Visualizing Complexity

Data Visualization (SARC 5400)

Professor Eric Field

9 May 2023

Escaping Chartsville

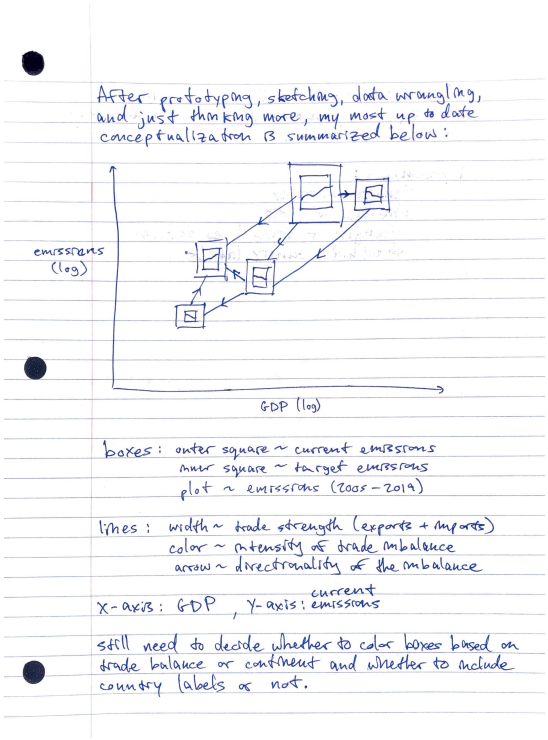
Throughout the course of this semester, I have struggled to break free from the plotting and charting mentality which has been so deeply ingrained in me during my data science career by aptly named libraries such as ggplot and Matplotlib. While these tools are useful in some contexts, they constrain the breadth and depth of expression which the good data visualizations we have observed since day one have striven to achieve. Through the past three major assignments of "Me, graphically", "Visual Data Analysis", and "Between Things", I have attempted to challenge the conventions upon which I have come to rely so regularly and progress towards something which excelled beyond the rudimentary and mundane, but, as has been noted in the feedback and critiques I received from the teaching staff, I was still failing to produce anything worthy of the title of being a true data visualization. Thus, when the theme of 'trust' was announced for the final project, I knew that I wanted to explore a question which would enable me to finally expand my horizons fully and prove to myself that I could create something both rich in content and striking in design. I chose to orient myself around the general topic of climate change, and, after contemplating how best to approach this area from the angle of trust, I decided to focus on CO2 emissions accountability.

At this point, however, I still did not have a well formulated question or clear notion of what data or visual elements I would need to make my project come to life. After reflecting on my past experiences and areas of interest, I realized that a very natural approach would be to pursue this analysis from a financial perspective. This was something which would fit nicely with my previous studies of economics as an undergraduate, and, by using international trade relationships as a means to show the interconnectivity of nations and the potential avenues for maintaining accountability, this opened up an exciting new graphical dimension which would add a valuable layer of complexity to my project design. I will refer one back to my submission for "Ugly Sketch & Prototype" in order to get a full glimpse into my preliminary

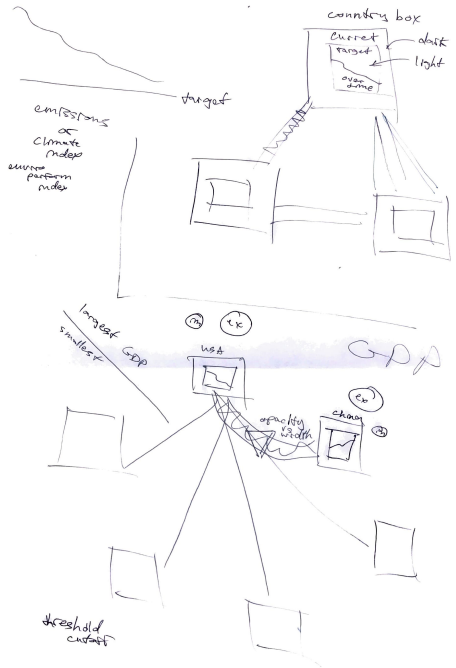
ideas surrounding the spatial organization, aesthetic mapping strategy, implementation, and conceptualization of this visualization, but the primary takeaway is that I knew the question I wanted to pursue was, "Which trade partners and economic allies can be trusted and relied upon to follow and adhere to their formally declared emissions commitments, and which countries are in a position to leverage their economic weight and prowess to positively incentivize lagging countries to improve their practices?" Although it would have been extremely convenient if a dataset already existed which was robust enough to address the subject matter at hand, given the complicated nature of this question, I ended up having to create my own datasets derived from fourteen different data tables pulled from the World Bank's DataBank. The first data frame I created consisted of the GDP, imports, exports, and CO2 emissions since 2005 for each of the top ten countries in terms of GDP. The second dataset I needed to construct would then contain the bilateral trade volumes for each of the top ten nations in terms of GDP. This notion of an international trade matrix has been discussed for decades, and a clear description of one can be found in this Federal Reserve article from 1972. Unfortunately, since international trade is a highly complex area of study that generally has far more levels of nuance that include breakdowns by specific industry and sector, there is no regularly documented and maintained world trade matrix that just measures the aggregate exports from one country to another. Therefore, I had to pull data from each country's individual trade partnership Excel spreadsheet and combine them all into a since reorganized matrix using pandas.

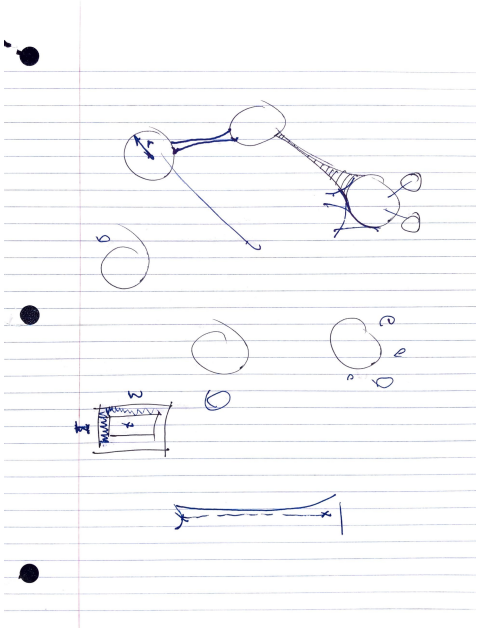
Once the data had been wrangled and processed into the desired form, I then began my visualization journey. My earliest sketches contained within my previous submission demonstrate the fact that I was still having trouble relinquishing my predisposition towards plotting as most of them are organized on a rudimentary Cartesian coordinate system and are therefore greatly constrained in their ability to encapsulate and display more unique and interesting relationships. I did come up with a neat concept fairly early on of stacking elements representing the current and target emissions on top of one another to show the gap between the two values in a similar manner to the rings we discussed in class and also reminiscent of the layout in "Land guzzlers," and this organizational scheme then allowed for

the interior to be used for what I initially thought would be some sort of mini lineplot tracking the emissions over time and for the color to indicate some other variables which I had at first thought to use for the continent. The sketch embedded to the right was the most updated version of my design going into the first peer review for this project.

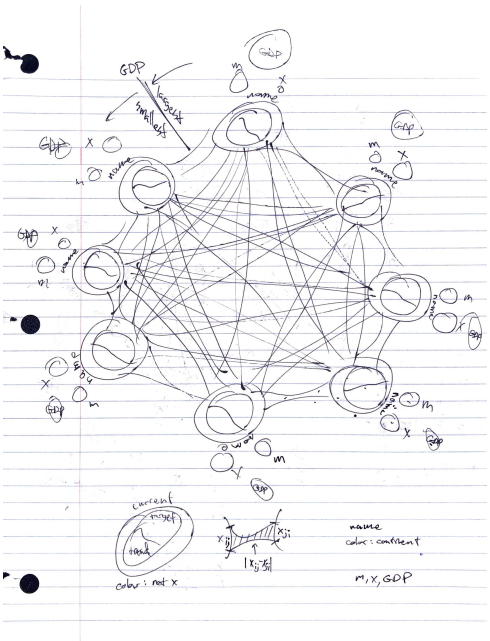


While taking a shower shortly before class that evening, however, I had my first major breakthrough idea; rather than relying on two linear axes to provide the primary framework for the visualization, if I were to organize the CO2 emissions elements around a radial axis and position them ordinally by GDP, this would then allow me to potentially add more layers of components around the perimeter if needed and would create an interior canvas upon which the international trade connections could be displayed. This latter point was a vital consideration since one of the primary issues I was facing at that point was the fact that any set of Cartesian axes would result in edges that both intersected each other and, more problematically, the other graphical nodes. By creating an interior region of the visualization which would be clear of all elements besides the trade network, this would allow for the network component of the visual to be analyzed and assessed far more easily and would put it front and center from a viewer perspective. I discussed this new idea during the live, in-person peer review session that evening, and through that process produced the adjacent iteration.





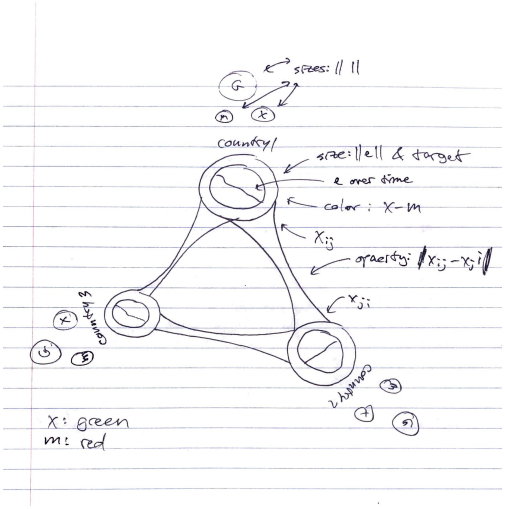
This was the first time I had attempted to display the information and data in this manner, and, through my discussions that evening, I came to a few more vital realizations. One aspect I had been grappling with at this stage was how to properly address the directionality of potentially imbalanced trade relationships. My first instinct was to simply use an arrow symbol to indicate the direction of the imbalance and to then color the line by the degree of imbalance, but, when one of my peers suggested the idea of having the width of the edge connecting the two country nodes vary such that the end widths represented the exports leaving a given country, I knew that this was a brilliant and elegant way of achieving my goal since the directionality could be visually inferred and the thickness of the line would still represent the aggregate strength of the relationship. Color and opacity were still aesthetics I had not fully resolved at that moment, but I immediately began reconceptualizing my overall design, and, by the time I left Campbell Hall that evening, I had resolved that the neighboring sketch I drew with annotations from Professor Field would be the overarching framework for my visualization.

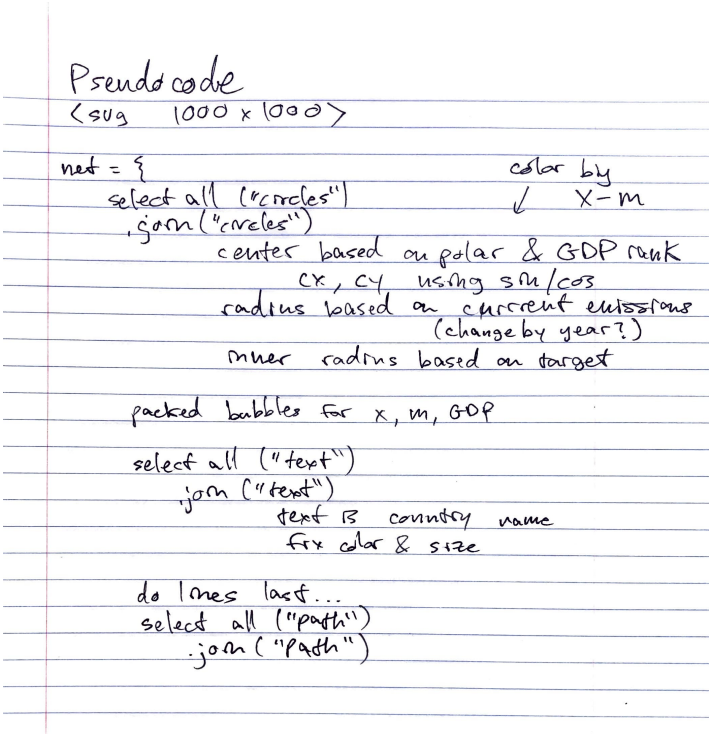


Shortly thereafter, I elaborated a bit more on this draft to produce the more detailed version displayed on the left. What this sketch showed was a radially organized scheme where the international trade relationships were laid out on the interior, the nodes around the perimeter would be the CO2 emissions elements around which the economic indicators would cluster. There was little formal mathematical description of the elements at this point, and I was a bit trepidatious about how to handle the varying width network lines, but, within 24 hours of my discussion

with Professor Field, he shared this Observable notebook with me which outlined how such an object could be achieved through the use of a filled Bezier path curve. This afforded me the confidence to proceed down this path with less uncertainty and allowed me to turn my attention to better organizing the node elements responsible for representing the CO2 emissions and economic indicators.

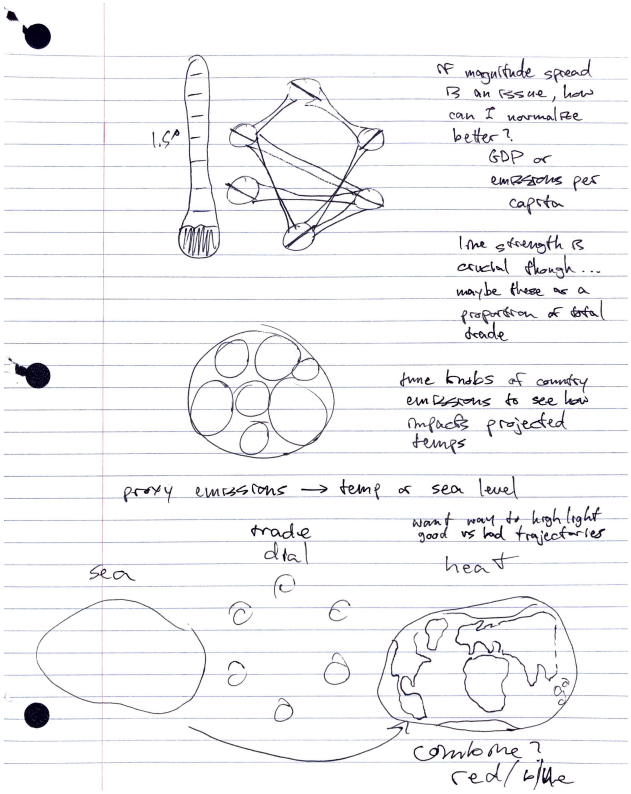
Once I had reached this iteration, there was a clear distinction between the three classes of elements forming, as shown by their loosely defined keys at the bottom of the sketch above, but there was not much clarity on how the economic indicators should be arranged and how they should relate to the CO2 emissions components. I began producing a series of sketches, like the one displayed here on the right, that stripped the various pieces of my visualization down to their base forms, and I started to feel a much stronger sense of certainty about the aesthetic





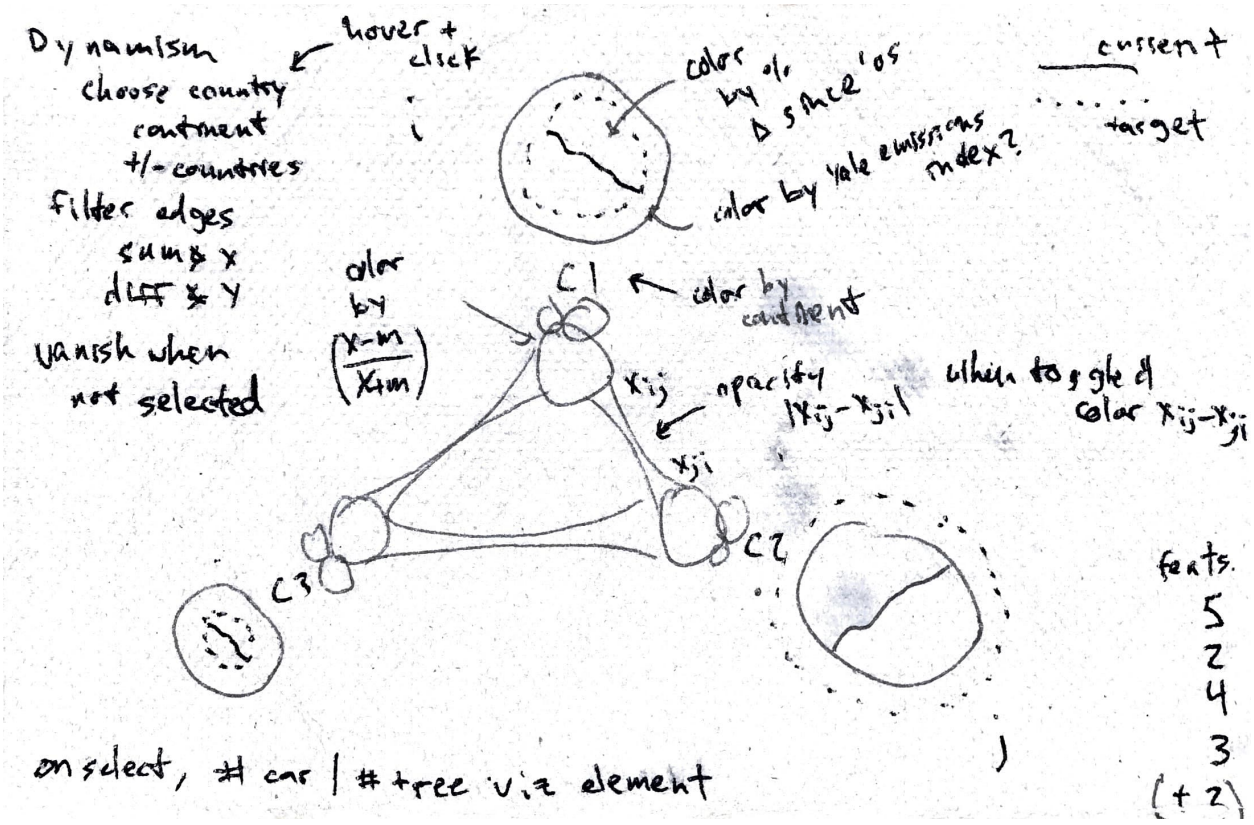
mappings which would be incorporated in its final form. I also had a clear enough plan in mind that I even started writing some basic pseudocode in order to start priming myself for figuring out how it would eventually be implemented in Observable. The example on the left clearly shows how I was mentally clustering these three crucial components and had begun laying them out logically in code space.

While I continued to churn through various ways in which I might be able to spatially and aesthetically organize the elements of my visualization, I also began considering more thoroughly how dynamics might play a role in my interactive application and how I could really emphasize the "So what?" takeaway aspect of my project. I knew from the start that I wanted users to be able to select a specific country and see its trade ties to other nations and how the countries were bound together from an economic perspective, but, since the broad topic I had initially decided to try to address was climate change, many of my initial thoughts on how to represent the environmental effects were to try and show how potential changes in CO2 emissions could impact global temperatures or, even more specifically, how sea levels and more local regional climates might be impacted. A few mock-ups of my very-early-stage thoughts on this matter can be seen on the right, but I quickly realized two issues that would render these exact configurations infeasible. The first was that there is no objective function that can directly map CO2 emissions to specific global climate impacts in such a trivial way, and the second was that, even though temperature is a more physical and understandable metric by which to interpret impact, it was still a bit abstract and did not feel as directly relevant as showing a value benchmarked against a more personal and tangible unit of measurement. After thinking about it a bit more and finding some inspiration during a peer review I was providing on another classmates' "Ugly Sketch & Prototype" submission, I began to formulate the idea of trying to measure the % change in CO2 emissions by a given country in terms of the equivalent number of automobiles' CO2 emissions or the equivalent number of trees' CO2 absorption. What I liked about this approach was that it would allow me to finally get at the



heart of the problem and demonstrate why even just small changes in emissions resulting from partner countries' trade incentives would be immensely valuable. I also enjoyed the fact that it would be a clever play on the 33,636 Gun Deaths Visualized Medium article we saw in class.

At this point, my final conceptualization was beginning to take shape and crystalize, but it wasn't until a solo dinner at Runk where I was struck with a series of epiphanies that I felt truly confident in the full model design for my visualization. The ideas formed during that meal resulted in the last sketch I did before starting to develop a codebase in Observable and allowed me to finally resolve many of the small details with which I had still been grappling at that time. The sequence of revelations came in quick succession, and each addressed a slightly different dilemma I had been having. The first realization I had was that the trade network edges should really be connected to the economic indicator elements (and GDP in particular) rather than the CO2 emissions ones since these were really the financial nodes linking the nations. This would then allow the CO2 emissions elements to orbit around the perimeter and be viewed and interpreted almost as a response variable to the economic interactions occurring within the network. The second idea born from this meal was that, rather than layering two filled circles on top of one another for the current versus target CO2 emissions levels, by using different line styles such as filled or dashed or dotted, I could then avoid any concerns over stacking the elements in the correct order when plotting and could also potentially incorporate additional concentric rings if needed. In order to add more layers of complexity to the CO2 emissions elements, I also began toying with the notion of incorporating multiple color scales for the rings versus the central object to then be able to map both percentage change in CO2 emissions and another value such as the Environmental Performance Index score from Yale simultaneously. In addition to the static organizational strategy, I also gained a clearer sense of how I wanted to approach the dynamic elements. That Runk meal was by far the most significant design session of this entire project, and, by the end of it, I had developed the sketch seen below on a random napkin and felt ready to embark on the journey to implement this conceptualized visualization in Observable.



Thus began my arduous voyage to successfully create from nothing a compelling and beautiful visualization in a foreign and unfamiliar coding environment. For the sake of brevity and avoiding tedium, I will refrain from delving too deeply into the details of my technical implementation, but I will say that I am greatly appreciative of Professor Field for his tolerance and courtesy with regards to my 2am debugging emails and for providing constant support and guidance while working to get the dynamism of this visualization off the ground. In all honesty, most of the elements used for the CO2 emissions and economic indicators components were fairly easy to produce procedurally with just a little trigonometry and some clever translations to ensure exact consistency of spacing and relations throughout. There are a few interesting aspects of these two element types that I will note, however. The first is that, while finalizing the design of the CO2 emissions element, I decided to modify it to now include three rings rather than two: the dotted one represents the 2005 levels, the solid one represents the current levels, and the dashed one represents the target levels. This would then allow viewers to quickly discern how the volume of emissions have been progressing since the Paris Agreement was made, but, in order to emphasize the notion of change, I also

derived a new feature for the % change in emissions since 2005. Mapping this feature to the slope of the trend line was achieved by letting the angle of the rotation transformation be equal to the result of applying a scaled hyperbolic tangent function to the % change, and, in order to doubly encode this salient attribute, I also used a linear color scale from blue to red to indicate whether the emissions had decreased or increased, respectively, in the past decade and a half. (Note that the specific function used to perform the mapping of % change to degree of rotation was 90*tanh() since this is bounded on the range (-90, 90), thus ensuring that, for any % change no matter how large, the resulting rotation would be asymptotically correct.) By organizing the trend bar and rings in this way, it now freed up the center of the element to be used for other purposes, so, for this I chose to use polygon symbols to represent the categorical variable of continent and the color scale from green to red to show the Environmental Performance Index score from 100 to 0, respectively. (Note that the size or area of the polygon is not directly related to a specific feature in the dataset, but, rather, the distance from the centroid of the regular polygon to the circle circumscribing it is set equal to 0.9 times the minimum of the three radii associated with the rings, thus implicitly defining the polygon's area.) In this way, I was able to elegantly and efficiently package together 6 distinct features into one simple and intuitive graphical element. One further aspect to note is that, while the area of the rings is the primary aesthetic being influenced by the magnitude of emissions, the linewidth of the rings is also a function of the emissions in order to make it so that even the lower emissions countries will have designs that are fully legible and distinguishable.

The second characteristic of these two component types that became pertinent to focus upon and address was that of radial versus vertical interpretive perspective. Since my visualization is laid out in a circular form, I worried that viewers might initially be inclined to assume that nodes for each country should be perceived and interpreted from a radial perspective with the center of the design as the origin. To mitigate the chances of this confusion occurring, I attempted to remove any potentially misleading instances of rotational symmetry. This presented in two fashions: the first being to place the county name labels below their respective emissions elements rather than having them be rotated in a manner

that caused them to be facing the center of the visualization, and the second being to maintain a constant horizontal alignment of the economic indicator circles since this would have posed the same issue if rotated based on their position around the visual. I had tried a number of different strategies for organizing and arranging the circles for imports, exports, and GDP for each country, but, once I landed on the horizontal layout, I felt confident with the result. This was for a number of reasons, but the primary motivators were the establishment of the vertical orientation as previously discussed and also the fact that, by placing them in this way and in this order, the imports and exports nodes would now act to delimit the endpoints of the color scale which was being used to define the color of the GDP circle bounded between them, thus producing a clear visual spectrum within which the net exports mapping would clearly fall. Having the exports and imports on opposite sides like this did reduce the ability to directly compare their areas which were determined by their magnitudes in US$, but, since the color of the GDP node between them already corresponded to this difference, I did not feel as though this posed too many issues for interpretability. By having the GDP node centrally placed in this way, I was also able to use these elements to define the entire framework and radius of the overall visualization since these were acting as the nodes to which the trade network was attached. This design configuration proved ideal for the reasons just discussed, and, overall, the element ended up being able to easily capture and represent another four features in a highly straightforward manner.

With these two types of components in place, it was then time to design and construct the international trade network. I will refer one to the makeWeb() function within my Observable notebook for this project for the specifics of how I was able to eventually fully implement this intricate and complex element in D3, but the high level overview is that all diagonals of a regular decagon can be computed based on the radius of the circle circumscribing it and also that all angles between vertices can be procedurally calculated so long as you track the index of the current node and those of all other nodes correctly. Even with the aforementioned notebook from Professor Field demonstrating how filled Bezier curve paths could be used to created thickened lines with varying widths, the implementation of this web creation process was still a highly technical and convoluted undertaking, and

configuring it to filter and change color dynamically for the interactive components was a behemoth of a task, but I am extremely proud of how beautifully it functions and behaves at this point, and I felt as though this part of my project was one of the most rewarding and fulfilling since it truly gave me the sense of creating something novel and inventive through the application of elegant mathematics and geometry in conjunction with the valuable technical tools of D3 and Javascript. From a visual perspective, I also found the final form of this network to be highly captivating, and I really appreciated how the convex curvature of the paths along with the varying opacities acted to both provide more texture to this component of the visual and also create a visceral sense of tension between the nations which I found quite appealing and compelling. Once this final fixed element was in place with the originating linewidths representing the exports from one country to the other, the opacity expressing the degree of imbalance of trade between the connected countries, and the color determined by the direction of the trade imbalance, I now had fifteen active features in total captured in one, single, comprehensive, complicated, and beautiful visualization. This was a feat of which I am immensely proud as one of the primary comments and critiques I had received throughout the semester was that my visualizations were not expressive enough and did not involve enough variables to be considered engaging, revealing, or compelling. In addition to feedback regarding the depth of my visualizations, I was also encouraged to explore more nuanced and innovative aesthetic mappings such as opacity and shape and both line width and style, and, in this design, I managed to incorporate all of these in ways that added substance to the meaning of my visualization without simply adding features and adornments superfluously. (While I have managed to touch on the most salient aspects of the visualization, there are a number of more technical details which I was not able to discuss fully, so, if there remain any outstanding questions or a need for an exact schematic of the measurements used to construct this visualization, feel free to reach out via email. My UVA address is kzj5qw@virginia.edu.)

Once all of the static components had been established and implemented successfully in code, I proceeded to layer on the additional dynamic and interactive features for the web application version of the visualization. This included scrolling to zoom, dragging

to pan, hovering to see tooltips which displayed all salient pieces of data for a given component type, and selecting a given country to display its specific international economic relationships and the directionality associated with each flow of trade to its partner nations. Upon selecting a country, I also created a graphical element which would display the equivalent environmental impact of a 1% reduction in the selected country's current level of CO2 emissions in terms of the number of average U.S. cars whose total volume of emissions to which this would correspond. To avoid latency issues or strange scrolling behavior when this element appears, I made it so that each circle rendered corresponds to either 10,000 or 100,000 average U.S. cars depending on how much 1% of current CO2 emissions equated to and also set it up so that the actual dimensions of the SVG object would expand or contract just enough to fit the new visuals being generated. These dynamics allowed for a number of enlightening discoveries which are discussed in my static visual and relate to the four countries of the United States, China, India, and the United Kingdom, but I will include those comments here as well for consolidation purposes.

The United States: With the largest economy in the world, the U.S. has a great deal of international influence. Although its emissions and environmental policies have been trending in the right direction, it is still the second biggest producer of CO2 emissions, and being a net importer with nearly every nation reduces its ability to incentivize other wealthy nations to change their practices through trade agreements alone.

China: Despite ranking second in terms of GDP, China is by far the largest generator of CO2 emissions on the planet. This is primarily due to the suboptimal environmental policies in place and the economy's reliance on the impressive production of goods at a massive scale. This focus on production has caused China to have remarkably high exports, but the potential economic sway it could have as a net exporter to nearly all of its trade partners does not seem as though it will be realized since China itself is already the worst emissions offender and has even set a 2030 target volume that is above its current levels.

India: Although this South Asian nation does not currently have the highest volume of CO2 emissions, it has experienced the largest % increase since 2005 (of over 100%)

and has blatantly ignored its target commitments. It also has one of the fastest growing economies, so this trend is likely to continue unless its net importer partner countries intervene through economic sanctions.

The United Kingdom: Leading the way both in terms of its environmental policies and its reduction in CO2 emissions since 2005, the U.K. has demonstrated exemplary behavior with regards to trying to improve the well-being of the planet. Its fellow European nations have followed a similar pattern, but, since these nations tend to be net importers with smaller overall economies than countries such as the U.S., China, and India, their ability to construct and impose trade-based environmental incentives is diminished slightly. Despite this limitation, other nations can still look to them for guidance on policies.

As important as my main takeaways, however, was the fact that these interactive elements allowed the functionality to come to life and helped to create a tool with which users could fully engage and use to explore their own questions with regards to the datasets I composed and created. While my specific question related to CO2 emissions accountability within the larger landscape of international trade relationships, this tool is robust enough in capability to support users' interests in a broader range of topics related to international relationships and will hopefully afford them the flexibility to explore and uncover some connections within and between entities which I had not even fully considered. (Additionally, since I had to compile this dataset manually, I was limited in the number of countries I could include, but there is no reason this structure could not scale to accommodate any number of nations so long as the formulas within the makeWeb() function are updated accordingly. For the scope of this project, I think limiting it to the countries with the 10 largest economies in the world was quite suitable and sufficient since these do produce the vast majority of the world's CO2 emissions, but, if this were to be extended further to include a larger pool of countries, that could create an even more sophisticated and impressive tool that could be used to analyze a larger segment of the global trade network and international supply chain in a fairly unique and novel way.)

Through this visualization, I also hoped to pay homage to some of the exemplary visualizations we have seen over the course of the semester, with the dotted and dashed rings orbiting the central polygons in the CO2 emissions elements evoking notes of Kepler's planets, and the imports, exports, and GDP circles being organized in such a horizontally aligned and tangential manner as to allude to the embedded diagrams of Saturn's moons in Galileo's notebooks. As we were challenged to aspire to be the best of all that we have seen, I also reflected back to the very beginning of the term and some of the qualities to which I found myself drawn in "the good, the bad, and the ugly." For my "good" visualization, I had selected a plot which attempted to reimagine a map of the world where the location of the countries and resulting clusters were determined by the relatively recent and mildly advanced method of t-distributed stochastic neighbor embedding (t-SNE). While, in retrospect, this visual is nothing more than a scatterplot with a cool embedding being used to determine the Cartesian coordinates, I still feel very much compelled by the idea that our perception and organization of the world can and should be redefined through the extraction and application of data driven methods. I strove to honor this idea of capturing and representing the relationships and interactions between nations in an inventive and novel way, but, in truth, the visualization I had selected as my "ugly" example was more meaningful for my growth than almost any of the "good" examples I have seen, and, in fact, the idea of plotting a trend bar rather than some sort of mini lineplot of emissions over the years was directly inspired by this visualization. The excerpt below contains the closing remarks I made regarding the impression this visualization left on me, and I feel as though these sentiments are what aided in allowing me to finally relinquish my dependency on the conventional plotting and charting techniques which plague the field of data science.

> "I found this plot to be the most fascinating and horrifyingly clever I have discovered, and, for that reason, I had to choose it as the 'ugliest,' although something tells me that the line between 'ugly' and 'beautiful' is very thin and potentially blurry, and, having spent a decent amount of time looking at this plot design at this point, I wonder whether there may be other non-standard and 'ugly' plot formats that might have some value in exploring some time down the line…"

As I suggested here, the line between horrifyingly clever and elegantly sophisticated can be quite thin and blurry at times, but I truly gave it my all to try to push my creative boundaries and discover this frontier through my final project. By embracing the experimental and untouched landscape of Observable where no built-in or established tools existed to artificially define the scope of or constraints on my visualization, I finally felt as though I was able to uproot the plotting mentality which has been so deeply implanted in our field of data science. Although there is always more work that can be done and areas which can be polished, I am immensely grateful for the opportunity this class has afforded me to try and produce seamless, intuitive, compelling, striking, engaging, and rich visualizations, and, thanks to this course and the incredibly supportive, critical, and passionate teaching staff, I feel as though I was able to begin finally taking my first steps towards escaping Flatland, or, as I prefer to think of it, Chartsville.