Global Trade Spectrum : WTO Merchandise Trade Visualization using Bokeh*

Deepak Khirey†
Indiana University, Bloomington

Gagan Arora†
Indiana University, Bloomington

Gautam Matkar†
Indiana University, Bloomington
(Dated: July 21, 2019)

Merchandise Trade in the modern world is typically driven by the difference in technology, resources, demand, Existence of Government policies and the existence of economies of scale in production. Merchandise trade [1] is characterized by merchandise exports and imports. Typically, there are two systems to record merchandise trade referred to as general trade and special trade. Special trade excludes certain trade flows. Merchandise trade is typically defined by general trade which covers inward and outward flows of goods through country or territory. Goods are merchandise which adds or reduces the stock of materials because of inward or outward flow through country or territory. The flow of good is valued at transaction values [2]. The trade values can be a good indicator of overall trade in International market, the trend for trade flows among developed and developing countries, growth gap between various economies, trade shares of different economies[3].

We are proposing here to build a visualization that can illustrate merchandise trade transactions over the years as per data collected by WTO. We describe our visualization as Global Trade Spectrum as it will provide 360 views of all transactions in different categories of trades.The data is visualized by 3 different charts-Line chart,GeoMap and Heatmap covering aspects of trending,display of worldmap and identify data patterns providing understanding of evolution of Global trade and factors influencing it since 1947.Bokeh was used to visualize the charts as it provides elegant and concise mechanism to plot highly complex chart and figure.

I. INTRODUCTION

In the last few decades, the global economy has opened and trade has increased with a rapid rate. As per report Latest trends in world trade 2017-2018 Chapter III [4], World merchandise exports increased to US$ 17.73 trillion in 2017, up from US$ 16.03 US trillion in 2016, World commercial services exports grew by 8% in 2017, reaching US$ 5.28 trillion, Asia was the top contributor to trade growth in volume terms in 2017, growing by 8%, World merchandise trade grew by 4.7% in 2017 in volume terms, driven by a rising demand for imports across the world. These numbers are massive but at the same time do highlight that these trends will change at a faster rate than any time before. These trends can provide great insight about the global economy. This data is quite intuitive for statisticians and economist but not intuitive for another genre. This information can be visualized in an intuitive way at a single place rather than residing in multiple manuals or papers. Another aspect is the capability of localizing or analyzing local trends rather than just global trends.These trend scan be analyzed based on time series and sliced diced by products, type of merchandise goods, export and import. In this study, we aim to create visualizations which bring together trends at a single place using Bokeh.

II. RESEARCH QUESTION

In this study, we will analyze WTO merchandise trade values annual dataset and visualize the data indicating the trade pattern and present indicators of Political policies, Trade expansion, Regional cooperation and trade balance via visualization.We are hoping to see evolving trends of Globalization era and understand pattern related to merchandise trade.There were three question which this research paper tries to answer and analyze:
1-How Export and Import trend over period of time?This will help to understand Export and import relation and factors affecting it,focus being individual WTO members.
2-How does Global trade evolve over World map?This can be great indicator of partnerships and evolution of trade over period of time.
3-How can granular information be obtained for different WTO member country?

* Bokeh is Python package
† MS Data Science, Summer 2019

III. RELATED WORK

Official website of the World Bank [5] has shown non-hierarchical treemap based visualization. We feel it does a decent job in representing the data. However, it doesnt clearly depict the relationship between various attributes of the data. I also lack the geolocation artifacts. Treemaps are good in representing the portion to the whole ration but might not capture the explicit comparison between data attributes. In this project, we hope to solve these issues. One way to look at this data is to represent in terms of trade balance which gives a fair idea of the gap between import and export for each country. While this is excellent graphics provided by WTO itself, it fails to recognize regional cooperation. It also does not show categorically as to how the center of world trade has shifted from the USA to China over the years. Here are examples of various visualizations provide by WTO which represents the data: [6] [7] [8] One of the inspiring visualization for our project is World Bank Global Development Sprint [10]. This visualization shows contracts awarded by all the countries of the world as per data collected by the World Bank. This is similar dataset to the current project. It talks about volumes of contracts over the years across various categories. It is built in D3.js framework and has an excellent representation of data. In addition to that, it has a nice animated way which keeps visualization engaging.

For this project we have used bokeh as a python library to process the trade transactions. Bokeh renders its charts and visualization using HTML and javascript thus Bokeh is perfect tool to build the interactive enriched visualization. Bokeh provides elegant and concise mechanism to plot highly complex chart and figure. Since we wanted to get the holistic view of trade transactions since 1947, bokeh is perfect candidate to have 360 degree of trade transactions. Official website of bokeh provides a very nice documnation of api and thus it was easy for us to adapt to bokeh.

IV. DATA AND METHOD

A. Dataset Summary

The data for this study is sourced from WTO — Trade Statistics - Bulk download of trade datasets with a specific focus on dataset WTO merchandise trade values annual dataset[9]. The data is in CSV format. The following table represents various columns and their sample distinct values to get the feel of the data. Highlighted columns are the target artifacts on which visualization would be focused. We have close to 959,601 data points.

B. Data Preparation

WTO has provided this dataset in CSV format with delimiter as comma. It does not have any NULL or NaN values. All fields are present in all rows.It is a very clean and neat data source from Data Quality perspective.However we need to make few changes in the dataset to make it suitable for our visualization.

Data Anomalies -

When we attempted to import this dataset into a dataframe, few rows were found to have more tokens than 14. This was happening because there was a comma present in the country name itself ie in Reporter Description and Partner Description fields. eg "Yemen, Republic of" So we had to replace such comma in country name with dash character. We did this activity manually with help of text editors. This helped us to resolve issue of tokenization.

Data Filters -

Once we imported dataset into dataframe, with further insights we realized that "Flow_Code","Unit","Flag","Source_Description","Note" fields do not have much data in it and we can safely drop those fields without affecting visualization outcome.

We also saw that few rows have trade value as zero. We decided to drop those rows as well because they will not show up on visualization. Both these filters helped us to reduce overall data size which we can manage for visualization.

Tidy Data -

The WTO dataset filtered in this way is already in 'Tidy Data' format. The Year and Trade Indicator columns are already melted and there is only one column which has actually dollar figures. This works very well for Visualization. Thanks to WTO Data source for making a useful copy of their data publicly available. We did not have to do any additional data operation to make it more efficient.

GeoJSON data -

Our next task was to map countries to actual geo spatial polygons. We referred "Natural Earth" site to get this information. (https://www.naturalearthdata.com/downloads/110m-cultural-vectors/). This site provides lots of ready to use map templates which we can plug in into our visualizations. we decided to use 1:110m resolution as it is best suitable for interactive maps. It also provides capability to zoom in which helps to have closer look at smaller countries in Europe or Island countries. This is a typical TOPOJSON file format. It stores all country boundaries in terms of vector polygons which makes it simpler to store in terms of size and usage. It has fields like country, country code and geometry. We noted that one of the geo-shape was for 'Antarctica'. Since we are dealing with Economic data here, we have no values for Antarctica, so we dropped this shape from our dataframe. We joined this geoshape dataframe with our WTO dataframe in order to get boundary polygons for

	Column Name	Distinct Count	Sample Values
0	Reporter_code	269	['AF' 'afr' 'AFR' 'ACP' 'AL']
1	Reporter_description	268	['Afghanistan' 'Africa' 'African, Caribbean and Pacific States (ACP)' 'Albania' 'Algeria']
2	Partner_code	52	['WL' 'OAF' 'CIS' 'NI4' 'IN']
3	Partner_description	49	['World' 'Other Africa' 'Commonwealth of Independent States (CIS), including associate and former member States' 'Four East Asian traders' 'India']
4	Indicator_code	31	['TO ' 'AGFO ' 'MA ' 'MAMTOF ' 'AG ']
5	Indicator_description	31	['Total merchandise' 'Food' 'Manufactures' 'Office and telecom equipment' 'Agricultural products']
6	Flow_Code	5	['X ' 'M ' 'RX' 'DX' 'RM']
7	Flow_Description	5	['Exports' 'Imports' 'Re-exports' 'Domestic exports' 'Retained imports']
8	Year	71	[1948 1949 1950 1951 1952]
9	Unit	1	['million USD']
10	Value	642296	[49. 58. 47. 52. 53.]
11	Flag	4	[' ' 'E ' 'B ' 'E']
12	Source_Description	1	['WTO']
13	Note	1	[' ']

FIG. 1. Data summary

countries representing economic data. We came across issues where country names were not matching. This is happening because naming conventions by WTO and that of natural Earth were different. eg. 'Chinese Taipei' in WTO dataframe is called 'Taiwan' in the map. So we created a mapping dictionary of such entries and replaced them in dataframe so that we get boundaries of those countries. Another issue was that Natural Earth does not give Boundaries for group of countries such as Euro zone or OPEC. It is a different process to gene rte polygons by adding geo-shapes, so we scoped it out due to time constraints. We decided to merge these dataframe as Left outer join between WTO dataset and GeoShape dataset based on Countries field. We replaced "No Data" value wherever the merged resulted into NULL value.

Data Types -

After importing the CSV data into Dataframe most of the columns were of Object Type and they were consuming large memory. This format is not very efficient for data operations behind the interactive visualization. So, we converted the data types of 'Reporter Description', 'Partner Description', 'Trade Indicator', 'Year' into 'Category' dtype. This proved to be very important as it reduced overall memory usage and our plot became fast and responsive.

C. Method for Visualization

1. *Line Chart*

Trending Line chart provides quickest way to get insights into trade performance in all WTO members and provides apt solution for multiple series of closely related series of data as it is light weight and provides opportunity for comparative study of trends and patterns. Line chart indicated that Global trade recorded its highest growth rate in six years in 2017, both in volume and value terms. But World trade faces strong headwinds in 2019 growing more slowly than expected in 2018 and possibility due to rising trade tensions and increased economic uncertainty. Line chart depicts this downward trend due to several factors including new tariffs and retaliatory measures affecting widely-traded goods, weaker global economic growth, volatility in financial markets and tighter monetary conditions in developed countries.For any economist it is of at most value if trend for export and import can be plotted on same chart against a time series providing opportunity to compare merchandise trade over certain duration and draw quick conclusion visually.

Plot Layout

'Export and Import trend' is one of tab on dashboard and it provides WTO country(which is sorted) as dropdown for interactive experience. The trending chart renders import and export dollar values from 1947 to 2018 based on selection for WTO country.Y-axis defines trade values in million dollars and X-axis defines financial years. Hover over tends to present dollar value with WTO country and year. Sometimes it is desirable to be able to hide glyphs by clicking on an entry in a Legend. In Bokeh this can be accomplished by setting the legend click policy property to "hide".Imports are represented by dashed blue line and Export is represented by green colored line.This facilitates to easily distinguish export and import line chart if printed.

Data processing on high level,implementation involves 3 main functions-

1- make or massaging data function-

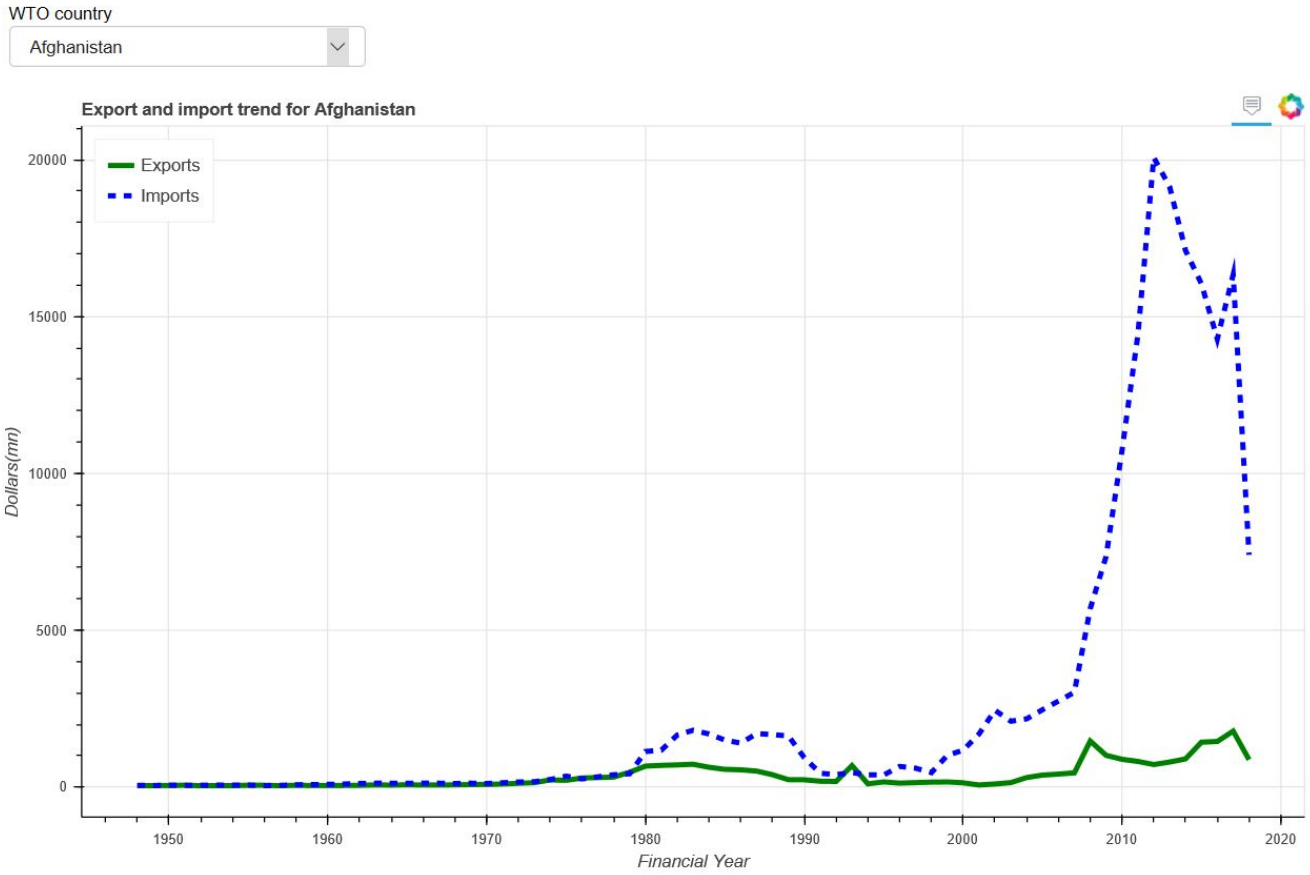

FIG. 2. Export and Import trend chart

In this process,separate dataframes are created for export and import based on Flow description attribute.In this dataframe and reporter is passed and it returns dataframe with data values based on user selection.
2-function to plot all the numeric data-
Initially plot is generated with default selection of WTO Country as 'India'. Using bokeh's plotting module line chart is plotted for export and import for different WTO members against time as an axis based on default selection.
3-update plot based on user selection-
Based on user selection,the line chart plot is updated to reflect new user selection.When user selects WTO country from dropdown to compare export import trending for a given WTO Country,here is a on.change event generated resulting in invoking and update function.This fuction captures user requested selection and returns dataframe based on new selection.This new dataframe is casted into ColumnDataSource format which is utilized by Bokeh interactive functionallity to render plot.

2. GeoMap

Why GeoMap?
We have used Choropleth map to visually represents the quantitative value of the trade across the globe. There are multiple advantages of using Choropleth map as it can provide the trade value across the countries in a single glance. For plotting we have used bokeh python library and used the enriched viridis color map to encode the colors on map. Other advantages include quickly understanding the relative difference in terms of trade value between countries. It also helps how neighbouring countries are doing trade wise in different commodities.

Plot layout
WTO GeoMap is the first tab on our dashboard. It depict the world map showing each countries with its boundaries. To visualize the trade data, we shaded the area with the viridis color map. Since trade value has great amount of divergence in its value, we have used the logarithmic color mapping to shade the countries. This chart has three interactive elements to it: Year [Slider], Commodity chooser [drop down] and trade flow[drop down]. In order to have interactive behaviour, we need to start the bokeh server. On the start of bokeh

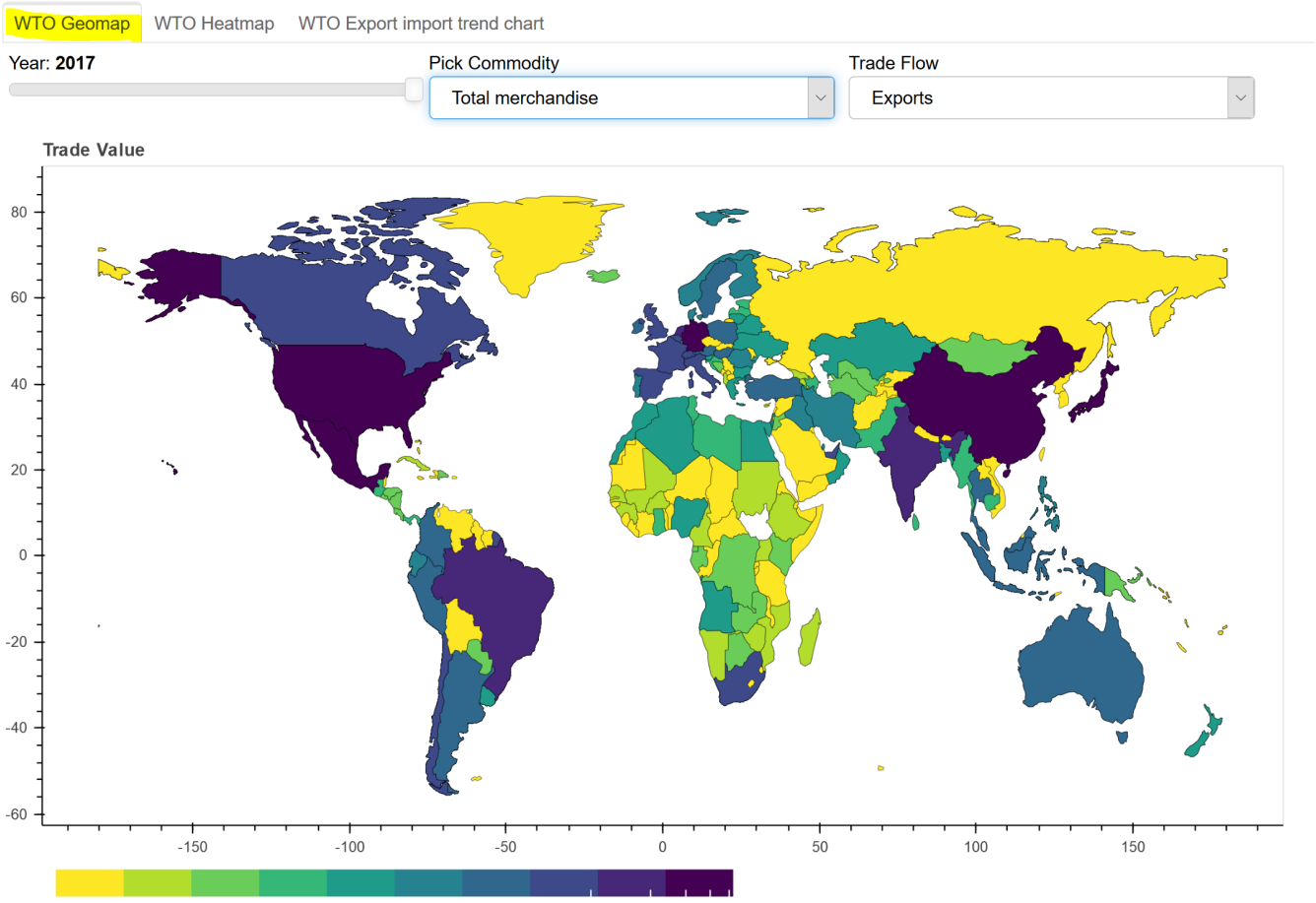

FIG. 3. Export and Import Geomap

server, we are showing the dashboard with three tabs on it. On the GeoMap tab - choropleth map update itself based on the user selected values: Year, commodity and trade flow. Interactive behaviour is achieved using the support for the bokeh call back function. On the user selection, subset of the dataframe is passed to the bokeh figure which manipulate the GeoMap dynamically.

Data Construction

As mentioned earlier, on the user selection, call back function selects the subset of the dataframe and pass it to the bokeh figure. In order to achieve this we have wrapped the functionality in separate function which listens to the user selection. One of the challenge which we have faced here is if user selects year and a particular commodity and if there is no export or import for particular country, bokeh figure was not plotting that particular country. So for rare commodity such as raw material, chemicals and etch many countries would vanish. In order to address this problem, we loaded the dataframe with all countries loaded and zero trade value for the selected commodity.

3. HeatMap

Why Heatmap?

Heatmaps are best suitable to represent large amount of data visually in single plot. It provides state of the data at a glance. For WTO Merchandise dataset, we planned to plot two heatmaps -one for export data and one for import dataset. This will show clear distinction in the patterns of transactions.

Moreover, the Geo-spatial plot could not reflect Economic groups on the map, mainly because their boundaries are complex and they are not readily available publicly. Also, since group membership changes over years, it is very difficult to represent them on an interactive plot. We were able to overcome this issue quickly by plotting heatmap as it does not require additional information like geospatial boundaries.

Plot Layout

WTO Heatmap is third tab on the dashboard and it provides 'Year' and 'Reporter Description' as two user input fields for interactive experience. User can choose year on a slider which ranges from 1947 to 2018. Reporter description goes by the same field name in the

FIG. 4. Export and Import Heatmap

dataset. It provides a drop down list of unique Reporter countries or economic groups. The list contains 419 entries and it is sorted in ascending order.

Two heatmaps are arranged side by side with same size of 450X450 pixels. It is most suitable for desktop of laptop screen. Every heatmap has a Title showing the year and reporter chosen by user. As user changes selection, the title updates itself to reflect the changes.

On X-axis of both plots, we have 'Merchandisee Trade Indicator Code' which shows economic field in which trade is taking place. eg Agriculture, Manufacturing etc. On the Y-axis, we have Reporter code which is the Trade partner of the transaction. It is essentially the list of countries where transaction is happening for a selected Reporter and year. For better readability and understanding, we have sorted categories on both axes alphabetically. We decided to use Trade code and Partner codes as axes ticks because they are short unlike actual descriptions which are too long.

We have adopted a consistent plot layout themed in 'viridis' color palette because it is suitable to show distinct color hues also it is color-blind friendly scheme. The color hue is set to logarithmic scale because the trade values are ranging from few thousand to to few billion. There is color bar under each heatmap which depicts the log scale and ticks representing the hue. To get detail of a particular transaction, we have chosen to display tootips. When user hovers mouse over heatmap cell, a tooltip appears which provides details of 'Trade Value', 'Trade Partner' and 'Trade Indicator'. This is a very effective way to visualize large information in a very compact manner.

Make_data

We have defined a make_data method for heatmap to manage underlying dataframe on which visualization is based. In this method, we pass year and reporter as argument and based on those we filter out the main source dataframe for particular selection. We also filter out 'Total merchandise' trade indicator because it is sum of all trade indicators, hence it is kind of repetition of information. Also, it reduces visibility of individual trade categories on the heatmap. Fortunately, since this dataset is already in tidy data format, we don't have to do any further data wrangling to generate visualization. Similarly, the heatmap shows the most granular information about this dataset, so we don't have to group dataset on any field for this plot.

Update

We have defined a dedicated update function which provides interactive capabilities to heatmap. When user selects year and reporter on the layout, there is a on_change event generated and in turn, it triggers update heatmap function. Job of update function is to receive changed user selection for year and/or reporter and pass it on to the make data function. It receives a

new dataframe based on new filter criteria from make data function which is then casted into ColumnDataSource format so that it can be fed to Bokeh interactive functionality. We also make sure that update function updates the Title of the plot based o user selection.

V. CONCLUSION

1. It is very tedious to create multiple plots for given dataset like WTO Merchandise Trade
2. It will be very inefficient way to create separate plots for each country or each trade category
3. Hence it is very useful to create interactive plots where visualization updates based on user selection
4. Bokeh provides very robust Interactive Dashboard capabilities to explore data visually
5. Different visualization methods like Line Chart, Geomap, Heatmap are useful to convey different properties of dataset like Temporal Trends, Regional co-operation etc.

VI. FUTURE WORK

In this project we made a bokeh Dashboard with three tabs on it: GeoMap, Heat Map and line chart. All three tabs are interactive with selectors on it. One limitation of this dashboard is, interactive behaviour is only possible when we start bokeh server. As the part of future work, we would deploy this dashboard as web server. For this we would need Flask - micro web framework for python which translate code to the hostable python code. Then python web application can be deployed to python supporting web servers such as pythonanywhere or Heroku. Other possible future work would be to create video or animation of a minute or two with trade values varying from 1947 to 2017 which will help to visualize the trade patterns over year.

We also observed that bokeh interactive behaviour is processing hundred thousand rows to plot the chart and is sometimes slow - this can be speed up if we can use caching which we think can be classified as future work.

ACKNOWLEDGMENTS

We wish to acknowledge the support of Prof. YY and Nathan Ratkiewicz through out the course and study material uploaded on Canavas.

[1] W. T. Notes, Nicita, a. (2014). united nations conference on trade and development. united nations publication. geneva 10, switzerland: United nations conference on trade and development. (2019).
[2] W. T. Organizations, World trade organization . (2019). wto statistical data sets - metadata. (2019).
[3] W. T. Organizations, World trade organization. (2018). latest trends in world trade 2017-2018. world trade statistical review , 28-39. (2019).
[4] W. T. Organizations, Bulk download of trade datasets. retrieved from www.wto.org: (2019).
[5] W. T. Organizations, Statistics on merchandise trade. retrieved from www.wto.org: (2019).
[6] W. T. Organizations, Country analysis. retrieved from wits.worldbank.org: (2019).
[7] W. T. Organizations, Detailed product analysis. retrieved from wits.worldbank.org: (2019).
[8] W. T. Organizations, Merchandise trade balance with partners. retrieved from wits.worldbank.org: (2019).
[9] W. T. Organizations, World integrated trade solution. retrieved from wits.worldbank.org: (2019).
[10] W. Koehrsen, Data visualization with bokeh in python, part i: Getting started (2018).
[11] W. Koehrsen, Data visualization with bokeh in python, part ii: Interactions (2018).
[12] W. Koehrsen, Data visualization with bokeh in python, part iii: Making a complete dashboard (2018).
[13] N. Earth, 1:110m cultural vectors (2019).