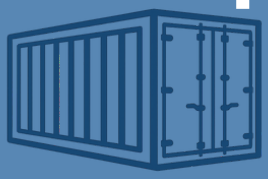
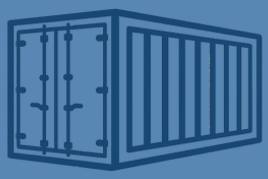
EPFL | COM-480 | 2025
Data Visualization

Acquiring Intuition on Global Trade

Team37

Huiyun Zhu

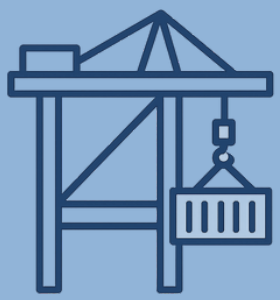
355235

Konstantinos Prasopoulos

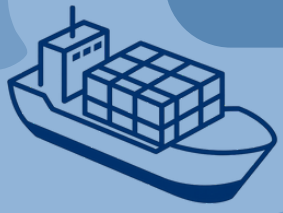
285813

Yuchen Qian

322420



Visit the project's website



Overview

Global goods trade has been in the headlines over the past few months, with many arguing that the world is transitioning into a new era that is yet to be determined. We find that the public discourse on the matter is quite abstract and aggregate, focusing on aggregate metrics and somewhat ideologically-driven interpretations of how global trade should look like.

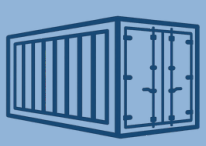
Without being experts on the subject, we nevertheless wanted to acquire and offer some intuition on how trade operates “under the hood”. To this end, we crunched publically available trade data that spans the last three decades to craft intuitive visualizations tied to real-world events and promote further exploration.

Motivation

Over the last decade, and especially in 2025, the collective brain of humanity has turned its attention to the difference between where goods like food and cars are made and where they are consumed.

Motivated by the desire to reduce transportation emissions or by national security concerns, humanity appears to be peeking under the hood of the global trade system, with the subject debated everywhere from government circles to taxi rides.

We wanted to better understand the subject and to encourage others to do the same. Specifically, we wanted to move from aggregate statements such as "country X has a trade deficit of Y billion dollars" and closer to statements such as "country X is not food independent" or "country X stopped exporting product Y because of event Z". In other words, we want to offer intuition as to what sort of events affect global trade. Equally importantly, we want to disaggregate the top line numbers as we believe that specific categories like food and energy are of higher importance.



Dataset

Dataset: CEPII BACI Bilateral Trade flows

CEPII is a French research center that specializes in the world economy. They publish the BACI Bilateral Trade Flows dataset, which is a cleaner, reconciled, and more readily available version of the UN Comtrade database. This dataset contains the value of all bilateral imports and exports between countries and for thousands of product categories. The dataset does not include data on the import and export of services.

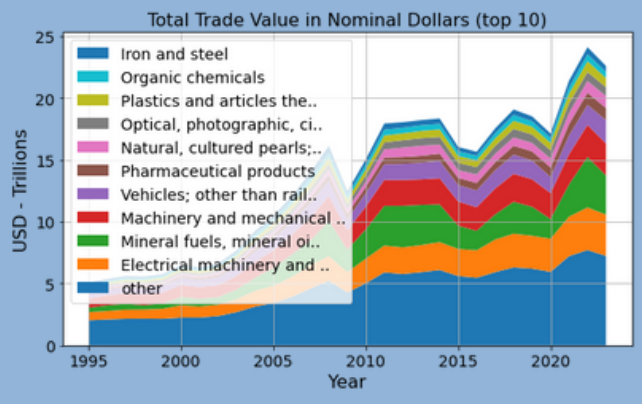
We are focusing on data following the "Harmonized System 1992 (HS 1992)" classification, which allows us to explore trade data between 1995 and 2023 to observe long-term trends.

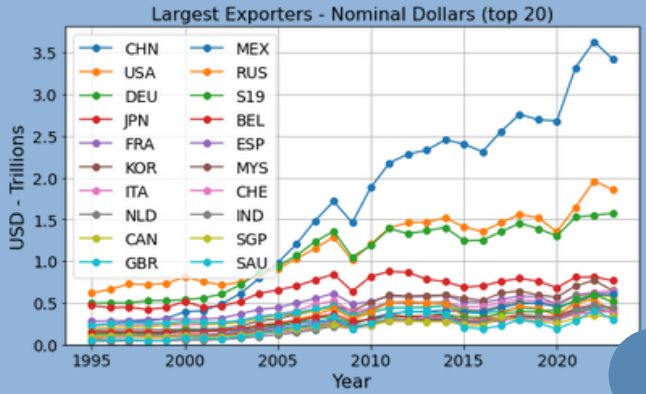
The dataset is ~8GB in size and has the following features for each year:

- Product Code (~5000 different categories)
- Exporting Country (238 Countries)
- Importing Country
- Yearly Value in Nominal US Dollars (thousands)
- Yearly Weight in metric tons.

Data Insights

- Total export tonnage increased from 7B Tons in 1995 to 15B in 2023
- The three largest exporters are: China, the US, and Germany, totaling ~30% of total exports. The three largest importers are: the US, China, and Germany, totaling again ~30% of total imports.
- China exports almost 2x as much as the next largest exporter while it was not even in the top 3 in 1995.

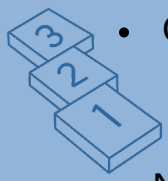




- The US exports ~60% of the dollar amount it imports while China imports ~60% of the amount it exports.
- China exports almost 2x as much as the next largest exporter while it was not even in the top 3 in 1995.
- When the war in Ukraine started, Ukrainian exports of Iron and Steel fell by ~50% compared to the average before the war.
- Syria's exports and imports fell by ~90% and ~70% after the beginning of the civil war in 2011 and did not recover at all by 2023.

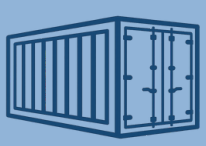
Storytelling / Layout

Our storytelling objective unfolds in three acts:



- Connect to the broad narrative the user is likely familiar with:
  - Show aggregate trade measures and connect it to material reality (e.g., how many million kilos of products are shipped every year).
- Narrow down to specific countries and stories involving trade:
  - How Saudi Arabia, an export giant, imports the vast majority of its food staples.
  - How the USA, famous for having the largest trade deficit in the world, is energy independent and exports its surplus.
- By now, we hope that, by observing some disaggregated trade scenarios, the user's interest is piqued. We therefore offer a sandbox experience for exploration through a set of interactive visualizations. The user can select among more than 500K combinations of countries, product categories, and years and observe trade flows temporaly and spatially through maps, sankey diagrams, timeseries, and bar charts.

Layout: We believe it is important to experience these steps in sequence and we, therefore, followed a strictly linear approach in terms of navigation. The user scrolls through three distinct sections that correspond to the themes described above. The middle section explores two distinct product categories, food and mineral fuels, and it is therefore



Visualisations

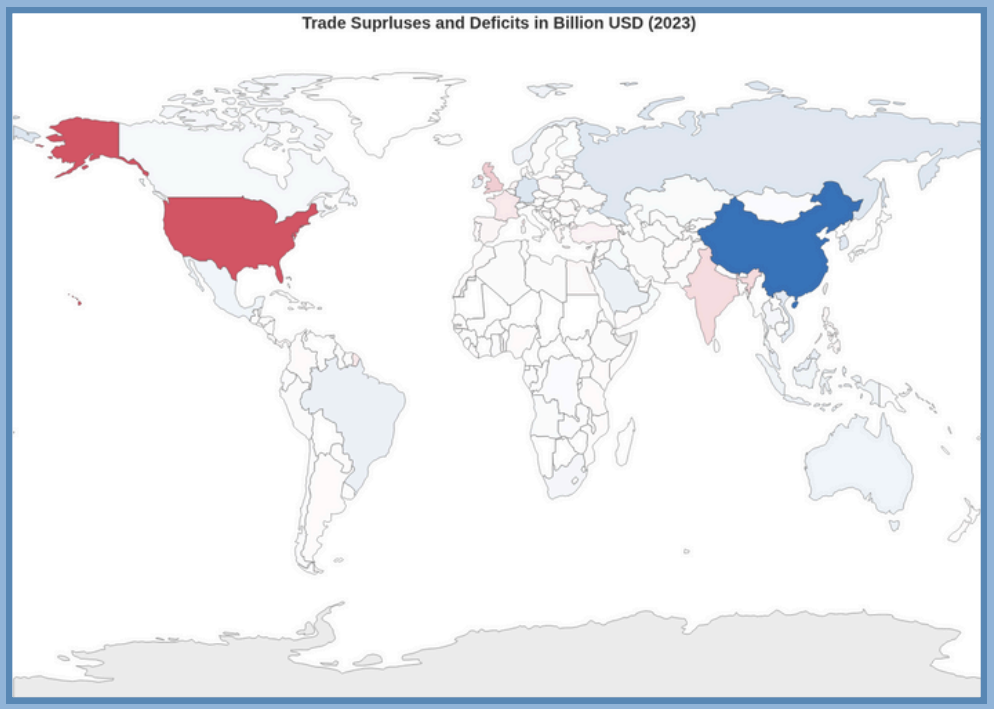
We will break this section down into the three storytelling themes

Introductory

In this section, we stuck with the sketches of millestone 2. There are two main types of visualization that help ease the user into the world of trade.

The first consists of annotated timeseries that show that more and more economic activity is conducted through trade.

The second is a world map that reflects the numbers thrown around in newspaper headlines: Total goods surplus/deficit for each country.



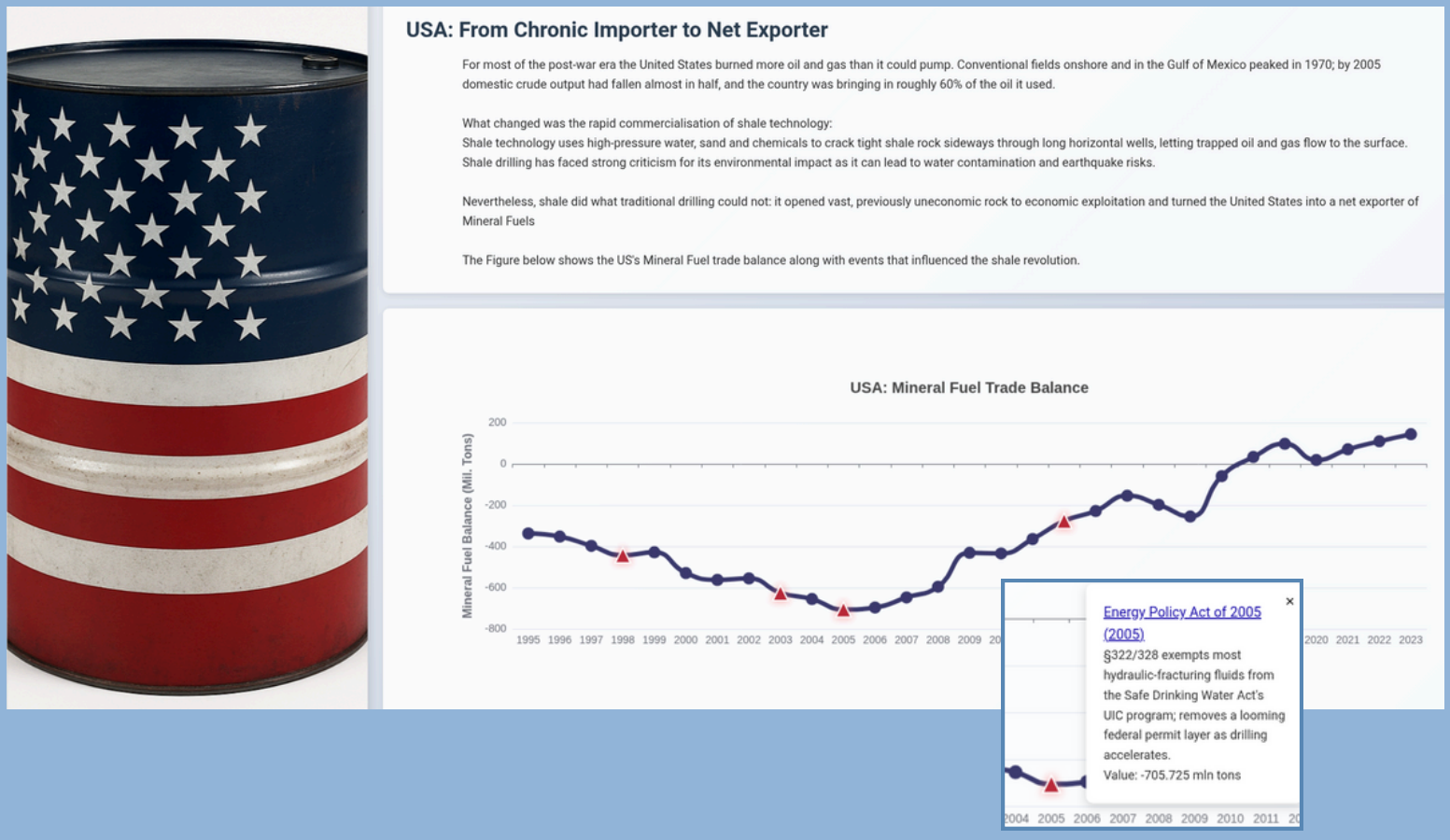
Case Studies

This section icnludes 6 country- and scenario-specific visualizations that form the linear narrative part and explore a specific scenario. Each page is split into three sections:

1) A banner on the left that references aspects of the material world that are connected to trade. We believe that it helps sets the tone and connects the discussion to experienced reality.

2) A text section that provides context and forwards the story.

3) A visualization that backs the story in an intuitive way. When appropriate, there are interactive elements that connect the timeline to specific events.



Each product category (food, mineral fuels), closes with an interactive map of the world that shows the trade balance of each country for the cateqory in question. The point is to let the user answer question that may have arisen during the narrative storytellin – for example: “if Ukraine is a large food exporter, who else is?”

Interactive Section

The goal of this section is to let the user freely and meanigfully visualize trade date for more than 500K combinations of countries, products, and years. The user selects the year through a slider, the product category via a drop-down menu, and the country by clicking on it on the map.

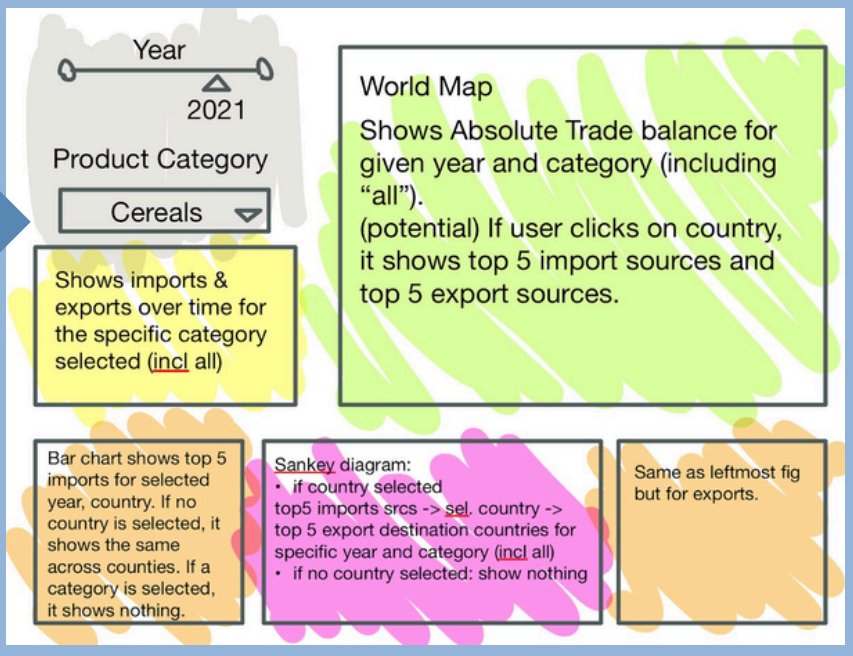
The following visualizations are then available to the user:

- [Across Space] **World map** showing the trade balance of each country for the specific product category and the specific year.
- [Across Space] **Sankey plot** showing the largest import sources (countries) and largest export destinations for the specific product category, year, and country.
- [Across Time] **Timeseries plot** of imports and exports from 1995 to 2023 for the given product and country.
- [Across Products] 2 x **Bar charts** showing the largest import and export categories for the selected country and year.

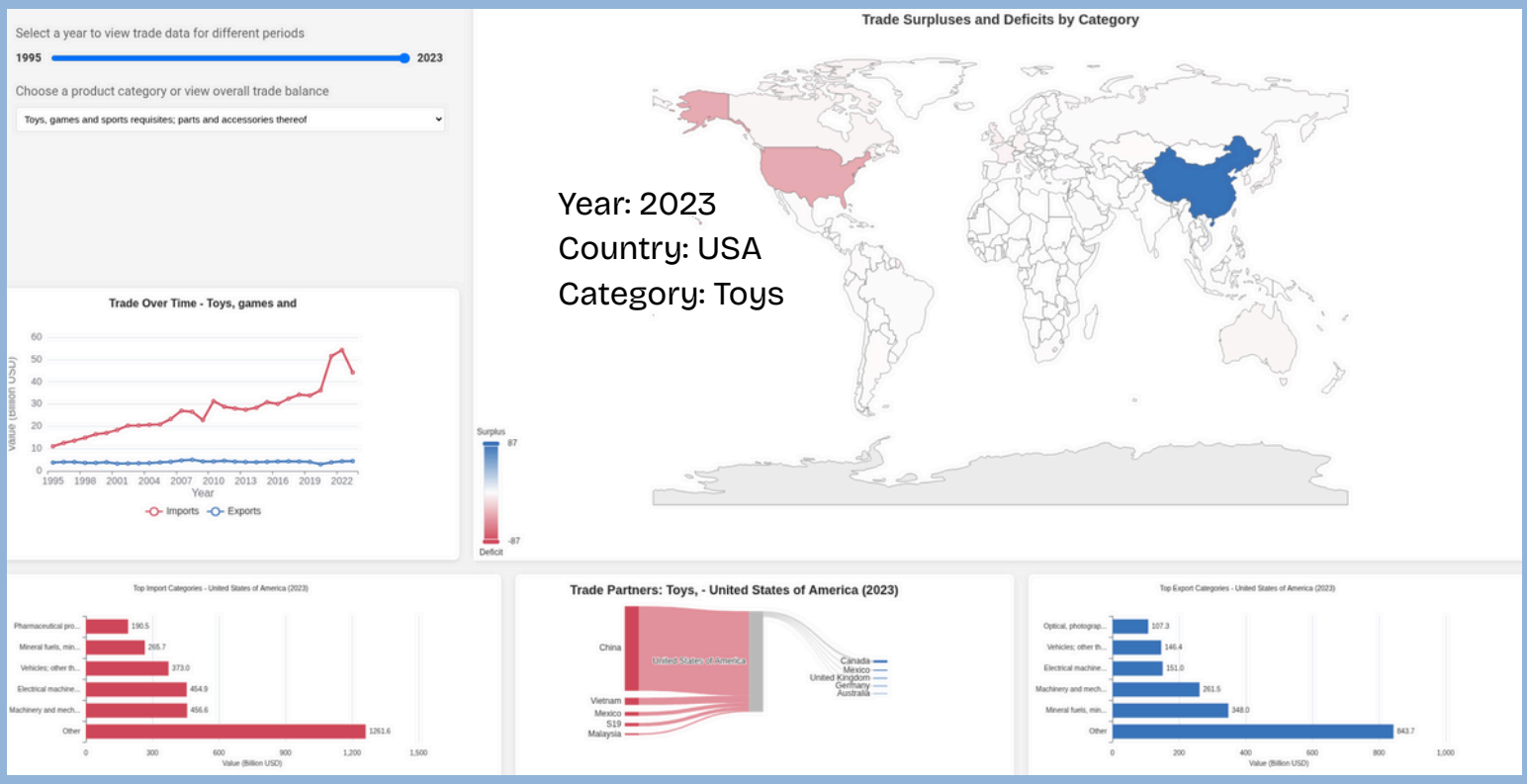
Using this tool, we concluded that it would be great to have it before doing research for this project, as it enables the exploration of a vast space.

Initial Sketch

Final Sketch



View of interactive section



Specific Challenges

Data Challenges:

- The dataset is too detailed in terms of product categories (~5000) for non-experts and for the purposes of this project. This fact also massively increases the overal size of the data, making loading times longer. We studied the HS92 format and reduced the category space to 87 as intended by the spec. Even then, there were 18 different categories of food – to fine-grain to distill intuition from. In this case, we aggregated manually.
- Although the data is very rich, it consists solely of trade flows. It does not include domestic consumption/production. When needed, we obtained additional datasets to normalize against reasonable metrics such as country-specific Gross Domestic Product series.
- One of the two trade value metrics of the dataset is Current US dollars (i.e., not adjusted for inflation). Because adjusting for inflation in a global context (i.e., not for any given country) is tricky and error prone, we avoided temporal comparisons using Dollars and either normalized to the GDP, or compared using weight within specific product categories.

Storytelling Challenges:

- As we decided to find specific cases and events that influenced the course of trade, the process was highly manual (and interesting). We went through dozens of countries and explored their trade history across product categories for the EDA of the project.

Visualization challenges:

- On the technical side, the combinations Year x Product Category x Country that we visualze interactively at the end of the site is very large and can yield poor performance if it is not handled carefully. To this end, for this prototype we followed two approaches:
  - Co-designing for function and performance: for example, only focusing on the top 5 peer countries reduces fetch volumes.
  - Pre-processing: We avoid almost all processing of the data in real time to make the website more responsive.

Peer Assesment

Huiyun Zhu worked on implementing the interactive maps across the website. Jointly with YuChen, she implemented the interactive visualization page, with emphasis on the interactive inputs and the map section. Huiyun documented decisions and challenges regarding these sections in this report.

Konstantinos Prasopoulos worked on the overal design of the website (storyline, presentation) and on executing and visualizing the handpicked stories in the middle of the website. He also processed the dataset and produced the backend inputs for all visualizations of the website. Konstantinos documented decisions and challenges regarding these sections in this report. He also designed and wrote the vast majority of this report.

Yuchen Qian designed and implemented the flow of the website as the user scrolls through the it. Jointly with Huiyun, he implemented the front-end part of the interactive visualization page; specifically he implemented the Sankey, the timeseries, and the bar plots. Yuchen documented his experience regarding these sections in this report.