Final Project Proposal

CSE 564: Data Visualization

April 16th, 2019

Project Information

Project Topic	Understanding the correlation between personal religious freedom and socio-economic backgrounds.
Technology	Python, D3.js, Bootstrap, Javascript
Team Members	Arun Swaminathan (SBU ID:112044697)
	Shubham Jindal (SBU ID:112129688)

Some Background...

As per Wikipedia, there are 4200 religions around the world [1]. At the same time, religion is an important factor when it comes to a person's functioning in society. Being an important facet of human civilization right from the early ages, it is important to understand the influences of religion in today's era, defined by the various macro and micro socio-economic factors.
Many developing countries today, are undergoing intricate situations when a religious political party comes into power. But what exactly happens to the status of its citizens? This is an important topic to put further light upon so that future policymakers can take this into account.

Mission and Hypothesis

We want to understand, whether a person is able to express his religious beliefs, how the socio-economic policies in that region are affected by religion, and in turn how religious freedom is affected by these policies.

Some of the questions which we would like to cover are -

1. Are crimes influenced by religion targeted at places which allow expression of religious beliefs?
2. What is the economic situation of a region, given that people in that area are allowed to express their religious beliefs?
3. What is the correlation between expressing religious beliefs and the happiness levels of people in that region?
4. If the dominant religion in that country is able to practice their religion freely, does it imply that the people happy? What about religious minorities?

We want to analyze these trends at a global level.

Literature Review

As the introduction says, in order to achieve what we set sails for, we started looking into a lot of resources like notable blogs of related work, existing kaggle datasets, Wikipedia articles, etc. to get an idea about how much data we have about the religions of the world, how is personal freedom calculated, which socio-economic factors are predominant and which of them are relevant to our study. These factors include development of a region, happiness index, threats as extreme as terrorism, etc.

We made sure that the datasets obtained are the latest available, relevant and is from a verified source. In order to bring out interesting insights, we reviewed and shortlisted the following datasets -

Human freedom index provided by Cato: [2]
This study tries to define relationships between freedom and different socio-economic phenomenon. It includes 79 distinct features encompassing a wide variety of topics such as law, trade, religion, security, government among others.
The data available is for 162 countries, and most of these features are on a scale of 0-10, where 10 is most free.

World, regional and national populations by religious beliefs - UMich [3]
This study provides details regarding distribution of population based on religious beliefs throughout the world on different global scales. Correlates of War (COW) project seeks to facilitate the collection, dissemination, and use of accurate and reliable quantitative data in international relations.

World Happiness Report - Gallup World Poll [4]
This dataset ranks 155 countries by their happiness levels at the global level. It is based on a few macro-level attributes such as GDP, Life Expectancy, Freedom, Generosity, etc.
It also provides a happiness score, on a scale of 0-10, where 10 means most happy.

Religious Terror Attacks [5]
This dataset gives a list of terror attacks which were motivated by religious beliefs. It includes the date, country, number of people killed/injured and a description of the attack.

Global Terrorism Database [6] [7]
It is an open source database on terrorist attacks from around the world from 1970 through 2017. It has greater than 100 attributes such as units, tactics, weapons used, etc. and covers more than 18000 attacks.
The data is provided by GTD Team at START headquarters at the University of Maryland. The codebook [6] defines the methodology employed to collect the data.

Datasets

The following is the list of datasets that we have found related to our problem. This dataset is only from our preliminary findings and should not be considered as an exhaustive list of the datasets used for the project.

DataSet	TimeLine (Year published)	Number of records	Major features/columns	Data Source
Human Freedom Index	2018 (Latest available)	1458 x 123	• 79 distinct indicators in the areas of Rule of Law, Safety, and Security, Religion, Expression, Freedom to trade internationally, regulation of credit, labor and business, etc. • Covers 162 countries and time series from 2008-2016	Ian Vásquez and Tanja Porčnik, *The Human Freedom Index 2018: A Global Measurement of Personal, Civil, and Economic Freedom* (Washington: Cato Institute, Fraser Institute, and the Friedrich Naumann Foundation for Freedom, 2018).
World Religion Data 1.1	2019 (Latest available)	1995 x 79	• Contains data about religions, religion family and religion adherence from 1945 to 2010. • Data presented from 1945 to 2010	The dataset was created by Zeev Maoz, University of California-Davis, and Errol Henderson, Pennsylvania State University, and published by the Correlates of War Project.
World Happiness Report	2018 (Latest available)	155 x 12	• Data present for 155 countries • Metric based scores from 0-10 • Date present from 2013-2016	Gallup World Poll
Religious Terrorist Attacks	2016 (Latest available)	29365 x 6	• City, Country, # of people killed, # people injured, Description	Kaggle Data scraped from Religion of Peace website - http://www.thereligionofpeace.com/
Global Terrorism Database	2017 (Latest available)	182000 x 135	• Terrorist attacks with 45-120 variables, including number of fatalities, injuries, weapons used, and perpetrators	University of Maryland, "Global Terrorism Database"

Approach and Implementation

Following is an overview of the steps of how we are going about this project. This is again preliminary as our approach as well as implementation would change based on the new datasets that we incorporate and the insights that we observe.

- We would start by doing an in-depth literature survey and try to analyze the information present.

- We would then perform the data cleaning step following the standard procedure we learned from class - import data in proper format (CSV), merge the datasets, rebuild missing data, standardize, normalize, deduplicate, verify and enrich it. If needed we would also go for data augmentation.

- We would also need to do the data reduction step and sample the datasets with various strategies, which is clear from the size of our datasets.

- Our data set has a lot of attributes, so we would probably have to do dimensionality reduction.

- We would then decide on the specific ways to visualize the data and try and find the correlations between the major attributes using techniques such as PCA.

- Based on the patterns observed we would check if we are able to prove our hypothesis and the related questions.

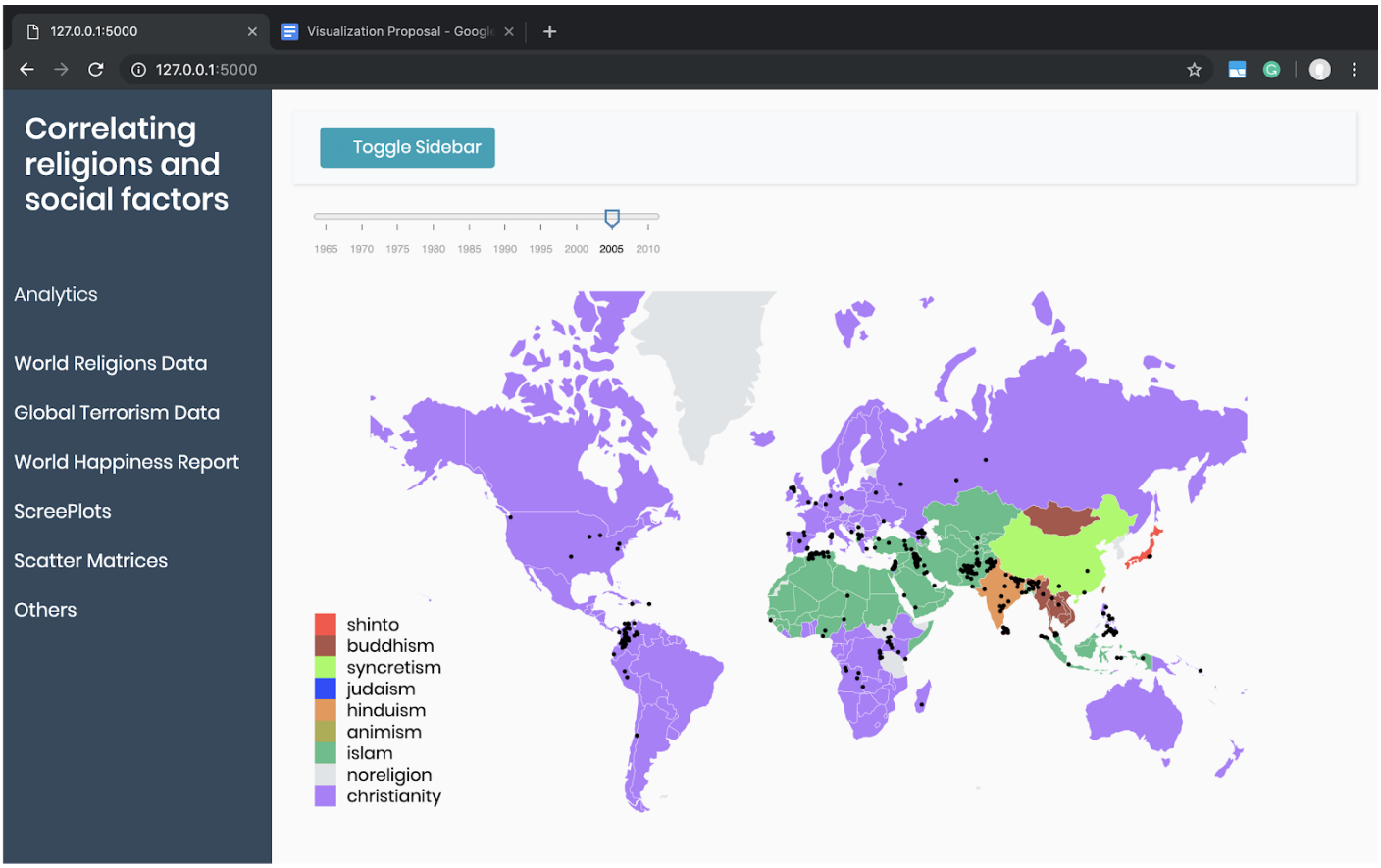
Conclusions

We hope to achieve the following with this project

- Show that our hypothesis defined above is in accordance with previously held beliefs and trends.
- By implementing a lot of what we learned from the classes, get our hands dirty and see, in practice, how actually visualization studies are done.
- Possibly provide a newer way to “look” into the socio-economic factors and it’s a correlation with religion for our policymakers to act wisely.
- World peace!

Preliminary Report

We are developing a dashboard which would help us analyze different aspects of how the ability to express one's religious beliefs correlate with different socio-economic factors. Different sections of the dashboard portray different metrics for visualization which would give us a unique angle to look at things.



Dashboard

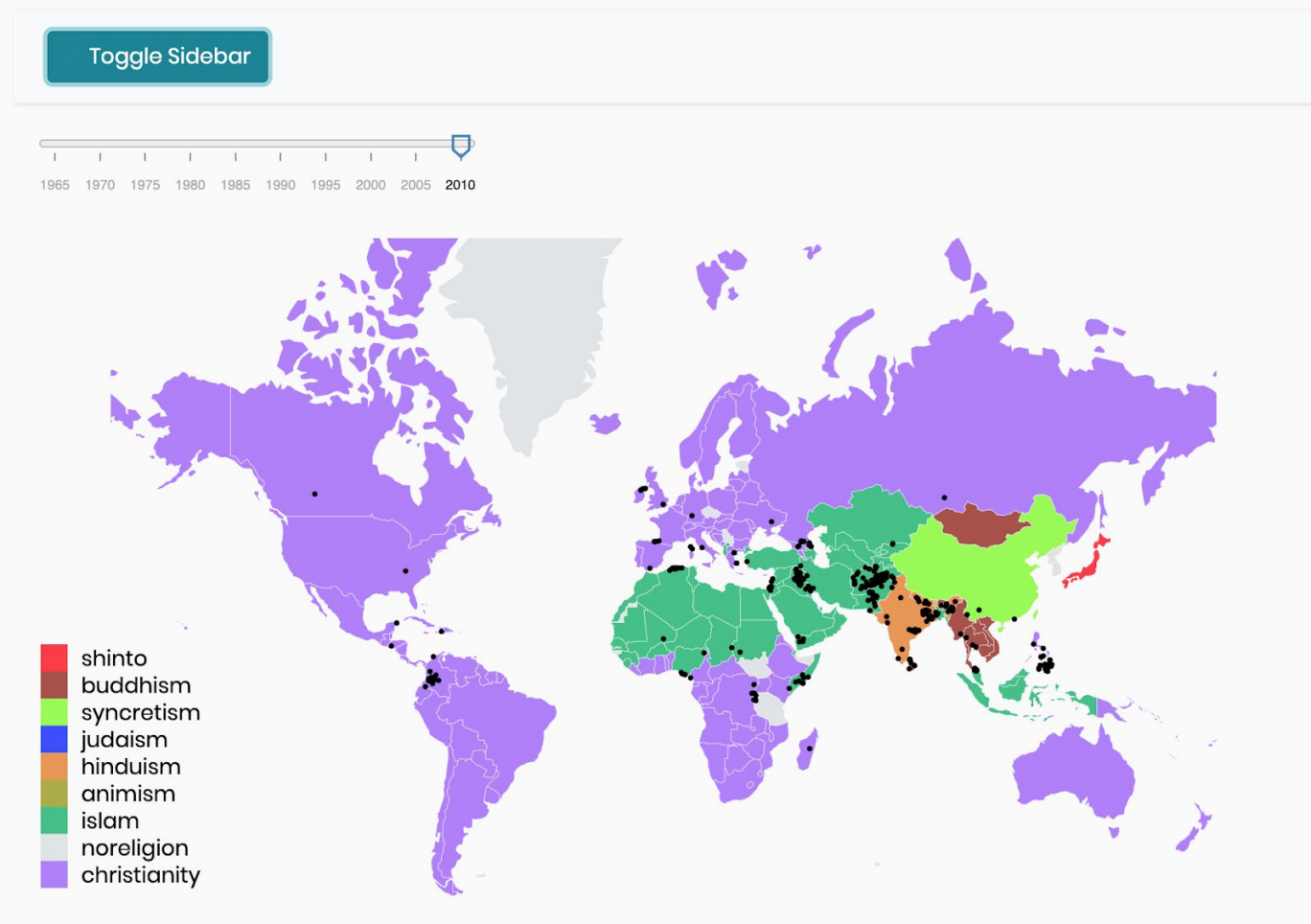
Data Preparation

We started by analyzing the distribution of religious population right from 1945 until the latest data available and figuring out the majority religion in each of the countries. We used the World Religion Dataset [3] to achieve this.

This dataset contains a distribution of different sub-religion populations for e.g. Catholics, protestants, Shia, Sunni, etc. In order to establish major religions in the world, we combined these sub-religions into the following 9 main categories -

- Christianity
- Islam
- Hinduism
- Buddhism
- Judaism
- Shinto
- Syncretism
- Animalism
- No Religion

Once this was done, we aggregated the population in each of the countries and found the majority religion for each of them. We then plotted them into a world map, on our dashboard, color-coding each of the countries according to the majority religion. The map changes as a time-series from 1945 to 2010.

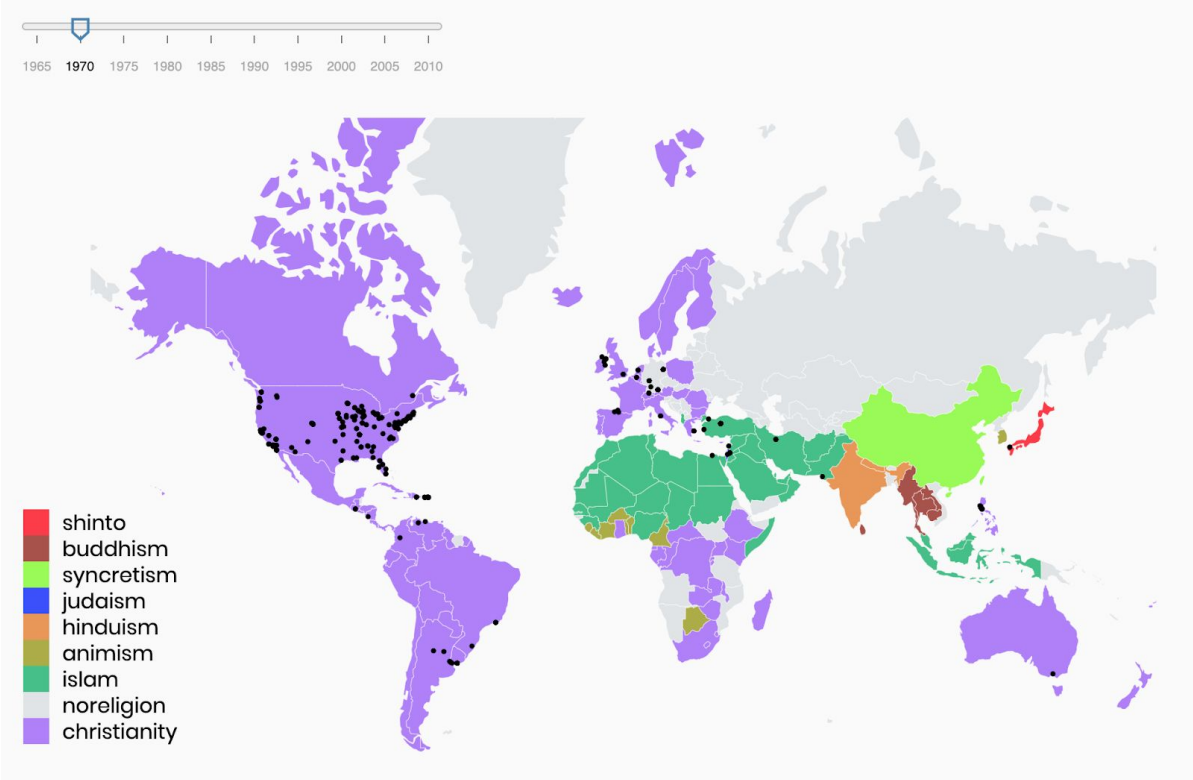


World Map of religious population distribution

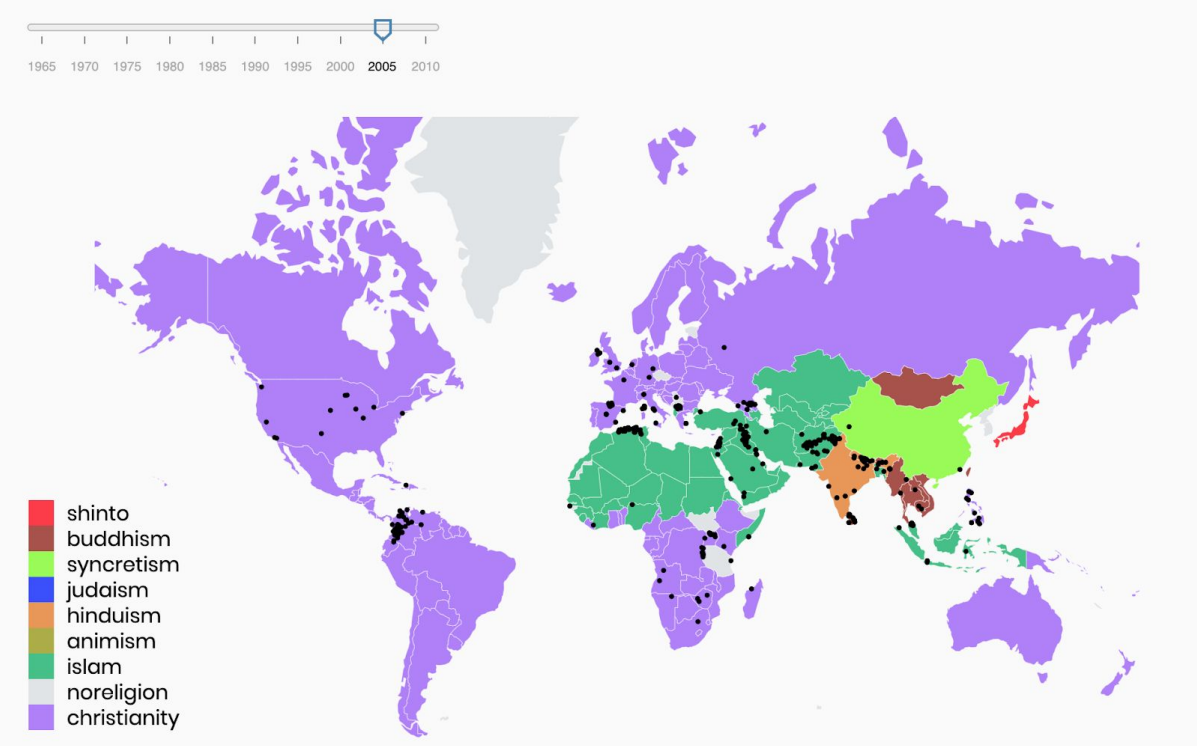
Hovering over a single country gives its major religion, population, and the name of the country.

A tooltip over a country

We then joined this dataset with the list of terror attacks from the global terrorism database, to analyze how this changes across time. The black dots which we can see above showcases terrorist attacks.
We observed that the cluster of terrorist attacks shifts from American regions to the Middle East as we move from the 1970s into the early 2000s. We can see that as below.



Terrorist attacks mapped to religion from 1965-1970



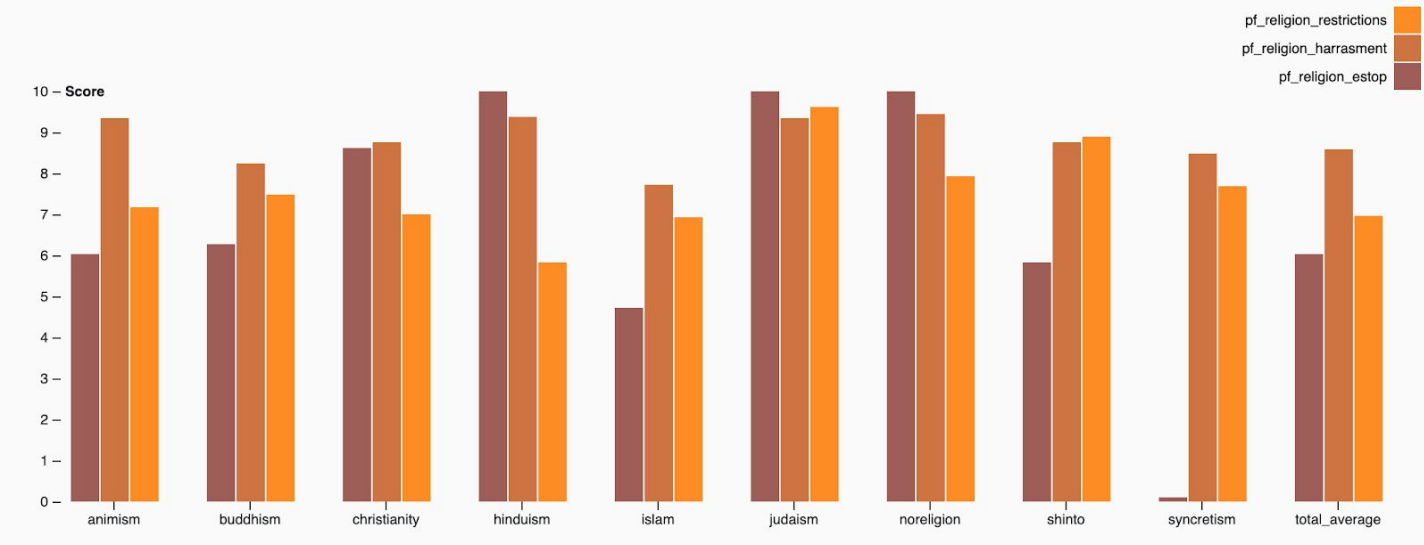
Terrorist attacks mapped to religion from 2000-2005

Personal Religious Freedom Dataset

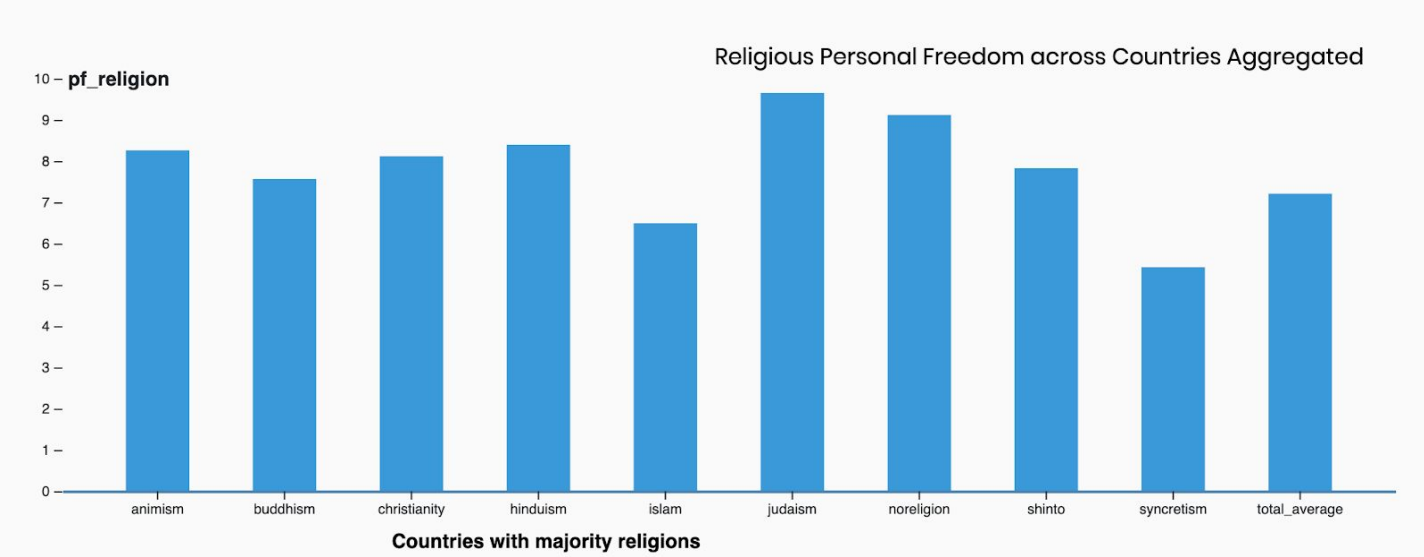
Next, we wanted to join the person religious freedom data from the human freedom index dataset and join it with the religious population distribution. Personal religious freedom is defined by 3 major factors

- Freedom to establish and operate a religious organization
- Religious Harassment
- Religious Restrictions

The factors are combined into a single score called Personal Religious Expression Score. Our religious population dataset and the human freedom index dataset mapped only for the year 2010. So we decided to visualize for this year.
On joining the 2 datasets, we cleaned it up by removing the NAN values and used grouped bar charts to visualize them.



All these scores were normalized according to the population of the country since we are talking about personal religious freedom here.
We also visualized the aggregated personal religious freedom as below.



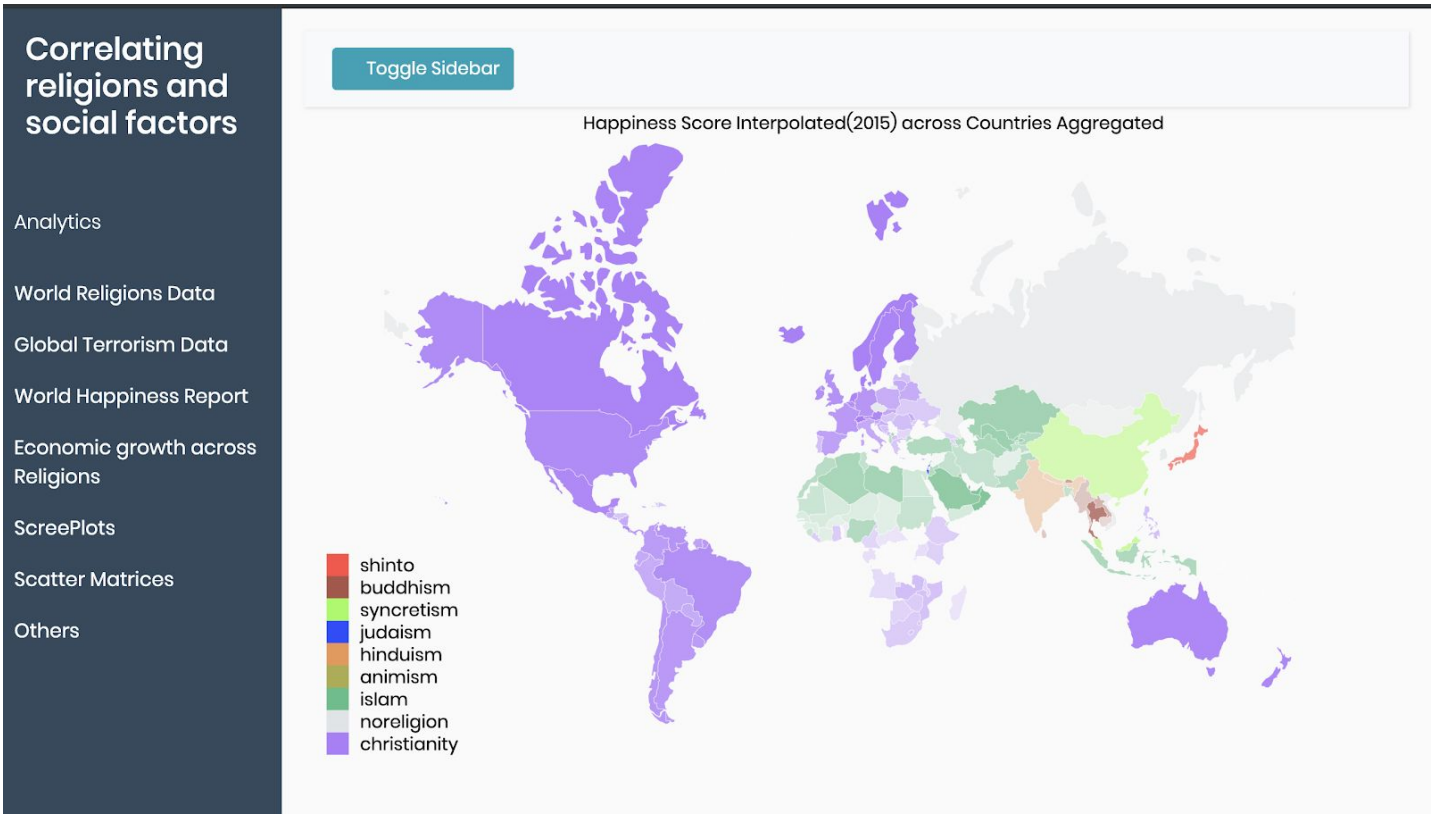
Happiness Score Dataset

We also did the preliminary analysis on the happiness dataset that we have wherein we faced a major conundrum.

Problem
The world happiness report was first published in 2012. So we do not have any data on the world happiness before that. The population by religion dataset that we have is from the 1940s to 2010.

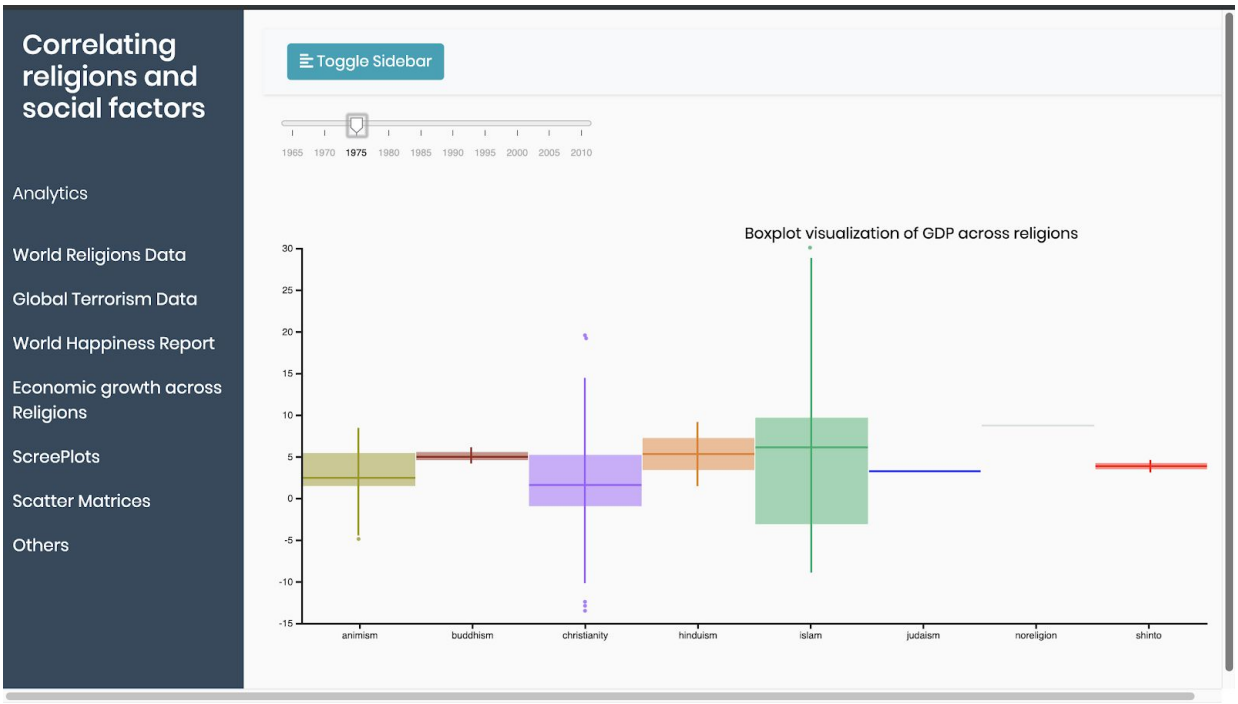
Solution
We planned to interpolate the data for the population to 2015 so that we could map the happiness score values with the religious freedom of a country.

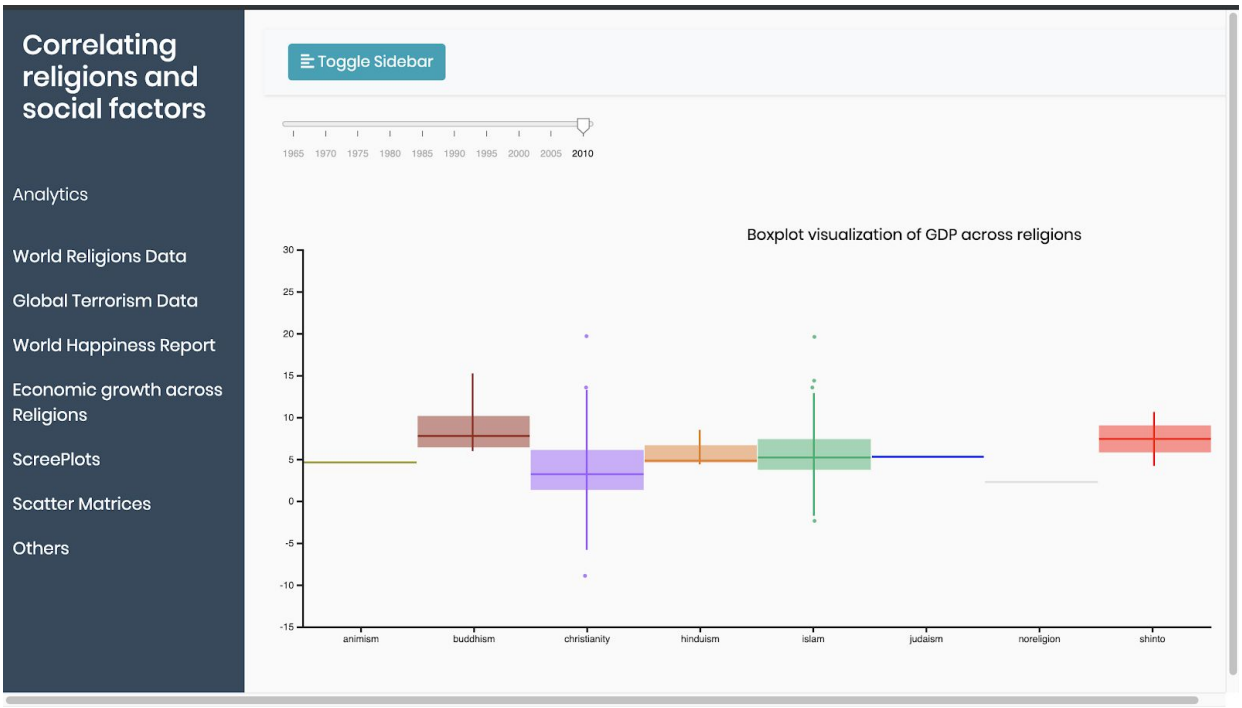
This would directly answer one of our questions.



Economic Growth across Religions

We wanted to find if there is any trend in how economic growth across religion is happening over time. For this, we added a new dataset provided by World Bank - **World Bank Growth DataSet [8]** which has the economic stats across countries from 1961-2017. We joined this with our World Religion Dataset, and use D3 BoxPlot visualization to observe relative growth trends across religions over time.





References

[1] Wikipedia contributors. "List of religions and spiritual traditions." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 16 Apr. 2019. Web. 16 Apr. 2019
[2] By Cato Institute, the Fraser Institute, and the Liberales Institut at the Friedrich Naumann Foundation for Freedom, "The Human Freedom Project", https://www.cato.org/human-freedom-index-new, 2018
[3] Zeev Maoz, University of California-Davis, and Errol Henderson, Pennsylvania State University, and published by the Correlates of War Project, University of Michigan, "Correlates of War: World Religions" http://www.correlatesofwar.org/data-sets/world-religion-data, 2019
[4] Sustainable Developments Network Solutions, "World Happiness Report", https://www.kaggle.com/unsdsn/world-happiness, 2018
[5] Felipe Argolo, "Religious Terrorist Attacks", https://www.kaggle.com/argolof/predicting-terrorism, 2016
[6] University of Maryland, "Global Terrorism Database", https://www.start.umd.edu/gtd/, 2017
[7] Terrorism: by Max Roser, Mohamed Nagdy and Hannah Ritchie, https://ourworldindata.org/terrorism 2018
[8] World Bank - Annual GDP Growth, https://data.worldbank.org/indicator/ny.gdp.mktp.kd.zg, 2018