

Data Visualization Project 1 : report

S M Abrar Jahin

INTRODUCTION

For this project I have downloaded different datasets from [this url](#). 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

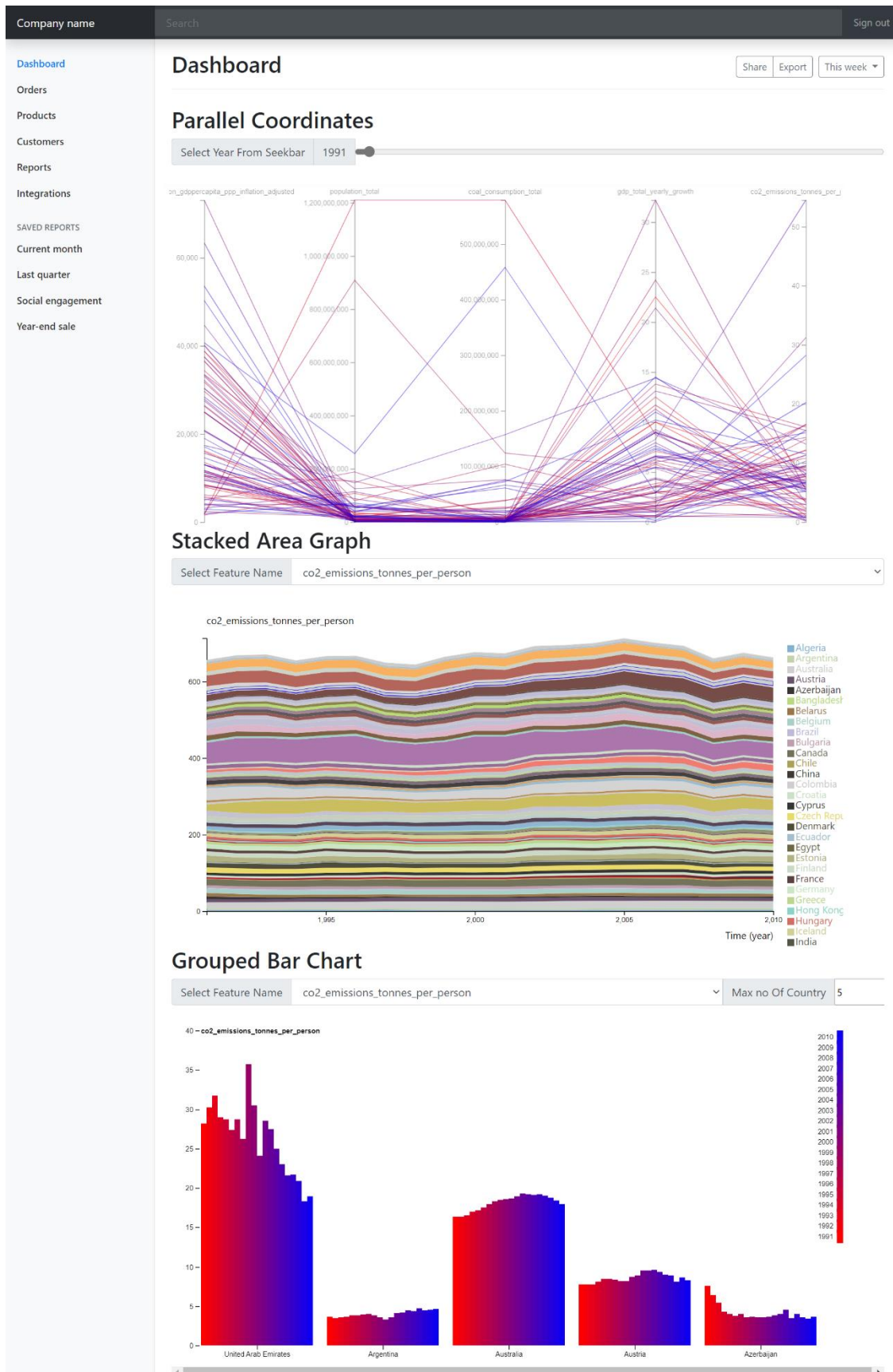
For data processing, I have written a python script. For that, I have selected all data and marged all 5 data to a single CSV file with storing year, country and 5 property name. So, my CSV is looking like this-

country	year	co2_emissi	coal_consumption_total	gdp_total	income_per	population_total
Argentina	1991	3.61	1050000	10	16200	33500000
Argentina	1992	3.46	694000	6.04	17300	34000000
Argentina	1993	3.55	607000	6.12	18100	34400000
Argentina	1994	3.67	638000	2.7	17400	34800000
Argentina	1995	3.82	668000	5.76	18100	35200000
Argentina	1996	3.86	527000	8.34	19300	35700000
Argentina	1997	3.87	708000	3.99	19900	36100000
Argentina	1998	4.02	850000	3.38	19000	36500000
Argentina	1999	3.85	646000	0.807	18600	36900000
Argentina	2000	3.58	529000	4.52	17600	37300000
Argentina	2001	3.3	394000	11.1	15500	37700000
Argentina	2002	3.54	481000	8.76	16700	38100000
Argentina	2003	4.08	819000	8.88	18000	38500000
Argentina	2004	4.15	943000	9.08	19400	38900000
Argentina	2005	4.45	853000	8.21	20800	39300000
Argentina	2006	4.39	1230000	8.39	22400	39700000

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

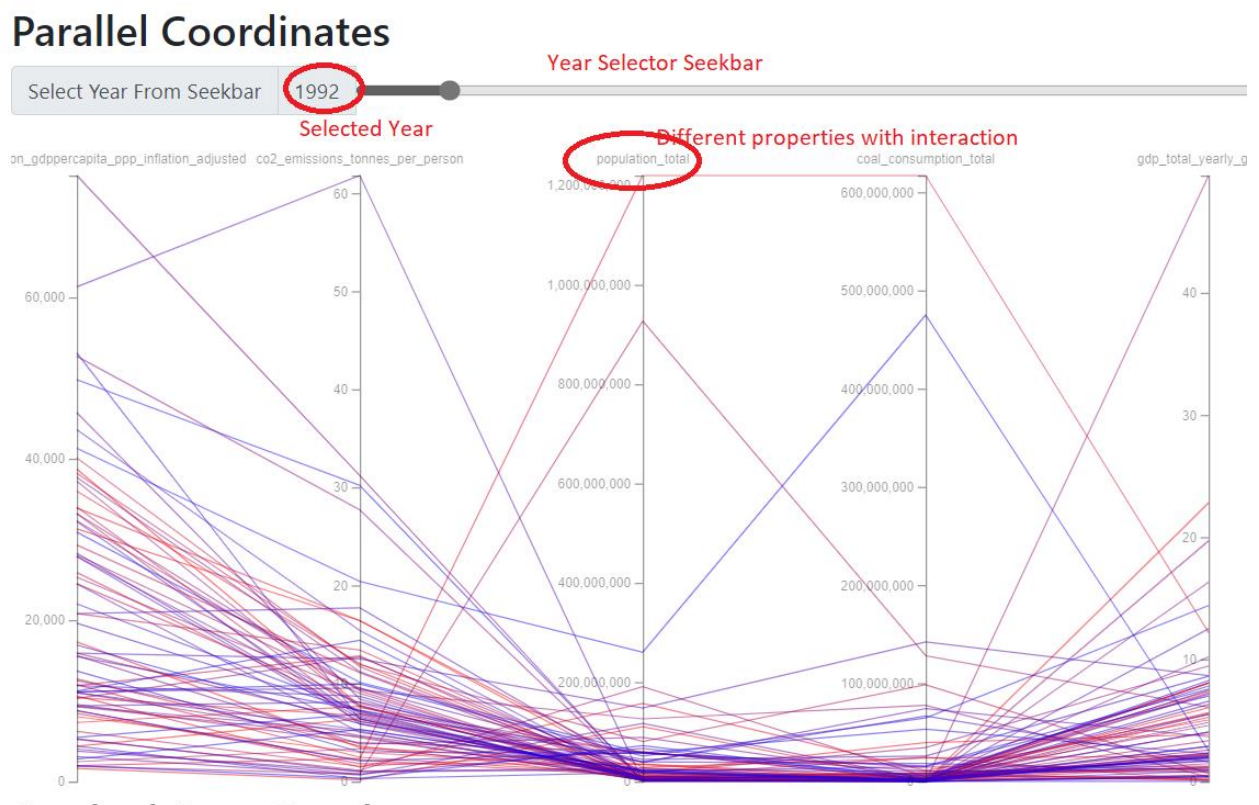
All the ranges of all 3 graphs are dynamically selected as values are changing. And all graphs are shown for same CSV data. As there are 3 kinds of charts in the dashboard, I like to talk about all of them 1 by one.

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.

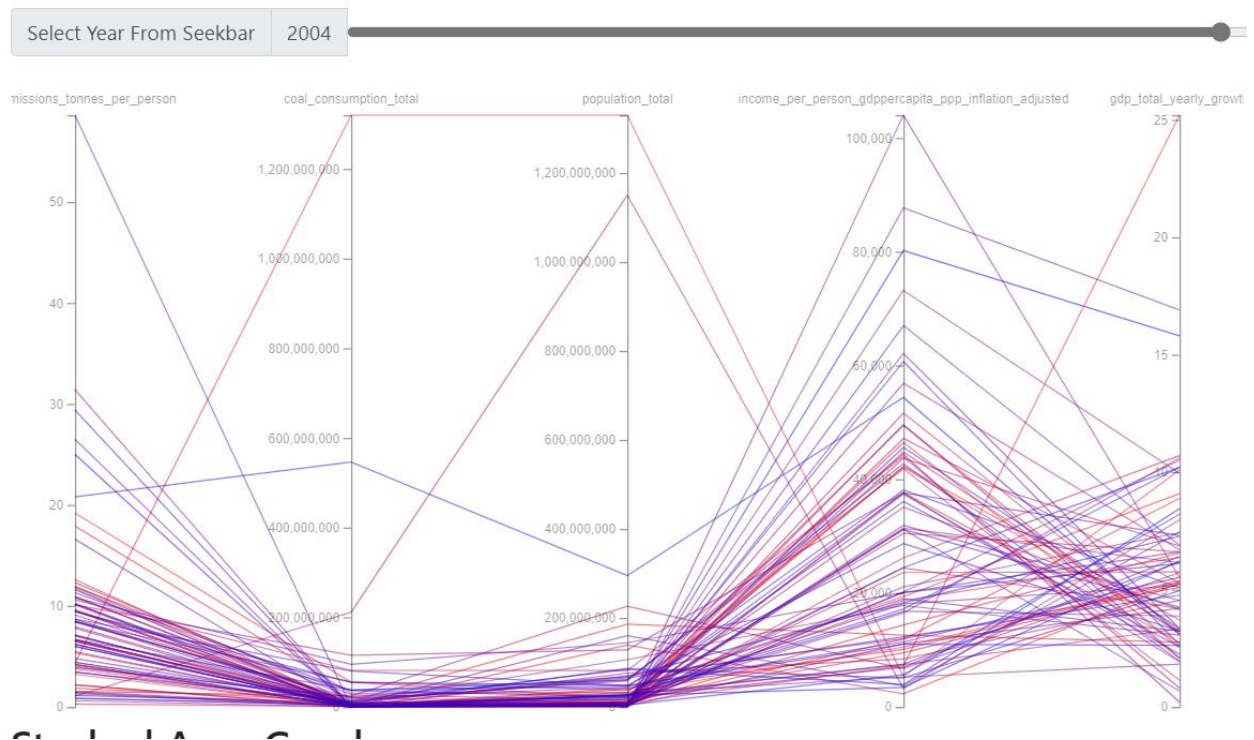
I am providing some parallel coordinate chart image for different year (selected from the seekbar) in here-

For year 1992



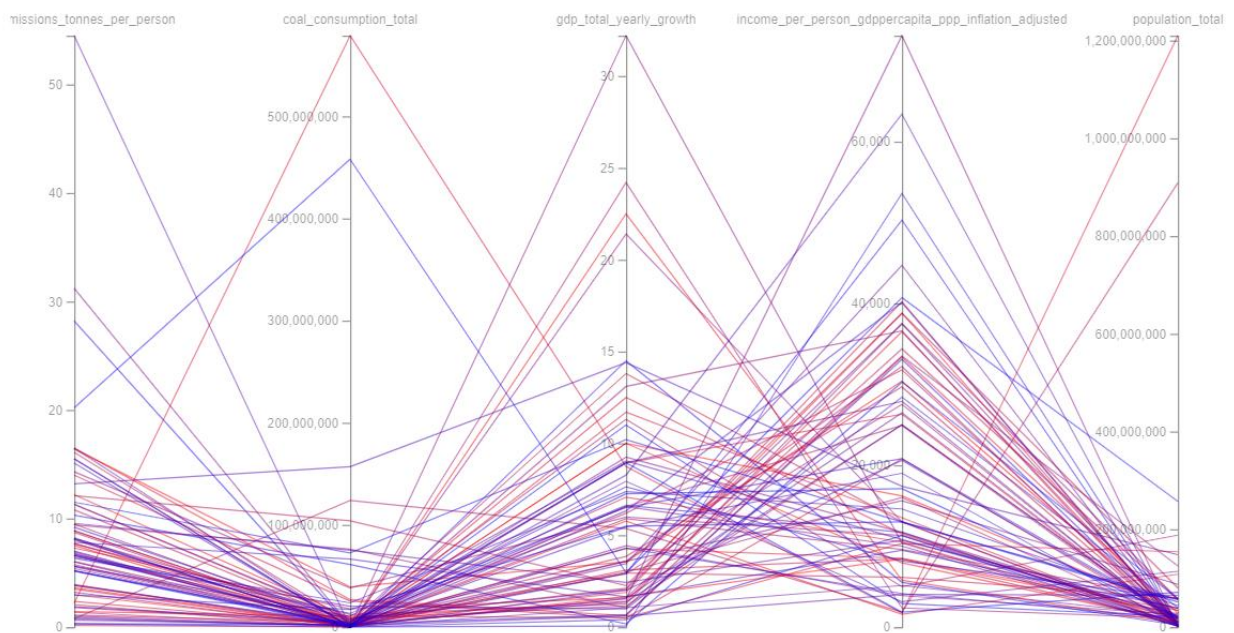
For year 2004-

Parallel Coordinates



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-

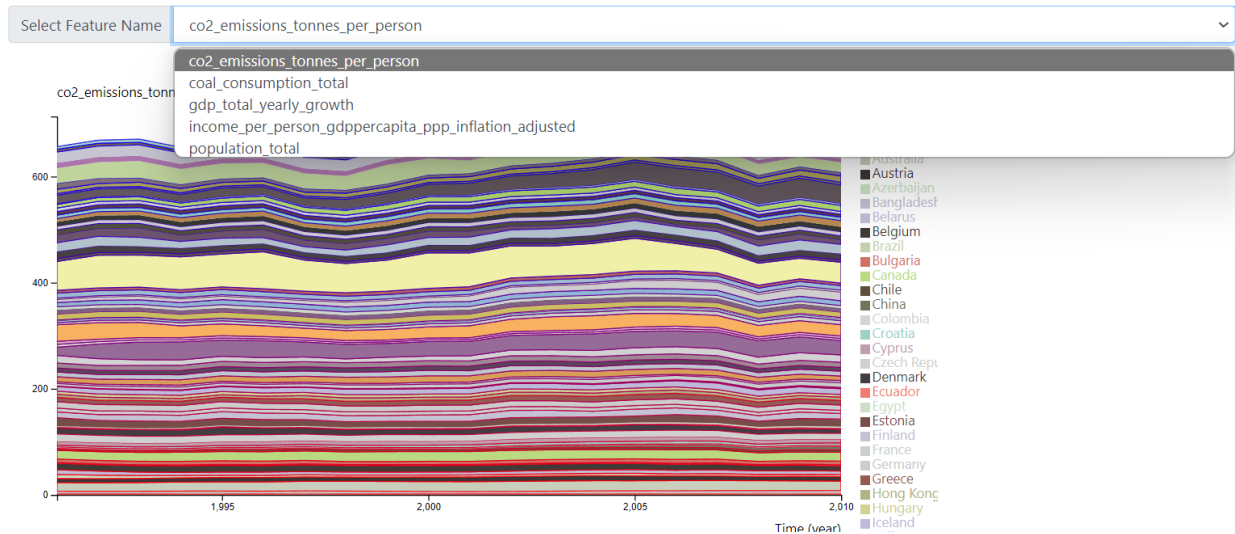


Feature highlighting, a single data highlighting, reordering of features showing are also implemented for better understanding and analysis.

2. Stacked Area Chart-

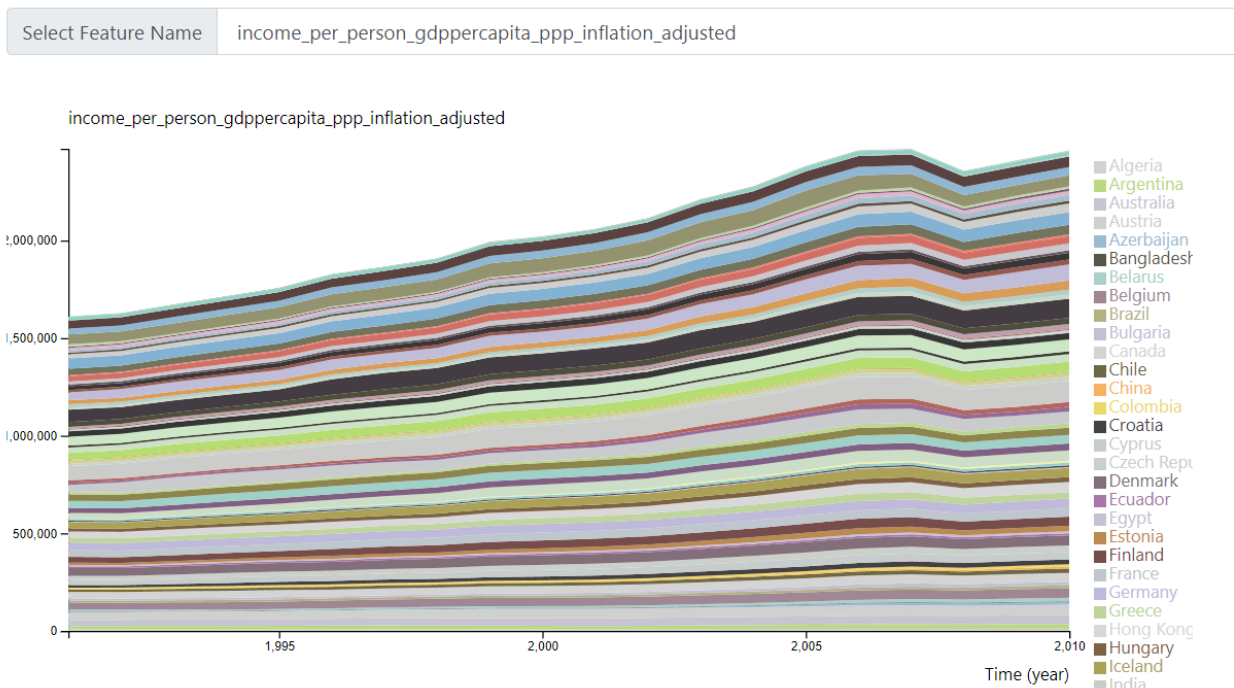
For this kind of chart, we can select different kind of features from the dropdown like this-

Stacked Area Graph

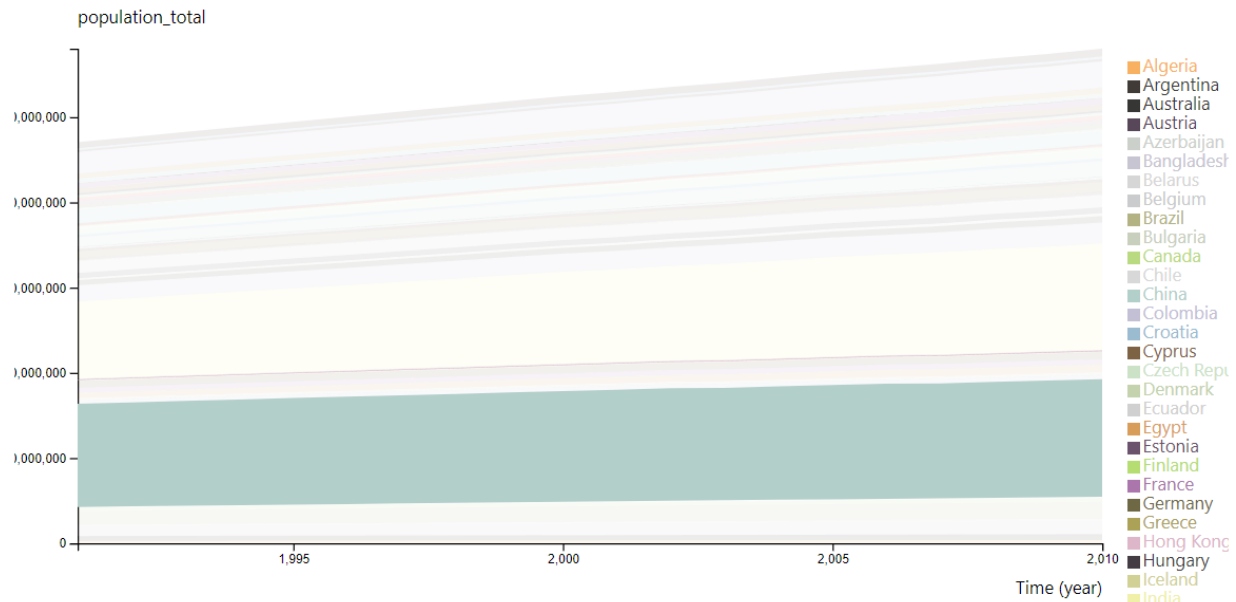


Color of different stacks are updated with my created random color generation function which also can be chosen from library and also can be set to incremental color values. If different features are selected, new feature values are shown-

Stacked Area Graph

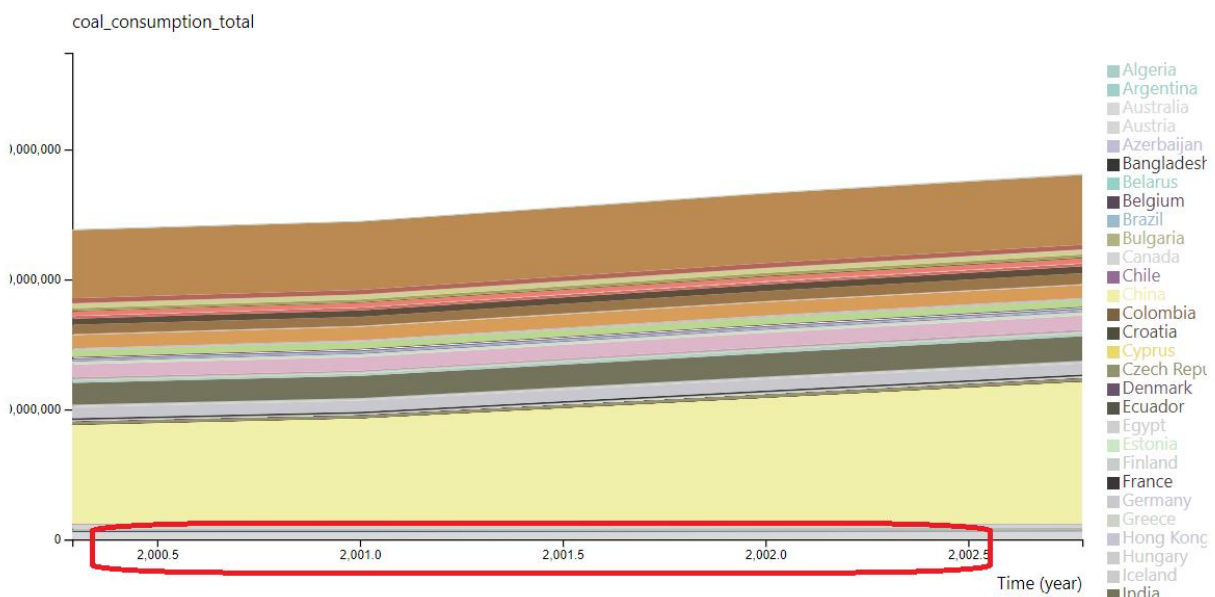


Moreover every country can also be highlighted like this-



Where population of China is highlighted.

Moreover brushing is also implemented like this-

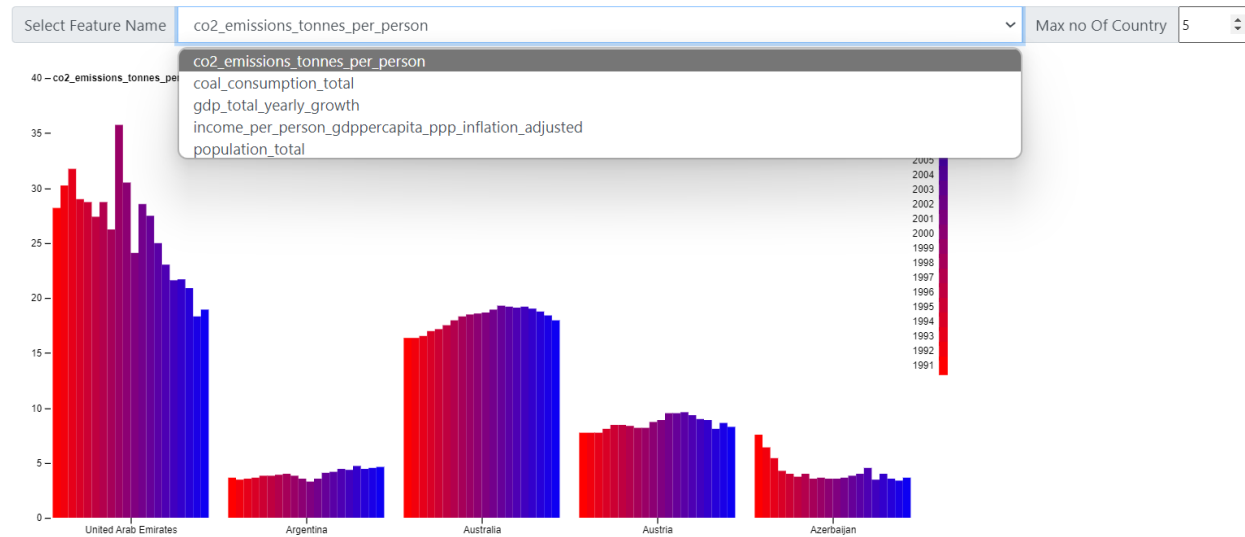


So, after brushing, the time has been extended for better visualization.

3. Grouped Bar Chart-

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

Grouped Bar Chart

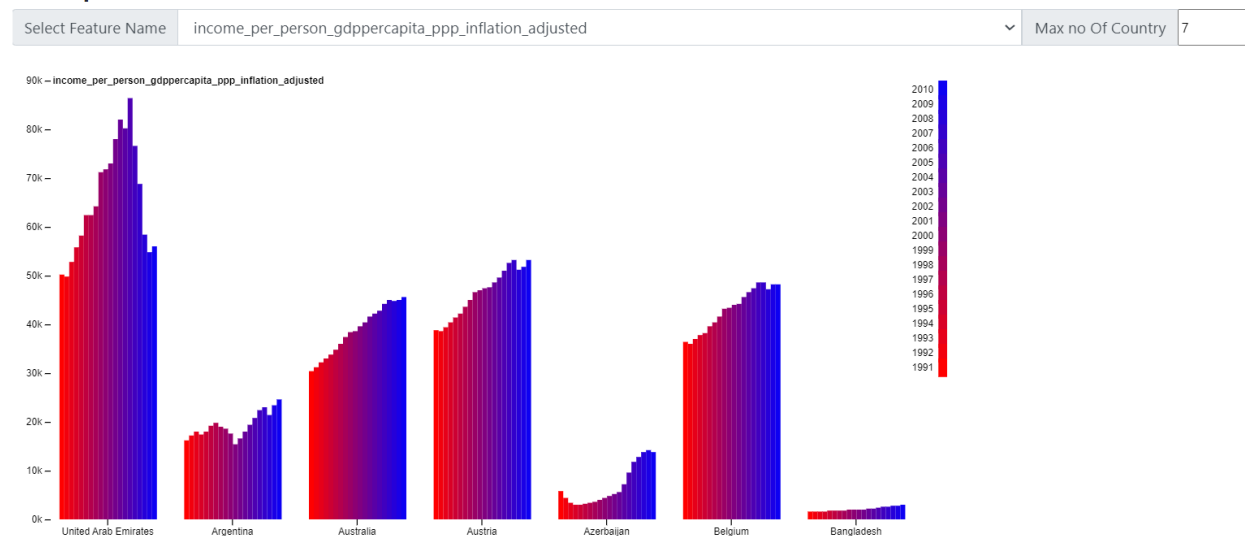


First one is features (similar with stacked area chart) and

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-

Grouped Bar Chart



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

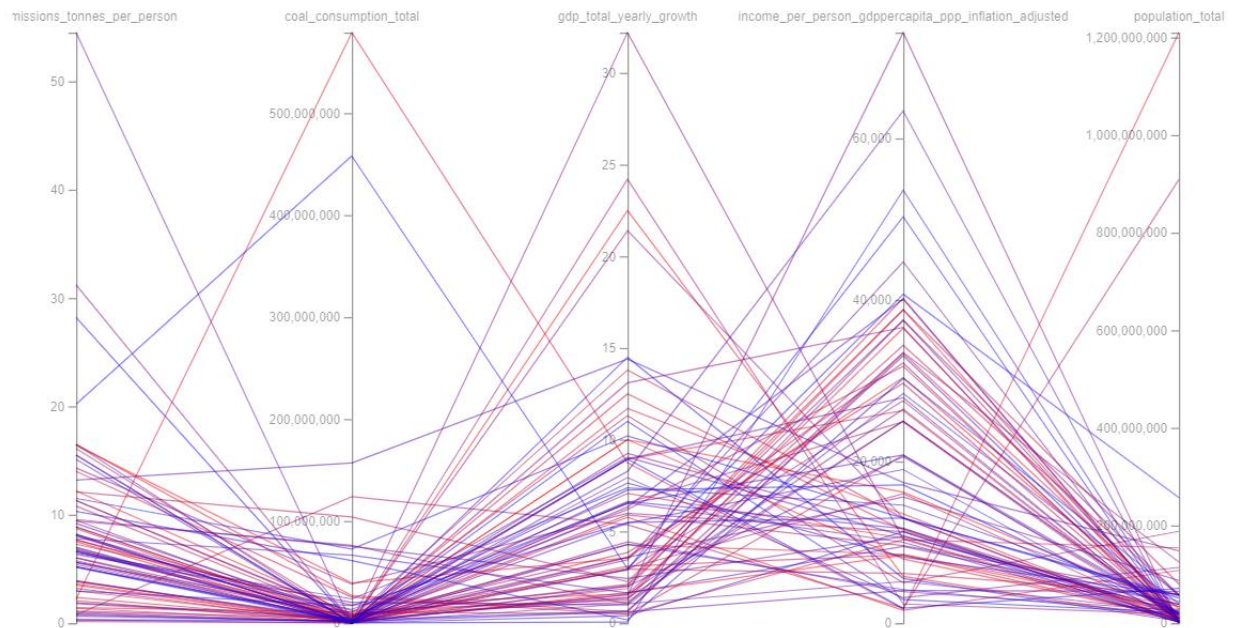
Moreover year range is represented with incremental color value from red to blue for better understanding. And countries are randomly chosen as we are not showing all data once as we have small space.

DATA ANALYSIS

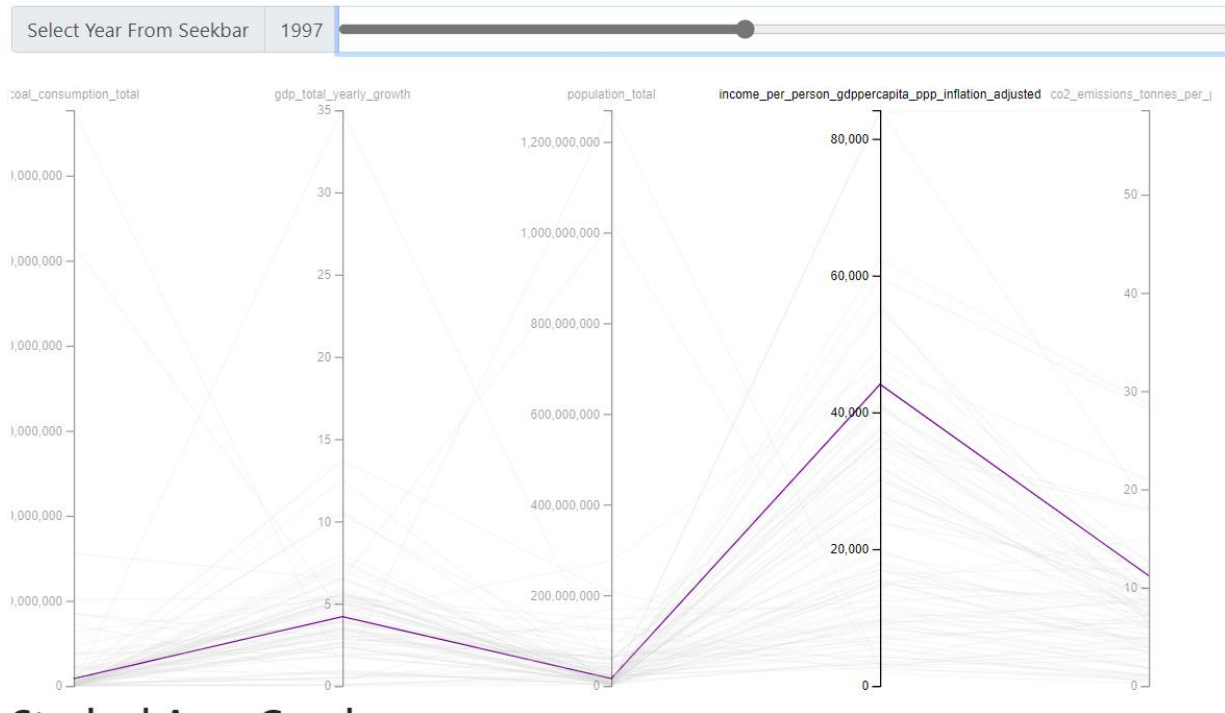
As there are 3 kinds of data, I like to give analysis one by one-

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.

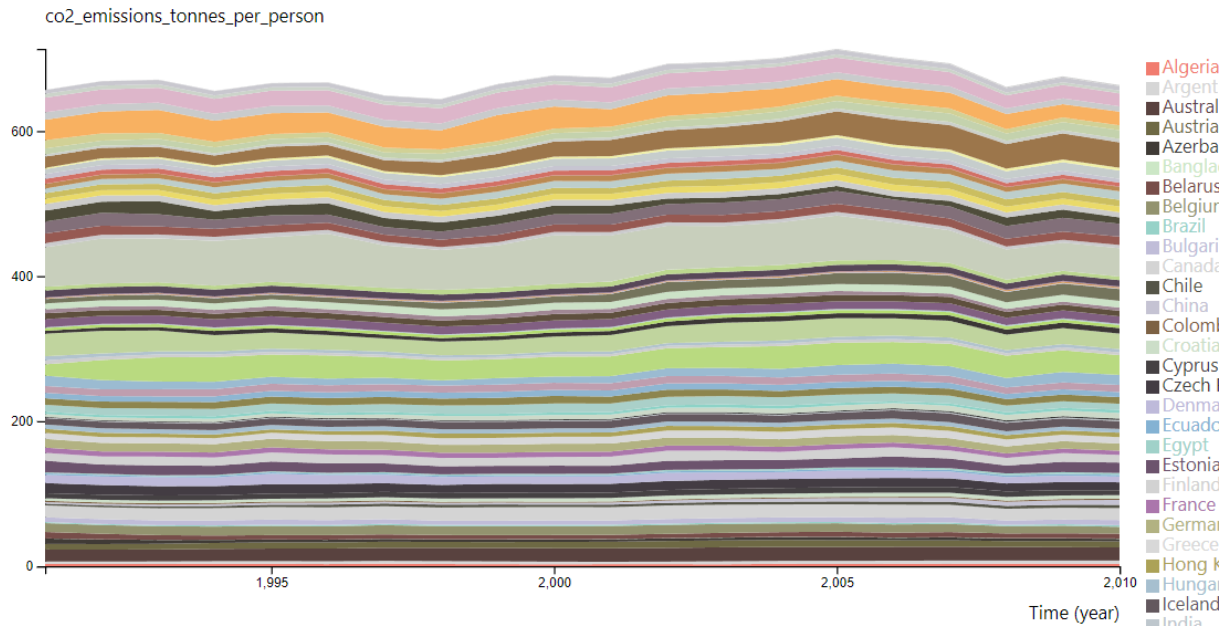


For majority of data, if population increases, per person capita income increase and CO2 emission increase. And GDP has a positive relation with coal consumption because if more coal is consumed, it is possible, more industries are in there.

2. 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-

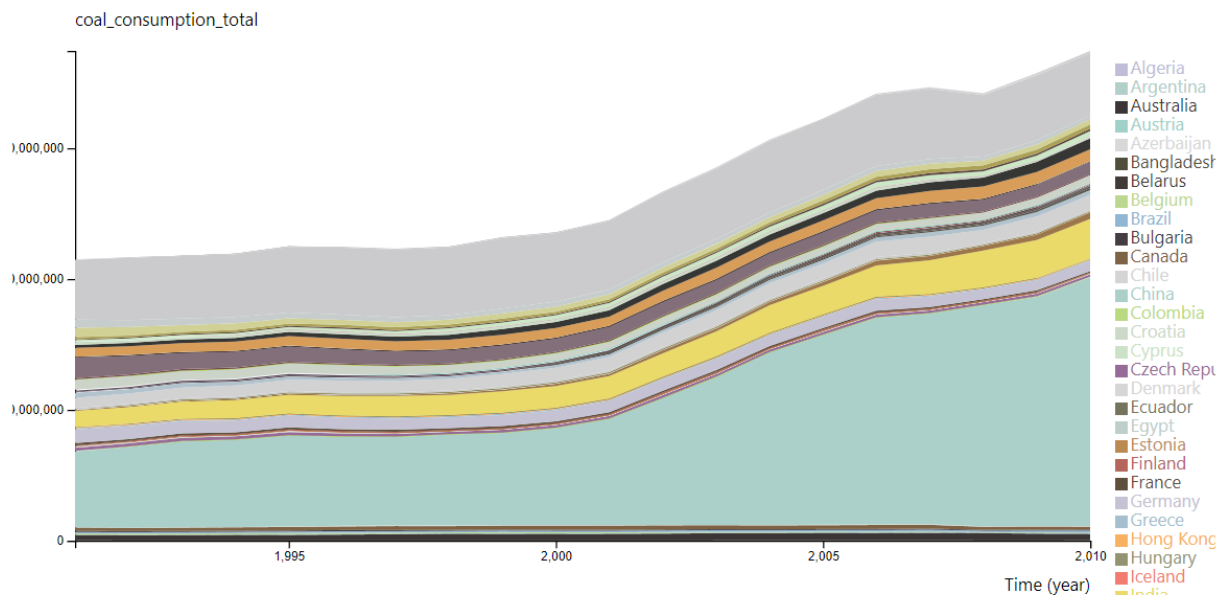
Select Feature Name co2_emissions_tonnes_per_person



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

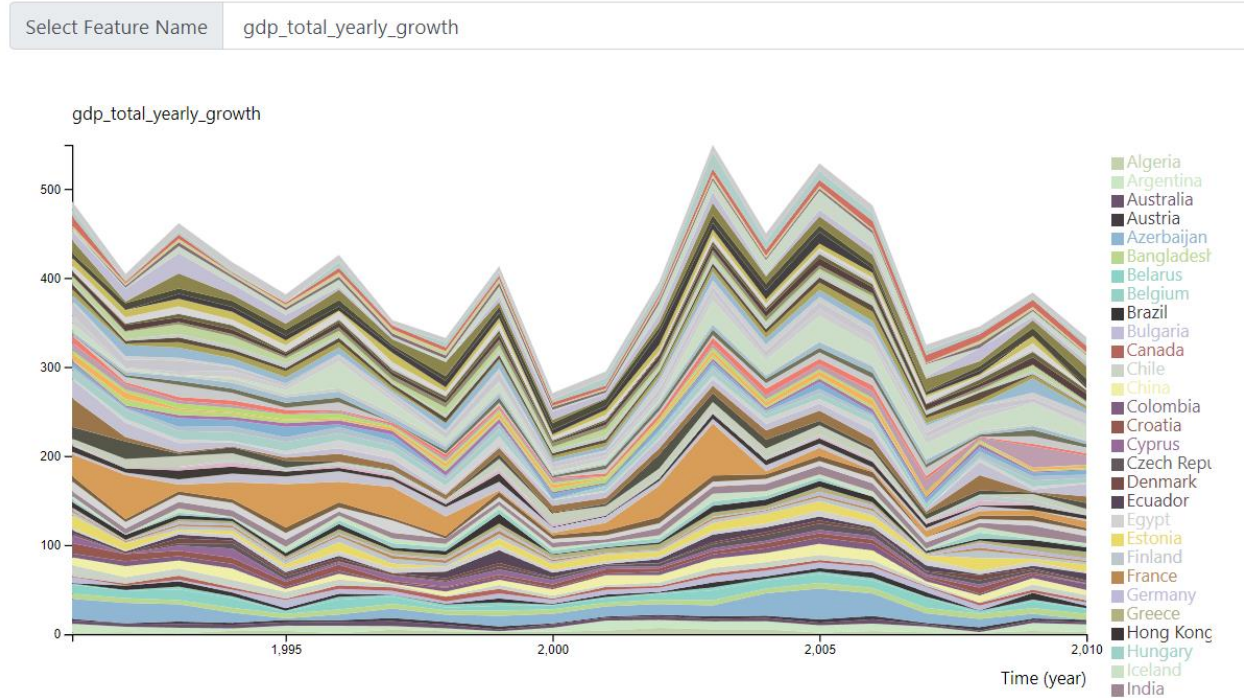
For overall coal consumption, we are seeing that total consumption is always increasing year by year-

Select Feature Name coal_consumption_total



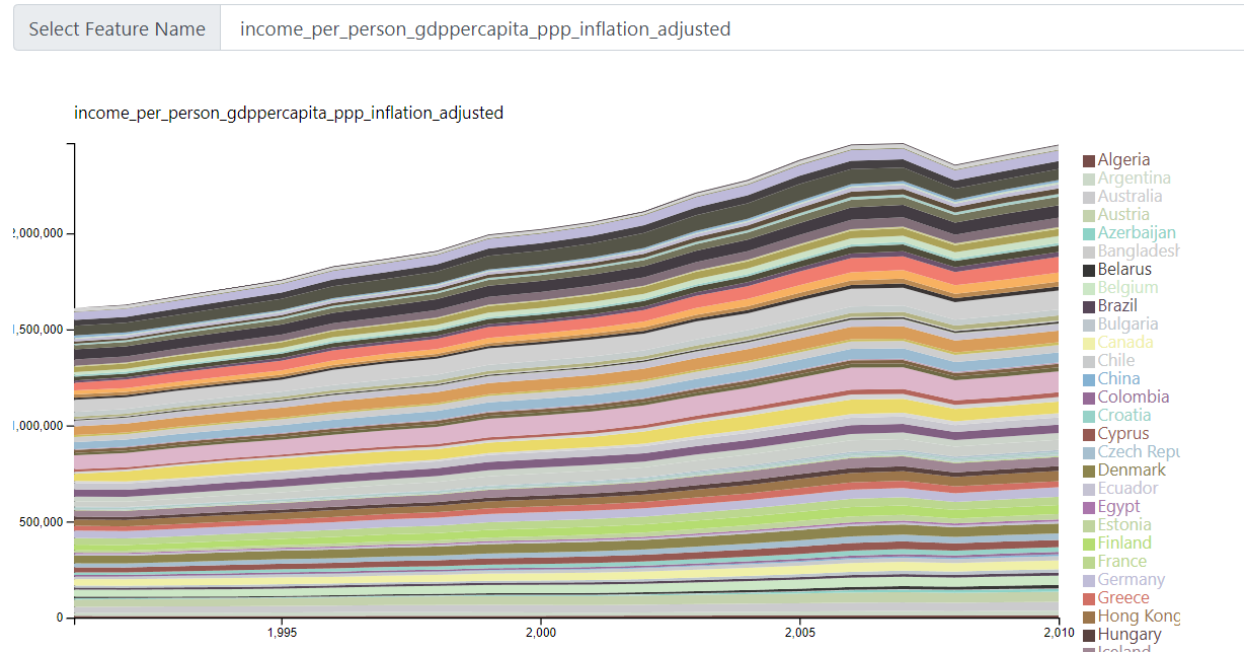
And after 2000, the increasing rate is higher.

The GDP total stacked area chart is looking like this-

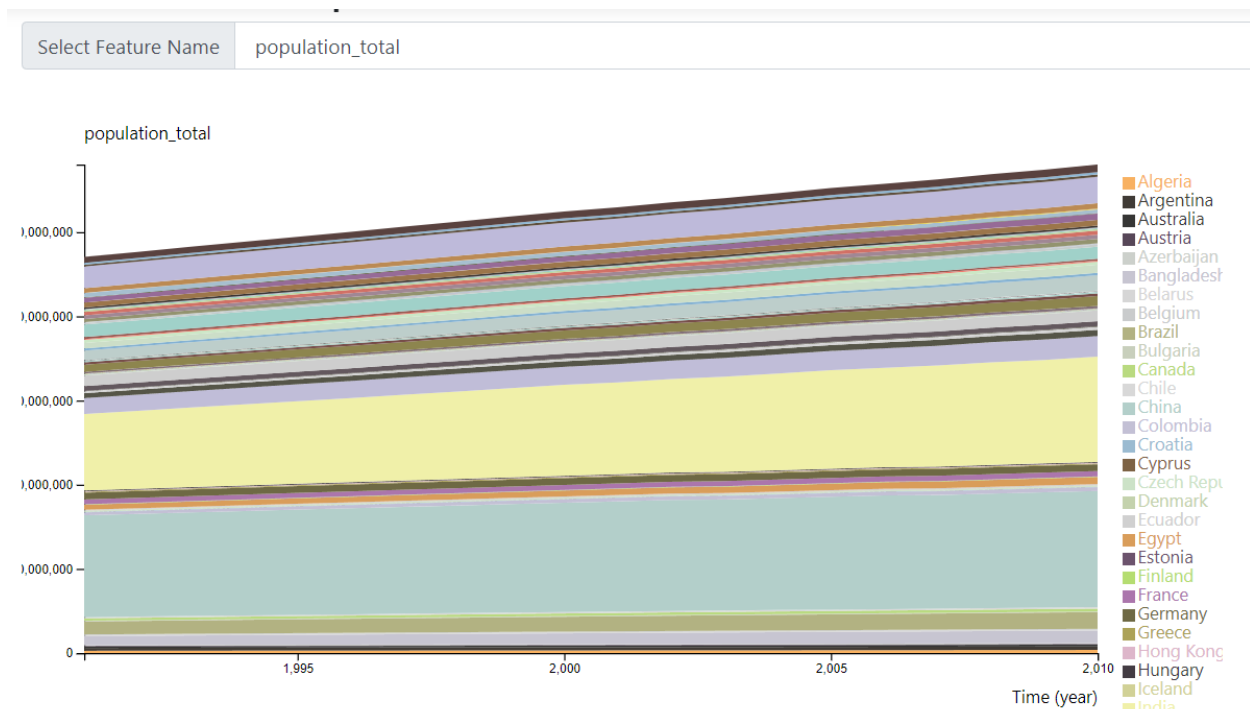


So, overall, income per person is always increasing with a little exception.

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



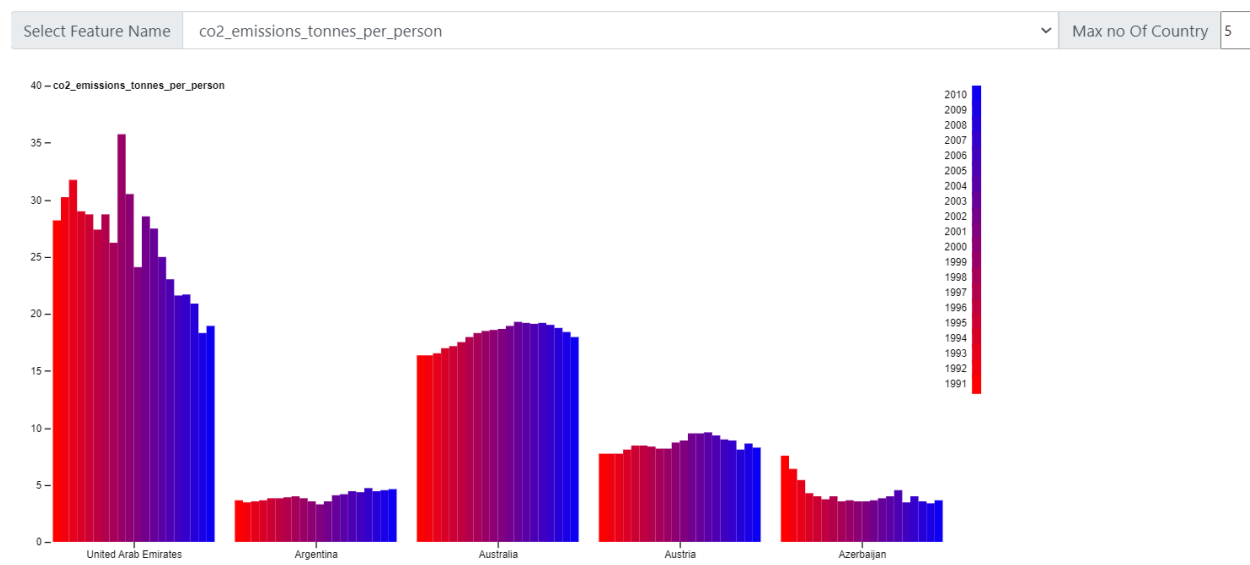
Last but not the least, population is always growing year by year like this-



3. Grouped Bar Chart –

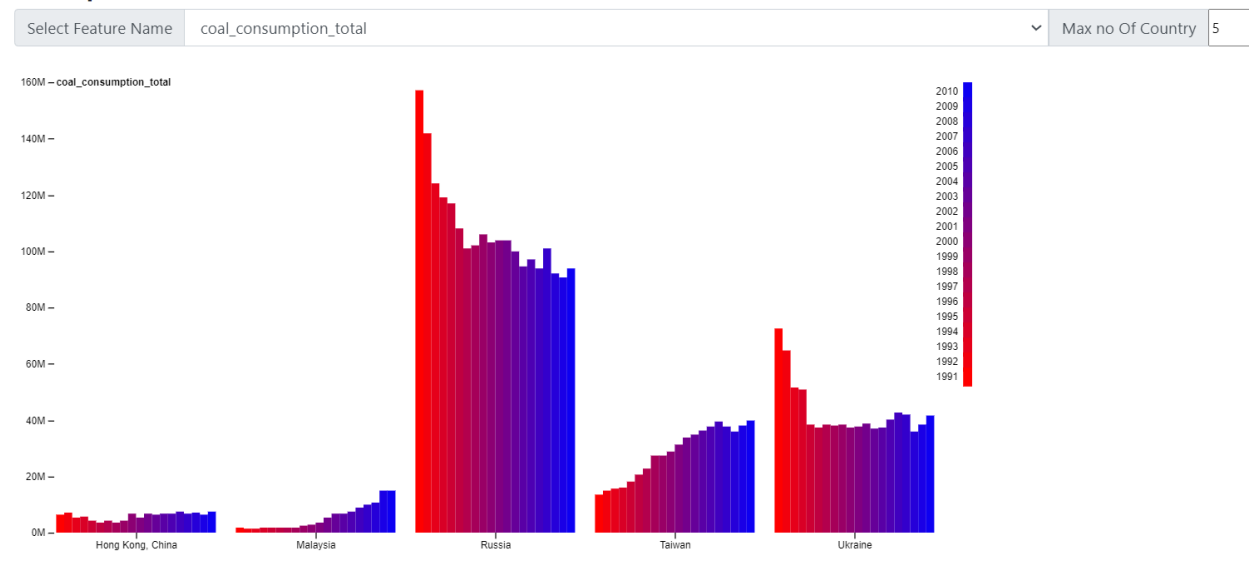
There are 5 properties in here-

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-

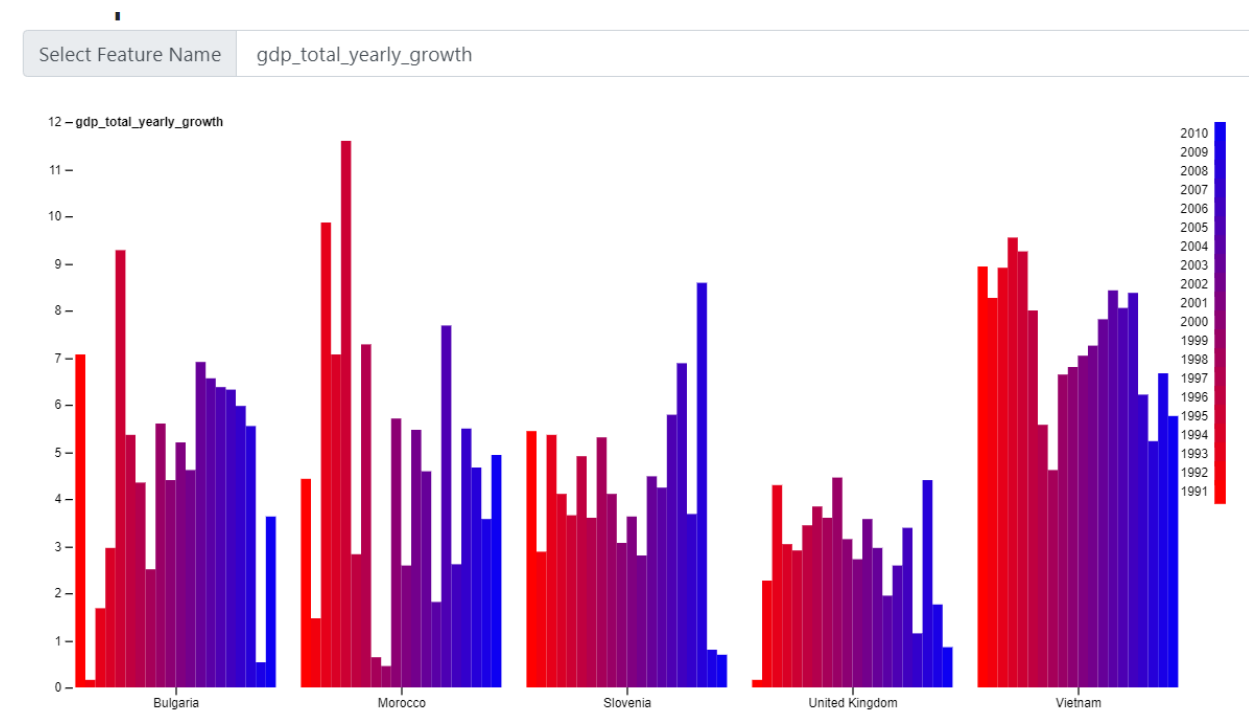


For coal consumption, it is overall increasing every year while Russia and Ukraine has decreased coal consumption in the during period for our data.

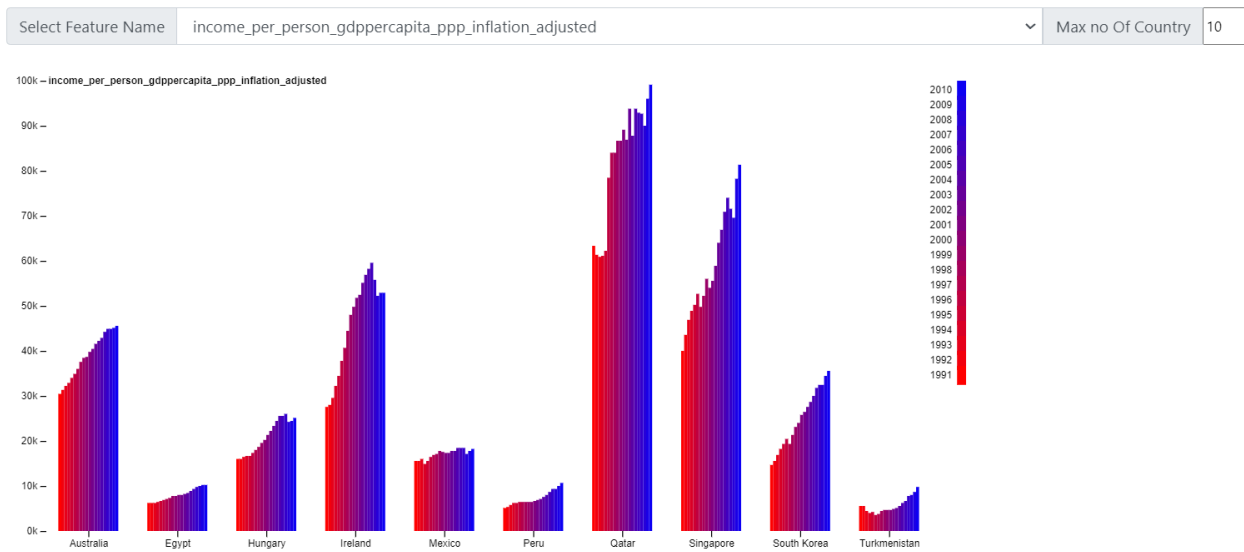
Grouped Bar Chart



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



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



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

