

# PROCESS BOOK – INTERNATIONAL TRADE

*DAVID DOWEY, ROSEANNE FENG, MATTHEW SILVA*

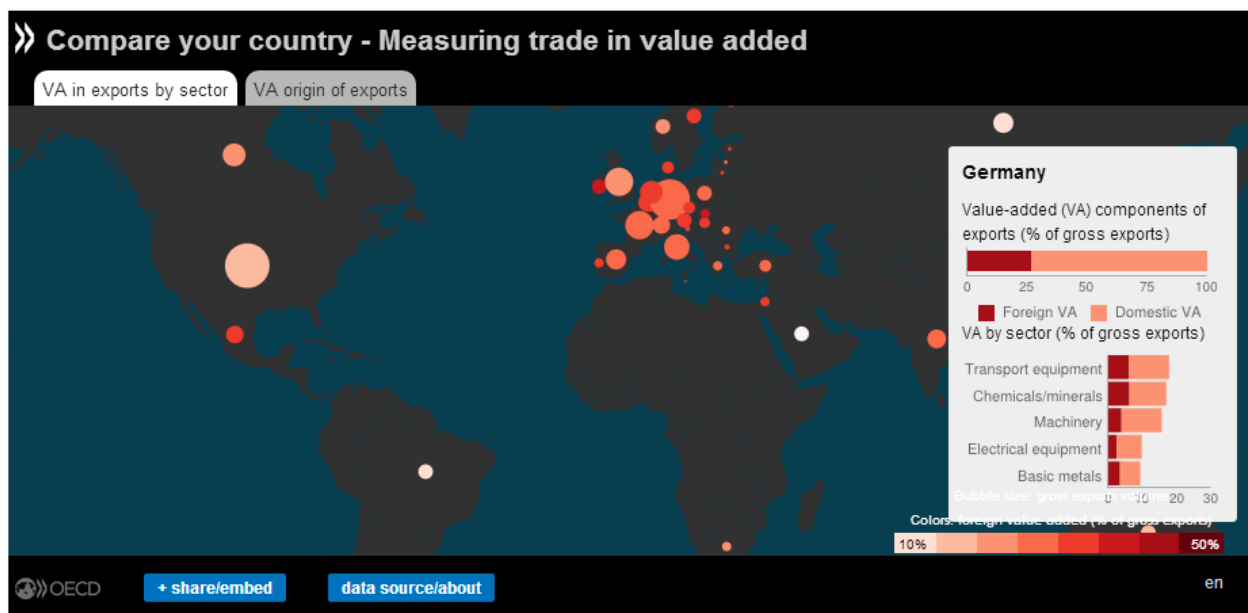
## OVERVIEW AND MOTIVATION

In a global economy, understanding international trade is important to understanding the economy as a whole. However, most data on international trade allots the total value of the final product to the country that manufactured it. This would be an accurate measurement if it were not for the fact that many countries do not produce the entirety of a product from raw materials, often importing components or labor from elsewhere. In this case, it is more useful to assign numerical values based on the percentage that has truly been contributed by that country, in order to show a different perspective on international trade and productivity than the traditional final product-focused data we are often presented with.

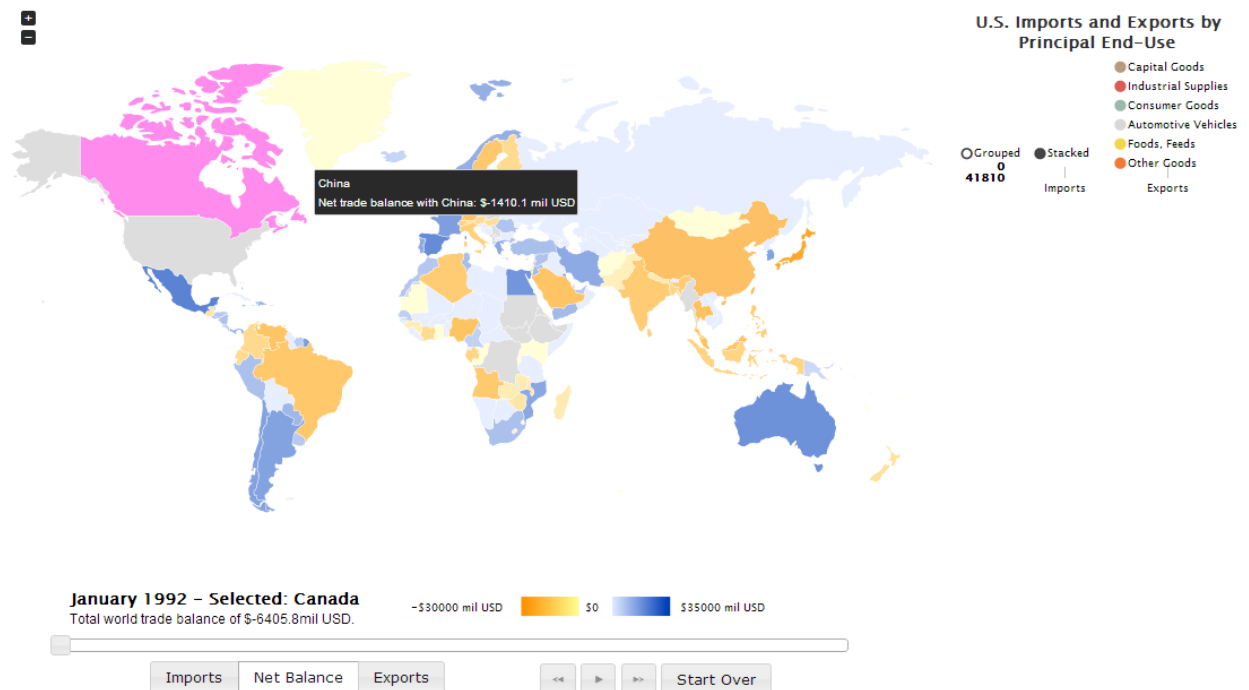
We will be looking at the Trade in Value Added initiative (more details under Data), which has brought together several existing areas of world trade to attempt to gain a better picture of flows and the overall economic impact – International Trade Statistics, Input-Output Tables and Bilateral Trade Database by Industry. The resulting database provides a substantive basis for some myth-busting analysis: e.g. how much of the value-added of your iPhone comes from China, and how much is actually value-added from other countries (such as design, research, raw materials, etc.).

## RELATED WORK

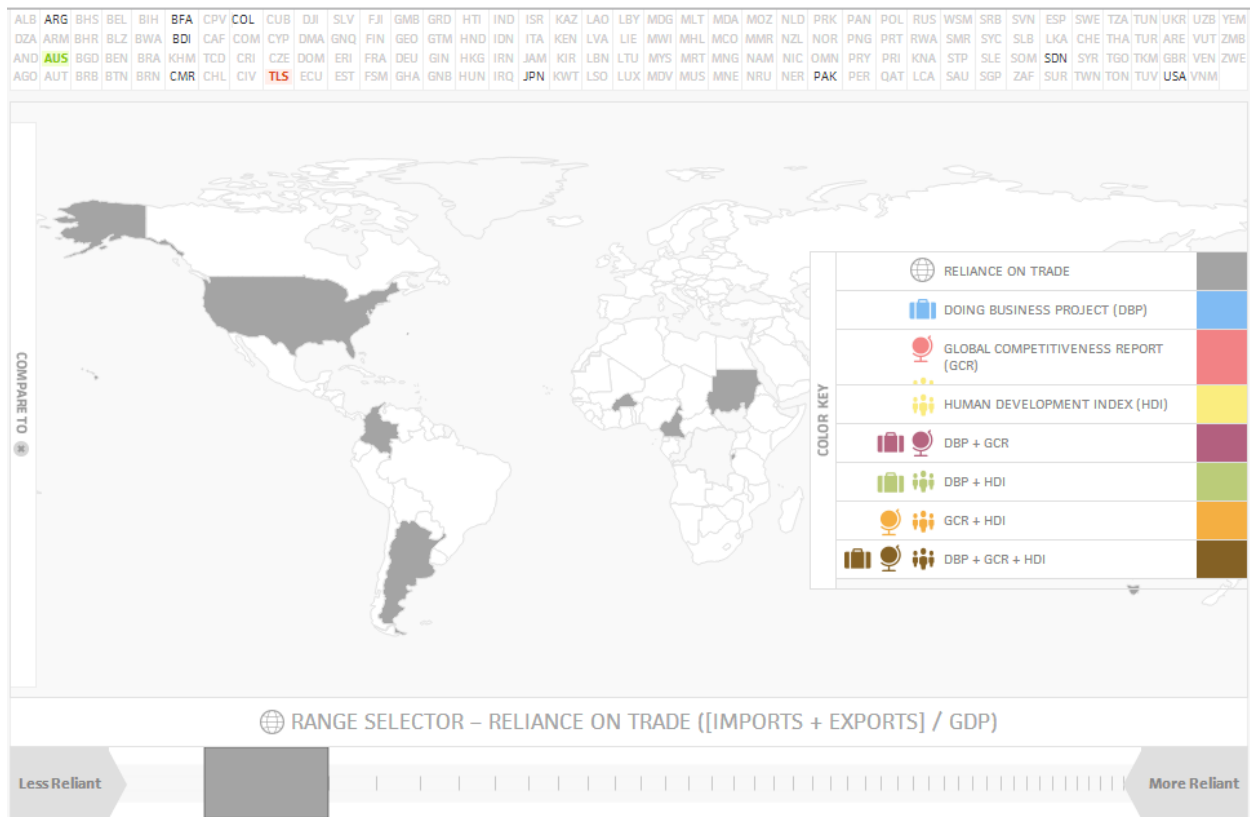
The OECD made their own visualization representing the same data set available at <http://www.oecd.org/statistics/datalab/trade-in-value-added.htm/>. While this visualization uses the same dataset as ours, we plan on adding additional information and encoding our data more effective visually than the simple bubble on the map with tooltips that is show on this visualization. We note that although some features are useful (the tooltips help to avoid from cluttering up the entire map, since they can be dismissed and removed from the canvas when not currently being looked at, and the ability to switch between two kinds of data, in this case exports by industry versus exports by country of origin), some of the other features make it hard to read the map, for example using the circles to represent gross export. While it is intuitive to connect these two characteristics, it makes certain parts of the map illegible, most notably in Europe.



A project from last years' class, created by Albert Young and Kevin Sun, also involved international trade data. This visualization is available at <https://googledrive.com/host/0B6YljmV-VTJZWlU4RU1hNUJOQzQ/index.html>. This visualization uses a different data set than ours. They used U.S. Census Bureau data, focusing on the US's exports, and they examined total value rather than the value added. However, there are similar features included, such as the ability to look at the data for different points in time and, again, the use of tooltips to encode more information without making the visualization unnecessarily cluttered. The use of detailed visualizations (not shown in screenshot) to further analyze the data for a given country was also a feature that helped improve the narrative of the visualization.



We also looked at a third visualization focusing on global trade, which can be found at <http://www.ideaslaboratory.com/projects/visualizing-global-trade/>. This visualization contains a high density of information that is at first overwhelming, even with the introduction. It also focused on different information, based around reliance on trade and making comparisons on a global level. However, looking at it gave us a better idea of what information we wanted to include and how we wanted to encode it, as well as helping us understand what kind of global narrative we wanted to show.



## QUESTIONS

We want to understand the productivity of each country in the international economy, which we will measure by looking at the value of its exports. However, we also understand that this data is often skewed. Our project will give us a clearer picture of how countries do interact in the global economy. In visualizing size of flows and linkages between countries, we will learn:

- How much of trade is value-added: and as source or receiver?
- What are the cross country linkages and how can we effectively show them on a map?
- What are the breakdowns by source country and by industry group (i.e. is there more value-added in services or raw materials?)?
- What are the special cases and details of trade in value-added flows?
- Are there ‘trade in value-added’ hubs? In what industries? How are they different from traditional ‘trade’ hubs?

## DATA

We are using data from the OECD-WTO: Statistics on Trade in Value added: [http://www.oecd-ilibrary.org/trade/data/oecd-wto-statistics-on-trade-in-value-added\\_data-00648-en](http://www.oecd-ilibrary.org/trade/data/oecd-wto-statistics-on-trade-in-value-added_data-00648-en). This data is available in a variety of formats such as XML and CSV. Our first attempt at using this data involved parsing the CSV file to generate a JSON file containing only what we needed. Later we decided to use the OData API to query the requested data each time the visualization is run. This allowed use to have more data available for the visualization without having to load too large of a file. The OData API is Beta version only, however for the purposes of the data calls so far, it seems to work fine. The API has reduced the amount of pre-processing work, since data can be loaded according to the views/filters of the visualization the user selects. There is one call for the chord diagram that is slow to load however and we need to find options for this.

We are using two different data sets that the OECD has available: the “Origin of Value Added” dataset TIVAORIGINVA and “Trade in Value Added” dataset TIVA\_OECD\_WTO. The TIVAORIGINVA is the database that breaks down trade in value added by source country and source industry (point #1 of ‘Data’ in our proposal). This is the value added embodied in gross

exports broken down by the value added by various source countries, and by the various source industries. The dataset provided interesting cross-tables, but the information is limited. We can basically see where value added is coming from – either the view by country of source, or the view by industry of source, or eventually both. The visualizations at a country-specific level are based on this database, as is the industry chord diagram and treemap.

TIVA\_OECD\_WTO is the database with the more extensive series related to trade flows. (points 2 and 3 of ‘Data’ in our proposal). Here the most interesting is the basic decomposition of gross exports (Gross exports = { Domestic value added embodied in gross exports + Direct domestic industry value added content of gross exports + Indirect domestic content of gross exports (originating from domestic intermediates) + Re-imported domestic value added content of gross exports } + Foreign value added content of gross exports. This larger database is interesting to get additional information for the visualization from: for example, the gross exports decompositions would be natural to show in a stacked barchart – i.e. gross export value broken down into its components of value-added. This can be done either for total gross exports, by share or exclusively for services. Other interesting information from this dataset are export ratios and the revealed comparative advantage. We have been thinking of how to use this additional data in order to better describe differences between industries.

## **EXPLORATORY DATA ANALYSIS**

We analyzed the data in excel and made a number of different visualizations using D3 to explore the data and find the best views to present for our final submission. We made choropleth maps for both datasets with selectors for all the different options available for querying the datasets. As noted above, one map visualization was based on the “Origin of Value Added”

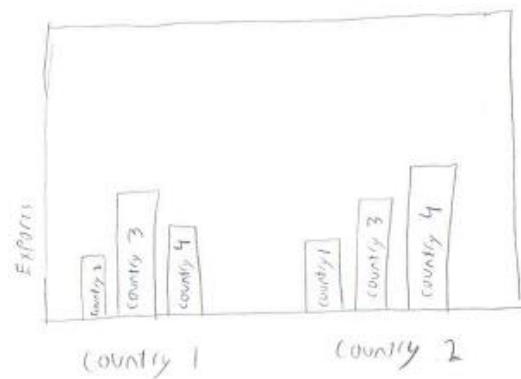
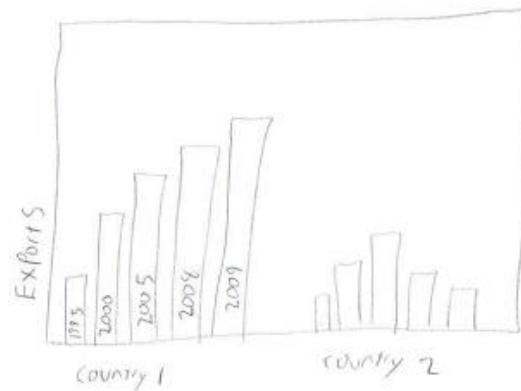
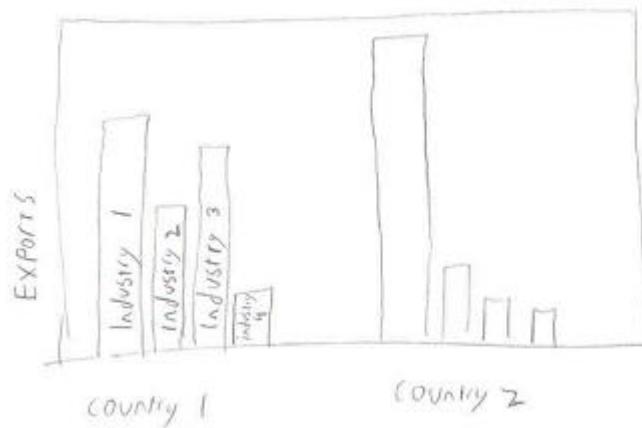
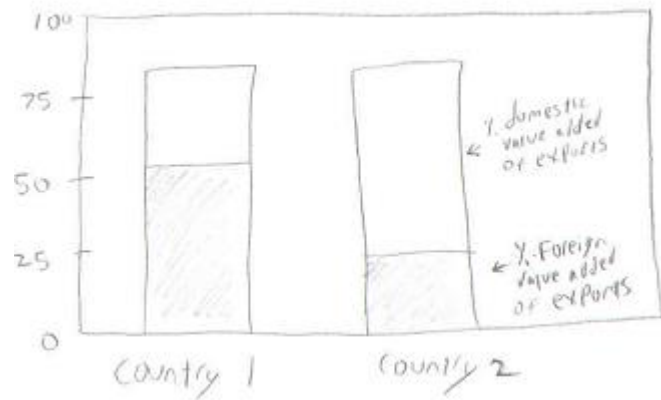
TIVAORIGINVA dataset with a map allowing you to select country, indicator, and source industry (row) from the option box selectors, time from the radio buttons and industry from the panel list – the map visualization is based on the corresponding row of the table. The dataset dimensions are ‘COU’, ‘COX’, ‘IND’, ‘INX’, ‘TIME’, ‘VAR’; for country, source country, industry, source industry, time and variable.

In another, the TIVA\_OECD\_WTO database was used in order to visualize a map with a range of data slices – selectors were created for most of the dataset dimensions which are ‘COU’, ‘IND’, ‘PAR’, ‘TIME’, ‘VAR’; for country, industry, partner, time and variable. While this map was not intended to be used in the final visualization, it allowed us to have a quick view on potential interesting cross-dimensions and stories to-be-told. We came to the conclusion that the TIVAORIGINVA dataset should form the core of our visualization, but that the larger TIVA\_OECD\_WTO dataset could be used as an first stage map for trade with partner “World” showing a breakdown of gross exports by value added, and then the dataset could be used to complement the core visualization with relevant detailed information.

## BEGINNING

Our initial idea was to construct a choropleth map (prior to Homework 4) which assigned colors based on that country’s gross exports, with tooltips for more data and a bar chart that would show that country’s exports compared to its value added, which would be a quick visual indicator of which countries relied more on foreign trade and which produced most if not all of their value. We also discussed other ways of implementing the bar chart, but in general thought that the bar chart would be a simple way to show these differences using Tufte’s design principles and because it would be easiest to see quantifiable relationships such as the exports to value added proportion. Concept sketches of these are included on the following page.





← Shows who we trade with.  
Where the exports go



## DESIGN EVOLUTION

Given the feedback received from our proposal, we have decided to consider many more features to be must-have and have attempted to increase the scope of the visualization. We have added chord diagrams, breakdowns by industry, as well as selection by year to our visualization for the first milestone after discarding the bar chart as too simple and not conveying enough information for the space it would require.

The core visualization on the “Origin of Value Added” is set up so that there is a country of focus. This set-up allows the country we’re focusing on to function as a driver for the story – basically, select a country and see what you can learn about its trade in value added. We needed to decide on the extent of the visualization elements that would be naturally country-dependent. As a first choice, we made a chord diagram exploring the value added to and from various industries for the selected country. The chord diagram is good for showing the connection source-destination and also the size of the value added contribution. For example, the chord shows the value-added from mining into chemicals. We may need to explain to users what it is they are seeing for this visualization element. We have also investigated a few different representations of bar charts and a table of top industries as a tooltip. Another possibility, but that we have not found an adequate solution would be to put arrows to show size and direction of value-added flows in addition to the colors of the map. It would be effective to visualize directions of flows: how does each country look as a source and receiving value-added trader?

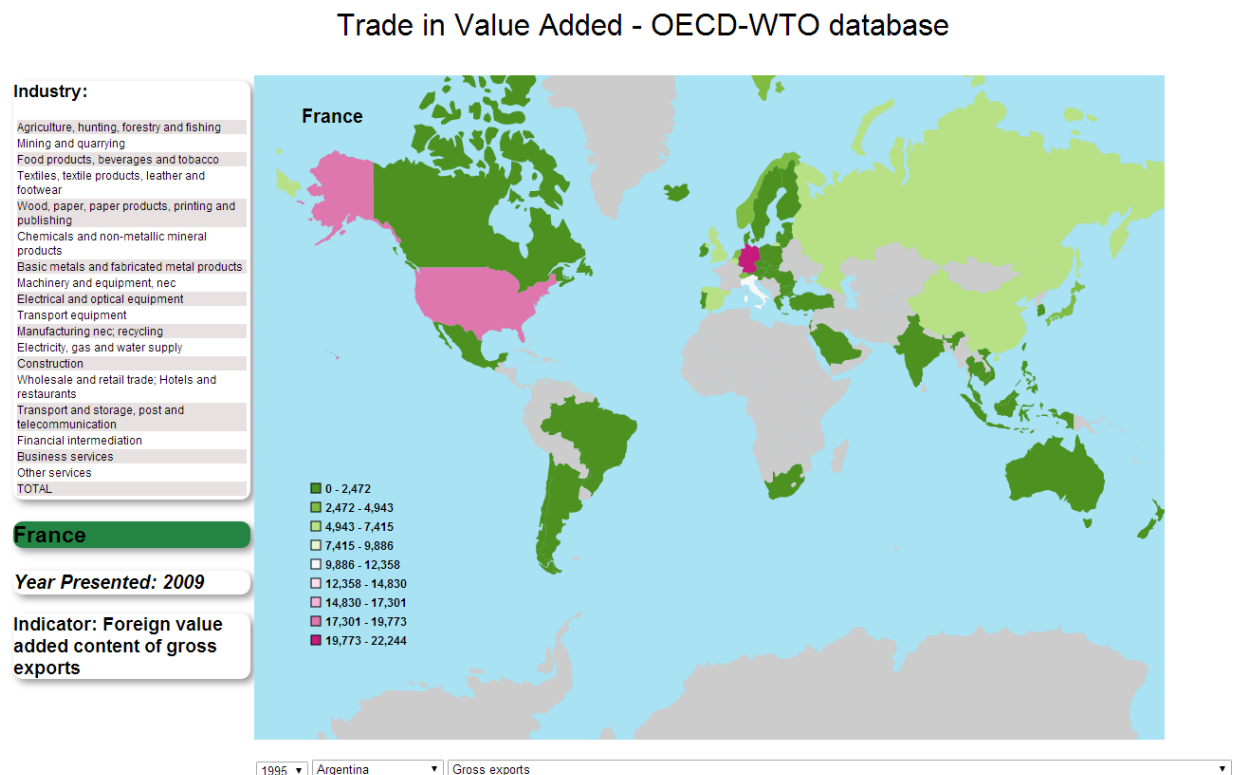
## DATA HANDLING

We initially struggled with how to process such a large database, and what form it would be most efficient to manipulate the data in. The OECD data is stored in different formats. Initially, we thought we would be able to use CSVs but realized that the files were huge and hard to work with, so we instead used the API to access the databases, save as JSON files, and load the objects into the site to begin building the visualization in our initial implementation. As we progressed and started adding more features, it was clear that this was time-consuming and inefficient, again in large part due to the size of the raw dataset to be accessed through the API. We switched from

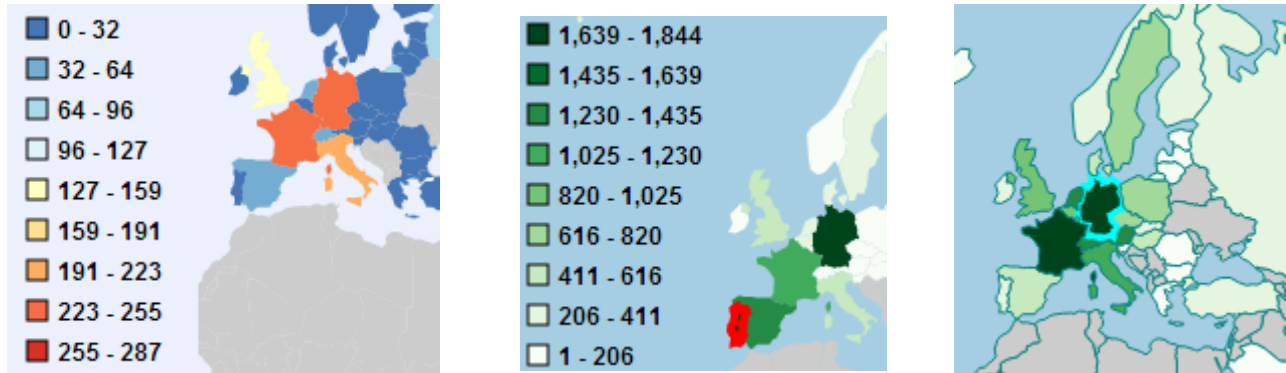
this to our current system, which involves making asynchronous queries for the individual countries when selected and querying again to load the information necessary whenever the country changes. For large data calls, such as the chord diagram, tooltips and treemap, we use cached data loaded from JSON files with some backup processing similar to our original method of gathering the data for the choropleth map.

## IMPLEMENTATION

The choice was made to continue working first on the “Origin of Value Added” core visualization. Our first design looked like this.



Several problems arose with the first version of the map:



- Scale for shading: the choice of scale impacts the way information is received from the map. While progressive shading is more natural, it doesn't produce the necessary contrast. A quantize scale did not produce enough differentiation. To get a more balanced breakdown of colours, a quantile scale with a progressive green scheme was chosen. Since the majority of source value-added comes from the selected country the domain of the scale should exclude it. As seen in the screenshot, the colour scale was originally a bicolour scale that passed through white, but we determined that the different colours were distracting, so green was chosen to stand out against the blue background and the grey countries for which we had no data.
- a better zoom: the zoom in to focus to fill the bounding-box on any country by click needs to work in tandem with the country selection on a different event. Currently it is also not possible to zoom out on laptops, since the zoom function is handled by the mouse's scroll wheel.
- the GeoJSON does not show clear borders. Also a consideration is to reduce the extent of the map to exclude Antarctica. We coded a function to highlight the border of the selected country in red to better differentiate it from its neighbours. We experimented with ways of showing borders, which was a dilemma because heavier weights on the other countries could be either helpful or distracting. For countries with no data, borders might give the impression that they should still be looked at despite being greyed out. For countries with data that weren't selected, we thought that since a lot of them were neighbours, especially in Europe, thicker borders might

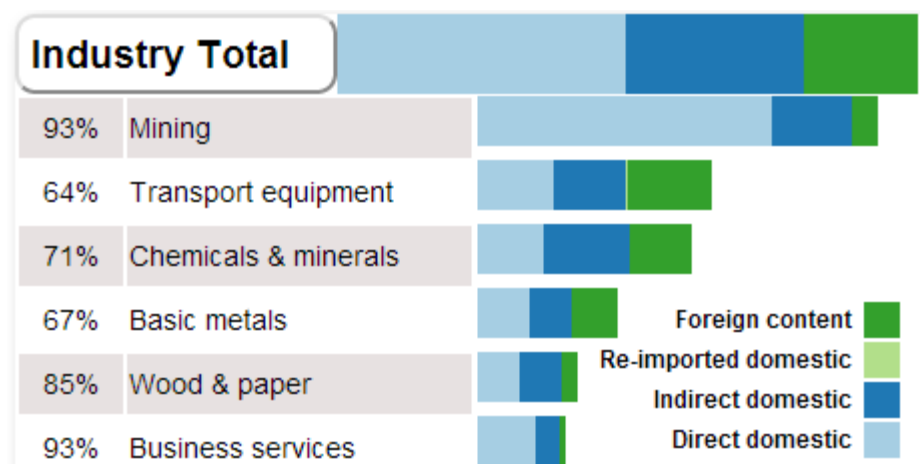
have made the visualization appear more cramped. This is clear in the second example under the color scale, where Portugal is relatively small and most of its image is replaced with the red border, impeding visibility. In the final version, the selected country's border is drawn in a thicker cyan line, while all borders are drawn in thinner, teal lines, which make it easy to distinguish small countries like Denmark or Brunei, which we do have data for, without being too distracting.

- Variations on the industry selector: to use colors more effectively – perhaps assign a color to each industry that can be used elsewhere in the visualization but, similar to the borders, this could be overwhelming, especially since the borders themselves could reasonably be reduced to two colors but we have 18 different industries to show.

One technical design problem is solving how to put multiple selector mechanism in the visualization. We did not consider the selectors while creating sketches of the visualizations, so when we initially created the files, we ended up coding in no less than three dropdowns, which was rough but allowed us to check the way we were processing the data. However, it was slow switching between different visualizations, and often required having to slog through the dropdowns. As we progressed, instead of using simple option list selectors, more intuitive solutions were envisioned and implemented, such as clicking on rows in a table of industries or using toggle buttons or radio buttons, shown below.



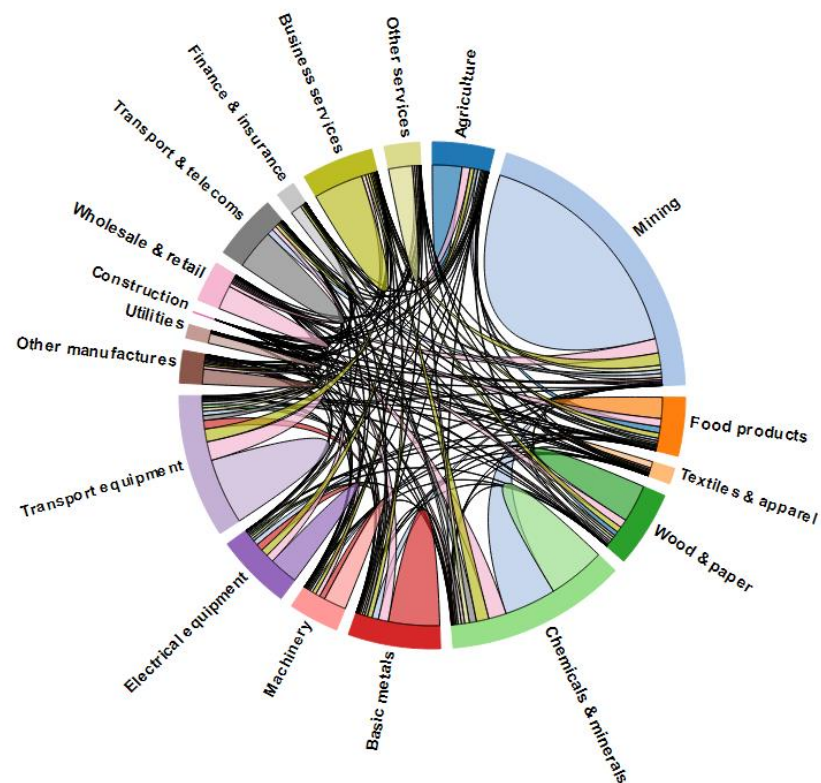
The industry selector was more complicated. The selector evolved into a panel with a table split into eighteen rows for each industry, as well as the total over all industries. In

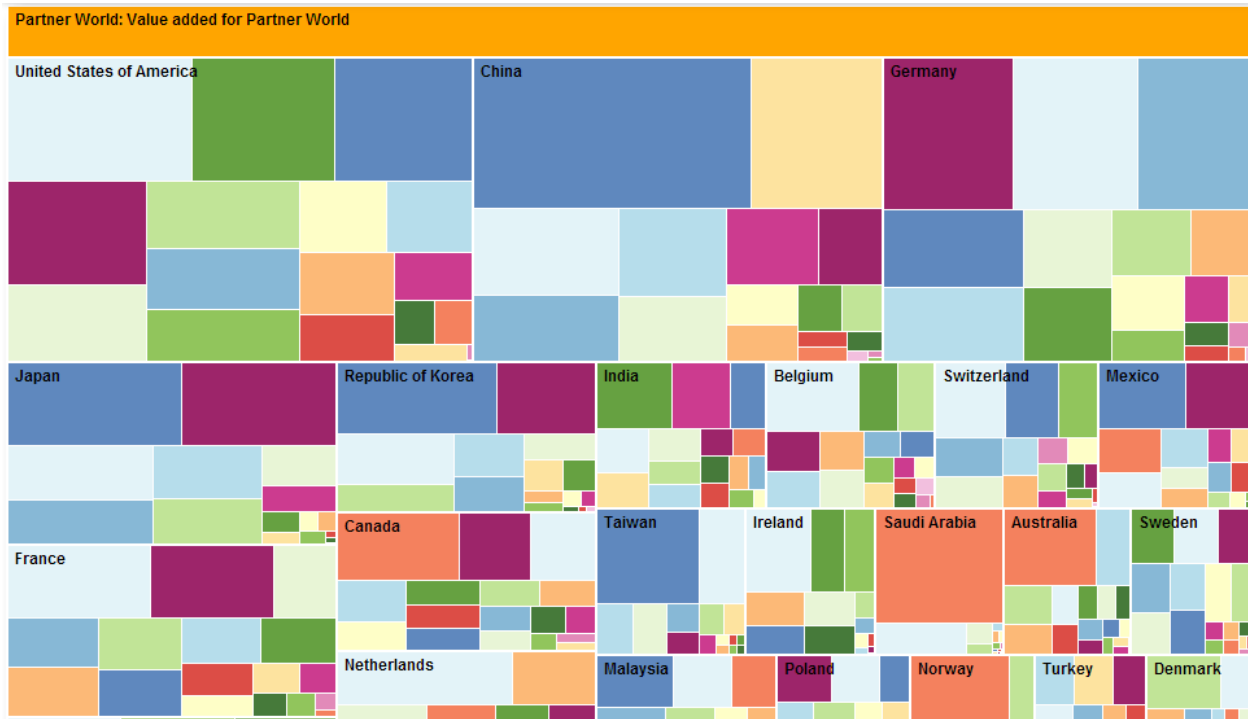


each row, the main part is clickable to select the industry, but the right column displays a bar divided proportionally by the percentage of that industry separated into the different amounts of value added over total value attributable to that industry as a source of value added (a pie chart was also considered but discarded because the area would have been more difficult to gauge visually, as well as needing comparatively more height for each row to be really effective). The table solved the problem of ordering the industries by some indicator – in this case, we simply ordered the table by the total value of gross exports. This is not only quicker to use than a dropdown, but has the added advantage of providing the user with more information about the industries specifically.

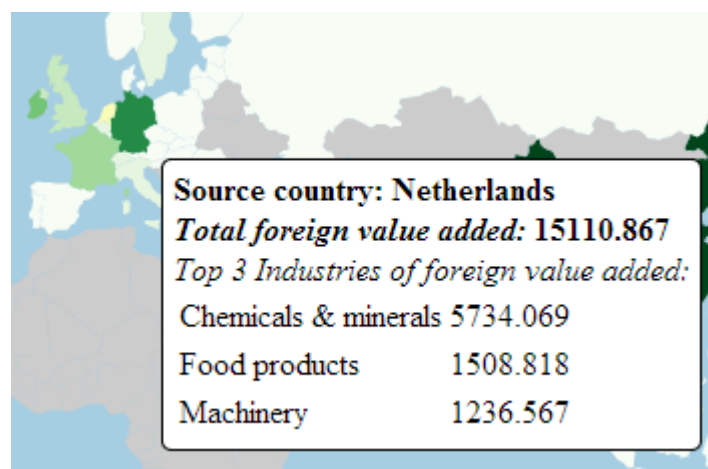
We also wanted to incorporate of relative importance of value added trade: the options were domestic added value export ratio and the revealed comparative advantage by industry for the selected country. In the end, the former was chosen to be encoded with a percentage indicator to left of the industry name, since it more effectively conveyed the difference between industries. We considered highlighting the industries in some progressive manner is possible with a small graphic showing the revealed comparative advantage for that country in each industry, but due to time constraints and a relatively smaller amount of information this would have provided, that feature was not included.

We also implemented a chord diagram, shown on the right, and a treemap, on the next page. The chord diagram breaks down the focus country into its industries and then shows the relationships between industries, which provides a clearer narrative of that country's exports and how the industries contribute to value added as well as to each other, while the treemap helps to visualize the proportions of a country's exports by industry on a global scale, with each box corresponding to a specific industry for that country.





For the map visualization, we need still to think about some other relevant data-driven symbology that might be relevant for mouse-overs on the map or other elements of the visualization. Tooltips were added that give small tables of top 3 source industries, appearing for the country that the mouse is hovering over in relation to the source country.



We also considered adding arrows to show direction of flow of exports, with the magnitude of the arrow representing the total volume of value from that country, but ultimately decided it was not very useful for our visualization. Since our visualization always has one focus country and displays the choropleth map based on the data from that country, we could already see what the source countries were, and adding extra arrows would have made readability harder.

We did not include additional linking and brushing to highlight countries because we wanted the users to be able to form their own narratives. The data provided is, we feel, rather intuitive, somewhat like net productivity. Additionally, it is a large dataset and due to the nature of the way we set up the focus country, would likely require us to hardcode in over 50 different sets of trade relationships by country alone, which becomes even more complicated when considering the different years that we could pull the data from. Instead, we encourage the user to focus on the countries that interest them while exploring the visualization and include suggestions for areas of interest in the introduction as a starting point.

## EVALUATION

In implementing our visualization, we gained a deeper understanding of the roles of various countries in producing exports, specifically how much of value that particular country actually contributed out of the total. It was very interesting to work with, as we can uncover and record trends such as countries that tend to have similar behaviors and are close geographically (notably, the European Union countries, but the developed/developing East Asian countries of China, South Korea, and Japan also have similar ratios of domestic and foreign values). We have learned how to quickly identify which countries have a lot of their export value originating locally and which, for example, import raw or unfinished products and improve them before exporting.

Using the five different views of the data that we have implemented, we are able to answer many of our questions just by exploring the data shallowly. For example, selecting various focus countries shows us a common trend of European countries trading heavily with Germany or electronics being a popular export in Asia with a large portion of those export values being foreign content/production. We can find outliers such as Saudi Arabia's unusually high

percentage of direct domestic value due to the mining industry, presumably due to its oil and other natural resources, or India's major export being business services, presumably due to foreign countries outsourcing for cheaper costs.

On a technical level, we produced multiple views coordinated with each other and that also support the changes in the country and time. We also addressed the second question, of how to best show cross-country linkages, by partitioning our data and using the industry table and chord diagram and treemap to better understand the breakdown for countries and more details about the industries of the focus country in particular. There are certainly improvements that could be made, such as a general streamlining of the visualization aesthetically (which was not high on our list of priorities), or designing a more useful way to implement the zoom and pan functions for the choropleth map. On the whole, we feel that, once the user goes through the introduction and understands what each view corresponds to, the visualization is very intuitive and shows not only relatively major partners of a given country in trade but also what industries are involved and how that country participates in a global market.