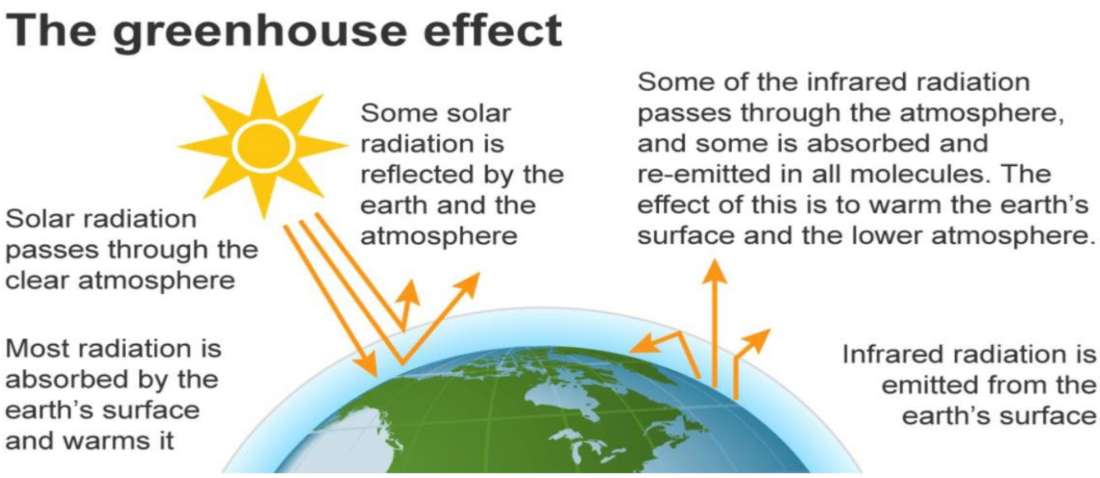
<u>Team Member's:</u>

1. **Jithin Raj Reghuvaran**
2. **Stanley Varghese**
3. **Mohammad Arif**
4. **Frank Kodua**
5. **Ashok Nandah Ramakrishnan**
6. **Singh Kunal**

1. Introduction

Global climate change is not a future problem. Changes to Earth's climate driven by increased human emissions of heat-trapping greenhouse gases are already having widespread effects on the environment: glaciers and ice sheets are shrinking, river and lake ice is breaking up earlier, plant and animal geographic ranges are shifting, and plants and trees are blooming sooner. Effects that scientists had long predicted would result from global climate change are now occurring, such as sea ice loss, accelerated sea level rise, and longer, more intense heat waves.

Gases that trap heat in the atmosphere are called greenhouse gases, which constitutes carbon dioxide (co2), Methane (CH4), Nitrous oxide (N2O), fluorinated gases.

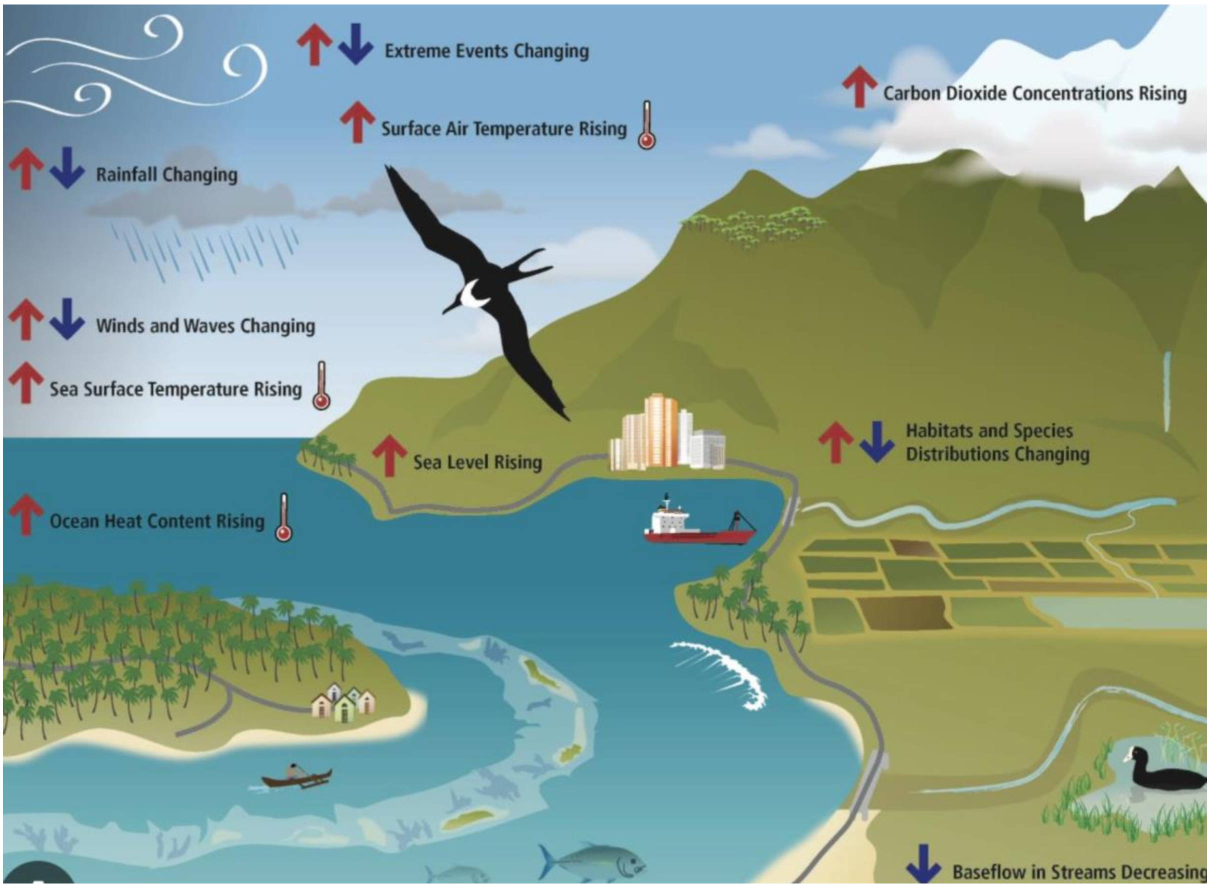


Among all greenhouse gases Co2 is the major contributor to greenhouse effect. Carbon dioxide is naturally present in the atmosphere as part of the Earth's carbon cycle (the natural circulation of carbon among the atmosphere, oceans, soil, plants, and animals). Human activities are altering the carbon cycle–both by adding more CO_2 to the atmosphere and by influencing the ability of natural sinks, like forests and soils, to remove and store CO_2 from the atmosphere. While CO_2 emissions come from a variety of natural sources, human-related emissions are responsible for the increase that has occurred in the atmosphere since the industrial revolution. The main human activity that emits CO_2 is the combustion of fossil fuels (coal, natural gas, and oil) for energy and transportation. Certain industrial processes and land-use changes also emit CO_2.

Current greenhouse gas emissions are too high to achieve the Paris Agreement goal of limiting global average temperature increase to 1.5 °C. People who contribute least to global emissions are already disproportionately affected by climate change, including through a range of direct and indirect impacts on health.[4] These impacts include the effects of increasing exposure to extreme heat, wildfires, floods and droughts, increased exposure to a range of infectious diseases, food insecurity, undernutrition, poverty and population displacement. Approximately 771 million people were responsible for nearly half of greenhouse gas emissions in 2019, while half of the world population emitted only about one tenth of total of global emissions.

In this analysis, we evaluate the accountability of historical Co2 emission from high GDP countries and rest of the world. Moreover, it also answers the research question *"Is historical Co2 emission from High GDP countries significantly different to rest of the world. ? "*

This analysis also focuses a descriptive analysis of climate change indicators.



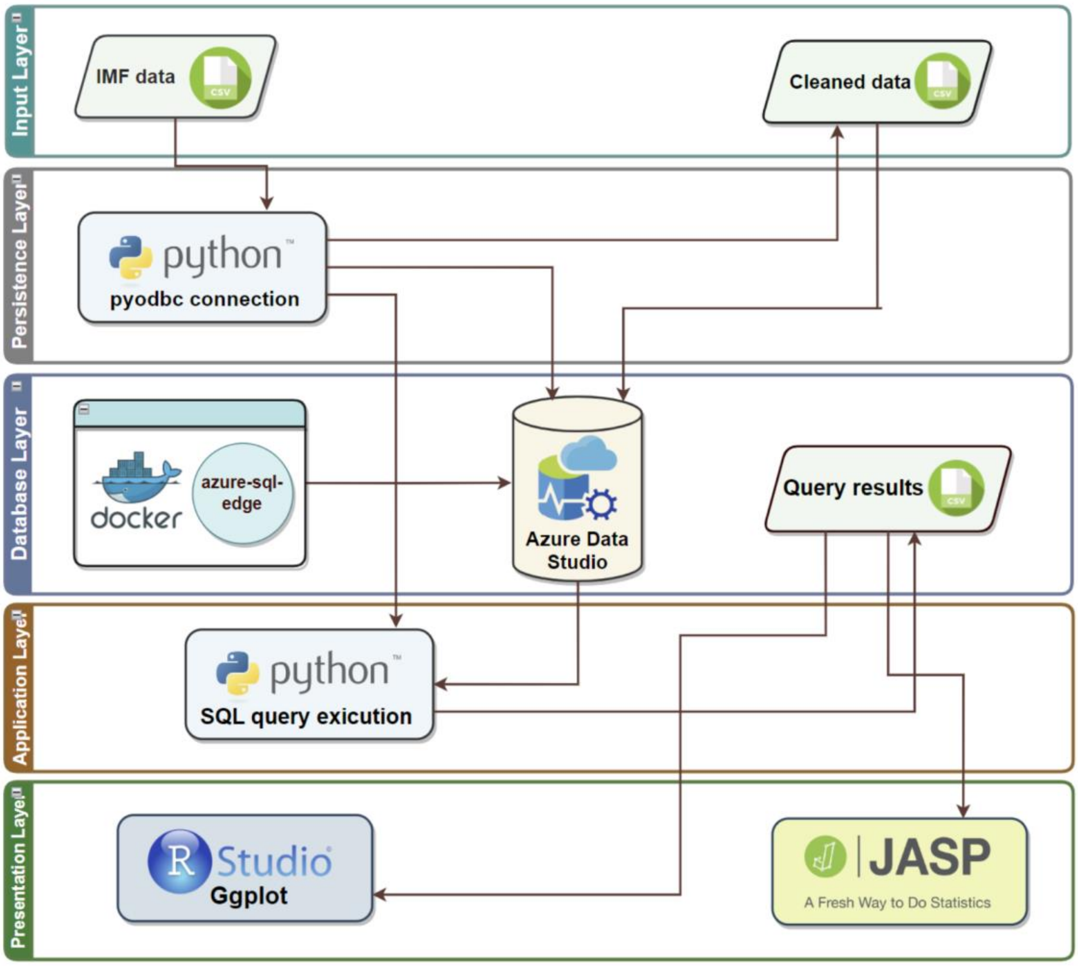
From the climate change indicators like sea surface temperature change, rainfall change, co2 concentration etc., this analysis focuses on Co2 concentration, greenhouse gas concentration, forest area and surface temperature.

2. Dataset

Climate data from International Monetary Fund is used for this analysis.
https://climatedata.imf.org/

3. Data FLow Diagram

Following Architecture is used for this analysis.



The technologies mentioned (Azure Data Studio, Docker, RStudio, and JASP) are incorporated into their respective layers.

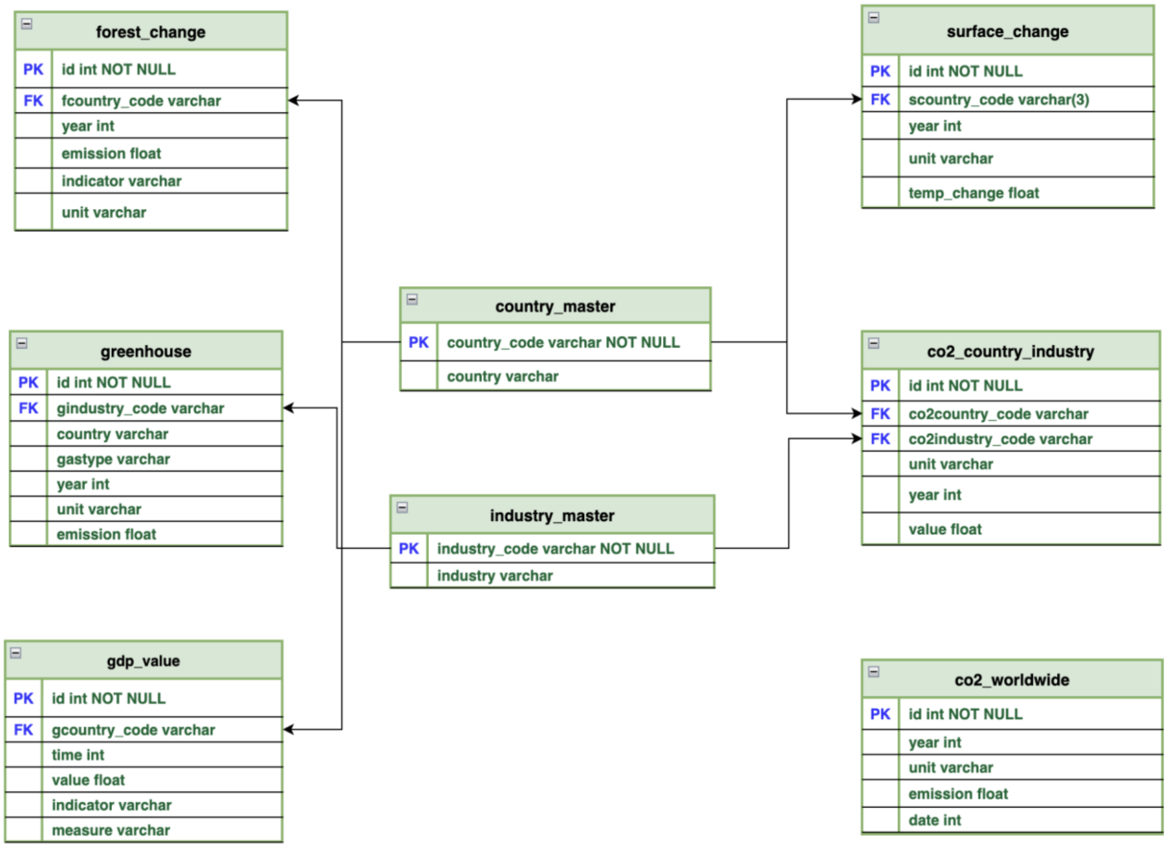
1. Input Layer:
   - IMF data is received in the Input Layer.
   - The data is cleaned and prepared for insertion into the database.
2. Persistence Layer:
   - PyODBC is used for data persistence in this layer.
   - The cleaned data from the Input Layer is inserted into the Azure Data Studio database via a PyODBC connection.

3. Database Layer:
   - Azure Data Studio is the chosen platform for database management.
   - Docker along with azure-sql-edge image is used to set up SQL database in Azure Data Studio.
4. Application Layer:
   - Python is used in the Application Layer for SQL query execution.

- Results of the executed queries are obtained and are collected in the database layer.

5. Presentation Layer:
    - The results from the Application Layer are presented in the Presentation Layer.
    - RStudio is used for visualizing and manipulating data.
    - JASP (Jeffreys Amazing Statistics Program) is employed for statistical analysis and presentation of results.

4. Entity Relationship Diagram



Below is the breakdown of the above presented Data schema.

1. Country Code Master File:
    - Primary Key: Country Code
    - This file serves as a master reference for country codes for all the world countries.
2. Industry Code Master File:
    - Primary Key: Industry Code
    - This file serves as a master reference for industry codes with their respective industries.
3. Forest Change File:
    - Foreign Key: Country Code

- Attributes related to changes in forest data.

4. Greenhouse File:
    - Foreign Key: Industry Code
    - Attributes related to emission according to the countries for different gas types.
5. GDP Value File:
    - Foreign Key: Country Code
    - Attributes related to the change in the country's GDP.
6. Surface Change File:
    - Foreign Key: Country Code
    - Attributes related to change in surface area temperature of different countries.
7. Co2 Country Industry File:
    - Foreign Key: Country Code and Industry Code
    - Attributes related to the emission according to the countries with for different industries.
8. Co2_worldwide:
    - Attributes related to the emission of Co2 worldwide through the years.

In summary, this database schema has two master files (Country Code and Industry Code) that contain the primary keys.
These keys are then used to connect with the remaining files which contain the foreign keys and contain the information on forest changes, GDP values, surface changes, greenhouse gas emissions, and Co2 emissions.
The relationships between these files allow for the retrieval and analysis of data based on both country and industry perspectives. The tables retrieved are used for the graphical analysis.
The Co2 worldwide file is used to describe the initial analysis about the change in Co2 emission through the years.

5. Queries for Table Creation

Tables are created in Azure data studio.

1.1. Database creation

```
CREATE DATABASE "CDMS"
```

A new database named "CDMS" is created for this project

1.2. Creating Table: country_master

```
CREATE TABLE country_master (
Country VARCHAR(50),
Country_code VARCHAR(6),
PRIMARY KEY (Country_code)
);
```

1.3. Creating Table: industry_master

```
CREATE TABLE industry_master (
   Industry VARCHAR(50)
   Industry_code VARCHAR(6),
PRIMARY KEY (Industry_code)
);
```

1.4. Creating Table: co2_country_industry

```
CREATE TABLE co2cuntry_industry (
 id INTEGER,
 co2Industry_code VARCHAR(6),
 Unit VARCHAR(6),
 co2country_code VARCHAR(6),
 Year INTEGER,
 Value float,
PRIMARY KEY (id),
FOREIGN KEY co2Industry_code REFERENCES industry_master(Industry_code),
FOREIGN KEY  co2country_code REFERENCES country_master(Country_code));
```

1.5. Creating Table: co2_world_wide

```
CREATE TABLE co2world_wide (
 id INTEGER,
 Unit VARCHAR(6),
 Date INTEGER,
 Value float,
 Year INTEGER,
PRIMARY KEY (id),
);
```

1.6. Creating Table: forest_change

```
CREATE TABLE forest_change (
 id INTEGER,
 Unit VARCHAR(6),
 Indicator VARCHAR(50),
 fcountry_code VARCHAR(6),
 Year INTEGER,
 emission float,
PRIMARY KEY (id),
FOREIGN KEY  fcountry_code   REFERENCES country_master(Country_code));
```

1.7. Creating Table: GDP_VALUE

```
CREATE TABLE GDP_VALUE (
Id INTEGER,
 GCOUNTRY_CODE VARCHAR(6),
 INDICATOR VARCHAR(10),
 MEASURE VARCHAR(10),
 TIME INTEGER,
 Value float,
PRIMARY KEY (id),
FOREIGN KEY  GCOUNTRY_CODE REFERENCES country_master(Country_code));
```

1.8. Creating Table: greenhouse

```
CREATE TABLE greenhouse (
 id INTEGER,
 Country VARCHAR(50),
 gIndustry_code VARCHAR(6),
 gastype VARCHAR(10),
 Unit VARCHAR(50),
 Year INTEGER,
 emission float,
PRIMARY KEY (id),
FOREIGN KEY  gIndustry_code  RFERENCES industry_master(Industry_code));
```

1.9. Creating Table: Surface_change

```
CREATE TABLE Surface_change (
 id INTEGER,
 Unit VARCHAR(50),
 scountry_code VARCHAR(6),
 Year INTEGER,
 temp_change float,
PRIMARY KEY (id),
FOREIGN KEY  scountry_code REFERENCES country_master(Country_code));
```

2. Monthly variation of atmospheric carbon dioxide concentrations in ppm

SQL code:

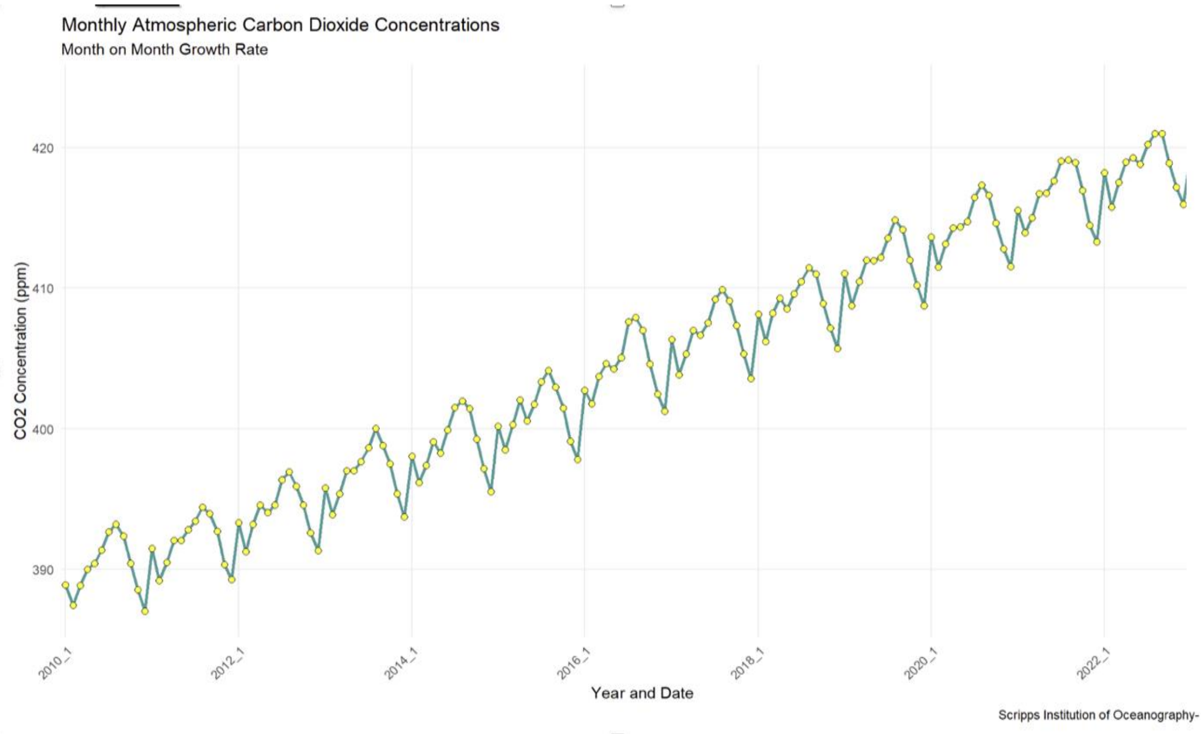
```
SELECT Year, Unit , value as emission, Date as date
FROM CO2world_wide
WHERE Unit IN ('Parts Per Million') AND Year >1990
ORDER BY Year;
```

CO2world_wide table is used for this query. WHERE and ORDER By clause is also used to get the required output.

Query Result:

	Year	Unit	emission	date
1	1991	Parts Per Million	354.93	1
2	1991	Parts Per Million	355.82	2
3	1991	Parts Per Million	357.33	3
4	1991	Parts Per Million	358.77	4
5	1991	Parts Per Million	359.23	5
6	1991	Parts Per Million	358.23	6

Analysis and interpretation:



Result from the sql query is used to plot month wise variation of Co2 concentration.

Key Takeaways:

- From the plot we can infer that atmospheric co2 concentration is on the rise.
- Co2 emission worldwide shows seasonal trend, where it drops in some months.

3. Monthly Variation of atmospheric carbon dioxide concentrations in percent:

SQL Query:

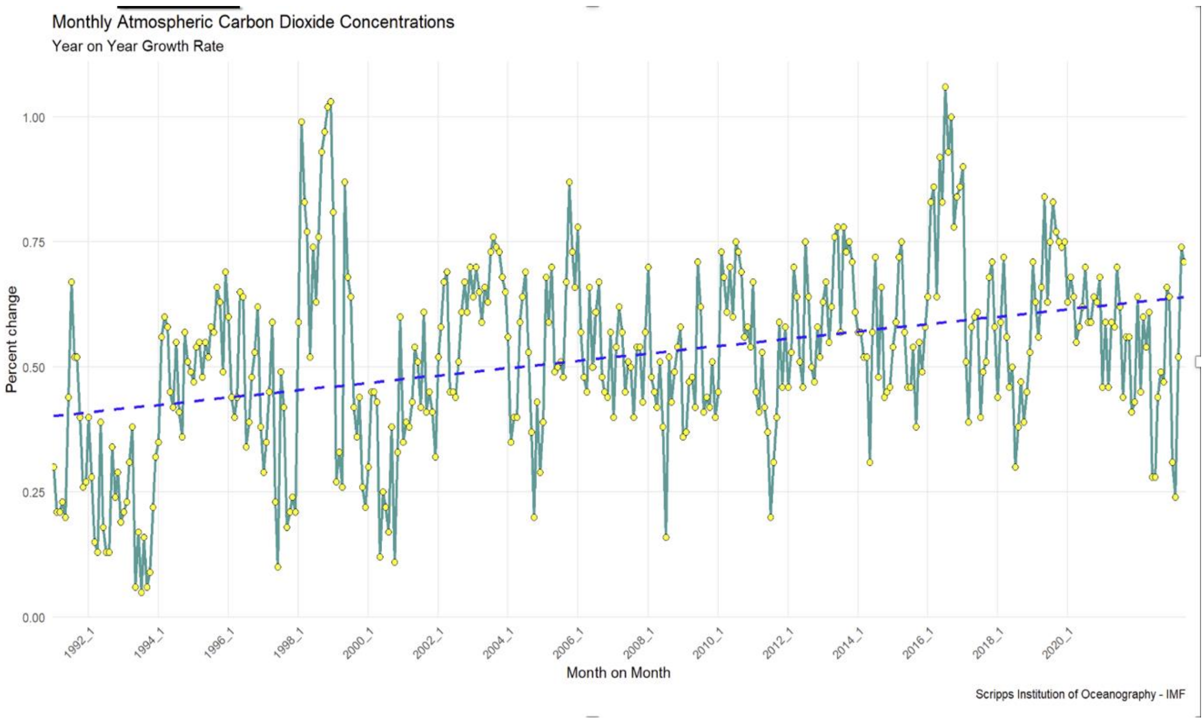
```
SELECT Year, Unit , value as emission, Date as date
FROM CO2world_wide
WHERE Unit IN ('Percent') AND Year >1990
ORDER BY Year;
```

CO2world_wide table is used for this query. WHERE and ORDER By clause is also used to get the required output.

Query Result:

	Year	Unit	emission	date
1	1991	Percent	0.3	1
2	1991	Percent	0.2	2
3	1991	Percent	0.44	3
4	1991	Percent	0.67	4
5	1991	Percent	0.52	5
6	1991	Percent	0.52	6
7	1991	Percent	0.4	7

Analysis and interpretation:



Result from the sql query is used to plot month wise percentage change of co2 concentration.
Key Takeaways:

- From the plot we can infer that atmospheric co2 concentration is on the rise.
- Co2 emission worldwide shows seasonal trend, where it drops in some months.

4. Average Temperature change over years

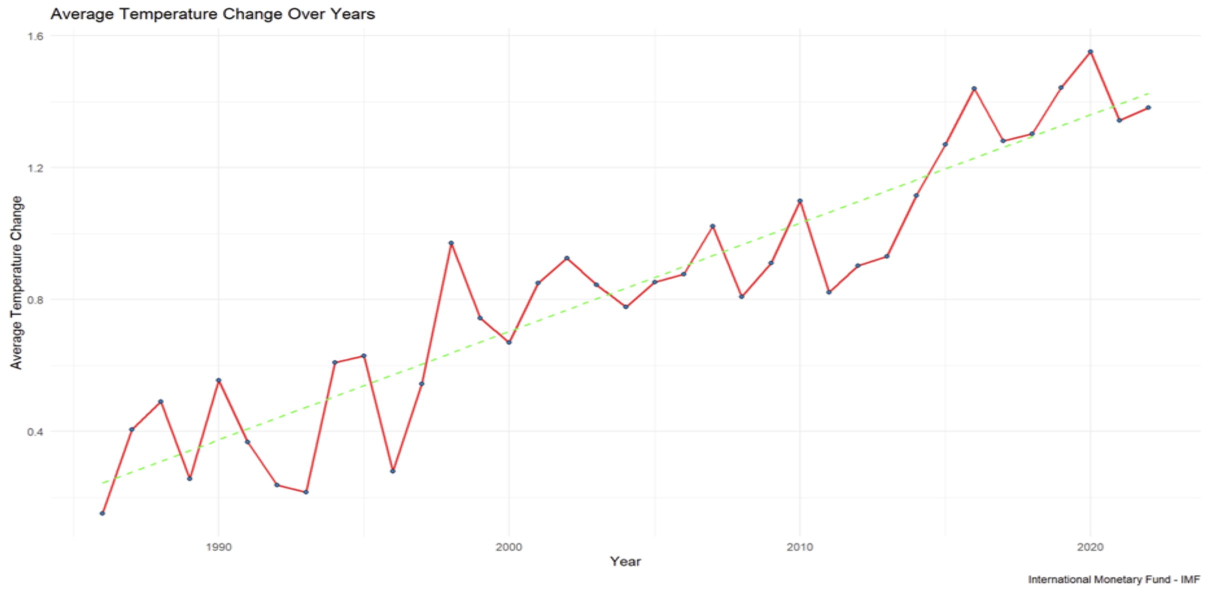
SQL Query:

```
SELECT Year, AVG(temp_change) as Temp_change
FROM surface_change
WHERE Year > 1985
GROUP BY year
ORDER BY year;
```

Temp_change table is used to filter yearly temperature change values.
Query Result:

	Year	Temp_change
1	1986	0.15115789473684219
2	1987	0.4050210526315791
3	1988	0.4900894736842106
4	1989	0.25616842105263155
5	1990	0.555259259259259
6	1991	0.36816489361702115
7	1992	0.23649038461538455

Analysis and interpretation:



Result from SQL query is stored as a csv, which is later read in R Studio to create line chart and chart showcasing temperature change over the years.

Key Takeaways:

- The world is warming up rapidly.
- Average temperature has increased by 1.36 degree Celsius globally since 1900's.

5. Average Temperature changes over years top 14 countries.

SQL Query:

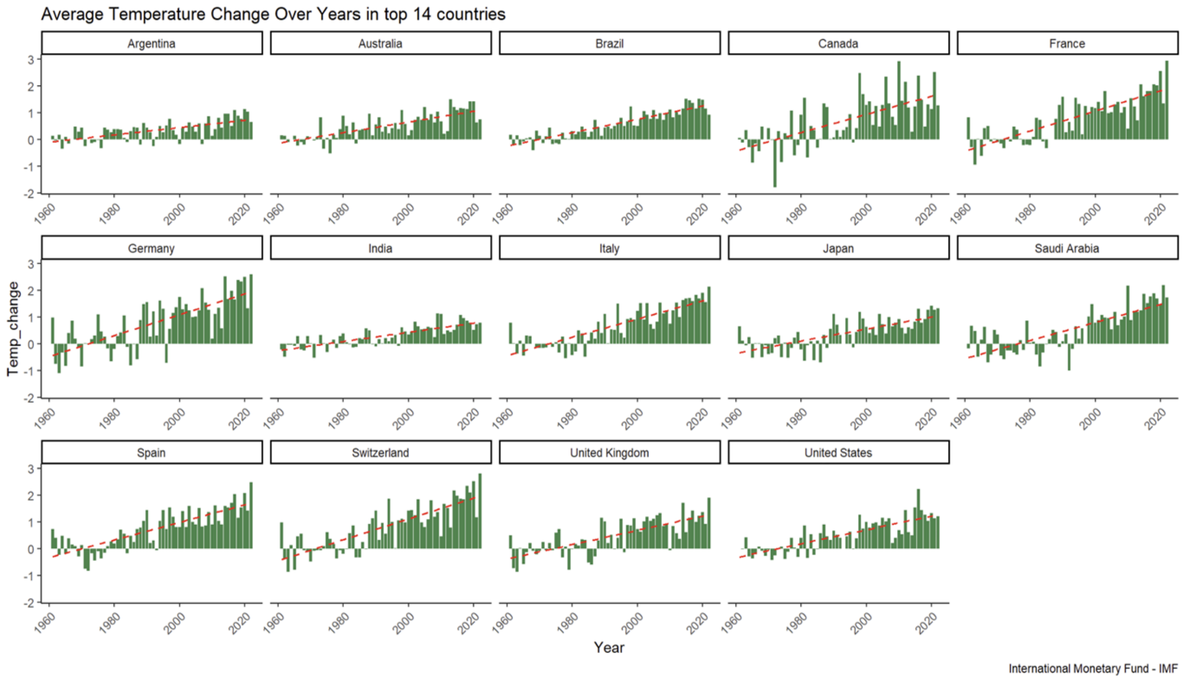
```
select c.Country, Year,AVG(temp_change) as tem_change From
surface_change
LEFT JOIN country_master  c on c.Country_code = scountry_code
WHERE c.Country in ('United States', 'China', 'Japan', 'Germany', 'United Kingdom',
    'India', 'France', 'Brazil', 'Italy', 'Canada', 'Australia',
    'Spain', 'Netherlands', 'Saudi Arabia', 'Switzerland', 'Argentina')
GROUP BY c.Country, Year
ORDER BY Year
```

In this query surface_change table is merged with country_master table using a left join. To create a plot of country wise surface temperature change of selected countries.

Query Result:

	Country	Year	tem_change
1	Argentina	1961	0.122
2	Australia	1961	0.157
3	Brazil	1961	0.167
4	Canada	1961	0.057
5	France	1961	0.827
6	Germany	1961	0.99
7	India	1961	-0.208

Analysis and interpretation:



All 14 countries in the graph have experienced an increase in average temperature over the past few decades.

Key Takeaways:

- The countries that have experienced the greatest warming are Argentina, Australia, Brazil, Canada, and the United States.

6. GDP and forest area: A countrywise analysis

SQL Query:

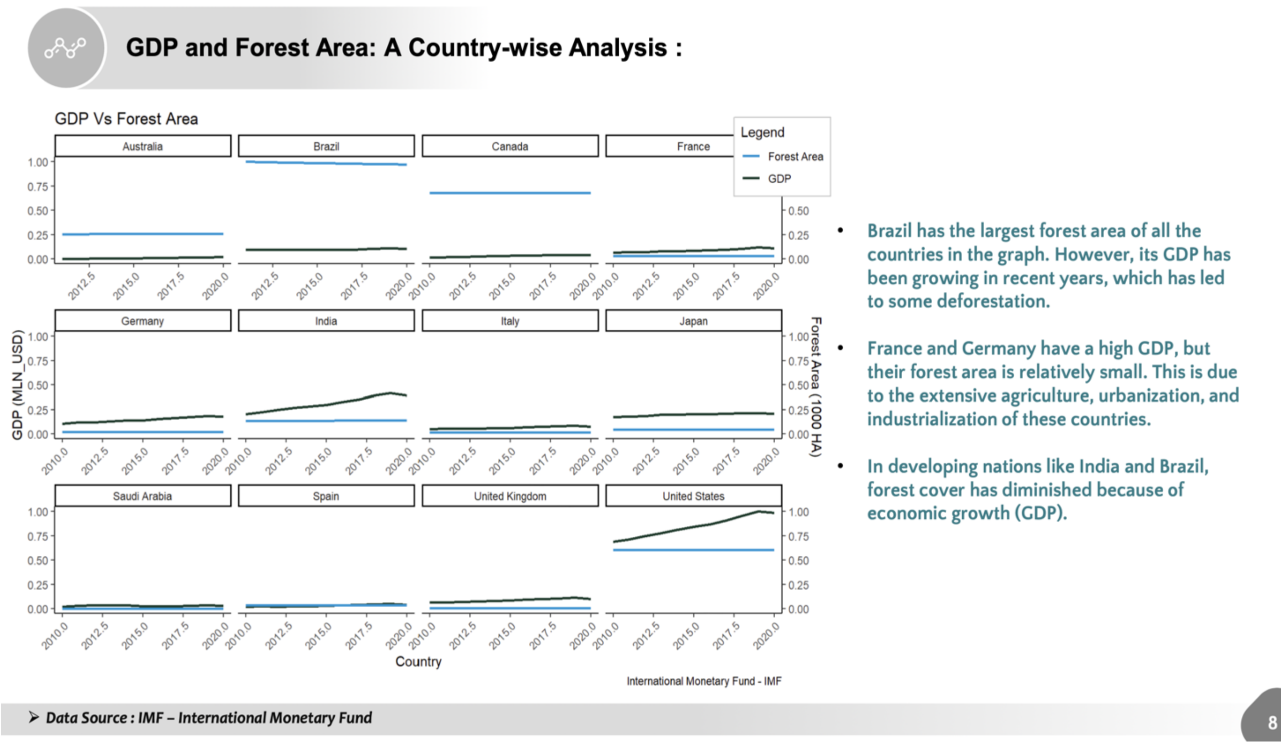
```
SELECT * FROM forest_change
select c.country ,forest_change.Year,forest_change.Indicator,AVG(emission) as Forest_change,
Sum(g.value) as gdp_value
from forest_change
LEFT JOIN country_master c on c.country_code = fcountry_code
LEFT JOIN GDP_VALUE g on (g.GCOUNTRY_CODE = fcountry_code AND g.TIME
=forest_change.Year)
where forest_change.Indicator In ('Forest area')
AND
c.Country IN (
    'United States', 'China', 'Japan', 'Germany', 'United Kingdom',
    'India', 'France', 'Brazil', 'Italy', 'Canada', 'Australia',
    'Spain', 'Netherlands', 'Saudi Arabia')
GROUP BY forest_change.Indicator,c.country,forest_change.Year
ORDER BY gdp_value
```

In this query forest_change, country_master and GDP_VALUE tables are merged using left join. Where, GROUP BY and ORDER BY clauses are used to create the required output.

Query Result:

	Id	Unit	Indicator	fcountry_code	Year	emission
1	1	Million tonnes	Carbon stocks in forests	AETMP	2010	56222.4507
2	2	1000 HA	Forest area	AETMP	2010	955178.5934
3	3	Index	Index of carbon stocks in forests	AETMP	2010	106.848883253871
4	4	Index	Index of forest extent	AETMP	2010	100.987268299979
5	5	1000 HA	Land area	AETMP	2010	2991130.036
6	6	Percent	Share of forest area	AETMP	2010	31.9337033797885
7	7	1000 HA	Forest area	AFG	2010	1208.44
8	8	Index	Index of forest extent	AFG	2010	100
9	9	1000 HA	Land area	AFG	2010	65223
10	10	Percent	Share of forest area	AFG	2010	1.85278199408184

Analysis and interpretation:



Result from SQL query is stored as a csv, which is later read in R Studio to create line chart. Year wise variation of Forest Area and GDP of 12 selected countries is plotted using Facetwrap feature of ggplot. Forest Area and GDP values are scaled between 0 and 1 using MinMax scaling. This allowed us to analyze the relative change of GDP and forest area.

Key Takeaways:

- worldwide forest area is on the decline.
- Increase in GDP accelerates deforestation, It is more evident in developing countries like India and Brazil where forest cover is reducing at higher pace than rest of the world.

7. Gas Emission Patterns Across the Regions

SQL Query:

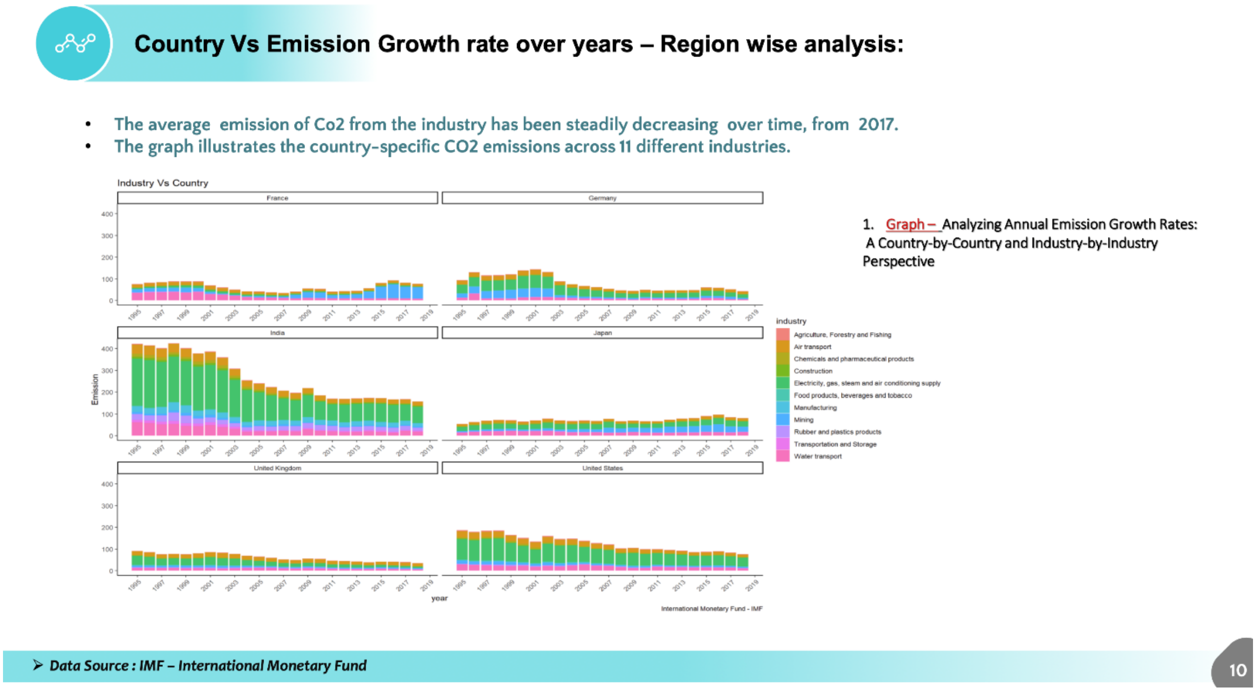
```
SELECT  country  ,gastype, Year,SUM(emission) as emission
FROM greenhouse
WHERE country IN('Asia','Africa','Americas','Europe','G20','Southern Europe','Advanced
Economies','Australia and New Zealand','G7','Southern Asia','Northern America')
GROUP BY country ,gastype, Year;
```

This query filter greenhouse for selected regions of the world. And emission value is aggregated by applying group by and sum.

Query Result:

	country	gastype	Year	emission
1	Northern America	Carbon dioxide	2010	12810.971499719999
2	Southern Europe	Nitrous oxide	2016	87.69140721400001
3	Asia	Carbon dioxide	2012	38772.223982200005
4	Africa	Methane	2013	2368.799659819
5	G7	Carbon dioxide	2014	19177.456568890004
6	Africa	Carbon dioxide	2020	2823.1374779149996
7	Advanced Economies	Methane	2012	2891.9163859029995
8	Southern Europe	Greenhouse gas	2013	2262.0041016699997
9	Americas	Greenhouse gas	2020	19902.0593437
10	Advanced Economies	Nitrous oxide	2013	1484.363794809

Analysis and interpretation:



Key Takeaways:

- Methane and C02 together contributes almost 80% of all greenhouse gas emission in almost all regions.
- Greenhouse gas emission from Asia has increased drastically over the years. While other parts of the world show a slow decline.

8. Emission Trend from Industry Across Selected Countries

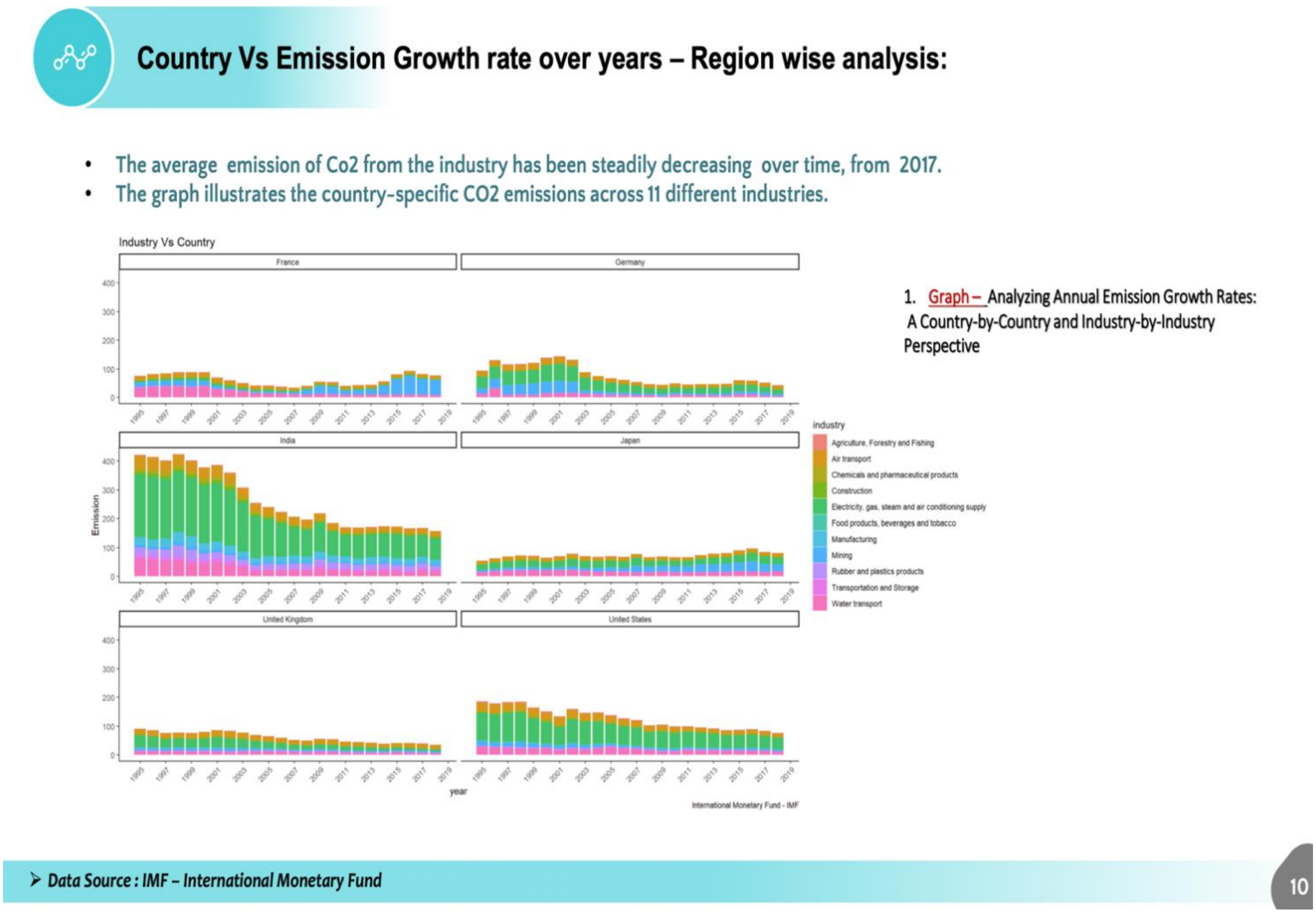
SQL Query:

```
SELECT c.Country,i.industry, Year, AVG(Value) as Emission
FROM CO2country_industry    -----Code changed need to resend
LEFT JOIN country_master as c ON c.country_code = CO2country_code
LEFT JOIN industry_master as i ON i.industry_code = co2Industry_code
WHERE c.Country IS NOT NULL
AND i.industry IN ('Agriculture, Forestry and Fishing','Construction',
'Electricity, gas, steam and air conditioning supply','Manufacturing','Mining',
'Transportation and Storage','Food products, beverages and tobacco','Air transport','Chemicals and
pharmaceutical products',
'Water transport','Rubber and plastics products')
AND unit IN ('Metric Tons of CO2 Emissions per $1million USD of output')
GROUP BY c.Country, i.industry, Year
ORDER BY Emission desc;
```

Query Result:

	Country	industry	Year	Emission
1	Kazakhstan, Rep. of	Rubber and plastics products	1995	267472.57935
2	Kazakhstan, Rep. of	Rubber and plastics products	1996	243890.43654999998
3	Kazakhstan, Rep. of	Rubber and plastics products	1997	180084.88069999998
4	Kazakhstan, Rep. of	Rubber and plastics products	1998	171415.6992
5	Kazakhstan, Rep. of	Rubber and plastics products	1999	129984.85070000001
6	Kazakhstan, Rep. of	Rubber and plastics products	2001	83538.167995
7	Kazakhstan, Rep. of	Electricity, gas, steam and air conditioning supply	1995	81937.88711499999
8	Kazakhstan, Rep. of	Electricity, gas, steam and air conditioning supply	2000	81256.440315
9	Kazakhstan, Rep. of	Electricity, gas, steam and air conditioning supply	1996	78749.72275
10	Kazakhstan, Rep. of	Electricity, gas, steam and air conditioning supply	1999	74194.66758499999

Analysis and interpretation:



Key Takeaways:

- Co2 Emission from various Industry has seen a decline over the years. It is more evident in India. Indian industries have reduced its Co2 emission by almost 50 percentage.

- Electricity, gas, steam, and air conditioning supply industry is the major contributor towards Co2 emission in almost all the countries.
- In France, Co2 emissions from mining has increased recently.

9. T test Independent group – High GDP vs Low GDP

SQL Query:

```
WITH GDP_Ranking AS (SELECT GCOUNTRY_CODE,
  RANK() OVER (ORDER BY SUM(GDP_VALUE.value) DESC) AS GDP_Rank
FROM
  GDP_VALUE
  GROUP BY
   GCOUNTRY_CODE)

SELECT C.Country, GDP_VALUE.GCOUNTRY_CODE,  SUM(GDP_VALUE.value) / 10 AS
gdp_value,SUM(co2country_industry.[Value]) AS Co2_emission,
CASE
 WHEN GDP_Rank <= 14 THEN 'HighGDP'
  ELSE 'LowGDP'
  END AS GDP_Group
  FROM
 GDP_VALUE
LEFT JOIN country_master C ON C.Country_code = GDP_VALUE.GCOUNTRY_CODE
LEFT JOIN co2country_industry ON co2country_industry.co2country_code =
GDP_VALUE.GCOUNTRY_CODE
LEFT JOIN GDP_Ranking ON GDP_VALUE.GCOUNTRY_CODE =
GDP_Ranking.GCOUNTRY_CODE
WHERE
   C.Country IS NOT NULL
GROUP BY
   C.Country, GDP_VALUE.GCOUNTRY_CODE, GDP_Rank
ORDER BY
  gdp_value DESC;
```

This query group the country as 'HighGDP' and 'LowGDP' countries based on sum of historical GDP values. Top 14 countries with highest cumulative sum is grouped into 'HighGDP' and rest of the countries are grouped into 'LowGDP' group.

Query Result:

	Country	GCOUNTRY_CODE	gdp_value	Co2_emission	GDP_Group
1	United States	USA	165810640758.00034	114251214.34599893	HighGDP
2	China, P.R.: Mainland	CHN	115007674278.9261	303156812.1552484	HighGDP
3	Japan	JPN	51627751445.28837	67719822.58772963	HighGDP
4	India	IND	44530268501.86546	166790528.6489276	HighGDP
5	Germany	DEU	37868745088.195595	65360220.01101689	HighGDP
6	France	FRA	26801667062.5123	64959677.45156631	HighGDP
7	United Kingdom	GBR	26437987285.22936	67684337.03822432	HighGDP
8	Italy	ITA	24066276818.434433	46846908.66195505	HighGDP
9	Russian Federation	RUS	21194240511.969284	188448836.48562482	HighGDP
10	Brazil	BRA	19897563654.586926	38892841.5419856	HighGDP
11	Indonesia	IDN	19633853836.77169	93074947.75971672	LowGDP
12	Mexico	MEX	18852257772.285774	110717296.83043173	LowGDP
13	Canada	CAN	15579156010.533731	125286697.10999851	LowGDP
14	Spain	ESP	15394842316.178143	93015534.36064012	LowGDP
15	Korea, Rep. of	KOR	15277691692.291157	118776535.72640142	LowGDP
16	Turkey	TUR	14759115521.821955	98302703.68232511	LowGDP
17	Saudi Arabia	SAU	14453859512.821753	368147673.0498141	LowGDP
18	Australia	AUS	10476871480.901333	131227245.83330497	LowGDP
19	Netherlands, The	NLD	8599188745.332903	72811598.76732506	LowGDP
20	Poland, Rep. of	POL	7648748115.735766	102602386.4118671	LowGDP
21	South Africa	ZAF	6968820393.342639	340388150.24388325	LowGDP

Analysis and interpretation:

Independent Samples T-Test

	t	df	p	Cohen's d	SE Cohen's d
Co2 Emission	1.055	52	0.296	0.357	0.346
GDP	6.342	52	< .001[a]	2.143	0.568

Note. Student's t-test.

[a] Brown-Forsythe test is significant ($p < .05$), suggesting a violation of the equal variance assumption

Assumption Checks ▼

Test of Equality of Variances (Brown-Forsythe)

	F	df_1	df_2	p
Co2 Emission	0.560	1	52	0.458
GDP	13.032	1	52	< .001

Descriptives

Group Descriptives

	Group	N	Mean	SD	SE	Coefficient of variation
Co2 Emission	HighGDP	11	$1.107\times10^{+8}$	$7.977\times10^{+7}$	$2.405\times10^{+7}$	0.721
	LowGDP	43	$8.568\times10^{+7}$	$6.750\times10^{+7}$	$1.029\times10^{+7}$	0.788
GDP	HighGDP	11	$5.026\times10^{+10}$	$4.716\times10^{+10}$	$1.422\times10^{+10}$	0.938
	LowGDP	43	$4.923\times10^{+9}$	$4.976\times10^{+9}$	$7.589\times10^{+8}$	1.011

This is the result of Indepent sample T-Test obtained using Jasp.

Null hypothesis: *Historical Co2 Emission from 'HighGDP' countries and 'LowGDP' countries are same*

Alternative hypothesis*: Historical Co2 Emission from 'HighGDP' countries and 'LowGDP' countries are significantly different.*

From the independent samples t-test result the p-value for Co2 Emission is .296, which is greater the .05. Hence, we accept null hypothesis. That means, historical Co2 emission from top 14 countries by GDP is not different from rest of the world.

10. Conclusion

- From the detailed analysis, we can conclude that historical Co2 emission from top 14 countries by GDP is not different from rest of the world. Top 14 GDP countries and rest of the world are equally accountable for the environmental consequences related to Co2 emission.

- Increase in GDP worldwide has directed impacted forest area worldwide. It is clear from data for developing countries, economic progress has an adverse effect on forest area.

- Co2 emission from Asia in on a steep rise, while other parts of the world show a slow decline.

- From the plots, it's clear that average temperature has risen by 1.36 degree Celsius from 1900's. This has serious consequence on the climate. If the world warms up by 1.5 degree Celsius, according to scientist and experts, the climate change is irreversible.