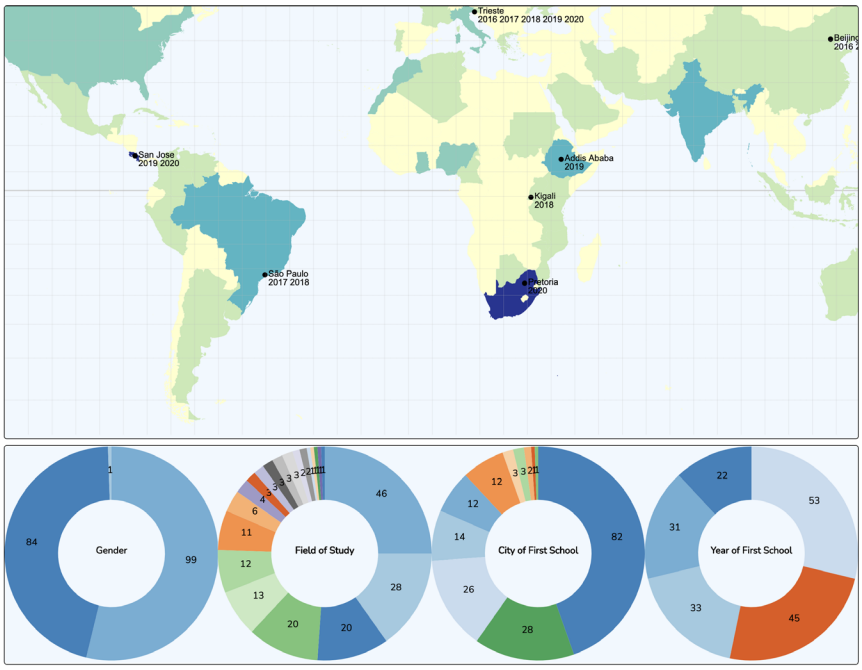
Project 2- CODATA RDA Project Map Documentation

Rob Quick, Srini Yerragolla, Andrew Kaplan, Lauren de Luna

INFO- H517 Visualization Design, Analysis, and Evaluation
Fall 2020, SOIC



1. **Cleaning the data**

   The data that was being used was being collected concurrently while the visualization was being developed. The data was being collected through a survey using google forms and exported into a CSV file. As a result, there was a fair amount of data scrubbing that had to occur before the data was usable for the visualization. For example, one of the more important fields for our visualization was the country that the respondent lives. We originally had this field as a text box where the respondent could right in their response. As a result, differences in spelling caused issues when integrated with the geomap JSON file that was used to develop the map. If the country was not an exact match with the code, the respondent's information would not appear on the map. When we noticed this issue, we made the country a multiple-choice option and merged the new data with the old scrubbed country data. A similar issue was encountered for the Field of Study field.

2. **Creation of the Map**

   We wanted the map to be the focus of the visualization. The map was going to visually show two of our most important variables, 1) which country each respondent currently lives and which country each respondent is originally from and 2) the location of the school for each year. We made the map span the majority of the screen, so it was large enough for smaller countries to be visible.

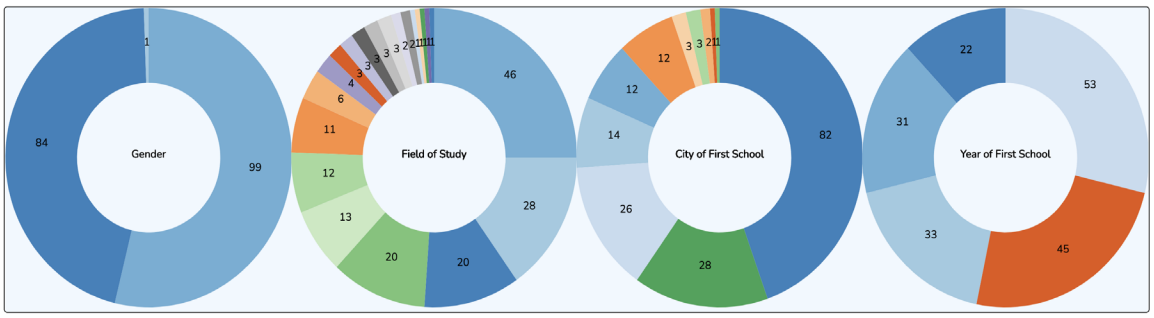
   To develop the map, we used an open source geoJSON code that included all countries. Using a monochromatic color scheme, we visually show which countries have a higher amount of responses currently living there by using a darker tone. In contrast, lighter tones are countries that do not have as many respondents. Lastly, we incorporated the location of the school with the years that those locations had a program. This was done with a simple CSV file that denotes the names of the cities and their longitudinal and latitudinal coordinates.

   We decided against outlining the countries. Since we were using color as our main visual indicator, the color needed to be easily visible. By adding outlines to the countries, the color for smaller countries becomes less visible and dominated by the outline.

   Lastly, we added a zoom function on the map. We made the map large however smaller countries were still hard to visualize. Adding a zoom function allows for easier visualization of data for respondents who may currently or previously lived in those smaller countries.

3. **Show additional information about gender, field of study, city of first school, and year of first school**

   The map was focused on the location of each student. We wanted to show additional information about the demographics of the cohort. To do so, we created donut charts which is a great way to show proportions of each response. Since some of the charts have many unique responses, such as the field of study which has 21 unique responses, we did not want to write the response on the donut chart. Therefore, we gave the number of responses that unique entry has because adding that many words would clutter the chart to make it virtually unreadable. That information is still important, so our solution was to make a hover function that provides the variable for that wedge and the relative percentage. For example, when you hover one of the wedges in the gender chart, you will get which gender that wedge is about while providing the relative percentage.



4. **Contact Table and interaction with the donut charts**

A major purpose for this visualization is to provide contact information about alumni of the school thus increasing the possibility of participant networking and collaboration after physically leaving the event. This is a great resource for people who are interested in learning more about the program from people who learned from the curriculum firsthand. We chose four relevant variables to display in the table: Name, ORCID number, emal, and Twitter handle. These are all great ways to get in contact with an alumnus of the program. At the time of submission, we had 185 entries and expected many more entries which will make a very overwhelming table. Therefore, we wanted a way to be more specific with the people who appear at the table. To do so, we created a function that makes the table entries interactive with the donut charts. When you click on a specific wedge on one of the donut charts (or country on the map), only those participants will appear in the table. For example, if you were to select the physics wedge from the field of study donut chart, only respondents who studied physics will appear.

5. **Color Choices**

We chose our color choices for the map using hexcodes from the colorbrewer 2.0 website. https://colorbrewer2.org/#type=sequential&scheme=YlGnBu&n=9

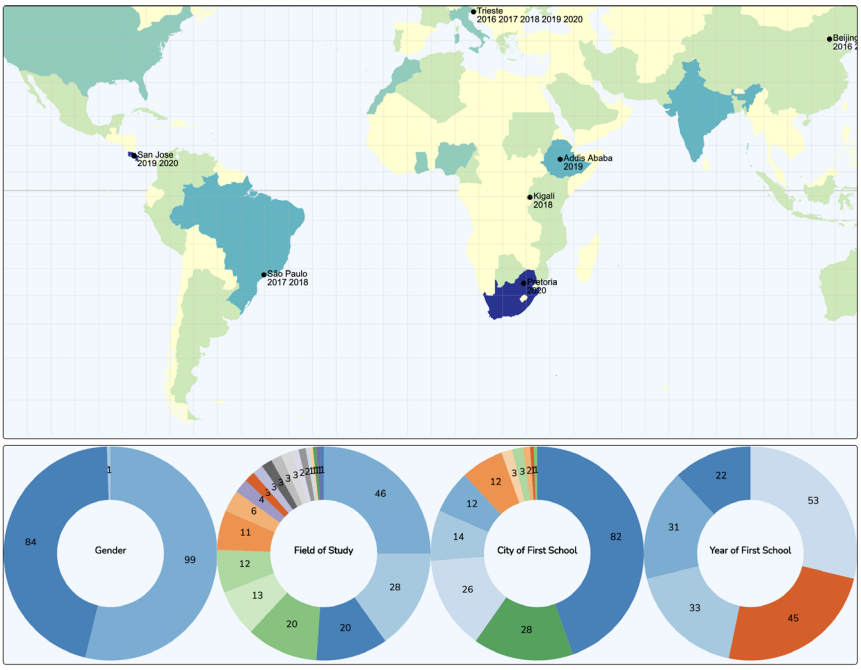
This website is a great resource for color schemes since they offer many colorblind safe color schemes. We contemplated choosing a monochromatic vs multi-hue color scheme and decided on a multi-hue. The main purpose was to easily visually display “density” differences. That is, we wanted countries who had a larger number of respondents (high density) to be darker while countries with fewer respondents (low density) to have a lighter color. Having both color and tonal differences helps depict that information more clearly.

Since we did choose a color scheme with multiple colors, we wanted to make sure that it is colorblind safe. Using colorblind filters on google chrome, we tested all variants of colorblindness. Monochromatic colorblindness was the only type of colorblindness where some difficulty reading the visualization occurred. The countries with no respondents (lightest color) merge with the water and can be tough to distinguish. However, this is the rarest colorblind variant.

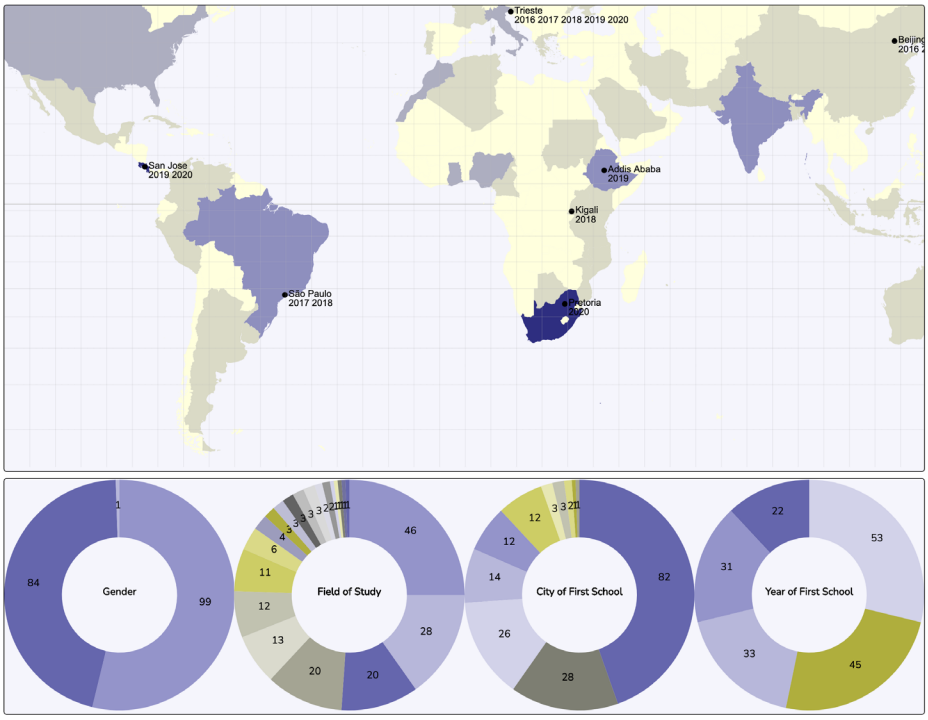
For the donut charts, we used the d3.scale.category20c function as our color scheme. This is a function that is included in d3 version 3. Since the number of unique responses for each variable may change over time, we wanted a color scheme that will automatically apply a new color when a new unique response is recorded. This is especially useful for our field of study variable which has 20 responses since it automatically applies a unique color for each unique response.

Colorblind Filter Photos

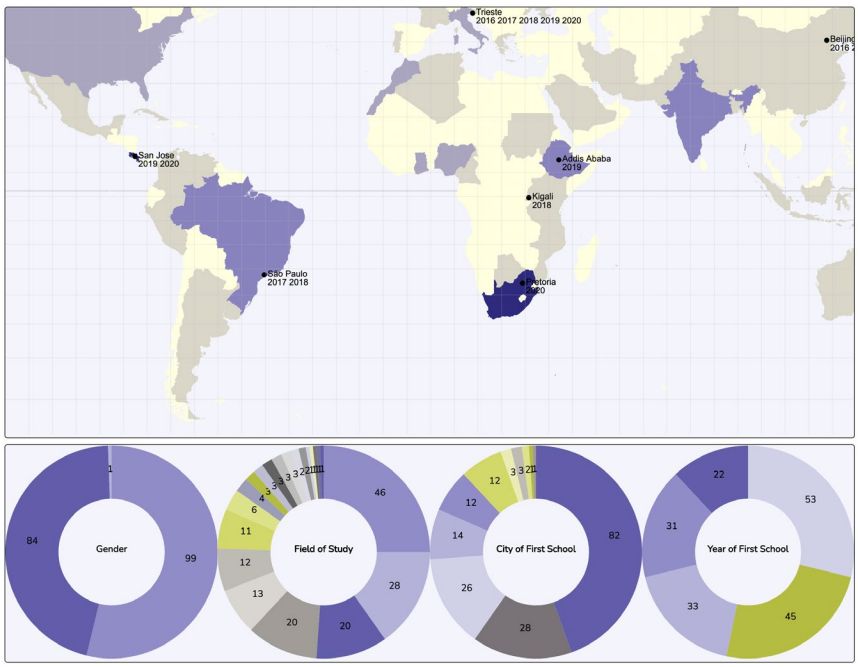
No Filter



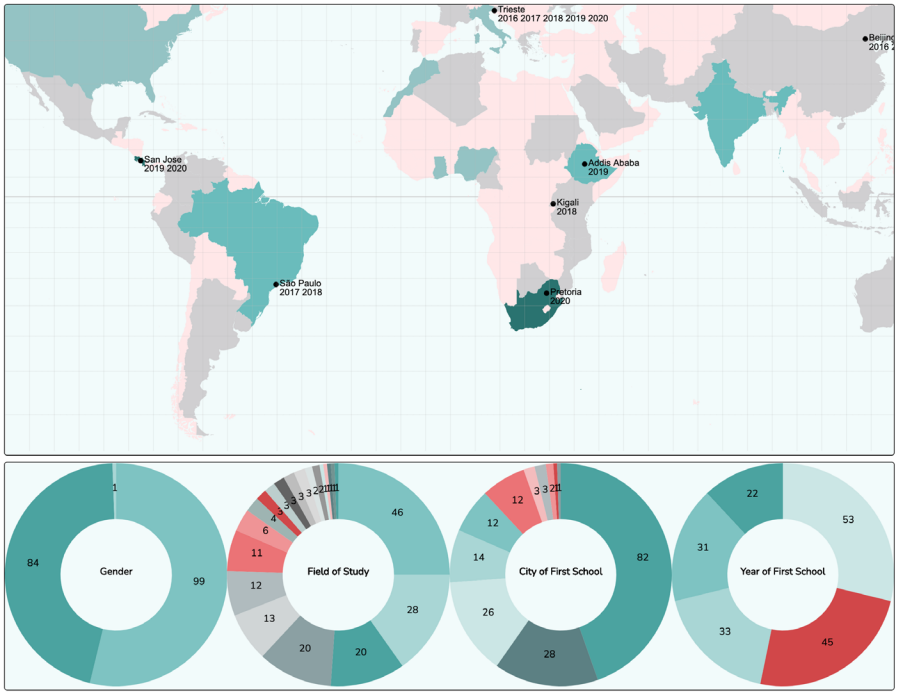
Red-Blind Filter



Green-Blind Filter



Blue-Blind Filter



References

[1] d3-Worldmap Template http://techslides.com/d3-map-starter-kit

[2] https://d3js.org/