Data Visualization

Visualization of International Debt

Group Member Names

18BCE0578 - Tanishk Aggarwal

18BCE0582 – Aditya Ruhatiya

18BCE0586 – Aditya Agrawal

Faculty Name

ANNAPURNA JONNALAGADDA
Associate Professor, SCOPE
VIT Vellore

Winter Semester

December 2019-June2020

Problem Statement

Visualization and Analysis of international debt of various countries using data from the International Monetary Fund. The data used is provided by International Monetary Fund. It contains debt statistics for several countries across the globe as recorded from 1950 to 2018.

Relevance of the Problem

This project tries to visualise an economic problem of the debt taken by different countries from the World Bank. In this project, we are going to visualize and analyse international debt data collected by The World Bank. The dataset contains information about the amount of debt (in % of their GDP) owed by developing countries across several categories. The data used in this project is provided by The World Bank. It contains both national and regional debt statistics for several countries across the globe as recorded from 1950 to 2018.

Outcomes

- To easily visualise the debts of different countries
- To find out country with highest and lowest debt
- To find whether geographical location affects the debts
- To see the change in debts of countries over a range of years

Detailed Literature Survey

World Bank Data Catalog-

Contains International Debt Statistics (IDS), successor to Global Development Finance and World Debt Tables which is used to visualize the international debt. The World Bank's Debtor Reporting System (DRS), from which the aggregate and country tables presented in this report are drawn, was established in 1951. World Debt Tables, the first publication that included DRS external debt data, appeared in 1973 and gained increased attention during the debt crisis of the 1980s. Since then,

the publication and data have undergone numerous revisions and iterations to address the challenges and demands posed by the global economic conditions.

Visualization of Spatial and Temporal Data with Dynamic Maps by Castronovo D-
Shows how to represent public health data over space and time through dynamic maps using Geographic Information Systems (GIS) and visualization software. Animation has potential to show the spread of infectious disease in relation to climatic factors at various spatial and temporal scales. However, dynamic maps represent a delicate balance between usefulness and incomprehensible information overload. This paper develops a conceptual framework for building informative dynamic maps using data on waterborne diseases and extreme weather events.

Geoplotlib: a Python Toolbox for Visualizing Geographical Data by Andrea Cuttone1, Sune Lehmann, Jakob Eg Larsen-
The paper introduces geoplotlib, an open-source python toolbox for visualizing geographical data. Geoplotlib supports the development of hardware-accelerated interactive visualizations in pure python, and provides implementations of dot maps, kernel density estimation, spatial graphs, Voronoi tessellation, shapefiles and many more common spatial visualizations. It describes geoplotlib design, functionalities and use cases.

Intelligent Geo Visualizations for Open Government Data (Vision Paper) by Auriol Degbelo, Christian Kray-
This paper talks about visualisation of Open government datasets (OGD). However, one key problem is the lack of flexibility of these visualizations, which severely limits their reuse in new scenarios. This article therefore proposes to increase the intelligence of existing geo visualisations by incorporating five features,to make better use of OGD: (i)automatic geographic data type recognition, (ii) generation of geovisualization designs, (iii) monitoring of users' understanding of geographic facts, (iv) self-optimization, and (v) user activity recognition. In addition to benefiting users of OGD, realizing these features presents rich scientific challenges and opportunities for Geovisualization
research, the OGD landscape (and beyond).

Contribution from each member of the team

The first step in the making of this project involves finding the dataset. The dataset includes finding the individual data set for each country of the world.

- **Tanishk Aggarwal**: He did research for finding the topic and the datasets related to this project. He studied various sources such as wikipedia and IMF Global debt website:
  https://en.wikipedia.org/wiki/List_of_countries_by_external_debt
  https://www.imf.org/external/datamapper/CG_DEBT_GDP@GDD/SWE
  He also helped in the paper work such as documentation for reviews etc.

- **Aditya Ruhatiya:** He found out all the python libraries that would be required for effective preprocessing of the data**.** He wrote the code that helped to understand the database in a better way and code to transform the database so as to make it perfect for visualization.
  Some python libraries required for above operations are:
  1. Pandas
  2. Numpy
  3. Matplotlib.pyplot

  He had the job of combining the dataset and the python modules to produce a basic framework on which a geospatial map can be built effectively.

- **Aditya Agrawal**: He found out all the python libraries that would be required for the proper implementation of the geospatial map. He wrote the code that would enable the actual visual representation of the debts of each country in the form of a geospatial world map.
  Some of python libraries required for above operations are:
  1. folium
  2. Seaborn
  3. Geopandas
  4. Bokeh

  He worked on integrating shape file, json file and the dataset to create a geospatial map of debts.

Schedule of the project

Title	Date of Completion
Topic and Domain selection	03/01/2020
Review 1	09/01/2020
Review 2	25/03/2020
Review 3	05/06/2020

Technology used in the project

We have taken the dataset in the form of a CSV file. This file will be processed using the 'panda' library of python. The dataset is transformed so that appropriate details of the things needed to plot a geospatial map is made available using the panda dataframe. We are also using the 'seaborn' library to show the change in central government debt varying from 1950 to 2018. The 'seaborn' library is also used to fit a regression model to the change in central government debt. This data is shown for some of the selected countries as it helps us to notice the pattern of debt taken by different countries over a range of years.

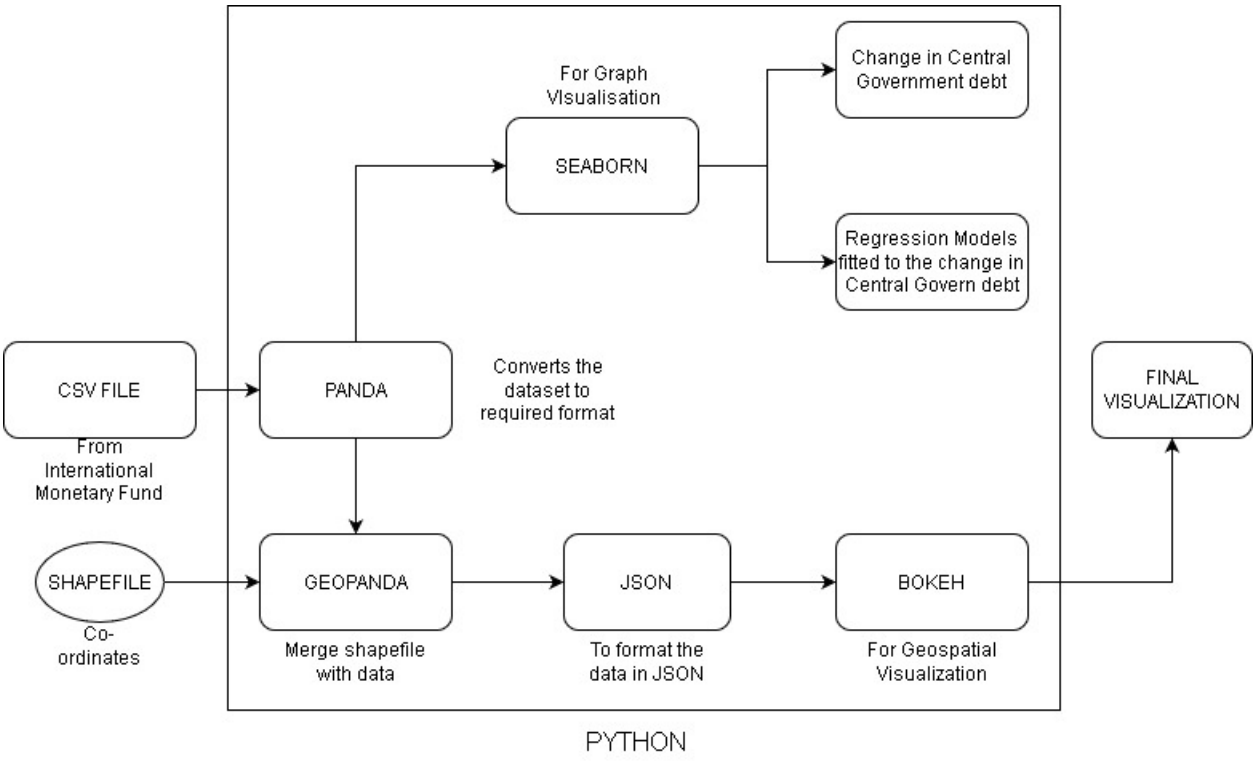
Now, after this we use a 'shapefile' containing the location coordinates of various countries of the world. We use the 'geopanda' library of python to integrate this shapefile with our processed dataframe. This merged data is converted to json format. This json file is used to plot our dataset on a geospatial map. We further add details like colour pallets, labels etc to enhance the visualization and make the map easier to understand.

How is this technology transforming

There are many powerful visualization tools that are coming up which can easily visualize this project. These tools eliminate the need of writing such long codes for visualization.

But the problem with these tools is that they are not free to use and are very complex for a beginner. Our code is a way in which we can visualize the data without spending the heavy price for these software.

Prototype of our project



Implementation aspects

1. Take CSV file from International Monetary Fund as data input.
2. Process this data file using ‘panda’ library of python. This gives us a processed dataset.
3. We can use ‘seaborn’ library of python to plot the change in debt. We will also fit regression model to this change in debt of some selected countries.
4. Now, to plot a geospatial map, we use the dataset processed by panda and merge it with a shapefile using geopanda library.
5. This merged file is then converted to GeoJSON format.

6. Finally, using this GeoJSON file as input, we plot a geospatial map using bokeh library.

Challenges faced in the implementation

- One of the major challenges we faced during this project was to effective pre-process the dataset. Because all further visualization depends on the processed dataset.
- We faced a challenge in choosing the best possible python libraries that would be required for the proper implementation.
- The next major challenge was to write a proper and the most efficient python code that would meet our needs.
- Integrating the json file was also a major challenge that we faced which allowed the proper visualization of the countries without any ambiguity.

Percentage of the work completed so far

- The dataset containing the international debts of all the countries has been found out.
- The data has been pre-processed so that effective visualization can be carried out.
- The proper visualization technique that we would use which would ensure visualization without any ambiguity has also been figured out.
- The python libraries that would be required have been figured out and majority of the python code has been written.
- Basic visualization of debts on a geospatial map has been obtained. Proper colouring has been provided and shown using legend.
- Added dashboard to switch years and see visualization of various years.

Verification / testing aspects

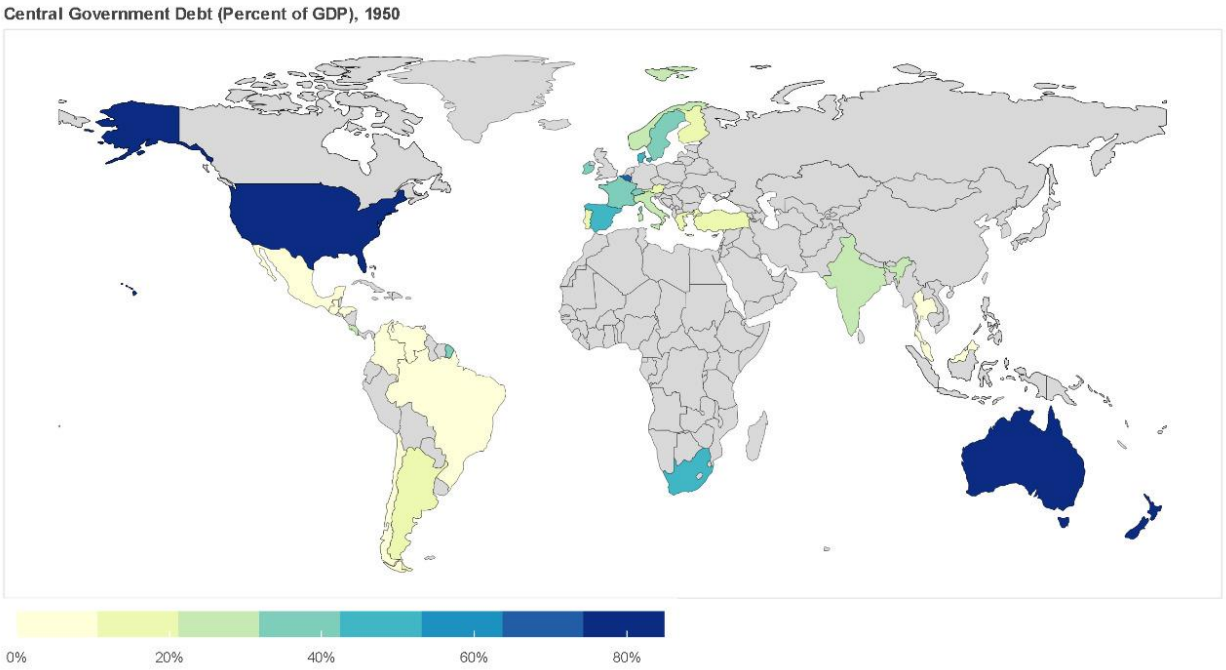
For testing we will take 5 to 6 countries randomly and compare it with the data set to check whether the visualized data matches the provided data. This could be done for any random 5 to 6 years and thus we can verify our visualization system.

Results obtained & Validation

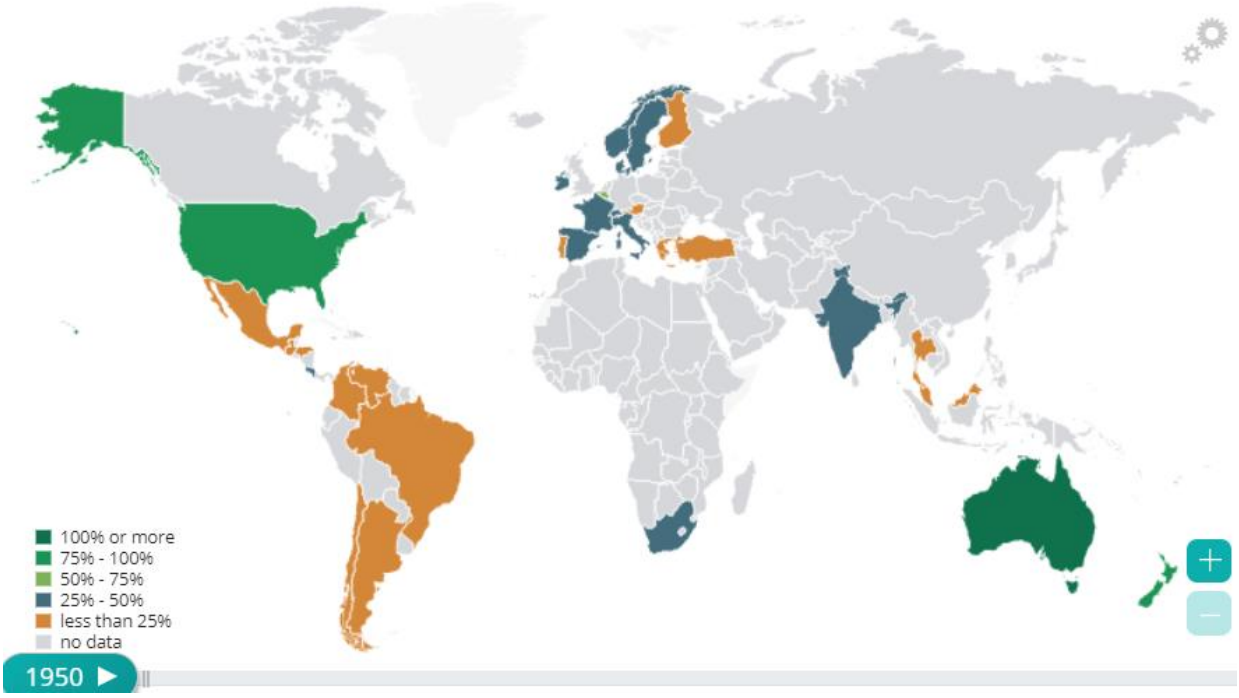
We are comparing our results with the actual map provided by the IMF.
This helps in Validation of our results.

For 1950:

Our map:

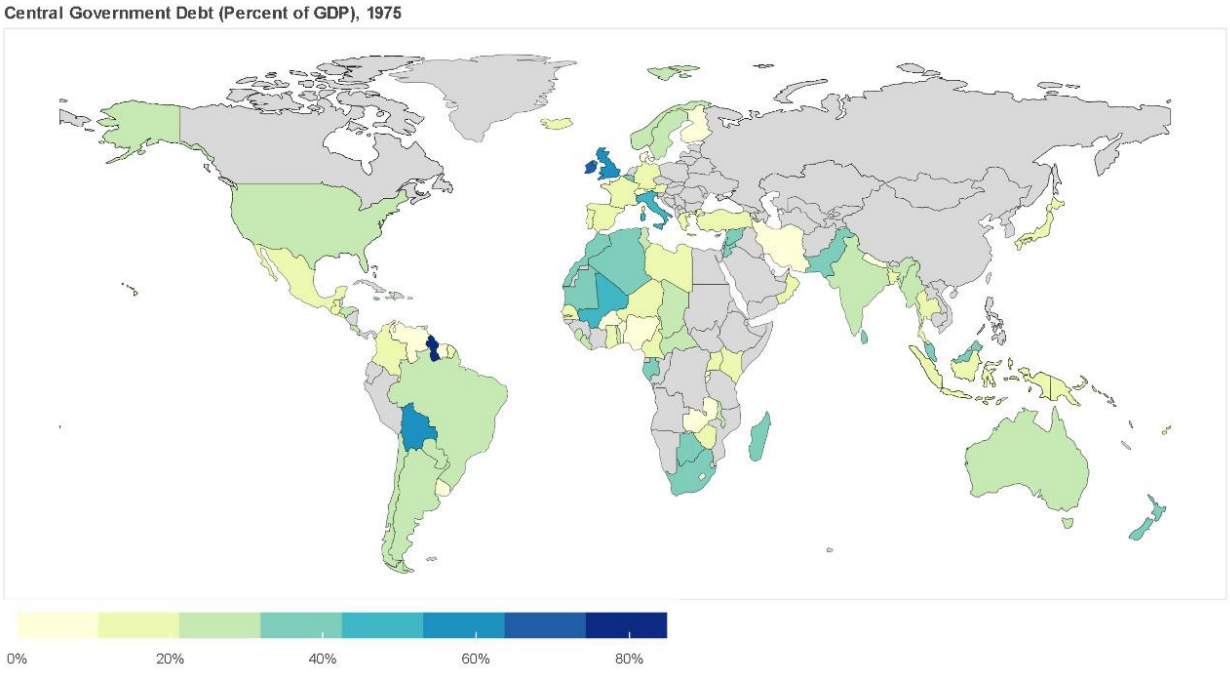


Map provided by IMF:

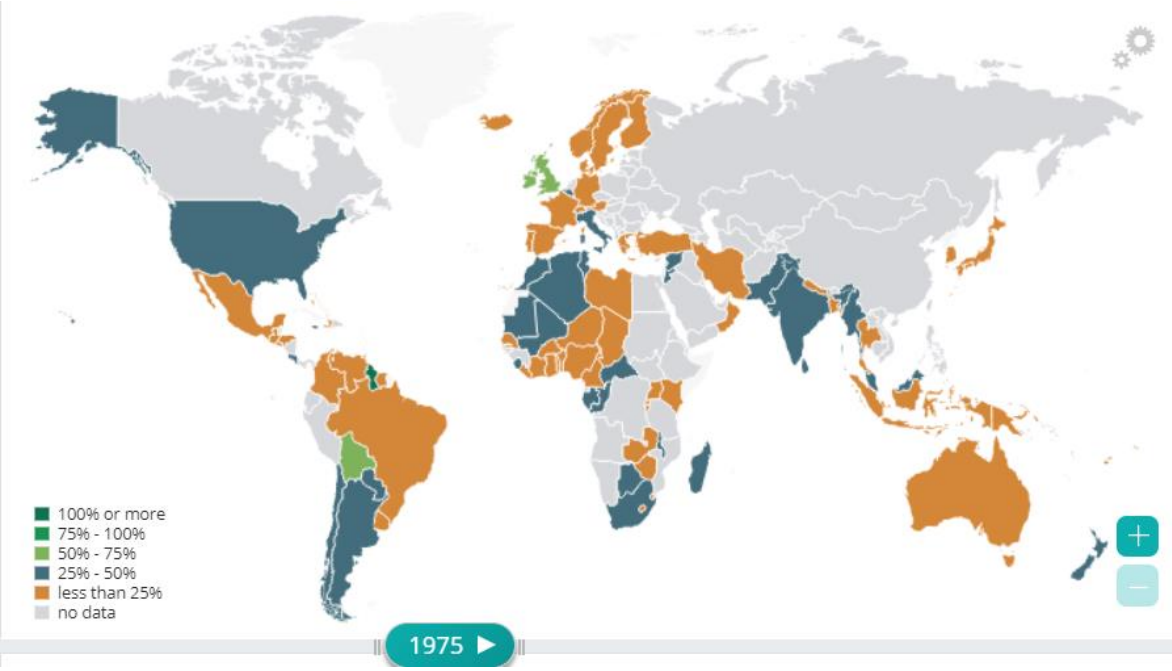


For 1975:

Our map:

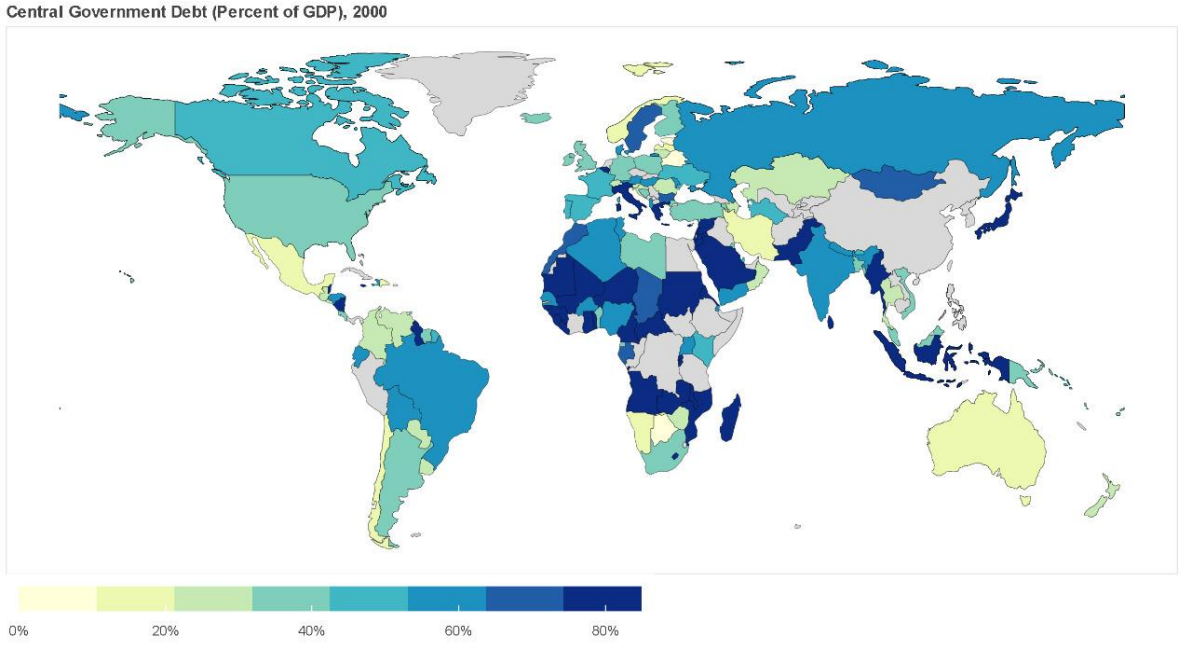


Map provided by IMF:

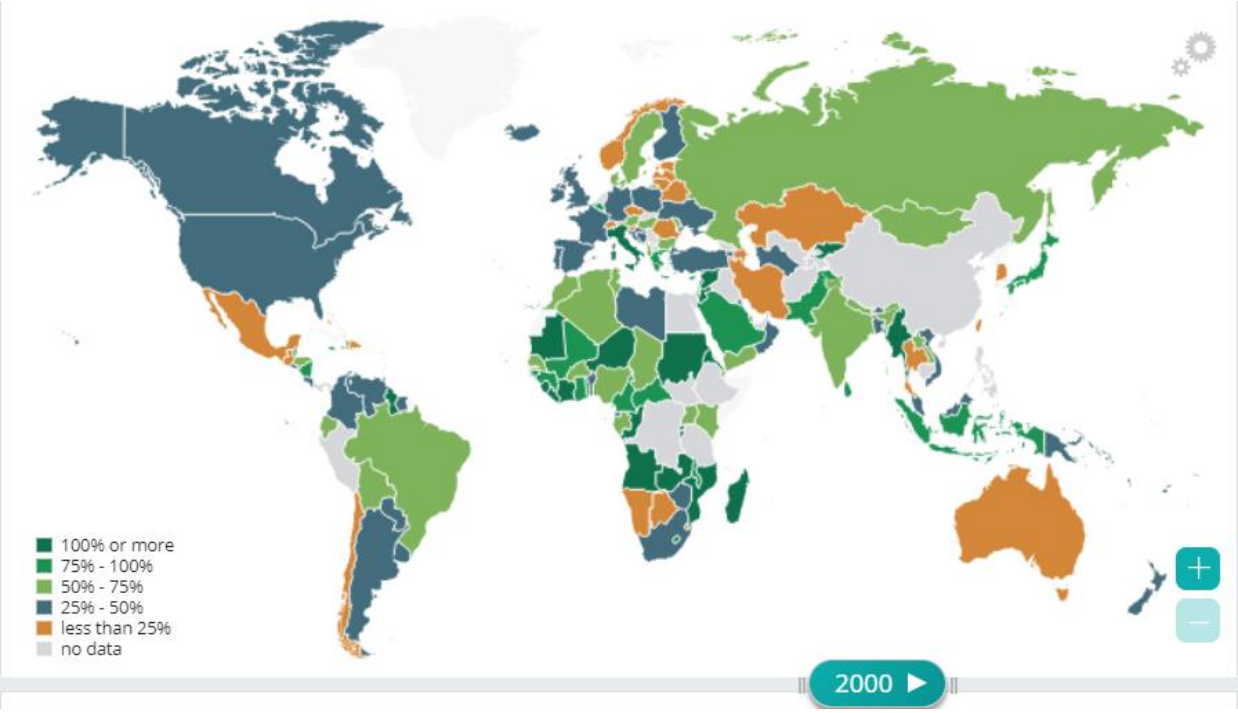


For 2000:

Our map:

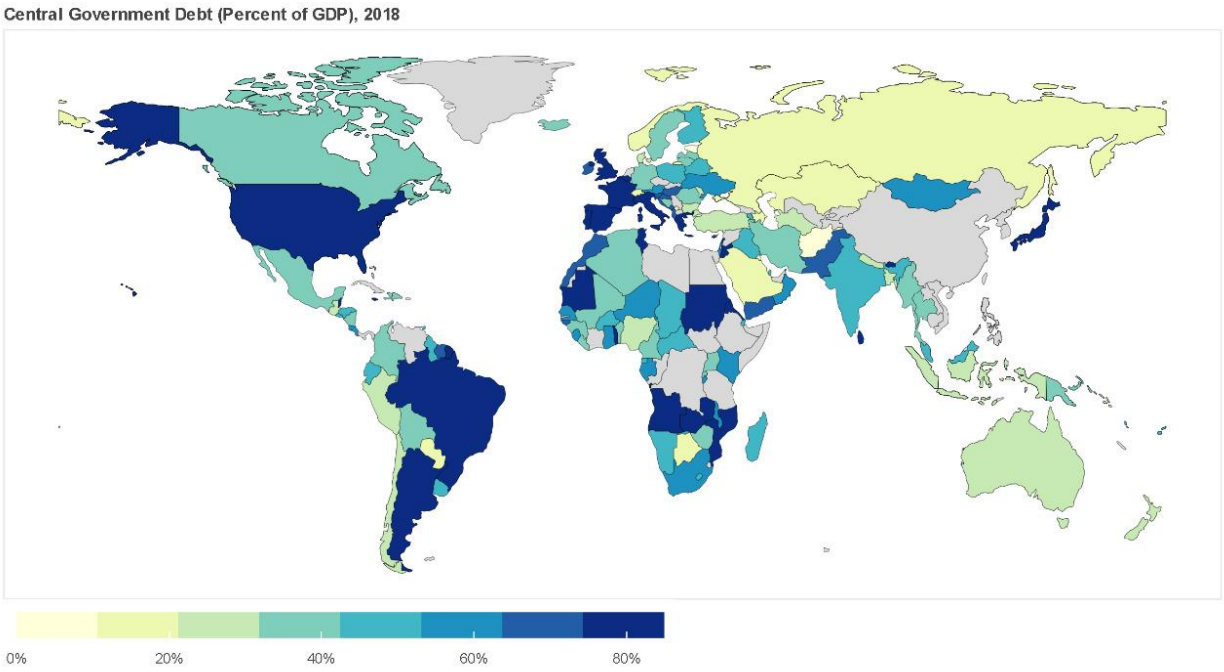


Map provided by IMF:

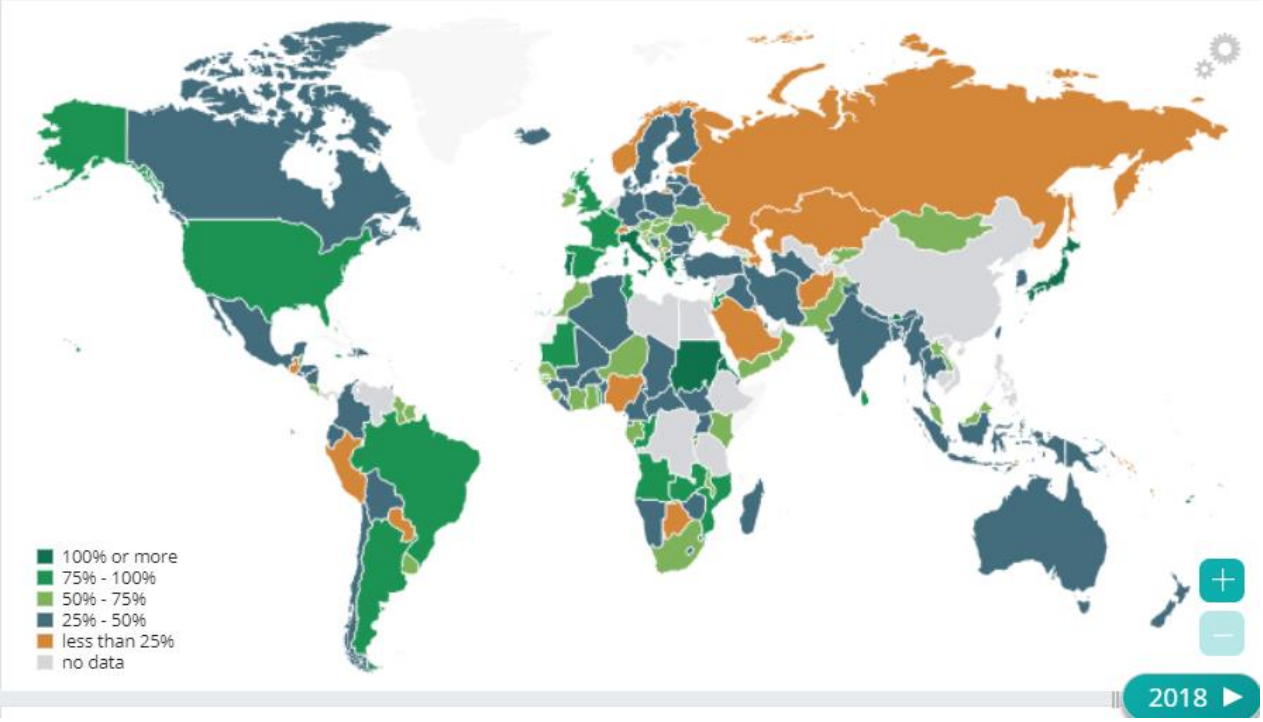


For 2018:

Our map:



Map provided by IMF:



Comparisons with existing Technologies

In terms of cost:

There are many powerful visualization tools that are there which can easily visualize this project. But the problem with these tools is that they are not free to use. Our code is a way in which we can visualize the data without spending the heavy price for these software.

Example: Visual Capitalist is website which is offering similar visualization in form for pie charts and they are charging $9.99 for this same visualization.
Our code is free of cost solution to similar visualization

In terms of Feasibility:

High ended software are much complex to use. Our project can also be used by a person having very less knowledge in this field.

Example: McKinsey is a website which uses Tableau for this similar visualization but it is very complex to use for a beginner.

Conclusion

We have successfully visualized the International debt of various countries across different years. Using this we can easily find out the country with highest and lowest debt for a particular year. This method is cheap as compared to other methods available in the market.

Also, this model can be used to visualize any dataset that contain data for every country. For example: in current times, we can use this model to visualize the cases of COVID-19 across different countries of the world.

References

1. https://journals.lww.com/epidem/fulltext/2008/11001/Visualization_of_Spatial_and_Temporal_Data_with.141.aspx
2. https://www.researchgate.net/publication/305983877_Geoplotlib_a_Python_Toolbox_for_Visualizing_Geographical_Data
3. https://www.academia.edu/37993790/Intelligent_Geovisualizations_for_Open_Government_Data_Vision_Paper_
4. https://datacatalog.worldbank.org/dataset/international-debt-statistics
5. https://www.mckinsey.com/business-functions/strategy-and-corporate-finance/our-insights/visualizing-global-debt
6. https://www.visualcapitalist.com/69-trillion-of-world-debt-in-one-infographic/