CS 6630 - Visualization for Data Science

Process Book

Visualization of Petrol Statistics by country

Team Members:
Ahmed Al Hikmani, Aniket Karna, Prajeet Reddy Vakiti

Basic Information

Project Title - Visualization of Petrol Stats by Country
Project Website - https://vprajeetreddy-404.github.io/
Project Repository - https://github.com/thatA7MAD/CS6630-project
Project Screen-Cast - Github Readme

Team Members Details-

- Ahmed Al Hikmani, u1305835@utah.edu, u1305835
- Aniket Karna, u1367570@utah.edu, u1367570
- Prajeet Reddy Vakiti, u1365431@utah.edu, u1365431

Overview and Motivation

Crude oil, also known as petroleum, is a naturally occurring liquid mineral that is the primary source of fuel for automobiles and, in some parts of the world, heating oil. The world used or consumed approximately 88.7 million barrels of oil per day in 2020. This figure represents a decrease in consumption when compared to previous years, and it is attributed to the COVID-19 pandemic.

The top ten oil consumers account for roughly 60% of total global oil consumption, with the rest of the world accounting for roughly 40% of total global oil consumption. The United States is both the world's largest producer and consumer of oil, mining approximately 11.5 million barrels per day and consuming more than 17 million gallons per day—more than the entire European Union (9.8 million). This amounts to about 15-20% of total global oil consumption. China consumes 14.2 million barrels of oil per day, followed by India, which consumes approximately 4.7 million barrels per day. The United States, China, and India account for more than one-third of global oil consumption. [src: worldpopulationreview]

Also, it is important that the petrol prices have become a global concern in the past few months.

We thought it would be interesting to visualize gas consumption because it has become such an essential commodity and consumption in each country is so high and diverse. Although there are numerous similar visualizations available, none of them provide a comprehensive overview of all parameters. The figure in *oilandgasinfo*, for example, shows the percentage of global oil consumption, whereas the figure in *worldpopulationreview* shows only the oil consumption per day by country in barrels. There were only a few good visualizations of global gas prices. So we decided to combine all of the parameters into a single interactive visualization. We combined several datasets to create one with the following parameters: daily oil consumption, world share percentage, price per gallon, and price per liter.

Related Work

Worldbank visualization - Pump Price for Gasoline(US$ per litre)

The worldbank.org website had a similar interactive visualization with a line chart, a bar graph and a map. However, we found many shortcomings for this visualization and wanted to make a better version of this visualization. Some of the shortcomings of this visualization which were addressed in our project are:

- At a time, only one of the plots were shown in the worldbank visualization
- The map wasn't interactive and did nothing when we clicked on a particular country in the map
- The bar graph visualization was static and we were unable to extract information about a specific country from it.
- The dataset could only be viewed after downloading it whereas we developed an interactive table that updates every time we click on a country in the map.

Our main motivation for this project was to visualize the increasing gas prices in the past years and gain valuable insights about the gas pricing in different countries.

Questions

The objective of this visualization is to tell the story of how much oil is consumed for each country in 2022.

The end goal is for the user to be able to answer these questions:

- How much oil is used globally per day?
- What countries are the top consumers of oil?
- What is the US share of the world's oil consumption?
- How much is a gallon of oil for each country?
- What is the difference in oil prices from country to country or region?

Some of the new questions that came up as we were working on the project:

- In which year was the gas price the highest for a specific country?
- Which countries had the highest and the lowest gas prices?
- How does country A's gas prices over the years compare to country B's gas prices?
- Which countries had the cheapest oil prices?

What are our learning goals?

- Drawing a dynamic map using Javascript and D3 library, and using the map to tell a story.
- How to add interaction effects into a visualization design while keeping human cognition and perception in mind (color, saturation, and transparency).
- Use what we have learned in the course to develop a complex effective visualization.

Data

The data is collected from:

1. [https://www.kaggle.com/datasets/zusmani/petrolgas-prices-worldwide]. Petrol/Gas Prices Worldwide by Zeeshan-Ul-Hassan Usmani
2. [https://data.worldbank.org/indicator/EP.PMP.SGAS.CD?end=2010&start=1991] German Agency for International Cooperation (GIZ)

The first dataset is in the following format:

- Country: the country from which oil data was recorded.
- Daily Oil Consumption (Barrels): the amount of oil consumption per day.
- World Share: the proportion of the country's oil consumption per day to world oil consumption.
- Yearly Gallons Per Capita: the total annual gallons per capita consumed.
- Price Per Gallon (USD): the price of a gallon of oil in that country.

The second dataset is in the following format:

- Country: the country from which oil data was recorded.
- Country ISO Code: a 3 digit ISO code of the country
- Indicator name - The measurement parameter - US$ per litre
- 1998 - Average gas price in the country in 1997-98.
- 2000- Average gas price in the country in 1999-00.
- 2020- Average gas price in the country in 2019-20.

Data Preprocessing

The data had country names columns and the ISO code columns were not present in the initial dataset. We needed the ISO code column for the map and thus we searched for a dataset that had country names with their iso codes and used pandas library to match and create a new column for the ISO code in our dataset.

We didn't expect substantial data cleanup or processing at the beginning of the project, however, we found many rows that had missing data in the second dataset and many countries for which there was no data. We decided to fill those missing cells with the average of the previous and the next cell in that row as the gas price couldn't have changed drastically in between those years for most of the countries. In addition, there were many countries that had inconsistent values(same gas price for 20 years), so we decided to remove those countries.

The second dataset had columns that represented year-wise gas price for each country, however, we had to use the pandas melt() function to break the column data into rows to develop the line-chart.

Exploratory Data Analysis

We performed EDA on our datasets using the pandas library.

First Dataset(processedDataset.csv):

```
df_first.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 187 entries, 0 to 186
Data columns (total 13 columns):
 #   Column                                Non-Null Count  Dtype
---  ------                                --------------  -----
 0   S.No                                  185 non-null    float64
 1   S#                                    185 non-null    float64
 2   Country                               187 non-null    object
 3   Daily Oil Consumption (Barrels)       187 non-null    object
 4   World Share                           187 non-null    int64
 5   Yearly Gallons Per Capita             187 non-null    float64
 6   Price Per Gallon (USD)                187 non-null    float64
 7   Price Per Liter (USD)                 187 non-null    float64
 8   Price Per Liter (PKR)                 187 non-null    float64
 9   GDP Per Capita ( USD )                187 non-null    object
 10  Gallons GDP Per Capita Can Buy        187 non-null    object
 11  xTimes Yearly Gallons Per Capita Buy  187 non-null    float64
 12  iso_code                              187 non-null    object
dtypes: float64(7), int64(1), object(5)
memory usage: 19.1+ KB
```

The first dataset is used for plotting the bar graph and the map. The interactive table shown on the website on the bottom right shows this dataset with the relevant columns.

Second Dataset(cleanedPetrolData.csv):

```
In [18]: df_second.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2712 entries, 0 to 2711
Data columns (total 5 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   Unnamed: 0        2712 non-null   int64
 1   Country Name      2712 non-null   object
 2   Country ISO Code  2712 non-null   object
 3   year              2712 non-null   int64
 4   pump price        2712 non-null   float64
dtypes: float64(1), int64(2), object(2)
memory usage: 106.1+ KB
```

The second dataset is used to plot the time-series line chart on the bottom left of the webpage. It stores the pump price of each country from 1998-2020. By default, the line-chart shows the countries shown on the bar-chart. If a country is clicked on the map, then the line-chart is updated with that country's pump price data points.

Visualization Design

Prototype 1

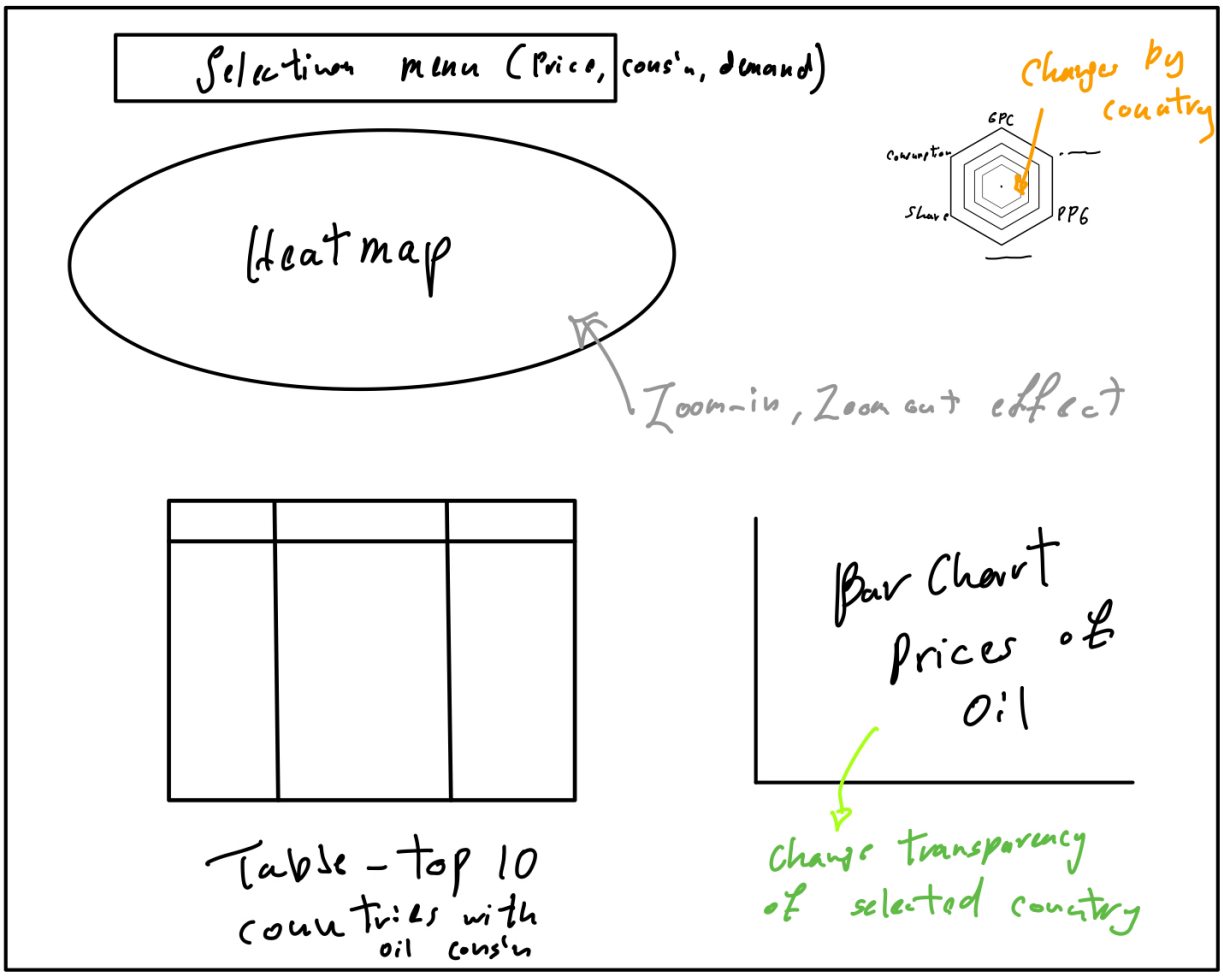


Figure 1: Prototype 1

- The global heatmap is in WinkleTriple projection to minimize area, distance, and angle distortions.
- The selection menu includes radio buttons to select a different scale (price, consumption, or YGPC) on the map, with a pre-selected option.
- The radar chart displays a selected country's data from either the heatmap or the barchart, with the option to add more selected countries for comparison.
- The table includes the countries with their data, and a selected country from the map should be displayed and highlighted in the table.

- The bar's length in the bar chart represents oil prices for each country and so the user can determine which country has the highest oil price.
- The bar chart should highlight the selected country from the heatmap and set the other bars transparent.
- The bars in the chart could be colored according to their country's YGPC as the length represents the price of oil.

Prototype 2

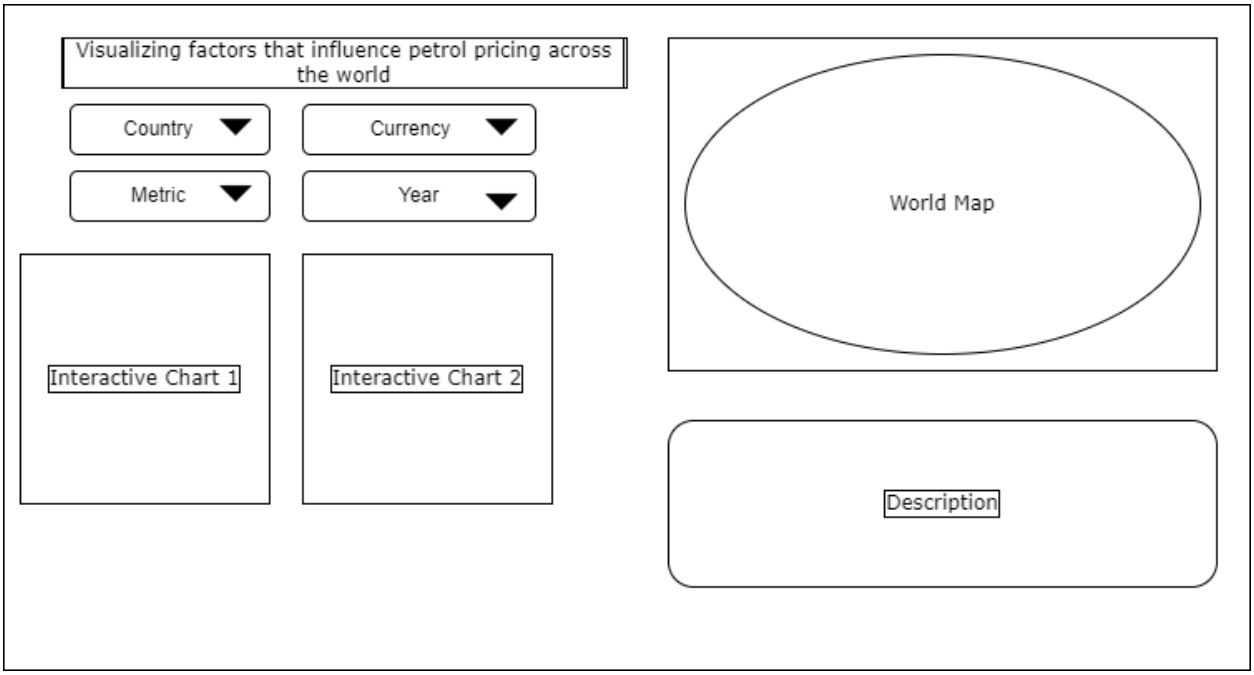


Figure 2: Prototype 2

- Drop-down selection is used because it helps conserve screen space.
- Interactive charts update on clicking a particular country on the world map.
- Using Leaflet, the map should support zooming-in and zooming-out features.

Prototype 3

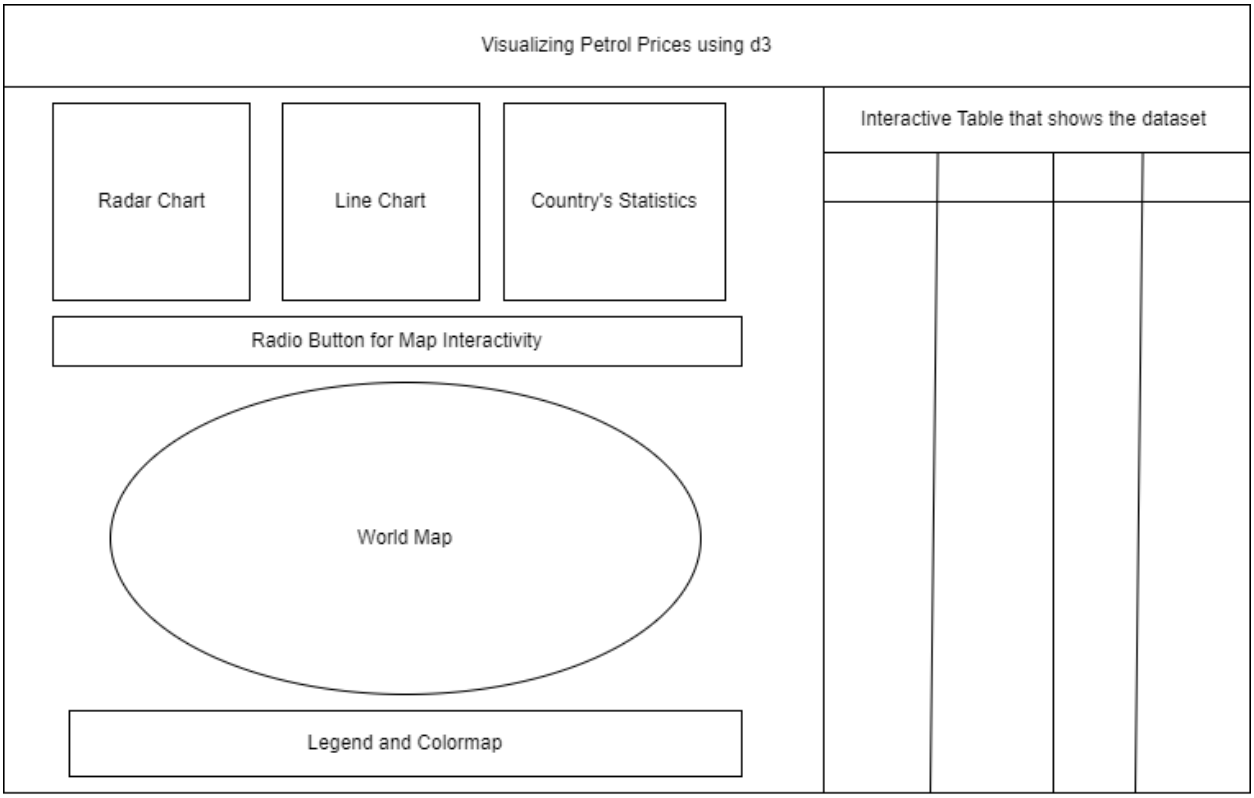
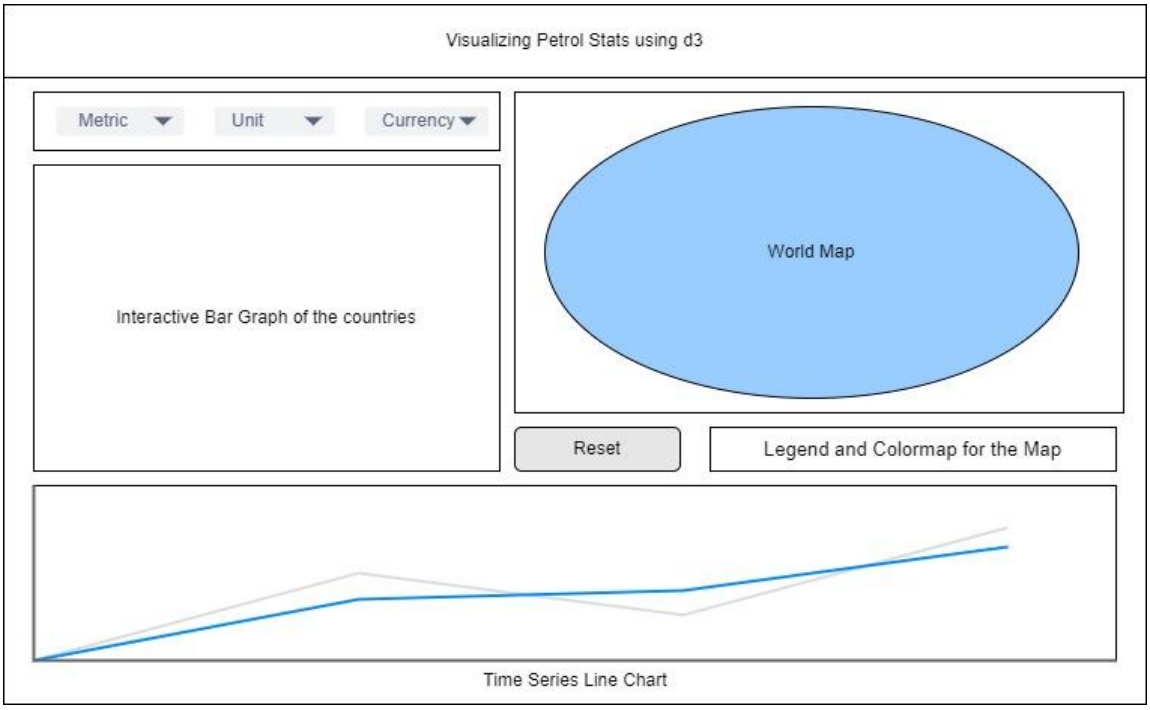


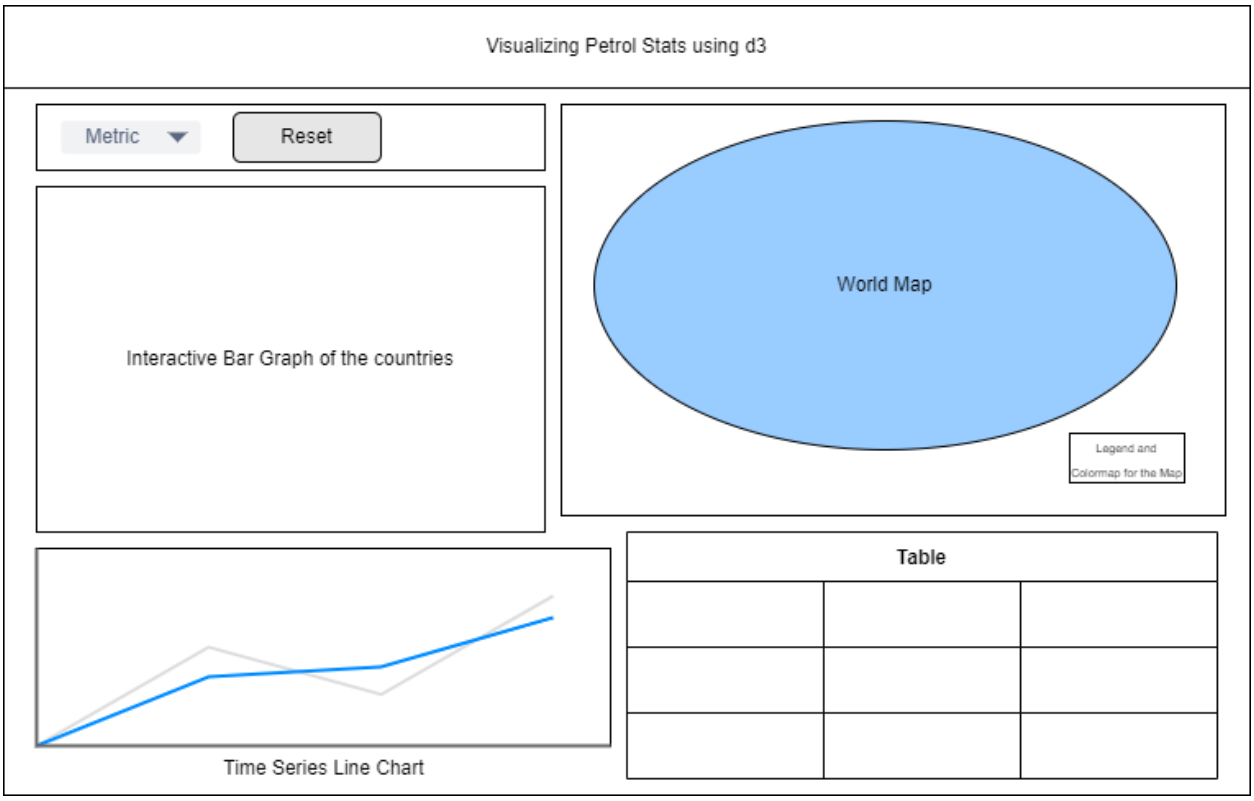
Figure 3: Prototype 3

- This prototype supports sorting countries on the interactive table according to the name, gas price, population, gdp, etc.
- The radar chart is used to compare two different countries' statistics.
- Top 10 and bottom 10 countries on the Line Chart, Country’s statistics, etc.

Initial Design



Final Design:



Evolution of our Design:

We considered many visualization designs shown in our prototype 1,2, and 3. We combined these three prototypes to develop our initial design at the time of project proposal. Our initial design had more drop down buttons but it didn't have a table. After getting the peer review feedback and inputs from our TA mentor Tripti, we decided to add an interactive table that would update based on the country selected on the map. We initially thought of a colormap which goes from yellow to red however after experimenting with more colormaps, we decided to go with the current colormap which goes from white to blue.

Features:

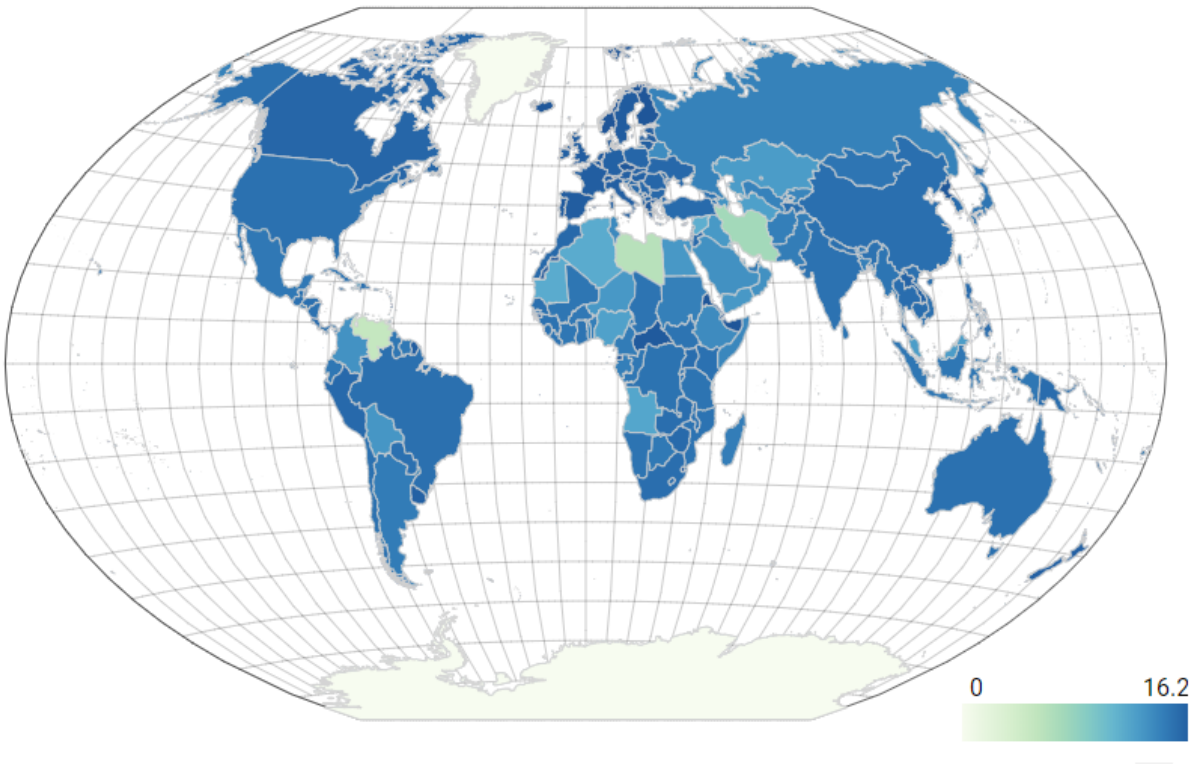
- The drop down menu on the top left includes a metric: The metric drop down has multiple options to choose from, like price per gallon, daily oil consumption, world share, yearly gallons per capita, etc.
- The global heatmap is in WinkleTriple projection to minimize area, distance, and angle distortions. It will be updated based on the input from the drop down menu selections.
- The bar chart shows the price for the top 10 countries with the highest gasoline prices across the world. This bar chart gets updated if a specific country is clicked on the map.
- The time series plot will include lines for each selected country from the world map.
- The table on the bottom right shows the dataset with columns as Country Name, Price of gas in USD, Daily Oil Consumption, and Yearly gallons per capita. The columns can be sorted after clicking on any of these four column names.
- The "Reset" button deselects all selections from the world map and shows the default views for each of the individual visualizations on the webpage.

Implementation:

Interactive Map:

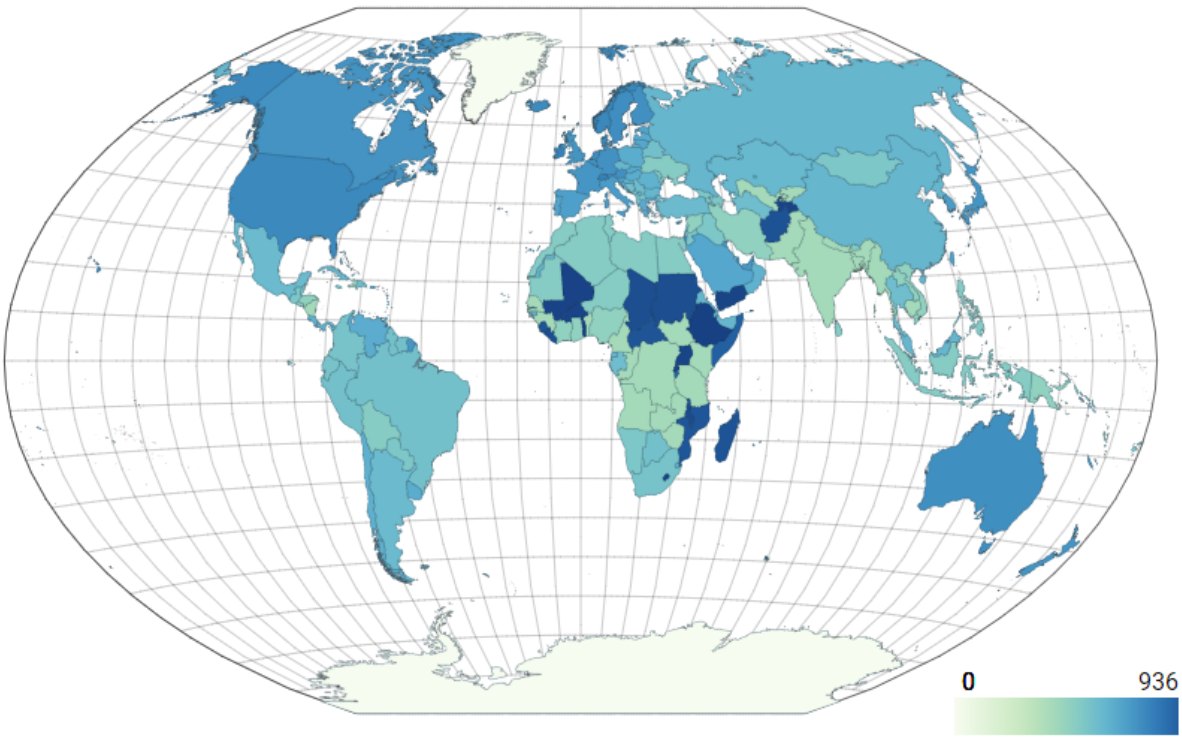
The first thing we implemented was the geowinkel3() map projection. We followed the same principles that were taught during the lectures to develop the map that changes every time the metric on the top left is toggled.

Map that shows the price metric distribution:



The map shows that gas prices were lowest in Venezuela, Libya, and Iran and highest in North Korea, and Scandinavian countries.

Map that shows the GDP per capita(USD) metric distribution:

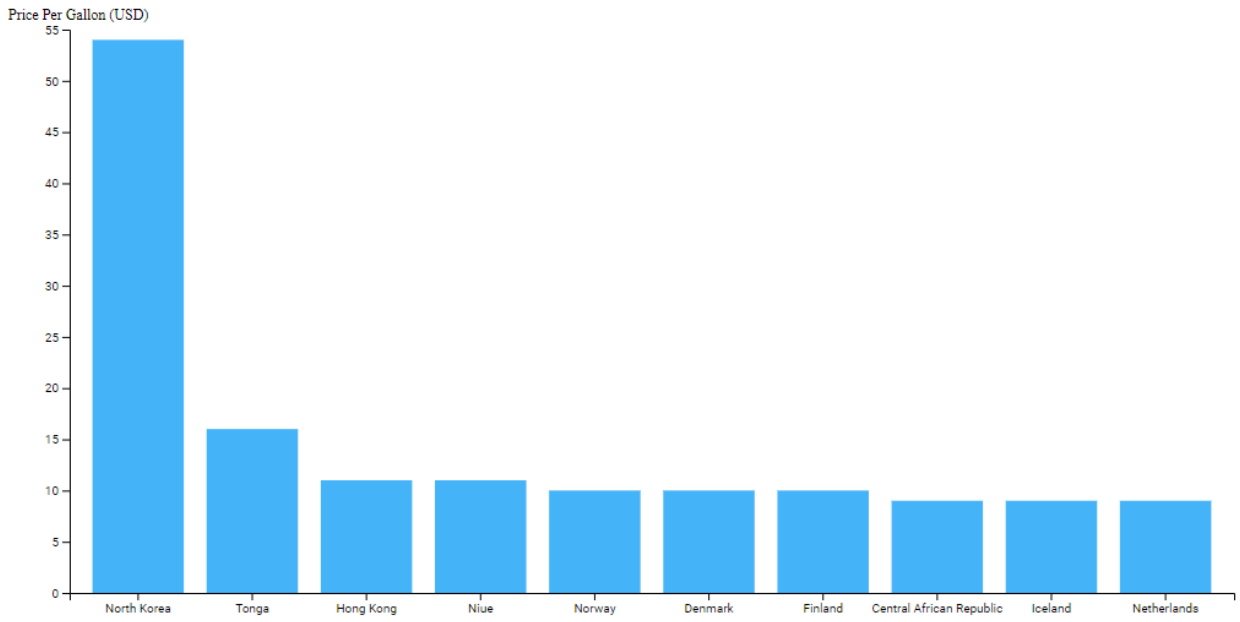


The map shows that GDP per capita was the highest in Ethiopia, Togo, and other African countries and was relatively low in India, Pakistan, Mexico.

Bar-Chart:

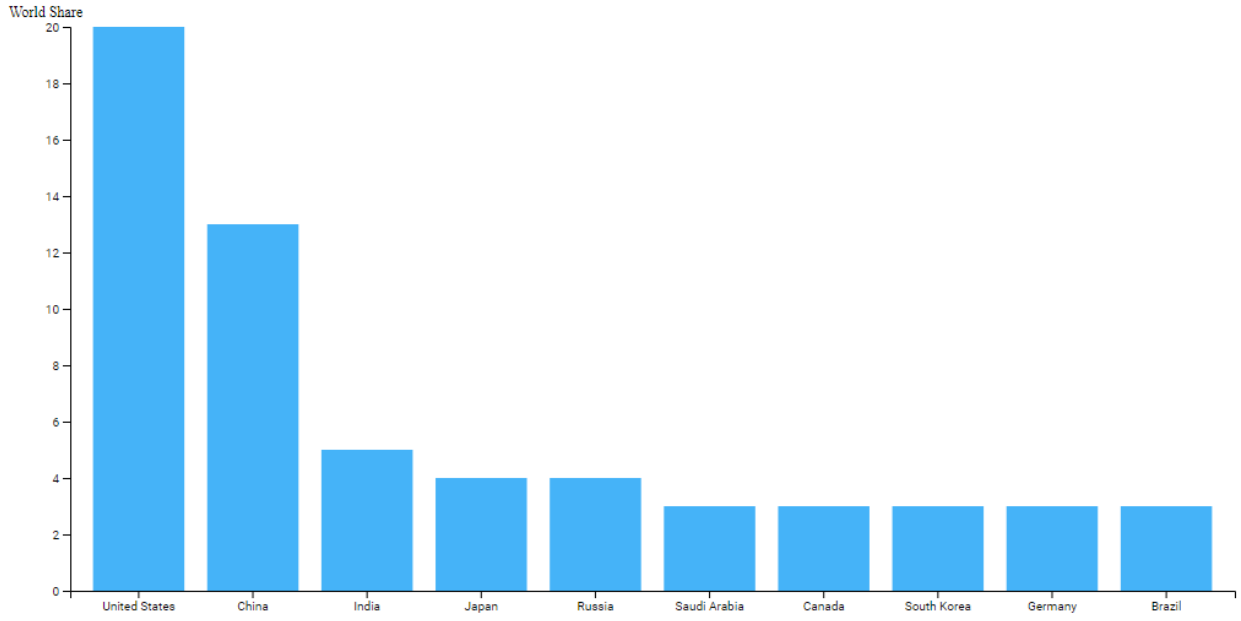
Next, we worked on the bar chart and displayed the top 10 countries for each metric. We worked on the interactivity of the map and the bar chart so that the bar chart updates every time we click on a country on the map.

Bar-chart that shows Price per gallon vs Country



North Korea had the highest gas price by far followed by Tonga and Hong Kong and Scandinavian countries.

Bar-chart that shows World Share vs Country



The chart shows that the US and China have the highest world share of gasoline followed by India which has around 4.5% of the total world share.

Interactive Table:

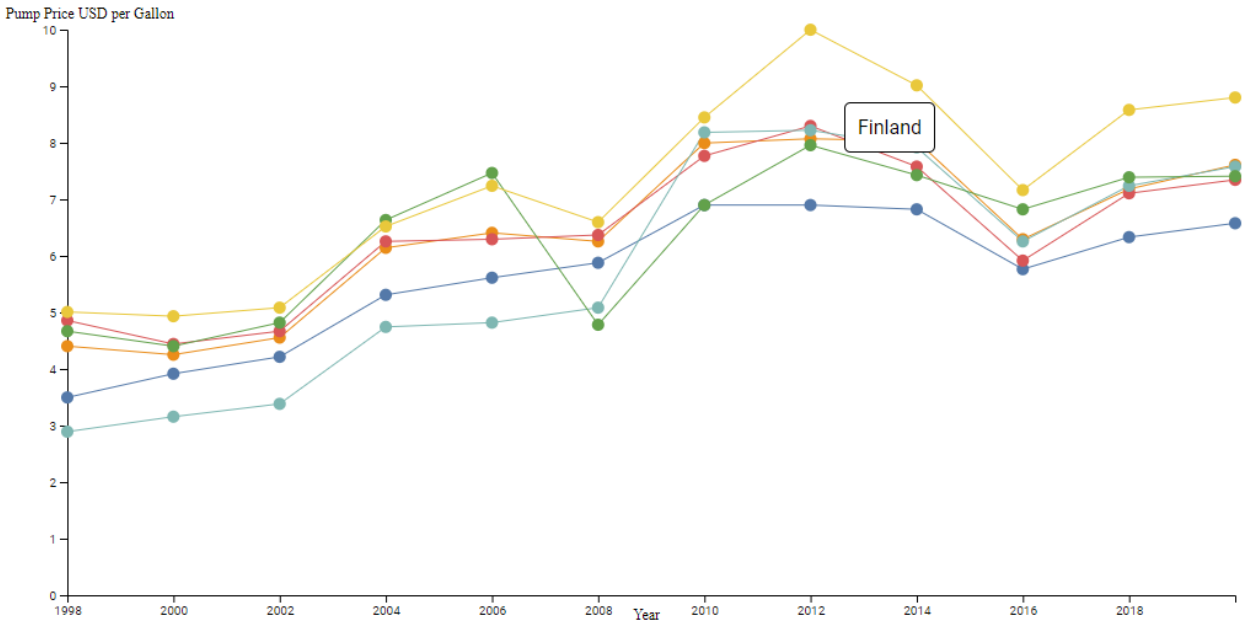
The table was implemented next by taking help from the homework assignment 5. We are displaying the dataset that has the country name, Price per gallon(USD), Daily oil consumption(Barrels), and Yearly gallons per capita as the columns.

Country	Price (USD)	Daily Oil Consumption (Barrels)	Yearly Gallons Per Capita
North Korea	54.89	18,000	10.9
Tonga	16.2	899	136.3
Hong Kong	11.35	408,491	864.5
Niue	11.43	51	484.4
Norway	10.22	204,090	595.8
Denmark	10.04	158,194	424.6
Finland	10.01	210,030	585.7
Central African Republic	9.06	2,800	9.5
Iceland	9.83	19,090	880.9
Netherlands	9.33	937,098	846
Greece	9.49	296,101	427.6
Switzerland	8.27	228,194	417.5
Singapore	8.71	1,357,000	3679.5
Sweden	8.7	322,109	502
Austria	8.07	262,352	459.8
Belgium	8.36	631,522	852.6

The table shows our dataset and gets sorted when we click on any of the column names. If a country is clicked on the map, the table shows the statistics for that particular country.

Time-Series Line Chart:

The time-series line chart was implemented next that shows the gas prices from 1998 till 2020 for each country clicked on the map.



The above time-series line chart shows the trajectory of gas prices in Greece, Saudi Arabia, Finland, Central African Republic, Denmark, Norway, Iceland. Each of the county's lines can be traced by hovering the mouse over their lines or by just clicking on that country on the map which updates the line chart.

Evaluation:

We can draw many valuable inferences from the dataset after using our visualization. Some of them are as follows:

- The gas prices were lowest in Venezuela, Libya, and Iran and highest in North Korea, and Scandinavian countries.
- We can compare the statistics of each country in the map with any other country.
- The US has a world share of around 20 percent of the total gas consumption.
- The time-series line chart can show the year in which the gas price for a country was highest and lowest.
- The daily oil consumption was the highest in the USA, China, India, Japan, and Russia and lowest in Antarctica, Greenland, Niue, Saint Helena, Montserrat.
- The yearly gallons per capita was the highest in Singapore(3679.5), Saint Pierre(1705.1), Malta(1652.2), and Saudi Arabia(1560.2) and lowest(less than 4) in Antarctica, Greenland, Burundi, and Congo.

From the line-chart, we can see that:

- The US had the highest gas price in 2012(around 4$) and lowest in 1998(1.4$) per gallon.
- Canada had the highest gas price in 2012(around 5$) and lowest in 1998(1.7$) per gallon.
- India had the highest gas price in 2012(around 5$) and lowest in 1998(2.3$) per gallon.
- China had the highest gas price in 2012(around 5.5$) and lowest in 1998(1.3$) per gallon.

This shows that the gas prices kept on increasing with time after 1998 and peaked in 2012 after which they started to decrease. The peak in 2012 was due to hurricanes, refinery outages, and tensions in the middle east.