GEOG 407

April 17, 2016

Team: The New Coders (Hanna Demyk, Nhattan Nguyen, and Elliot Tan)

**Assignment 5: Conduct data journalism using D3JS in the cloud**

**Important Note:** D3JS used in our assignment does not work with Chrome. Please start at index.html

A. Choosing the topic and visualization

    Some of our first thoughts revolved around the panama papers leaks. However, as the issue had just surfaced, we figured data on the leaks were still being uncovered and therefore limited. Our first real topic consideration was the Syrian refugee crisis. We were a bit hesitant at first as other teams had considered this topic, but after discussing their plans, we decided to explore the topic further as we would be working with different data. We narrowed down the topic to Syrian refugee flows; where were Syrian refugees going? This topic turned out to be very difficult and surprisingly in the first few days we could not find any data specifically on that topic. We then tentatively considered another topic: the demographic transition around the world as we knew government data existed on this despite it being a more boring topic. Finally however, we managed to dig out a Vox article which referenced UNHCR data that contained the numbers we were looking for.

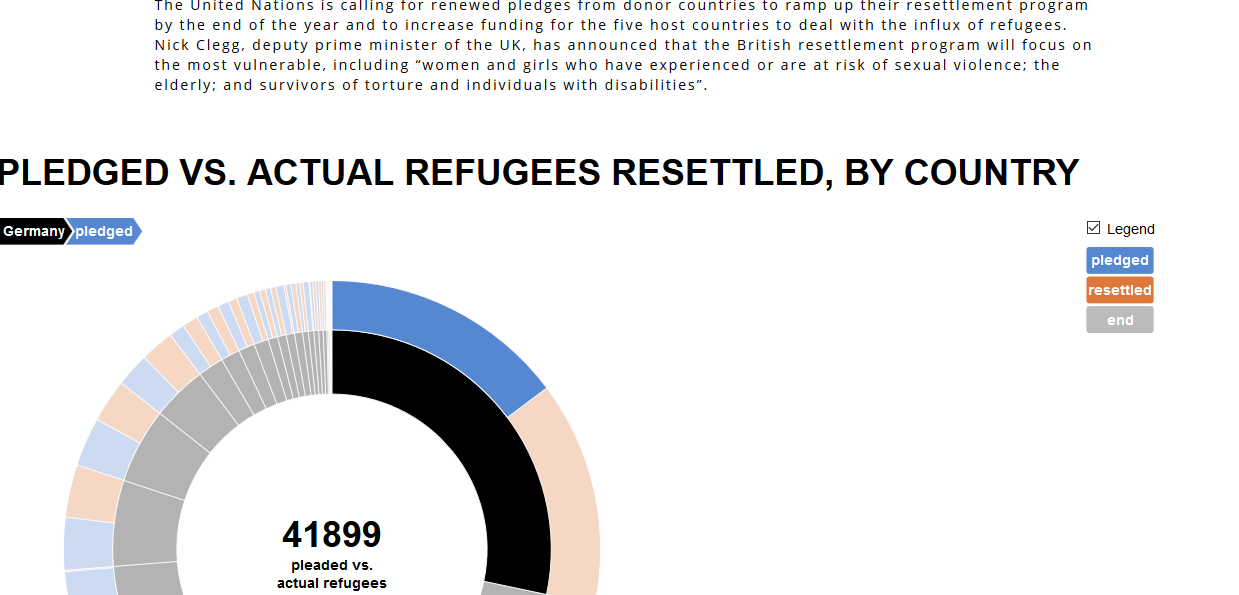
    We then proceeded to consider a visualization. Indeed, there was no point in choosing a visualization first as our choice of visualization would depend heavily on the data we had. We went through the D3JS website to look at different examples, with a preference for interactive ones, and then selected the sunburst as explained in the following sections.

B. Visualizations Considered

* **Global Arcs**: when we first thought of the topic of the refugee flows, we had in mind a visualization illustrating the flow of refugees from Syria to other countries on the globe. However, this was essentially a cartographic-based visualization so we disqualified it. Example here: <http://bl.ocks.org/dwtkns/4973620>
* **Bar chart**: We had also considered a bar chart to illustrate the difference in numbers. However, we couldn’t find any that were visually appealing. Example here: <http://bl.ocks.org/mbostock/3885705>
* **Sunburst**: We finally chose the sunburst. We considered a sequence sunburst (<http://bl.ocks.org/kerryrodden/7090426>) and a zoomable sunburst (<https://www.jasondavies.com/coffee-wheel/>). We found the sequence sunburst to be more straightforward to work with for our purposes.

C. Visualization choice

    We first proceeded by elimination, considering the criteria of interactiveness and non-cartographic-based visualizations. The bar chart templates we found did not satisfactorily differentiate pledges both between countries and then the pledges vs. actual resettled refugee numbers within countries. On the other hand, the sunburst was able to do that as it had multiple ring layers that we could each use to display different information. A sunburst is like a super pie chart (because Dipto doesn’t like pie charts). Consequently, this makes the sunburst superior to a cartographic approach as it allows us to better visualize differences between countries and the discrepancies of numbers within a country. We also wanted to emphasize how important the numbers were, and a cartographic approach would have distracted the focus from how many refugees countries were resettling to how close countries were to Syria.



**Figure 1.** Screenshot of webpage with visualization

Annex 1 - Hanna’s reflection

1. **Most difficult step:** The most difficult step of this process was finding a visualization to suit our needs and display the data we wanted to convey in a meaningful way. Current open-source D3JS “templates” available from websites are geared toward specific datasets and are usually not one-fits-all in their capabilities. We needed our dataset to show even the smallest numbers in a visible way, and the D3JS visualization we chose was the only one that was able to permit a user to highlight over everything, even the thinnest sliver in the chart, to see the number of pledged/resettled Syrian refugees by country. Bar charts would completely hide these small numbers, so this “sunburst” visualization was our best option for the task at hand. Even so, the chart shows irrelevant numbers on the first inner-most ring, as the number shown is pledged and resettled added together. Unfortunately, despite digging through the code, this was not something we had the capacity to change. That is why finding the right visualization turned out to be the most difficult step in this assignment. Even when the best is found, it is hard to overcome errors when incorporating your own data and customizing the D3JS “template”.
2. **Improvements to assignment:** This assignment was well-structured and creating a non-cartographic visualization was interesting, considering the rest of the semester was very cartographic-based. I would potentially add the option to use a cartographic-based visualization with justification as to why it was better over other options for a team’s topic and data. There isn’t anything I would leave out from the instructions given, as they were clear but also very open in terms of what we can do with our visualization. The only thing is, incorporating D3JS into a webpage that has its own set of styling can cause issues, so I would suggest leaving out the website necessity in order to allow more time to work with the visualization.
3. **Working with visualizations:** D3JS is easy to use after completing previous assignments that used often more complex coding you had to sort through. The problem with D3JS right now is that it is fairly new and visualization examples have not yet been made for easy re-use in projects aside from original use by the author. This makes it hard to incorporate your data with someone else’s open-source D3JS template. The strengths of D3JS are that the visualizations are incredibly dynamic, the possibilities are vast, and you are not stuck using Google Maps or any other standard-looking 2D map. D3JS feels similar to Google Earth in its 3D capabilities. Map APIs we used for previous assignments were useful in showing specific geographic information for points, polygons or line segments. You are, however, limited in how you can display your data, by using balloons, overlays, or similar. D3JS allows you to really bring your data to life in more ways than mapping techniques used before, but you have to learn a whole new set of code to truly open D3JS’s potential and build your own template. Starting from scratch in D3JS would bring forward a lot of strengths in using this visualization technique over others we have used.

Annex 2 - Nhattan’s reflection

1. **The most difficult step**: The most difficult step was figuring how the D3JS coding worked together and with the data. Despite having a template, the instructions provided were limited so it took some playing around to set the example up to work. It then became challenging how to get the code to extract the data from the dataset, and that also required playing around and searching through the code, as well as tweaking the way we worded terms in the dataset. However, surprisingly, the most frustrating step turned out to be looking for the dataset we wanted. We had our hearts set to looking at Syrian refugee resettlement flows. For the first few days it was near impossible to find the data. We found where Syrian refugees were being temporarily hosted in neighboring countries, we knew how many refugees were leaving Syria, we knew how many refugees host countries were receiving every year, but we couldn’t find how many Syrian refugees there were exactly in each country. UNHCR data appeared to be limited. It was quite odd as we didn’t have the data we wanted for our story. Finally, after doing visualization searches, we stumbled on a Vox article which had that data sourced from UNHCR. We also found a UNHCR report documenting resettlement pledges from similar countries.
2. **Improving the assignment**: Overall, I think this was a really well structured and balanced assignment. Compared to previous assignments, I don’t really have anything to suggest adding or removing. The required output and time to complete was also mostly reasonable. The D3JS examples were quite limited, it would have been nice to access more open-sourced examples. We also had little time to practice with D3JS, so it would have good to do an exercise in class to use a D3JS visualization and integrate data into it. It would actually have been interesting to use an interactive cartographic visualization, but I’m not sure if that would be too complicated.
3. **Working with visualization**: What I found to be very interesting was the interactiveness of the visualization. This helps condense a lot of information rather than having multiple charts. Furthermore, interactivity with D3 seems to show so much potential to visualize and analyze complex data, especially when there is too much information to initially grasp when presented as a single image. Comparing visualizations like D3JS to the maps we’ve used in previous assignments, D3JS allows for much more handling of numeric data, but the downside seems to be that the model is fixed and limited. In other words, you can only play around with the dataset provided. In comparison, the map applications we’ve used allow for much more manipulation: I can drag markers around and plot out calculations from any point on the planet. There is also obviously the more straightforward ability to visualize data spatially (in cartographic sense). The downside with the map applications is that comparing with data is not as efficient and you can’t use as much data or else it will block the map.

Annex 3 - Elliot’s reflection

The most difficult step was learning how to navigate D3. Although the format was still very much like a tree (branch – branch – leaf etc.), I found that the functions themselves were at times confusing. While practicing using D3 and spreadsheets, I also found it was a bit difficult to find the correct cell I was looking for (although that may have been an issue with the spreadsheet I was using, as I constructed it pretty poorly). I also found that I was having some difficulty understanding how to use the data within the D3 functions/ displaying it at first. Finding the right visualization was also difficult, as there were so many to choose from that we felt fit our project. When we found the visualization we wanted to use, it was in a different format as previous assignments, in that the template was through an HTML file which called a Javascript and CSS file for running functions/ styling. Matching up the different variable names between files was a little confusing at first. For example, in the HTML file, to show the center number, it called #percentage from the javascript file, which had D3 functions conducted on it.

I feel that this assignment could be better improved with some examples in class working through different types of basic visualizations. It was a little overwhelming having to choose a visualization, but not knowing how it worked at a base level.  Besides that, I felt it was a fair assignment for learning how to use D3 in data visualizations.

I think that data visualization through means besides a map is extremely useful. It offers more avenues to display data in aesthetic and clear ways. For example if we used a map to display our issue, it would probably be a map with numbers on it, or bar graphs sticking out (both extremely messy). Using our method, we display geographic rooted data in a clear, and aesthetically pleasing manner. SVGs also allow much more freedom than traditional GIS software data for displaying data, in terms of aesthetics and interactivity. Data visualization methods are essentially really good for displaying complex data, that won’t look messy. D3 in itself is really interesting, as SVG has so much potential for flexible data visualizations. However I find that there is a much higher learning curve to D3 (especially the syntax), and thus it could be fairly inaccessible. Using a visualization to display geographic data is also at the mercy of the creator, especially if the data is reliant on geography, and that’s not well displayed. Overall I think that alternate data visualizations are an essential and powerful tool for any neogeographer, but it will take time to really master it.