Visualization for Data Science Project Proposal

- ## Basic Info.
  - The project title, your names, e-mail addresses, UIDs, a link to the project repository.
    - Trade Wars
    - Jake Osterloh; jake.osterloh@gmail.com; u0645881
    - Avram Twitchell; avram.twitch@gmail.com; u1193523
    - Oleks Korshak; alexokorshak@gmail.com; u1210377
    - https://github.com/strikerz76/dataviscourse-pr-trade-wars

- ## Background and Motivation.
  - Discuss your motivations and reasons for choosing this project, especially any background or research interests that may have influenced your decision.

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.

- ## Project Objectives.
  - Provide the primary questions you are trying to answer with your visualization. What would you like to learn and accomplish? List the benefits.

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.**
  - From where and how are you collecting your data? If appropriate, provide a link to your data sources.

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 data, we plan on using the Gapminder data from homework 4.

- **Data Processing.**
  - Do you expect to do substantial data cleanup? What quantities do you plan to derive from your data? How will data processing be implemented?

Fortunately, the OEC site has already done much of the data cleaning, so there really isn’t much cleaning that we need to do.

The sheer size of the dataset poses a problem for us. There are a few billion rows in the data. This makes sense, since the data roughly represents every combination of the 251 countries, for every of the 1000 products tracked, for every year since the 1960s.

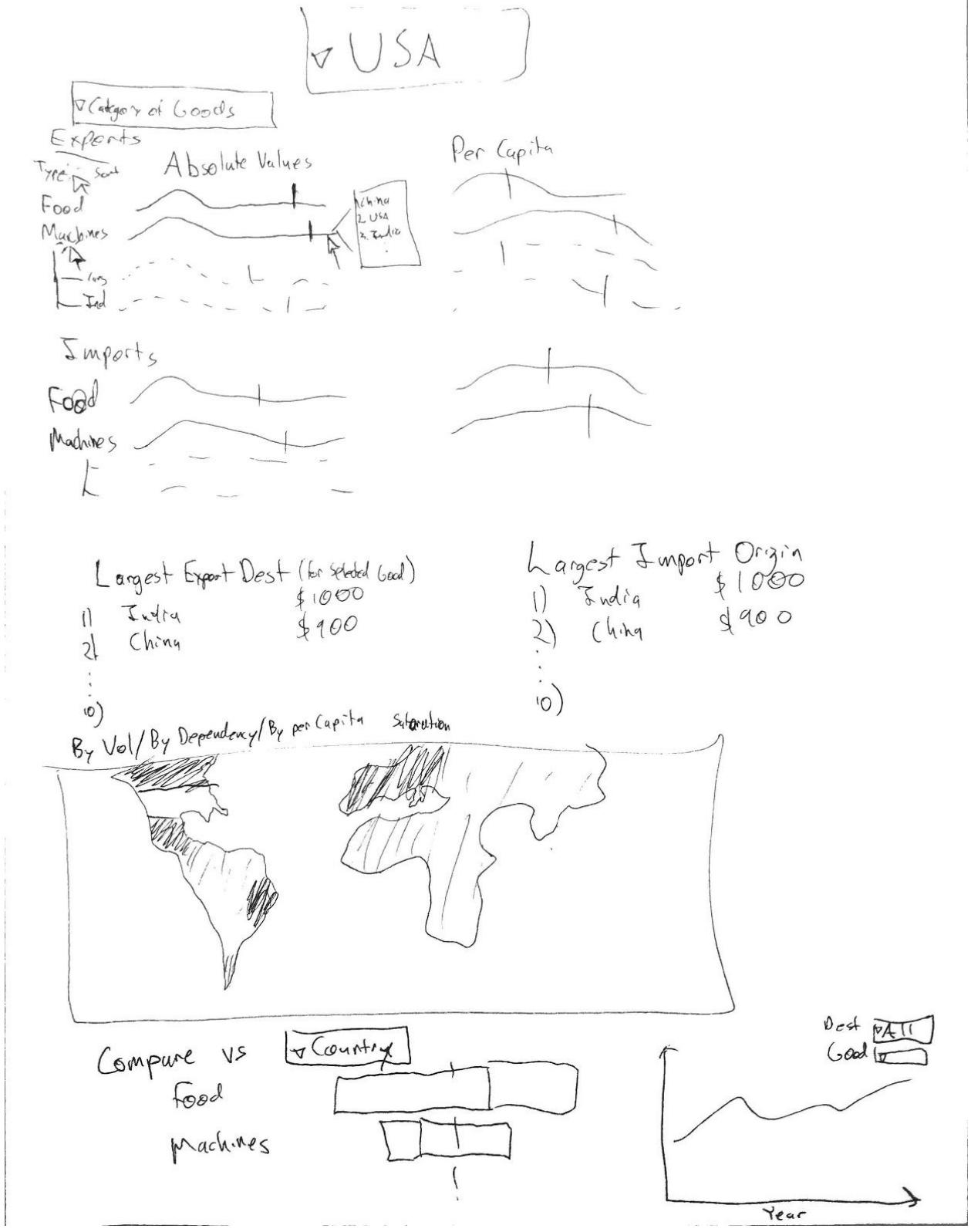
We are considering different ways to reduce this data to make it more manageable. We could focus on a single country’s exports/imports, we could focus on a single year, or we could focus on just a few products. For now, we anticipate aggregating the product data into a few broad categories. This means a loss of detail at the product level, but allows for a fuller exploration of each country.

We do plan on integrating population data in order to calculate per capita trade quantities. This will be a simple, one time calculation of dividing trade values by the population of the country.

Furthermore, we plan on deriving a few values from the data itself. These include percentages of exports (e.g. how much of country A’s food exports go to country B?), and similar measures for imports. This again we plan on calculating before hand for each observation.

Dependency: Attempt to show how dependent a country is on other countries for trade (who they rely on import, who they rely on to buy their exports)

- ## Visualization Design.
    - How will you display your data? Provide some general ideas that you have for the visualization design. Develop **three alternative prototype designs for your visualization**. Create **one final design that incorporates the best of your three designs**. Describe your designs and justify your choices of visual encodings. We recommend you use the Five Design Sheet Methodology.


USA
Category of Goods
Exports
Absolute Values
Per Capita
Type
Sort
Food
Machines
1. China
2. USA
3. India
Imports
Food
Machines
Largest Export Dest (for selected Good)
1) India $1000
2) China $900
10)
Largest Import Origin
1) India $1000
2) China $900
10)
By Vol/By Dependency/By per Capita Saturation
Compare vs
Country
Food
Machines
Dest ALL
Good
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.

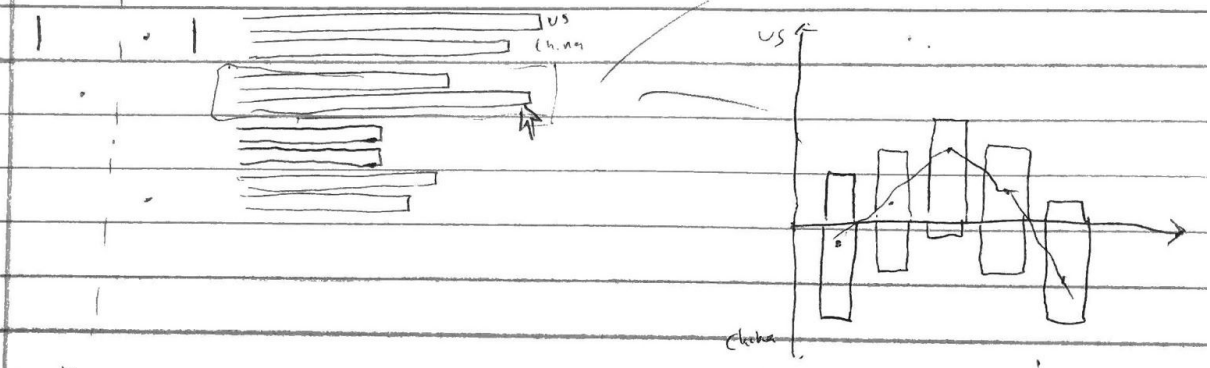
One idea (above) was to have horizontal bar charts for each category, and a dashed line down the middle that indicates balanced trade. If for that category, trade is balanced, they will meet exactly in the middle. If country A sends a lot more of the good to country B than vice versa, then the A side 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.

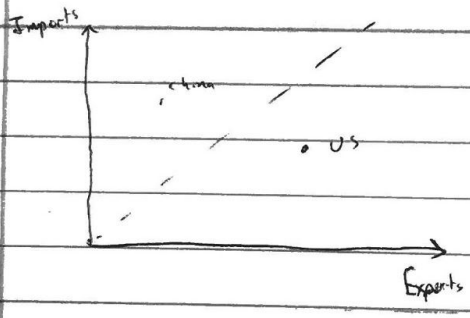
The line chart (above) would be a simple chart that shows how the relation was over time. The user could use the drop down to show imports, exports, and deficits.

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

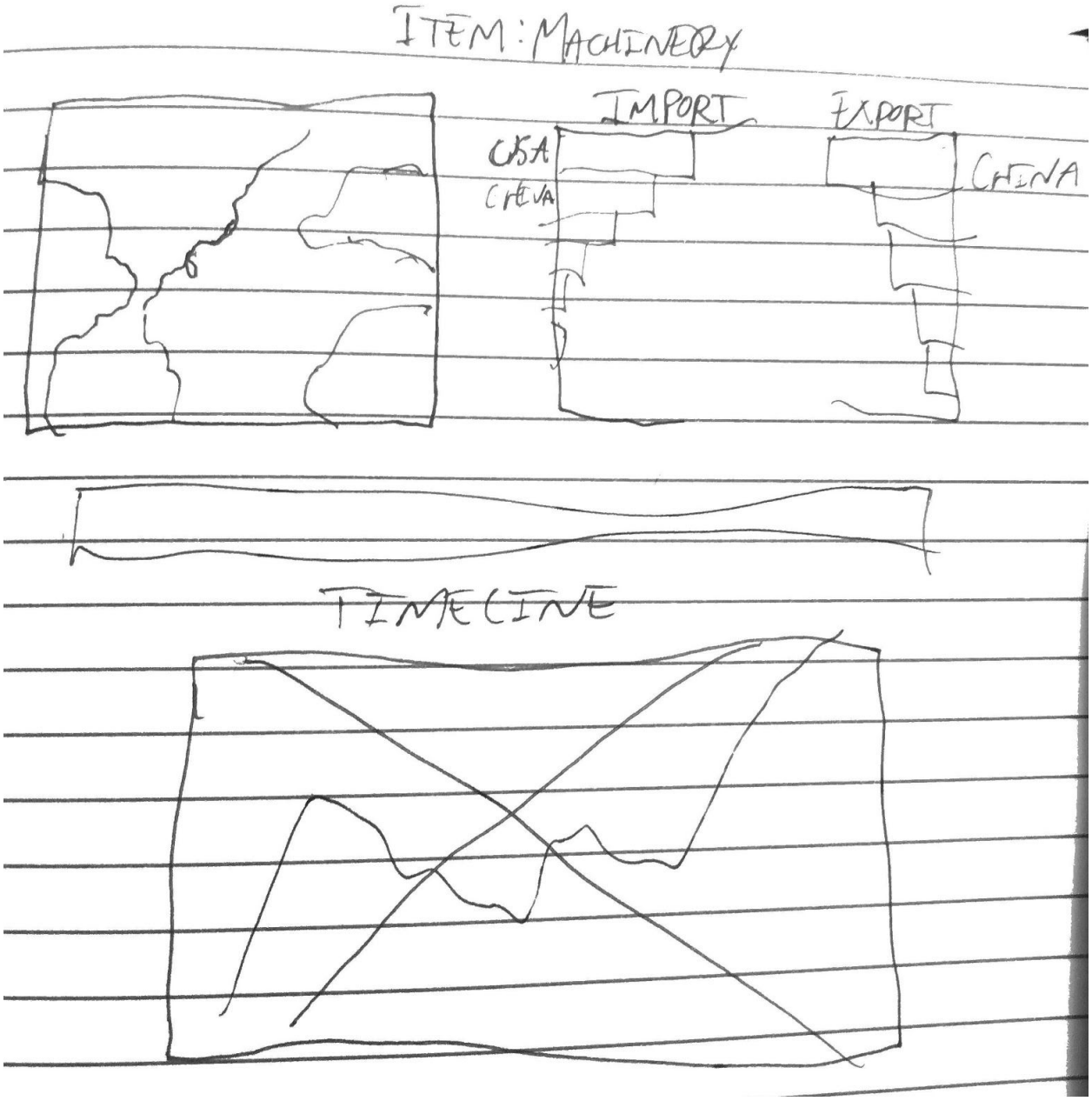
Broad Category, 1990





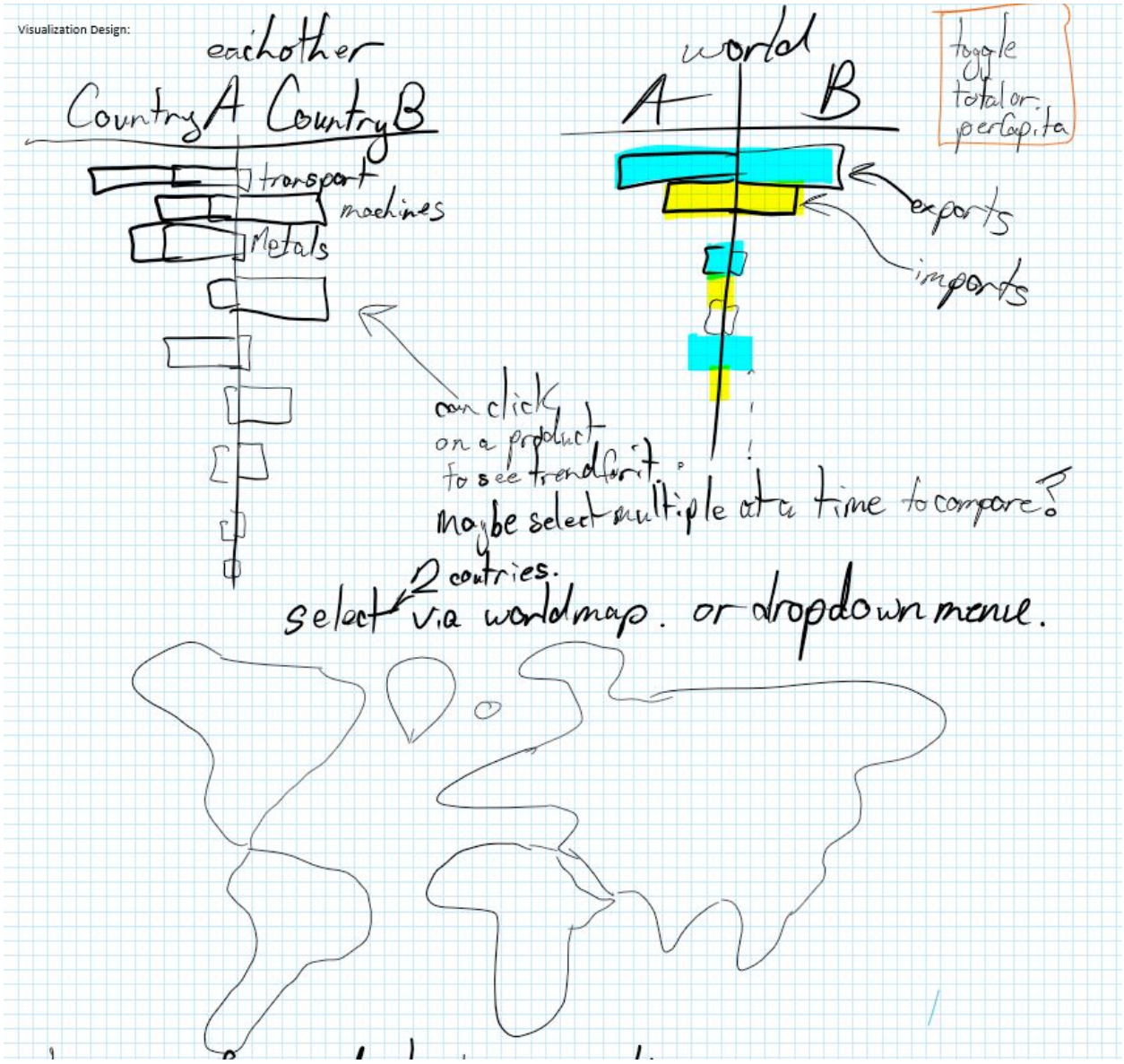
Countries ↓ countries Years SITC Codes

251 x 251 x 56 x 986 = 3,478,663,216

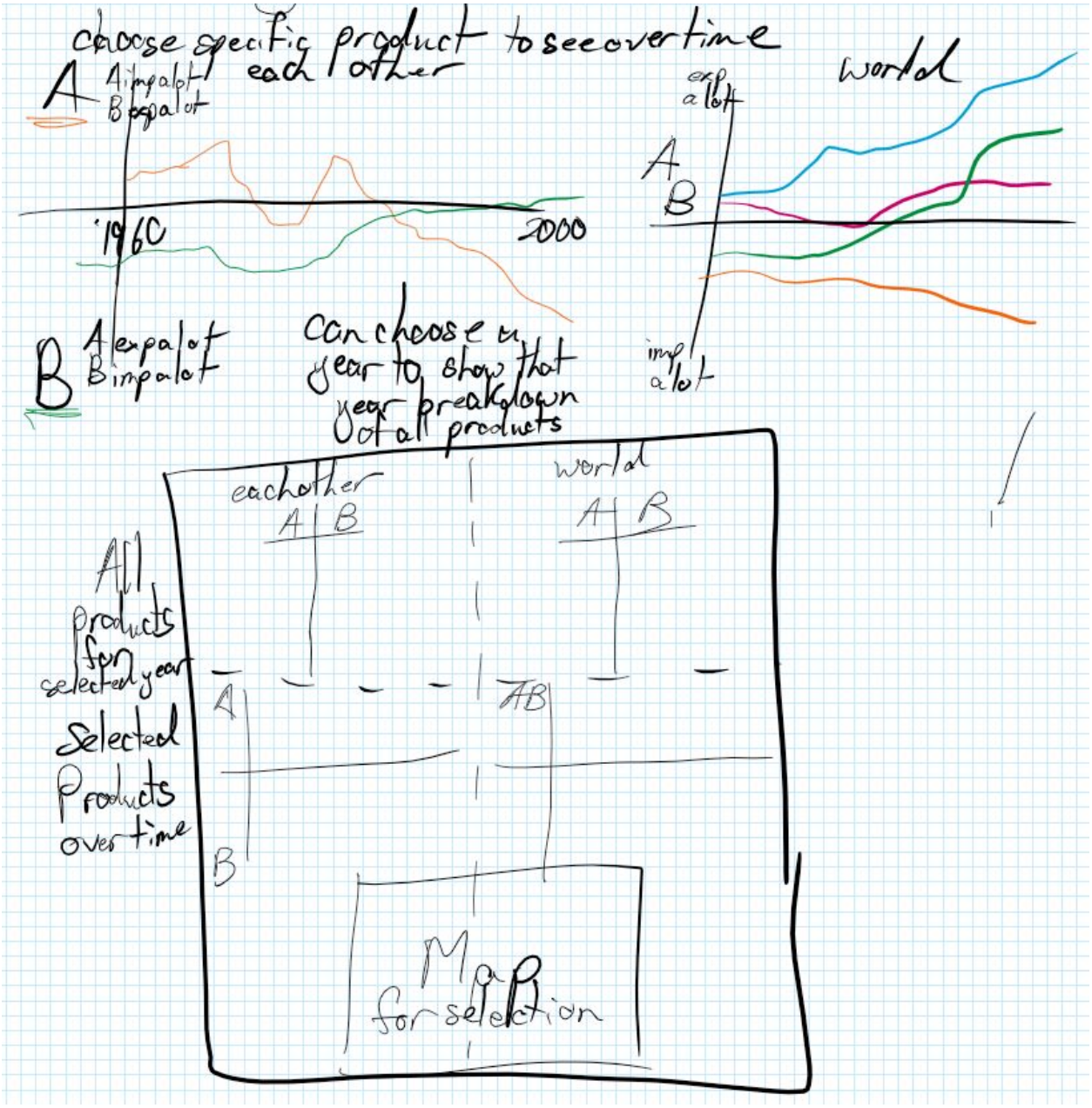


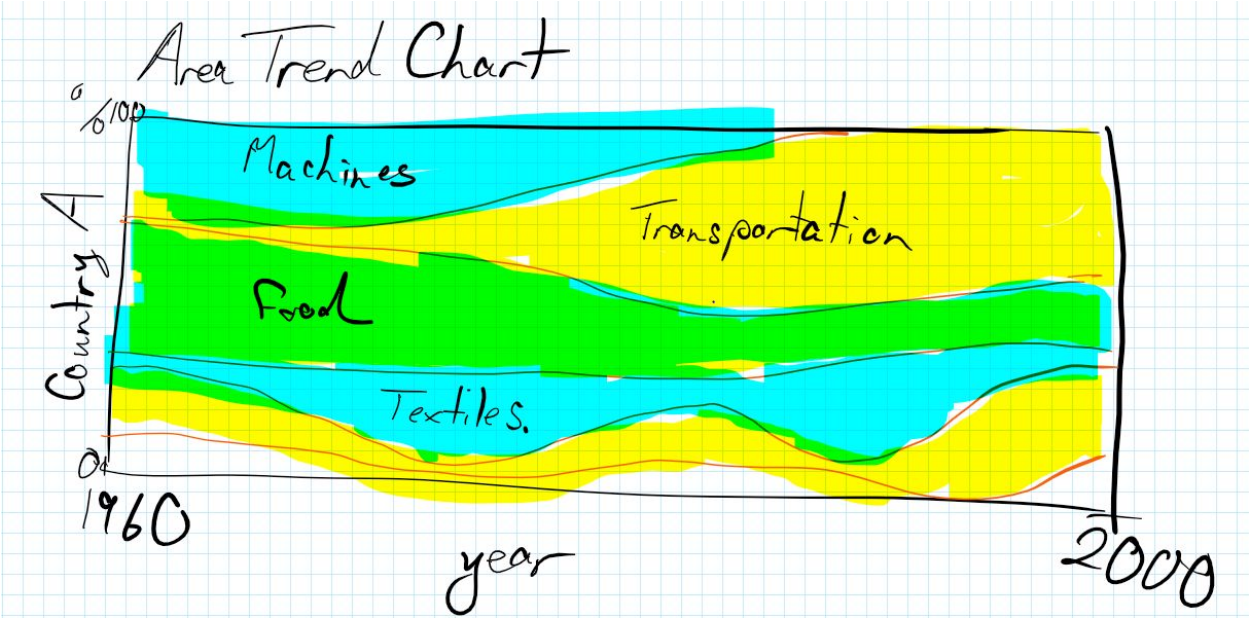
The design above is a visualization that focuses solely on one traded good. Based on the good selected the visualization would update. The heat map on the top left would show the magnitude of where the selected good is imported and exported from. The bar chart on the top right shows the rankings of countries 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.

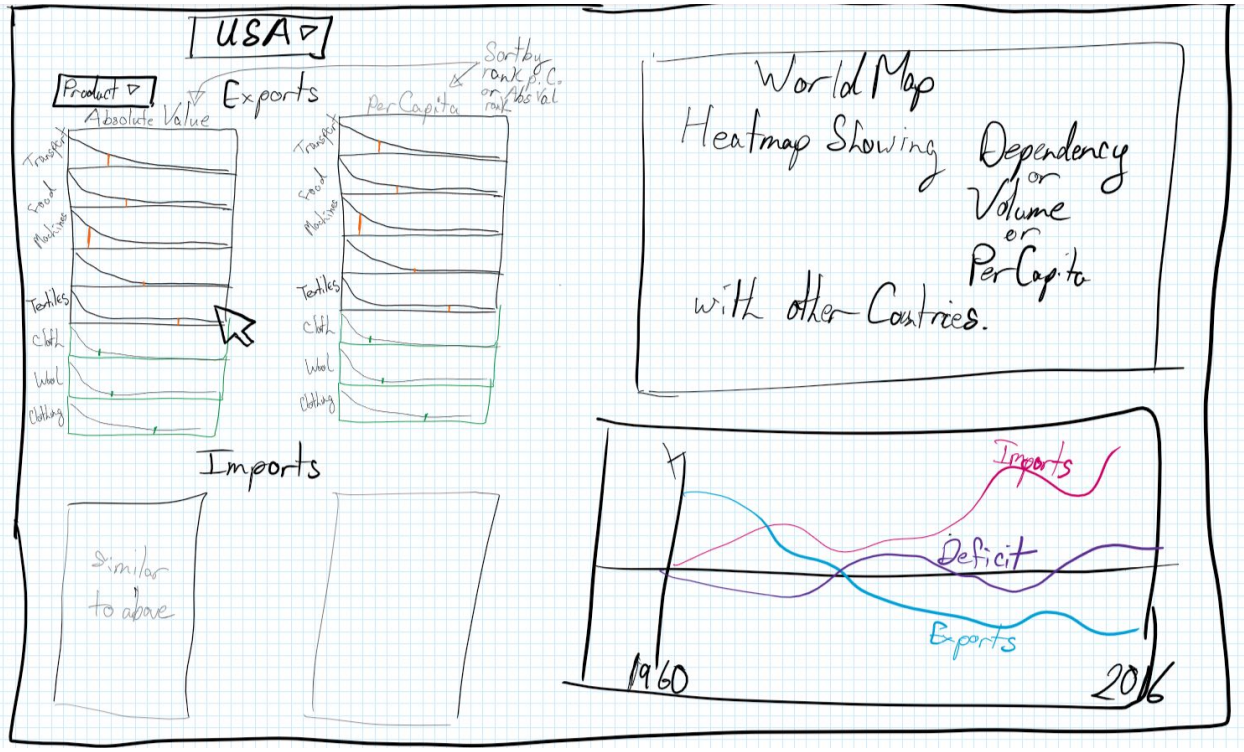


The trend line charts show how things have changed over time for specific products and for in general. This part also has both a country A with country B trading and their total trading with the world. The map is used for mostly easy selection but might also have a heatmap or something.


choose specific product to see over time
each other
A
A imp a lot
B exp a lot
1960
2000
B
A exp a lot
B imp a lot
exp a lot
imp a lot
A
B
world
Can choose a year to show that year breakdown of all products
each other
A B
world
A B
All products for selected year
Selected Products over time
A
B
AB
Map for selection



Synthesis of our different design ideas.



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.

- **Must-Have Features.**
  - List the features without which you would consider your project to be a failure.
    - Select a country, get a breakdown
    - Have some trends shown over time
    - Top exports/imports of a country

- **Optional Features.**
  - List the features which you consider to be nice to have, but not critical.
    - World heat map by product
    - Place in distribution chart
    - A comparison (head to head) with another country
    - A “storytelling” portion where we highlight interesting historical trade properties
    - Timeline of Imports/Exports/Deficit with two countries
    - Be able to select specific products and see info on subproducts

Consideration: for the world map, how to avoid losing small countries to oclusion or because they are too small to see clearly on the map.

- **Project Schedule.**
    - Make sure that you plan your work so that you can avoid a big rush right before the final project deadline, and delegate different modules and responsibilities among your team members. Write this in terms of weekly deadlines.

Meeting on October 17:
- Brainstorming about project idea and dataset.
- Decided on dataset and assigned homework to come up with a few different visualizations.

Meeting on October 22:
- Shared visualization designs for dataset.
- Discussed possibilities and decided on an initial design.
- Assigned tasks for project proposal.

Week Oct 21 - 27

Proposal due Friday October 26

Week Oct 28 - Nov 3

Oleks is gone November 1 to November 5
Pre-process and filter data (Avram)
World Map -- Create Framework (Oleks)
Import/Export Distributions -- Create Framework(Avram)
Trend chart -- Create Framework(Jake)
“Story” -- Research interesting trade trends to showcase with visualization (All)

Week Nov 4 - Nov 10

Oleks is gone November 1 to November 5
World Map -- Implement Heatmap, bind with data (Oleks)
Import/Export Distributions -- Create simple line chart, create distributions(Avram)
Trend chart -- Implement lines, make comparison country selectable(Jake)
“Story” -- Create static visualizations that show interesting relationships. (ALL)
Project milestone due November 9

Week Nov 11 - Nov 17

Link all charts to Country selector (TBD, likely a group effort)
World Map -- Tooltips, implement multiple metrics (Oleks)
Import/Export Distributions -- Tooltips, enable clicking to show sub-products(Avram)
Trend chart -- Create ways to focus or zoom in on trends (Jake)
“Story” -- (potentially) find way to have story progress by scrolling (All)

Week Nov 18 - Nov 24

Oleks is gone November 22 to November 25 and November 29 to December 3
- Can work remotely November 22 to November 25.

Smooth the edges -- Clean up code, prettify charts, touch up stylings, make dynamic (All)

Week Nov 25 - Dec 1

Oleks is gone November 22 to November 25 and November 29 to December 3

- Can work remotely November 22 to November 25.

Do screencast, finishing touches
Final project due November 30