Project Title: Who Trades With Whom - Exploring International Trade Relationships Over Time – <mark>Project Milestone</mark>

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/

<mark>**Current Status for Milestone**</mark>
We are on schedule with our implementation and the project is moving forward nicely. Our web site is live at the link above. Currently our visualizations are not showing on the live site. We think this is due to the large amount of data processing that is occurring. We are refining our data wrangling algorithm in hopes to get this up and running soon. No other issues or setbacks to report.

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

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.

<mark>**Update for Project Milestone**</mark>

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

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 used to working. 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.

<mark>Update for Project Milestone</mark>

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.

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. - Conduct 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 interactions between different visualizations (interactive design) - 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 - Fully 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 - Final Project Due

Project Milestone: We are currently right on track with our proposed schedule. Our layout, interaction design, and file/class structure is complete. We are currently working on creating the different views that make up our visualization.

Visualization Design

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 (ie. is there a better, more effective way to represent the data?).

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

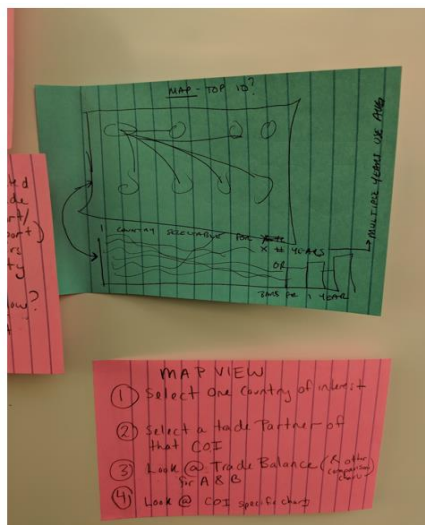# Sketches developed during brainstorming
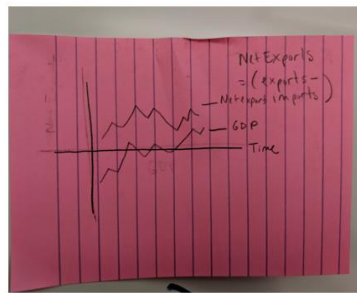
Line Chart:
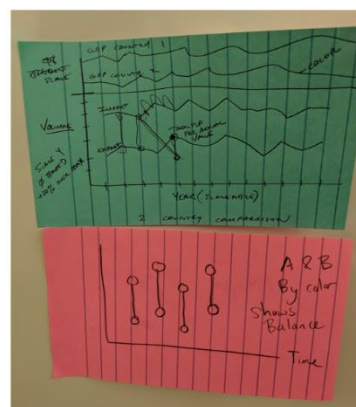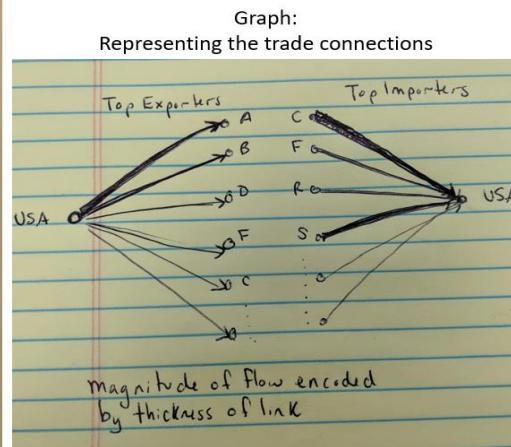Comparing Net Exports and GDP over time

Bar Chart:
Compare Global Balance between two countries
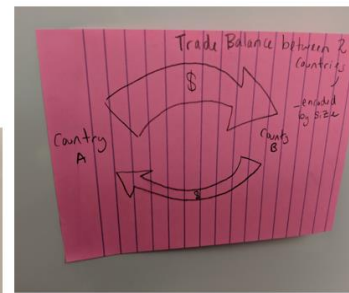
Trade Flow Diagram:
Shows flow between countries encoded by size

Graph:
Representing the trade connections

Maps:
Representing the trade connections

Multiple Bar Charts:
Top import sources and export destinations
(% of total imports/exports) for one country

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

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.

Visualization: World Map (Design and Draft)
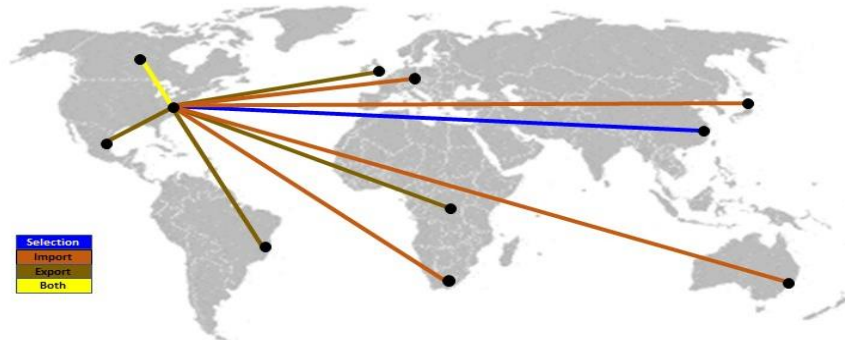
Screenshot
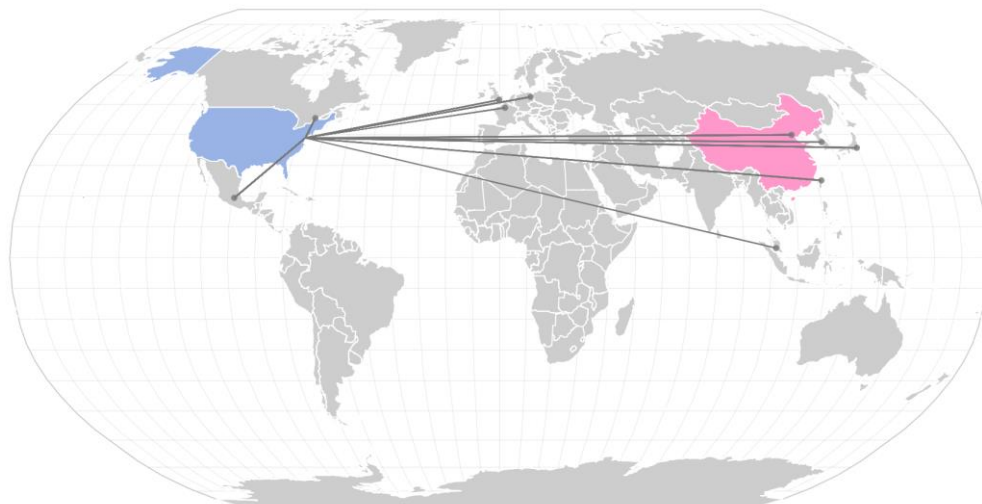
Default View: United States (primary), China (secondary) – both highlighted in different colors.
Representation: Show the top 10 trade partners of the primary country selected.
Interaction: There will always be 2 countries selected (primary and secondary).
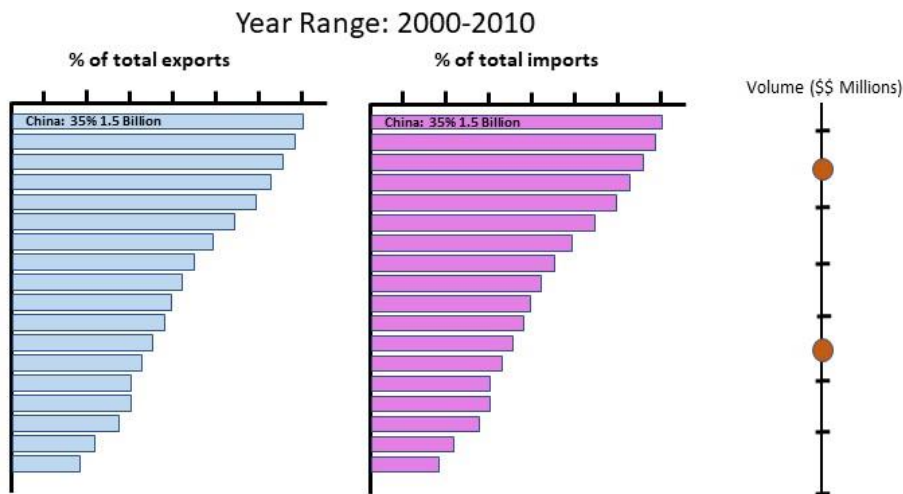Choice 1: Click on the primary country – do nothing.
Choice 2: Click on the secondary country – toggle as primary country.
Choice 3: Click any other country will transition the secondary country to the primary country and the new country as the secondary.
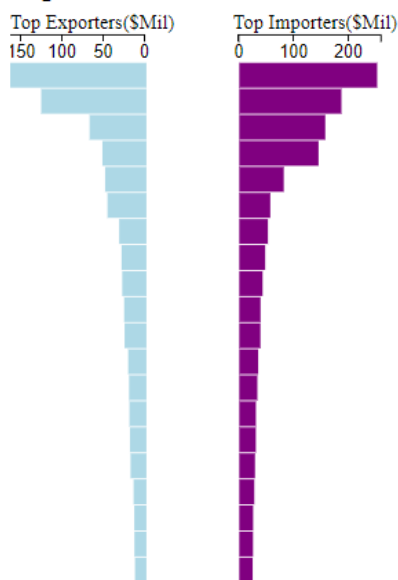**Constraints:** None
Implementation: 80%

Visualization: Top Traders (Design and Draft)

Year Range: 2000-2010

% of total exports

China: 35% 1.5 Billion

% of total imports

China: 35% 1.5 Billion

Volume ($$ Millions)

Top Trade Partners

Screenshot

Top Exporters($Mil)

150 100 50 0

Top Importers($Mil)

0 100 200

We will implement parallel coordinates on this view to show the relationship between top exporters and top importers.

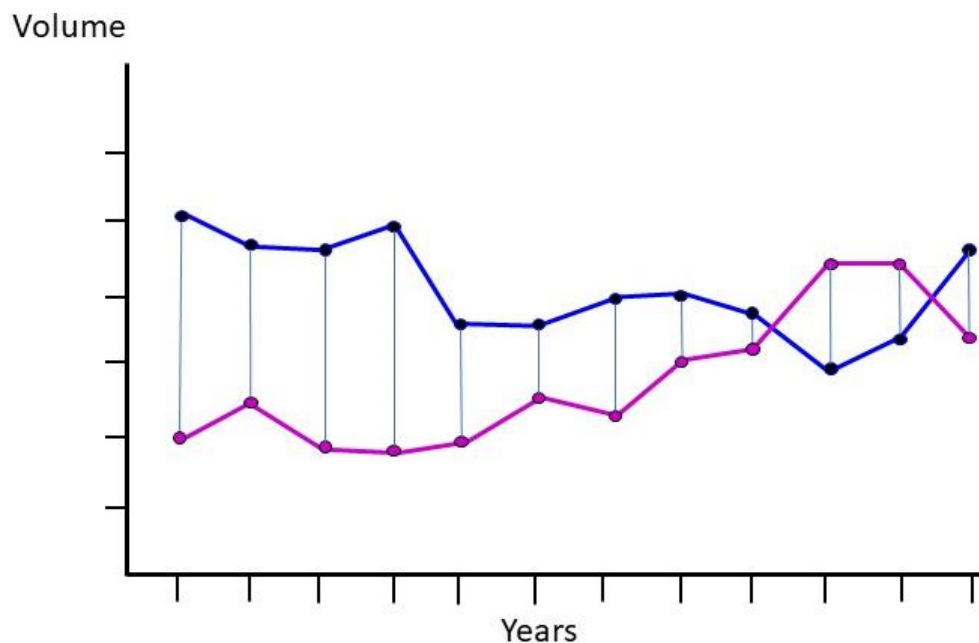**Default View:** United States (primary) – 2000-2005 (6 years selected).
Representation: Shows top import and export countries in descending order by percentage of total imports and exports. This view shows the major trade partners of the primary country.
Interaction: Highlighting over any bar will show a tool tip with detailed trade information. Highlighting over any bar will also highlight the country on the map and the corresponding import/export bar. Clicking on a bar will change the primary/secondary countries the same as a map selection. We will also add a parallel coordinates graph connecting the export and import bars to show corresponding countries.
Constraints: None.
Implementation: 80%

Visualization: Trade Balance – Single (Design)

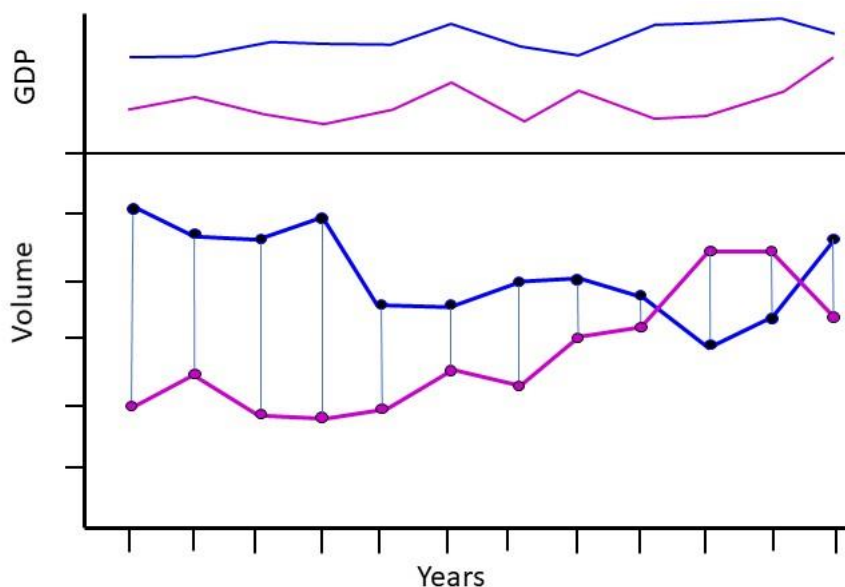**Default View:** United States (primary) – 2000-2005 (6 years selected).
Representation: Shows the total volume of imports and exports over time. X-axis is time. Y-axis is volume in US dollars. One line represents total imports, the other total exports. The line between shows the difference.
Interaction: 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: Five-year range minimum (we don't want a chart with less than five years).
Implementation: Coming soon.

Visualization: Trade Balance – Double (Design)

Default View: United States (primary) and China (secondary) between 2000-2005 (6 years selected).

Representation: Shows the volume of imports and exports over time between two countries and the deficit/delta. 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 for both countries. This view provides a complementary view to the trade view to see how GDP correlates with trends.

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: Five-year range minimum (we don't want a chart with less than five years).

Implementation: Coming soon.

Visualization: Global Trade Balance (Design and Redesign)
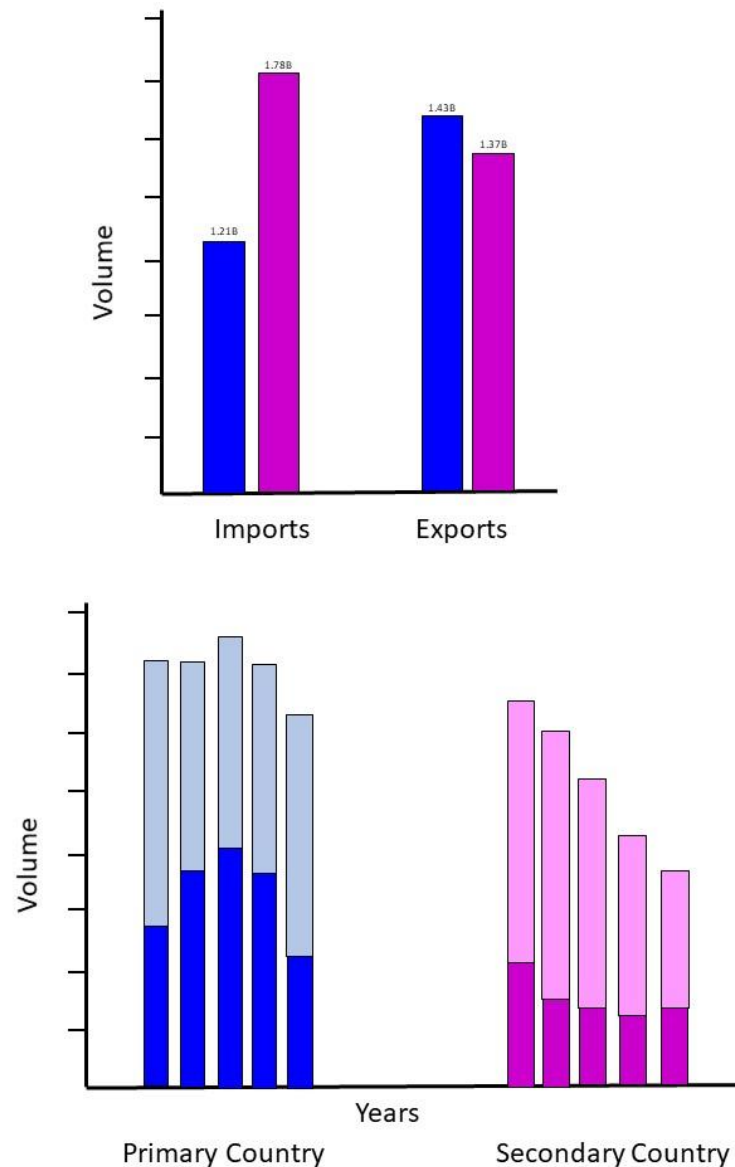
Default View: United States (primary) and China (secondary) between 2000-2005 (6 years selected).

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 trade is happening with the rest of the world. Y-axis is volume in US dollars. Plan for scale to be zero based.

Interaction: Updated based on selection of primary and secondary countries as well as changes to the year selector.

Constraints: None.

**Implementation:** Coming soon.

Putting it All Together - Layouts
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.
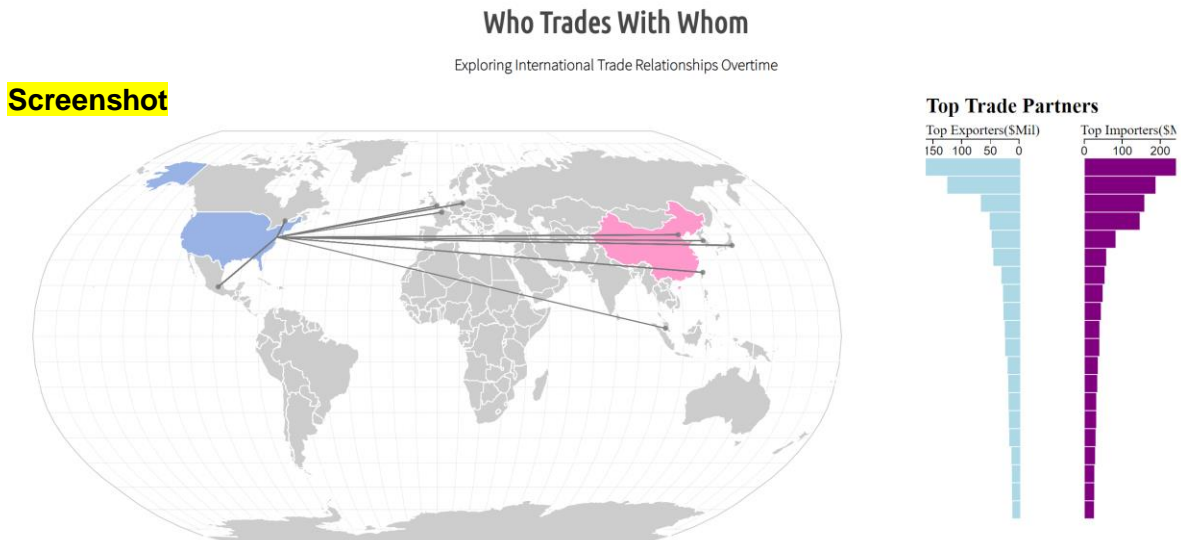

Trade Balance-Single, Trade Balance-Double, and Global Balance visualizations will all go below these two primary visualizations. We discovered that having the more central in the view was critical to good interaction. The map is the driver in selecting countries and will eventually be joined by the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

Draft Layout Selection:
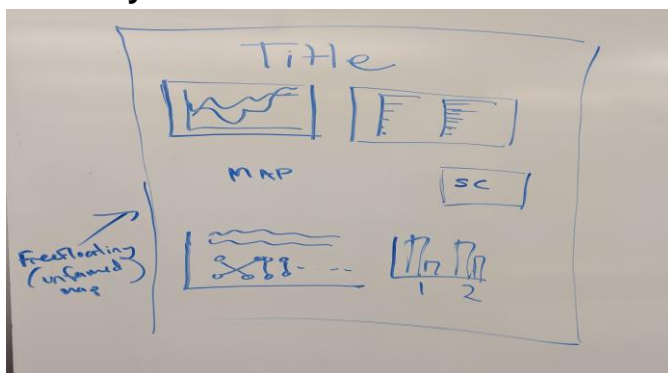
Design Alternatives:
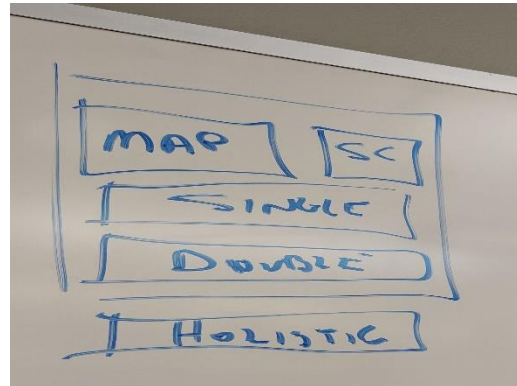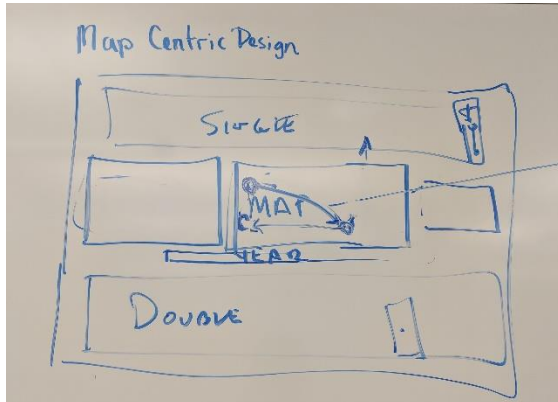
Linear Design 1:
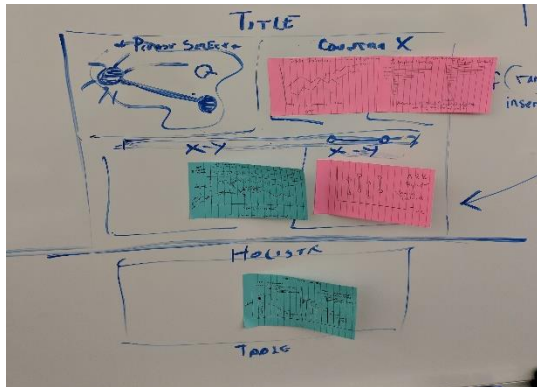
Linear Design 2:

Map Centric with "Holistic":

No Map Design: