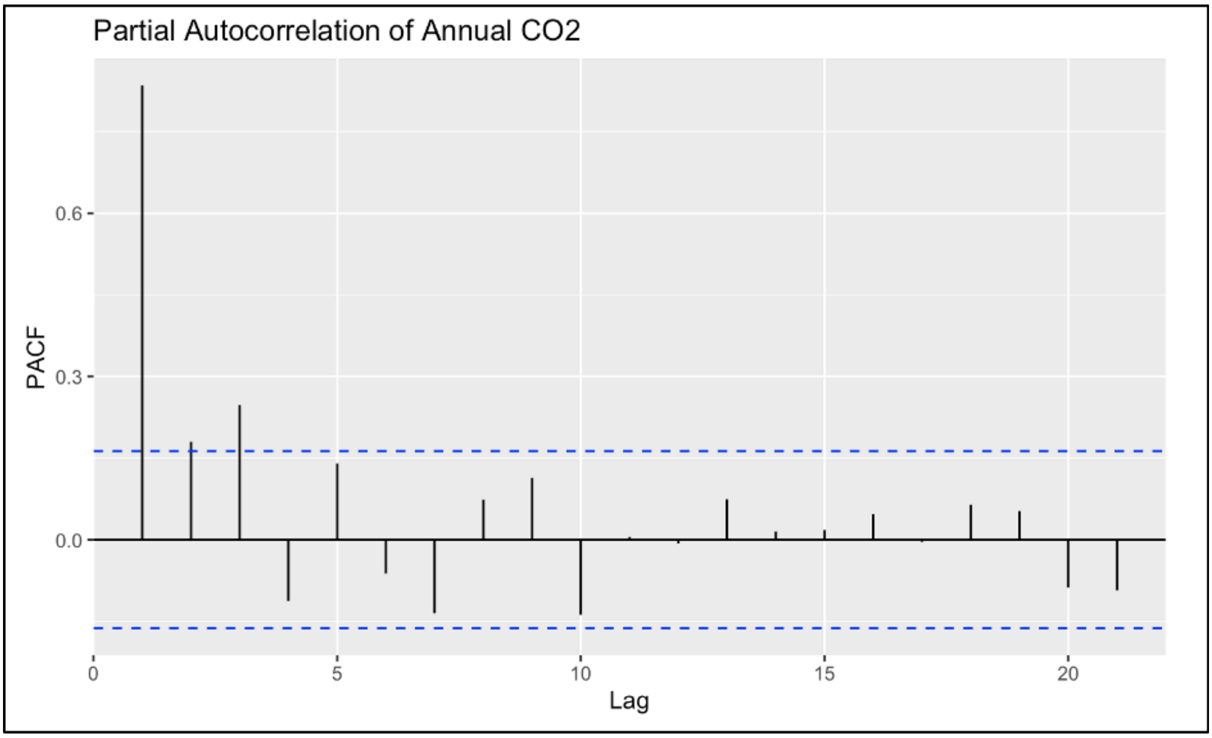
Regression analysis for GDP per capita, Annual CO2 and Cumulative CO2 emissions from various sources - 2
```

Key findings

- Adjusted R square suggests that cumulative CO2 emissions from various sources in NZ can explain a 98% variation in GDP per capita.
- The P-values illustrate that CO2 emissions from flaring (0.000917), industry (0.008273), oil (0.008245), coal (0.004926) and gas (0.0247793) are statistically significant to GDP per capita at 5% significance level.
- The p-value of the f-statistic is 2.2e-16, suggesting the model is the best fit and some independent variables are statistically significant with GDP per capita.
- For each unit increase in flaring, coal, and gas, the GDP per capita decreases by 0.0016310, 0.0003142, and 0.0003775.

Analysis 4: Conclusion and limitations

The correlation analysis demonstrates that most cumulative CO2 emissions from various sectors highly influence GDP per capita in NZ. Moreover, regression analysis shows that cumulative emissions from flaring, industry, oil, coal, and gas are statistically significant in NZ's GDP per capita. The correlation analysis and regression analysis show that cumulative CO2 from various sectors significantly influences GDP per capita in NZ. However, more variables may be needed to refine the model and study their impact on GDP per capita.

Analysis 5 - Autocorrelation and Partial Autocorrelation of GDP per capita, Annual GHG, Annual CO2 and GHG per capita

The report further employs Autocorrelation and Partial correlation analysis to analyze the relationship between the time series data. Autocorrelation measures if past values influence future values (Dondurur, 2018). Partial Autocorrelation helps identify lags by controlling the effects of intervening lags (Mariano & Ozmucur, 2020).

Since this paper focuses on time series data of variables like GDP per capita, Annual GHG, Annual CO2 and GHG per capita in New Zealand, it makes sense to employ Autocorrelation and Partial Autocorrelation.

1. GHG per capita emissions over time in NZ

```
#### autocorrelation and partial autocorrelation analysis

```{r}

#GHG per capita

ghg_pc_ac <- read.csv("GHG_per_capita.csv")

ghg_pc_ac <- ghg_pc_ac %>%
  filter(Entity == "New Zealand")
```

```{r}

ghg_per_capita_ts <- ts(ghg_pc_ac$Per.capita.greenhouse.gas.emissions.in.C

autoplot(ghg_per_capita_ts) +
  ggtitle("GHG per Capita Emissions Over Time") +
  xlab("Year") +
  ylab("GHG per Capita")


#autocorrelation - AC

ggAcf(ghg_per_capita_ts) +
  ggtitle("Autocorrelation of GHG per Capita Emissions")

#partial autocorrelation - PAC
ggPacf(ghg_per_capita_ts) +
  ggtitle("Partial Autocorrelation of GHG per Capita Emissions")
```
```

```
```



GHG per Capita Emissions Over Time
GHG per Capita
Year
25.0
22.5
20.0
17.5
15.0
12.5
1990
2000
2010
2020

Fig 1


Autocorrelation of GHG per Capita Emissions
0.8
0.4
0.0
ACF
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
Lag

Fig 1.1


Partial Autocorrelation of GHG per Capita Emissions
0.75
0.50
0.25
0.00
-0.25
-0.50
PACF
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
Lag

Fig 1.2

Key findings

- The GHG per capita emissions chart shows a noticeable decline in emissions over time.
- ACF shows that past GHG per capita emission values strongly influence future values as the plot shows positive autocorrelations up to lag 5.
- The PACF shows a significant lag at point 1.

2. GDP per capita emissions over time in NZ

````
```{r}
#GDP per capita

gdp_pc_ac <- gdp_pc

```

```{r}

gdp_per_capita_ts <- ts(gdp_pc_ac$`GDP per capita`, start=min(gdp_pc_ac$Y

autoplot(gdp_per_capita_ts) +
  ggtitle("GDP per Capita Over Time") +
  xlab("Year") +
  ylab("GDP per Capita")


#autocorrelation - AC

ggAcf(gdp_per_capita_ts) +
  ggtitle("Autocorrelation of GDP per Capita")

#partial autocorrelation - PAC
ggPacf(gdp_per_capita_ts) +
  ggtitle("Partial Autocorrelation of GDP per Capita")
````
````

```
```


GDP per Capita Over Time
GDP per Capita
45000
40000
35000
30000
1990
2000
2010
2020
Year

Fig 2



Autocorrelation of GDP per Capita
ACF
0.5
0.0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
Lag

Fig 2.1



Partial Autocorrelation of GDP per Capita
PACF
0.5
0.0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
Lag

Fig 2.2

Key findings

- The GDP per capita trend in NZ is growing.
- The GDP per capita values highly depend on the previous year's values, up to lag 8.
- PACF suggests that the relationship between the previous GDP per capita values is minimal beyond the first lag.

3. Annual GHG trends over time in NZ

````
```{r}

#Annual GHG Emissions

ghg_ac <- read_csv("Ann-GHG-NZ.csv", show_col_types = FALSE)

```

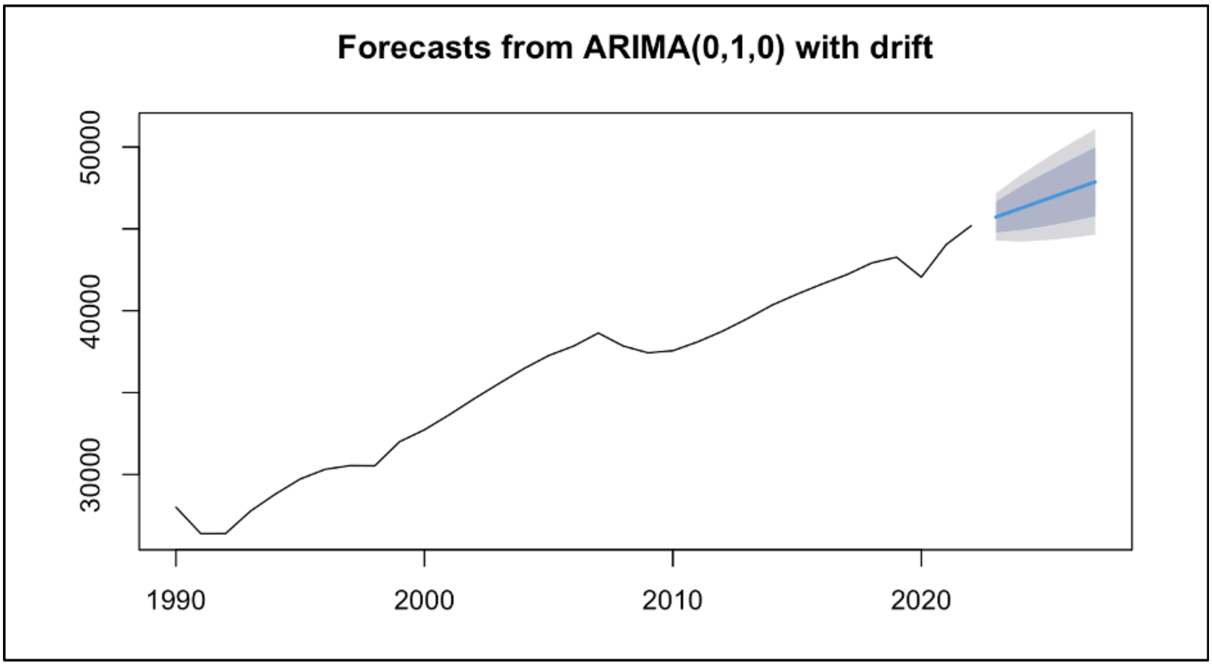
```{r}

ghg_ts <- ts(ghg_ac$`Annual greenhouse gas emissions in CO2 equivalents`, s

autoplot(ghg_ts) +
  ggtitle("Annual GHG Over Time") +
  xlab("Year") +
  ylab("GHG")


#autocorrelation - AC
ggAcf(ghg_ts) +
  ggtitle("Autocorrelation of Annual GHG")

#partial autocorrelation - PAC
ggPacf(ghg_ts) +
````
````

```
  ggtitle("Partial Autocorrelation of Annual GHG")
```



Fig 3


Autocorrelation of Annual GHG
ACF
1.00
0.75
0.50
0.25
0.00
0
5
10
15
20
Lag

Fig 3.1


Partial Autocorrelation of Annual GHG
PACF
1.00
0.75
0.50
0.25
0.00
0
5
10
15
20
Lag

Fig 3.2

Key findings

- The plot illustrates a steady increase in GHG over time, peaking in the middle of the 20th century and dropping in the later part with a high fluctuation after 2000.
- ACF shows a high degree of correlation, demonstrating that GHG emissions strongly depend on the previous year's values.
- PACF shows a substantial spike on lag 1 and low correlations beyond that.

4. Annual CO2 emissions over time in NZ

````
```{r}
#Annual CO2

co2_ac <- read_csv("Ann-CO2.csv", show_col_types = FALSE)

```

```{r}
co2_ts <- ts(co2_ac$`Annual CO2 emissions including land-use change`, start

autoplot(co2_ts) +
  ggtitle("Annual CO2 Over Time") +
  xlab("Year") +
  ylab("CO2")


#autocorrelation - AC

ggAcf(co2_ts) +
  ggtitle("Autocorrelation of Annual CO2")

#partial autocorrelation - PAC
ggPacf(co2_ts) +
  ggtitle("Partial Autocorrelation of Annual CO2")
````
````

```
```


Annual CO2 Over Time
8e+07
6e+07
4e+07
2e+07
CO2
1900
1950
2000
Year

Fig 4


Autocorrelation of Annual CO2
0.6
0.3
0.0
ACF
0
5
10
15
20
Lag

Fig 4.1



Fig 4.2

Key findings

- The plot shows a gradual increase in CO2 emissions followed by fluctuations from the mid-20th century.
- ACF illustrates a strong correlation in the initial lags followed by gradual decay over time. This suggests that CO2 emissions are highly dependent on past values.
- PACF demonstrates a significant spike at lag 1, followed by smaller subsequent spikes, demonstrating that the most influence comes from the past.

Analysis 6 - Forecasting GDP per capita, Annual GHG, Annual CO2 and GHG per capita for the next 5 years using ARIMA model

Time series analysis is a part of Predictive analytics (tableau, n.d.). It helps predict future events by analyzing historical trends and patterns. Forecasting GHG and CO2 emissions can assist governments and policymakers in developing plans to achieve carbon neutrality early (Huang et al., 2023). Moreover, GDP prediction can help balance the economies by providing insights on where to invest and spend (Liberto, 2024).

> ARIMA (Autoregressive integrated Moving Average) model helps in forecasting possible future values of a time series and is one of the widely used models for time series forecasting (Noble, 2024).

Using the historical data, the report further predicts the following five-year values: Annual GHG, GDP per capita, GHG per capita and Annual CO2 with the help of a few ARIMA models.

1. GDP per capita

````
```{r}
#GDP per capita

gdp_arima <- auto.arima(gdp_per_capita_ts)
summary(gdp_arima)

# Forecast for the next 5 years
gdp_forecast <- forecast(gdp_arima, h = 5)
plot(gdp_forecast)

```
````



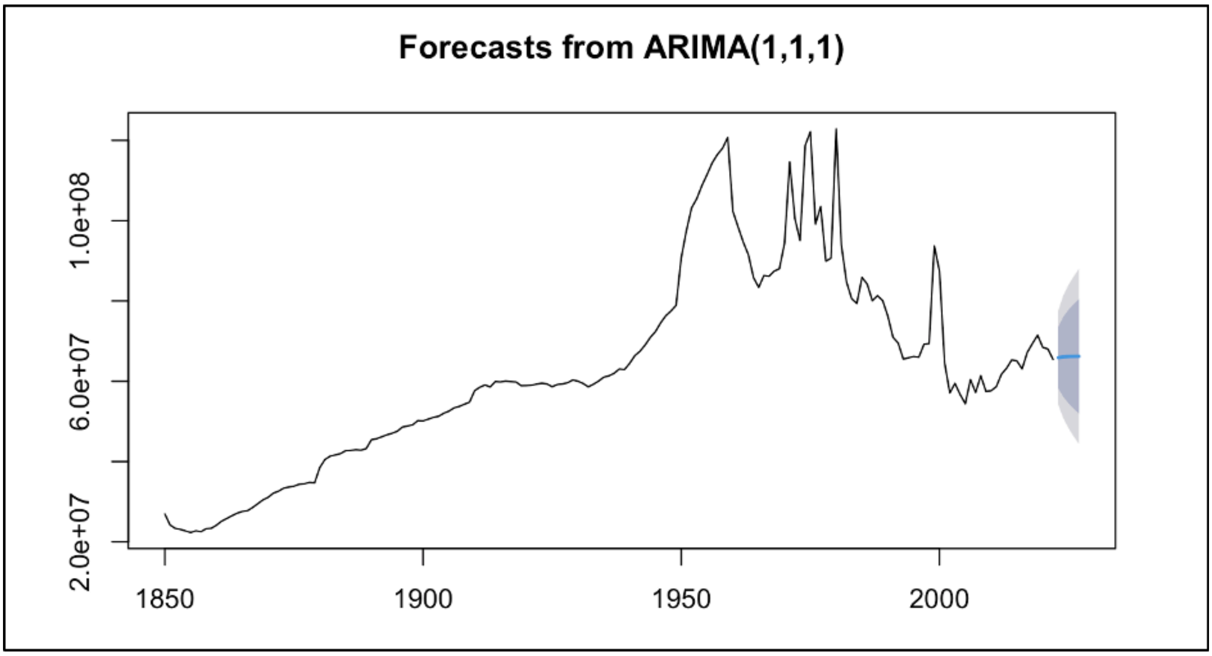
2. Annual GHG

```
```{r}
#Annual GHG


ghg_arima <- auto.arima(ghg_ts)
summary(ghg_arima)

# Forecast for the next 5 years
ghg_forecast <- forecast(ghg_arima, h = 5)
plot(ghg_forecast)


```
```



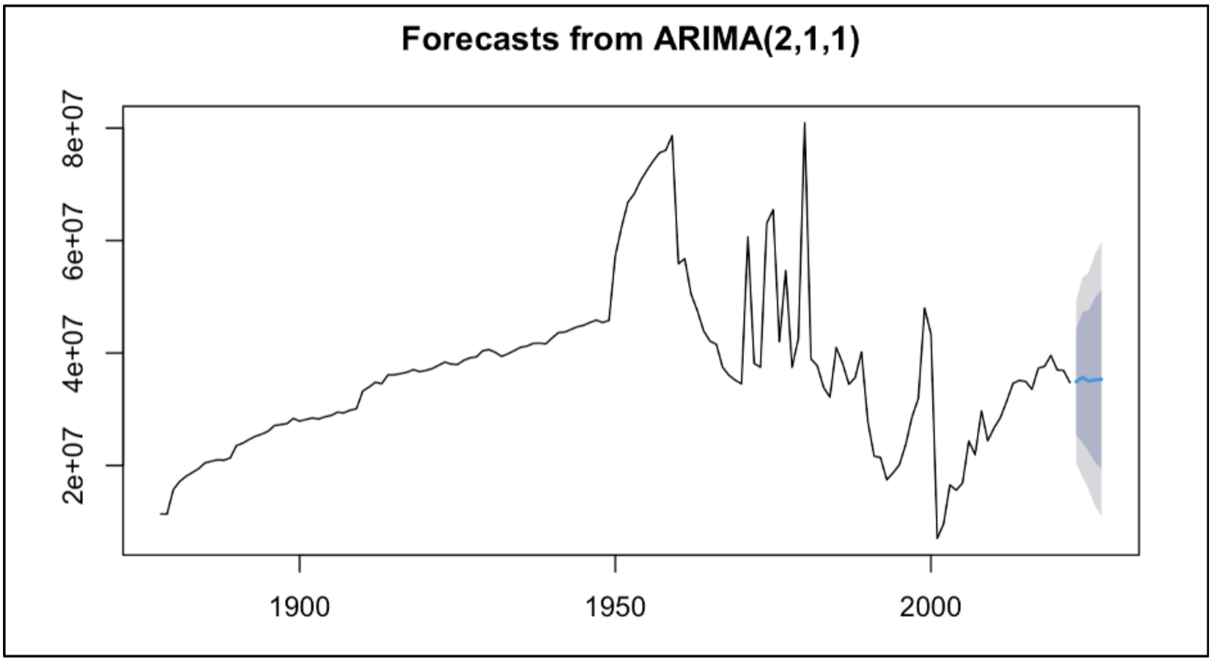
3. Annual CO2

```{r}
#annual CO2

co2_arima <- auto.arima(co2_ts)
summary(co2_arima)

# Forecast for the next 5 years
co2_forecast <- forecast(co2_arima, h = 5)
plot(co2_forecast)


```



4. GHG per capita

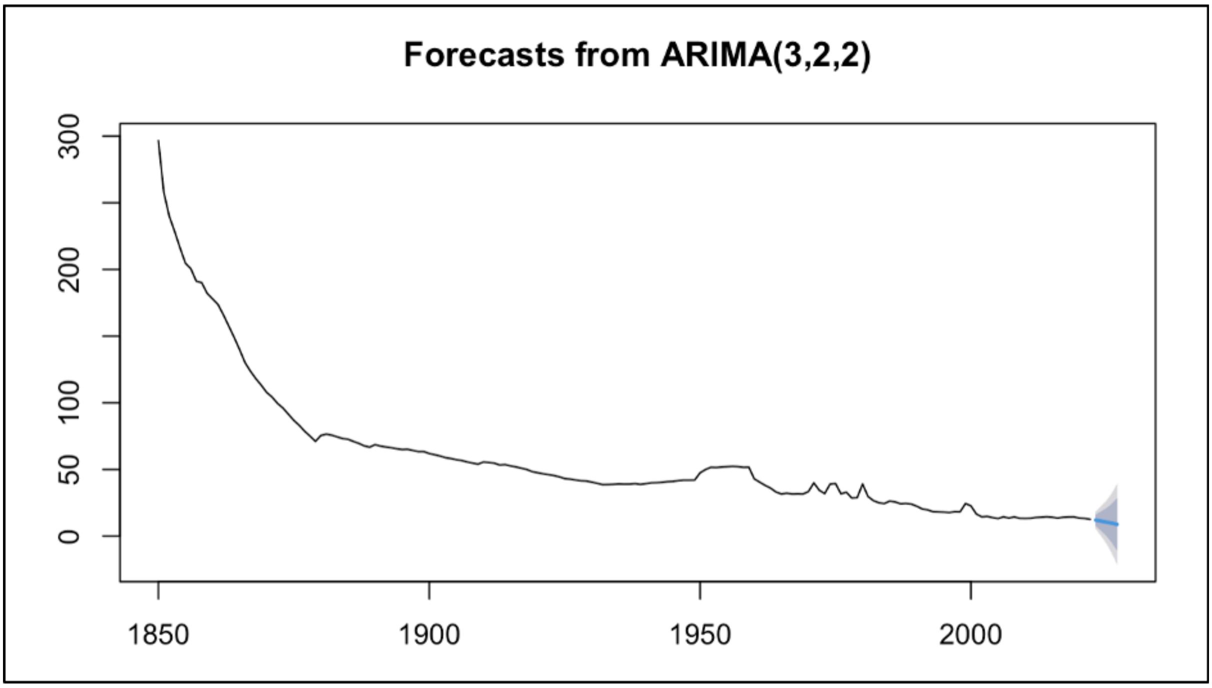
```
```{r}

#GHG Per capita

ghg_per_capita_arima <- auto.arima(ghg_per_capita_ts)
summary(ghg_per_capita_arima)

# Forecast for the next 5 years
ghg_per_capita_forecast <- forecast(ghg_per_capita_arima, h = 5)
plot(ghg_per_capita_forecast)

```
```



Key findings:

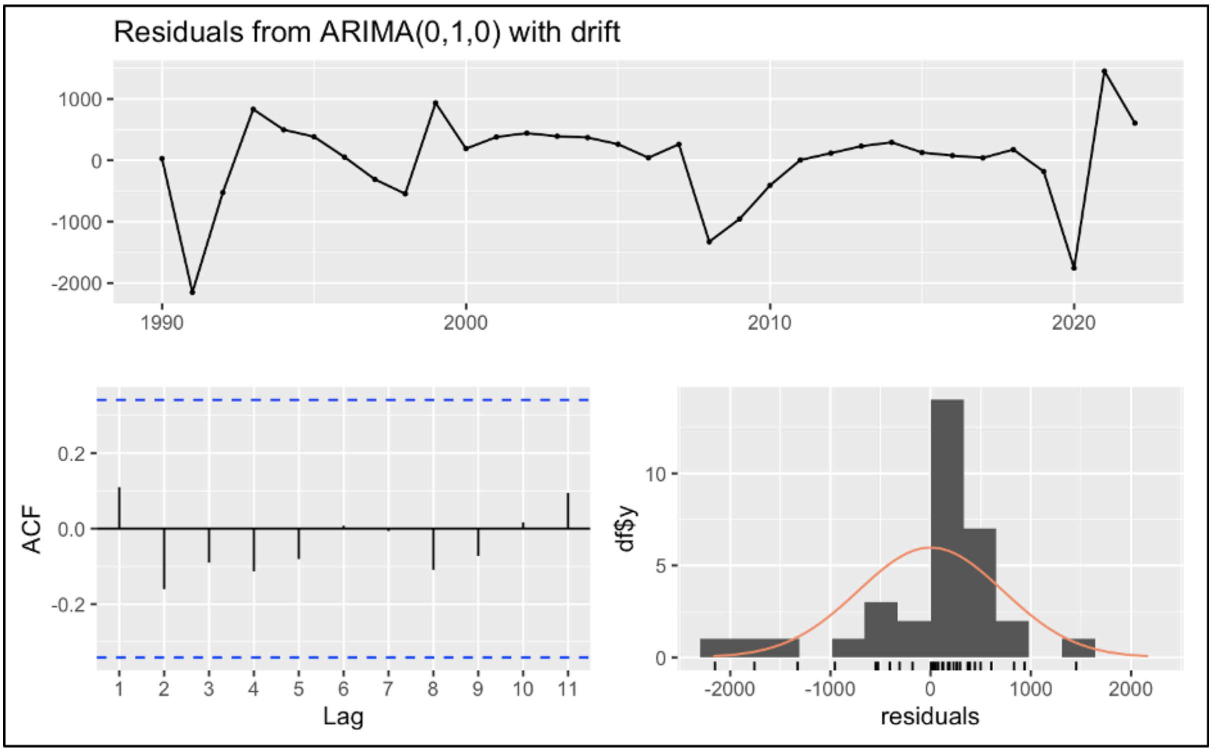
- GDP per capita is supposed to increase, showing an upward trajectory.
- Annual GHG emissions show stability due to trend and volatility.
- Annual CO2 emissions demonstrate flatness due to peaks and drops in historical data, suggesting consistent trends.
- GHG per capita displays a downward trend due to a long-term decline pattern, suggesting reductions in future as well.

Analysis 7 - Model performance evaluation

Evaluating the models is necessary to determine their accuracy (Hyndman & Athanasopoulos, n.d.). The report further evaluates the models of predicted variables.

1. GDP per capita

```
# Check residuals for GDP per Capita model
checkresiduals(gdp_arima)
```

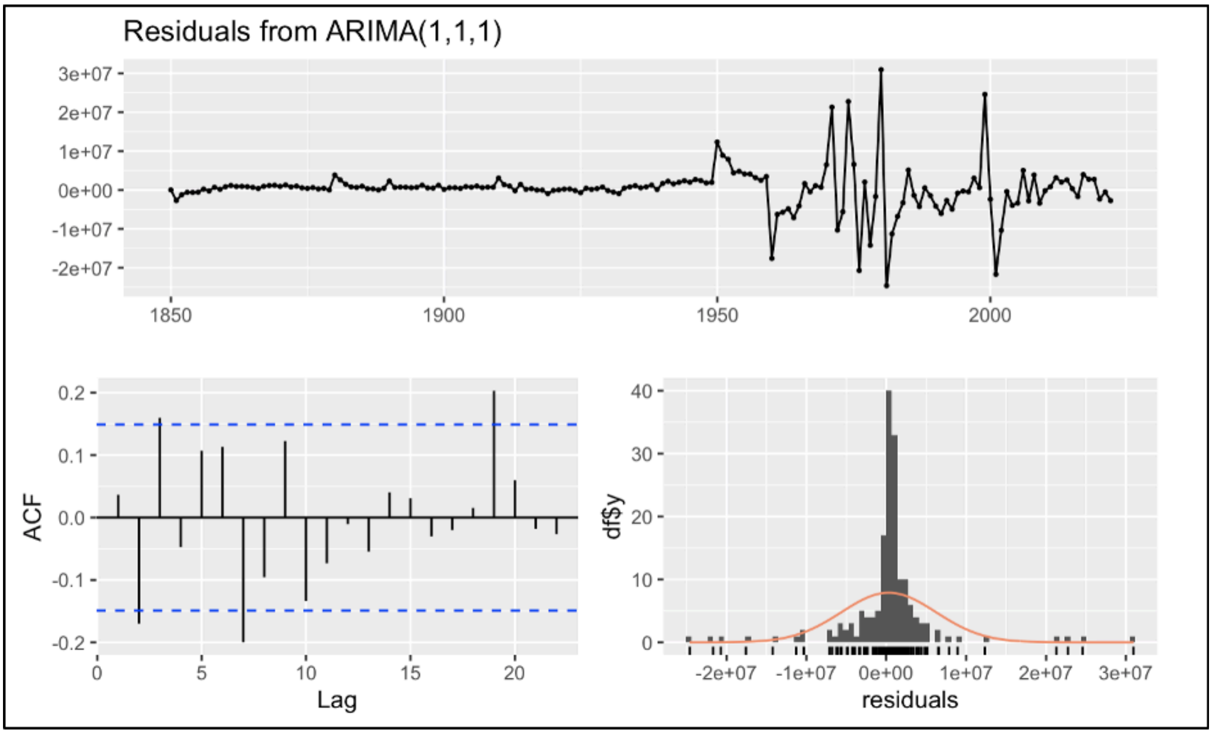


Key conclusion

The residuals exhibit no significant autocorrelation and are approximately normally distributed with some outliers and skewness. The model generally captures the data structure.

2. Annual GHG

```
checkresiduals(ghg_arima)
```

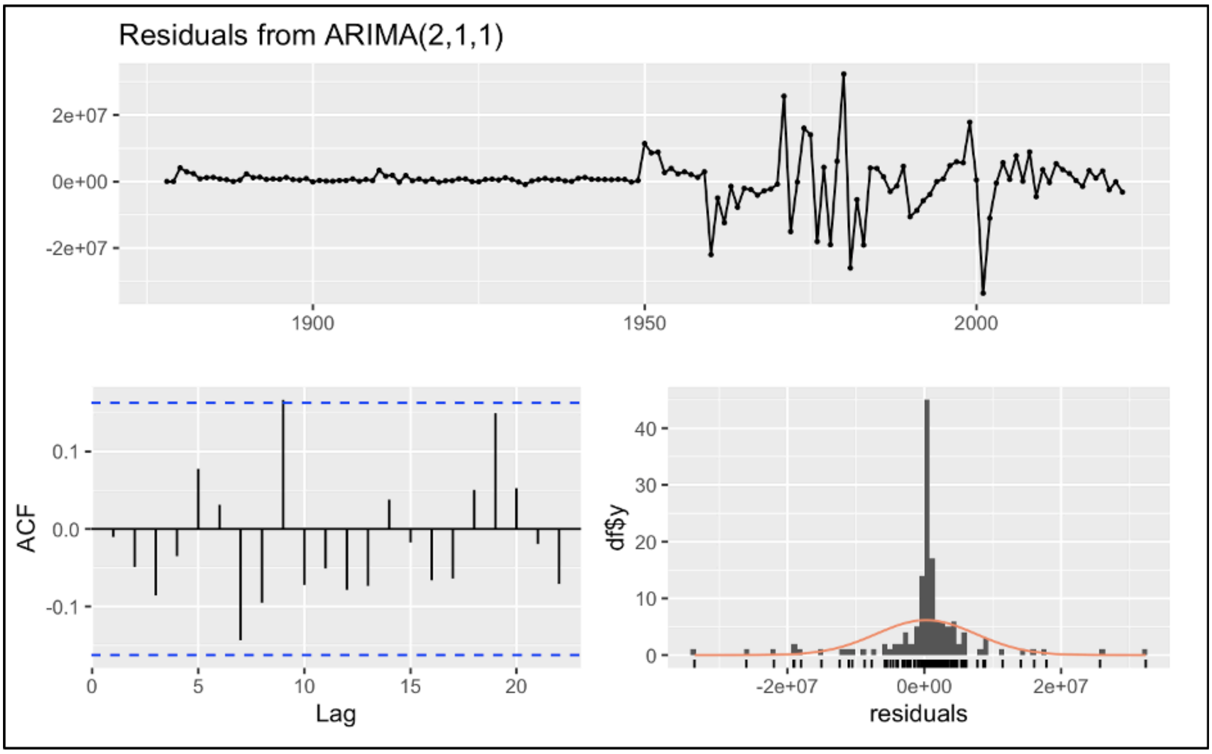


Key conclusion

The residuals demonstrate no clear patterns and follow a rough normal distribution. The ACF shows minimal autocorrelation. This indicates that the model has captured the data structure well.

3. Annual CO2 emissions

```
checkresiduals(co2_arima)
```

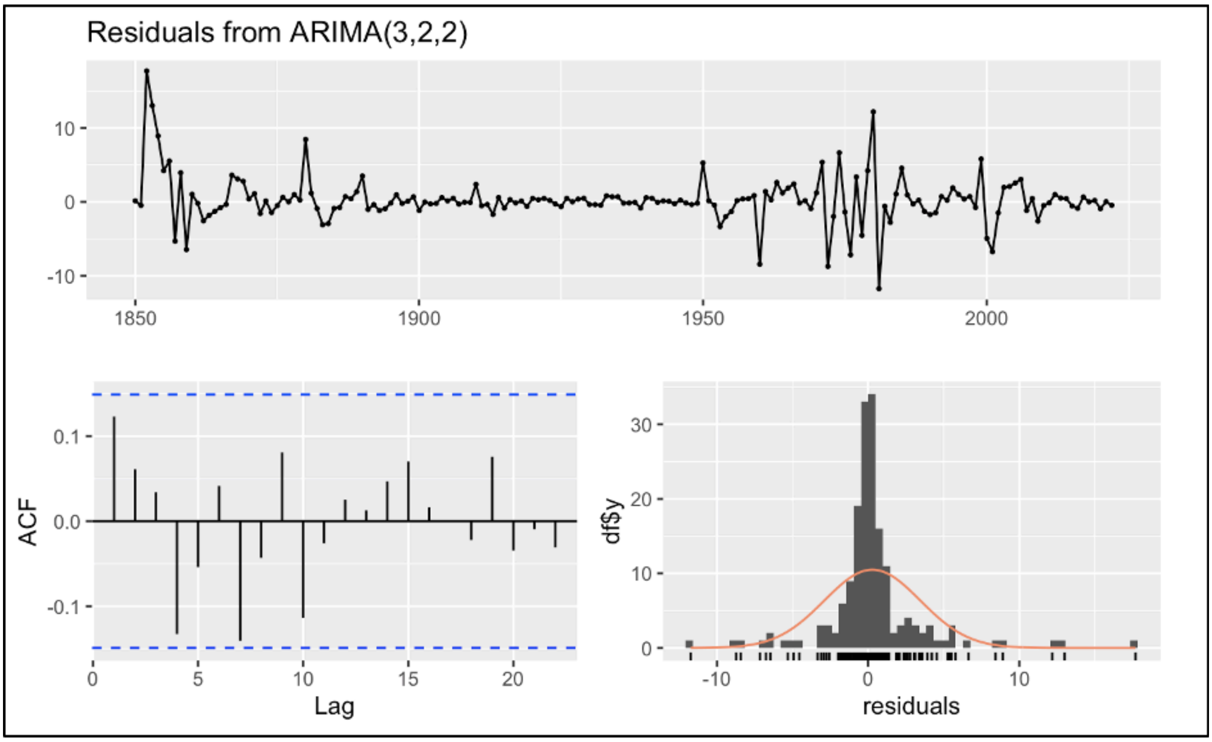


Key conclusion

- The residuals show no clear pattern, have minimum autocorrelation, and demonstrate rough normality. This suggests that the model captured data structure effectively.

4. GHG per capita emissions

```
checkresiduals(ghg_per_capita_arima)
```



Key conclusion

There is no clear pattern in the residuals. Moreover, ACF shows minimum autocorrelation and the data is approximately normally distributed. This suggests that the model captured data structure effectively with minimal residual errors.

Analysis 7 - AIC of GDP per capita, Annual GHG, CO2 and GHG per capita emissions

AIC (Akaike information criterion) is used to evaluate how well the model fits the data it was created from (Bevans, 2023). This technique determines which model is the best fit for the given data. It is often used to select the best model by comparing the various models.

Below are the AIC outputs of the four variables.

```{r}
# Compare AIC values
AIC(gdp_arima, ghg_arima, co2_arima, ghg_per_capita_arima)

```

	df <dbl>	AIC <dbl>
gdp_arima	2	516.1399
ghg_arima	3	5859.6980
co2_arima	4	4967.9031
ghg_per_capita_arima	6	895.2392

4 rows

Key conclusion

- The GDP per capita has the lowest AIC value (516.14), suggesting it is the most effective fitting data.
- Annual GHG has the highest AIC value (5859.69) and the least favourable fit.
- GHG per capita model has a moderately low AIC value (895.24), outperforming CO2 and GHG.
- GDP per capita model efficiently captures the data well.