**Introduction**

World trade can be affected by productivity, supply and demand, diplomatic relations, as well as global incidents such as a pandemic. We are intrigued by the international trade datasets from UN Comtrade because we would like to investigate how the composition of international trade changes for the US in terms of both import and export. The change could be in the share of imports and exports, could be in the countries that it cooperates with, and could be in the contribution of different commodity types. We consider understanding the changes in the structure of international trade between the US and the world to be crucial because it could provide insights into the US’s industrial structure as well as its economic relationship with other countries.

To further guide our research, we decided to focus on the following questions: How does the annual total trade value between the US and different countries change over the time period 2016-2020, for both import and export? What is the trend in trade value and quantity for different commodity categories between the US and the world, for both imports and exports? How does the ranking of commodity categories measured by trade value between the US and the world  change before and after COVID-19 outbreak, for both imports and exports? What is the impact of exchange rates on international trade between the US and other countries over the time period 2016-2020?

To address these problems, we build five different types of data visualizations, which includes a choropleth map, a treemap, a race bar chart, a slop chart, as well as a linked line and stacked bar chart to better dissect data and figure out the messages hidden behind. We expect to see the US to be strong trading partners with Canada, Mexico, EU countries, and China. Moreover, it is expected that the US would rely on imports heavily for commodity goods such as electronic equipment, mechanical equipment, automobiles, etc. Lastly, we reckon a decrease in total trade value is likely to occur for 2020, due to shutdowns caused by COVID-19 outbreak.

**Method**

The treemap is designed to showcase trade value and quantity rankings of commodity categories for US import and export from 2016-2020. The raw data used consists of 10 csv files with a total of 1,000,000 rows (100,000 rows per file), and 35 columns in each file. Below are steps for data preprocessing:

* Change column names to snake-case.
* Drop unwanted columns, leaving only *partner, year, commodity\_code, commodity, qty, and trade\_value\_\_us\_\_*.
* Aggregation: compute the sum of trade quantity for all root categories (C*ommodity\_Code* with a value between 1 and 99) based on the quantities of sub-categories (*Qty*).
* Filter desired rows: *Partner* = “world” and keep only root categories .
* Drop all rows with either 0 quantity or trade value.
* Convert the unit of trade value to US$ billion.
* Generate a new feature called *world*, with only one string type value “Commodity Categories”, in order to have a single root node in the treemap.
* Sort the dataframe by [“*commodity*”, “*world*”].
* Apply a custom function that allows text to wrap.

The resulting cleaned dataframe contains 474 rows and 7 columns. The primary unit of analysis is commodity, which contains the name of all commodity categories. The most important variables are trade\_value\_\_us\_\_ and qty, which contains the trade value and quantity for each root commodity category, respectively. qty is a derived variable. In the original data, there are thousands of sub-categories under 99 root categories, and only sub-categories have data on individual trade quantities. Since the treemap would be difficult to interpret if all sub-categories were plotted, only root categories are included, and their quantities are obtained by aggregating sub-category quantities.

Treemaps on import (fig.3) and export (fig.4) are plotted separately. We felt treemap is very effective in showing the ranking of commodity categories in terms of trade value and quantity at the same time. The color of each category reflects its total trade value, and the area of each category is proportional to its trade quantity. Since both trade value and quantity are numeric and continuous variables, area and color are intuitive ways to encode them. Furthermore, since most commodity category names are quite lengthy, labeling them would be a mess for other types of visualizations due to not enough space. This is not a problem for treemap since there is plenty of space in each rectangle to display the name of the category.

In terms of the design, a dropdown menu that allows the audience to select data of different years is featured. Since we have 5 years of data and one treemap can only show 1 year of data at a time, a dropdown menu is a necessary practice to avoid plotting a figure for each year of data.

Moreover, a tooltip is used to display details about each category when the cursor hovers on it. It contains additional information that includes trading year, trading value in US$ billion, and quantity. Instead of displaying all the information in category boxes directly, showing it in a hovering tooltip makes the visualization more concise.

In terms of visual encoding choices, sequential color encoding is used to show the difference between trade values. The darker the color is, the greater the trade value. We chose two different color themes for import and export treemaps: teal for import and orange for export. We reckon that the strong contrast between teal and orange, one being cool and one being warm, allows the audience to distinguish between import and export better.

My original goal was to showcase how trade value and quantity are distributed across different commodity categories using percentage values to represent share in the economy. However, during the making of the prototype, I felt it is better to use absolute values for both variables because it gives a better sense in terms of the scale of a commodity category. Apart from this, I sticked to my original plan.

**Results**

1. The category Nuclear Reactors, Boilers, Machinery and Mechanical Appliances has the greatest contribution in trade value throughout the years for both import and export. However, its trade quantity is not as high, as the rectangle for Nuclear Reactors is often quite small, which has a strong contrast against its dark color.
2. Electrical Machinery and Equipment and Mineral Fuels are always among top 3 in terms of trade value for export
3. The category Vehicles always has the third highest contribution in trade value for import. It has a much lower contribution to export.
4. Mineral Fuels, Inorganic Chemicals and Cereals, surprisingly, have the top 3 contributions in trade quantity over the years for export. However, Electrical Machinery and Equipment replaces Cereals in terms of trade quantity for import. Also,

**Discussion**

1. The absence of Vehicles in top categories for export and its high contribution to import suggests that the US has a huge market as well as demand for foreign cars
2. The interesting comparison between the dark color and the small size of Nuclear Reactors, Boilers, Machinery and Mechanical Appliances indicates that products in this category are extremely expensive.
3. The US has a great demand as well as supply of mineral fuels/oils and inorganic chemicals, which could suggest that the US has a robust chemical production industry.