

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)

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 biggest 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.

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.

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	<ul style="list-style-type: none">• 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	<ul style="list-style-type: none">• 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	<ul style="list-style-type: none">• Views creation in progress• Interactivity in progress	11/9 - Project Milestone Due
11/10 - 11/16	<ul style="list-style-type: none">• Complete Views/Interactivity• Project Website hosted on GitHub pages	11/16 - Fully Functioning Draft Complete
11/17 - 11/23	<ul style="list-style-type: none">• Testing (peer testing) - Final adjustments and error handling	11/23 - Visualization Complete
11/24 - 11/30	<ul style="list-style-type: none">• Complete and embed screencast into project	11/30 - Final Project Due

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.

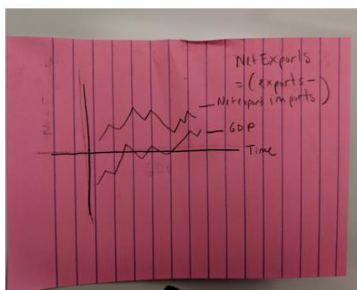


Next, we filtered, categorized, and combined the ideas. We discussed the pros and cons of the ideas. We further questioned whether the idea was answering the questions we posed as the purpose for our visualization (ie. is there a better, more effective way to represent the data?).

Some of the most important discussion points were as follows:

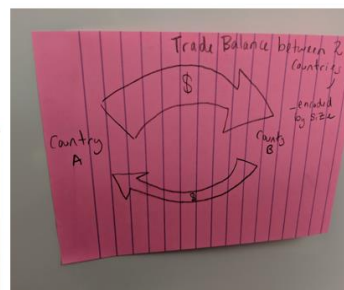
- Using a map instead of a standard graph (or omitting connection information completely)
- Representing top trade partners
- Appropriately showing trade balance
- User interaction – how will the user focus in on details

Sketches developed during brainstorming:

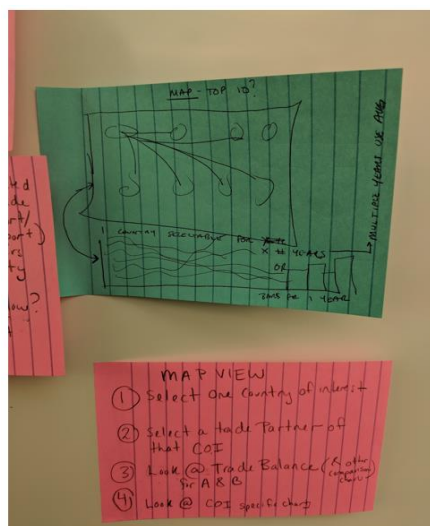


Line Chart:
Comparing Net Exports and GDP over time

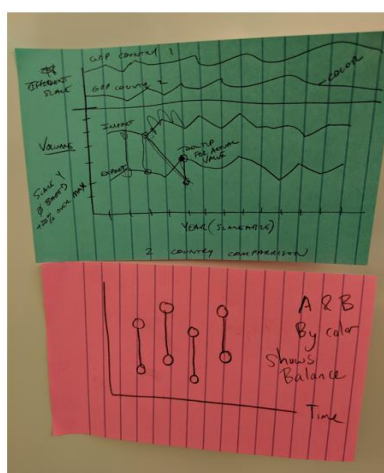
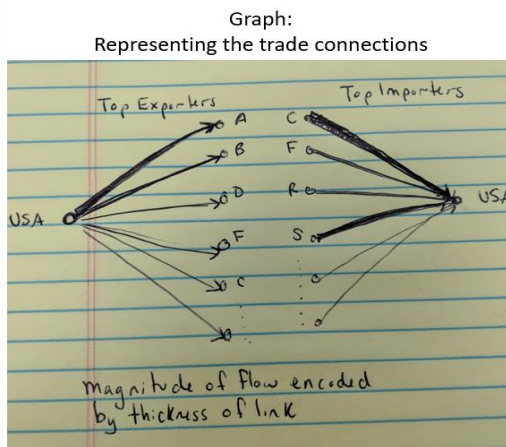
Bar Chart:
Compare Global Balance between two countries



Trade Flow Diagram:
Shows flow between countries encoded by size

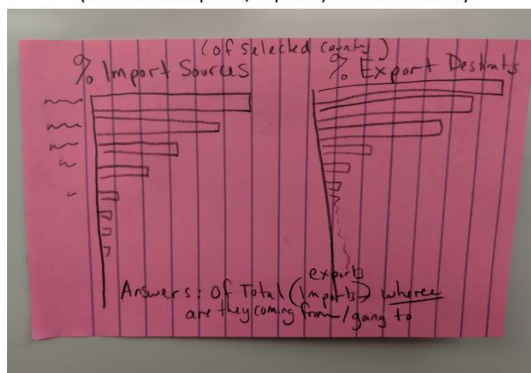


Maps:
Representing the trade connections



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

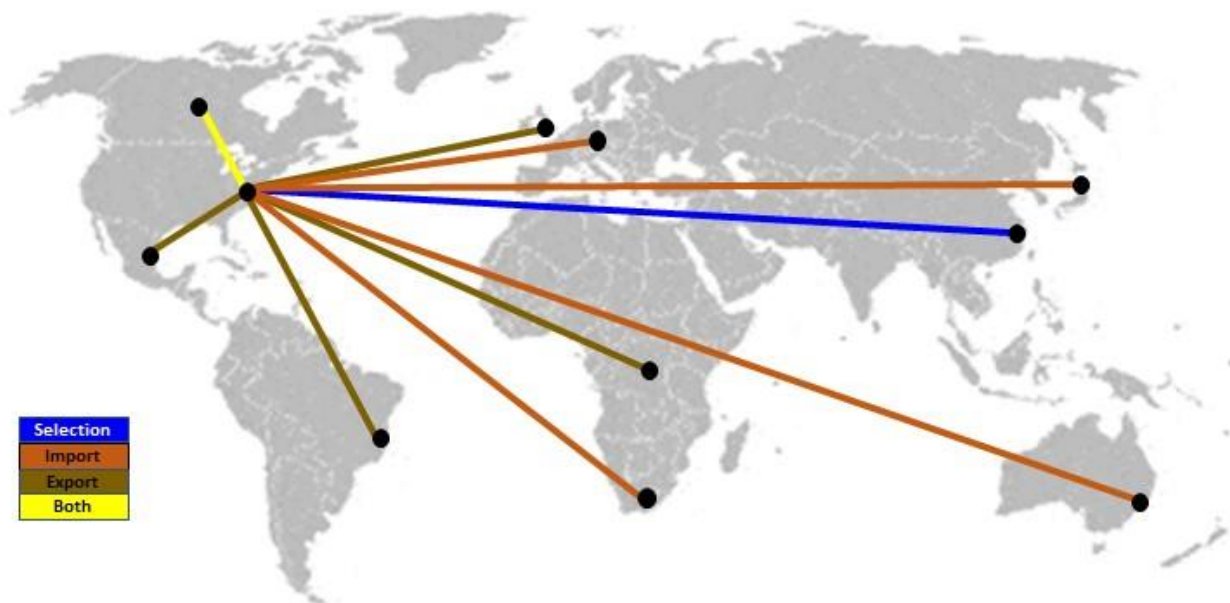
Multiple Bar Charts:
Top import sources and export destinations
(% of total imports/exports) for one country



Putting it All Together - Sub-Visualizations:

At the end of the process, we decided on the five best visualizations to be included as parts of our whole visualization. These visualizations are described in detail below.

Visualization Name: World Map



Default View: United States (primary), China (secondary).

Representation: Show 10 connections. The highlighted (blue in this example) shows the connection with the secondary country. The other 9 connections are top imports/exports to other countries. We will likely use 3 colors for these connections (import/export/both). Color of the connection may also correspond to volume (brighter hue = high volume).

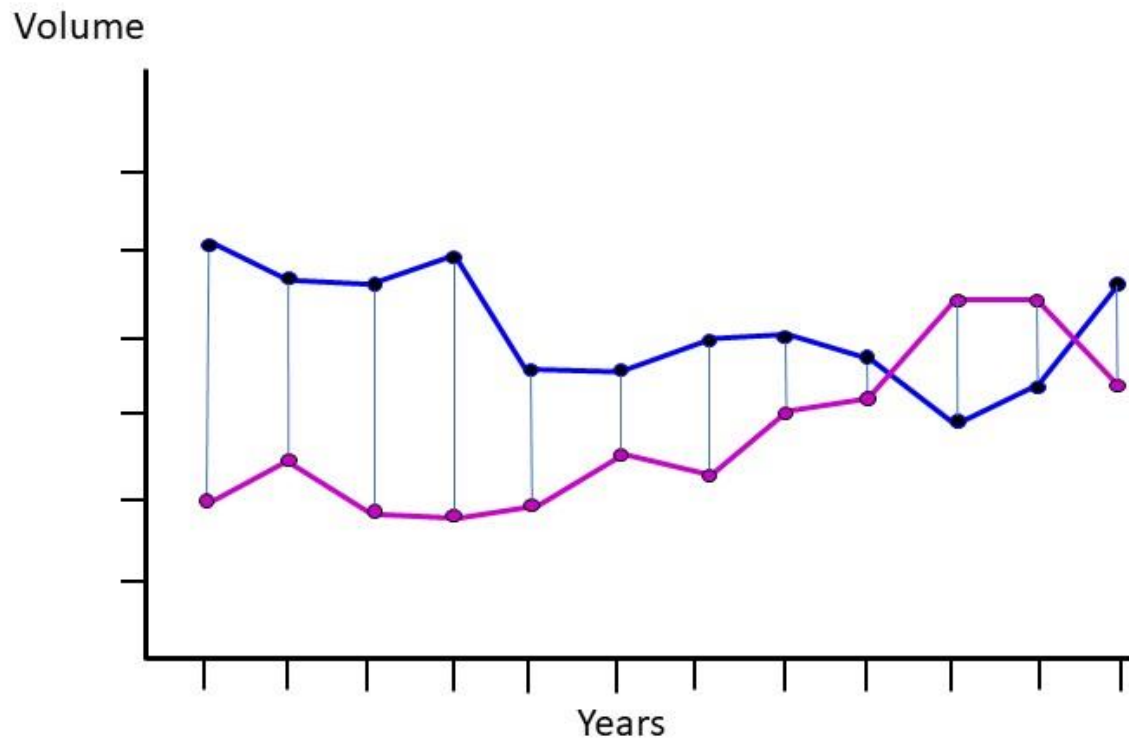
Interaction: There will always be 2 countries selected.

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.

Visualization Name: Trade Balance (Single)



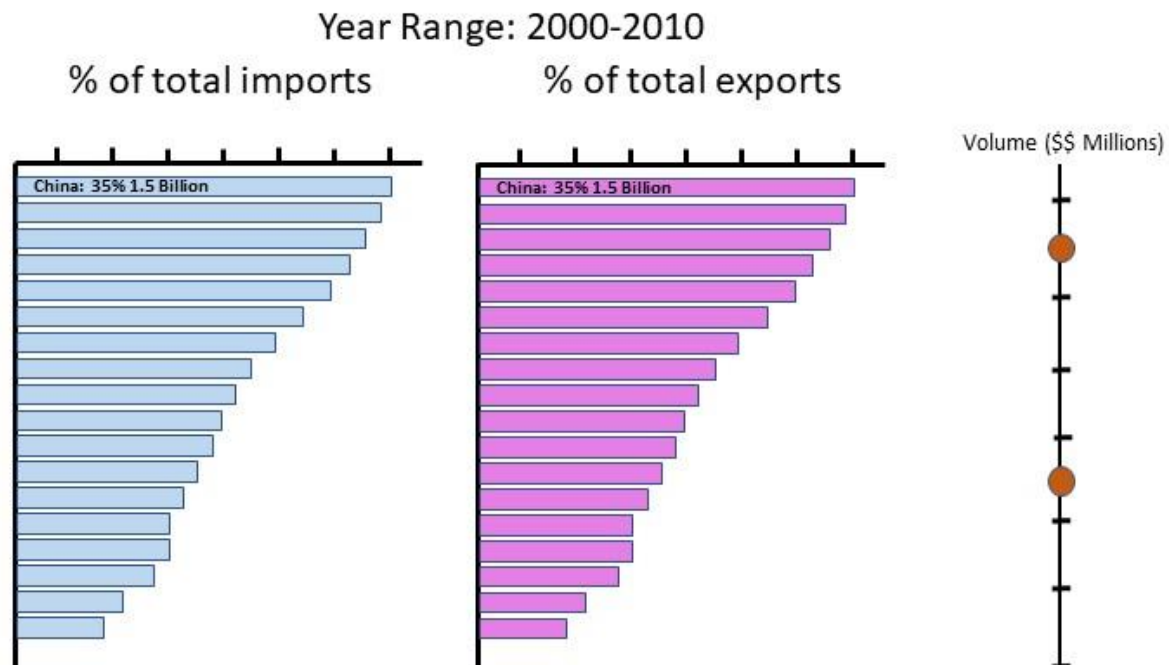
Default View: United States (primary) – 2000-2010 (10 years selected).

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: May impose a 5 year minimum on the year range.

Visualization Name: Top Traders (Single)



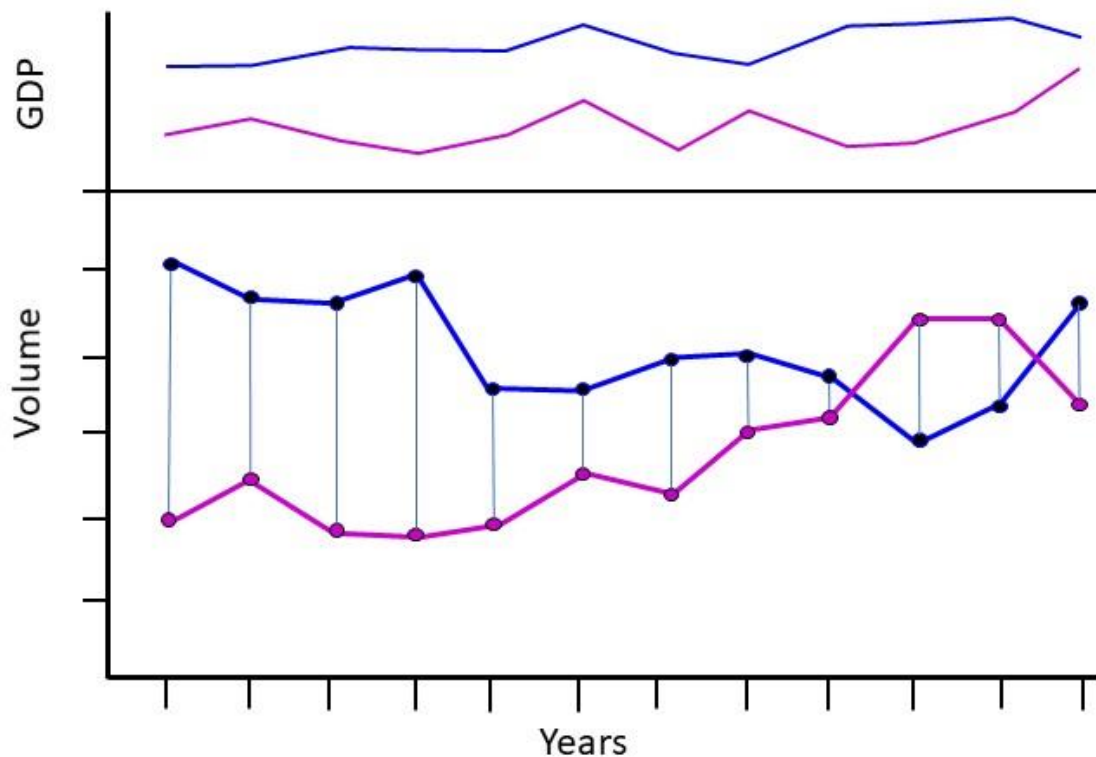
Default View: United States (primary) – 2000-2010 (10 years selected).

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. The default will be to show all the countries. When all countries are represented the bar charts will be very thin and not able to contain data (we will provide a tool tip). When the bars are large enough we will encode data within them.

Interaction: Changing the volume slider will narrow the focus to trade partners within the specified volume (if you want to find trade partners that trade between x million and y million you can customize. Additionally, clicking on a bar chart will change the primary/secondary countries the same as a map selection.

Constraints: None.

Visualization Name: Trade Balance (Double)



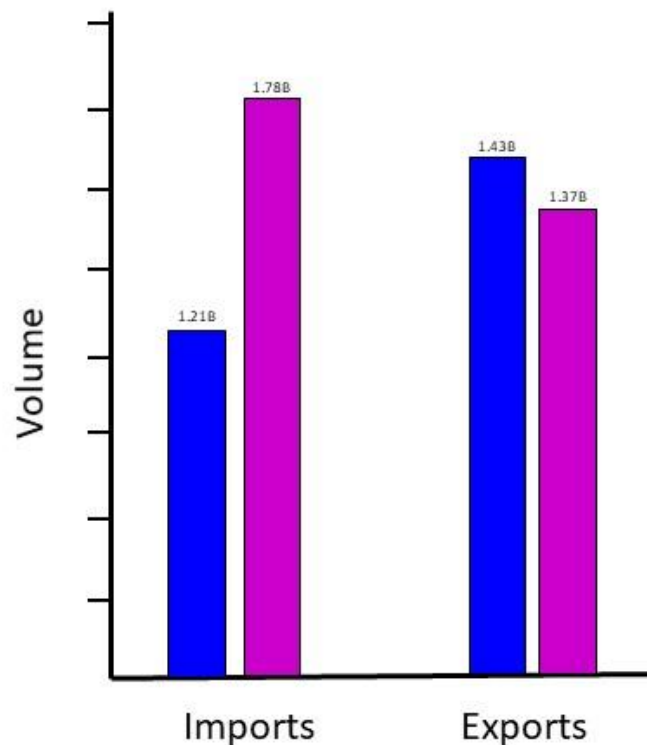
Default View: United States (primary) and China (secondary) between 2000-2010 (10 years selected).

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: May impose a 5 year minimum on the year range.

Visualization Name: Global Balance (Double)



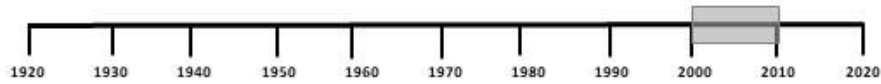
Default View: United States (primary) and China (secondary) between 2000-2010 (10 years selected).

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.

Visualization Tool: Year Selector



Default View: 2000-2010 (10 years selected).

Representation: Customizes the data to be relevant for the years selected.

Interaction: Will likely implement with a brush for ease of scaling and manipulating a specific year range.

Constraints: 5 year minimum.

Putting it All Together - Layouts:

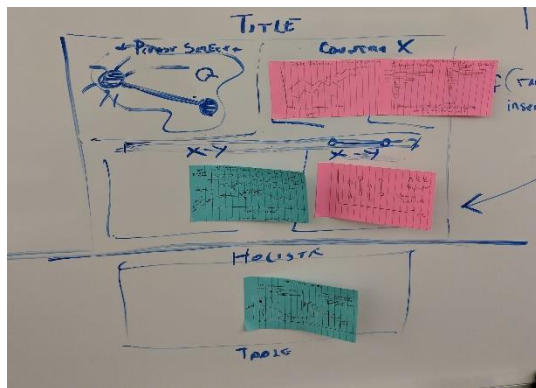
After discussing how we wanted to visualize the data by determining the parts of the whole, we then discussed various layout alternatives incorporating these visualizations.

Current Favorite ('Final' Design):

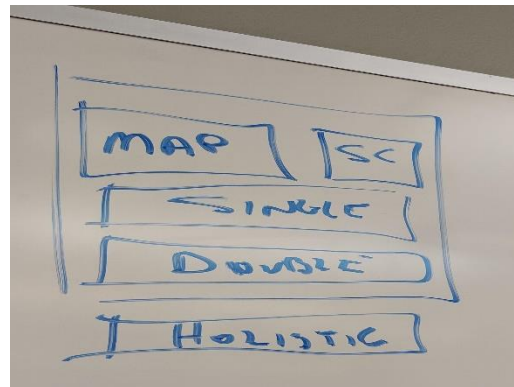


Design Alternatives:

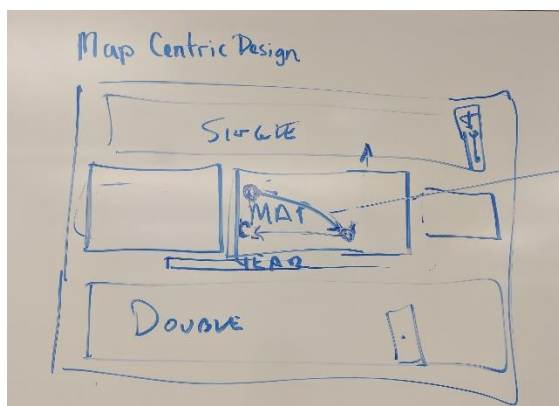
Linear Design 1:



Linear Design 2:



Map Centric with "Holistic":



No Map Design:

