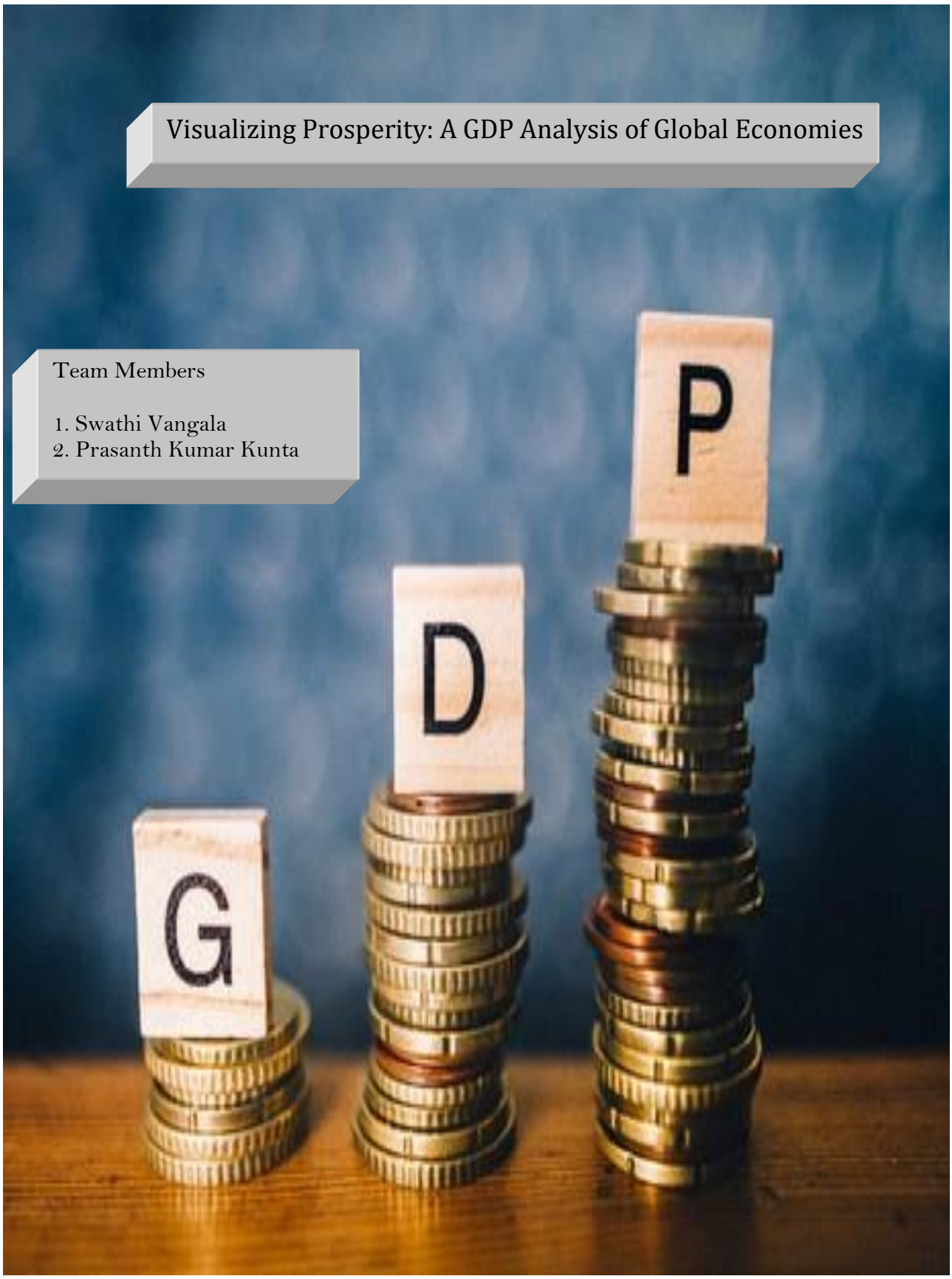
Visualizing Prosperity: A GDP Analysis of Global Economies

Team Members

1. Swathi Vangala
2. Prasanth Kumar Kunta



BACKGROUND

Why GDP?

Our objective was to choose a topic that not only allows us to demonstrate the data visualization techniques we had learned, but also supports a compelling narrative that unfolds as we delve into the data.

Gross Domestic Product (GDP) emerged as an ideal choice due to multiple reasons. First, GDP data is widely available from a variety of reliable sources. Second, it offers a wealth of insights into the economic health of countries, enabling us to rank them and uncover surprising trends.

We were intrigued by the potential stories that GDP data could tell. For instance, does GDP accurately reflect a country's overall economic health? Can it illuminate how international trade can transform a country's internal and external economic landscape? Are there discrepancies between the perceived wealth of a nation and its actual GDP?

These were some of the compelling questions that we hoped to explore as we began our data-driven journey into the world of economics.

As we move forward, we aim to delve deeper into these queries and more, using the power of data visualization to bring these economic narratives to life.

With this project, we hope to uncover surprising insights and challenge common misconceptions about wealth and economic health on a global scale.

In doing so, we wish to emphasize the importance of data-driven decision-making in the realm of economics and beyond.

Data

Data Sources

For our project, we relied on data from the following reputable sources:

1. Kaggle:

   A data science community that provides numerous publicly available datasets.

2. CIA World Factbook:

   An annual publication by the Central Intelligence Agency of the United States with comprehensive details about the countries of the world.

3. World Bank:

   An international financial institution that provides access to extensive financial and economic data about countries worldwide.

The datasets contained information of about 266 countries and regions from 1960 to 2020.

Data Cleaning

Our data cleaning process was meticulous and thorough, ensuring that our analysis was based on high-quality, reliable data. Here are the steps we took:

Handling Missing Values:

The datasets contained NaN values that required careful handling. Depending on the proportion of missing values for a particular country, we either dropped the data or replaced the NaN values.

Dropping Countries:

We opted to exclude countries with data missing for more than 25% of the years. This decision was made to ensure the robustness of our analysis.

Filling Missing Data:

For the remaining missing data, we employed linear interpolation, forward fill, and backward fill techniques to fill the gaps. These methods were chosen over replacing with mean or median values to maintain the chronological integrity of the data.

Data Type Conversion:

We converted the GDP data type for all countries to 'float' for consistency throughout our analysis.

Slicing the Data:

To accommodate the maximum number of countries with GDP data, we restricted our dataset to the years 1980 to 2020.

By adhering to these steps, we ensured that our data was clean, consistent, and ready for meaningful analysis.

Snapshot of final cleaned data

Country Name	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	
Aruba	4.054634e+08	4.054634e+08	4.054634e+08	4.054634e+08	4.054634e+08	4.054634e+08	4.876025e+08	5.964236e+08	6.953044e+08	7.648871e+08	8.721
Africa Eastern and Southern	1.603781e+11	1.546694e+11	1.597575e+11	1.460219e+11	1.304398e+11	1.470252e+11	1.797395e+11	1.890033e+11	1.945434e+11	2.123361e+11	2.207
Africa Western and Central	2.110035e+11	1.871637e+11	1.381152e+11	1.142627e+11	1.165073e+11	1.074975e+11	1.103218e+11	1.089435e+11	1.017688e+11	1.218022e+11	1.174
Angola	5.553824e+09	5.553824e+09	5.787824e+09	6.135166e+09	7.558613e+09	7.076794e+09	8.089279e+09	8.775116e+09	1.020792e+10	1.123628e+10	1.123
Albania	1.857338e+09	1.857338e+09	1.857338e+09	1.857338e+09	1.897050e+09	2.097326e+09	2.080796e+09	2.051236e+09	2.253090e+09	2.028554e+09	1.099

VISUALIZATIONS

Choropleth Graph

Technique description:

1. **Geo-Data Loading and Projection**:

The script uses the D3.js `d3.queue()` function to load the geographic data (worldmap) and the population data (worldpopulation). The world map is represented in GeoJSON format, a type of JSON file that includes geographic data. The GDP data is in CSV format. The map is projected using the Robinson projection (`d3.geoRobinson()`), a compromise projection used for world maps.

2. **Choropleth Coloring**:

The script uses a threshold scale (`d3.scaleThreshold()`) to map GDP values to colors. This scale divides the domain (the GDP values) into segments based on the provided thresholds and maps each segment to a color from the specified color range (`d3.schemeSpectral[7]`).

3. **Tooltip Functionality**:

A tooltip is an overlay that displays additional information when users hover over specific elements. In this map, when a user hovers over a country, a tooltip appears displaying the country's name and its GDP. This is achieved using mouseover and mouseleave event listeners.

4. **Zoom Functionality**:

The script includes a zoom function that scales and translates the map based on the user's selection. When a user clicks on a country, the map zooms in on that country. When the user clicks again, the map zooms out to the initial view.

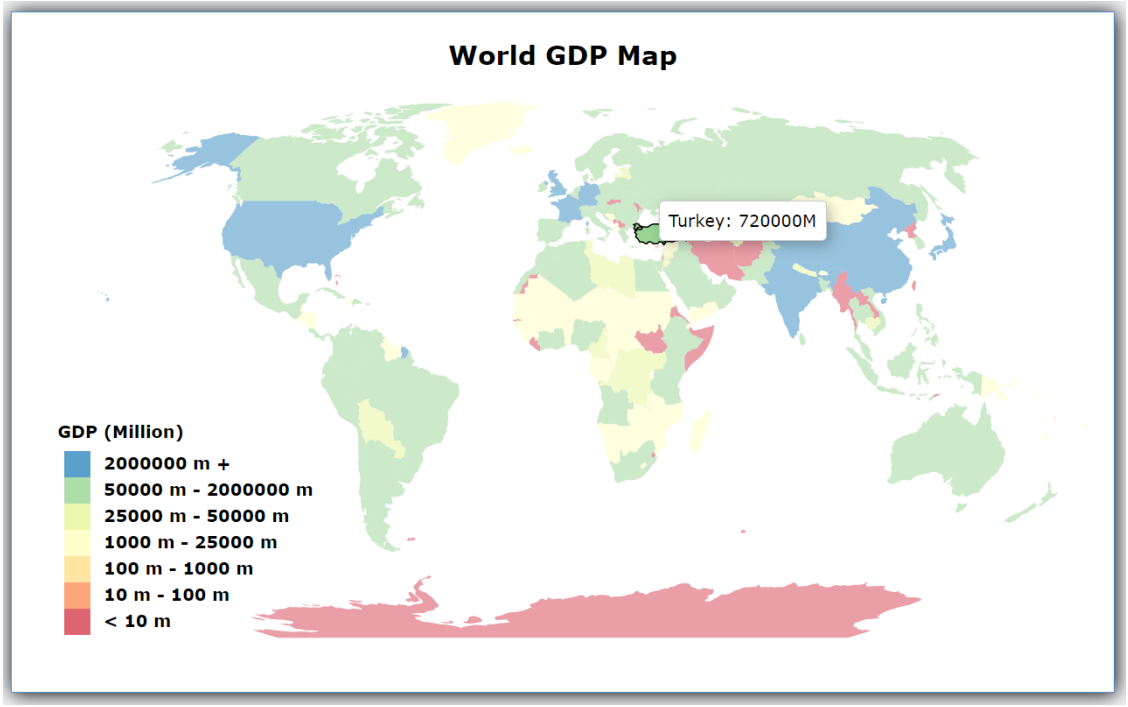
5. **Dynamic Legend**:

The script generates a legend that explains the color coding of the map. The legend is created dynamically based on the color scale and the GDP thresholds.

In summary, this script provides a powerful and interactive way of visualizing GDP data across different countries. The use of a Choropleth map allows for easy comparison of GDP across countries, while the tooltip and zoom functionalities provide additional context and detail.

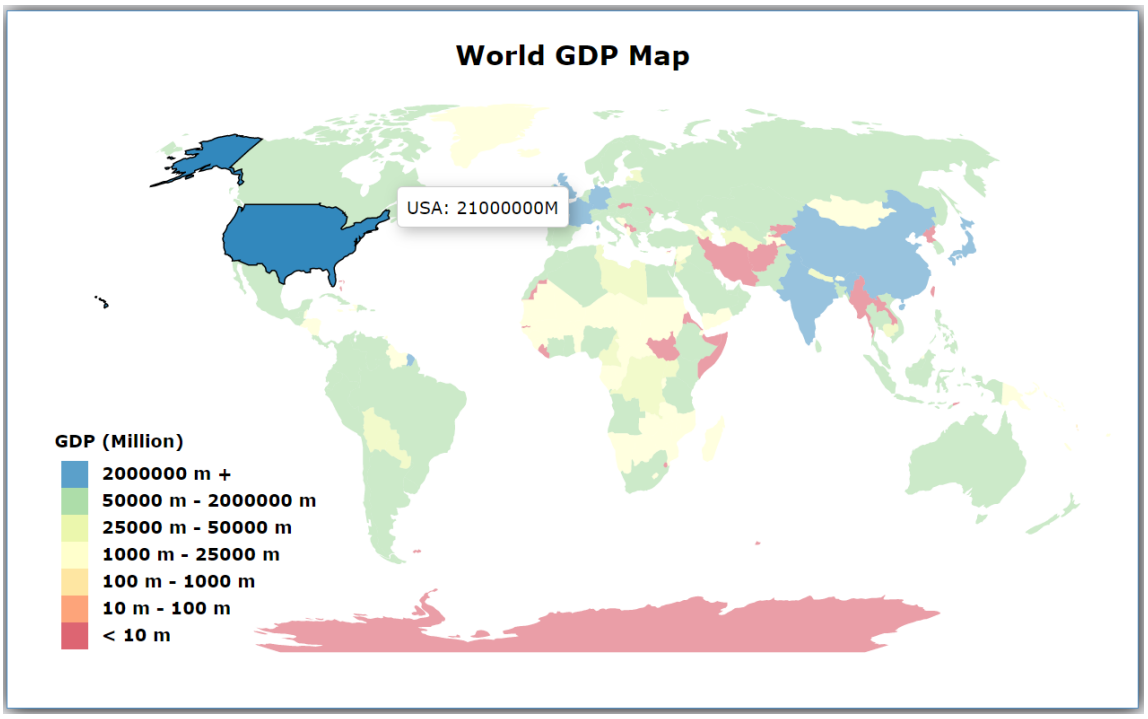
Results and interpretation:

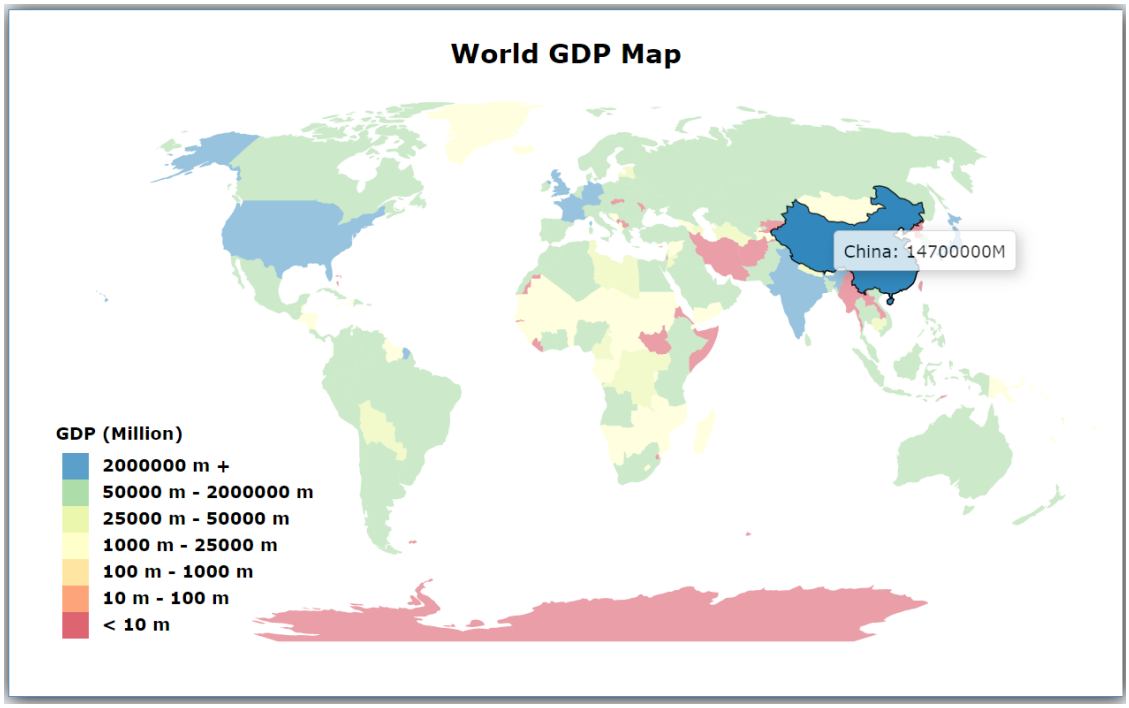
- "Gross Domestic Product (GDP) serves as a key metric for measuring a country's economic output. It is calculated by taking into account both expenditure and trade to determine the overall value of goods and services produced within a country during a specific period, typically a year. Both industrialized and developing countries contribute substantially to the global GDP, which stands at nearly 75 trillion dollars.
- Our choropleth graph provides a visualization of countries according to their GDP in 2020. For instance, Turkey ranked 20th with a projected GDP of 719.92 billion dollars.
- The United States, representing nearly 18 percent of the global GDP, is currently the largest contributor.



The global gross domestic product (GDP) is However, China, one of the BRIC countries (Brazil, Russia, India, and China), is not far behind. These emerging economies are gaining significant momentum on the global economic stage.

Interestingly, projections indicate that China is on track to surpass the United States by 2030, thus becoming the world's largest economy. This shift underlines the dynamic nature of global economic power and the rising influence of developing economies."


World GDP Map
USA: 21000000M
GDP (Million)
2000000 m +
50000 m - 2000000 m
25000 m - 50000 m
1000 m - 25000 m
100 m - 1000 m
10 m - 100 m
< 10 m


World GDP Map
China: 14700000M
GDP (Million)
2000000 m +
50000 m - 2000000 m
25000 m - 50000 m
1000 m - 25000 m
100 m - 1000 m
10 m - 100 m
< 10 m

Brushable Histogram

Technique description:

1. **Data preparation**:

The code generates random data for demonstration purposes. It creates an array of objects, each containing a 'country' (a random string of letters) and a 'value' (a random number). This data is then sorted by the 'value' property.

2. **Scale and Axis creation**:

The code sets up several D3 scales and axes to control the positioning and sizing of the bars in the chart. There are separate scales for the mini-chart ('mini') and main chart ('main'), and for the x and y dimensions. A color scale is also set up to color the bars according to their 'value' property.

3. **Chart and Brush creation**:

The code creates SVG elements for the mini and main charts, and appends bars to these charts based on the data. It also sets up a D3 brush on the mini-chart, which allows the user to select a subset of the data by clicking and dragging.

4. **Interaction Handling**:

The code defines several functions (`brushstart`, `brushmove`, `brushcenter`, and `scroll`) that update the charts based on user interactions. The `brushmove` function is particularly important, as it updates the main chart to reflect the data subset selected by the brush. The colors of the bars in the mini-chart are also updated to indicate which data is currently selected.

5. **Gradient Creation**:

The `createGradient` function creates a linear gradient that is used to color the bars in the mini-chart.

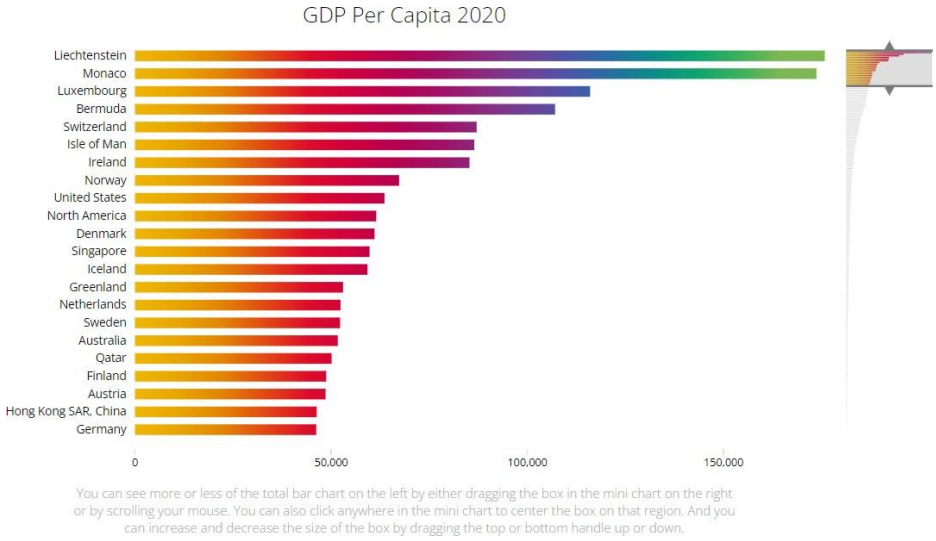
6. **Utility Functions**:

The `makeWord` function is a utility function that generates random strings of letters for the demo data.

Overall, this script demonstrates how to create rich, interactive visualizations with D3.js, and how to use brushing to enable detailed exploration of large data sets.

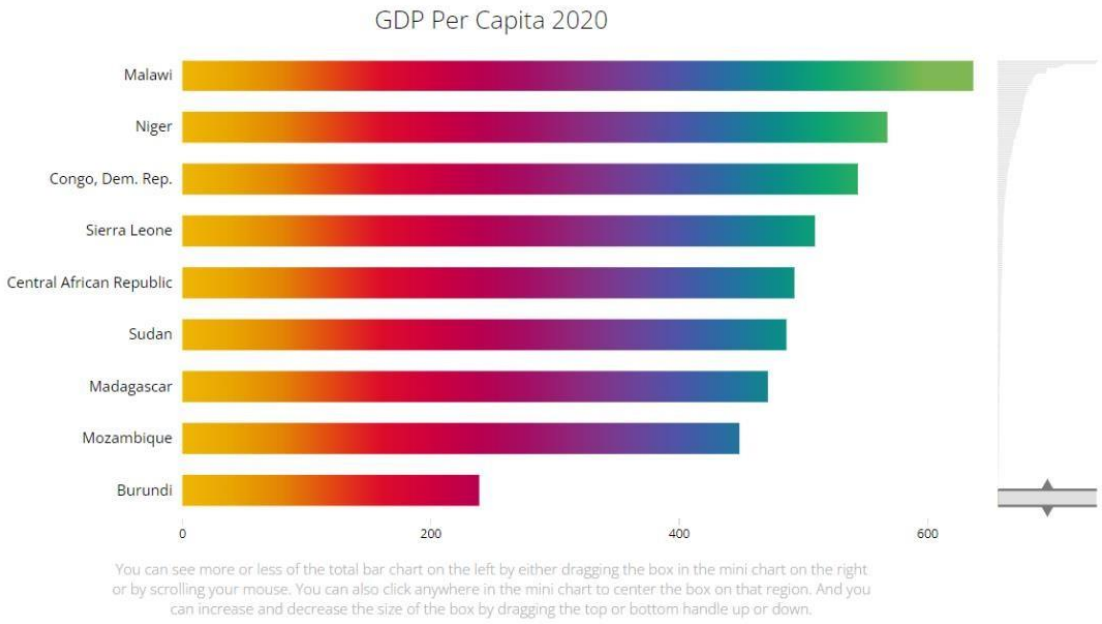
Results and interpretation:

- "GDP per capita, calculated as a country's total GDP divided by its population, serves as a critical indicator of the economic output attributable to each citizen. It's commonly used to gauge the standard of living, prosperity, and overall well-being within a country. A high GDP per capita typically suggests a high quality of life, whereas a low GDP per capita can imply a struggle to fulfill the basic needs of its citizens.
- Our brushable histogram illustrates the countries with the highest GDP per capita in 2020. Germany, for example, was positioned around the 20th rank, with an anticipated GDP per capita of 46,215.6 US dollars.
- Countries such as Luxembourg, Norway, Qatar, and Switzerland reported the highest per capita GDP globally in 2020. For comparison, the United States' GDP per capita was projected to be around 58 thousand dollars, approximately half of Luxembourg's.



- While a large GDP indicates a robust economy, the GDP per capita serves as a more accurate reflection of economic advancement. The United States, despite its vast GDP, had one of the highest unemployment rates among major developed and emerging economies in 2013, standing over 7.3 percent.
- Norway stands out as a country with one of the highest GDP per capita globally. Over the last decade, Norway's GDP per capita has surged by more than 43 thousand dollars, nearly doubling its earlier level, indicating a thriving and stable economy.

- On the other end of the spectrum, Burundi and Sudan reported the lowest GDP per capita in 2020. These countries, grappling with underdeveloped infrastructure and low living standards, face significant economic challenges.



- Sudan, with its 13 million inhabitants, has endured chronic poverty, high mortality rates, civil strife, and violence for years, leading to deplorable living conditions. Despite the country's reliance on agriculture and the majority of the population residing in rural areas, frequent famines and diseases exacerbated by conflict, food shortages, and inadequate water supply persist.

- Despite the country's small population, Sudan's GDP per capita doesn't even reach four figures, reflecting its struggle to produce and export goods efficiently. While NGOs and government agencies strive to mitigate their plight by offering funding and education, significant progress remains to be seen."

Exploding Radial Graph

Technique description:

1. **Setting up the SVG canvas:**

   The SVG (Scalable Vector Graphics) canvas is the area where the graph is plotted. This code sets up an SVG canvas with a specified width and height.

2. **Defining scales and layout:**

   D3's pie layout is used here to create the circular structure of the graph. The scale functions are used to map the data values to visual variables. In this case, a square root scale (`scaleSqrt`) is used to calculate the radius of each arc in the pie based on its data value.

3. **Loading and preparing data:**

   The data is loaded from a CSV file using the `d3.csv` function. The data is then prepared and transformed into a suitable format.

4. **Drawing the graph:**

   The pie layout is applied to the data to create an array of arc descriptors. These descriptors are then used to create SVG `path` elements for each slice of the pie. The `d3.arc` function is used to generate the `d` attribute of the `path` elements, which describes the shape of the arc. When these `path` elements are appended to the SVG canvas, they appear as a pie chart.

5. **Animating the graph:**

   A `step` function is defined to update the graph and advance the "time". This function is run at regular intervals when the "Play" button is clicked, creating the animation effect.

6. **Interaction:**

   The graph is interactive. The user can control the animation using the "Play/Pause" button and the slider. When the slider is moved, the `time` variable is updated and the graph is redrawn to reflect the data at the new time point.

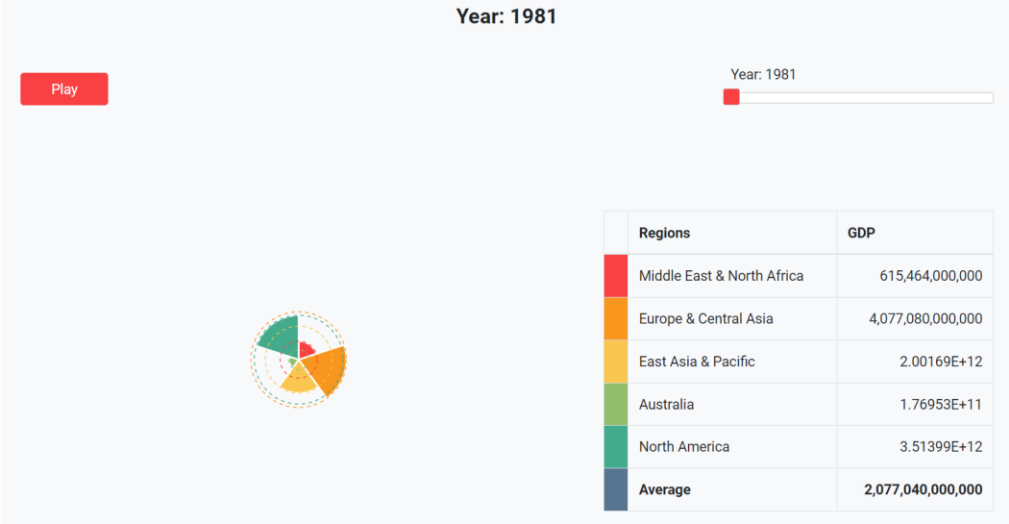
7. **Updating the visualization:**

   When the data is updated, the `update` function is called to redraw the graph. This function updates the radii of the arcs and the "average" line based on the new data.

The "exploding" effect is created by the way the data is visualized in the graph. Each data point is represented by an arc radiating from the center of the circle. The radius of the arc is determined by the data value, so larger values result in larger radii. This makes the data points appear to "explode" out from the center of the circle.

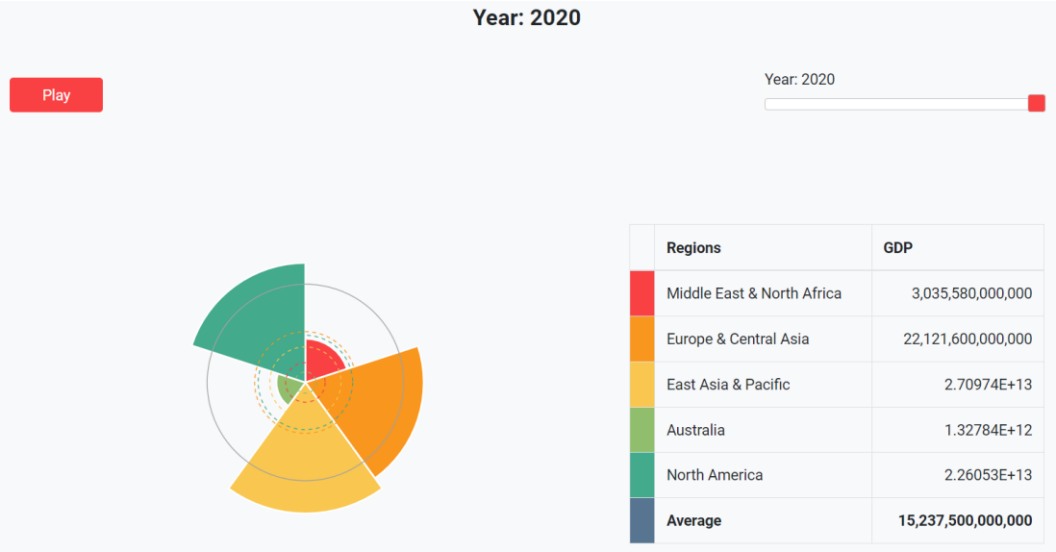
Results and Interpretation:

- "Initiating our exploration of regional GDP trends, we were intrigued to see how specific parts of the world compared to North America. We hypothesized that North America would rank highly, if not at the pinnacle, yet the data visualization unfolded a different narrative.



Regions	GDP
Middle East & North Africa	615,464,000,000
Europe & Central Asia	4,077,080,000,000
East Asia & Pacific	2.00169E+12
Australia	1.76953E+11
North America	3.51399E+12
Average	**2,077,040,000,000**

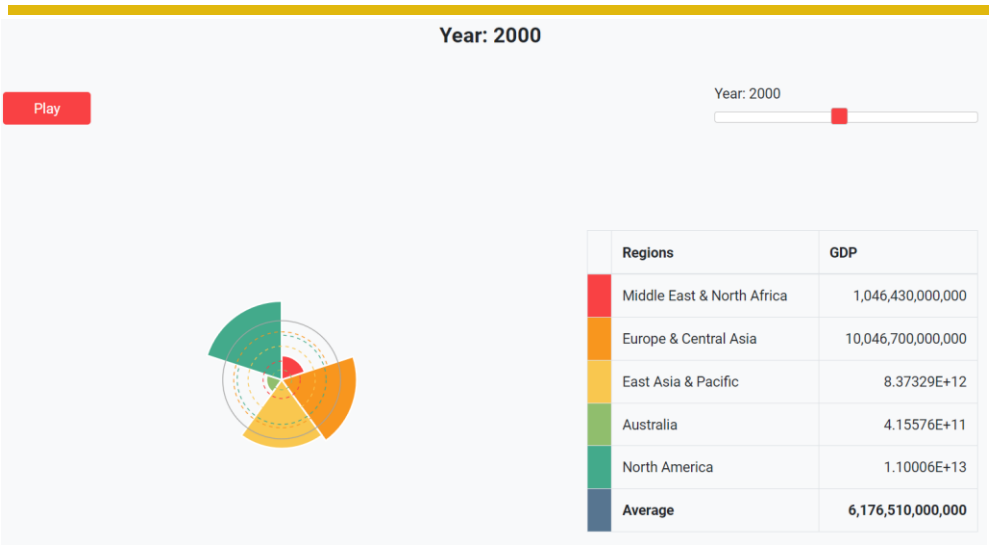
- Contrary to our expectations, regions such as Australia and East Asia (comprising China, Japan, and Southeast Asian countries) demonstrated similar GDP levels to North America. This unexpected parity sparked our curiosity, leading us to delve into the potential driving forces behind this economic trend.



Regions	GDP
Middle East & North Africa	3,035,580,000,000
Europe & Central Asia	22,121,600,000,000
East Asia & Pacific	2.70974E+13
Australia	1.32784E+12
North America	2.26053E+13
Average	**15,237,500,000,000**

As we animated the visualization over time, a key factor emerged: the influence of free trade. The late 1990s and early 2000s marked a significant upswing in free trade agreements, which extended far beyond the widely acknowledged China and Mexico to include countries like Australia.

Interestingly, despite the formation of the European Union, Europe and Central Asia were unable to make the same economic strides as Australia. This observation highlights the profound impact of free trade on global economic landscapes. By fostering income distribution and increasing interdependence among nations, free trade has effectively fueled global economic integration and cooperation."



Streamgraph

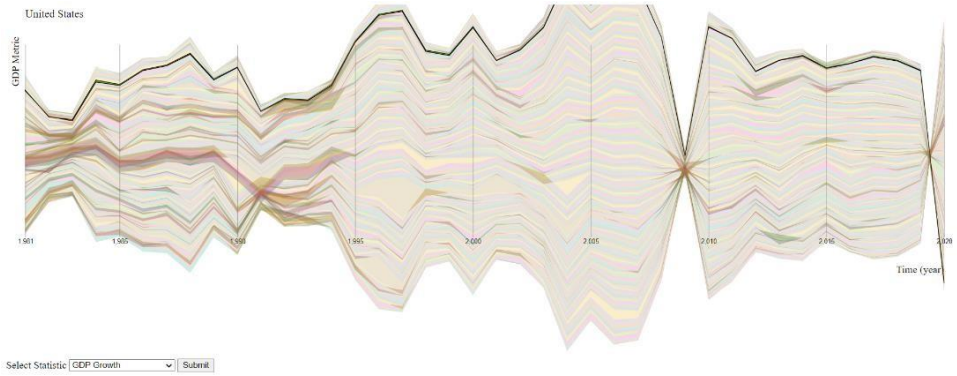
Technique description:

1. **Data Parsing**: The data is fetched from a CSV file with `d3.csv()`. The columns of the CSV file represent the different layers of the Streamgraph.
2. **SVG Creation**: An SVG is appended to a selected div in the webpage. This SVG will contain the Streamgraph.
3. **Axes Creation**: Both X and Y axes are created using `d3.scaleLinear()`. The X axis represents time and the Y axis represents the metric you're measuring (GDP growth or GDP per capita growth).
4. **Stacked Data Calculation**: The `d3.stack()` function is used to create a new stack layout. The stack layout is a way of creating a new data structure suitable for area charts. The `d3.stackOffsetSilhouette` offset is used to create a Streamgraph that is symmetrical along the horizontal axis.
5. **Area Definition**: The `d3.area()` function is used to create the shape of the areas that will be stacked to create the Streamgraph.
6. **Layer Creation**: The layers (or 'streams') of the Streamgraph are created by appending 'path' elements to the SVG for each data key (or 'group'). These paths are filled with color based on the data key they represent.
7. **Tooltip Creation**: Tooltips are created to enhance interactivity. They appear when a user hovers over a particular layer or stream, providing more specific information about the data that the stream represents.
8. **Interactivity**: Mouseover, mousemove, and mouseleave events are added to each layer. These change the appearance of the Streamgraph and the tooltip when the user interacts with the graph.
9. **Form Selection**: The form allows you to select different datasets to visualize with the Streamgraph. On clicking the 'Submit' button, the Streamgraph is updated with the selected dataset.

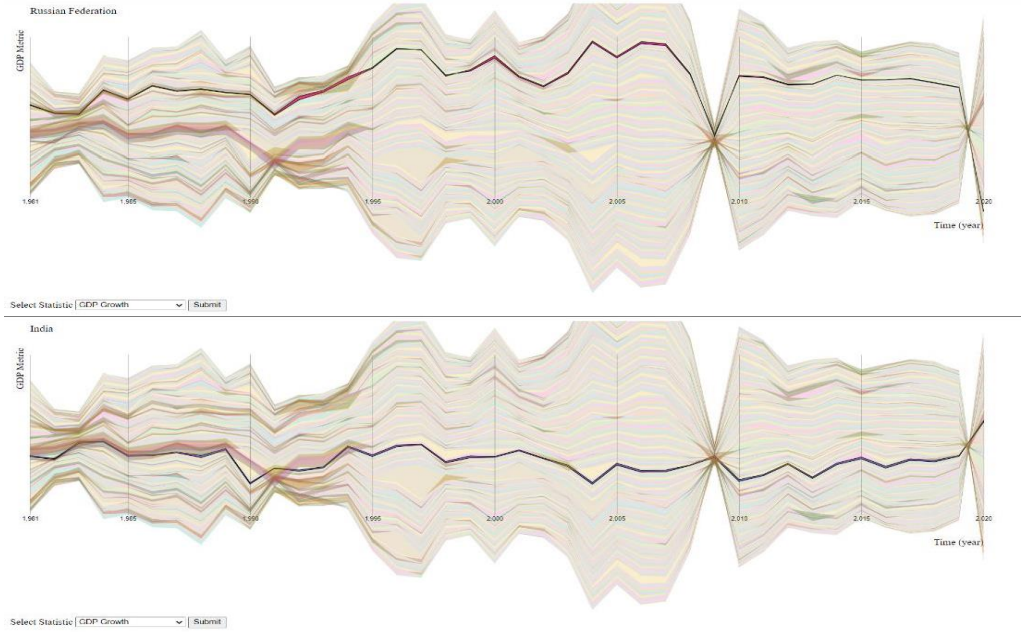
In summary, the technique used in this code not only visualizes the selected economic data but also offers a level of interaction, allowing users to focus on specific parts of the graph for more detailed information.

Results and interpretation:

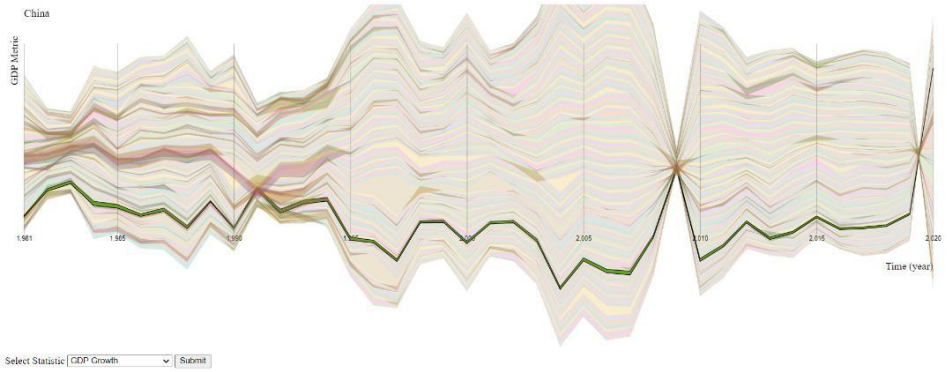
- At its core, annual GDP represents the monetary value of all goods and services produced within a country during a given year. Consequently, an upswing in GDP suggests an increase in the production of goods and services. However, the implications of GDP fluctuations extend beyond this simplistic interpretation.
- Looking back to 2020, the United States' real gross domestic product (GDP) saw an impressive upturn of 5.7% in 2021. This annual growth rate surpasses the average recorded over the preceding decade. The precipitous dip in GDP growth in 2020, induced by the coronavirus (COVID-19) pandemic, began to recede in 2021, signaling the onset of recovery.



- Despite boasting the highest GDP globally, the United States does not lead the pack in terms of real GDP growth rate among the world's major economies. This distinction belongs to China, which reported a real GDP growth rate of 1.5% in the final quarter of 2019, overshadowing the 0.53% growth rate observed in the United States. Both Russia and India showcased superior growth rates than the US, even though the US outpaced other major industrialized economies in real GDP growth.



- Among the world's seven largest economies, the United Kingdom bore the brunt of the economic impact of the coronavirus (COVID-19) pandemic. With its GDP growth rate plummeting to negative 9.6% in the third quarter of 2020 in comparison to the previous year, the UK's economy took a significant hit. Moreover, India's GDP also contracted by 7.5%. During this same period, China emerged as the sole major economy boasting a positive GDP growth rate of 4.9%.



Summary

"Gross Domestic Product (GDP) is a critical indicator of a country's economic prowess, offering insights into development patterns over specified periods. It serves as a vital tool for analyzing the impact of past policies on nations with diverse economic profiles, from powerhouses like China to developing countries such as Sudan. This analysis can guide future strategies to enhance living standards and stimulate economic growth.

It's important to acknowledge that a weak GDP can exacerbate social instability, potentially creating a breeding ground for extremism. This phenomenon underscores the imperative of sound economic management for ensuring social harmony and political stability.

However, it's equally crucial to recognize the limitations of GDP as a sole determinant of a nation's wellbeing or resilience. The COVID-19 pandemic exemplified this, as countries across the GDP spectrum, from the wealthiest to the poorest, found themselves equally vulnerable to the virus's devastating impacts. This scenario serves as a stark reminder that wealth does not necessarily confer immunity from global crises or guarantee overall societal health and wellbeing.

As we move forward, it's essential to continue refining our understanding and use of GDP as a measure, while also acknowledging the need for a more holistic set of indicators that can accurately reflect a nation's overall health, prosperity, and resilience."

Contribution:

1. **Swathi Vangala**: she was responsible for data collection, where [he/she] meticulously gathered GDP data from several reliable databases and international organizations. This required extensive research to ensure accuracy and comprehensiveness of the data used in the project. Swathi also took the lead in the development of the choropleth graph and stream graph. She not only handled the programming and design aspects but also optimized the visualizations for enhanced user experience. She contributed significantly to data interpretation for these graphs, drawing inferences and explaining trends.
2. **Prasanth Kumar Kunta** : He played a pivotal role in pre-processing the collected data. This involved cleaning the data, handling missing values, and formatting it appropriately for analysis. He also worked on creating the brushable histogram and exploding radial graph. His contribution went beyond programming and design, as he also focused on improving the user interface and ensuring the visualizations were intuitive and interactive. prasanth also led the data interpretation for these visualizations, providing comprehensive explanations of the observed patterns.

Both of us equally shared the responsibilities. We collaboratively worked on understanding the trends and patterns, drawing insightful conclusions about the GDP growth and per capita GDP of various countries. Our combined efforts have resulted in a comprehensive and detailed study of the GDP statistics across the globe.