Trade by the Decade

CS 5630/6630 Visualization for Data Science

Jake Osterloh
jake.osterloh@gmail.com
u0645881

Avram Twitchell
avram.twitch@gmail.com
u1193523

Oleks Korshak
alexokorshak@gmail.com
u1210377

[Project Github Page](#)
[Project Screen Cast](#)

Background and Motivation

Contemporary public debate deals a lot with the idea of international trade, and whether it helps or hurts. In recent years, many major political issues deal directly with trade, such as TTIP/TPP, Brexit, and tariffs. There are many interesting facets of the debate. Economic and historical theory suggest trade is almost always beneficial. Yet political pundits argue that trade deficits can hurt certain industries, and may lead to strategic vulnerabilities.

The motivation for this visualization is to use international trade data to contribute to this public debate by showing the connections that trading countries create. Admittedly, we come from a pro-trade point-of-view, but we believe that there is value in seeing what we export and their destinations, as well as what is imported and their origins. We also want to stray away from an adversarial "deficit" framework of trade, and focus more on what an individual country brings in or sends out.

Questions

What is interesting about trade between countries? Are there any anomalies or interesting trends that can be observed between countries? Which imports and exports are important for a country and what kind of dependencies do countries have based on their trade? Teach viewers how trade works by demonstrating with real world examples. This is less of a argumentative tool and more of an exploration tool so it is important that we show the potential of the visualization through the story telling aspect and then let viewers have fun exploring the rest of the data. It should be easy for people to find answers to questions like, "Which country is the largest exporter of X good?"

Data

The data comes from this site: https://atlas.media.mit.edu/en/resources/data/. It is an open source visualization of trade data produced at the MIT Media Labs. They have cleaned and readily accessible international trade data, going back to the 1960s. A single row in the data represents the total trade of a specific good between two countries in a given year. Ultimately, the data comes from a few sources: The Center for International Data from Robert Feenstra, UN COMTRADE, and BACI International Trade Database. The OEC site is simply a convenient place to access all this data.

For population and world geographical data, we used the Gapminder data.

Fortunately, the OEC site has done a good job of cleaning the data, so we did not have to. The challenge we faced is the sheer size of the dataset. The data represents every combination of the 251 countries for each of the 1000 products tracked for every year from 1960 to 2010. With a few billion rows of data, we considered approaches to reduce the data to make it more manageable. Some of our considerations revolved around removing one of the degrees of freedom in the data by narrowing it down to only one country or year or product. The downside of this solution is that the resulting dataset will be narrow and uninteresting. Instead, we compromised between reducing data and maintaining interesting information by aggregating the 1000 products into 9 general categories. This reduced the size of the dataset by a factor of 100.

After the reduction we separated the data by country and by year so that data loading could be compartmentalized. Before this separation the visualization would take several seconds to load all of the data before displaying it. Now that it is possible to load only the specific data that is selected the visualization loads in around 100ms on average. By caching the specific data in memory, we were able to eliminate the load time for previously loaded data.

Initially, we planned to precompute per capita trade quantities to reduce computational overhead, but since the primary bottleneck is the size of the dataset we decided to compute per capita values on the fly.

Exploratory Data Analysis

One of the first things we did when considering what our visualization should look was look at other visualizations that have used similar data. We knew that trade data has been visualized before, but not for the specific purposes that we were interested in visualizing so we wanted to see what we could learn from preexisting visualization to see how we could improve our own.

A lot of what we found were visualizations where the area of a rectangle was used to represent the import and export data. We felt this was ineffective though because the numbers in trade were so large they didn't seem to provide as much information as a ranking system would. By visualizing rankings based on on a country's trade we could quickly determine their prominence in any category. We also felt that we could draw more interesting conclusions by using an actual map in our visualization in order to see relative distance between countries. We also felt that being able to see

the overall trade trends of countries over time would provide additional insight about the current year being analyzed by examining previous or future years of the data.

Design Evolution

Our design underwent a major directional change during the planning process. Initially we had not decided what the main focus of the visualization was going to be so we had some debate between comparing countries side by side or visualizing a single country's trade at a time. This debate was resolved once we synthesized our preliminary design ideas into a cohesive team of views: a trend chart, a world map, and a set of sorted bar charts.

Drafts

The following photographs are sketches of ideas for visualizations. Each design comes with a blurb explaining it.
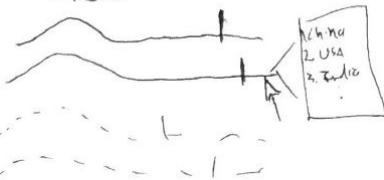
▽ USA

▽ Category of Goods

Exports

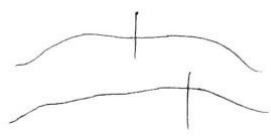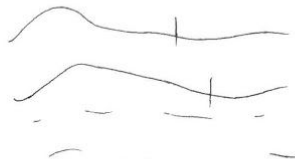Type: Sort
Food   Absolute Values   Per Capita
Machines
                                          China
                                        2 USA
      log   3. India
      Ind

Imports

Food

Machines

Largest Export Dest (for selected Good)   Largest Import Origin
                          $1000 $1000
1) India   1) India
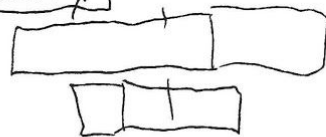2) China $900   2) China $900
...   ...
10)   10)
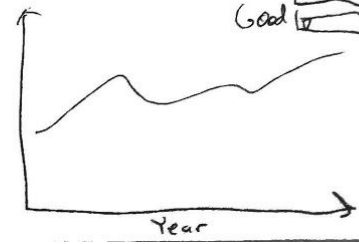
By Vol/ By Dependency/ By per Capita   Saturation

Compare vs ▽ Country   Dest ▽ All
Food   Good ▽

Machines

Year
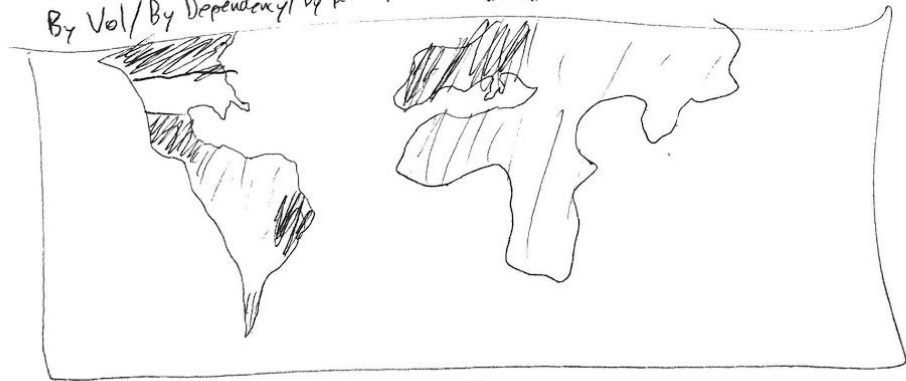
Here, the idea was to try to use the data to create a "spotlight" on any selected country. This gets away from the adversarial "trade war" take on international trade, and focuses on an individual country.

Going from top to bottom, we use a "distribution" chart that shows how much the selected country exports for each category of good. A vertical line would be used to indicate that. We would also include a shaded curve to show the distribution of the entire world for that product. That way, while we are focusing on the selected country, we can see that in the context of the rest of the world. We'd also separate this into two charts: one for absolute trade values, and one for per capita trade values. The idea is that certain countries (e.g. China, India), will probably be top on almost every good, simply because they are by far the most populous countries in the world. Showing per capita values will help deal with this problem. The user will be able to select what category of goods they want to look at, and clicking on any category will show immediate sub categories, if applicable. A similar chart will be done for imports.

Below that would be a simple list showing the largest export destinations and import origins for the country. For this, I chose simple text, just because the idea is to simply show which countries are the "most important" for the selected country. I considered something to visualize the magnitude of the trade amount (beyond the number), but decided that the world heat map would convey that information fine. Also, I want to keep away from a "winner/loser" framing of this data, so I opted to not show that.

The world map would be a "world heatmap". The idea is that your selected country would be highlighted, and the magnitude of your selected countries trade relations would be shown with saturation. The user could select how this heatmap is encoded: imports or exports, by volume, per capita, or by dependency.

I toyed with several ideas of how to show a trade deficit. These ideas are at the bottom of the above image, and all the ideas in the next image. The overarching implementation would be to have the user select a comparison country, and the comparison charts would update.

One idea (above) was to have horizontal bar charts for each category, and a dashed line down the middle that indicates balanced trade. If for that category, trade is balanced, they will meet exactly in the middle. If country A sends a lot more of the good to country B than vice versa, then the A side bar is large, and the dashed line goes through the country A bar. This way, the magnitude of the trade relation is encoded in the bar, and the dashed line/position of the bars encode the deficit.

The line chart (above) would be a simple chart that shows how the relation was over time. The user could use the drop down to show imports, exports, and deficits.

Below are just more ideas of how to encode this same idea. I tried to find a good chart that could show the relationship over time (e.g. the top right chart in the image below).

Goals: Compare flow of Goods between countries
Show trade balance / trade balance trends
Pro trade
Drill down on details

Broad Category, 1990
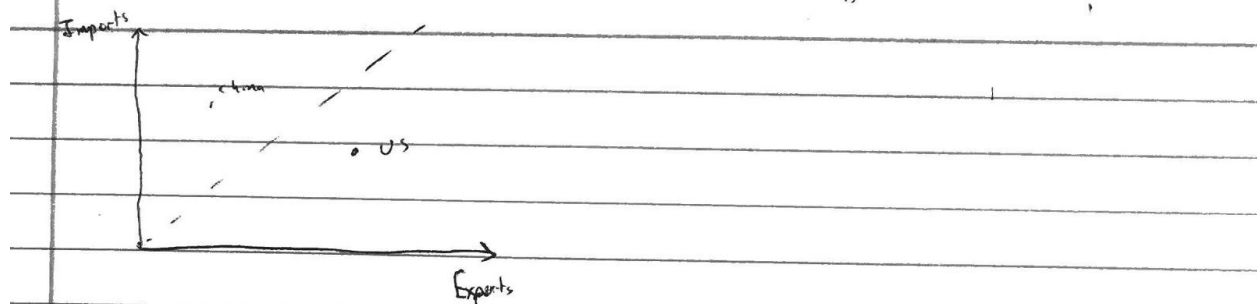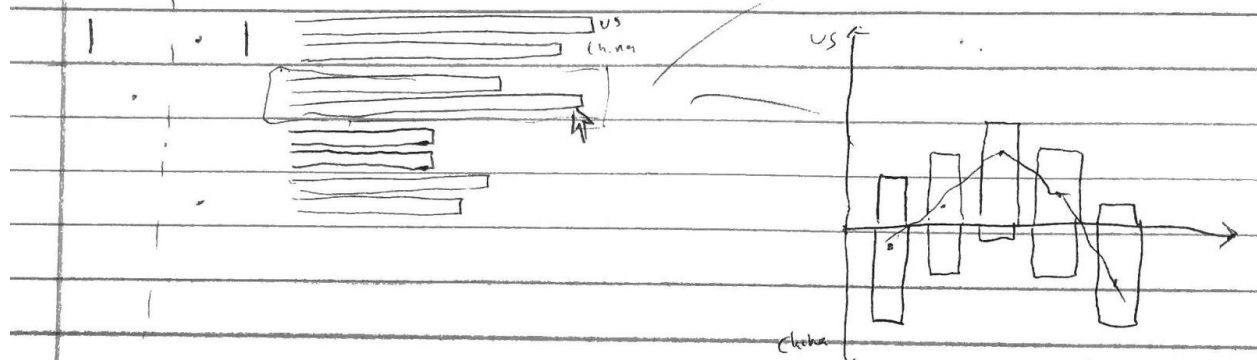
Countries   Years   SITC
                             Codes
Countries   Countries

$$251 \times 251 \times 56 \times 986 = 3,478,663,216$$
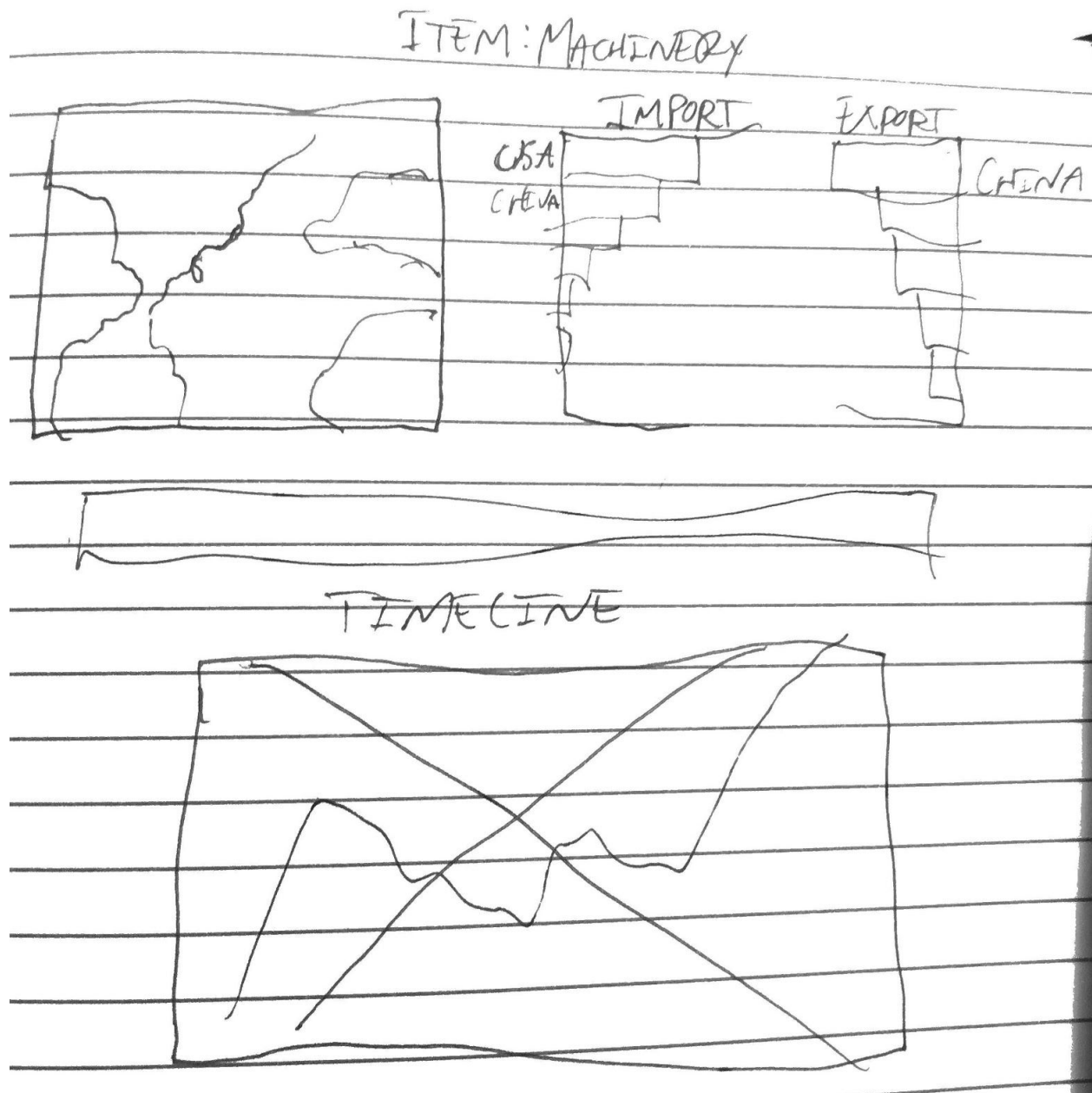
The design above is a visualization that focuses solely on one traded good. Based on the good selected the visualization would update. The heat map on the top left would show the magnitude of where the selected good is imported and exported from. The bar chart on the top right shows the rankings of countries who import and export the selected good. The bar chart would allow for easy comparison and the heat map can easily show location information of the products. If we were solely interested in where goods are originating and being exported to, this visualization would allow us to easily see those trends.

This is a very head to head country vs country visualization idea. It shows how two countries trade with each other and how they compare when trading with the world.
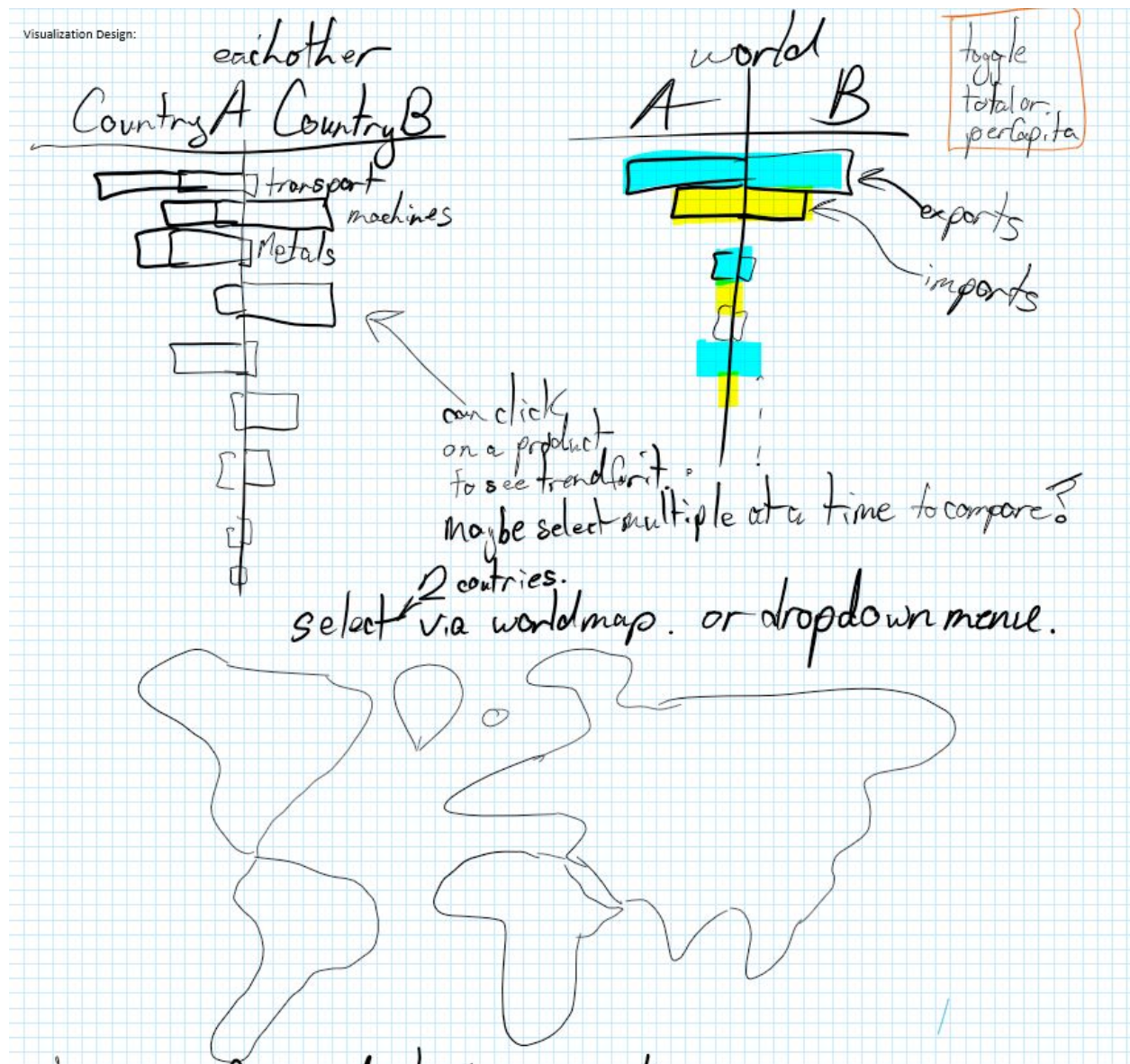
The trend line charts show how things have changed over time for specific products and for in general. This part also has both a country A with country B trading and their total trading with the world. The map is used for mostly easy selection but might also have a heatmap or something.
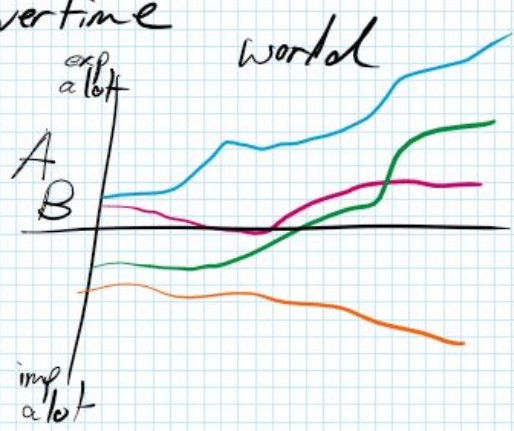
choose specific product to see overtime

A Aimpalot each other
B espalot
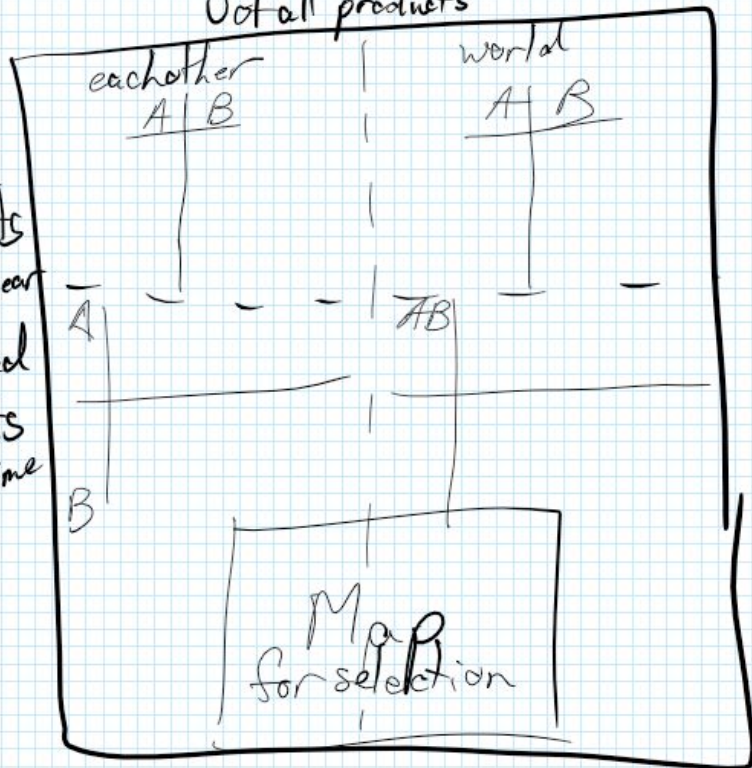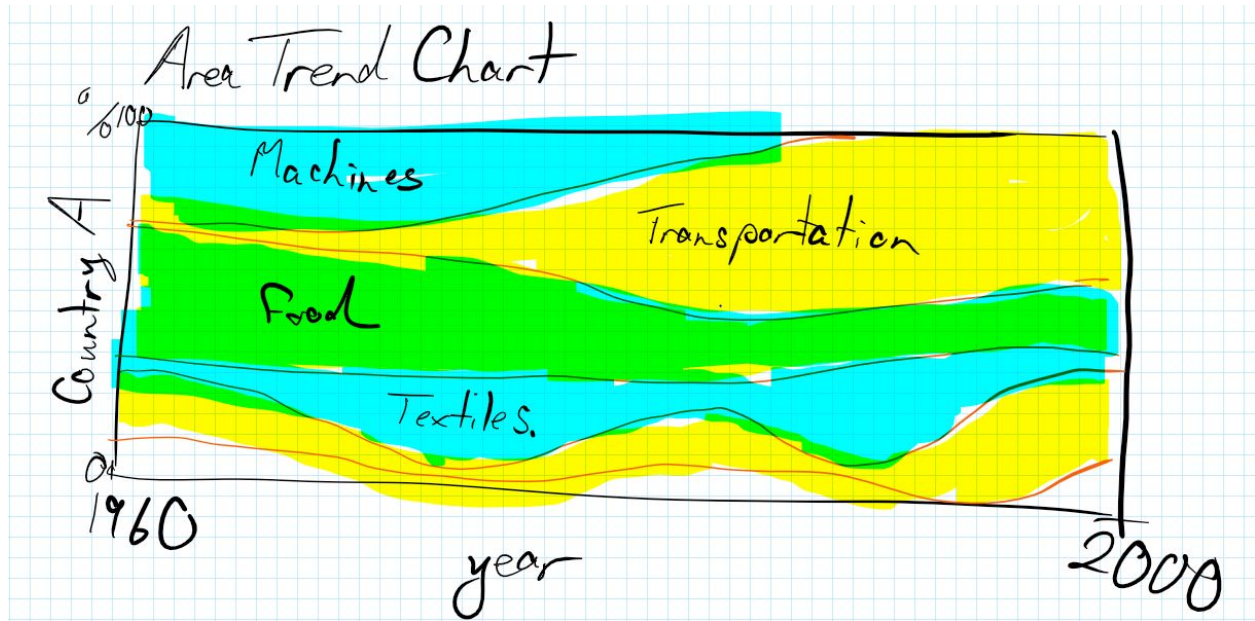
'1960   2000

exp
a lot

world

A
B

imp
a lot
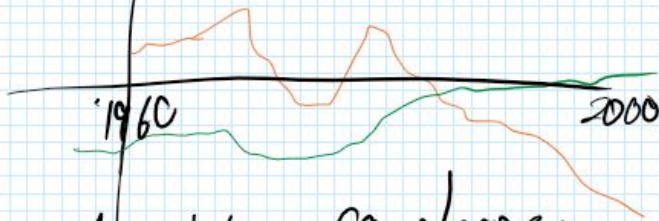
B Aexpalot
Bimpalot

Can choose a
year to show that
year breakdown
of all products

eachother
A | B

world
A | B

All
products
for
selected year

Selected
Products
over time

A

AB

B

MaP
for selection

Area Trend Chart

Country A — Machines, Food, Transportation, Textiles — year 1960 to 2000, %100 to 0

Synthesis of the design ideas

USA

Product — Exports — Absolute Value — Per Capita — Sort by rank p.c. or abs val

Transport, Food, Machines, Textiles, Cloth, Wool, Clothing

Imports — Similar to above

World Map Heatmap Showing Dependency or Volume or Per Capita with other Countries.

Imports, Deficit, Exports — 1960 to 2016
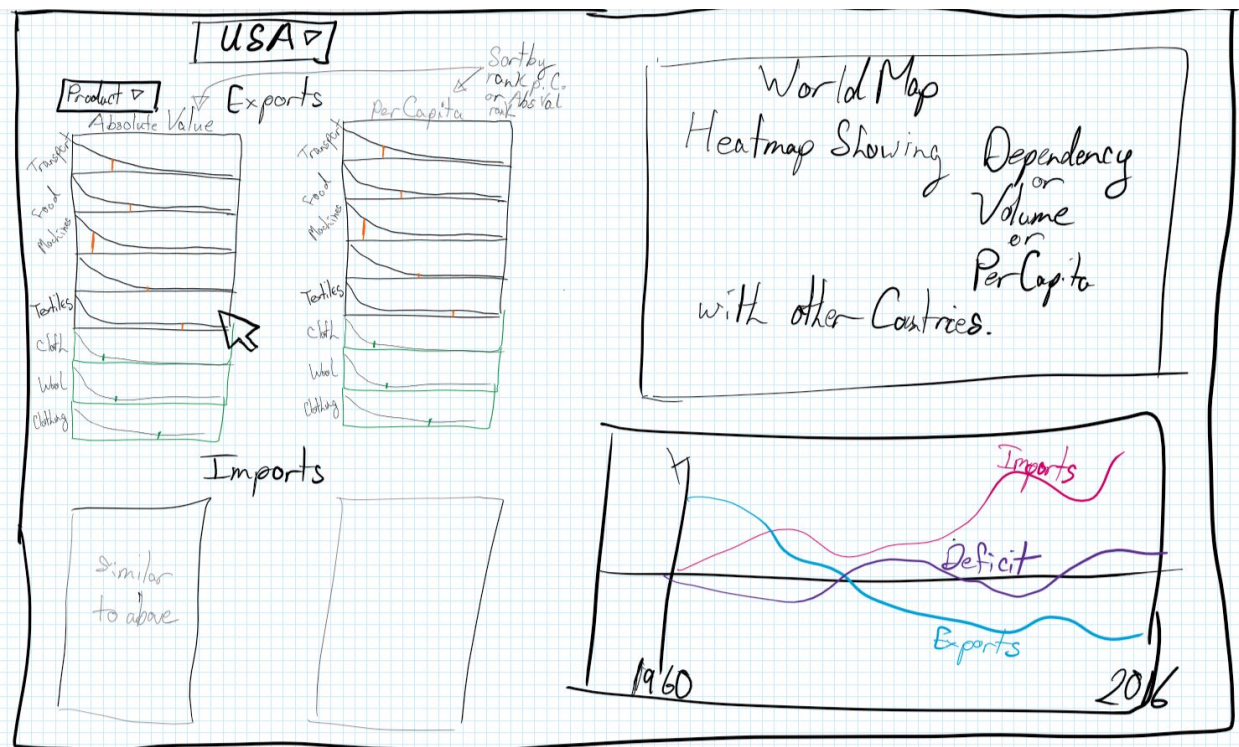
For now, we are planning on a simple design layout of 3 primary charts: a distribution chart (for both exports and imports), a trend line chart (as a comparison with other countries), and a world heatmap.

We plan on having multiple points of interactivity for the user. The primary interaction will be the country selection, which we anticipate will be a dropdown menu and potentially a search feature. We also want to allow selecting specific products for the distribution graph. This may potentially be limited, depending on how much data we have to aggregate, but we do want to be able to go down at least one "layer" of data, if possible.

The world heatmap will allow for the encoding of different metrics, such as absolute imports/exports, per capita imports/exports, deficits, and dependencies.

The trend chart will allow for either the selected country with the entire world (the default), or a comparison with a selected country.

Intermediate Implementation

The distribution chart uses two encodings: an area chart and a single vertical line. The area portion is simply each country ranked, with the x axis being ranks, with the last place country on the left and the #1 position on the right. The y axis encodes the actual values (e.g. food exports). The line represents the position of the selected country in the distribution. The colors used at this point in the implementation were simply placeholders.
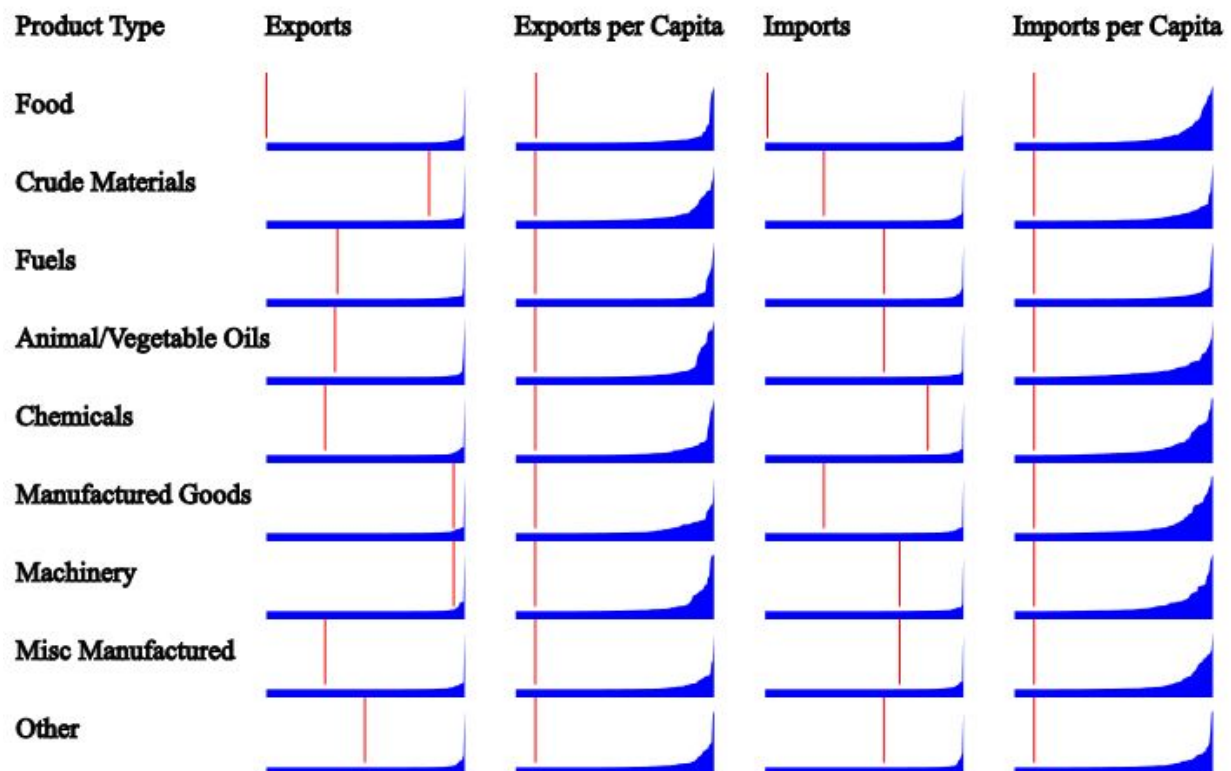
The world heat map uses a gradient going from black to red with the most trade mapping to red and the least trade mapping to black. In order to better represent the differences between countries with a 0 value and countries with a small trade value the 0 values are drawn in white so they blend in with the background. This makes it very clear that for the selected product and year, Egypt trades with Australia, Norway, and Brazil but not with Russia or Argentina.
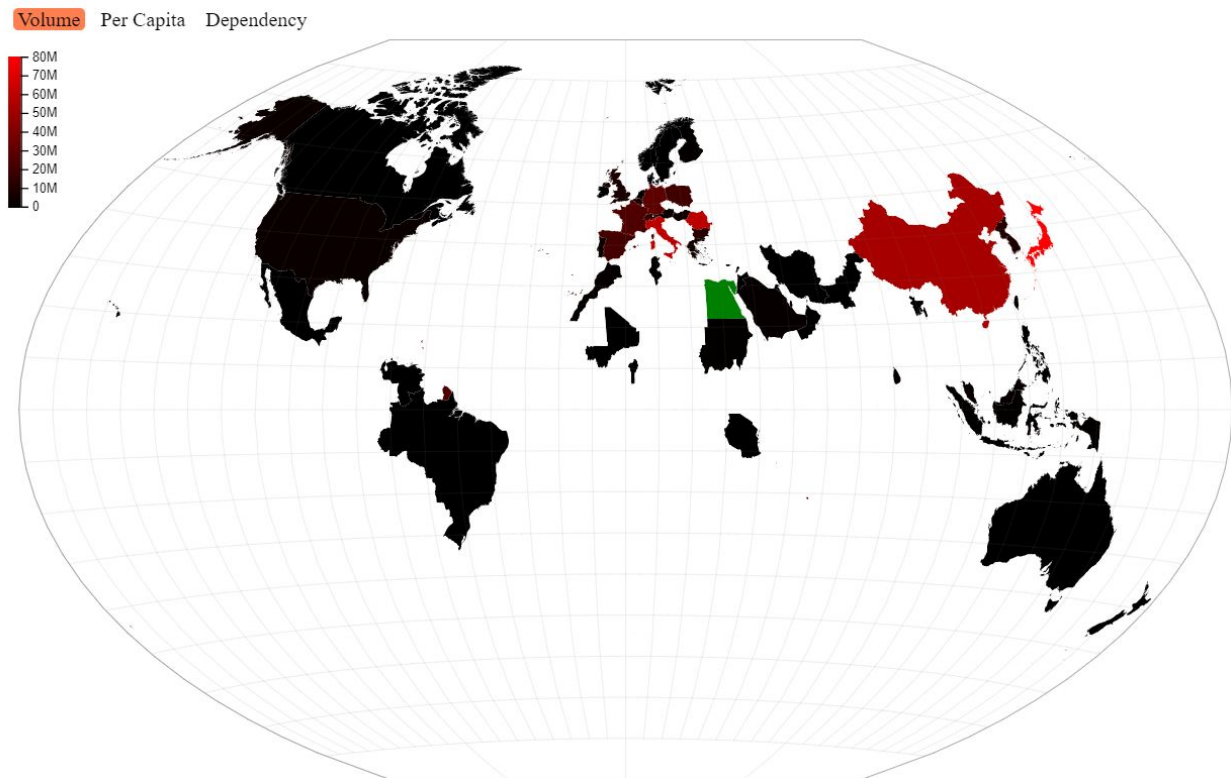
After switching to the "Per Capita" view you can see that the amount of exports to the USA is actually small relative to its size while the exports to Italy and Romania are much higher per capita.
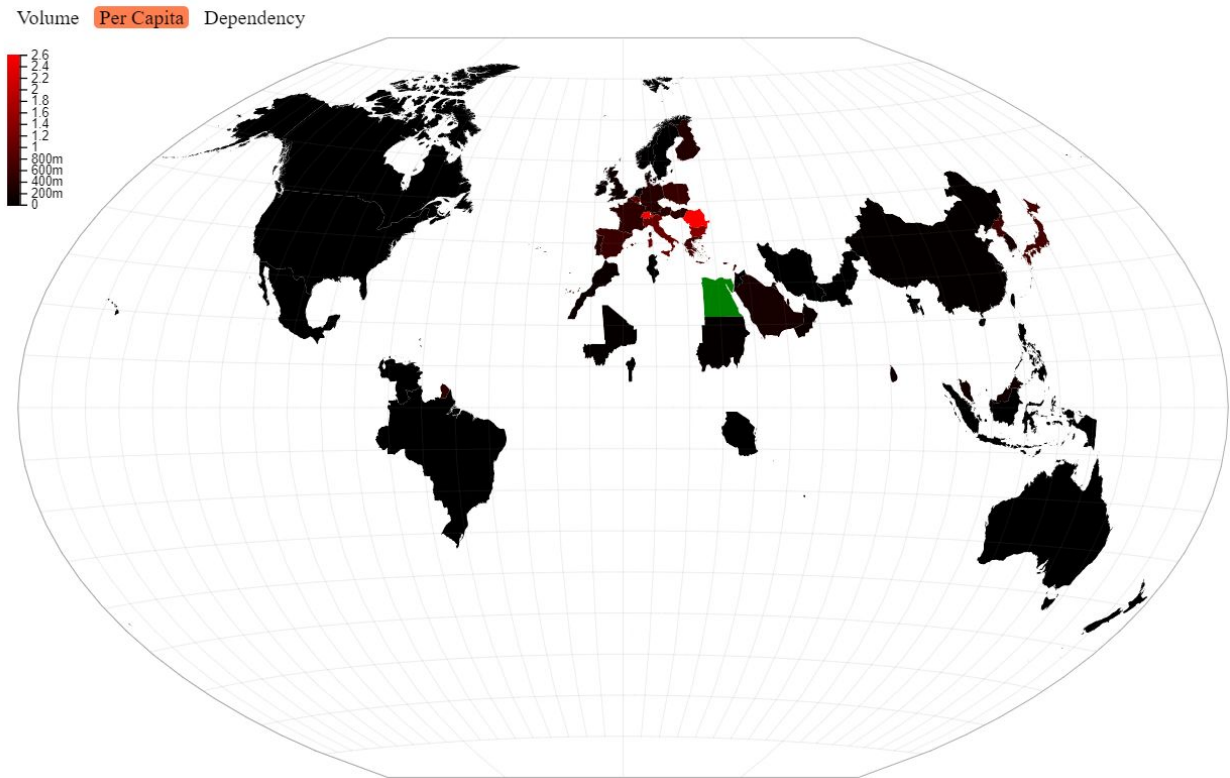
The trend chart uses a scatter plot to show the import and export trends of a country over time. The red circles show the export data and the blue circles show import data. The x-axis shows time in years while the y-axis shows the value of the importing and exporting data. For most countries we can see a clear correlation between imports and exports and visualize the deficit between the two over time.
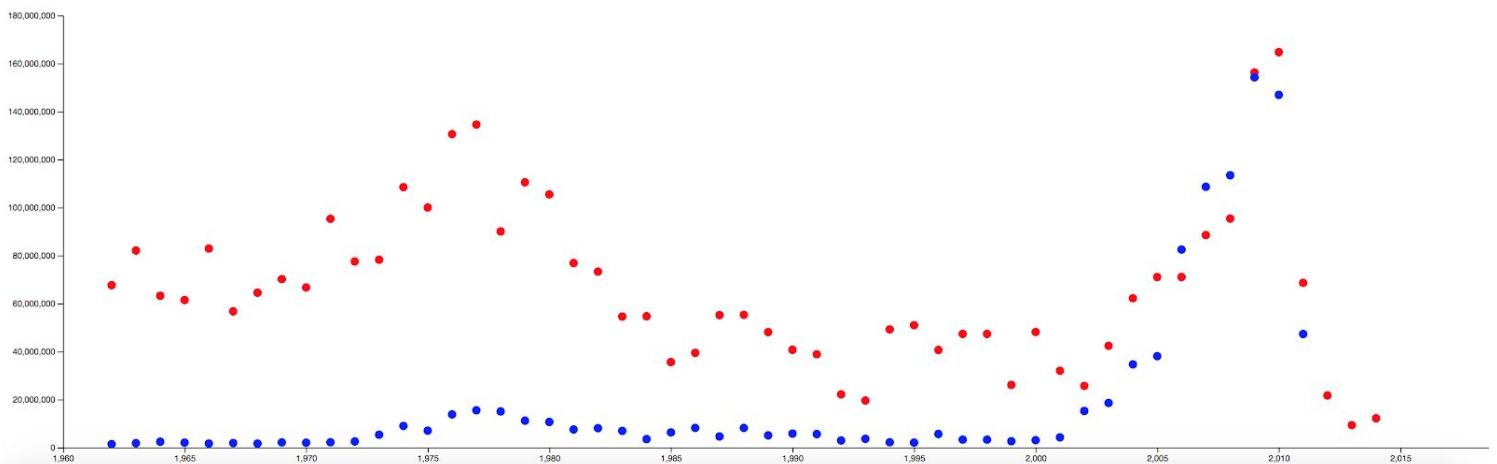
Final Implementation

After the first milestone of the project we improved the interaction by allowing selection of year, product, and country using the three different views. Clicking on one of the rows in the set of bar charts makes the world chart and trend chart display data for only that product. Likewise, selecting a year on the trend chart make the bar charts and world chart display data for only that year. This method of filtering is easy for users to learn and utilize.

The country selection drop down menu was a crucial addition which enables users to select the tiniest of countries which are otherwise too small to click on the world map. (Pro Challenge: try selecting Singapore using the world map.)
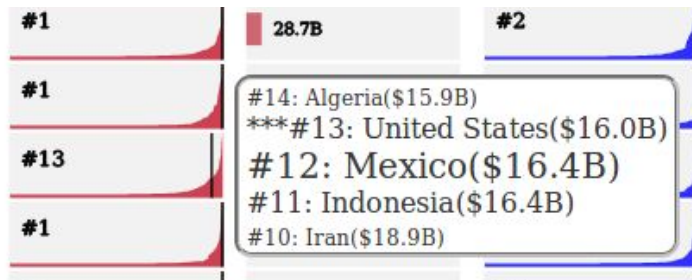
Country:
United States

We realized that we needed encoding for the values for each product other than the distribution chart, so that the users can see what products a country is exporting/importing relative to other products. To do this, we implemented bar charts adjacent to the distribution charts that encode the values for the selected country.

Product Type	Export Ranks	Exports	Import Ranks	Imports
Machinery	#1	379.4B	#1	491.2B
Misc Manufactured	#2	89.3B	#1	183.0B
Chemicals	#1	87.4B	#1	76.9B
Manufactured Goods	#1	73.7B	#1	125.2B
Crude Materials	#1	28.7B	#2	22.6B
Other	#1	23.8B	#2	47.9B
Fuels	#13	16.0B	#1	135.9B
Food	#1	15.2B	#1	14.3B
Animal/Vegetable Oils	#1	12.4B	#2	3.9B

For added context, text labels were used: the number of the country's rank for the distribution charts, and the rounded value (whether volume or per capita) on the bar charts.

We have tooltips for the distribution charts and the bar charts. The distribution chart shows the ranks of countries in the distribution, along with how much they import/export of the product. The tooltip changes based on the x position of the mouse, showing the 5 countries at that portion of the chart. The selected country has prepended three asterisks (***). The tooltip for the bar charts shows the

top 5 countries that the selected country exports to/imports from, along with the associated values.


We also added a sorting functionality. When the user clicks on the headers in the first column (export ranks, exports, import ranks, imports), the rows are sorted by that column. For the ranks column, the rows are sorted by the selected countries rank. For the bars, they are sorted by value of the bar. Clicking on the header multiple times reverses the sorting order.


Furthermore, clicking on the product names in the left column causes the other charts in the visualization to filter their data based on the selected product.

For increased usability, we made untraded countries light gray instead of white. That way users can always see the big picture rather than an incomplete jigsaw puzzle of countries. Other additions include choosing between visualizing Exports and Imports. This allows analysis both outgoing and ingoing trade for countries. From these screenshots you can see another clear difference between viewing total volume of trade and trade normalized by population. Canada imports at a higher rate from Guyana per Guyana citizen than from the US per US citizen although the magnitude of imports from the US is much greater. When mousing over a country, a tooltip is displayed showing the exact quantity of the selected type of trade. It would be impossible for a human to determine exact amounts based off of the red-black gradient.

Volume Per Capita Exports Imports

Egypt
Imports: 24.0K
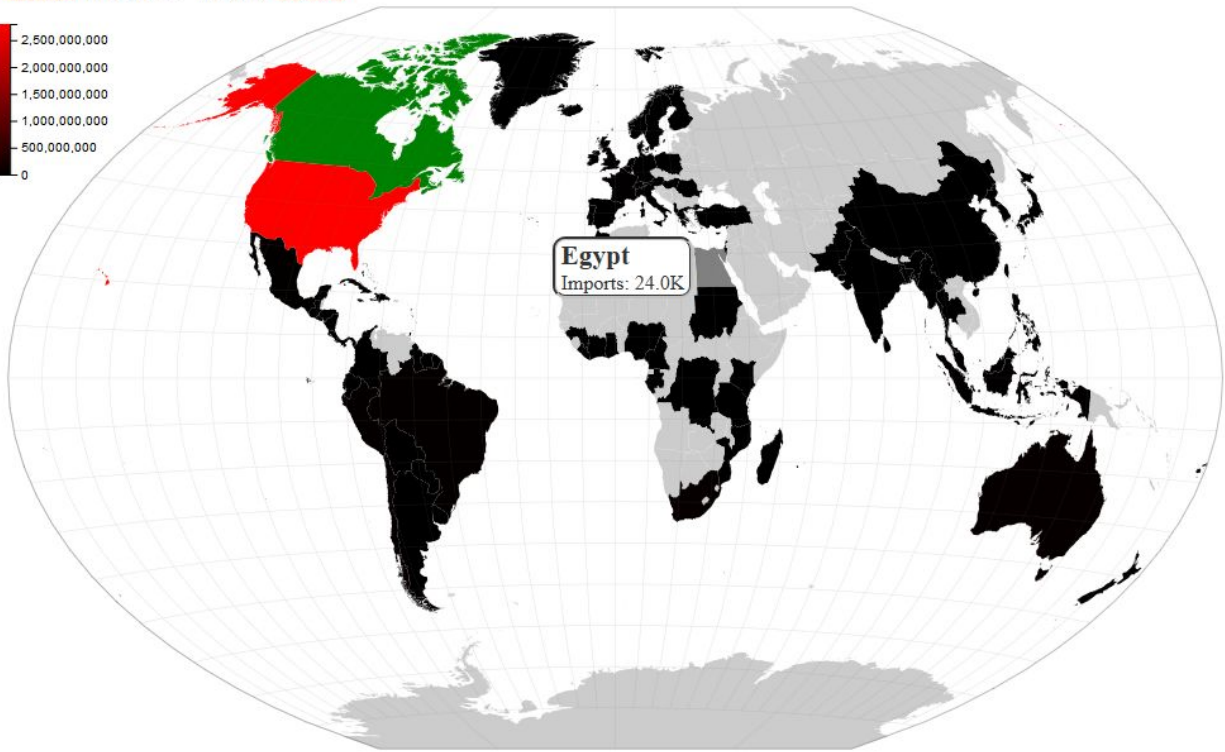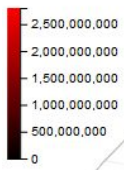
Volume Per Capita Exports Imports
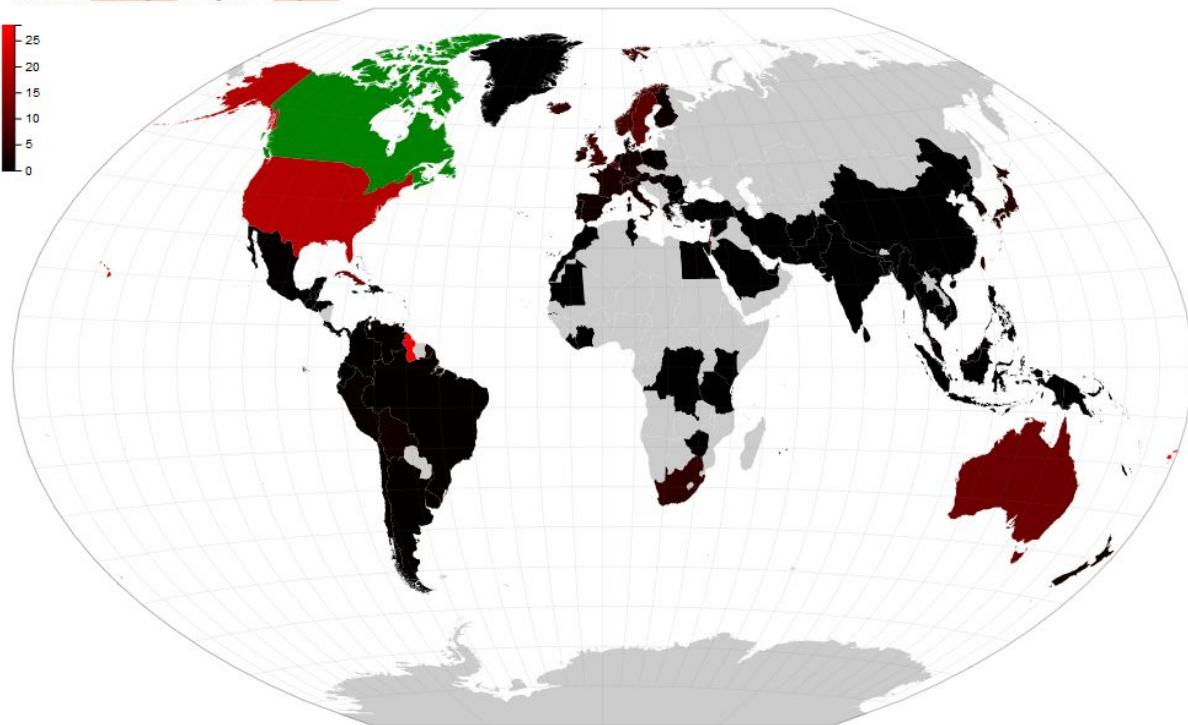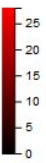
Many changes were made to the trend chart in order to make it more readable and interactive. Now users can select a year by clicking on a circle that will update the rest of the plots. The selected year circle is enlarged to show the users selection and when hovering over a circle a tooltip is displayed and an outline is created around the circle to show intractability. Additionally, there is now a display on the trend plot that shows the selected country, year, product type, and corresponding import and exports numbers. With these changes users can clearly decipher what information is being shown and make interesting conclusion more quickly without being confused about what they are looking at.
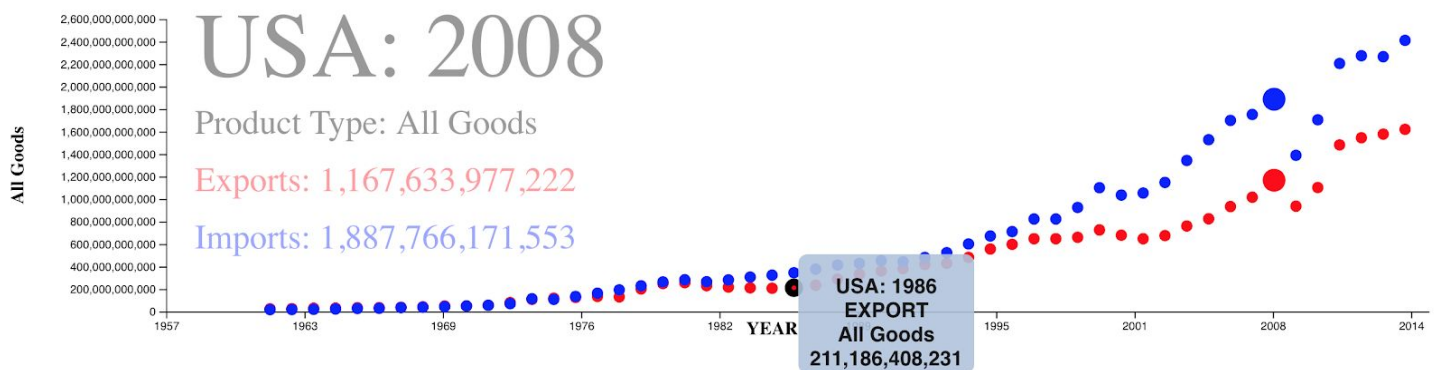
Evaluation

What did you learn about the data by using your visualizations? How did you answer your questions? How well does your visualization work, and how could you further improve it?

Before starting this project we thought countries were a lot more specialized in trade but after fiddling with the visualization you can see that although some countries like Egypt and Mexico export proportionally more fuels than other products, most countries dabble in a variety of product types both as exports and imports. Another realization was that certain product types are two way streets while others are more one directional. For example, machinery tends to have high import quantities and high export quantities for the same country while fuels is usually either highly exported or highly imported.

We feel that this visualization does a great job of zooming in on a single country. The user can understand how the country compares to the rest of the world, where it exports/imports the different products, and understand over time how its exports/imports have changed. One thing that our visualization doesn't do well is compare to other countries. The user can see how the selected country fits in the world context, but they can't see how two countries, or a group of countries, relate to each other.

There are a couple of ways that we could further improve this visualization in the future. One thing is to add functionality to compare two or more countries to each other in terms of a selected metrix

(e.g. exports, volume, products). We considered adding an "adjacency matrix" where the user could select multiple countries, and each square in the matrix would be encoded by the metric selected.

We also wanted to try to format the data in such a way that we could include finer categories (there were about 1000 product codes, we just used the top level categories). However, a naive filtering of the data led to about 1.2 GB of data total, so we opted not to do that. Still, there may be a good way to format the data to make this all possible. This would allow even more detailed exploration by the user.