Rachel Thorp
Carina Sauter
Info247 - Information Visualization and Presentation

D3 Assignment writeup

Links to final work:
Parallel Coordinates: https://rachelthorp.github.io/InfovizD3/
Radar Chart: http://people.ischool.berkeley.edu/~carinasauter/

Overview:
We decided to split the assignment - one person (Rachel) coding the parallel coordinates and the other one (Carina) doing the Radar Chart visualization.
While we briefly considered displaying both visualizations in one website, we found that given the different approaches we had taken to do the visualizations, there was no easy / good way to combine them. Therefore, we kept the two visualizations separate on two websites.

Data preparation:
In order to ensure that we would visualize comparable information, adjustments were made on the original dataset, and both partners used the same edited dataset to work on our visualizations.

In detail, the changes made to the original dataset were the following:
- Only data from the latest year (2016) was retained to simplify the visualizations
- Only the top 200 universities (as determined by the rank number) were retained to prevent both an overwhelming number of schools' data being displayed, and odd gaps in the parallel coordinates visualization (as schools with a rank over 200 were binned in terms of rank number and left without a specified total score).
- Formatting types were updated for easier processing—e.g., turned ratios written in colon format to percentages, and removed some unnecessary symbols (like '%', '=', and commas) in numerical entries
- All empty numerical entries were filled with the median of the corresponding column
- The *year* and *country* columns of the original data set were removed, because (1) an entry for the year of the data is no longer meaningful with only one year included in the data, and (2) we did not find the countries of the schools to be an

important component of university rankings, and retaining the values would complicate the visualization process

The data preparation was exclusively done by Rachel and shared with Carina for the coding step. The code used to clean the data and output a new csv file can be found at https://github.com/rachelthorp/InfovizD3/blob/master/data_transform.py.

Parallel Coordinates Analysis - Created by Rachel Thorp:
Borrowed Code:
In building the parallel coordinates chart, I used Jason Davies's *Parallel Coordinates* chart (https://bl.ocks.org/jasondavies/1341281) as a starting template for the structure of the chart, and I then referred to source code of a previous final project for the course by Anand Rajagopal, Anubhav Gupta, Joshua Appleman, and Juan Shishido (http://people.ischool.berkeley.edu/~japple/jeopardy/) to include line highlighting and a corresponding tooltip. In order to bring the highlighted line in front of all other lines, I referred to the source code of a chart by Mike Bostock (https://bl.ocks.org/mbostock/3709000) found online.

Modifications & Motivations:
I chose a different color palette than what was originally used in either of the two references above — one that is neutral but warm, and is easy to differentiate between the three line types (light lavender background, rosy-brown foreground, and a darker rust color for the highlighted line), while also being intuitive as to color indications (i.e., lavender is the lightest, and corresponds to those unselected lines in the background, whereas the darkest color is used to highlight the current university in question).

In referring to source code to help with implementing a tooltip and line highlighting, I had to adjust a number of corresponding components, as the starting data differed in formats — for instance, while the final project included data from 3 different sources with foreground color coding to distinguish the sources, I chose to keep the data of this project under one primary group, with just a single highlighted school at a time to compare against the rest.

I also made sure to bring the highlighted line to the front of the drawn data, as the sheer number of schools on one chart limited the visibility of most universities when highlighted. I also updated the appearance of the tool tip — or more specifically, I included a white box with a black border around the tooltip label for improved visibility, while adjusting font style and size for uniformity with the rest of the chart. Finally, I rearranged the data scales such that while all other dimensions were charted from high values (top) to low values, the school rank parameter was charted from low values (top) to low values, as lower rank numerical values tend to suggest higher numerical values in other fields.

Radar Chart Analysis - Coded by Carina Sauter:

Code Contributions:
As a scaffolding, the radar chart implementation of Nadieh Bremer was used because of its modern look, improved readability, and relatively easy adjustability to our data: http://bl.ocks.org/nbremer/6506614http://bl.ocks.org/nbremer/6506614

Nadieh Bremer's code, in turn, builds heavily on the code from Alvaro Graves - the original radar chart implementation:
https://github.com/alangrafu/radar-chart-d3

Finally, the drop down menu code was taken from Paul Glenn's section of the code for a similar assignment for this class:
https://github.com/pdglenn/UniversityVisualizationD3
This code, in turn, leverages the d3.dispatch implementation found here:
https://bl.ocks.org/mbostock/5872848 which was used to get the code to work with the data format and radar chart scaffolding I had chosen.

Code Modifications:
Most of the time coding this piece of the assignment was spent figuring out how these different pieces of code could be made to fit together and how the data could be loaded into the code in its current form. While I played around with different color choices, I finally decided not to change it because I liked the contrast of the orange and blue to clearly see which data points belonged to which school.
Additional changes were done to the thickness of the lines of the radar chart, the position of the labels of the five categories, and the opacity of the field on mouseover.
I spent some time modifying the different parts (size and position of the radar chart, size and position of the drop down menu) for the website to make it easily readable and usable.

Our process:
We decided to each code one of the visualizations separately because we were both going to spend Spring Break away from Berkeley. After Rachel shared the modified the dataset as a common base, each team member worked separately on the visualization of their assigned part. Once the coding was done, each team member shared their result for review with the other, at which point we exchanged feedback about some areas that could be improved upon; each team member then returned to their charts to make the suggested improvements (whenever possible with our limited d3 and javascript experience).

Strengths and Weaknesses:
One of the major weaknesses in our delivery is having two charts that are on two different websites that are not connected, as the information and insights gained from each cannot easily be combined. In effect, both visualizations stand for themselves.

The parallel coordinates chart is a user-friendly, intuitive tool for observing trends in the data. Using the format of a line chart where data from *all* universities are displayed at once makes it easy to see large-scale trends from one parameter to the next. There is no limit to the number of categories that can be included, which allows for a higher density of information being visualized. Furthermore, the chart interactivity allows users to (1) rearrange columns, (2) select and deselect any subset of data along any combination of the columns, and (3) highlight one school at a time, which offers more opportunities for users to explore the data in depth and reach their own inferences about patterns and trends. While there is a possible limitation in connecting inherently non-sequential data with a line across a large number of categories, this ability to rearrange categories should still make this visualization a useful and insightful tool. It should also be noted that the format of the chart is limiting in that it does not allow users to easily compared two or more specific universities against each other (unless very careful selection of data subsets is done along one or more of the columns), and a user cannot easily search for a specific university—instead, they must browse through all the lines and hope they find the name they are looking for.

The radar chart implementation is good at providing a tool that easily allows the comparison of two universities in the ranking across 5 major categories. The interactivity of the chart is fun to play with and engaging. The chart design and color choices make it easy to read and the two universities chosen and their data points clearly distinguishable. The radar chart in general does a good job in focussing one's attention on areas of difference (where one color dominates) rather than focus on areas where rankings are very similar. This helps to retrieve meaningful differences between universities quickly. However, the design also limits comparison of a few parameters - not all data of the ranking can be visualized effectively. All data needs to be in the same format (% in this case) so that means that absolute numbers cannot be included efficiently.

Overall, we feel that these two distinct visualizations complement one another in their respective strengths and weaknesses — the parallel coordinates chart may allow a user to easily visualize more information at once (both in terms of the number of schools being plotted and in the number of categories), but the radar chart allows for a closer and more concise comparison between two hand-selected schools of the user's

choosing. If given more time, we would have liked to place both charts next to each other on the same web page, with connected interactivities such that any two schools selected in the radar chart would be highlighted on the parallel coordinates plot.