PSTAT 100 Course Project

December 13, 2023

1 PSTAT 100 Course Project Global Emissions Report

by Michael Chen, Sayan Andrews, Dimitri Manolis, Anjan Chakravarti

1.1 Background

The data set we are going to dive into is the Historical Emissions data set. The source of the data is Climate Watch. Climate Watch is a platform developed by the World Resources Institute (WRI) to provide open access to a wide range of climate-related data. Climate Watch collects its data from a variety of sources, including official reports submitted by countries, international organizations, research institutions, and other reputable sources. The historical emissions data set has 195 observations. Each observation is associated with a certain region in the world. Some of these regions are as specific as countries, and others are continents or the world as a whole. We tidied the data set however, and we're left with 5,850 observations because we created a column for the years rather than having each row with 30 columns. The data set records all greenhouse gasses for each observation, and the total emissions recorded are from all sectors including LUCF (Land Change and Forestry). The unit of measurement used to record the amount of emissions is "Metric tons of carbon dioxide equivalent". For each of these observations there is the total amount of greenhouse gasses in Metric tons of carbon dioxide equivalent per year from the years 1990 - 2019 inclusive.

There is one n/a value in the data set, it being from Namibia in 1990.

[39]:
```
import numpy as np

import pandas as pd

import altair as alt

import statsmodels.api as sm

climate = pd.read_csv('data/historical_emissions/historical_emissions.csv')

climate.head()
```

[39]:
```
                Country    Data source                 Sector      Gas    Unit  \
0                 World  Climate Watch  Total including LUCF  All GHG  MtCO e
1                 China  Climate Watch  Total including LUCF  All GHG  MtCO e
2         United States  Climate Watch  Total including LUCF  All GHG  MtCO e
```

```
3               India  Climate Watch  Total including LUCF  All GHG  MtCO e
4  European Union (27)  Climate Watch  Total including LUCF  All GHG  MtCO e

       2019      2018      2017      2016      2015  …      1999      1998  \
0  49758.23  49368.04  48251.88  47531.68  46871.77  …  35101.90  35099.21
1  12055.41  11821.66  11385.48  11151.31  11108.86  …   4028.58   4095.97
2   5771.00   5892.37   5689.61   5743.85   5665.21  …   6210.12   6208.83
3   3363.60   3360.56   3215.07   3076.48   3003.07  …   1440.38   1362.33
4   3149.57   3295.53   3379.38   3364.77   3019.49  …   3874.40   3949.25

       1997      1996      1995      1994      1993      1992      1991  \
0  35537.18  34179.33  33805.61  33015.04  32729.06  32588.09  32670.51
1   3977.65   3982.11   3960.71   3557.37   3397.80   3168.05   3039.15
2   6160.86   5901.00   5729.69   5661.57   5567.55   5456.12   5372.08
3   1331.88   1272.74   1223.65   1158.48   1114.22   1081.28   1056.25
4   3983.29   4058.46   3947.64   3894.63   3908.95   3981.08   4120.94

       1990
0  32523.58
1   2891.73
2   5417.32
3   1002.56
4   4187.90

[5 rows x 35 columns]
```

[40]:
```
climate = climate.drop(columns = ['Data source', 'Sector', 'Gas', 'Unit'])

climate
```

[40]:
```
                 Country      2019      2018      2017      2016      2015  \
0                  World  49758.23  49368.04  48251.88  47531.68  46871.77
1                  China  12055.41  11821.66  11385.48  11151.31  11108.86
2          United States   5771.00   5892.37   5689.61   5743.85   5665.21
3                  India   3363.60   3360.56   3215.07   3076.48   3003.07
4    European Union (27)   3149.57   3295.53   3379.38   3364.77   3019.49
..                   …       …         …         …         …         …
190         Cook Islands      0.10      0.10      0.10      0.09      0.08
191                Nauru      0.07      0.07      0.07      0.07      0.06
192               Tuvalu      0.03      0.03      0.02      0.02      0.02
193                 Niue      0.01      0.01      0.01      0.01      0.01
194                 Fiji     -0.16     -0.24     -0.22     -0.41     -0.54

         2014      2013      2012      2011  …      1999      1998  \
0    46881.78  46238.61  45597.14  45041.81  …  35101.90  35099.21
1    11228.48  11168.26  10675.66  10388.48  …   4028.58   4095.97
2     5779.54   5734.28   5593.25   5811.96  …   6210.12   6208.83
```

```
3     2984.52   2804.34   2740.40   2584.75  …   1440.38   1362.33
4     2962.68   3109.55   3188.85   3250.07  …   3874.40   3949.25
..        …         …         …         …  …         …         …
190      0.09      0.09      0.09      0.09  …      0.05      0.05
191      0.06      0.05      0.05      0.04  …      0.10      0.10
192      0.02      0.02      0.02      0.02  …      0.02      0.02
193      0.00      0.00      0.00      0.00  …      0.01      0.01
194     -0.66     -0.25     -0.37     -0.29  …     -0.63     -0.60

          1997      1996      1995      1994      1993      1992      1991  \
0    35537.18  34179.33  33805.61  33015.04  32729.06  32588.09  32670.51
1     3977.65   3982.11   3960.71   3557.37   3397.80   3168.05   3039.15
2     6160.86   5901.00   5729.69   5661.57   5567.55   5456.12   5372.08
3     1331.88   1272.74   1223.65   1158.48   1114.22   1081.28   1056.25
4     3983.29   4058.46   3947.64   3894.63   3908.95   3981.08   4120.94
..        …         …         …         …         …         …
190      0.05      0.05      0.04      0.04      0.04      0.03      0.03
191      0.11      0.11      0.11      0.12      0.12      0.12      0.13
192      0.02      0.02      0.02      0.02      0.02      0.02      0.02
193      0.01      0.01      0.01      0.01      0.01      0.01      0.01
194     -0.56     -0.56     -0.60     -0.66     -0.71     -0.77     -0.81

          1990
0    32523.58
1     2891.73
2     5417.32
3     1002.56
4     4187.90
..        …
190      0.03
191      0.13
192      0.02
193      0.01
194     -0.82

[195 rows x 31 columns]
```

[41]:
```
melted_climate = pd.melt(climate, id_vars=['Country'], var_name='Year',␣
↪value_name='Emissions')
melted_climate
```

[41]:
```
                 Country  Year  Emissions
0                  World  2019   49758.23
1                  China  2019   12055.41
2          United States  2019    5771.00
3                  India  2019    3363.60
4     European Union (27)  2019    3149.57
```

```
…                    …    …         …
5845         Cook Islands  1990      0.03
5846                Nauru  1990      0.13
5847               Tuvalu  1990      0.02
5848                 Niue  1990      0.01
5849                 Fiji  1990     -0.82

[5850 rows x 3 columns]
```

[42]:
```
missing_values = melted_climate.isnull().sum()
print(missing_values)
```

```
Country      0
Year         0
Emissions    1
dtype: int64
```

[43]:
```
index_of_missing_value = melted_climate.loc[melted_climate['Emissions'].
 ↪isnull()].index

print(index_of_missing_value)
```

```
Index([5777], dtype='int64')
```

[44]:
```
missing_row = melted_climate.loc[index_of_missing_value]


print("Country:", missing_row['Country'].values[0])
print("Year:", missing_row['Year'].values[0])
```

```
Country: Namibia
Year: 1990
```

[46]:
```
melted_climate['Year'] = melted_climate['Year'].astype(int)

sorted_climate = melted_climate.sort_values(by='Year')

sorted_climate.reset_index(drop=True, inplace=True)

sorted_climate
```

[46]:
```
           Country  Year  Emissions
0             Fiji  1990      -0.82
1            Kenya  1990      30.65
2          Romania  1990     224.74
3           Greece  1990      94.31
4     North Korea  1990     161.17
```

```
…            …   …         …
5845    Lithuania  2019      18.30
5846       Malawi  2019      19.34
5847        Gabon  2019      19.68
5848      Namibia  2019      21.22
5849        World  2019   49758.23

[5850 rows x 3 columns]
```

Variable name	Description
`Country`	Name of Country
`Year`	Year That Data Corresponds To
`Emissions`	Metric tons of Carbon Dioxide Emission

1.2 Inquiries

Using this data set we will be able to come to impactful conclusions about the world as it relates to greenhouse gas emissions. We will be able to examine how emissions have changed across different parts of the world over thirty years.

More specifically we will be asking the following questions:

How have the world's biggest emissions changed over the years, specifically India, USA, China, Russia, European Union ?

How do countries within Europe compare to one another over the years?

How do countries within North America compare to one another over the years?

How do countries in South America compare to one another over the years?

How do countries in Asia compare to one another over the years?

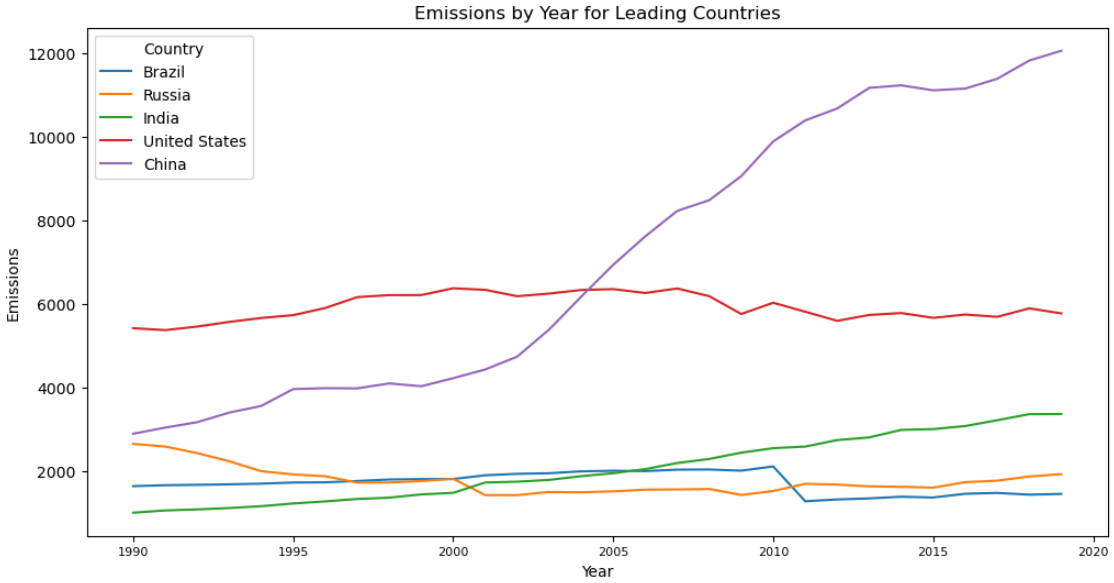
How do countries in Africa compare to one another over the years?

How do all the continents stack up against one another? Asia, Africa, Australia, North America, South America, and Europe?

1.3 How have the world's biggest emissions changed over the years, specifically India, USA, China, Russia, European Union?

The USA, Russia, and the European Union have similar trend lines with a relatively flat slope. This indicates that from 1990 to 2019, these nations have had a similar amount of industrial activity with no large changes over time. This makes sense as these nations have been industrialized and developed for a longer time and haven't had drastic growth. China has an extremely large growth going from 3000 to almost 12000 in just 30 years. This makes sense as China has had an extreme growth in developing which makes it now the leading country in emissions worldwide. India has also had a steady increase in emissions as well with quite a linear trend. Russia and Brazil have both appeared to have a slight decrease over time but with steep drops rather than a slow trend. China is almost double the activity of the second greatest country which is the US.

[29]:
```
leading_countries = [
    'United States', 'Russia', 'China', 'India', 'Brazil', 'European Union'
]

leading_data = sorted_climate[sorted_climate['Country'].isin(leading_countries)]

plt.figure(figsize=(12, 6))
sns.lineplot(x='Year', y='Emissions', hue='Country', data=leading_data)
plt.title('Emissions by Year for Leading Countries')
plt.xlabel('Year')
plt.ylabel('Emissions')
plt.legend(title='Country')
plt.xticks(fontsize=8)
plt.show()
```

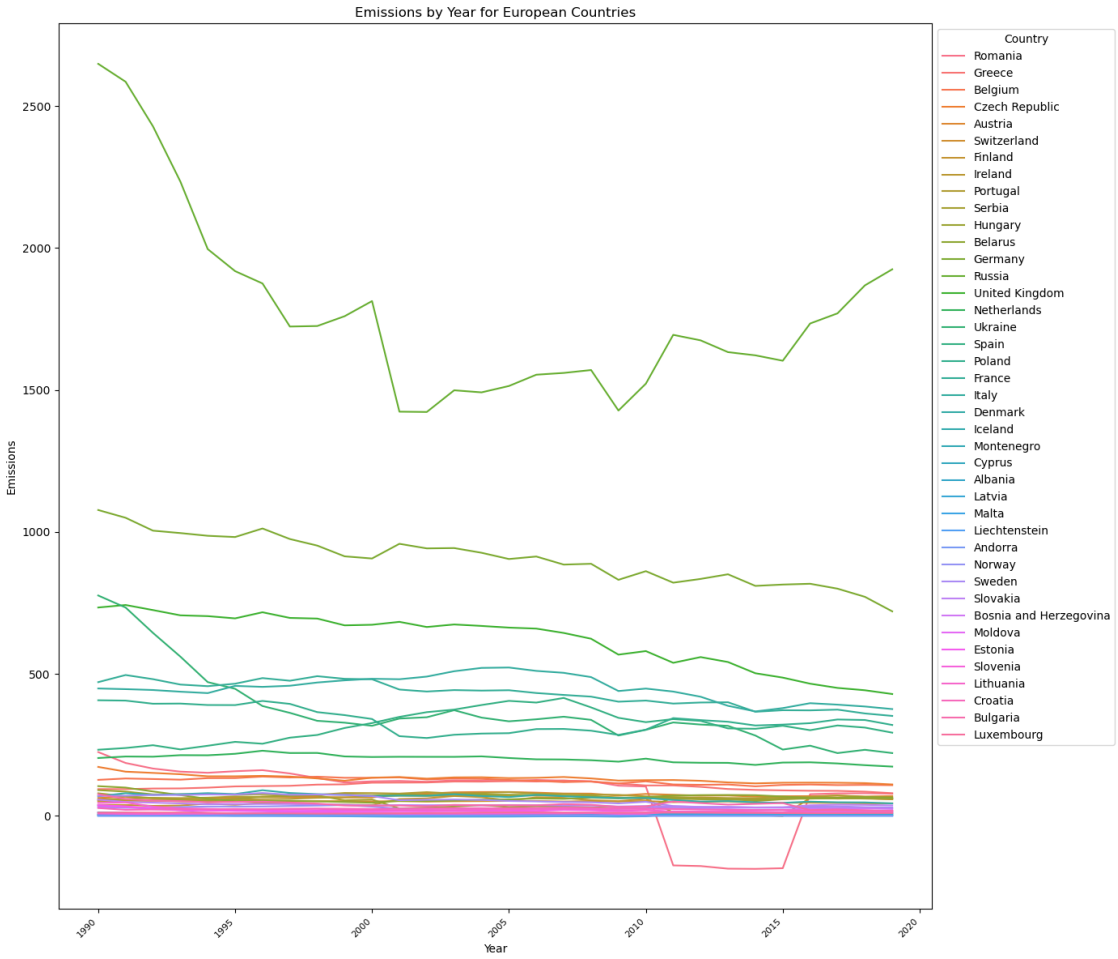


1.4 How do countries within Europe compare to one another over the years?

Out of all countries geographically inEurope (not European Union) Russia is the largest contributor to emissions. Considering Russia has the greatest population of all these countries it is no surprise. They are also a country with a lot of industry. Overall though, the european countries emissions actually have a slight negative trend. This is good and means that Europe has strong sustainable initiatives.

[16]:
```
european_countries = [
    'Albania', 'Andorra', 'Austria', 'Belarus', 'Belgium', 'Bosnia and␣
 ↪Herzegovina',
    'Bulgaria', 'Croatia', 'Cyprus', 'Czech Republic', 'Denmark', 'Estonia',␣
 ↪'Finland',
```

```
    'France', 'Germany', 'Greece', 'Hungary', 'Iceland', 'Ireland', 'Italy',␣
 ↪'Kosovo',
    'Latvia', 'Liechtenstein', 'Lithuania', 'Luxembourg', 'Malta', 'Moldova',␣
 ↪'Monaco',
    'Montenegro', 'Netherlands', 'North Macedonia', 'Norway', 'Poland',␣
 ↪'Portugal',
    'Romania', 'Russia', 'San Marino', 'Serbia', 'Slovakia', 'Slovenia',␣
 ↪'Spain',
    'Sweden', 'Switzerland', 'Ukraine', 'United Kingdom', 'Vatican City'
]


european_data = sorted_climate[sorted_climate['Country'].
 ↪isin(european_countries)]

plt.figure(figsize=(14, 12))  # Increase the height of the figure
sns.lineplot(x='Year', y='Emissions', hue='Country', data=european_data,␣
 ↪legend='full')
plt.title('Emissions by Year for European Countries')
plt.xlabel('Year')
plt.ylabel('Emissions')


plt.xticks(fontsize=8, rotation=45, ha='right')


plt.legend(title='Country', bbox_to_anchor=(1, 1), loc='upper left')

plt.tight_layout()
plt.show()
```

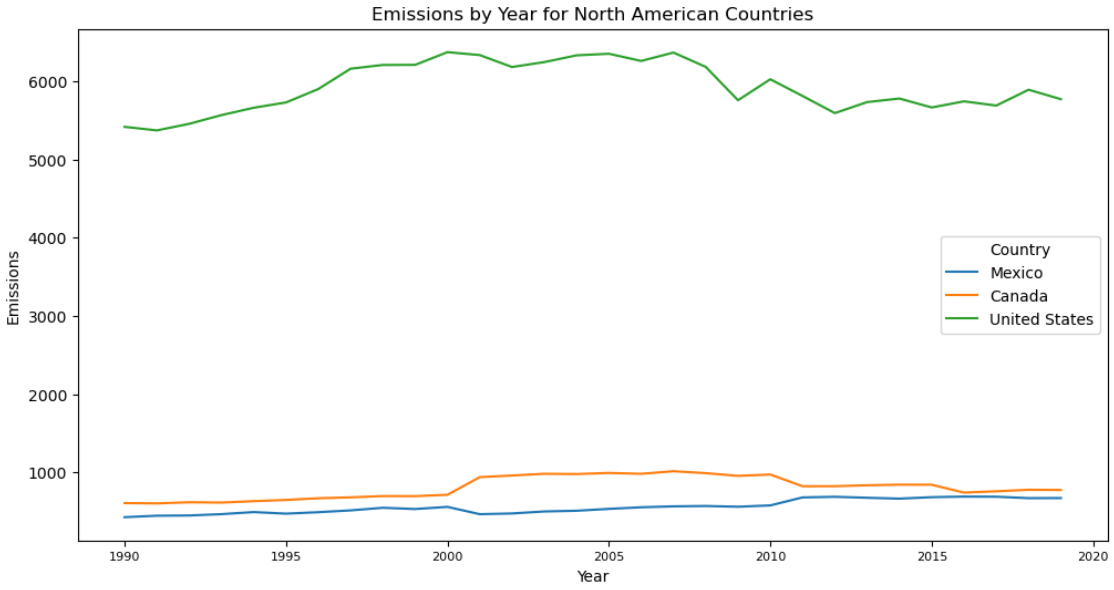


1.5 How do countries within North America compare to one another over the years?

The United States is far and away the front runner when it comes to greenhouse gas emissions in North America. Being a country with 331.9 million people as well as being an industrial powerhouse this is no surprise. Compared to Canada which has a population of 38 million the discrepancy is no surprise. Something surprising though is that Mexico, which has a population of 126.7 million people, has the least emissions. This could be due to sampling errors, Mexico being somewhat of a third world country. Especially considering that garbage burning is common practice in Mexico this seems suspicious. Another thing to note is that the emissions for all of these countries are slightly trending in a positive direction. This is an indication of a lack of sustainable initiative.

[15]:
```
import matplotlib.pyplot as plt
import seaborn as sns

north_american_countries = ['Canada', 'United States', 'Mexico']
```

```
north_american_data = sorted_climate[sorted_climate['Country'].
  ↪isin(north_american_countries)]

plt.figure(figsize=(12, 6))
sns.lineplot(x='Year', y='Emissions', hue='Country', data=north_american_data)
plt.title('Emissions by Year for North American Countries')
plt.xlabel('Year')
plt.ylabel('Emissions')
plt.legend(title='Country')
plt.xticks(fontsize=8)
plt.show()
```

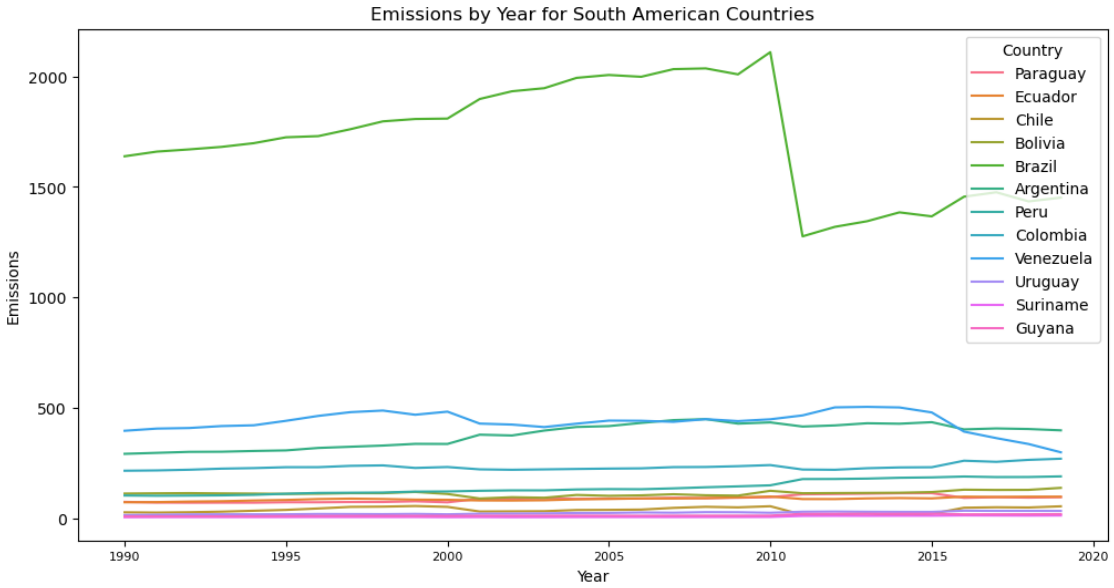


1.6 How do countries in South America compare to one another over the years?

Brazil is the frontrunner in South America when it comes to emissions in South America. Of course this is no surprise considering it is the most populated country in South America (214 million) followed by Columbia with a population of 52 million people. There is a curious feature in this data where there is a sudden drop in emissions for Brazil from the year 2010 - 2011. This seems unnatural. Overall for the countries in South America their emissions seem to be trending positively very slightly. An indication of lack of sustainable initiative.

[17]:
```
south_american_countries = [
    'Argentina', 'Bolivia', 'Brazil', 'Chile', 'Colombia', 'Ecuador', 'Guyana',
    'Paraguay', 'Peru', 'Suriname', 'Uruguay', 'Venezuela'
]
```

```
south_american_data = sorted_climate[sorted_climate['Country'].
 ↪isin(south_american_countries)]


plt.figure(figsize=(12, 6))
sns.lineplot(x='Year', y='Emissions', hue='Country', data=south_american_data)
plt.title('Emissions by Year for South American Countries')
plt.xlabel('Year')
plt.ylabel('Emissions')
plt.legend(title='Country')
plt.xticks(fontsize=8)
plt.show()
```

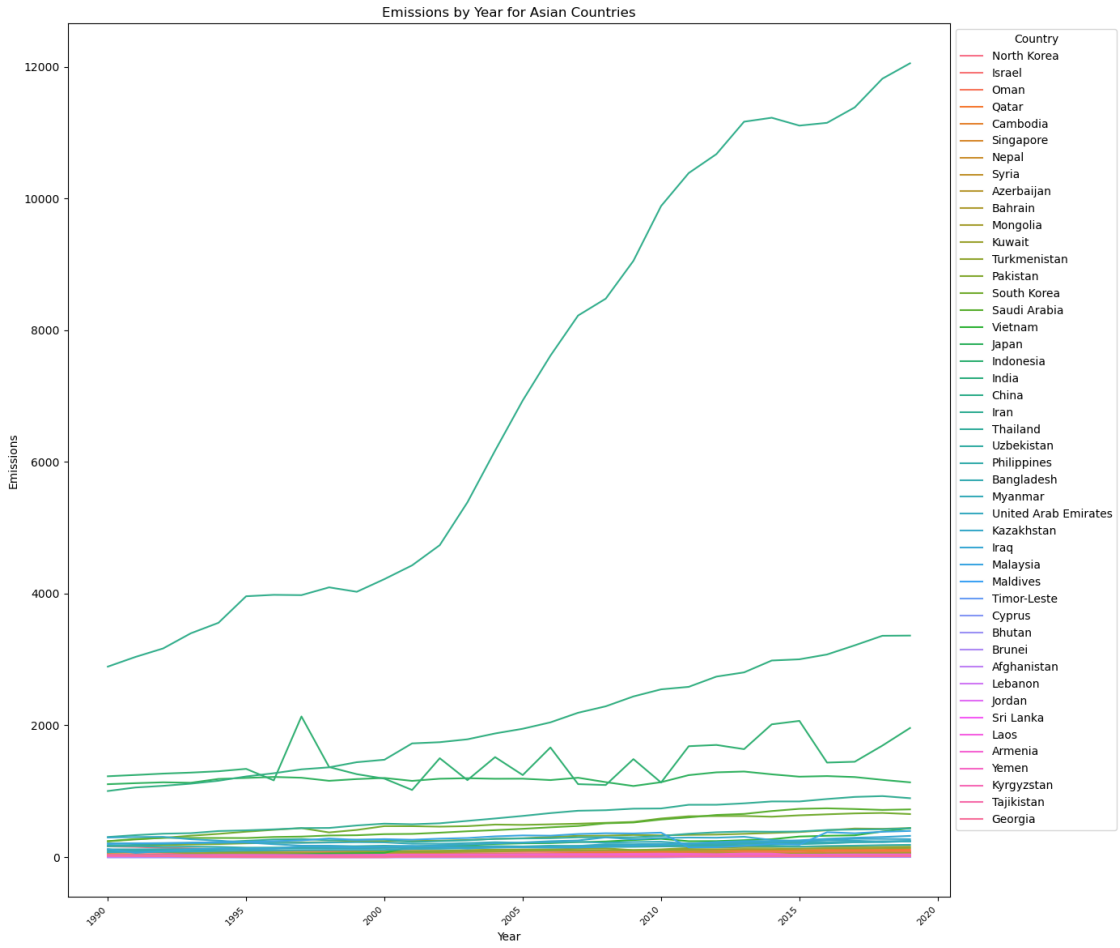


1.7 How do countries in Asia compare to one another over the years?

The country in Asia with the most emissions is China(population 1.4 billions). With almost 12,000 metric tons of carbon dioxide equivalent. Second was India(population 1.4 billion). Had a little over 2,000 metric tons of carbon dioxide equivalent. It is interesting here that two countries with very similar populations had such stark differences in emissions. This highlights how bad China's greenhouse gas emissions are. Because India is not necessarily a "good" country in this sense yet they are blown out of the water by China. In addition a majority of these Asian countries display an extreme positive trend. This is a sign of blatant disregard for sustainable initiatives.

[19]:
```
asian_countries = [
    'Afghanistan', 'Armenia', 'Azerbaijan', 'Bahrain', 'Bangladesh', 'Bhutan',␣
 ↪'Brunei',
```

```
    'Cambodia', 'China', 'Cyprus', 'Georgia', 'India', 'Indonesia', 'Iran',␣
 ↪'Iraq', 'Israel',
    'Japan', 'Jordan', 'Kazakhstan', 'Kuwait', 'Kyrgyzstan', 'Laos', 'Lebanon',␣
 ↪'Malaysia',
    'Maldives', 'Mongolia', 'Myanmar', 'Nepal', 'North Korea', 'Oman',␣
 ↪'Pakistan', 'Palestine',
    'Philippines', 'Qatar', 'Saudi Arabia', 'Singapore', 'South Korea', 'Sri␣
 ↪Lanka', 'Syria',
    'Taiwan', 'Tajikistan', 'Thailand', 'Timor-Leste', 'Turkey',␣
 ↪'Turkmenistan', 'United Arab Emirates',
    'Uzbekistan', 'Vietnam', 'Yemen'
]

asian_data = sorted_climate[sorted_climate['Country'].isin(asian_countries)]


plt.figure(figsize=(14, 12))

sns.lineplot(x='Year', y='Emissions', hue='Country', data=asian_data,␣
 ↪legend='full')

plt.title('Emissions by Year for Asian Countries')

plt.xlabel('Year')

plt.ylabel('Emissions')


plt.xticks(fontsize=8, rotation=45, ha='right')


plt.legend(title='Country', bbox_to_anchor=(1, 1), loc='upper left')

plt.tight_layout()

plt.show()
```

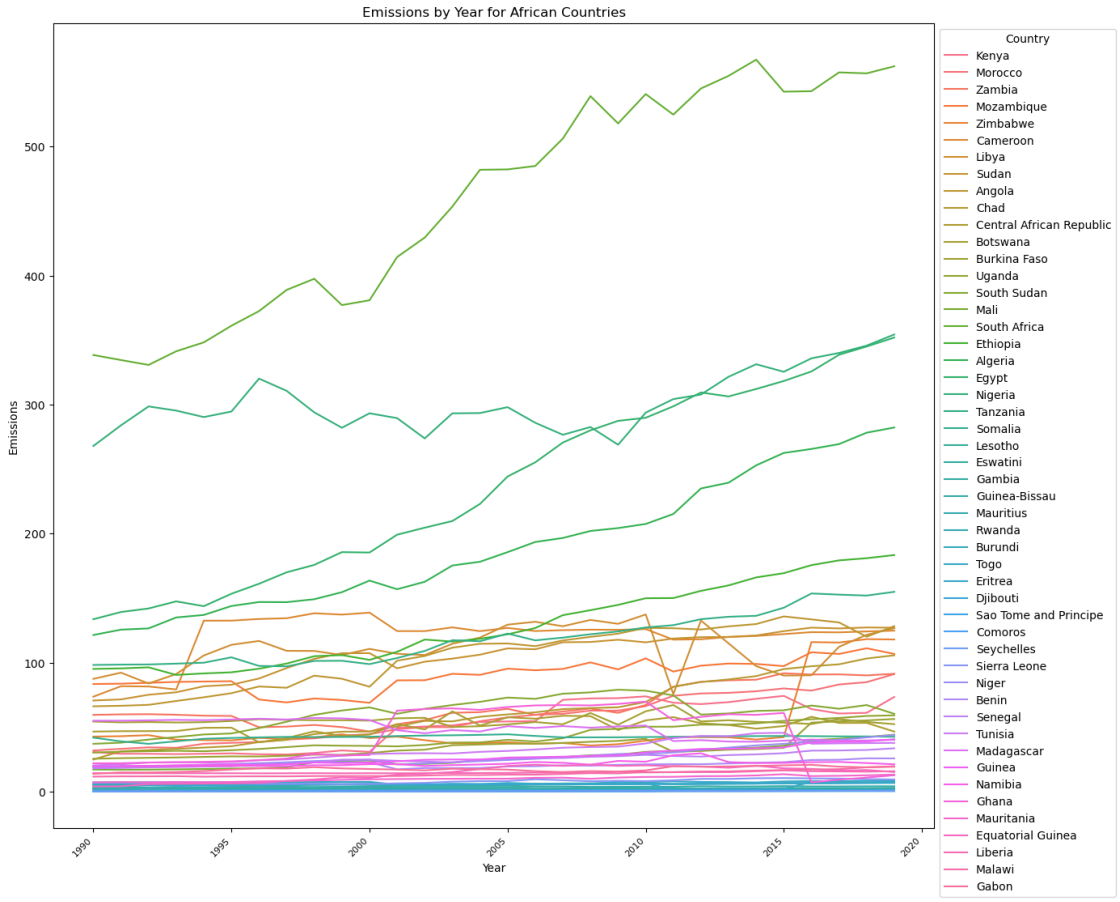


1.8 How do countries in Africa compare to one another over the years?

Based on the graph of Emissions by Year for African Countries, we can see that countries with historically greater emissions have increasing values as the year becomes later while countries with little emissions historically have little change. South Africa has by far the highest emissions of any country and has increased from about 350 to almost 550 from 1990 until the present. Egypt is the second highest and appears to have stayed relatively constant over the years. The 3rd and 4th countries of Nigeria and Algeria have extremely large increases of almost 150 units. Another takeaway is that the majority of the countries in Africa have very low emissions and little change over time indicating a lack of industrial progress as high emissions are often a sign of developed nations with fully functioning transportation networks and industrial activity.

[28]:
```
african_countries = [
    'Algeria', 'Angola', 'Benin', 'Botswana', 'Burkina Faso', 'Burundi', 'Cabo␣
↪Verde', 'Cameroon',
    'Central African Republic', 'Chad', 'Comoros', 'Congo', 'Djibouti',␣
↪'Egypt', 'Equatorial Guinea',
```

```
    'Eritrea', 'Eswatini', 'Ethiopia', 'Gabon', 'Gambia', 'Ghana', 'Guinea',␣
 ↪'Guinea-Bissau', 'Ivory Coast',
    'Kenya', 'Lesotho', 'Liberia', 'Libya', 'Madagascar', 'Malawi', 'Mali',␣
 ↪'Mauritania', 'Mauritius',
    'Morocco', 'Mozambique', 'Namibia', 'Niger', 'Nigeria', 'Rwanda', 'Sao Tome␣
 ↪and Principe', 'Senegal',
    'Seychelles', 'Sierra Leone', 'Somalia', 'South Africa', 'South Sudan',␣
 ↪'Sudan', 'Tanzania', 'Togo',
    'Tunisia', 'Uganda', 'Zambia', 'Zimbabwe'
]

african_data = sorted_climate[sorted_climate['Country'].isin(african_countries)]

plt.figure(figsize=(14, 12))  # Increase the height of the figure


sns.lineplot(x='Year', y='Emissions', hue='Country', data=african_data,␣
 ↪legend='full')

plt.title('Emissions by Year for African Countries')

plt.xlabel('Year')


plt.ylabel('Emissions')


plt.xticks(fontsize=8, rotation=45, ha='right')


plt.legend(title='Country', bbox_to_anchor=(1, 1), loc='upper left')


plt.tight_layout()

plt.show()
```



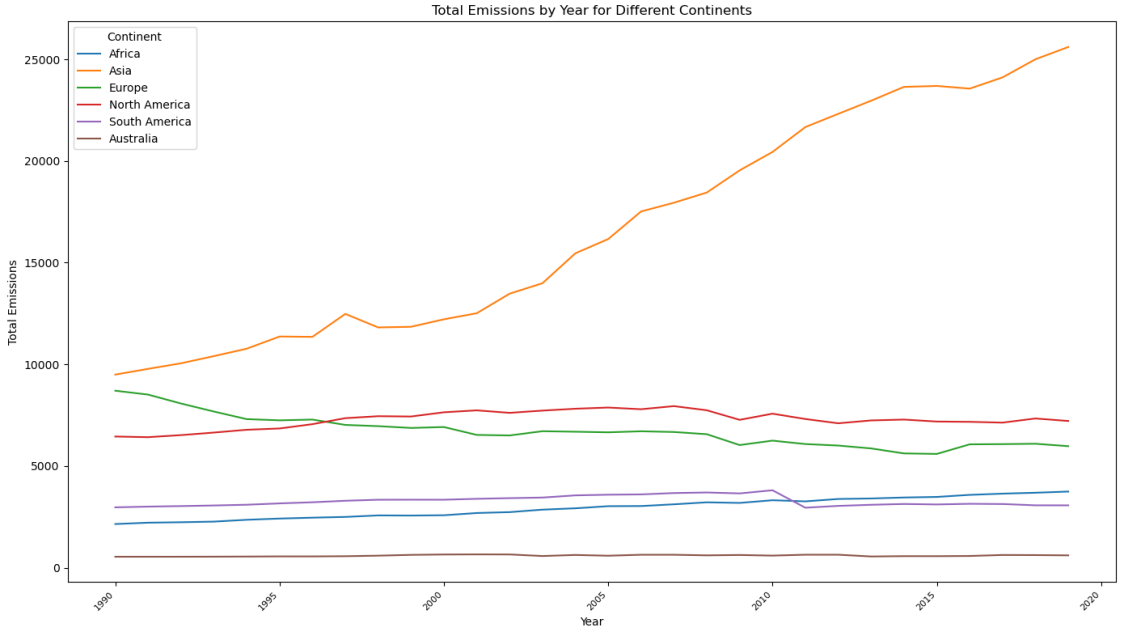
1.9 How do all the continents stack up against one another? Asia, Africa, Australia, North America, South America, and Europe?

Emissions by continent for North America, South America, and Africa have a flat slope overall and indicate that there has not been a large change in emissions values over the last 30 years. North and South America intuitively makes sense as most of the nations had development before 1990. Europe's line has been a small decrease which is emblematic of the emissions control and regulation's the EU is known for. The outlier is Asia which was leading all continents in 1990 but has grown almost 2.5 times to almost 25,000 after starting at just 10,000 in 1990. This makes sense as both China and India have steep increases in emissions values which represents the growing industrialization efforts of both nations and represents the pollution such activities cause. Australia has a flat slope with little change and the emissions value is by far the lowest of all nations at less than 500. The only continents with significant emissions change is Asia with a large increase and europe with a modest but steady decrease

[32]:
```
australian_continent = [
    'Australia'
]
```

```
australian_data = sorted_climate[sorted_climate['Country'].
 ↪isin(australian_continent)]

african_sum = african_data.groupby('Year')['Emissions'].sum().reset_index()

asian_sum = asian_data.groupby('Year')['Emissions'].sum().reset_index()

european_sum = european_data.groupby('Year')['Emissions'].sum().reset_index()

north_american_sum = north_american_data.groupby('Year')['Emissions'].sum().
 ↪reset_index()

south_american_sum = south_american_data.groupby('Year')['Emissions'].sum().
 ↪reset_index()

australian_sum = australian_data.groupby('Year')['Emissions'].sum().
 ↪reset_index()


plt.figure(figsize=(14, 8))
sns.lineplot(x='Year', y='Emissions', data=african_sum, label='Africa')

sns.lineplot(x='Year', y='Emissions', data=asian_sum, label='Asia')

sns.lineplot(x='Year', y='Emissions', data=european_sum, label='Europe')

sns.lineplot(x='Year', y='Emissions', data=north_american_sum, label='North␣
 ↪America')

sns.lineplot(x='Year', y='Emissions', data=south_american_sum, label='South␣
 ↪America')

sns.lineplot(x='Year', y='Emissions', data=australian_sum, label='Australia')

plt.title('Total Emissions by Year for Different Continents')

plt.xlabel('Year')

plt.ylabel('Total Emissions')

plt.xticks(fontsize=8, rotation=45, ha='right')

plt.legend(title='Continent')

plt.tight_layout()
```

```
plt.show()
```


Total Emissions by Year for Different Continents
Continent
Africa
Asia
Europe
North America
South America
Australia
Total Emissions
0
5000
10000
15000
20000
25000
1990
1995
2000
2005
2010
2015
2020
Year