World Leader Religious Affiliations and their Sociopolitical Effects

Process Book

Final Project
CS 6630 Visualization for Data Science
University of Utah
Fall 2020

YouTube Link
https://youtu.be/p7uaOZ3lrrc

Github Repo
https://github.com/bmanmo14/Data-Vis-Project

PROJECT SITE
https://bmanmo14.github.io/Data-Vis-Project/project.html

TEAM MEMBERS
Kyle Price - u1018878 - kyle.d.price@utah.edu
Brandon Mouser - u0962682 - u0962682@utah.edu
Brandon Ward - u0838890 - brandon.ward@utah.edu

Table of Contents

Overview & Motivation

According to Pew Research, in 2017, around 84% of people across the world claim a religion. A region's predominant religious affiliations can affect the political, sociological, and environmental aspects of the regions in profound ways. We want to understand and visualize these effects over time (using the above metrics).

For the most part, the religious inclinations of any particular region stay consistent over time— and when they do change, it's usually gradual. We want to provide not just a gradual view of these societal effects, but a year-by-year comparison. Since religious populations usually shift slowly, we decided to use the religion of the leading political figure rather than strictly the predominant religion in the region. We realize this may not be representative of the region as a whole, but then again, what is? Political figures have the strongest effect on the organizational changes that occur in a society. Thus, we hope that some interesting patterns will arise.

We do realize that it wouldn't be fair to claim causation in this project—meaning, for example, that since the religion of the leading political figure shifted from X to Y in year Z for some region, and that some societal aspect improved or deteriorated in the remaining years, that the religious shift can be thanked. However, we do hope to point out correlations.

Project Objectives

The main objective and topic question we are looking to answer for this project is if there are any correlations to be made between world leaders' religious affiliations and the effects it has on socio-political issues for the countries in which they are the leaders of.

Benefits of answering this topic question?

- Interesting
    - If nothing else, the combinations of the datasets we are using in this project introduce interesting questions. There can be claims made that predominant religious affiliations in a particular area or region can have great effect on socio-political and environmental issues. This is also a topic that can be analyzed in many different forms and ways. We feel that data visualization is a perfect form of analysis for this subject, as it can help quickly educate and inform on this topic; something that may seem to have many layers of ambiguity.
- Create the ability to draw conclusions on the topic
    - In addition to being an interesting question to answer, another benefit (or benefit we hope comes as a result of this project) is the ability to draw conclusions on the relationships of the datasets. As stated in the previous section, we hope that our visualization(s) convey data in a way that will allow the level of ambiguity on this topic to be simplified in a way that can invite/create stimulating and conclusive thoughts by the visulizaton's user.

We also hope to accomplish the task of creating a visualization that allows quick learning and viewing for the viewer. Since this project utilizes the combination of two different datasets that may not be typically used in relation to each other, we feel it is important to design our layouts' views and their marks and channels in such a way that will allow for quick learning. Doing this will help answer our questions listed above as well.

Data

The data we use in this project comes from two different datasets; Environment Social and Governance (ESG) Data from WorldBank and Global Religion Dataset (GRD) from the Association of Religion Data Archives (ARDA). The GRD dataset includes a list of world leaders dating back to 1946 and their religious affiliation while the ESG dataset includes a wide variety of statistics for countries dating back to 1960. Many statistics in the ESG dataset are irrelevant for our project and to ensure we are not misrepresenting the data, we decided to include attributes that were measured in terms of percentages, from 0 to 100.

We made the decision to include only percentages because it gives us a constant y-axis for any graph we choose to include in our visualization. Many attributes in the data are not percentage based and instead are a number on an arbitrary scale. We concluded that having a different y-axis for different attributes would be confusing, and many attributes would not be comparable as the scales differ. For those reasons, we chose to use these 18 attributes for each country:

Environment	Economics and Infrastructure	Health	Education	Social Protection and Labor
Electricity production from coal sources	Individuals using the Internet	Prevalence of overweight	Literacy rate, adult total	Children in employment
Renewable electricity output	GDP growth	Cause of death, by communicable diseases	School enrollment, primary	Unemployment
Renewable energy consumption	Poverty headcount ratio at national poverty lines	Population ages 65 and above	Government expenditure on education, total	

Fossil fuel energy consumption		Unmet need for contraception	Proportion of seats held by women in national parliament	

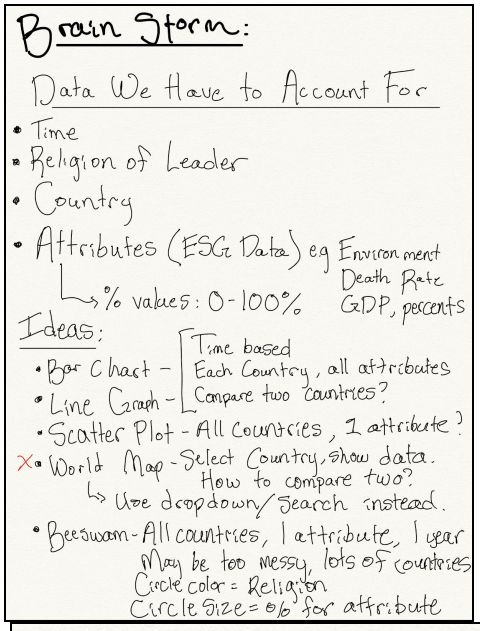
Notice the column headers, these are our *topics*. The topics group together a specific category from the data. Each topic includes two or more attributes, which will be the data in our visualization.

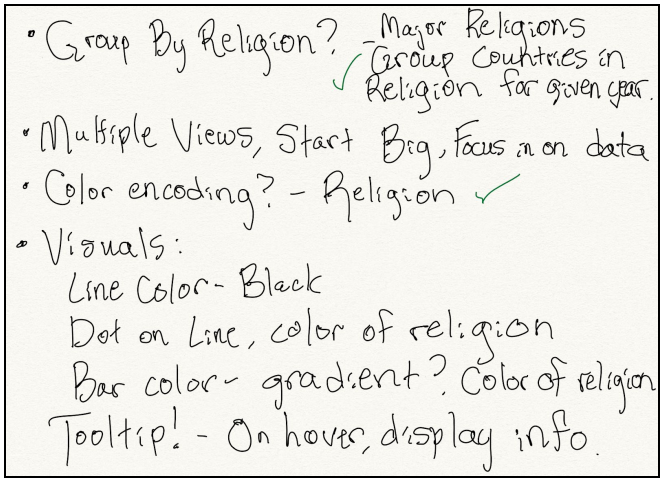
Data Processing

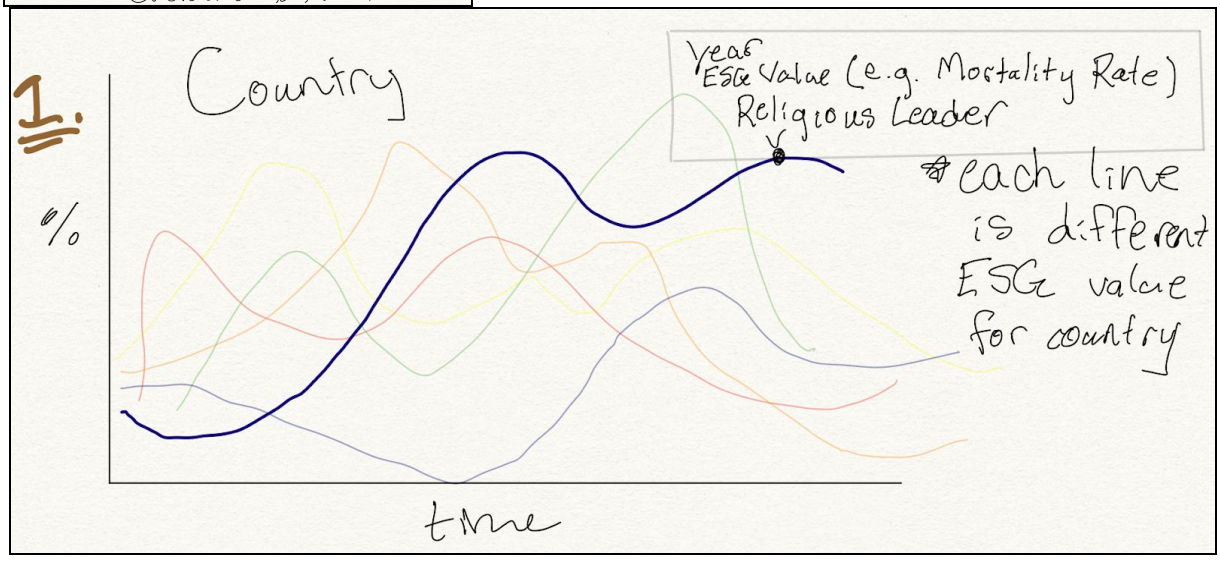
Before combining the two datasets, we first had to remove the unused attributes for each country in the ESG dataset, keeping only the columns we want. Afterward, country names need to match in both datasets before they can be joined. The ESG dataset includes a legend for country names and abbreviations which can be used when joining the two datasets. We plan on using Python to create one combined dataset with the eighteen topics listed above, as well as the country, year, leader, and leaders religious affiliation.

Visualization Design

We first have to consider all of the dimensions for the design based on the data. The GRD dataset provides a mapping from countries to their leader's religious affiliation from 1946 to 2018. The ESG dataset provides attribute data per country over varying topics (for example, for topic "Environment", one attribute is "renewable energy consumption"). We want to show that if a new political leader is elected with a different religion from the previous leader, and some change occurs among the attributes, it can be detected in the visualization. We also want this to show these effects over time. The following images show our general brainstorming notes.



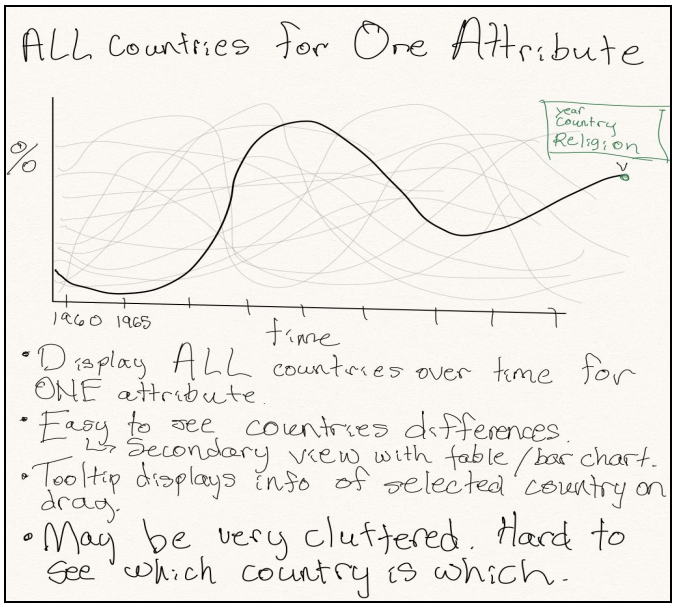




Naturally, we started looking at line charts. Our earliest brainstorming idea was similar to that of the New York Times COVID-19 visualization shown in class. We found that table to be intuitive, informative, and visually pleasing. In our case, the x-axis would show time and the y-axis would be the attribute percentage for the specified country (recall that we only selected attributes with percentage-based values so that they can reside on the same axis).

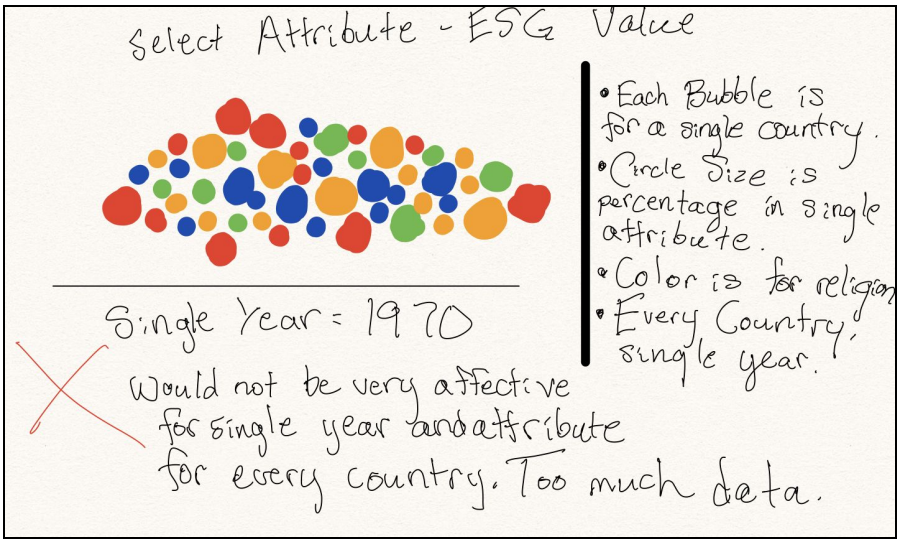
The idea was that each attribute would appear as a line and when the line is hovered over, it would "pop out". We realized two things: 1) we weren't showing religion, and 2) we weren't providing an adequate way to analyze across countries.

We next added each country as a line rather than the attributes:



We quickly discarded this idea. With +150 country lines, the chart would be horrendously busy.

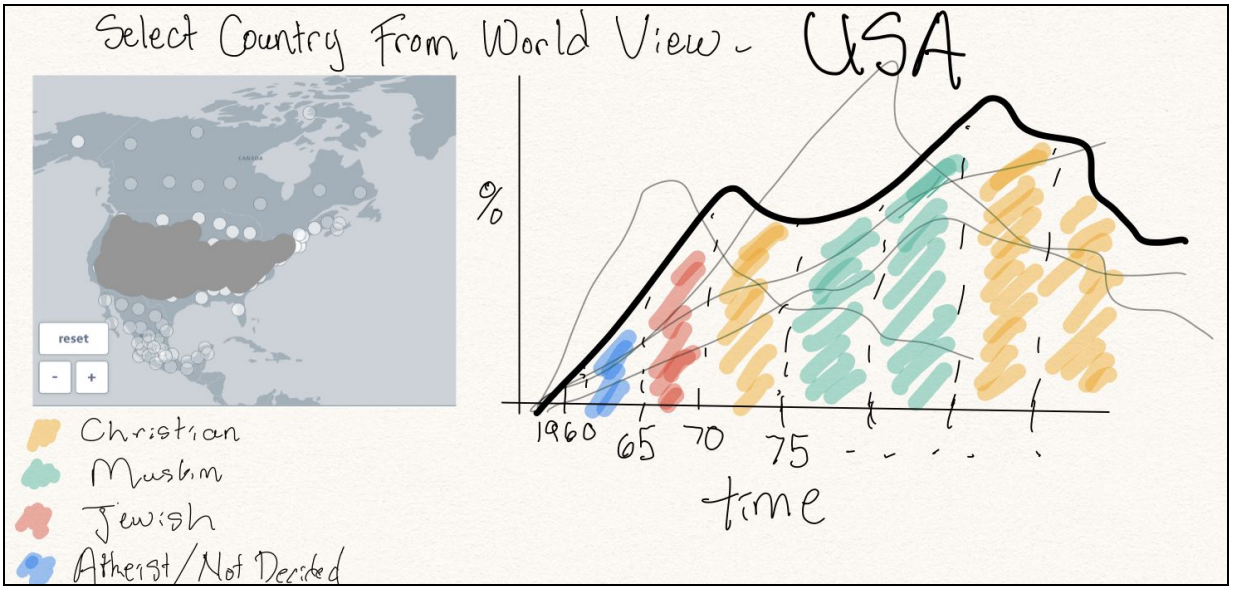
We then looked at beeswarm charts as a way to allow us to show more data without cluttering it up too much.



Each bubble was to be a country, the circle sizes were to encode the value of a single attribute, and the colors were going to be used to encode religion. In this process, we looked into the data more in-depth and realized that there were hundreds of different religions and that trying to encode each and every one of them would be tedious and likely unnecessary. This is when we decided to just bucket each religion into five buckets. The four leading religions plus an “other” category. This would allow us to encode religion with color.

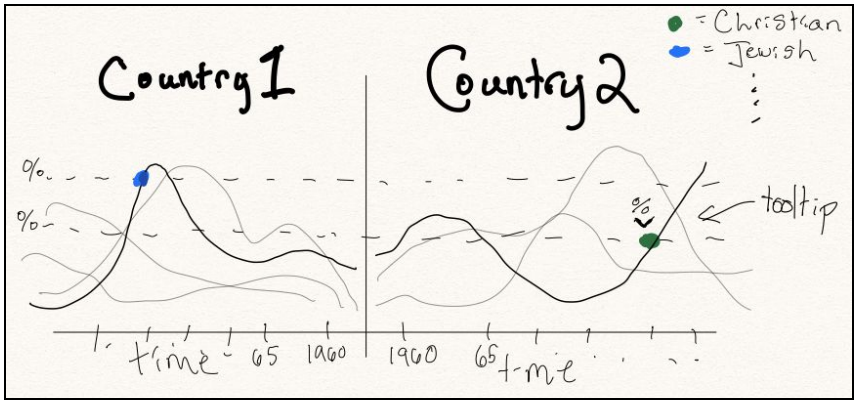
Again, we realized we were trying to show too much data. There are just too many countries. Additionally, since diameter is a relatively poor way to encode magnitude, we felt okay about abandoning this idea.

We went back to the line chart idea. We wanted to allow the user to filter based on country so we thought of adding a country map chart that would allow functionality.



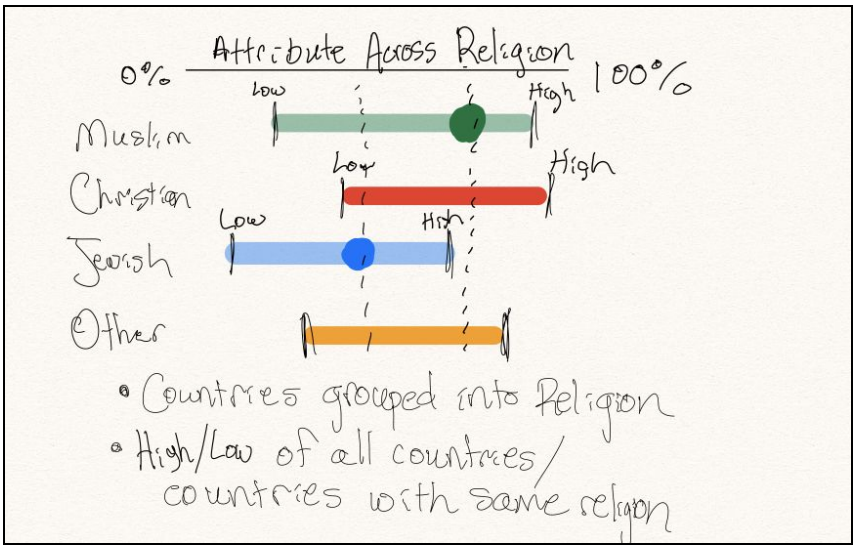
We took the color-religion encoding idea from the beeswarm section and extended it to the line chart. The area under the curve would be filled with the color that corresponded to the leader's religion throughout that year on the x-axis. When a different line (attribute) was selected, the area would increase or decrease to match the area under the new line. However, since this chart represents a single country, the position on the x-axis of the color fills wouldn't actually change.

We liked this idea, but we wanted to allow comparison. Since the religions were bucketed, it didn't make sense to allow general comparisons. We decided it would be better to compare by country. Therefore, we added a second chart to the right of the original that allows just that.



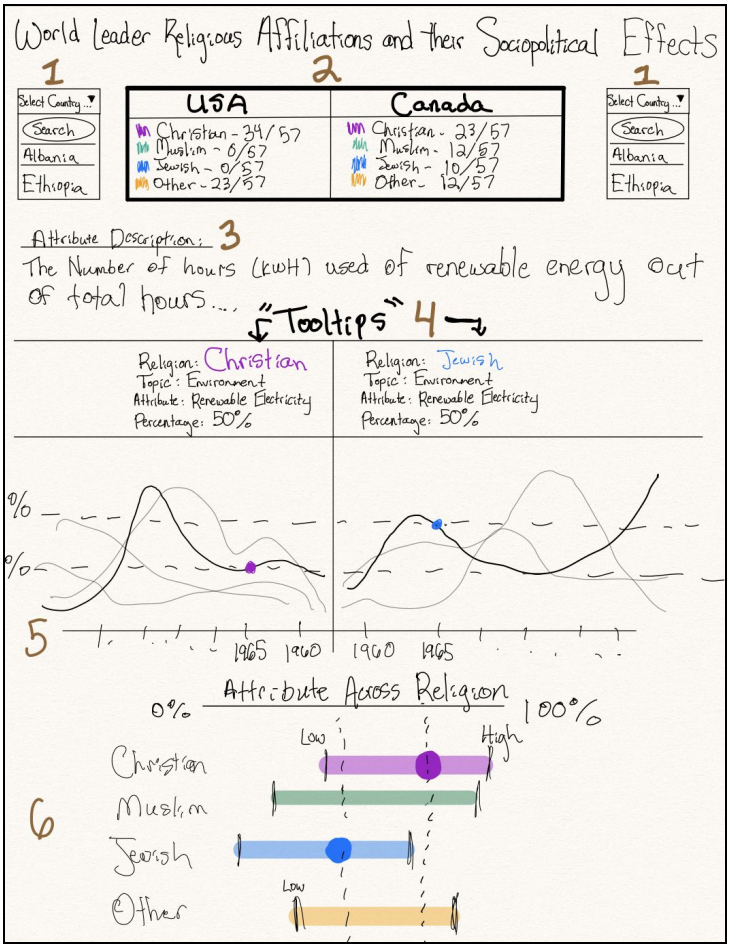
We also decided that a circle could appear on the line where the user's cursor is positioned. Another circle would show up on the graph for country 2 for the same year. These circles would be colored to encode the religions of the two countries in that year. Additionally, as the user drags the cursor along the line, the circles follow it and change color to match the religion in that time period. We also played with the idea of showing a dotted line that highlights the difference between the two in that year. This ended up being the basis for the main visualization.

We wanted to provide more detail for the religious aspect. We decided to design a secondary view that shows a more detailed view of this aspect.



The above view was initially thought to be used as a more condensed tooltip of sorts that would appear above the cursor. We decided it would serve better to place it in its own location below the two line charts. As a user cursor moves along the line, this window populates. The window corresponds to a year, a particular attribute, and the two selected countries. It will simply be a collection of 5 bar type charts—one for each religion. Each religion's high/low range across all countries will be encoded as a bar and the two country's positions will appear as a circle on the bar that corresponds to their respective religion. The scale of the bars will have to stretch from 0-100 so as not to misrepresent the range, but that bars length will encode the religion's low/high range.

Our final design can be found below.



The following points indicate the purpose of each labeled (brown) item:

1. Allows the user to select the country for the left or right chart.
2. Show the selected countries. It also shows the number of years for each religion that shows up in that country. This window is static across attributes.
3. Shows a detailed description for the selected line (attribute).
4. Acts as a static tooltip. Like (3) it changes when a new attribute is selected, but it also shows specifics for the particular year. Namely, it shows the religion, topic, attribute, and magnitude.
5. Is the two countries line graphs for all of their attributes. Selecting on a single line will highlight the line and users can scroll to different years on the line.
6. Is the bar chart that provides detailed religion information across all countries (as well as the selected countries position).

Features

Must-Have Features

The following is a list of must-have features that we plan to include in our project.

- The ability to compare two countries for a particular attribute and year, with all attributes on the same graph.
- When hovering over a specific attribute, the second view is updated to include the high/low values for each country in the particular religion for that year.
- An information pane showing details about the two countries being compared and a description of the selected attribute.
- The ability to select countries via a dropdown menu or search bar.
- A bar chart showing, for each religion, the low and high for the selected attribute currently being viewed for two countries and where those two countries stand in comparison to the rest of the world.

Optional Features

The following is a list of optional features that we would like to add to our project, time permitting.

- For the two country line charts, we would like to shade the areas under the curve so that for each year, the shading encodes that country leader's religion.
- Brushing: We could not find a really efficient way to implement brushing with the way we have set up our graphs and plot, but it would be a nice feature to have. Maybe brushing over a length of time for two countries shows the average values for their attributes of the span of time, as well as the average values for all countries in each religion over the selected time.
- Show only selected topics. Instead of showing all of the attributes for the two countries being compared, let the user select a specific topic. Only the attributes in the topic would be shown in the line graph and bar graph.

Project Schedule

The following is the project schedule we laid out for a timeline of approximately 5.5 weeks. (October 25 - December 2)

Week 0 - October 25 - October 31:

- Project Proposal - All group members

Week 1 - November 1 - November 7:

- Data pre-processing - Kyle
- Begin HTML/Javascript layout - Brandon M.
- Start Process Book - Brandon W.

Week 2 - November 8 - November 14:

- Data structure design - All Members
- Begin initial visualizations - All Members
- Update Process Book - All Members

Week 3 - November 15 - November 21:

- Encoding and styling - Kyle
- Tools for searching and filtering data - Brandon W.
- Continue work on visualizations - Brandon M.
- Update Process Book - All Members

Week 4 - November 22 - November 28 (Holiday Short Week):

- Continue task from prior week - All Members
- Update Process Book - All Members

Week 5 - November 29 - December 2 (Project Due Date):

- Finishing up code-related work on respective assigned tasks - All Members
- Create/set up site - Brandon M.
- Record screencast - Brandon W.
- Complete Process Book - All Members

Evolution

Feedback Meeting

We first met with other students in the class to critique and evaluate their project design as well as ours. The students we met with were William Ludwig and Jacob Blomquist. The following are some notes, suggestions, questions, and feedback they gave us in the meeting.

> **Comments and Feedback:**
>
> **What years are you going to include?**
>
> **Suggestion they gave us:** Specific date feature, somebody wants to look at a specific year for a country, more easily see trends for a certain country. Maybe look into brushing for a specific year range. Doing this may make it easier to see the dataset, zoom into that point.
>
> **Where is the data coming from?**
>
> **Answer:** Data is coming from two different datasets.
>
> **Other notes:** Good amount of detail and interactivity, any more may affect the layout and visualizations in an unexpected way. Risk adding to many widgets, it becomes too overwhelming.

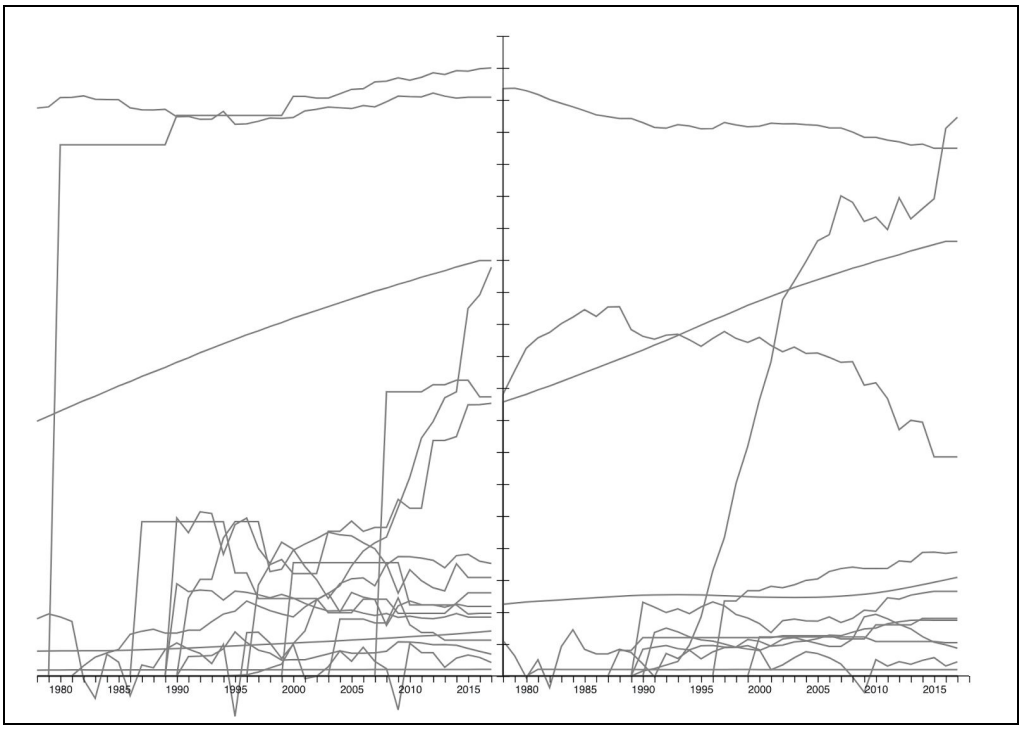
- See feedback_excersie.md in project repo.

Design and Implementation

First Steps

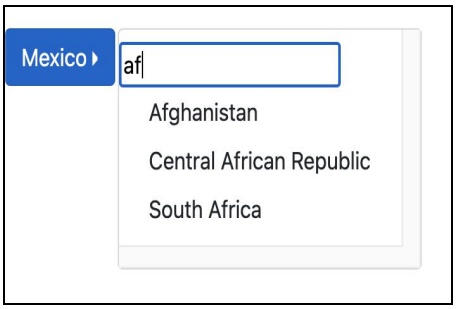
To stay true to the goals and objectives we highlighted in the visualization design section, we began by creating a visualization that compares two countries side-by-side. These countries can be selected by two drop-down menu buttons on each side of the visualization. By clicking on the drop-down button, a country can either be selected via drop-down selection, or searched for in a search bar.

The first change that we made from our initial designs is with the main country comparison chart. In our initial sketches, we had the country comparison chart diverging at 0, where 0 is the starting year of our data. As we began to actually code out the design, we realized that visually it didn't make sense to have an x-axis for a timeline to read from right to left, when typically any type of timeline or anything that is intended to be visualized in chronological order is usually left to right. So instead of having both charts diverge at the starting year in the middle, we adjusted the x-axis for both countries being compared to read from left to right. The following screenshots show the main line plot view.



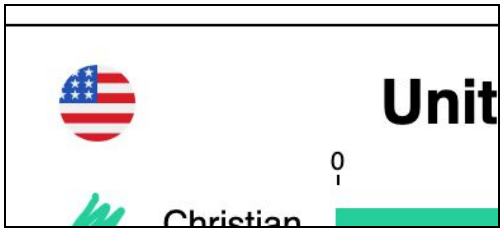
Notice that some of the lines dip below zero. We learned after the fact that GDP can be negative. We'll have to handle this in some way that doesn't distort how people perceive the results. We could either pull GDP out of the dataset, floor it to zero, extend the y-axis to allow negative results, or encode negative regions in some other manner.

We also created a type/combo box input area for each plot that can be used for country selection. A screenshot can be found below.



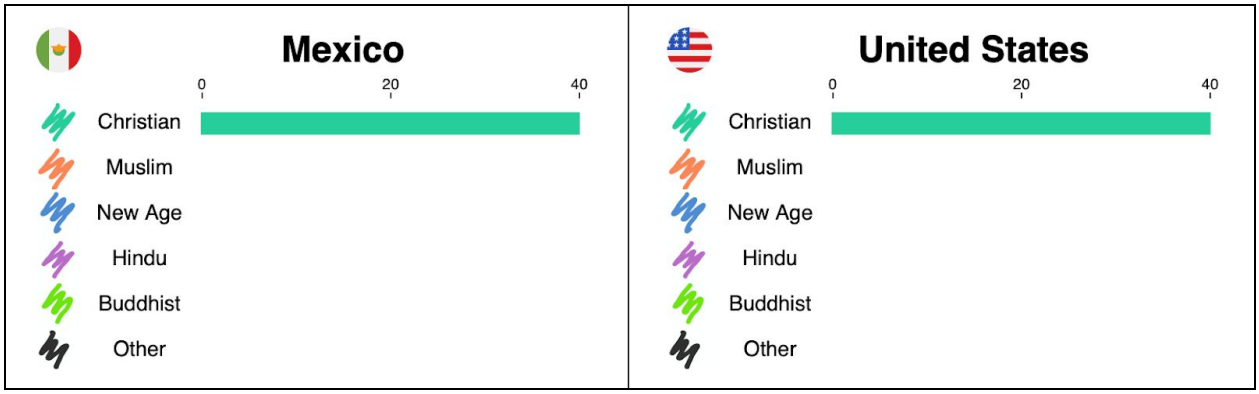
Country Selection Table

For the most part, the country selection table stayed true to the original design sketches. We did however make some minor changes that we feel improved the design and tied the whole visualization together nicely. The first change we made was changing for each religion the year counts from a count out of the total number of years to a bar on a scaled axis from 0-40 years. We also added an svg image of the selected country's flag to the left of the country's name as seen in the screenshot below.



The next addition we made to the tables was to add indicators for each religion. We did this because we felt that these tables needed some color and visual identifiers for each religion, especially in the cases when a religion does not have any value for its corresponding bar chart. We decided on a "squiggle" colored in each religion's color

coding as the indicator. We felt that this gives a nice "organic" touch to a design that features mostly clean squared lines. These "squiggles" as indicators can be seen in the following screenshot.

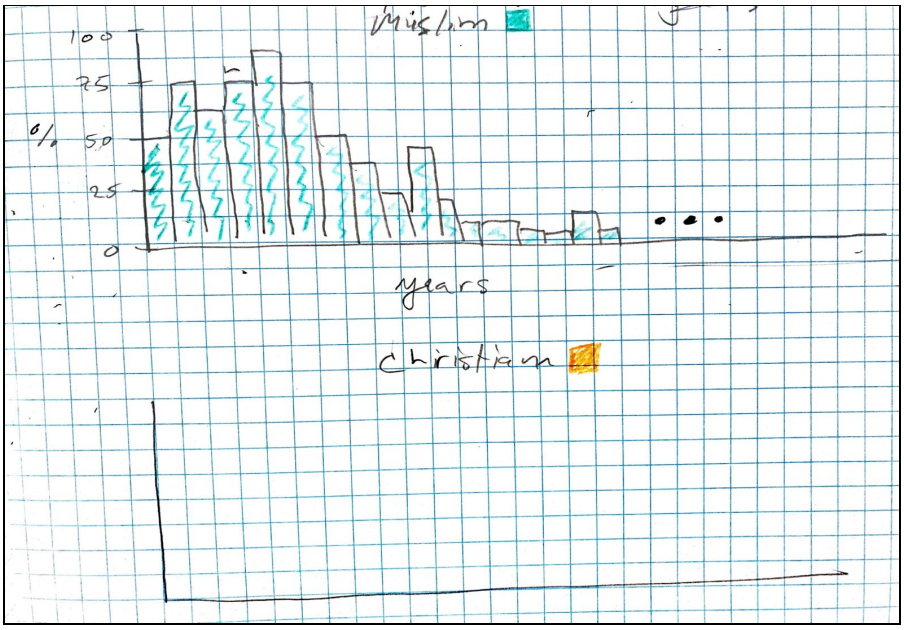


Religion Plot

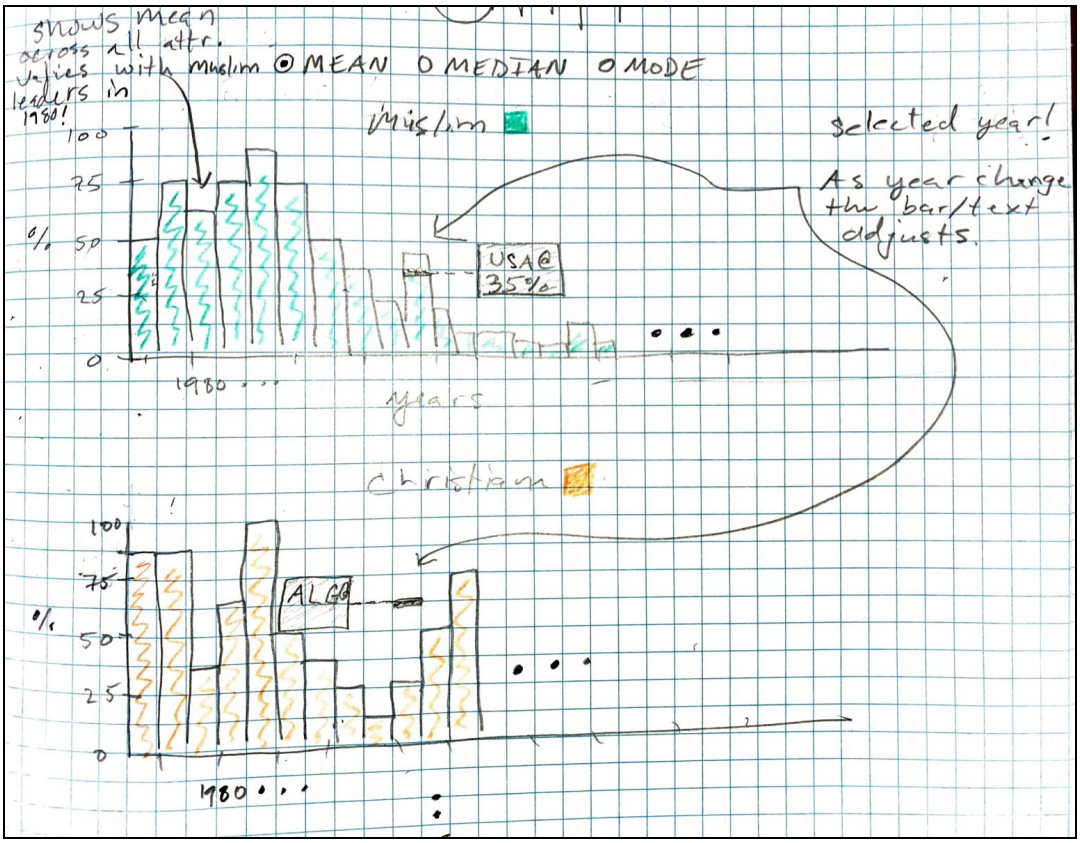
As we considered how to create the religion plots, we started discussing what value the plots were actually bringing to the table. We wanted to provide a global view for each religion for the selected topic/attribute/year. By solely relying on the bars with color gradients to indicate the distribution of the different selected attribute values found within the countries whose leader ascribed to the selected religion in the selected year (as well as where the selected countries sit in that distribution), we realized we weren't providing that much information. We decided that we wanted to:

1. Show the distributions across a timespan to visualize temporal change.
2. Include different metrics for the data (e.g., mean, median, mode).

We considered alternatives to the simple 1D bars that we initially planned. Our first (and final) idea was to use a histogram. Histograms provide a 2D representation that would naturally allow for time and attribute values to be visualized. Our initial design can be seen below.



For each religion, a histogram would be created and they would be **stack**ed on top of each other. Since we also wanted to allow the user to analyze with different metrics, we considered creating one **stack** for each metric. We quickly realized that this would make the page too tall and cumbersome to navigate. We decided to create a set of radio buttons that would allow the user to toggle between the different metrics. Additionally, like in the original design, we wanted to show where the selected countries fell in the distribution. We decided to display two small text boxes that would follow the selected year along the x-axis and would indicate each country's attribute value for that year. The boxes should be displayed in the correct religion plot that their leader prescribed to in that year. The updated sketch can be found below:



This sketch drove our design for the final product. We justify this deviation from the original design by pointing out that the new design allows us to show much more in less space. It shows a temporal change for a given attribute for each religion whereas the bar design restricted us to a single year. Furthermore, since length is a much better form of visual encoding than color gradients, it actually shows the attribute magnitude much more effectively.

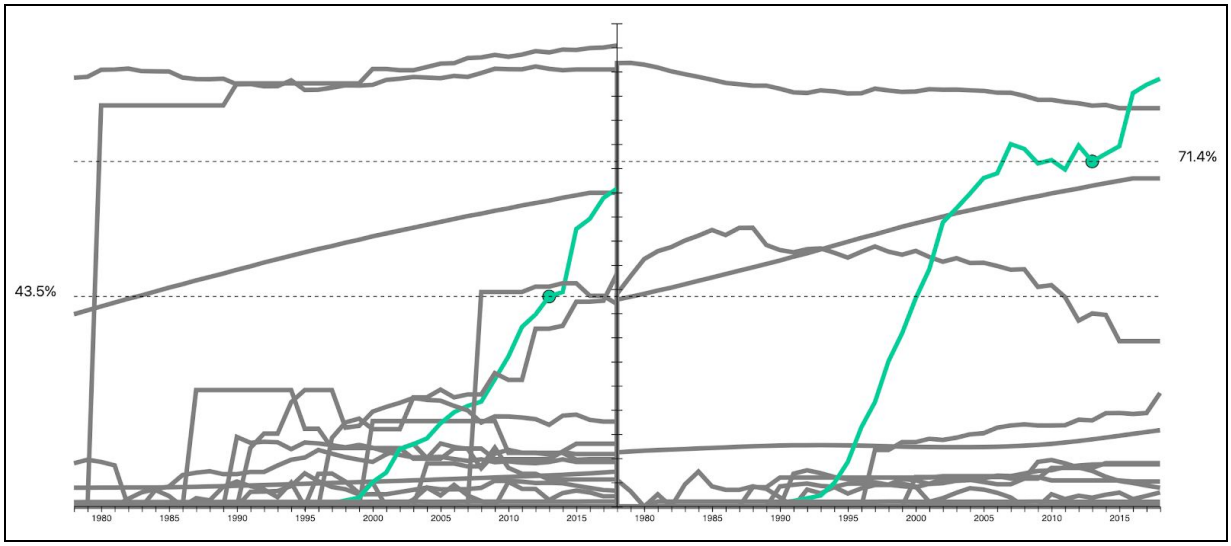
Line Graph

The line graph went through many stages of development to implement the required features we specified in our project proposal. We wanted to make the line graph interactive so that users could compare countries, topics, and attributes for a specific year.

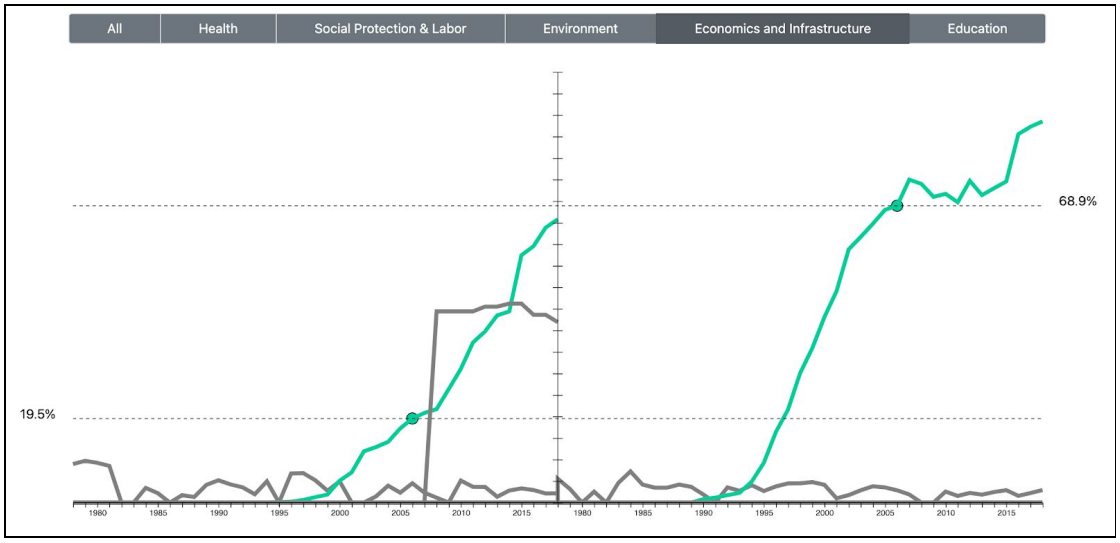
To accomplish this task, we implemented highlighting, dragging, and dotted lines for both selected countries to highlight the difference between the countries for a given attribute and year. At first, we implemented an ‘onClick’ handler for each of the lines. It soon became apparent that dragging would be a lot better than clicking on a specific place on the line, so we also added a ‘drag’ feature to the lines.

As part of the initial design, we wanted to include a dotted line to show the differences between two countries attributes for a specific year. We ended up adding two lines that spread across the entire graph to easily convey this information to users. At first we had the percentages on one side of the graph, but realized differences would be easier to see if the percentages were on their respective country side, especially when two lines were close together.

When selecting a particular line/attribute on the line graph, we also change the religion plot to convey to users how a particular country compares to all countries with the same religion in a particular year. For countries or years that did not have values, we made the decision to use the values from the previous year, as going down to 0% made the graph look messy and jagged. There are a few ‘null’ values in the graph that use the previous years value for an attribute, but overall a lot of the data for most countries was there. The image below shows the highlight and dragging features we implemented, along with the dotted lines.



We also chose to implement one of our optional features, allowing users to select between various topics, and only showing attributes for each country that fall within the selected topic. To do this, we added a line of buttons above the line graph that users can select to show attributes for a particular topic, or to show all attributes. When a button press occurs, the tooltip and religion plot are updated accordingly.

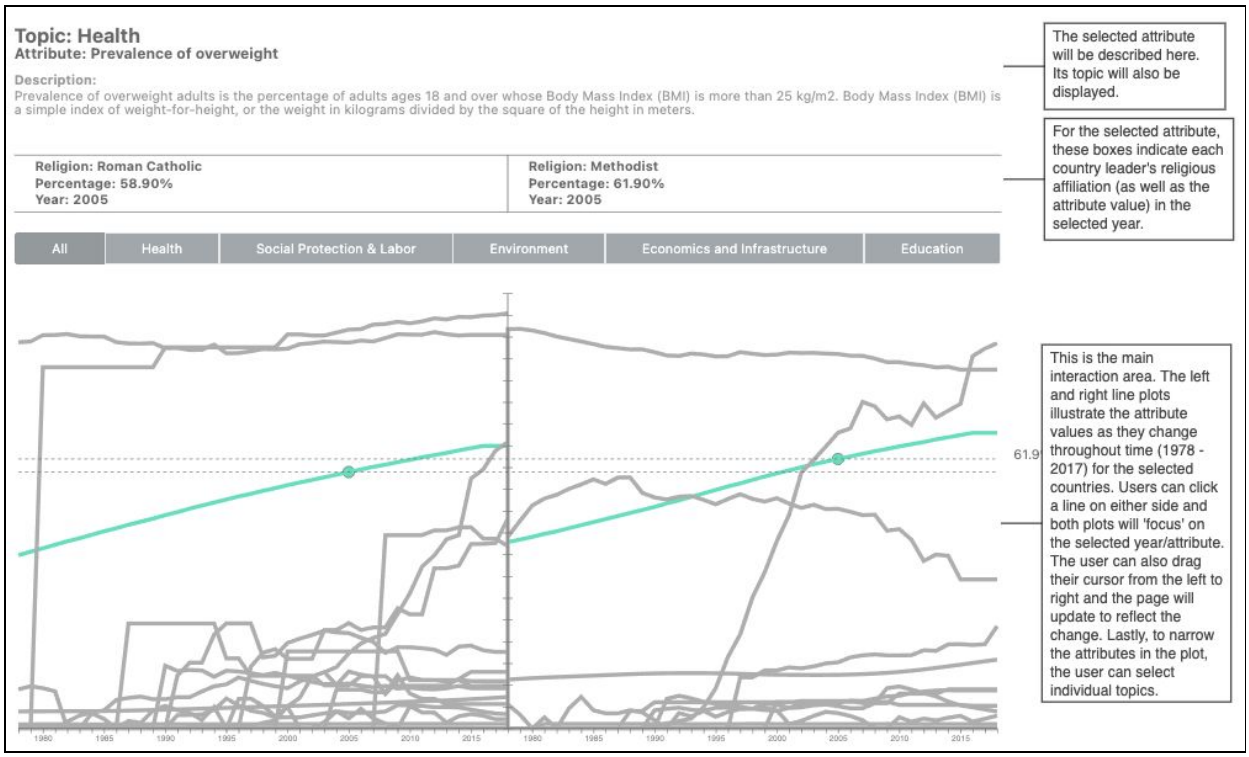


While implementing the line graph, we noticed that one attribute, GDP, can have negative values. We discussed various solutions including extending the graph to show negative percentages, coloring the negative regions of the line a different color, and just flooring the negative GDP to 0 for the specific year. When laying out designs for the page, we soon realized that having a negative axis for only one attribute would not be an efficient use of space within the page, so we decided to make all negative GDP values go to 0% on the graph. We also decided that changing the color of the line for the years when GDP is negative might be confusing to users, so we left the color of the line the same and show the negative GDP in the tooltip.

Help Button

Because the site has multiple interactive elements, we decided to add a help button that shows users how to navigate the website. Adding the help button addresses one of the comments we received during our first project milestone to add some descriptive text showing users how to use the website.

The help button borrows ideas from the storytelling portion of Homework 6. When clicked, an overlaid 'div' covers the screen with a white 'film' that disables normal interaction. Additionally, boxes are displayed next to each main aspect of the site; country selection, topic and attributes description, tooltip, line graph, and religion plot. The text within each box describes what each part of the site does and informs users how interaction works on the site. When a click occurs, the help screen disappears and users can navigate the site as normal.



Evaluation

In this section we highlight and evaluate what we learned about our data at each step of the implementation process. We also discuss the resulting effects from the things we learned about our data had on our visualizations.

Things Learned from Data Pre-Processing

When we first began creating our data structures to be used to hold the data for the visualizations, we noticed some things that we needed to clean up before we could use the data. First, we realized in the dataset that contained the features that pertain to the religious affiliations for each of the world leaders that the names of the religious leaders had a unicode encoding that was not easily recognizable. Because of this, we had to remove the leader's name from our data structures. In our preliminary design planning and proposal we never intended to use any of the leaders' names, but this unicode issue took away any possibility of adding them into our design.

We also discovered that the country data's country codes are not the same as that of religious data's country codes—something we thought was universal when we first elected to use these two datasets. As a result, we only took selected countries to use for our design that contained the same country code in both the religious dataset and the country data set. This resulted in our visualization now only encoding 176 countries.

Visualization Efficiency and Improvements

While D3 is a powerful data visualization tool, we soon realized that making clean visualizations is not easy without a clean dataset to work with. A lot of the problems within our project stemmed from the way the data initially imported, which forced us to do quite a bit of preprocessing before implementing the visualizations.

After the data was imported and made into classes, we were able to start building the visualizations. At first we were concerned that reading in the large CSV files and creating

the classes in Javascript would be very time consuming and noticeable to the user when the page loaded, but it turns out the page loads fairly quickly. Even when incorporating the visuals like the country flags and colors for each religion, the page loads quickly. While building interactions into the page, we wanted to make sure selecting a new country, dragging on a line, selecting a different topic, or changing the metric being viewed did not result in high latency on the page. Because of the pre-processing done on the data, and the way we implemented the classes in Javascript, interactions on the page occur smoothly. One reason for this is because all of the data is pre-processed and classes are built only once in the very beginning when the page is loaded, which makes interaction quick because values are accessed almost instantly from these data structures.

When we completed implementing the visuals, we were able to compare two countries and their topics/attributes/religion for every year from 1978-2017. While testing by cycling through various countries, we came across several interesting comparisons. For example, Sweden and the United States. Sweden alternates between Christian and “New Age”. When it does, there seem to be noticeable effects (or causes) from the attributes included under the “Environment” topic. It was interesting to see how a country's leader's religious affiliation changed over the years, and how various attributes of that country changed in response. While we understand there are many aspects besides a leader and their religion that affect a country’s socioeconomical state, we found viewing a country’s attribute in terms of a leader's religion an interesting way to analyze both a religion and a country.

While we designed and implemented many interactions on the page to analyze the state of a country in relation to the leaders religious affiliation, some of our stretch-goals were to incorporate a brushing feature to view a set of attributes across multiple years. This would allow users to compare countries for a set of years and attributes instead of one year and attribute. However, due to the unplanned deviation from the original design (bar charts to histograms (to focus on global metrics across religions), and due to the limited time period of data we included, we decided that including the brush wouldn’t provide enough value.

Overall, we are very pleased with how the visualizations turned out. After implementing the visualization, we realized the dataset was not as complete as we would have liked, but

the final visuals still show the differences between countries for different years and religions.