
Cross-Religious Denomination Analysis

William Guo

Department of Computer Science, Economics
wrg2@rice.edu

Abstract

One may find the fact that constituents of different religions are, on the whole, very similar in their views of religion and ethics. Using this as a jumping-off point, I examined how groups of differing levels of religiosity emerge within a single denomination and compare the similarity of these groups across denominations. In particular, I am interested in seeing how socioeconomic factors impact people of different denominations. Using multiple component analysis (MCA), a special form of nonlinear principle component analysis (PCA), I synthesized my own measures of religiosity to create a more easily understandable, lightweight regression model.

1 Introduction

My initial interest stemmed from class discussion, where the point that conventional measures of religiosity, particularly religious service attendance, is increasingly misleading. For example, the unaffiliated population has no religious service to attend, but may practice a set of ethics as fervently as any conventionally religious person. And just as postulated in Weber's secularization hypothesis, Hout and Fischer (2002) note that the unaffiliated population has been growing. Furthermore, even within the affiliated population, church attendance has been decreasing as secular opportunities take precedence and other factors, such as the Internet, allow believers to practice their faith in unconventional ways not captured by this metric.

Other works in this literature, namely Woodberry (1998), have also observed discrepancies in the behavior of church-goers relative to their non-attending counterparts. Woodberry in particular focuses on the fact that church-goers are more cooperative and easier to extract answers from than those who do not attend. While we cannot correct this problem if it is the case, he proposes changing the loss function so that the answers between the two disjoint groups are weighted differently. My main takeaway, however, is the fact that current religiosity measures are not adequate to make the generalizations that people do with them.

Finally, the church-like and sect-like denominations Iannaccone (1988) introduced helped crystallize my initial exploratory data analysis because of how he sought to find patterns in the data not captured by conventional labels.

To better capture the ethical and religious views of the changing American landscape, I focused less on religious denomination itself and more on how people view others of similar or different religious beliefs through cluster analysis. Interestingly enough, my initial analysis showed slight clustering based on denomination, but nothing conclusive enough to consider religious groups as completely disjoint sets of people.

I also tackled the problem of inadequate measures of religiosity by synthesizing religiosity variables that each tackled different aspects of religiosity while remaining topical. By doing so, I hoped to create a religiosity measure that addressed most of the shortcomings of each individual measure.

2 Data Analysis

To reaffirm work done by Barro et al. (2010) on conversion rates across different religious denominations, I ran t-SNE on nine hand-picked survey questions from the World Value Survey (the first nine variables listed in the following table) regarding their views on their own religion and how they felt about people of different religions.

Table 1: Variables of interest

Variable code	Description
V9	Importance in life: religion
V41	Would not like to have as neighbors: People of a different religion
V70	It is important to this person to think up new ideas and be creative; to do things one's own way
V73	It is important to this person to have a good time; to "spoil" oneself
V78	Looking after the environment is important to this person; to care for nature and save life resources
V79	Tradition is important to this person; to follow the customs handed down by one's religion or family
V106	How much you trust: People of another religion
V150	Meaning of religion: To follow religious norms and ceremonies vs To do good to other people
V151	Meaning of religion: To make sense of life after death vs To make sense of life in this world
V152	How important is God in your life
V153	Whenever science and religion conflict, religion is always right
V154	The only acceptable religion is my religion
V156	People who belong to different religions are probably just as moral as those who belong to mine

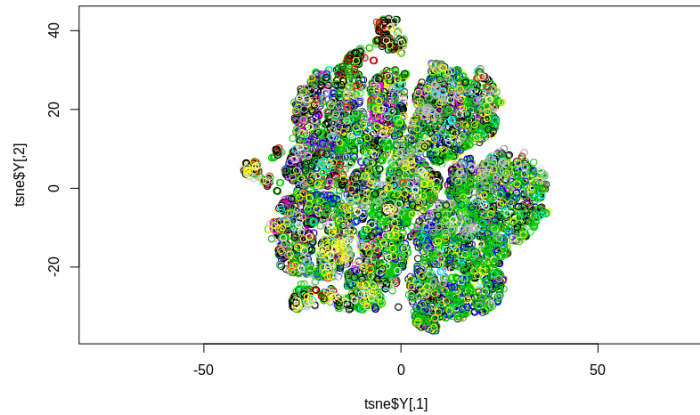


Figure 1: t-SNE plot of religious and ethical opinions clustered by denomination

Although my initial conclusion was that there was little clustering based on denomination, a closer examination and a change in variables revealed that there were slight groupings of ostensibly similar religions. For example, we see that the unaffiliated subjects are closely tied to the Buddhist subjects. Since Buddhists do not espouse a belief in a particular deity, they are commonly seen as one of the more secular religious people.

Given these preliminary results, I took the thirteen religiously motivated WVS questions and chose n of them for analysis. In doing so, I hoped to find the combination of variables that created the most noticeable separations across denomination. While many of the variables directly asked the subjects about their personal religiosity, I made sure to include questions that would force them to consider how they framed their religious preferences upon others and the world around them. More

specifically, I wanted to capture their attitudes towards their personal responsibility in being pious and their attitudes towards how pious others should be and how much their religion should be respected.¹

As one would expect, experimenting with different combinations of these thirteen questions led to varying levels of separation. The combination of more similar variables, and conversely the exclusion of the more exotic questions, led to better clustering. More specifically, the questions that asked more direct questions about religiosity led to better clustering, while the questions that dealt with more ambiguous ethical questions led to minimal cluster separation.

One could attribute this benefit to the fact that the questions I excluded in the “best” t-SNE plot did not adequately measure religiosity,² but I also believe the nature of how the questions can be answered could have limited the range of answers the subjects could have given. In the given example, the subject is presented with a choice between two options, whereas the variables that were kept allow the subject to more freely express their personal preference.

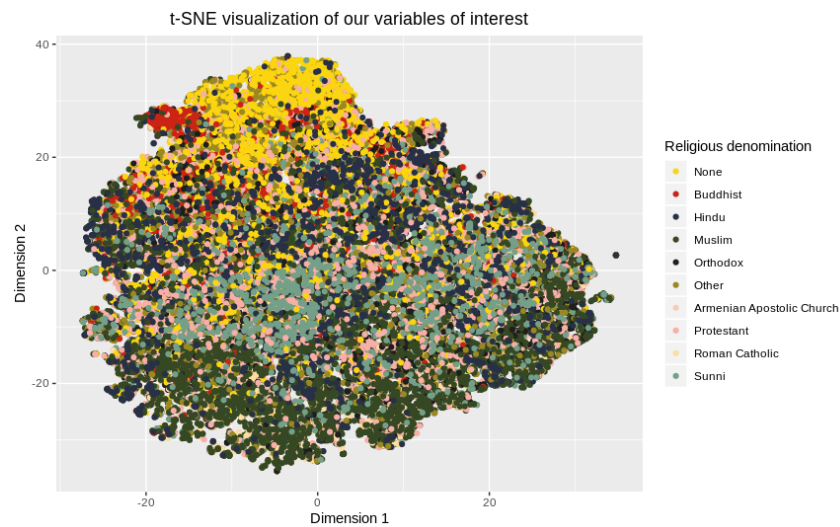


Figure 2: Plot of master variable list excluding V151

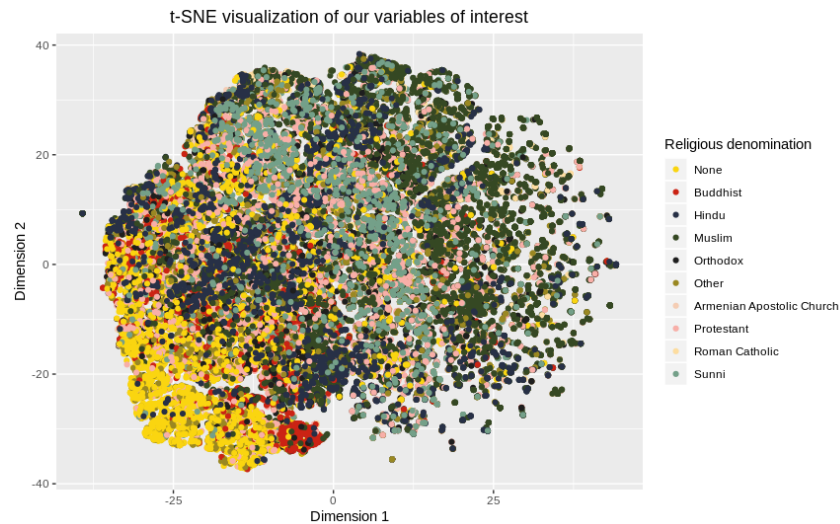


Figure 3: Plot of master variable list excluding V151, V70, and V73

¹A good example of the latter case would be the question that asked whether the subject believed if people of other religions were just as moral as those in their religion.

²From the master table of variables I excluded V70 and V73.

With these two plots, we can see how the general religious denomination clusters are still mostly the same, but there is a degree of separation in the second plot not found in the first.

3 Dimensionality reduction

While one would think that using the dimensions of the t-SNE calculations for some regression model would be a natural next step, they are unfortunately unusable. Due to the stochastic nature of the algorithm, there is no deterministic mapping between the high-dimensional space the algorithm takes and the 2D/3D space it projects the final plot in. Therefore, while the results of Section 2 are descriptive, we cannot use them for any kind of serious inference. Instead, I used the variables that led to the maximal separation in the dimensionality reduction algorithm of my choice, which I explain in the next passage.

Since t-SNE is a nonlinear visualization technique, it would not make sense to perform PCA, which maps high-dimensional data linearly into a lower-dimensional space. PCA is naturally incompatible with categorical data as there is no concrete scale with which to measure the strength of feeling towards certain questions. Instead, multiple correspondence analysis (MCA) is performed. It is possible to binarize each of a variable's answer choices and perform PCA on the modified matrix, but doing this procedure by hand does not make much sense when an automated alternative exists. The following represents the variance captured by pushing the aforementioned variables through MCA. I have also included the same for the variables that appeared in the control t-SNE plot.

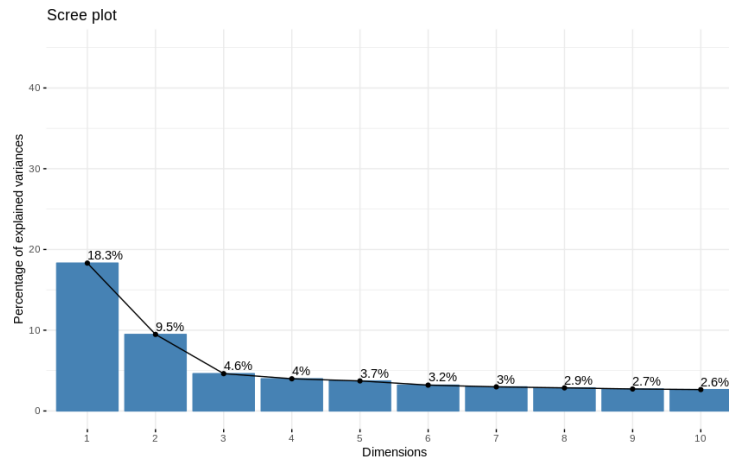


Figure 4: Amount of variance captured by the components of the good set of variables

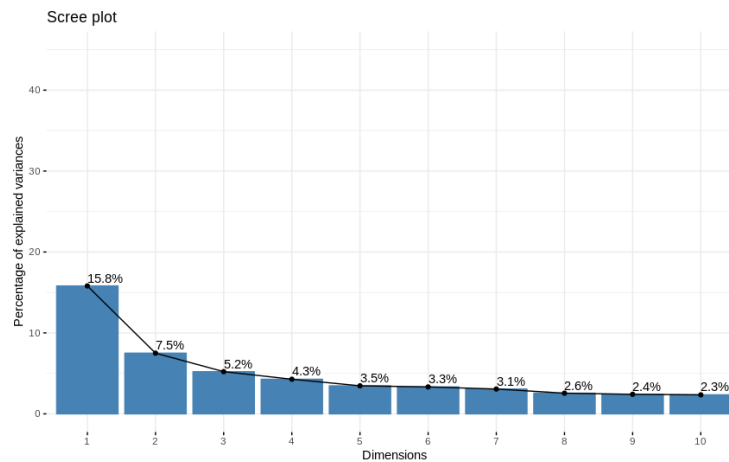


Figure 5: Amount of variance captured by the components of the control set

Looking at these graphs, we can confirm our previous intuition as the amount of variance explained by the first two components is greater for our smaller, more religion-oriented variables. With this in mind, we can use our chosen set of variables for model inference with confidence.

4 Regression model

Looking at the biplot that results from running MCA on the good variable set, we see that some variables, namely V153 and V154, are mostly unexplained by either component individually, reflecting a failure to adequately explain the variance in the data. This result is not unexpected, however; with $n=55140$ and a relatively low number of religiosity measures, the data is most likely too sparse. As such, we will treat the regressions as more of a thought exercise.

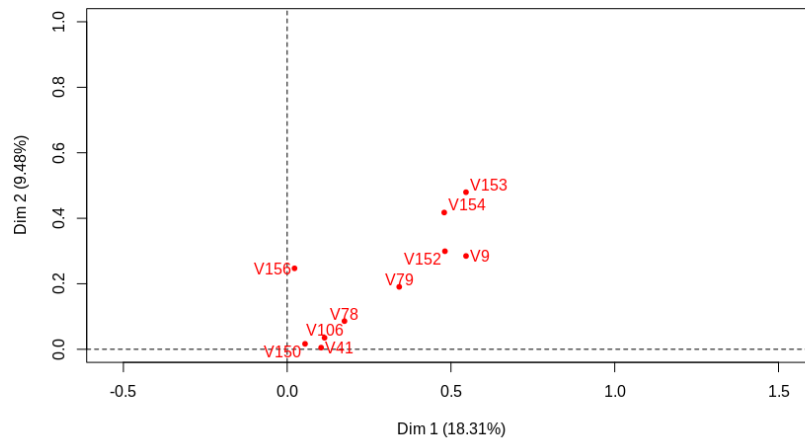


Figure 6: Biplot clusters based on how much each dimension explains the given variable

I will represent the first component as the subject's personal beliefs about members of other religions (now referred to as *outer*). We can see a small cluster of a few variables, namely V150, V106, 41, and 78. What these variables have in common is the fact that they all ask the subject about things external to their own personal decision to belong in a certain faith, whether it is to care for the environment around them or judge members of other religions more critically.³

³Because the second dimension captures less variance than the first and does not have any noticeable variable clusters, I will forego including it in our model.

5 Regression results

To clarify, the `libertarianism` variable was a question that asked the subject to measure their feelings between the sentiment that the government should take more responsibility to ensure that everyone is provided for or that people should take personal responsibility. `income_eq` represents their feelings on whether incomes should be made more equal, and `competition` asks whether competition is good or harmful. The rest of the variables, e.g. `age`, `sex`, `educ`, all represent what their names actually mean.

Table 2

	<i>Dependent variable:</i>
	outer
<code>age</code>	0.003*** (0.0001)
<code>sex</code>	−0.016*** (0.004)
<code>educ</code>	0.054*** (0.001)
<code>income_eq</code>	−0.031*** (0.001)
<code>libertarianism</code>	0.019*** (0.001)
<code>competition</code>	0.020*** (0.001)
Constant	−0.422*** (0.012)
Observations	55,140
R ²	0.107
Adjusted R ²	0.107
Residual Std. Error	0.506 (df = 55133)
F Statistic	1,099.744*** (df = 6; 55133)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

Perhaps due to the huge sample size, we interestingly see that all of our coefficients are significant. In particular, education seems to have a strongly positive correlation with values of `outer`.⁴ Age is also weakly positively correlated, which confirms our previous discussion about how people tend to become more secular as they enter their adulthood and become more religious as they move towards retirement.

Interestingly enough, those who lean “libertarian” and believe competition is good for the overall marketplace are also more secular, while those who believe in more income inequality are more religious. Considering the current political climate of the US, I was expecting the signs on `libertarianism` and `income_eq` to be flipped, but since the data is over the entire world, it is not surprising to see such a result. Another contradictory result lies in the coefficient on the `sex` variable, which indicates that women are more secular than men. Given our first assignment was based on the *a priori* assumption that women attended church more often (and were therefore more religious), I was a little surprised by the result.

⁴The scale of the original variables mapped lower integer values to lower levels of religiosity, so education here seems to be correlated with lower religiosity

6 Conclusion

My regression results mostly confirm common ideas we have discussed throughout the semester: age is loosely positively correlated with secularism and education is strongly so. However, like I mentioned previously, since the first component itself was not very representative of the variables I named, I do not have much confidence in any of my results, whether they confirmed or contradicted my *a priori* beliefs regarding one's personal religiosity.

Given the variables I introduced and subsequently excised, ethical questions do not seem like a good way to measure religiosity. While ethics and religion are naturally intertwined, the relationship captured in the WVS data was not strong enough to make including the ethical questions a worthwhile addition. The entire set of variables fail both the eye test of making distinct clusters as well as capturing as much variance as the good set of variables. That being said, the good set of variables also fails to truly capture enough variance to be meaningful, which could point to problems in the data itself. Also, while I had initially thought that religious denomination would create few or no clusters, they ended up clustering in somewhat reasonable ways. Given that the data used is a huge sample drawn from groups all across the world, one could argue that regional effects might have washed out. However, having never used nonlinear PCA/MCA or t-SNE in practical applications, I am happy that I got the chance to put the algorithms to use and unpack common issues that we discussed in class but never examined in depth.

References

- Barro, R., McCleary, R., and Hwang, J. (2010). Religious conversion in 40 countries. *Journal for the Scientific Study of Religion*, 49:15–36.
- Hout, M. and Fischer, C. S. (2002). Why more americans have no religious preference: Politics and generations. *American Sociological Review*, 67:165–190.
- Iannaccone, L. R. (1988). A formal model of church and sect. *American Journal of Sociology*, 94:241–268.
- Iannaccone, L. R. (1992). Religious markets and the economies of religion. *Social Compass*, 39:123–131.
- Iannaccone, L. R. (1994). Why strict churches are strong. *American Journal of Sociology*, 99:1180–1211.
- Inglehart, R. and Baker, W. E. (2000). Modernization, cultural change, and the persistence of traditional values. *American Sociological Review*, 65:19–51.
- Montgomery, J. D. (1996). Dynamics of the religious economy: Exit, voice, and denominational secularization. *Rationality and Society*, 8:81–110.
- Sahgal, N. (2018). Being christian in western europe. *Pew Research Center*.
- Stark, R. and Iannaccone, L. R. (1994). A supply-side reinterpretation of the "secularization" of europe. *Journal for the Scientific Study of Religion*, 33:230–252.
- Woodberry, R. D. (1998). When surveys lie and people tell the truth: How surveys oversample church attenders. *American Sociological Review*, 63:119–122.