Process Book

N'yoma Diamond, Felix Chen, Luke Gebler

Overview and Motivation:

For our project we decided to visualize trade data, specifically between Brazil and the US. One main reason for this choice was the availability of large amounts of datasets at our disposal. There were multiple websites with a huge amount of data we were able to download and use, this gave us freedom in visualizations as we could choose and manipulate this massive amount of data as we wished.

Related Work:

We were able to find multiple visualizations with similar goals to ours that we could draw inspiration from. One of the most in depth examples of this is http://globe.cid.harvard.edu/?mode=gridSphere&id=US#. This website includes a plethora of different visualizations all interconnected, showing how many different ways this data can be explored. Another website similar to this from the same group is https://atlas.cid.harvard.edu/explore which has a variety of visualizations available.

Another website that was able to explore effectively the same dataset but in a different format is https://oec.world/en/profile/country/nld?subnatTradeValueSelector=tradeScale1&tradeFlowSelectorComparision=flow0. This site is able to focus a lot more on each individual country rather than how each country interacts with each other. This was useful for our project as we were only going to be focusing on two countries, the US and Brazil.

Questions:

Our main question when starting this project was figuring out how the trade relations between the US and Brazil changed during the past decade. As the project progressed we were able to attack much more specific questions as we were aware of what our dataset would be. Our next big question was how would we be able to show all of the different product types we have in our dataset in an easily digestible way. This was difficult for multiple reasons, not only were there more categories of products you could list in one graph but they each had a different impact, stemming from their trading value.

Data

The process for gathering data was lengthy and a key aspect of what our group worked on for the first few days of meeting. It was important we chose a dataset which had detailed enough data for multiple visualizations while also having an easy to use format. We eventually chose to

use WITS (World Integrated Trade Solution) in order to get our dataset. WITS has a built in query creation tool where you can select specific products/years/countries and more based on the data you want. Our dataset consisted of the trades, both imports and exports, between Brazil and the US from 2008 to 2019 of every product listed on the site. This outputted a csv with around 16 thousand rows, giving us plenty of data to use.

One main reason for choosing this data was because of the hierarchical order the products were categorized into. We chose to use SIC (Standard Industrial Classification) which orders the products based on a 4-digit code, the more digits, the lower into the hierarchy you are, and in order to find the parent of a product, you just remove the digit in the one's place. The main issue with this type of data classification is that it relies on the usage of leading zeros, which excel as well as many other data manipulation tools will remove due to their inferred uselessness. Fortunately we were able to quickly overcome this by just reading in the data as text and never saving it through excel. Some cleaning and alteration was also done using Tableau Prep to generate new data subsets, making the data easier and faster to access.

As a potential improvement for the future, the dataset we used could be swapped for one of Harvard's Dataverse trade data repositories, such as this one: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/H8SFD2. These repositories have much more data as well as using the SITC classification system, which offers better granularity, as opposed to the SIC system. The problem with this data set is that it is pretty unwieldy, and would require more time just to get accustomed to the dataset itself as it is organized differently from the WITS dataset and requires much more explanation.

Luke Gebler:

Exploratory Data Analysis:

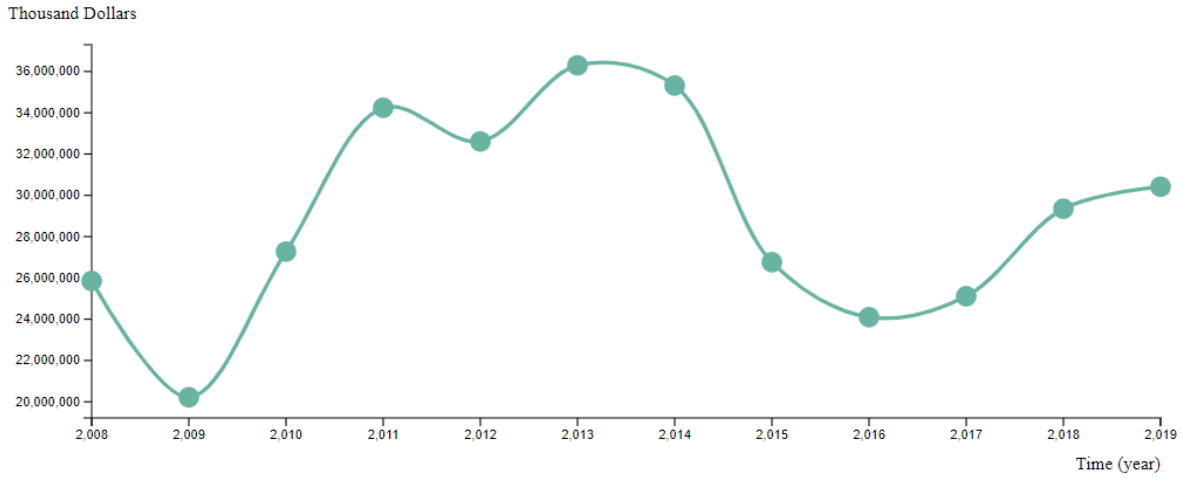
When first using the data we gathered, our first visualization was a scatter plot consisting of sixteen thousand data points all connected to each other. As you can imagine, this created a clustered and unintelligible graph which only on a technicality actually represented our data. From here we thought of different ways we could segment the data to show different statistics. This data was already ordered by aggregation so that was one avenue to pursue, which ended up becoming our stacked area chart. Another way to combine data is by grouping it based on one column. A few of our options at this point were grouping by year, trade flow (export/import), product, and more. Our scatterplot ended up grouping by year as we figured that was the most digestible way to show this huge amount of data.

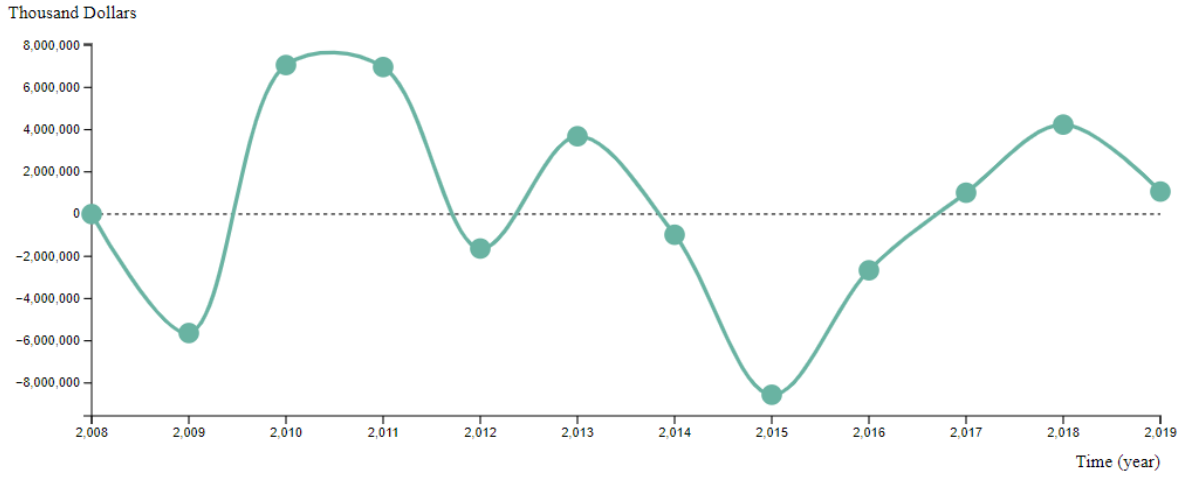
Design Evolution:

After our initial implementation of a scatter plot to represent that data, it was clear changes needed to be made, specifically to what we wanted to represent on each axis. A good starting point was time vs value. The year on the x axis and the value of the product on the y axis. This did end up panning out to our final iteration of the visualization but more options were tried in between. There were multiple attempts at switching value with either weight or quantity of the product, and even having a function amongst those, such as value/quantity * weight. While including those figures would technically be representing more data, we decided that simplicity overrules complexity when trying to manage and accurately portray a large amount of data. This did end up staying true to our proposal as the same overarching idea held true.

Implementation:

Below is the implemented visualization. The intent of this visualization is to give an overarching idea of how well the trade between the US and Brazil has been doing in the past decade. There are only two buttons which can swap between import and export as well as Total and Growth gain. Total references the total amount of value in the products either imported or exported that year while growth refers to the gain compared to the year prior. Minor details were added such as a color change between import and export as well as a dotted line at the zero mark when necessary in order to give the viewer a better experience at understanding the graph.





Evaluation:

Since this visualization shows the yearly progress of Brazil's economy, we were able to get some insight on when Brazil was struggling. In both 2009 and 2015 the graph takes a huge dive when comparing to the year prior. The 2009 dip was due to the international financial crisis while the 2015 dip was a much more targeted Brazil issue as their GDP plummeted especially when compared to other large economies. Another easily found piece of information is that Brazil imports from the US more than they export to the US which is interesting as the US famously imports a significant amount more than they export.

Overall, this visualization does a decent job at representing the data in a manner that is easily grasped and intuitive to use. Although, time was our enemy on this project and the graph could definitely be improved upon in many aspects. One thing that could be added is extra buttons or options to choose from to represent given data. As talked about earlier, we also had the weight and quantity of the products which could be switched out for the y-axis. Another idea that was tossed around but never was able to come to fruition due to time constraints was an interactive

bar chart we could include when clicking on a data point. Since the data represented is an overarching compilation of the yearly exports or imports, including a more detailed graph which breaks down that years products would be a nice addition, one caveat to this is that it may end up providing duplicate information given our other visualizations.

Felix Chen:

Exploratory Data Analysis:

The data we were looking at had each of the individual data points categorized according to import and export codes found at https://www.census.gov/foreign-trade/reference/codes/sic/esic.txt. These codes have up to four digits, where the first digit is the most broad descriptor and the following digits get more descriptive. This gave the data a hierarchy and having lots of different data points made the data perfect for a treemap.

Design Evolution:

Putting the data into a treemap made sense because the data we were looking at was hierarchical and there were a lot of different categories which would fill out a treemap nicely. With our initial treemap implementation, we only used the most specific level of aggregation, but we noticed that with this level of aggregation, it was difficult to see many of the rectangles and that the data was a bit overwhelming to view. To address this, we decided to add a slider which changed the level of aggregation. This made the treemap a lot easier to read, as you could start with more generalized data, and then change the slider to look at more specific data within each larger category.

Implementation:

The treemap implementation was meant to show the viewer a breakdown of all the total imports/exports for a given year and allow them to see what categories were the most important to either the US or Brazil. It shows similar data to our stacked bar chart, but allows looking at more data at once. There are sliders which allow the viewer to look at the breakdown of goods by both year and aggregation level. If the user wants to also see the value of an individual cell, they can mouse over that cell and see the category and the value of the goods traded that year.

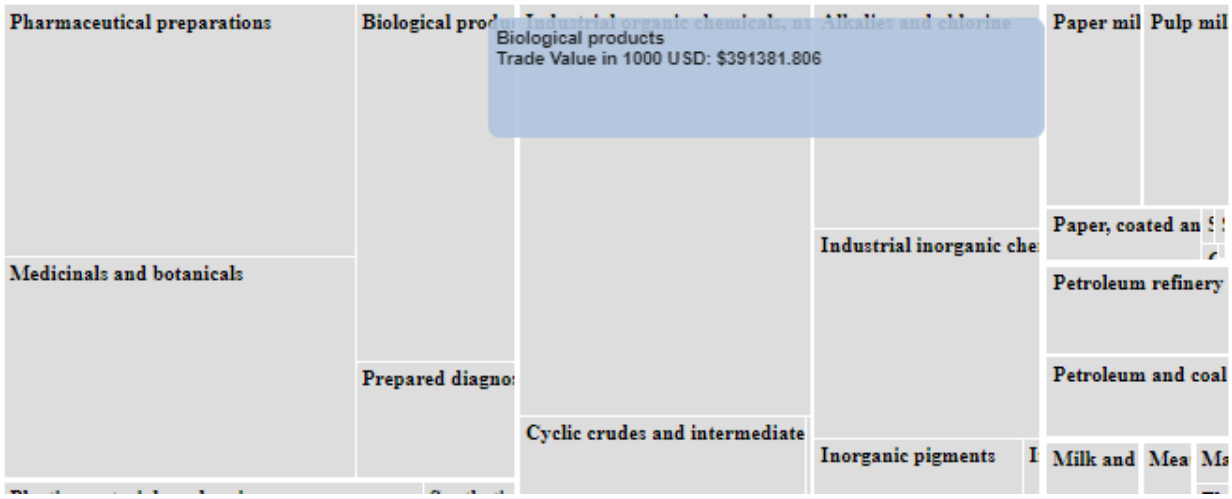
Breakdown of Brazilian Imports from the USA: 2009

Move this slider to look at different years. Hover over a product for more details:

Year: 2009

Move this slider to look at different aggregation levels:

Aggregation Level: 4



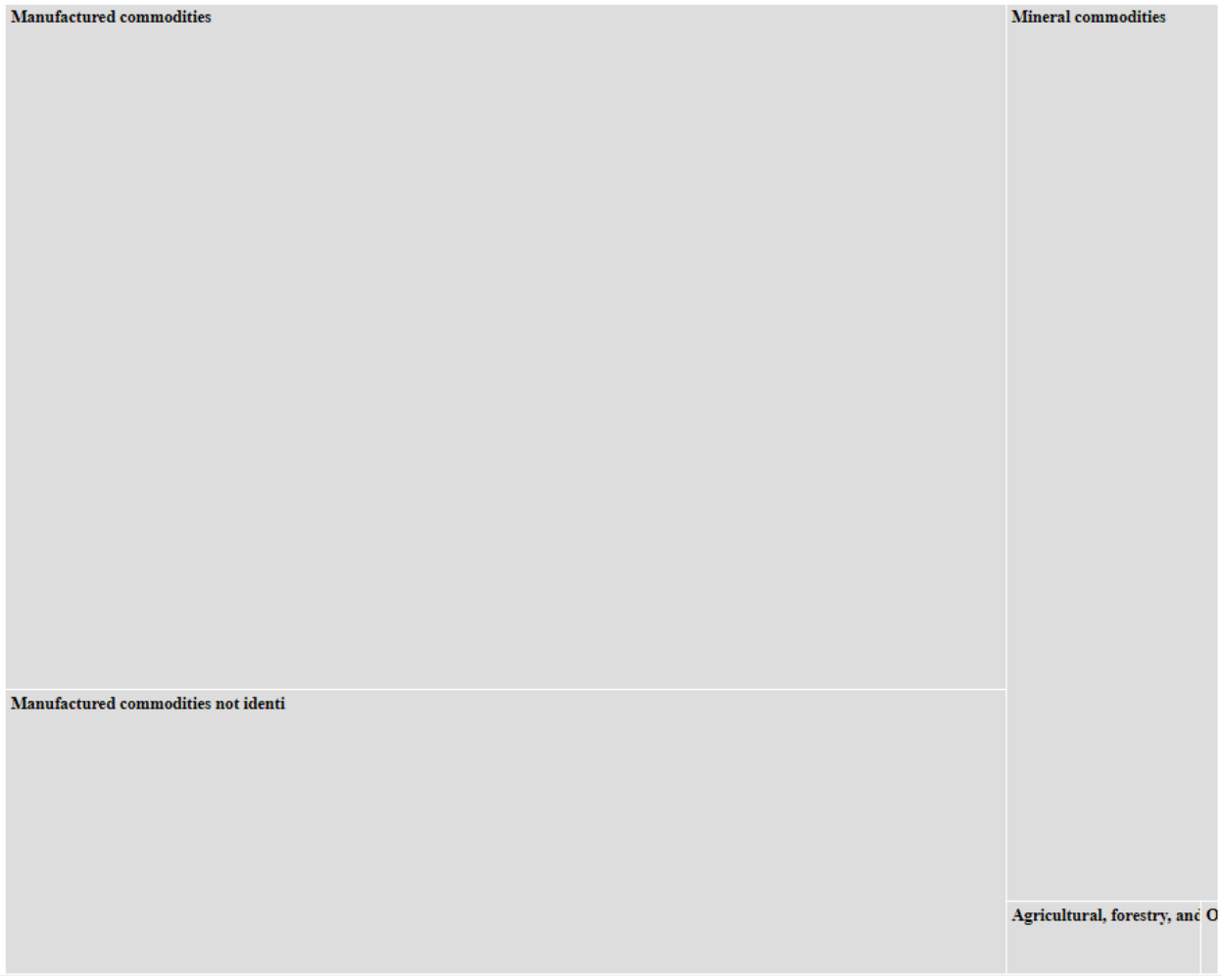
Breakdown of Brazilian Imports from the USA: 2009

Move this slider to look at different years. Hover over a product for more details:

Year: 2009

Move this slider to look at different aggregation levels:

Aggregation Level: 1



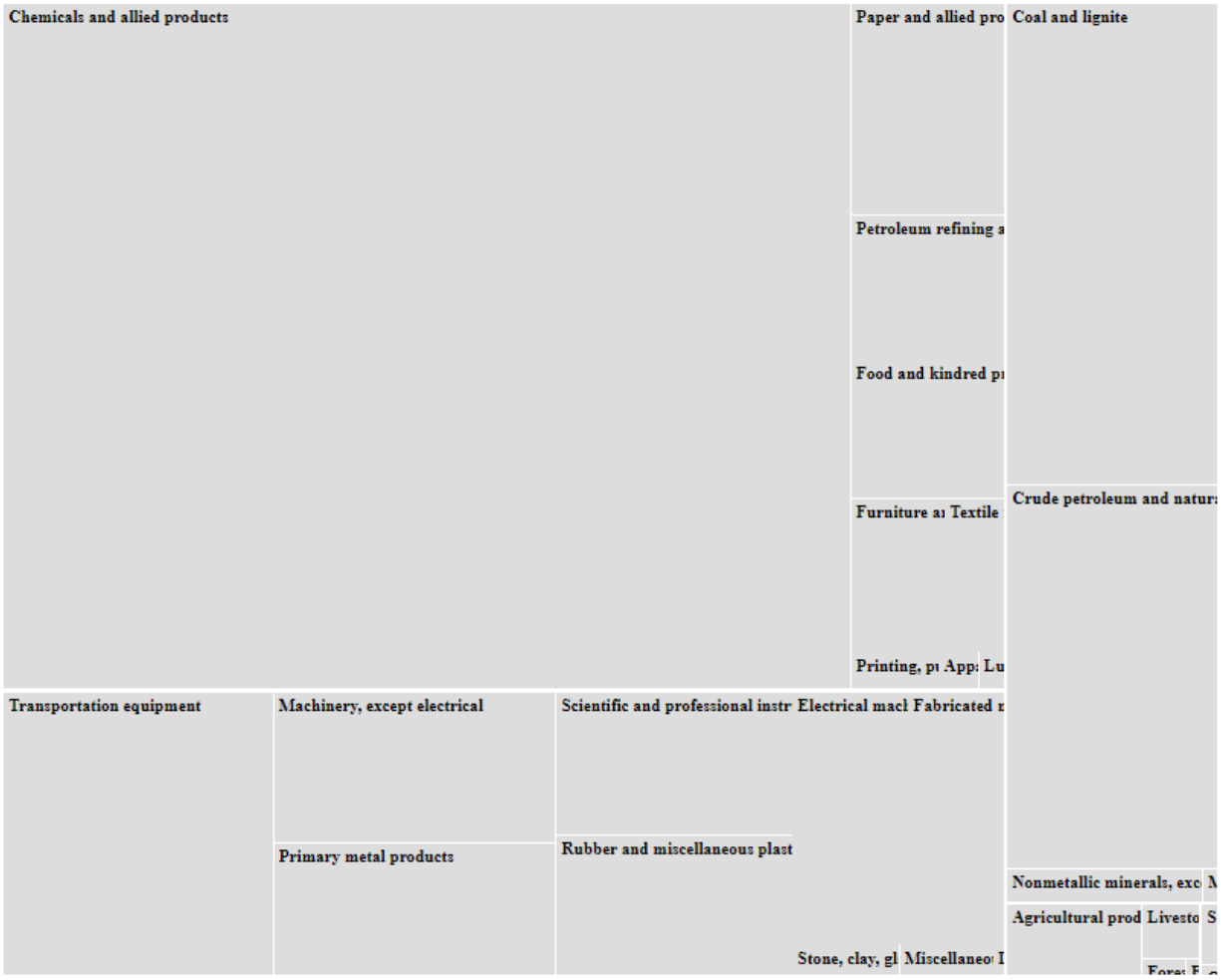
Breakdown of Brazilian Imports from the USA: 2009

Move this slider to look at different years. Hover over a product for more details:

Year: 2009

Move this slider to look at different aggregation levels:

Aggregation Level: 2



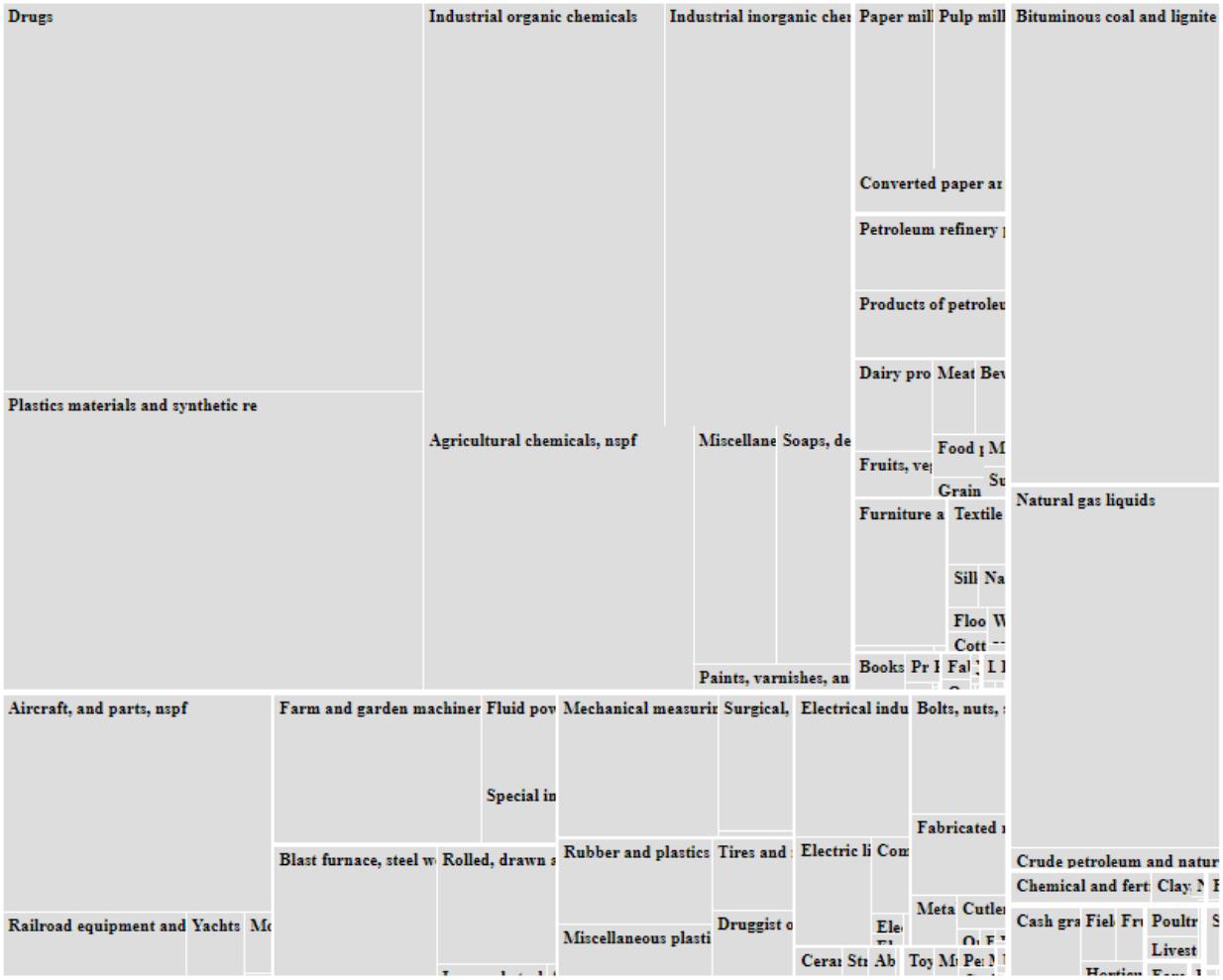
Breakdown of Brazilian Imports from the USA: 2009

Move this slider to look at different years. Hover over a product for more details:

Year: 2009

Move this slider to look at different aggregation levels:

Aggregation Level: 3



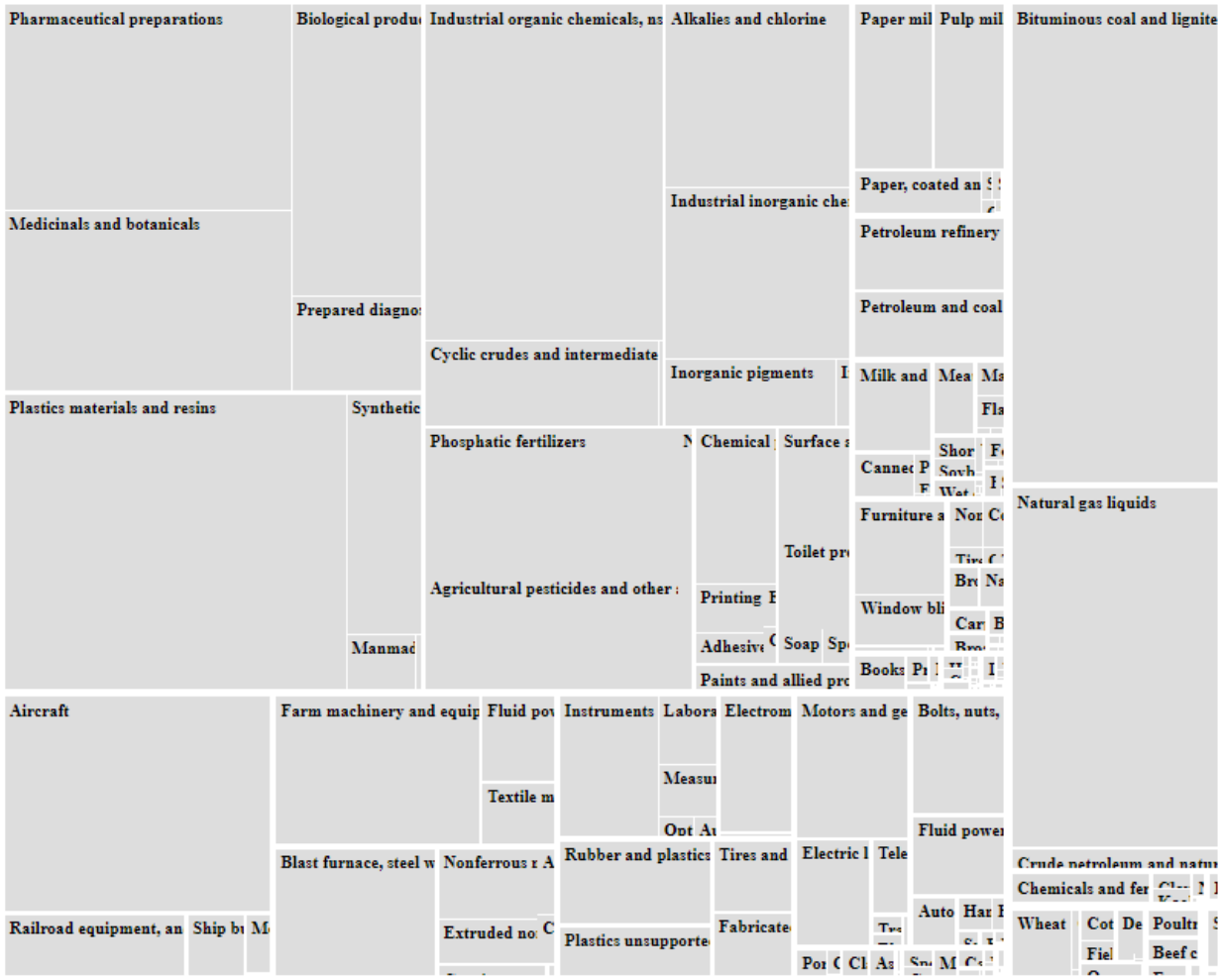
Breakdown of Brazilian Imports from the USA: 2009

Move this slider to look at different years. Hover over a product for more details:

Year: 2009

Move this slider to look at different aggregation levels:

Aggregation Level: 4



Evaluation:

When comparing imports and exports, you can see that Brazil exports significantly more agricultural, forestry, and fishery goods as well as mineral commodities than it imports from the US, and it imports a lot more manufactured commodities than it exports. Looking at specifically agricultural goods, Brazil mainly exports fruits and tree nuts within that category while importing a lot of wheat from the US.

The visualization functions reasonably well, but there could be several major improvements. In the treemap, there are certain categories which are extremely small to the point where it is

impossible to read what import/export they are. To address this, a zooming function could have been added so that the viewer could zoom into different categories to get a better look at exactly which products made up that category. Additionally, all of the categories are currently the same color which could make it difficult for a viewer to find a particular category they were interested in. Adding in different colors for different categories would make it a lot easier for viewers to tell the difference between different goods. Another minor issue is that the dataset that we used had some abbreviations which are not easily interpreted. The data could have been cleaned up prior to building the visualization so that all labels made sense.

N’yoma Diamond:

Exploratory Data Analysis:

In building the stacked area chart, the first thought was to construct a scatter plot to look at trends over time with all the data. The problem with this was that the data has different levels of aggregation available and combining every entry (of which the aggregation levels vary) into one visualization ends up being more detrimental toward the visualization than helping it. That said, this helped in identifying that it would be useful to make it so you could look at a sector’s performance over time broken down into its components and potentially compare imports with exports.

Design Evolution:

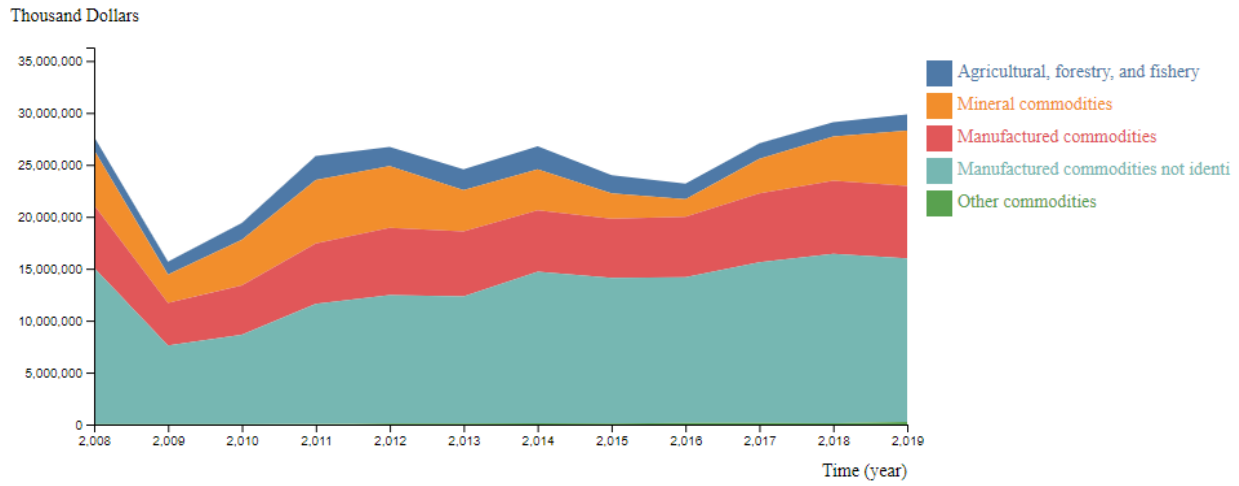
As stated earlier, we started with a scatter plot with all the data, but this ended up being more detrimental than helpful. After that first attempt using scatter plots we quickly transitioned to a stacked area chart, as we felt not only could we get a better idea of how different industries and product types relate to each other, but also provide the user with the ability to easily drill down into specific product categories and sectors. This stayed nicely in line with our proposal as in the end it was functionally the same as the scatter plot and much better from a practicality standpoint with regards to showing trends over time.

Implementation:

Part of the goal with the stacked area chart is to allow the viewer to quickly build an understanding of how Brazil trades with the United States at a glance. This was handled using intuitive interaction. Hovering over the legend or the graph highlights the corresponding data, making it easier to see and interpret; if a category can be drilled down into the cursor changes to a pointer, can be drilled into be left clicking on the corresponding category, and can be drilled back up from by right clicking anywhere on the graph; and an exports/imports toggle is available allowing the user to toggle between looking at import data and export data with a smooth transition, allowing for comparison of imports versus exports.

Brazilian Exports to the USA: Overall

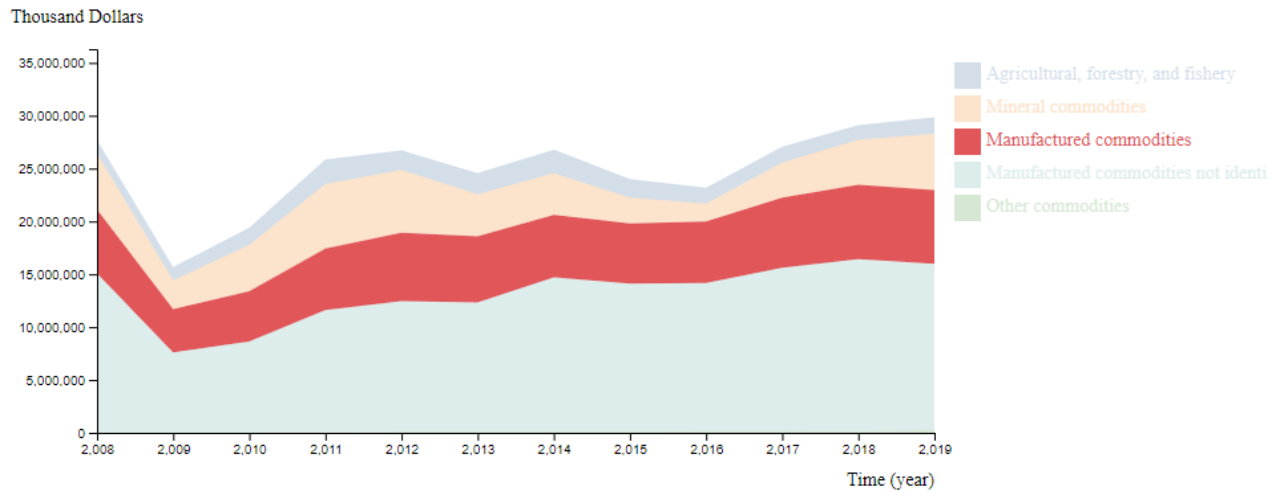
Left click to drill down. Right click to go back to previous aggregation level



Imports

Brazilian Exports to the USA: Overall

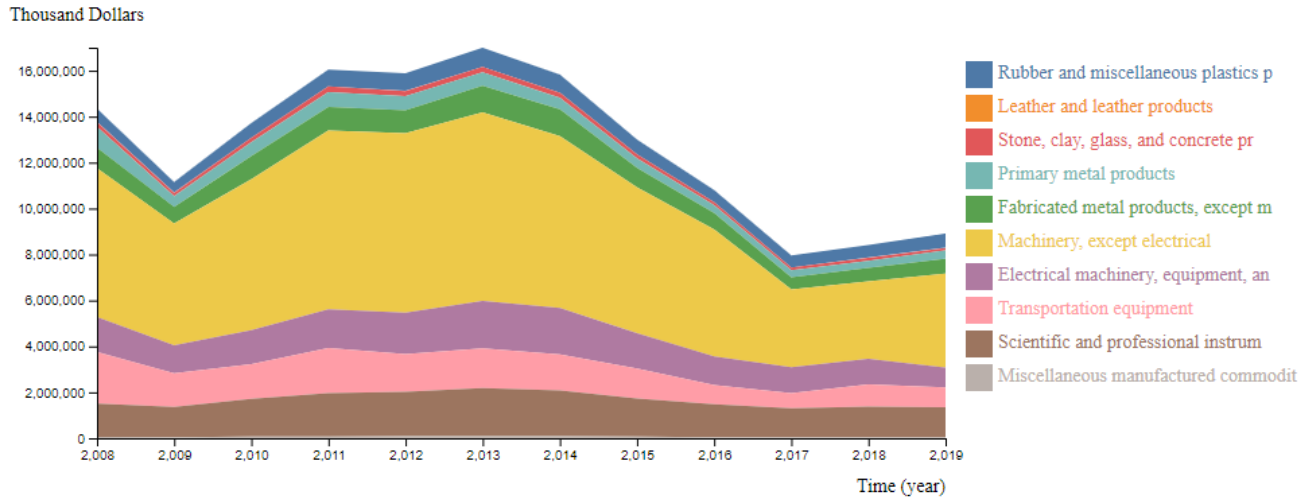
Left click to drill down. Right click to go back to previous aggregation level



Imports

Brazilian Imports from the USA: Manufactured commodities not identi

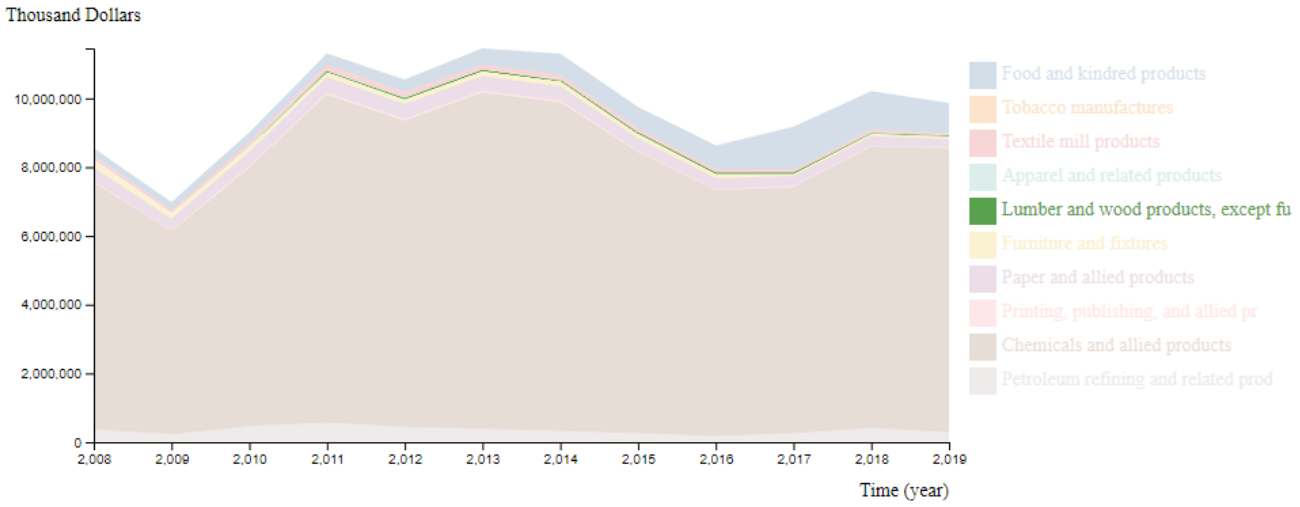
Left click to drill down. Right click to go back to previous aggregation level



Exports

Brazilian Imports from the USA: Manufactured commodities

Left click to drill down. Right click to go back to previous aggregation level



Exports

Evaluation:

Something that was really interesting to see is that in some sectors there is a vast difference between imports and exports of certain products; for example for products in the “Chemicals and allied products” category Brazil imports anywhere from 6 to 10 billion dollars of goods per year while only exporting roughly 1 billion. This is reasonably expected as some companies

inherently produce more of certain products than others and as such will export more of what they have a quality surplus of; the reason this is interesting is that for some products the difference may be an order of magnitude while in others its only a small difference. The other interesting thing to see is that despite there being large differences in import versus export values on the individual product or product category level, the overall trade balance between the US and Brazil is relatively even.

Overall we are pretty happy with this visualization. We got most of what we wanted out of it and if possible it would be very easy to expand just by providing larger datasets. We believe the interaction is clean and intuitive and it is easy to get a grasp on the data using this visual.

This said, there are also some things we feel could reasonably be improved about this visualization: The first thing that we would like to improve would be transitions when drilling up or down. As it stands the code is not built in a way that would allow for transitions while drilling up or down, but it is theoretically possible. The second thing would be adding tooltips so that the user could hover over a section of the chart and see exact values for the year and category they are hovering over. The third thing that could be improved would be data granularity and zooming: As it stands while we have a lot of granular data with regard to product categories, the data used for this visualization is only yearly and only tracks the last roughly 10 years. If more data could be acquired it might be possible to build a method for panning across and zooming into the data so the user could look at both long term and short term trends in a more meaningful manner.