Data Visualization Project 1 : report

S M Abrar Jahin

# Introduction

For this project I have downloaded different datasets from [this url](https://www.gapminder.org/data/). From different dataset, I have tried to find datasets of 5 different features which has no missing data for at least year 1991-2010 and country of at least 20. After trying different datasets, I have come up to this 5 datasets-

1. co2\_emissions\_tonnes\_per\_person : Total CO2 emissions data for different country
2. coal\_consumption\_total: Total coal consumption for different country by year
3. GDP\_total\_yearly\_growth: Total GDP growth for different country by year
4. income\_per\_person\_gdppercapita\_ppp\_inflation\_adjusted: Per person capita income for different country by year
5. population\_total: Total population for different country by year

# Data Processing

For data processing, I have written a python script. For that, I have selected all data and marged all 5 data to a single CSV file with storing year, country and 5 property name. So, my CSV is looking like this-

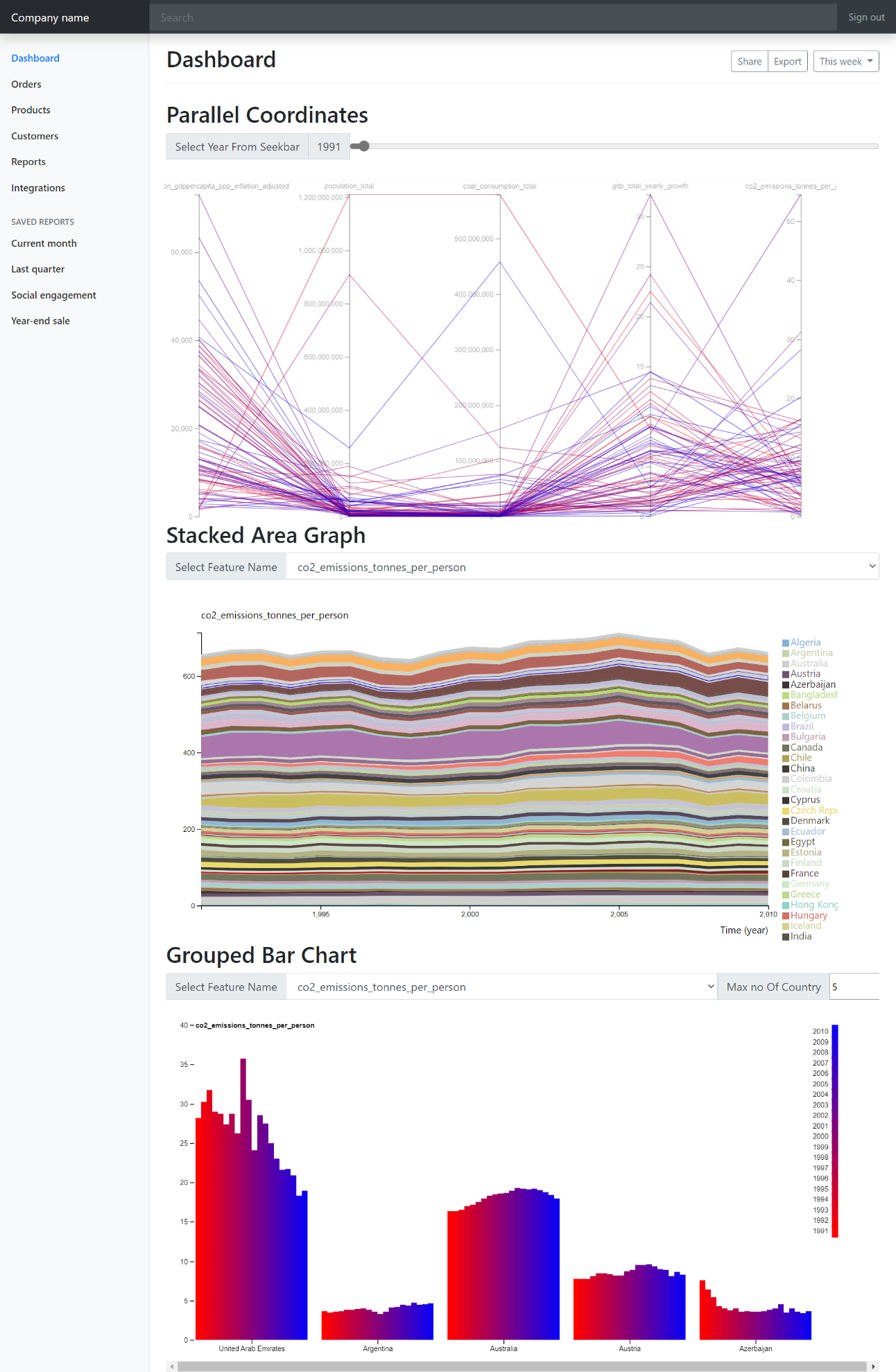
Table

Description automatically generated

For my data, I have found 80 data which don’t have missing data. So, I have removed all country data which has any missing values during the time period of 1991-2010.

# Data visualization Dashboard

In this project, I have visualized data in a dashboard. In the dashboard, there are some dynamic interactive options to visualize data from different angle and different range. I have created a dashboard like this-



# Data visualization Features

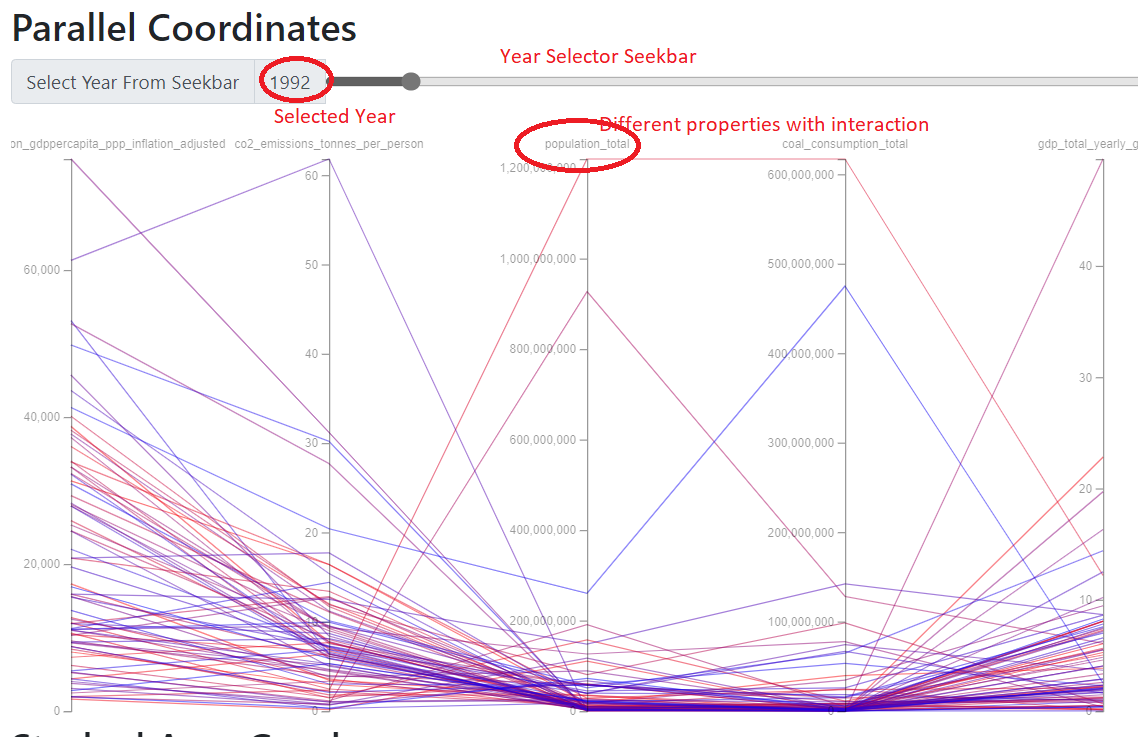
All the ranges of all 3 graphs are dynamically selected as values are changing. And all graphs are shown for same CSV data. As there are 3 kinds of charts in the dashboard, I like to talk about all of them 1 by one.

1. **Parallel Coordinates**-

For parallel coordinates, I have plotted all data for country for different 5 attributes on a selected year. The year can be changed by seek bar provided on top of the parallel coordinate graph. Stock color of different data points are set color from an incremental color from red to blue.

I am providing some parallel coordinate chart image for different year (selected from the seekbar) in here-

For year 1992



For year 2004-

Chart, radar chart

Description automatically generated

So, if value is selected from the seek bar by seeking, the selected year is showing in the bar for better understanding and visualization.

If we like, we can also rearrange the property orders for better visualization like this-

Chart

Description automatically generated

Feature highlighting, a single data highlighting, reordering of features showing are also implemented for better understanding and analysis.

1. **Stacked Area Chart**-

For this kind of chart, we can select different kind of features from the dropdown like this-

A screenshot of a computer

Description automatically generated

Color of different stacks are updated with my created random color generation function which also can be chosen from library and also can be set to incremental color values. If different features are selected, new feature values are shown-

Chart

Description automatically generated with medium confidence

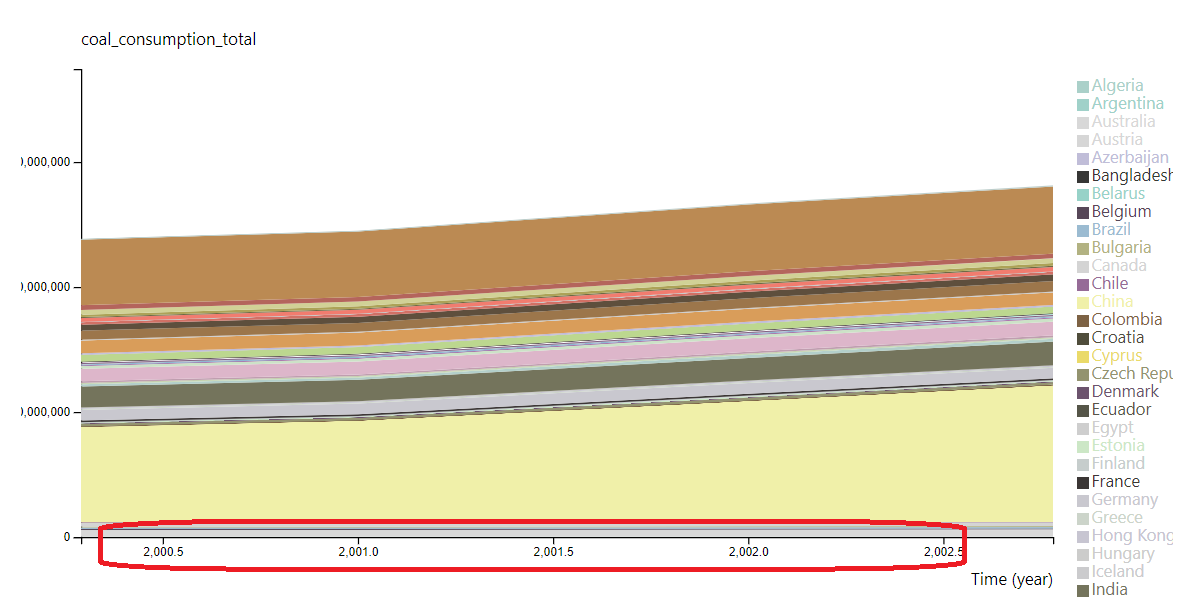
Moreover every country can also be highlighted like this-

Graphical user interface, application

Description automatically generated with medium confidence

Where population of China is highlighted.

Moreover brushing is also implemented like this-



So, after brushing, the time has been extended for better visualization.

1. **Grouped Bar Chart**-

For grouped bar chart, 2 selector is added like this-

Chart, bar chart

Description automatically generated

First one is features (similar with stacked area chart) and

The second one is no of years to show.

As we are showing 20 years of data for each country, it is not possible to show all country data in this small space, so I have made a selector which can determine max no of country to show. Lets change property and change max no of country to see the visualization-

Chart, bar chart

Description automatically generated

For this case, I have changed max no of countries to 7 for better visualization.

Moreover year range is represented with incremental color value from red to blue for better understanding. And countries are randomly chosen as we are not showing all data once as we have small space.

# Data analysis

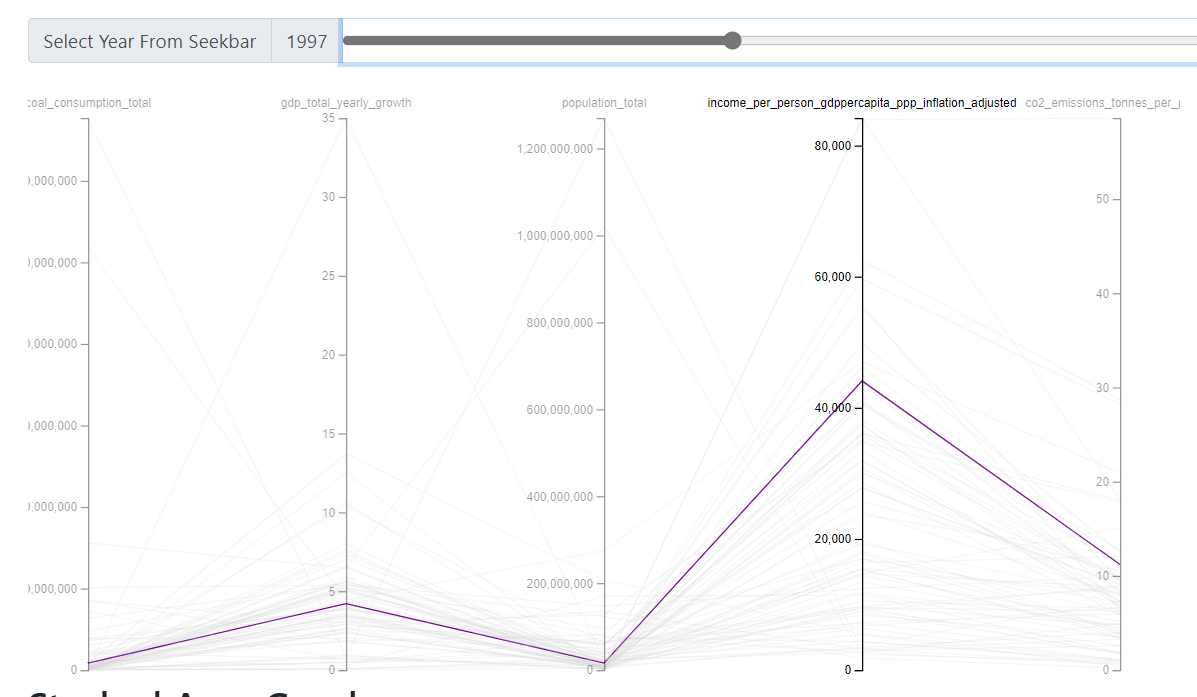
As there are 3 kinds of data, I like to give analysis one by one-

1. **Parallel Coordinates** –

For parallel coordinate, what I am seeing from different year data, the property of data are in average like the selected line-



So, CO2 emission and total coal consumption is co related, population is also co-related with CO2 emission, but coal consumption is related with CO2 emissions.



For majority of data, if population increases, per person capita income increase and CO2 emission increase. And GDP has a positive relation with coal consumption because if more coal is consumed, it is possible, more industries are in there.

1. **Stacked Area Chart** –

From the chart, we are having 5 feature based stacked area chart. For CO2 emission, I am getting a chart like this-

A picture containing background pattern

Description automatically generated

So, if year is increasing from 1990 to 2010, CO2 emission has a slidably increasing every year till 2005, after 2005, CO2 emission has a slidably decrease.

For overall coal consumption, we are seeing that total consumption is always increasing year by year-

Chart

Description automatically generated

And after 2000, the increasing rate is higher.

The GDP total stacked area chart is looking like this-

A picture containing graphical user interface

Description automatically generated

So, overall, income per person is always increasing with a little exception.

In the stacked area chart of per person GDP capita income is always increasing year by year.

A picture containing chart

Description automatically generated

Last but not the least, population is always growing year by year like this-

Chart

Description automatically generated with medium confidence

1. **Grouped Bar Chart** –

There are 5 properties in here-

For CO2 emissions, for the selected 5 countries, the CO2 emission is overall increasing for Australia, Austria and Argentina, but decreasing for United Arab Amirates and Azerbizan-

Chart, bar chart

Description automatically generated

For coal consumption, it is overall increasing every yerar while Russia and Ukrain has decreased coal consumption in the during period for our data.

Chart, histogram

Description automatically generated

For GDP yearly growth of this 5 countries, data is not showing any specie properties as data is showing scatter properties.

Chart, bar chart, histogram

Description automatically generated

For income, it is increasing for all of mentioned 10 countries-

Chart, bar chart, histogram

Description automatically generated

For population of these 7 countries, it is overall increasing-

Chart, bar chart

Description automatically generated