Visualization of Expenditure from G-20 Countries

DATS 6401 Visualization of Complex Data

Xi Zhang

G28490493

Instructor: Nima Zahadat, Ph.D.

Agenda

- Data Source
- Data Preprocessing
- Programming Introduction
- Analysis by Category
  - Overview
  - Education Expenditure
  - Healthcare Expenditure
  - Military Expenditure
  - GDP Impact
- Conclusion

Data Description

❖Data Source:

World Bank: https://data.worldbank.org/

❖Research Catalogs:

GDP; Population; Education; Healthcare; Military

❖Selected Countries:

USA, UK, France, Canada, Japan, China, India, Mexico, Australia, South Korea

❖Selected Years:

2010 – 2016

Data Preprocessing

Read data and filter the last ten years and selected countries.

```
df = pd.read_csv(path + Document, index_col=0,header=2)
# filter the selected countries
df = df.loc[['United States', 'United Kingdom', 'France', 'Canada',
             'Japan', 'China', 'India', 'Mexico', 'Australia', 'Korea, Rep.']]
# filter the data since 2005 and keep the country code
Code = df.loc[:,'Country Code']
df = df.loc[:, '2010':'2019']
```

	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
United States	48466.82338	49883.11398	51603.49726	53106.90977	55032.95800	56803.47243	57904.20196	59927.92983	62794.58565	nan
United Kingdom	39435.83990	42038.50048	42462.71491	43444.56484	47417.63505	44966.10193	41074.16728	40361.41738	42943.90227	nan
France	40638.33400	43790.73205	40874.70351	42592.93409	43008.64738	36613.37522	36962.22229	38679.12715	41463.64402	nan
Canada	47450.31847	52101.79609	52542.34666	52504.65570	50835.51118	43495.05439	42279.90082	45069.92725	46232.98962	nan
Japan	44507.67639	48167.99727	48603.47665	40454.44746	38109.41211	34524.46986	38794.33094	38331.97940	39289.95843	nan
China	4550.45360	5618.13249	6316.91886	7050.64592	7651.36604	8033.38804	8078.79047	8759.04158	9770.84709	nan
India	1357.56372	1458.10353	1443.87953	1449.60591	1573.88149	1605.60543	1729.26802	1981.26871	2009.97886	nan
Mexico	9271.39823	10203.42085	10241.72792	10725.18332	10922.37605	9605.95235	8739.75604	9278.41817	9673.44367	nan
Australia	52022.12560	62517.83375	68012.14790	68150.10704	62510.79117	56755.72171	49971.13146	54066.47127	57373.68668	nan
Korea, Rep.	22086.95292	24079.78852	24358.78218	25890.01867	27811.36638	27105.07623	27608.24743	29742.83886	31362.75147	nan

Obviously, there are many missing values in 2019. Then I check the data quality.

Data Preprocessing

Check the duplication, missing values and fix the problems.

```
# duplicate values
print(df.duplicated())
```

```
Country Name
United States      False
United Kingdom     False
France             False
Canada             False
Japan              False
China              False
India              False
Mexico             False
Australia          False
Korea, Rep.        False
dtype: bool
```

```
# number of missing values
for COL in df.columns:
    print(COL + ':', len(df) - df[COL].count())
```

```
2010: 0
2011: 0
2012: 0
2013: 0
2014: 0
2015: 0
2016: 0
2017: 0
2018: 0
2019: 10
```

```
# drop columns with too many missing values
df = df.drop(['2019'], axis=1)

# fill missing values by mean values
df = df.fillna(df.mean())
# drop duplicated values
df = df.drop_duplicates()
```

```
# save the cleaned data
df.to_csv(path + "/cleaned_data/" + "0" + Document, index=None)
```

After traversing all the datasets, missing values always focus on 2017 – 2019. Therefore, I only use data from 2010 -2016 for my following researches.
For other small parts of missing, I fixe them by filling by mean values.

Data Preprocessing - Calculation

```
# Libraries ========================================================================================================== #
import pandas as pd

# Data reading ======================================================================================================= #
path = "/Users/Fox/Desktop/Data Science/Data Vis/Project1/Data/"
NAME = "Government expenditure on education, total (US$)"
Document = "cleaned_data/"+ NAME + ".csv"
GDP_ = "GDP"
GDP__ = "cleaned_data/"+ GDP_ + ".csv"
GDP = pd.read_csv(path + GDP__, index_col= 0)
POP_ = "Population, total"
POP__ = "cleaned_data/"+ POP_ + ".csv"
POP = pd.read_csv(path + POP__, index_col= 0)
df0 = pd.read_csv(path + Document, index_col= 0)

# Calculate the education expenditure in US$
# df = GDP * df0 / 100

# Calculate the per capita data
# df = df0 / POP

# Calculate the growth rate
df= ((df0["2016"] - df0["2010"])/df0["2010"])/7

NEWNAME = "Education Growth Rate"
# save the cleaned data
df.to_csv(path + "/cleaned_data/" + NEWNAME + ".csv", index=True)
```

I load the associated data and write formulas to calculate expenditures in different units and growth rate for every categories .

Programming Introduction

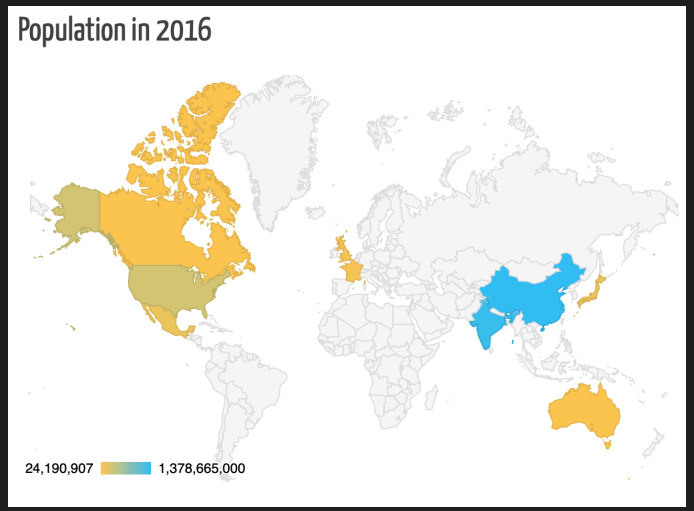
```
google.charts.load('current', {'packages':['corechart']});
google.charts.setOnLoadCallback(drawAllSheets);

function drawAllSheets() {
  drawSheetName('Pop2016', 'SELECT A,B',
                PopulationHandler);
  drawSheetName('GDP', 'SELECT A,H',
                GDPHandler);
  drawSheetName('Spending2016', 'SELECT A,B',
                SpendingHandler);
} //drawAllSheets

function drawSheetName(sheetName, query, responseHandler) {
  var queryString = encodeURIComponent(query);
  var query = new google.visualization.Query(
    'https://docs.google.com/spreadsheets/d/1SoWB4JJiK67XChpglo8AjLStyZZSDlc4M3vYT2TdNEM/gviz/tq?sheet='
    + sheetName + '&headers=1&tq=' +queryString
  );

  //https://docs.google.com/spreadsheets/d/1SoWB4JJiK67XChpglo8AjLStyZZSDlc4M3vYT2TdNEM/edit?usp=sharing
  query.send(responseHandler);
} //drawSheetNames
```

Google API

Data Site:
https://docs.google.com/spreadsheets/d/1SoWB4JJiK67XChpglo8AjLStyZZSDlc4M3vYT2TdNEM/edit?usp=sharing

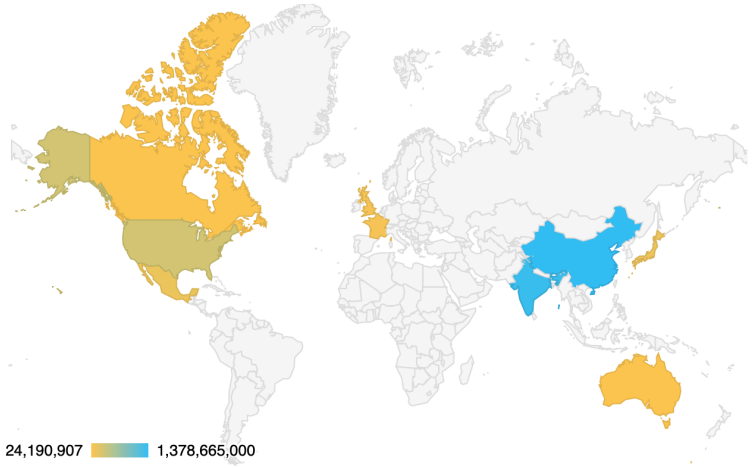
Programming Introduction

```
function PopulationHandler(response){
  //get the data
  var data = response.getDataTable();

  //set the options
  var options = {
      height: 400,
      colorAxis: {colors: ['#ffc850', '#35BDF5']}, //yellow - skyblue
      title: 'Population in 2016'
  };
  var chart = new google.visualization.GeoChart(document.getElementById('pop2016_div'));
  chart.draw(data, options);
} // Population map
```

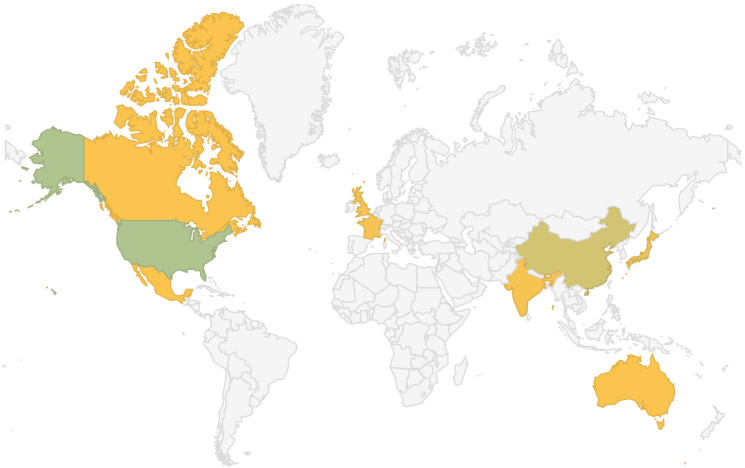


Draw charts by define functions

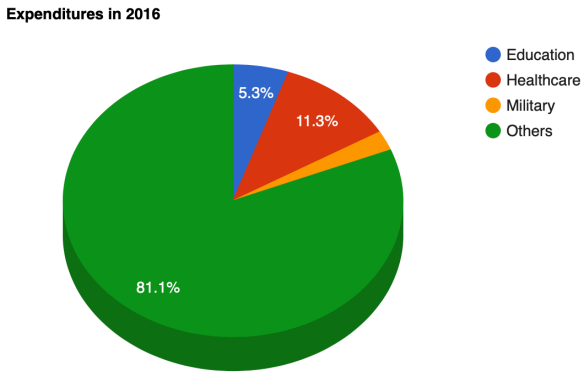
Population in 2016



GDP in 2016 (USD)



Expenditures in 2016 (USD)



Geo Chart

```
var options = {
    height: 400,
    colorAxis: {colors: ['#ffc850', '#35BDF5']}, //yellow - skyblue
    title: 'Population in 2016'};
var chart = new google.visualization.GeoChart(document.getElementById('pop2016_div'));
```

Pie Chart

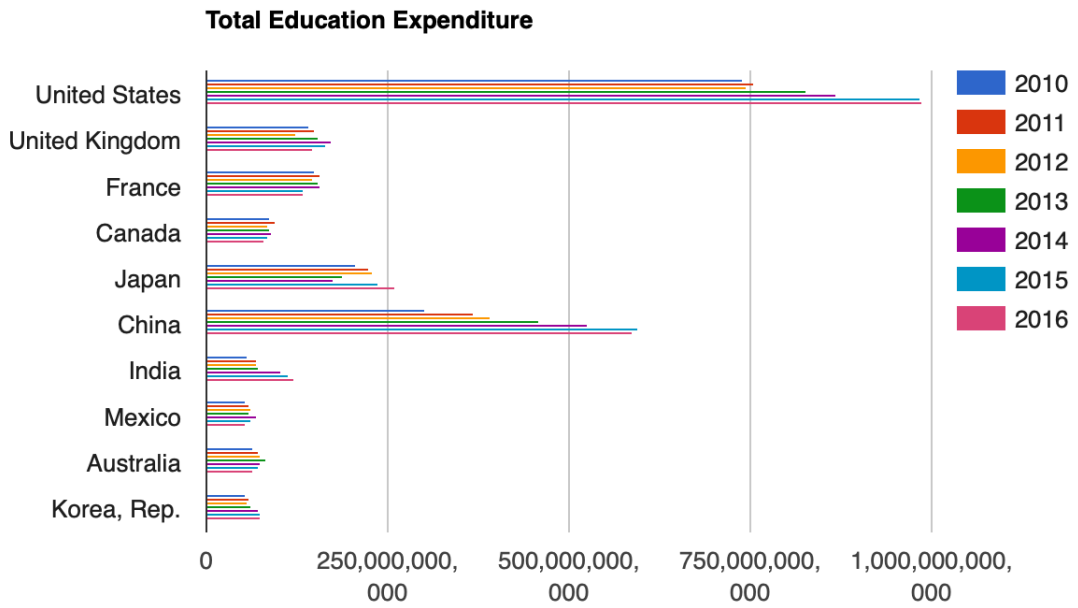
```
var options = {
    title: 'Expenditures in 2016',
    is3D: true,};
var chart = new google.visualization.PieChart(document.getElementById('piechart_3d'));
```

Overview

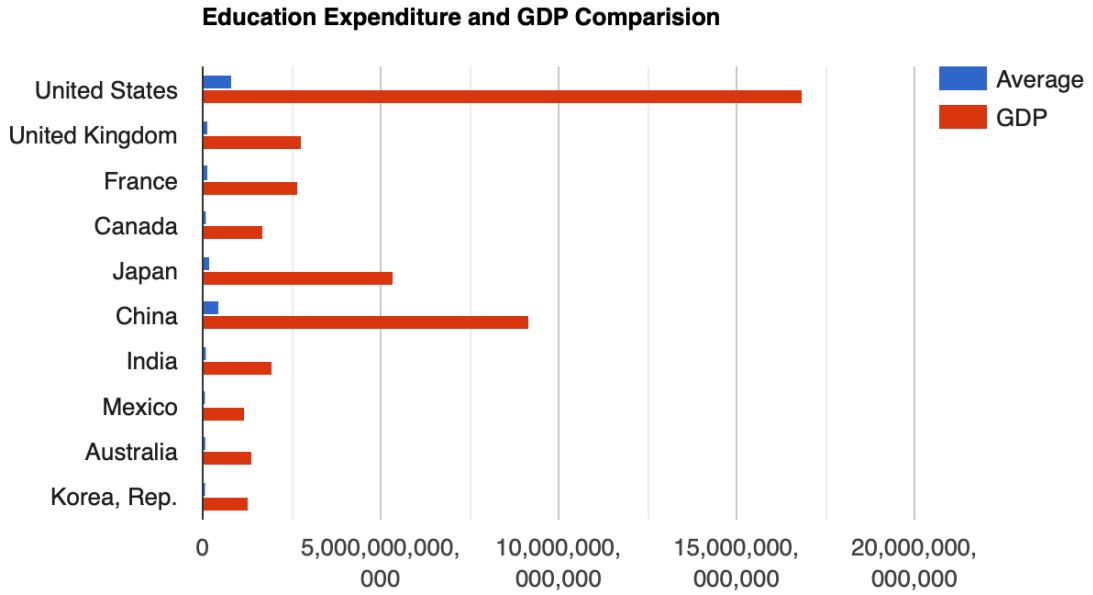
Education

Both the United States and China rank among the top. This is closely related to the relatively developed GDP of these two countries.

Overview of Education Expenditure (USD)



Education Expenditure and GDP Comparison (USD)

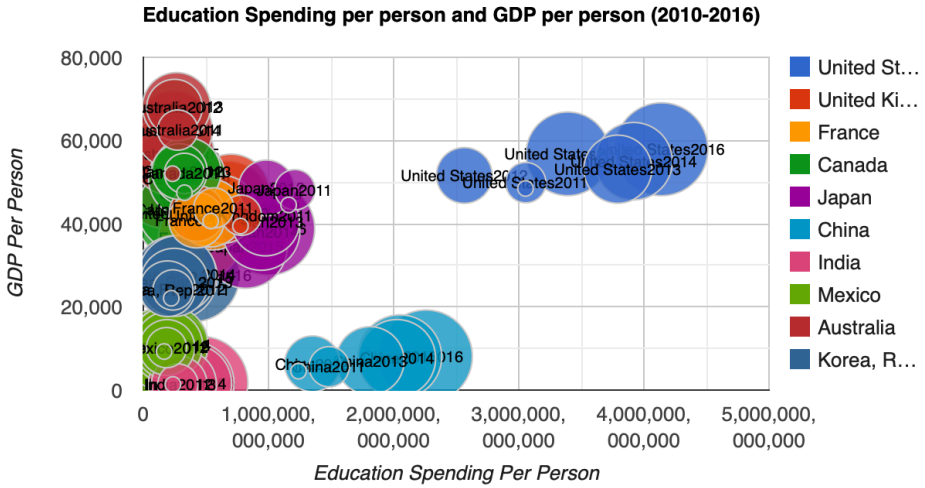



```
var options = {
    height: 400,
    colorAxis: {colors: ['#ffc850', '#35BDF5']}, //yellow - skyblue
    title: 'Total Education Expenditure'

var chart = new google.visualization.BarChart(document.getElementById('Education_div'));
```

Bar Chart

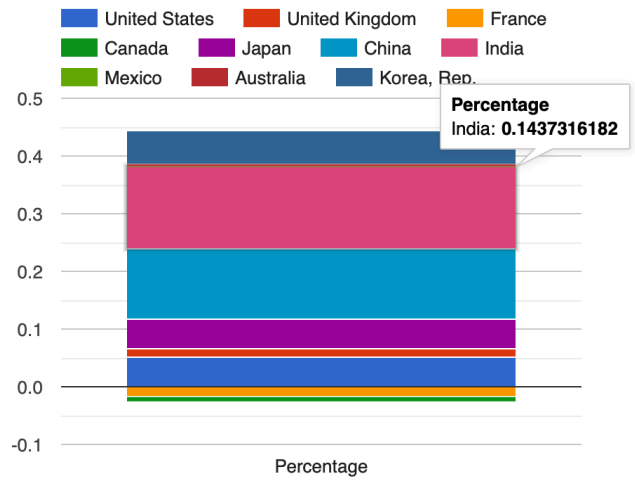
Education

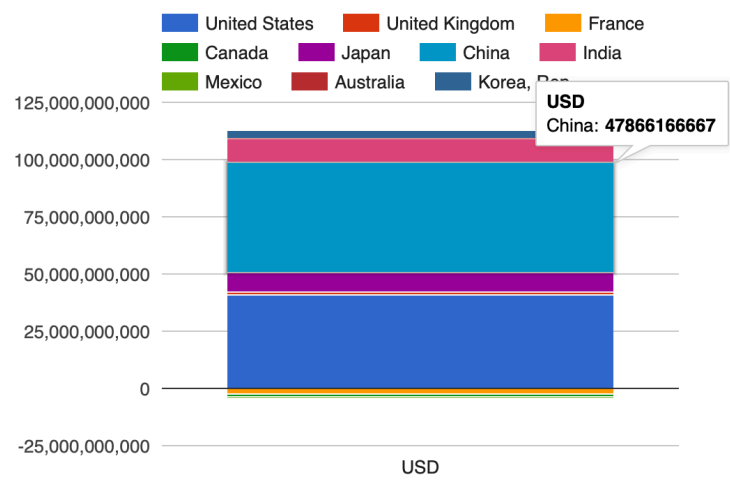
How much an individual values education?




```
var options = {
    title: 'Education Spending per person ' +
        'and GDP per person (2010-2016)',
    hAxis: {title: 'Education Spending Per Person'},
    vAxis: {title: 'GDP Per Person'},
    bubble: {textStyle: {fontSize: 10,auraColor: 'none'}}      };

var chart = new google.visualization.BubbleChart(document.getElementById('Education_GDPer_div'));
```

Bubble Chart





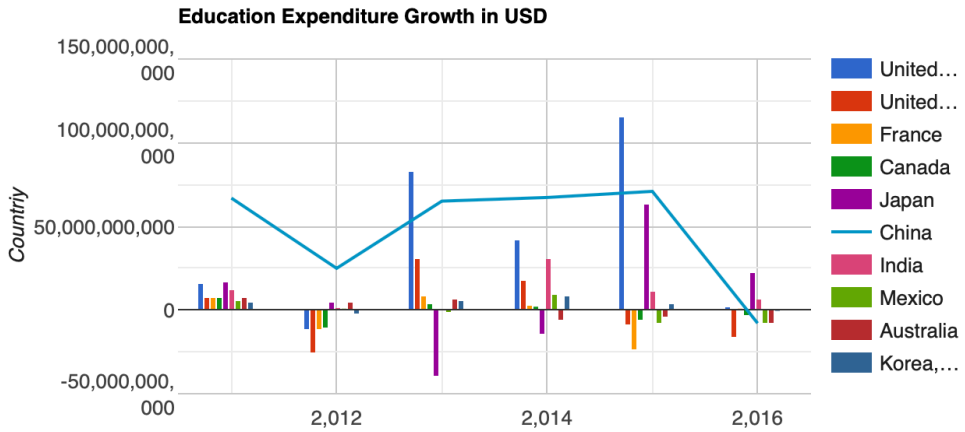

```
var options = {
  width: 600,
  height: 400,
  legend: { position: 'top', maxLines: 3 },
  bar: { groupWidth: '75%' },
  isStacked: true,};
var chart = new google.visualization.ColumnChart(document.getElementById('Education_fast1_div'));
```

Column Chart

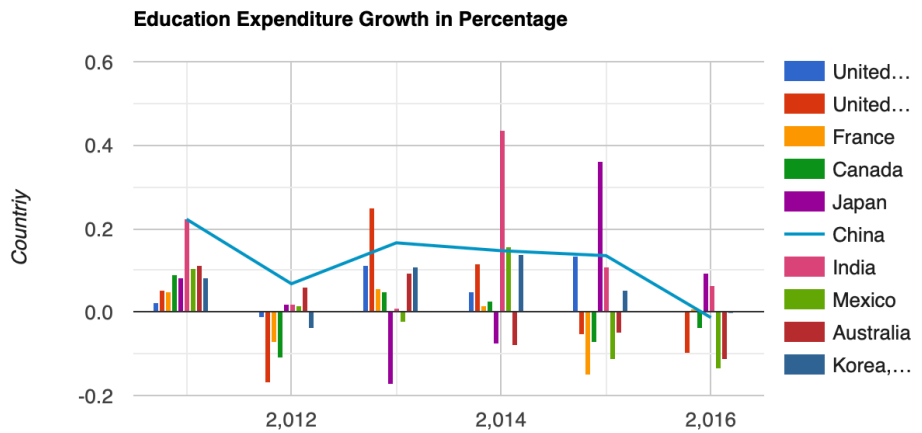
Education

Forecast

Education Expenditure Growth (USD)



Education Expenditure Growth (%)

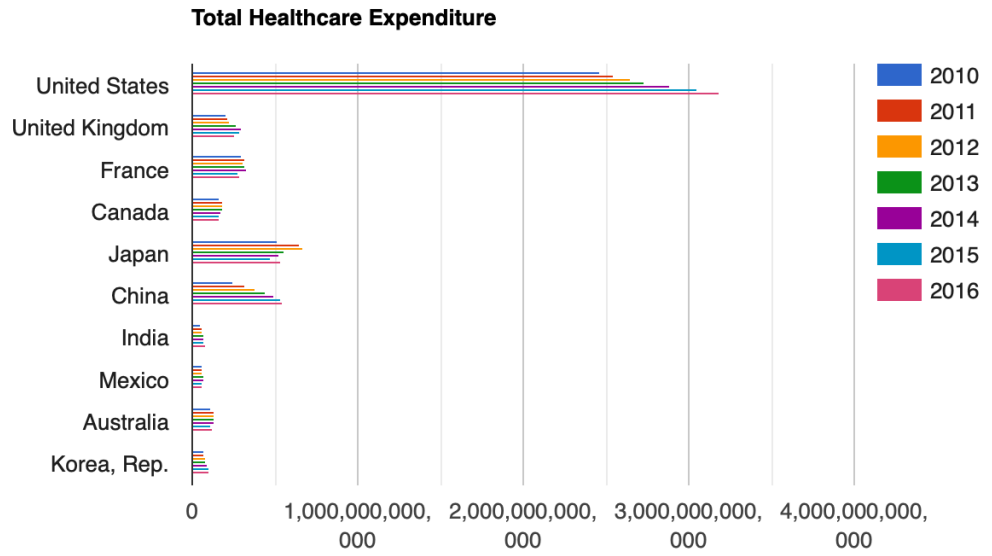



```
var options = {
    title : 'Education Expenditure Growth in Percentage',
    vAxis: {title: 'Countriy'},
    hAxis: {title: 'Year'},
    seriesType: 'bars',
    series: {5: {type: 'line'}}        };
var chart = new google.visualization.ComboChart(document.getElementById('Education_GrowPer_div'));
```

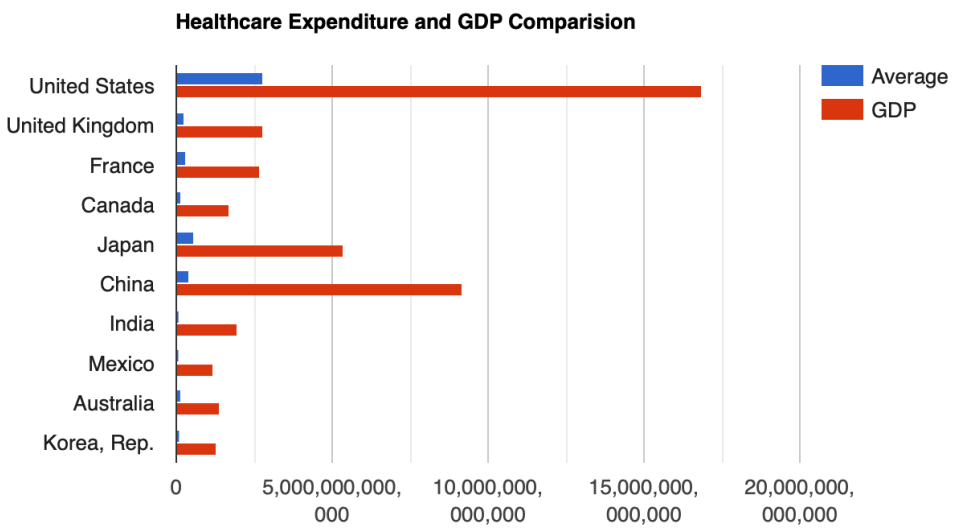
Combo Chart

Healthcare

Overview of Healthcare Expenditure (USD)



Healthcare Expenditure and GDP Comparison (USD)



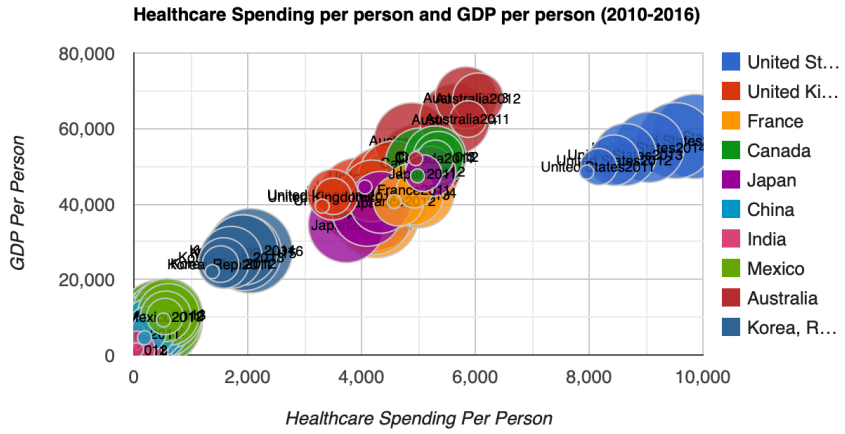

```
var options = {
    height: 400,
    colorAxis: {colors: ['#ffc850', '#35BDF5']}, //yellow - skyblue
    title: 'Total Education Expenditure'

var chart = new google.visualization.BarChart(document.getElementById('Education_div'));
```

Bar Chart

Healthcare

How much an individual values healthcare?

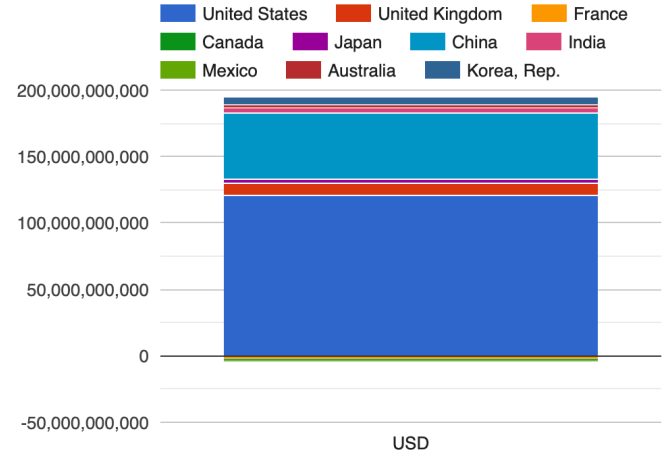
Healthcare Expenditure and GDP (per capita USD)




```
var options = {
    title: 'Education Spending per person ' +
           'and GDP per person (2010-2016)',
    hAxis: {title: 'Education Spending Per Person'},
    vAxis: {title: 'GDP Per Person'},
    bubble: {textStyle: {fontSize: 10,auraColor: 'none'}}      };

var chart = new google.visualization.BubbleChart(document.getElementById('Education_GDPer_div'));
```

Bubble Chart

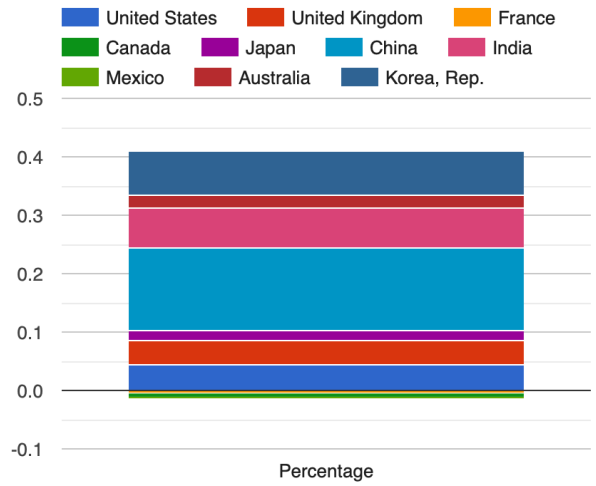
Average Growth Rate 2010 - 2016(USD)




```
var options = {
  width: 600,
  height: 400,
  legend: { position: 'top', maxLines: 3 },
  bar: { groupWidth: '75%' },
  isStacked: true,};
var chart = new google.visualization.ColumnChart(document.getElementById('Education_fast1_div'));
```

Column Chart

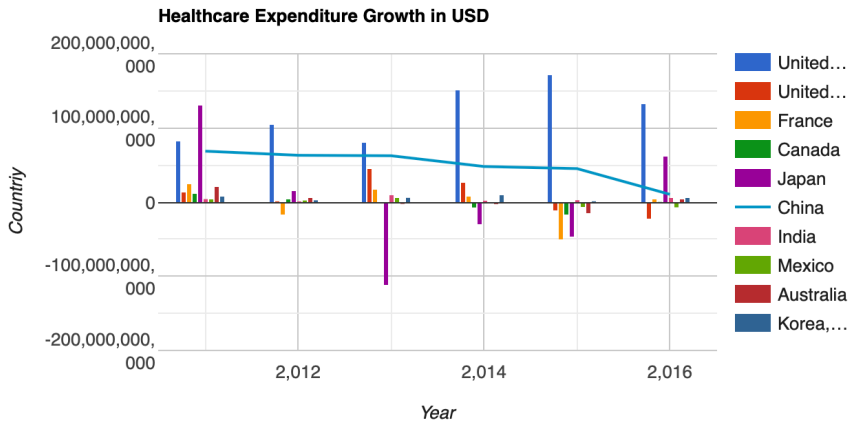
Average Growth Rate 2010 - 2016(%)



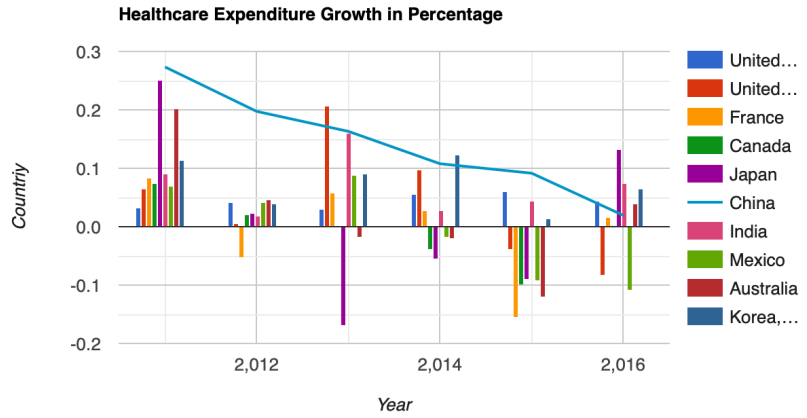
Healthcare

Forecast

Healthcare Expenditure Growth (USD)



Healthcare Expenditure Growth (%)

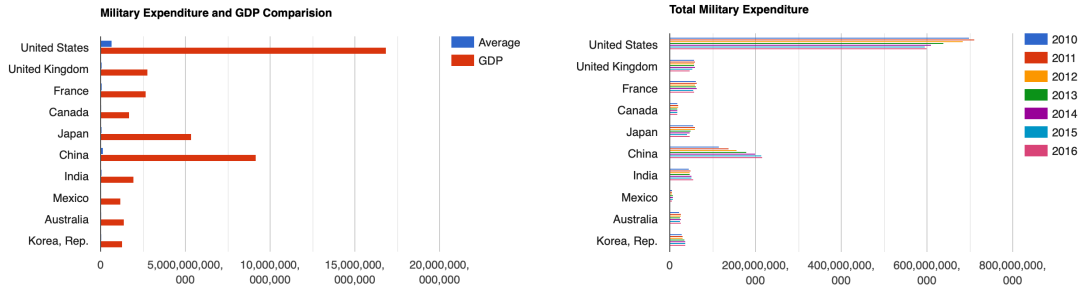


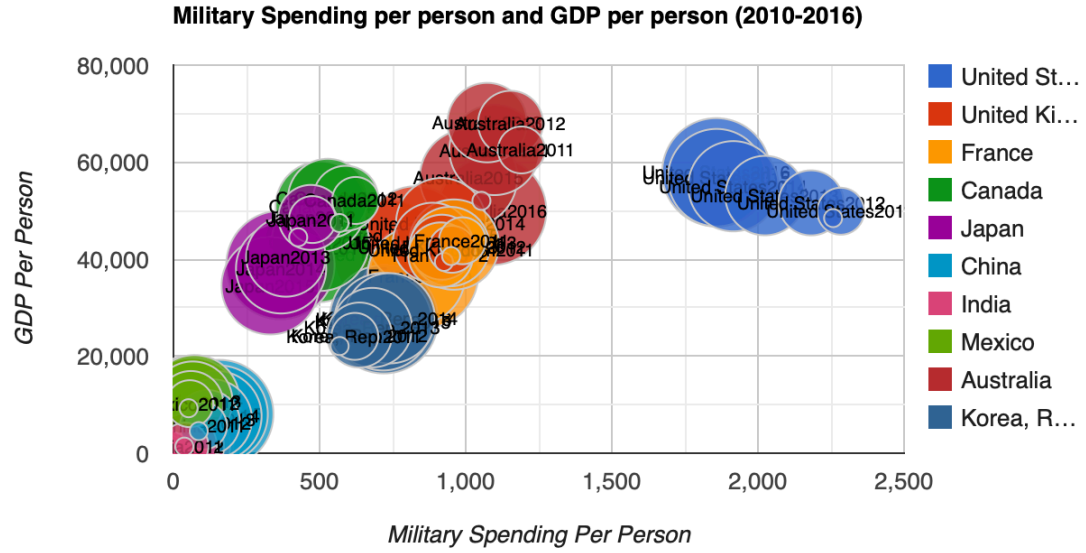

```
var options = {
    title : 'Education Expenditure Growth in Percentage',
    vAxis: {title: 'Countriy'},
    hAxis: {title: 'Year'},
    seriesType: 'bars',
    series: {5: {type: 'line'}}        };
var chart = new google.visualization.ComboChart(document.getElementById('Education_GrowPer_div'));
```

Combo Chart

Military

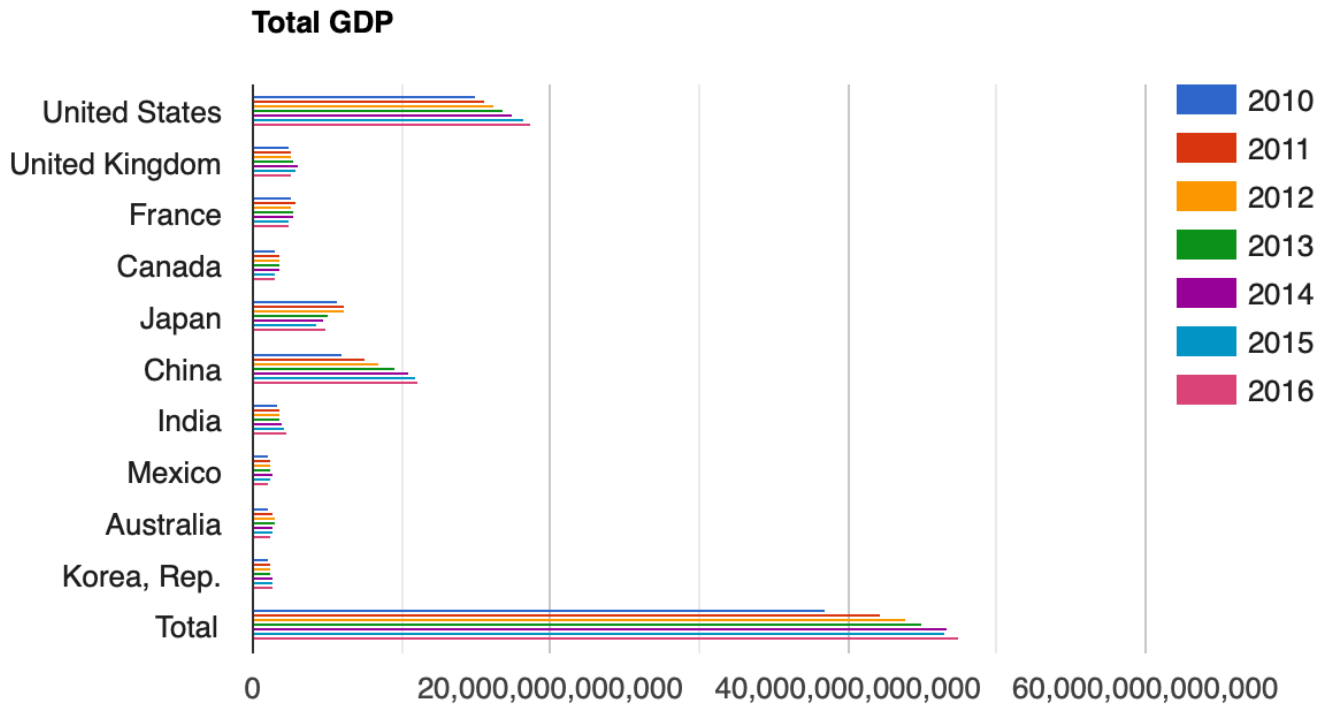
- In military part, we can also see the strong relations between expenditure and GDP. Therefore, I tried to study more about GDP impact.





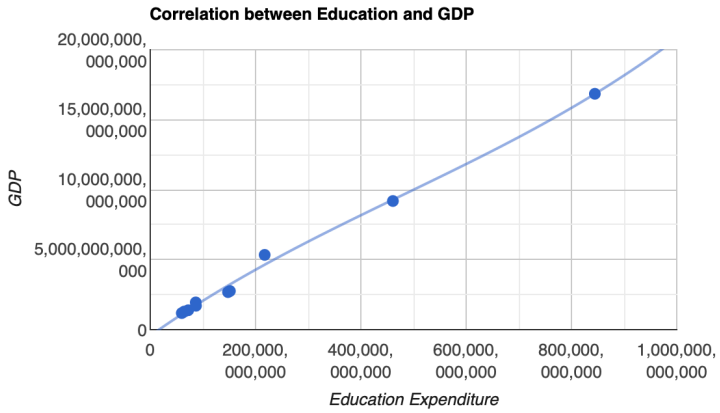
GDP Impact

Overview of GDP (USD)

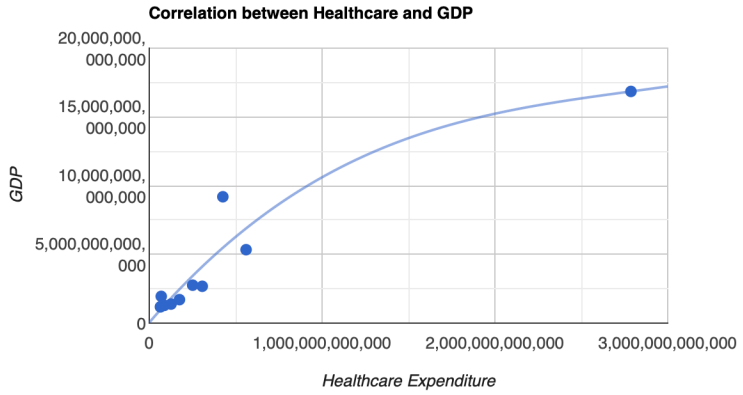


- This bar chart demonstrates that the United States and China rank high in the world in terms of overall GDP.
- According to the total value, world GDP is growing year by year.

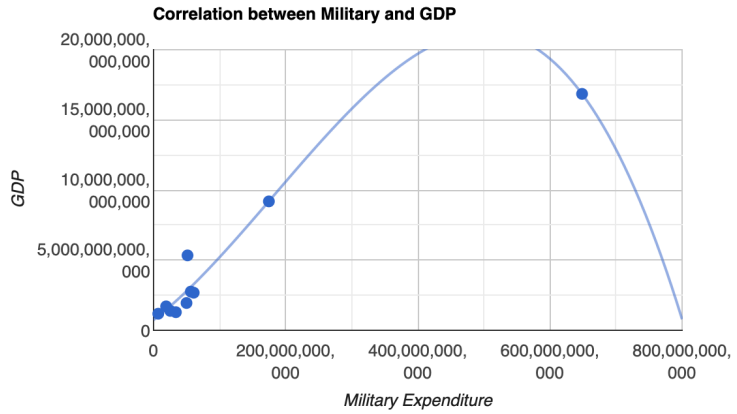
Correlation between Education and GDP (USD)



Correlation between Healthcare and GDP (USD)



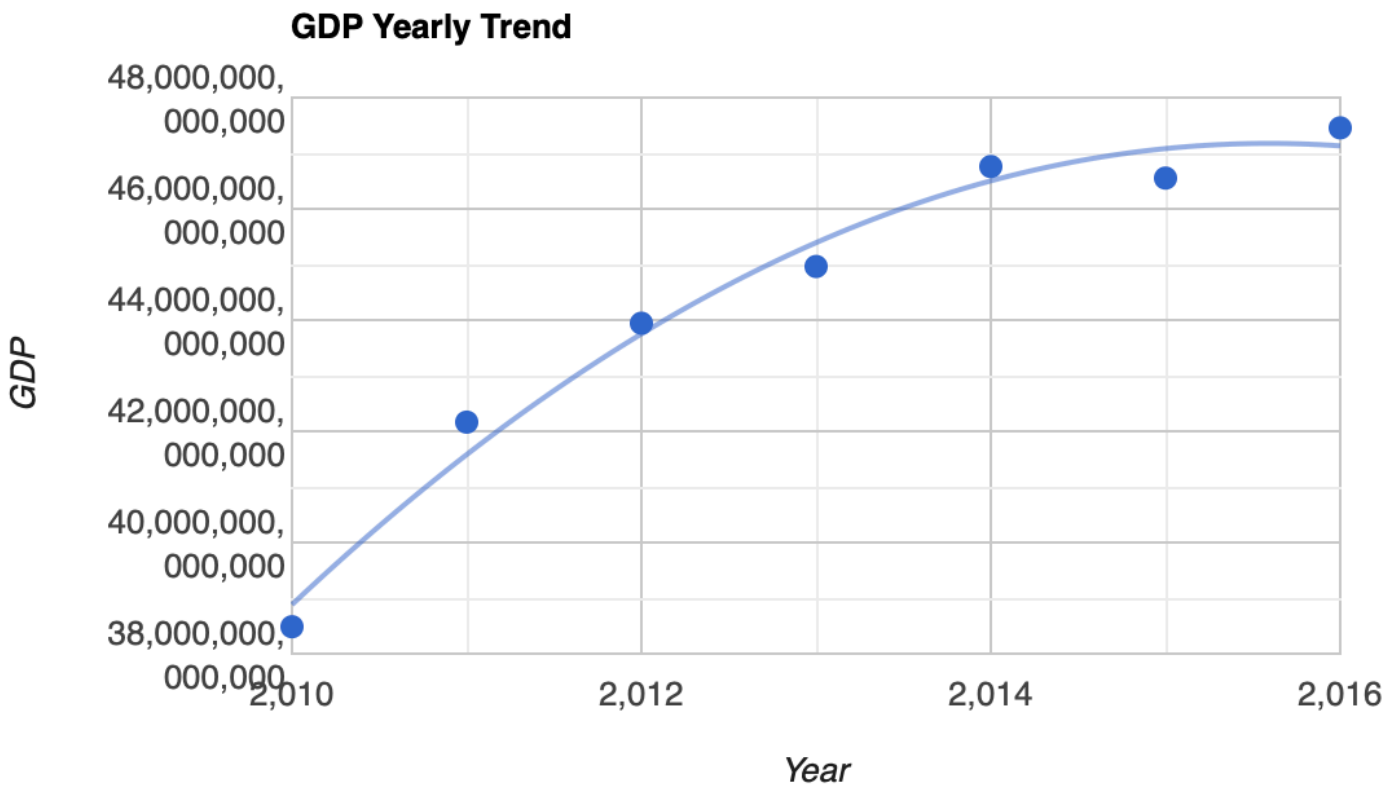
Correlation between Military and GDP (USD)



GDP Impact

- As can be seen from the bar chart, the overall GDP increased year by year from 2010 to 2016. The US has the largest share of GDP, while Mexico and South Korea have the fewest.
- We can see from the graphs of scatter and trendline that spending on education and health care is proportional to GDP. Military spending, on the other hand, begins to move inversely to GDP at a certain point.

GDP Impact - Forecast



- The trend of GDP growth is slow year by year. Based on the relationship between education, healthcare expenditure and GDP, I predict that education and healthcare expenditure will also decline in the future.
- The current peak of GDP has not yet reached the turning point of military, so I believe that military spending will also decline with GDP.

Conclusion

1. The expenditure of Healthcare takes the largest part of all the three categories.
2. All three expenditures have a high correlation with GDP, and the correlation between education and GDP is the highest.
3. The United States is always the top spenders because of its absolute advantage in GDP.
4. China has seen the biggest growth in expenditure and GDP in recent years.
5. China's individual spending on education is outstanding, which is probably the main reason for its fast GDP growth. The relationship between education and GDP bears this out.
6. World GDP will slow down in the future.

References

- World Bank Group: worldbank.org
- Google Charts: developers.google.com/chart
- Medium: medium.com
- Flaticon: flaticon.com
- Google Developer: developers.google.com
- W3School: w3schools.com/

Thanks for watching

Publish Links

- GitHub.io: https://xzhangfox.github.io/
- GitHub: https://github.com/xzhangfox/Visualization-of-Expenditure-from-G-20-Countries
- Zenodo: https://zenodo.org/record/3695940#.Xl7GW5NKjOR