Project Process Book

Who Trades With Whom
Exploring International Trade Relationships Over Time

Submitted in Partial Fulfillment
Of the Requirements Of
CS 6630 – Visualization for Data Science

Submitted By:

Scott Gale, u1203422
Sudie Roweton, u1210082

Table of Contents

Project Title: Who Trades With Whom - Exploring International Trade Relationships Over Time

Team Information

Scott Gale	u1203422	scottdgale@gmail.com
Sudie Roweton	u1210082	sudie.roweton@gmail.com

Project Repository Link: https://github.com/scottdgale/DataVisProject.git (public)
Project Website Link: https://scottdgale.github.io/DataVisProject/

Background and Motivation

We both have a connection to the Department of Defense, and we decided to make a group to consider a defense related topic. As we were looking for viable data sets, we came across the Correlates of War Project Trade Data Set. This data set sparked our interest, and we pivoted from defense to trade, specifically considering international trade relationships over time. Examining international trade data in the context of trade relationships over time is an interesting analysis and lends itself to rich visualization possibilities. In essence, the data shaped our project topic decision.

Another motivating factor for our choice is that international trade is a major topic in the contemporary political environment. The United States is currently in the process of renegotiating trade agreements from many allied countries. Additionally, the United States is imposing tariffs on imported goods and services from many long-time trade partners.

We would like to build a visualization to help people understand more about international trade supporting the user tasks of presentation, exploration, identification, and comparison. This will help to provide context and understanding on the volume and interactions of international trade between the countries of the world.

Project Objectives

Questions to be answered by our visualization
1. What countries trade with other countries?
2. For a given country – who are their top importers?
3. For a given country – who are their top exporters?

4. What is the trade deficit (or surplus) between two countries (in US dollars)?
5. How has trade between countries changed over time?
6. What effects do tariffs have on trade (nice to have implementation)?

Learning/Accomplishment Goals
1. Create intuitive visualization that accurately represent the data.
2. Design highly interactive visualization that allow users to explore the data set.
3. Design an effective visualization that takes advantage of visualization theory principles and guidelines learned during the course (especially expressiveness/effectiveness).
4. Create custom (never used in class) interactive data manipulation tools (both direct and indirect) that make navigation simple.

Data

Initial Entry - Proposal

We currently have one comprehensive data set. Below is the link and an overview as described from the source.

"The trade dataset is the result of the effort to code trade flows between states (as defined by the Correlates of War project) for the period 1870-2014. The data include information on both bilateral trade flows and total national imports and exports. The dyadic trade dataset describes import and export data in current U.S. dollars for pairs of sovereign states. The National (Monadic) Trade dataset contains information on individual states import and export levels in current U.S. dollars."
Link: http://www.correlatesofwar.org/data-sets/bilateral-trade

An additional possible source of data comes from https://wits.worldbank.org/. We can get AHS Weighted % and MFN Weighted % (Tariff Data) from a selected country to other countries/regions. This data only goes back to 1991 - so we would have a significant gap in tariff data from 1870 - 1990. This may, however, still be useful.

We will also use the GDP per capita data that we already have from homework 4 to explore trade balance vs GDP.

Separate from the trade specific data, we have acquired data to support our world map. From homework 4, we have a world data file (world.json) that we can utilize. Additionally, to link trade partners, we also have latitude/longitude data for the capital cities. Link:https://esa.un.org/unpd/wup/cd-rom/WUP2014_XLS_CD_FILES/WUP2014-F13-Capital_Cities.xls

We are in the process of identifying additional data sets that contain information on tariffs, trade agreements, and other geopolitical factors that we could possibly use in our project to provide greater clarity and understanding on international trade. Additionally, we are looking for data on trade over the last 3 years since the data set we are using is only current through 2014.

Update – Project Milestone

While the tariff data that we acquired was a 'nice to have', we have narrowed down our datasets to support our 'must have' features. The datasets that we have and are using to support our objectives are the following:

- Correlates of War data set – Dyadic and National from 1990 to 2014.
  Link: http://www.correlatesofwar.org/data-sets/bilateral-trade
- GDP dataset from Homework 4
- Capital Cities data set
  Link: https://esa.un.org/unpd/wup/cd-rom/WUP2014_XLS_CD_FILES/WUP2014-F13-Capital_Cities.xls
- World Map dataset from Homework 4 – world.json

This data can be accessed in our github repository in the "Data" folder.

Data Processing

Initial Entry - Proposal

Our datasets (correlates of war, GDP, tariff data, and lat/long data) will not require substantial data cleanup. The data appears to be clean and well organized. In addition, all three are in a format (.csv) with which we are familiar. However, we will still perform a data cleaning process on our datasets to make sure there will be seamless integration into our visualization.

We expect to use most of the data as is. However, the visualization will handle summing quantities when multiple years are selected. For example, the total amount of imports from country B from year 1 to year 5 will be a summation. This will be done on the fly and not a part of processing the data ahead of time.

Our data processing implementation will include checking for duplicate rows, trimming leading/trailing whitespace in columns, checking for spelling and capitalization inconsistencies, and using a 'no-data' flag for applicable cells. Additionally, our primary dataset from Correlates of War contains several columns of data that we are not using,

so we will remove these extraneous columns. All of these cleaning/processing tasks can be done using Excel.

Update – Project Milestone

The data processing step was more intensive than we first thought. We had to include a column in each of our datasets that indicated the three-letter country code rather than just the country name. We needed a way of linking all of our datasets so that there would be synchronization in our visualization. The three-letter country code addition made that possible. The map already includes the country code, so we had to add it to the dyadic, national, gdp, and capital cities datasets.

In addition to removing extraneous columns and performing basic data cleaning, we also filtered the dyadic and national data to include only the years from 1990 to 2014. This was a result of our design decision based on our peer review to limit the number of years to avoid major map change issues. We then created separate .csv files, one for each year for both dyadic and national. We did this for easier data loading on year selection. The cleaned, usable datasets can be found in the Data folder in our github repository.

As we have been implementing the visualizations, we ran into a couple of issues in regards to using the data once it was loaded. Our visualizations require aggregation of the exports and imports over the selected time periods. Currently, we are using d3.nest() and d3.rollup() to aggregate the values over the selected time periods, but we are considering alternatives to increase performance.

Update – Dyadic Data Preprocessing Overhaul

To increase the performance of our visualization and after discussion with our assigned TA, we decided to perform further preprocessing of the dyadic data. Previously, the dyadic data was broken down by year into separate .csv files.

Dyadic_1990.csv	11/3/2018 2:09 PM	Microsoft Excel Co...	575 KB
Dyadic_1991.csv	11/3/2018 2:10 PM	Microsoft Excel Co...	679 KB
Dyadic_1992.csv	11/3/2018 2:12 PM	Microsoft Excel Co...	714 KB
Dyadic_1993.csv	11/3/2018 2:12 PM	Microsoft Excel Co...	757 KB

When different time ranges were selected, our update function would pull from these files and perform processing detailed in the above section. Now, the dyadic data is broken down by primary country and year into separate .csv files.

AFG_Dyadic_1990.csv	11/15/2018 10:01 …	Microsoft Excel Co…	10 KB
AGO_Dyadic_1990.csv	11/15/2018 10:01 …	Microsoft Excel Co…	9 KB
ALB_Dyadic_1990.csv	11/15/2018 10:01 …	Microsoft Excel Co…	9 KB
ARE_Dyadic_1990.csv	11/15/2018 10:04 …	Microsoft Excel Co…	11 KB

We used a Java library called OpenCSV to read our old .csv files and write new .csv files based on country and year. The code can be found on our github repository under OpenCSV.

```java
    */
csvReader = new CSVReader(new FileReader("countries.csv"));

// List for holding all the rows
List<String[]> countries = new ArrayList<String[]>();

countries = csvReader.readAll();
// Read individual row from List of rows
for (String[] country : countries) {

    for (int i = 1990; i < 2015; i++) {
        CSVReader yearReader = null;
        yearReader = new CSVReader(new FileReader("Data/Dyadic/Dyadic_" + i + ".csv"));

        List<String[]> data = new ArrayList<String[]>();
        List<String[]> filteredData = new ArrayList<String[]>();
        data = yearReader.readAll();

        for (String[] datum : data) {
            // System.out.println(Arrays.toString(datum));
            String[] temp = new String[7];
            if (datum[2].equals(country[0])) { // importer 1 is pri -- flow1 import flow 2 export
                filteredData.add(datum);
            }
            if (datum[4].equals(country[0])) { // importer 2 is pri -- flow2 import flow1 export
                temp[0] = datum[0];
                temp[1] = datum[3];
                temp[2] = datum[4];
                temp[3] = datum[1];
                temp[4] = datum[2];
                temp[5] = datum[6];
                temp[6] = datum[5];

                filteredData.add(temp);
```

Based on the primary country selected and the year range, our update function loads a much smaller amount of data, thus, increasing performance.

Update – Custom Data Processing

Because we are computing trade data over time we read in as many files as there are years selected and perform data aggregation and sorting. We average imports and exports over the selected year range and then sort this data to compute who are the top importers and exporters of the primary country. Additionally, we compute total imports and total exports from the primary country to be able to articulate the percentage of trade from a single country.

To calculate this information, we had to iterate through the data several times. After many refinements and improvements to our algorithm we were able to minimize this processing to an acceptable level. One of the biggest gains in performance came from trimming the data early in the process. Immediately after summing the total imports and total exports for the primary country we eliminate all countries from our data processing except the top 30 trade partners. This resulted in an 85% improvement in processing time.

Must-Have Features

1. Interactive Map
2. Single Country Elements:
    a. Trade Balance over time (comparing imports to exports)
    b. Top trade partners - From Imports and Exports
3. Dyadic Elements:
    a. Compare import/exports correlated with GDP Per Capita between 2 countries from the perspective of the primary country
    b. Global Trade Balance of the 2 Countries

Optional Features

1. Holistic data representing the entire data set regardless of selection
2. Tariff Data Set (US centric) – will only show when US is primary country
3. Storytelling Element - 'Show Me Over Time' button that animates the visualization through the years

Project Schedule

Week	Plan	Important Deadlines
10/19 - 10/26	- Conduct individual research in topic area and prepare additional data sources. - Individual brainstorming for vis ideas - Meet as a team on Tues 10/23 to work on Project Proposal - Final dataset(s) decision - Data Cleaning / Processing	10/26 - Proposal Due
10/27 - 11/2	- Peer evaluations - Use feedback to refine visualization sketches - Draft visualizations complete - Finalize protocol for interactivity of views - HTML layout complete - File / class structure complete	10/30 - Peer Evaluations
11/3 - 11/9	- Views creation in progress - Interactivity in progress	11/9 - Project Milestone Due
11/10 - 11/16	- Complete Views/Interactivity - Project Website hosted on GitHub pages	11/16 - Functioning Draft Complete
11/17 - 11/23	Testing (peer testing) - Final adjustments and error handling	11/23 - Visualization Complete
11/24 - 11/30	Complete and embed screencast into project	11/30 - Project Due

Scheduling Update

We remained on schedule through the entire process. Establishing a schedule upfront was very helpful and allowed us to gauge how we were progressing throughout the project. The project information on the course website was very helpful in understanding the scope of the project and all the required components.

Visualization Design

Overview and Brainstorming

Overview: In order to determine the direction of our visualization design, we had a brainstorming meeting using some ideas from the Five Design Sheet Methodology to guide our process. This process allowed us to zero in on the sub-visualizations -- what we call elements -- that when combined, make up our total visualization. We will describe the process, what elements we decided were the most important, the alternative layouts we discussed, and the current favorite (i.e. the final design that incorporates the best of the alternatives).

Brainstorming Process: To start, we started generating ideas by sketching different ways to visualize our data set on post-it notes and sticking them to a whiteboard.

Next, we filtered, categorized, and combined the ideas. We discussed the pros and cons of the ideas. We further questioned whether the idea was answering the questions we posed as the purpose for our visualization (how can we better represent the data?).

Some of the most important discussion points were as follows:
- Using a map instead of a standard graph
- Representing top trade partners
- Appropriately showing trade balance
- User interaction – how will the user focus in on details

Sketches developed during brainstorming
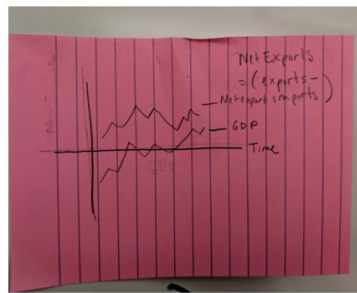
Line Chart:
Comparing Net Exports and GDP over time

Bar Chart:
Compare Global Balance between
two countries
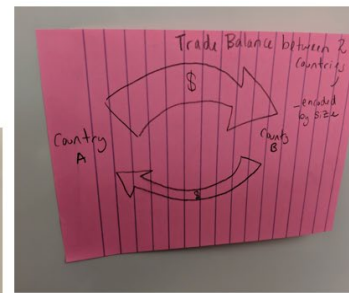
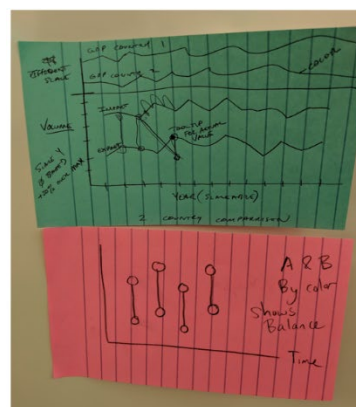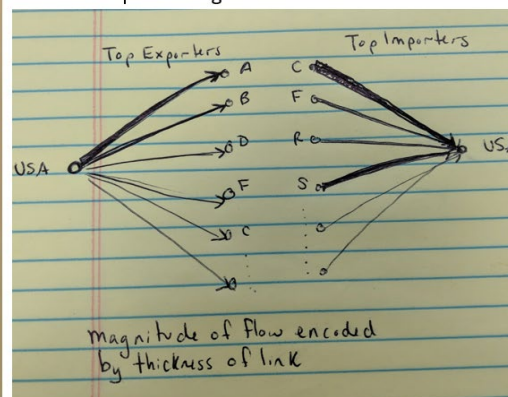Trade Flow Diagram:
Shows flow between countries
encoded by size
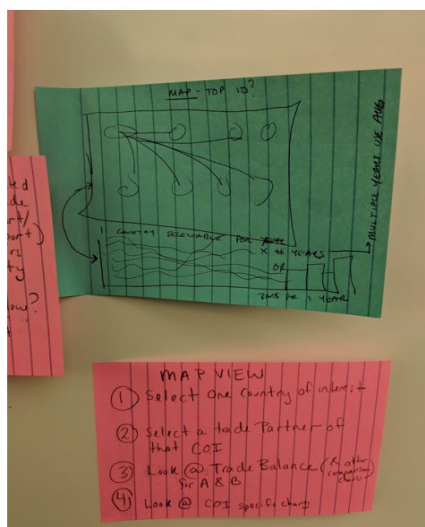
Maps:
Representing the trade connections

Graph:
Representing the trade connections

magnitude of flow encoded
by thickness of link

Stacked Line Chart:
Correlating GDP and imports/exports
between two countries

Multiple Bar Charts:
Top import sources and export destinations
(% of total imports/exports) for one country
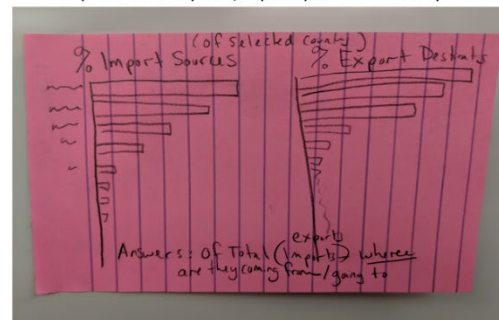
Putting it All Together - Sub-Visualizations:

We selected five separate visualizations to be included in our project. They are detailed below. Based on positive peer feedback and more time to refine our visualizations we have made several adjustments since submitting our proposal. Below is a brief description of each visualization and the evolution in design.

Visualization: World Map

Initial Design:

Map Implementation for Milestone:
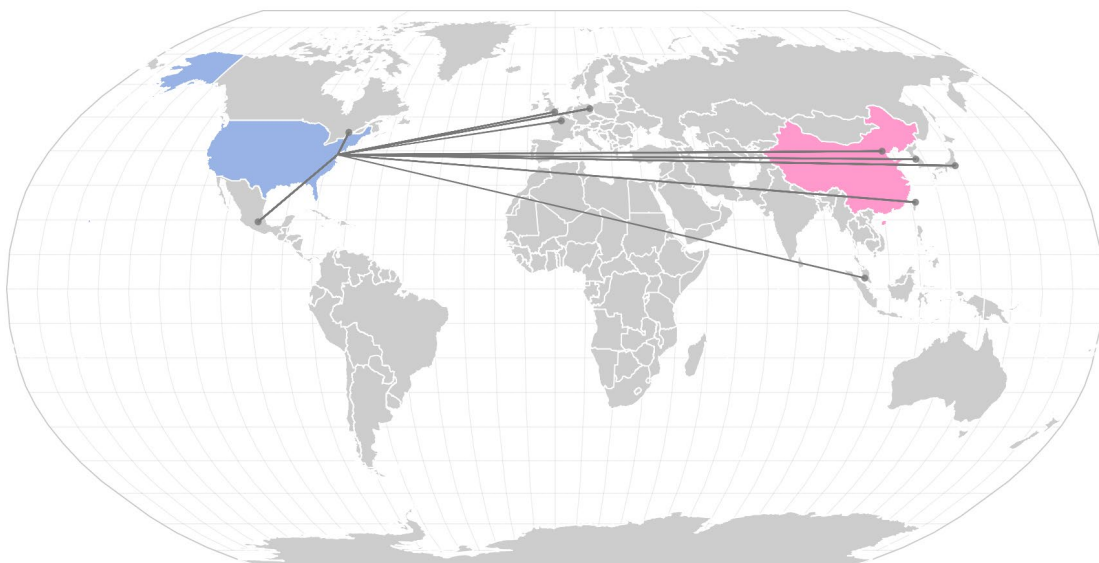
Map Final Implementation with Descriptive Info Box:

Primary Country: United States of America
Secondary Country: China
Year Range: 2010 - 2014
*Clicking on the primary/secondary country will toggle, otherwise a new secondary country will be selected
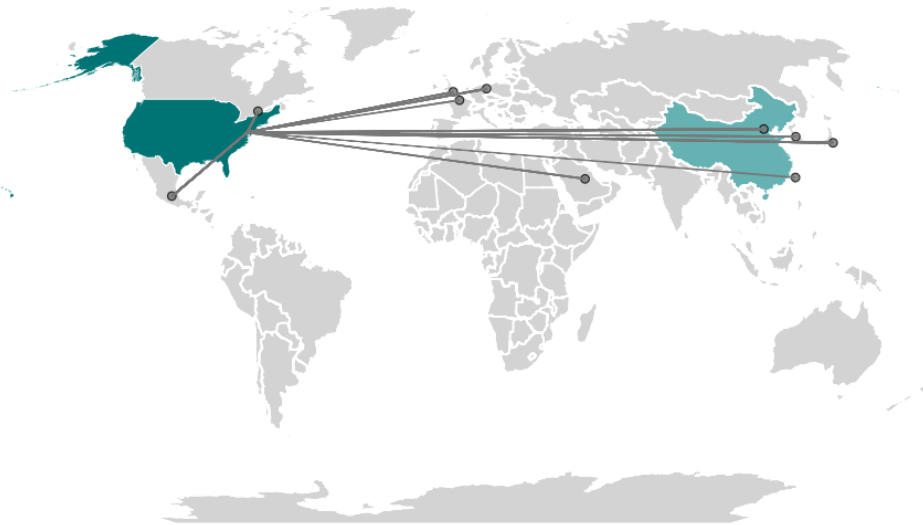
Default View: United States (primary), China (secondary) – different colors.
Representation: Show the top 10 trade partners of the primary country selected.
Interaction: There will always be 2 countries selected (primary and secondary).
Choice 1: Click on the primary country – toggle as secondary country.
Choice 2: Click on the secondary country – toggle as primary country.
Choice 3: Click on a non-selected country will set that country as the secondary country.
Constraints: None
Design Evolution: As implementation progressed, the map went through a few modifications from the original design. First, we removed the coloring on the links between the primary country and it's top 10 trade partners. Since we have the top traders chart next to the map, the link coloring was redundant. Second, we changed the coloring of the primary and secondary countries. We are using a darker teal for the primary and a lighter teal for the secondary to indicate order.

Visualization: Top Traders

Year Range: 2000-2010

% of total exports

China: 35% 1.5 Billion

% of total imports

China: 35% 1.5 Billion

Volume ($$ Millions)
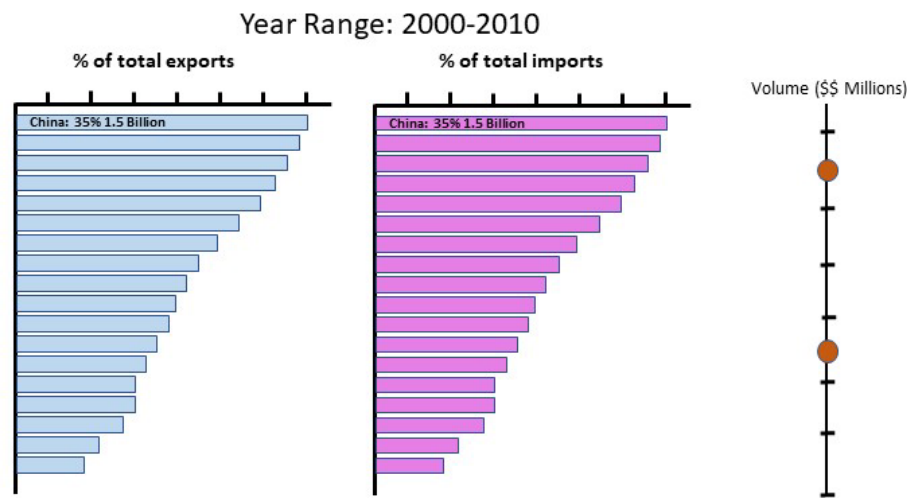
Top Traders Implementation for Milestone:
We had planned to do a parallel coordinates view with this chart but ran into a problem with screen space. We made the decision is was more important have text as part of the display then requiring the user to hover over the bars to know what country they were looking at.

Top Trade Partners

Top Exporters($Mil)   Top Importers($Mil)
150 100 50   0   0   100   200
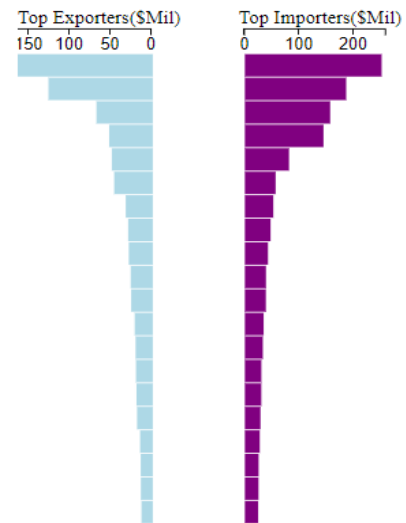
Top Traders Final Implementation:

Top Trade Partners

Top Exporters($ Millions of US Dollars)
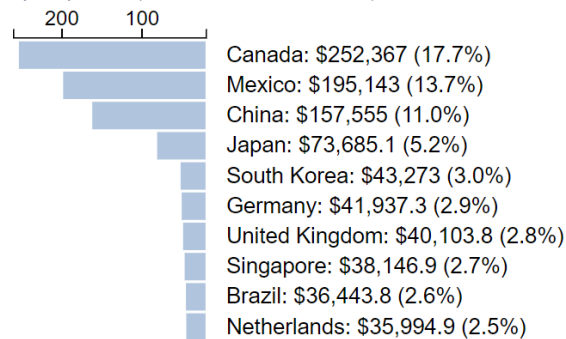
200	100

Canada: $252,367 (17.7%)
Mexico: $195,143 (13.7%)
China: $157,555 (11.0%)
Japan: $73,685.1 (5.2%)
South Korea: $43,273 (3.0%)
Germany: $41,937.3 (2.9%)
United Kingdom: $40,103.8 (2.8%)
Singapore: $38,146.9 (2.7%)
Brazil: $36,443.8 (2.6%)
Netherlands: $35,994.9 (2.5%)

Top Importers($ Millions of US Dollars)

400	200	0

China: $439,617 (19.6%)
Canada: $322,819 (14.4%)
Mexico: $270,912 (12.1%)
Japan: $136,497 (6.1%)
Germany: $107,100 (4.8%)
South Korea: $60,949.7 (2.7%)
United Kingdom: $53,307 (2.4%)
Saudi Arabia: $47,775 (2.1%)
France: $43,207.4 (1.9%)
Taiwan: $40,010.9 (1.8%)
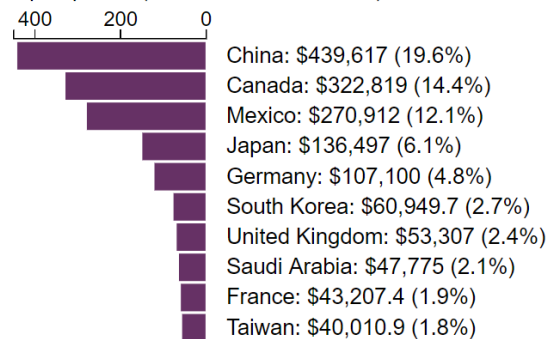
Default View: United States (primary) – 2010-2014 (5 years selected).
Representation: Shows top 10 export and top 10 import countries of the primary country in descending order by percentage of total imports and exports.
Interaction: Hovering over any bar will highlight the corresponding country on the map and in the export/import section. Clicking on a bar will select that country as the secondary country.
Constraints: None.
Design Evolution: We went through three major changes on this view. We are very satisfied with the final result and think it best represents the data to the user.

Visualization: Primary Country Import and Export Volume

Initial Design:
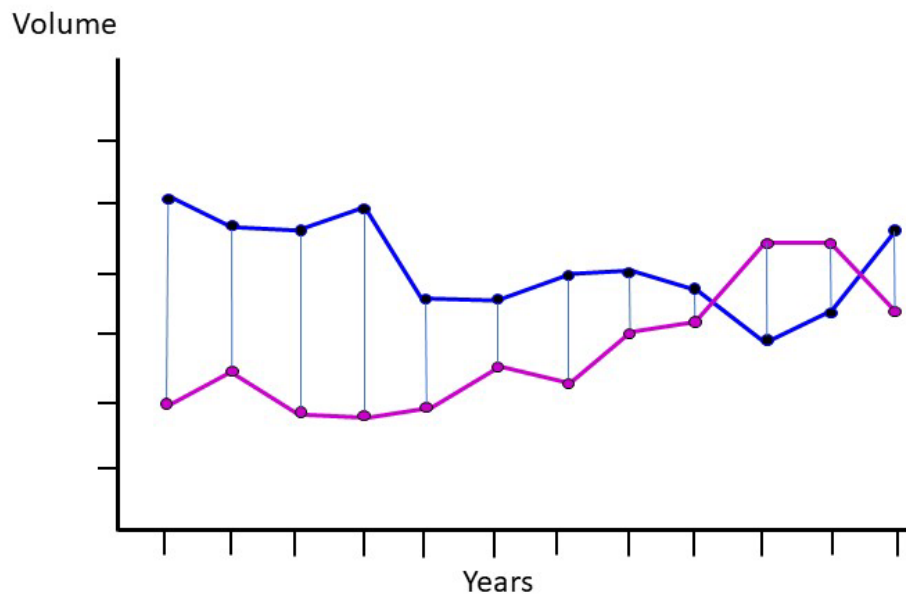
Primary Country Import and Export Volume Final Implementation:
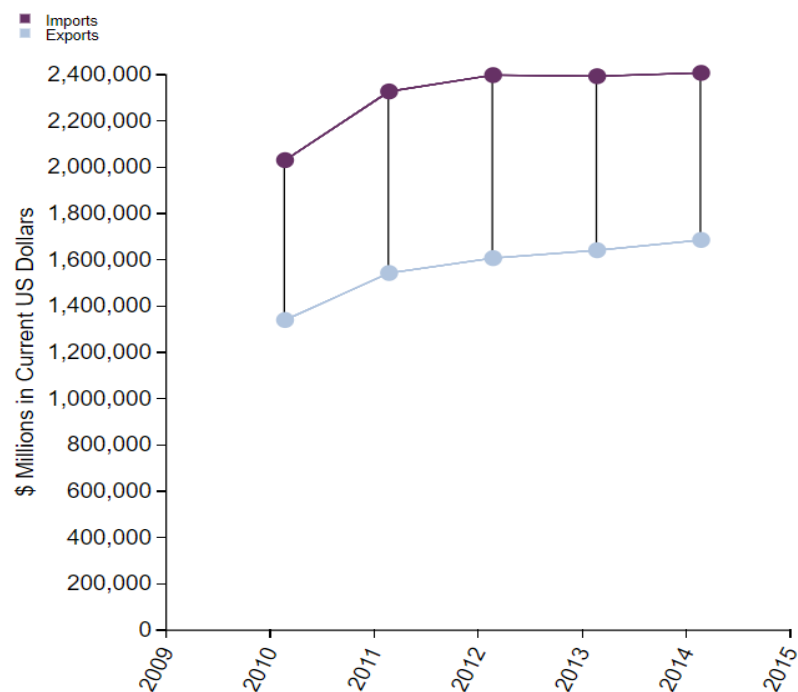
Default View: United States (primary) – 2010-2014 (5 years selected).
Representation: Shows the total volume of imports and exports over time. X-axis is time. Y-axis is volume in US dollars. One line represents total imports, the other total exports. The line between shows the difference.
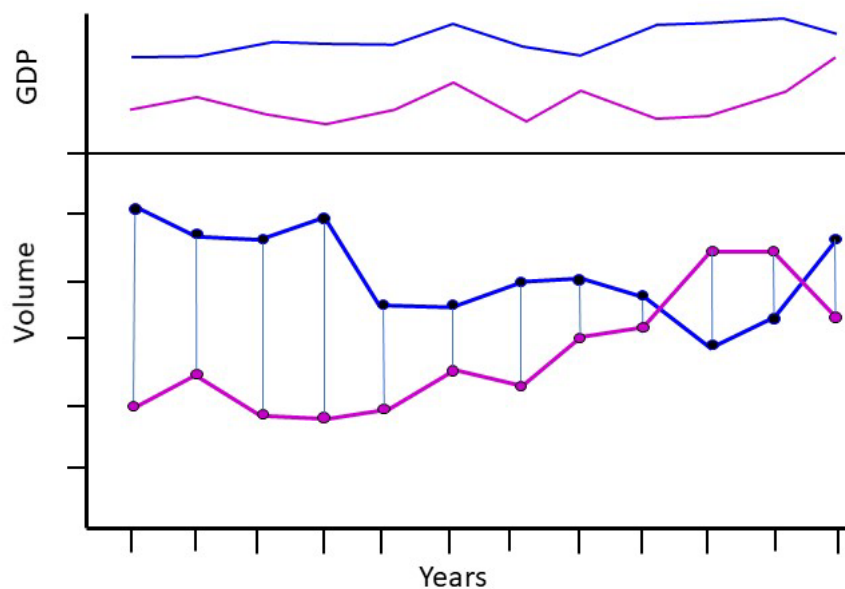**Interaction:** Highlighting over any point or the line between imports and exports will show a tool tip with detailed data. Changing the year selector will resize the x-axis, adjust the scales appropriately, and update the graph.
Constraints: None
Design Evolution: Previously, we had a constraint of a minimum of 5 years being selected for this chart. However, we removed that constraint, allowing users to select a single year.

Visualization: Primary and Secondary Country Comparison

Initial Design:

Primary and Secondary Country Comparison Final Implementation:
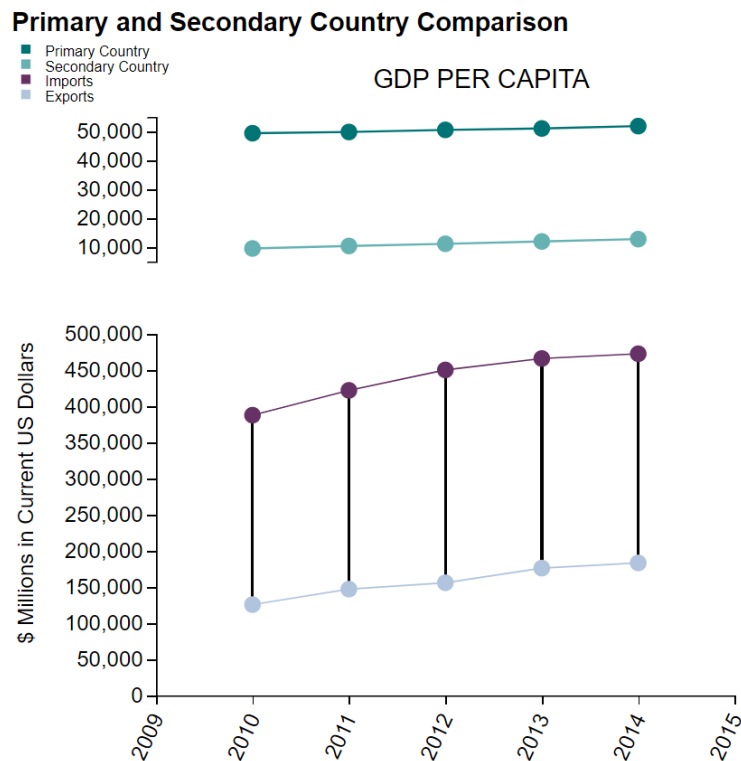
Primary and Secondary Country Comparison

GDP PER CAPITA

Default View: United States (primary) and China (secondary) between 2010-2014 (5 years selected).

Representation: Shows the volume of imports and exports over time between two countries and the deficit/surplus. X-axis is time. Y-axis is volume in US dollars. One line represents imports TO primary country from secondary country, the other line represents exports FROM the primary country to the secondary country. The line between shows the difference. The top view shows the GDP per capita for both countries.

Interaction: Highlighting over any point or the line between imports and exports will show a tool tip with detailed data. Changing the year selector will resize the x-axis, adjust the scales appropriately, and update the graph.

Constraints: None.

Design Evolution: No major changes from initial design.

Visualization: Global Imports / Exports Comparison
Initial Design:
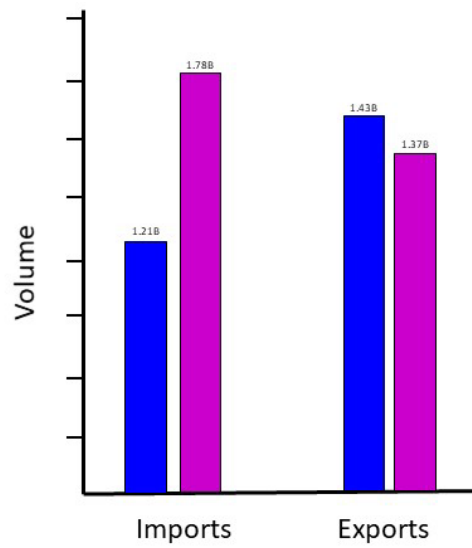
Updated Design after Peer Feedback:
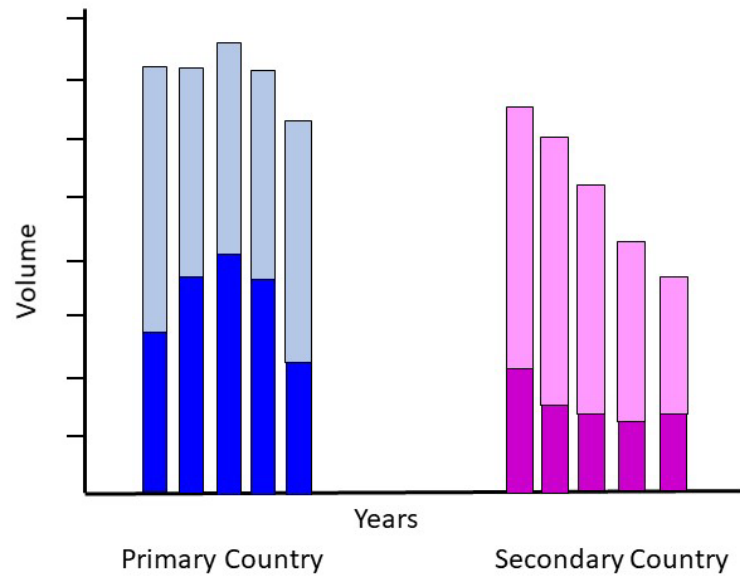
Global Imports / Exports Comparison Final Implementation:

Primary and Secondary Country Global Imports Comparison
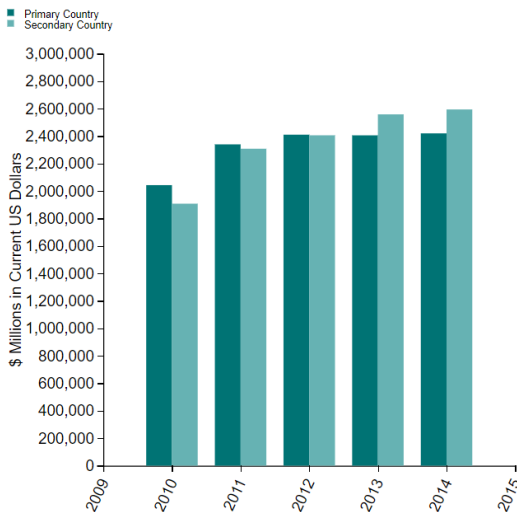
Primary and Secondary Country Global Exports Comparison
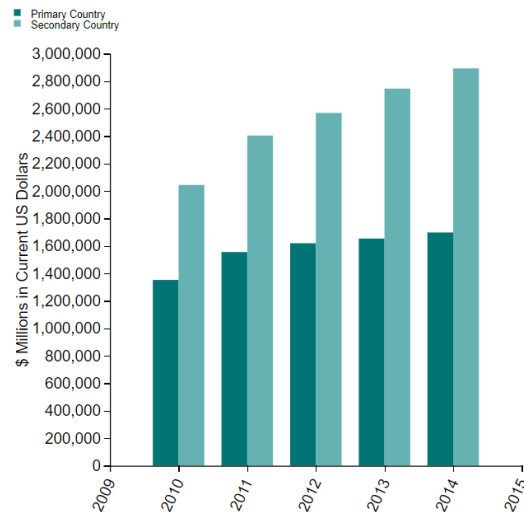
Default View: United States (primary) and China (secondary) between 2010-2014 (5 years selected).

Representation: Compares the sum of global imports and exports between the primary and secondary country. Data is calculated as a sum (or average) over the range of years. This allows users to better understand how much the primary and secondary country is trading with the rest of the world.

Interaction: Hovering over a bar will display a tool tip with detailed information.

Constraints: None.

Design Evolution: In the end, we decided to have two bar charts side by side. One compares global imports and one compares global exports of the primary and secondary countries for the selected time range.

Putting it All Together – Layouts

Initial - Proposal

Draft Layout Selection:
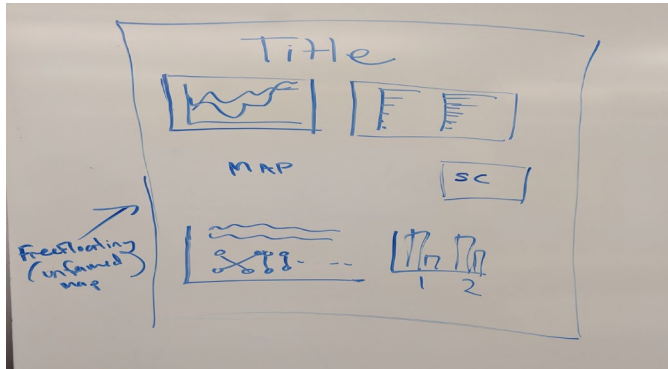
Design Alternatives:

Linear Design 1:
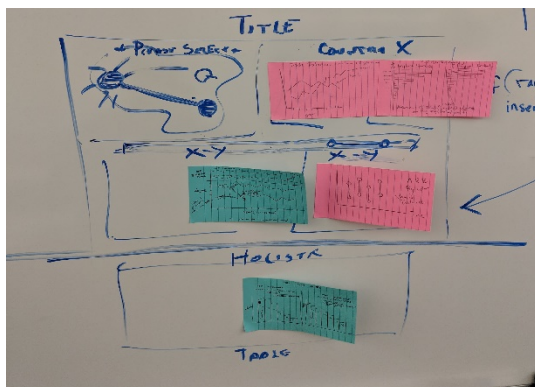
Linear Design 2:
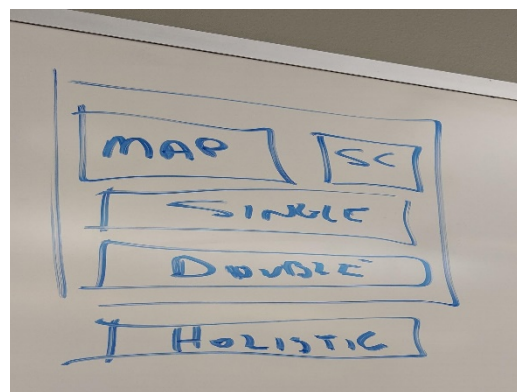
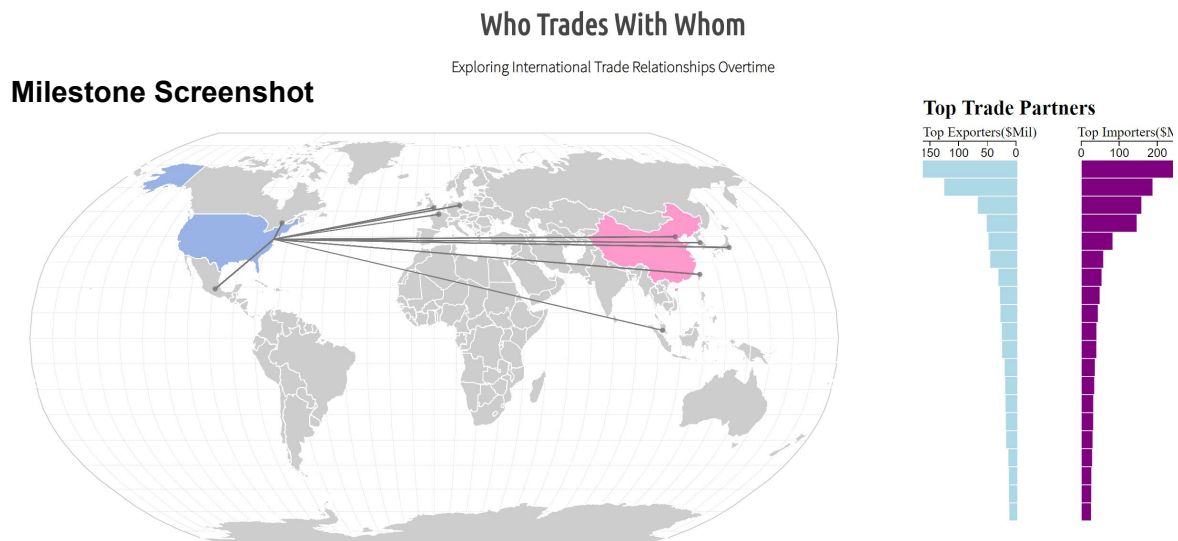Map Centric with "Holistic":
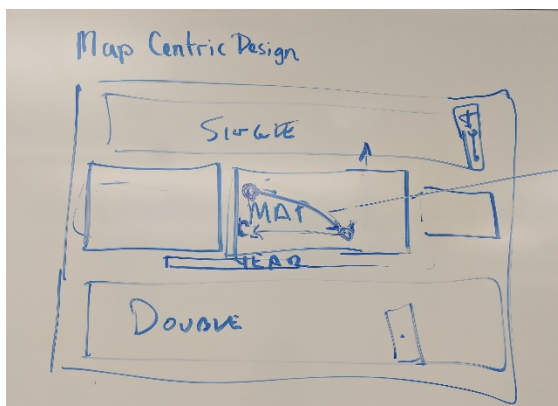
No Map Design:
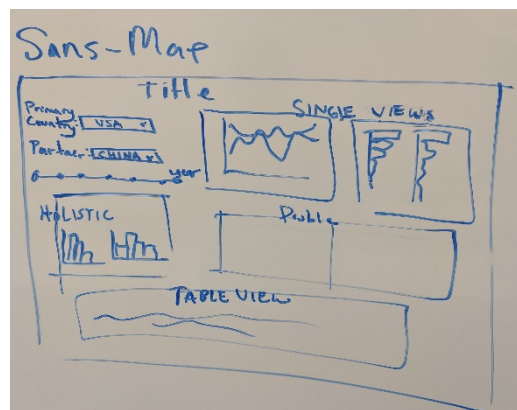
Update – Project Milestone

After several iterations, peer feedback, and some testing we have adjusted our layout. This is a current screen shot of our project as of 9 November 2018. We will continue to refine as we complete all of our visualizations.

Trade Balance-Single, Trade Balance-Double, and Global Balance visualizations will all go below these two primary visualizations. We discovered that having the more central in the view was critical to good interaction. The map is the driver in selecting countries and will eventually be joined by the year selected bar immediately below. We kept our original design drafts in this document to show the progression to where we are today.

Final Layout

Evaluation

In the design and implementation of this project we learned a lot about our data. It became evident very quickly that there are about 10 countries in the world that dominate international trade. Each of our views addressed one of our questions. With more time and creativity we could have combined multiple views into one view and perhaps answered additional questions and provided additional insights.

Our visualization functions properly under all tested conditions. The map works well for larger countries; however, is less effective for small countries, such as those in Europe, due to the size and proximity to their trade partners. Our views are very common – bar charts, line charts, etc – but they communicate the data in an effective way and address the questions we set out to answer.

This project illustrates a very high level of international trade. It could be improved upon by adding categorical trade information such as agriculture, steel, automobiles, etc and allowing the user to drill down into details. A nice feature would be able to visualize the top exporters of a certain type of item. Or to see who imports the most of a certain type of item. Additionally, adding tariff information would be very valuable in understanding "fairness" or existing constraints that exists between two trading countries.