Synthesizing a better religiosity measure

William Guo

Rice University

April 11, 2019

Initial concerns

- Traditional measures of religiosity suffer from misclassifying the growing secular population around the world
- Many of the questions about religion are more biased towards a Western audience, e.g. church attendance, belief in Hell
- **Question:** how can we disentangle religious denomination from measures of religiosity?

Motivation

- When measuring religiosity, there is a tradeoff between how easy a question is to measure and how accurately it pinpoints one's true religiosity
  - E.g. attendance is easily measured, but is not comparable between members of different religions
  - On the other hand, more ambiguous questions that can address many different religions are then less quantifiable
- Adding more variables will decrease the interpretability of our models

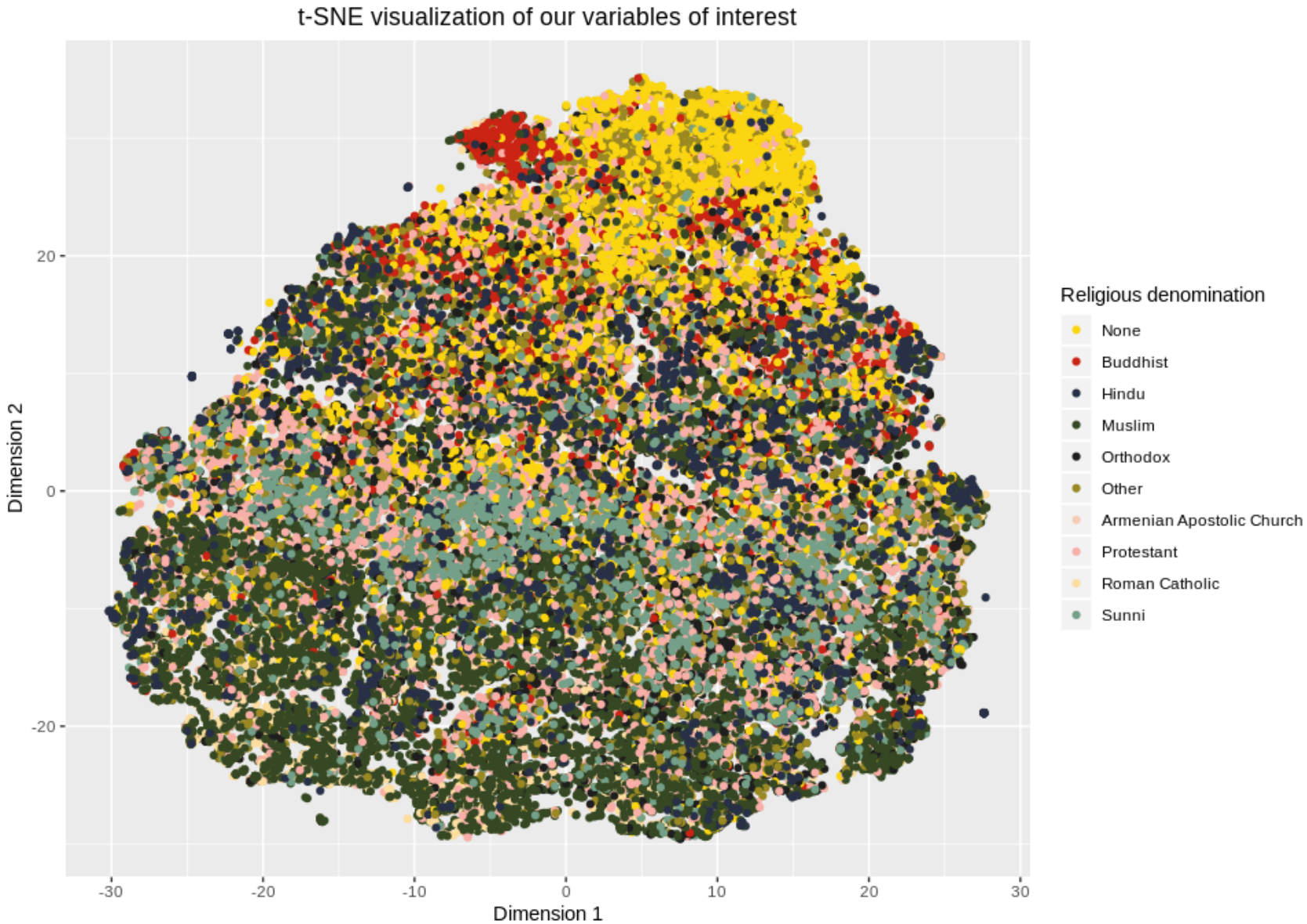
Solution

- We can synthesize our own variables through factor analysis!

Exploratory data analysis

- Performed t-SNE on a group of questions in the WVS regarding religion and ethics to visualize whether any denomination-based clusters emerged from the data
  - Finds and minimizes the KL-divergence between the joint probability distribution of two points in the original, high-dimension space and two points in the lower-dimensional space (usually 2D for visualization purposes)
- **Question:** do we expect members of one religion to answer differently from another's?

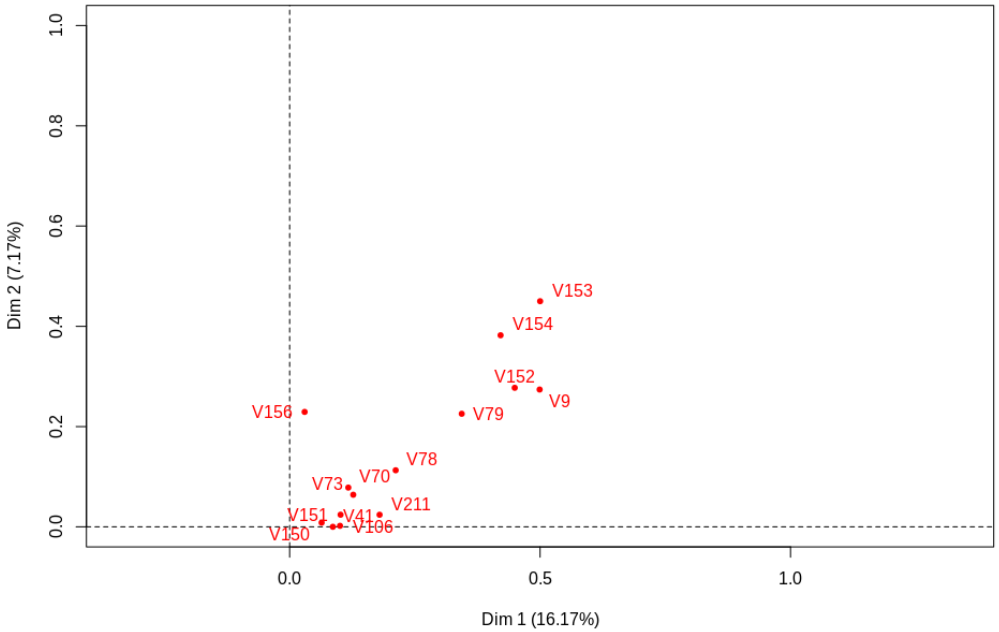
Exploratory data analysis (cont.)

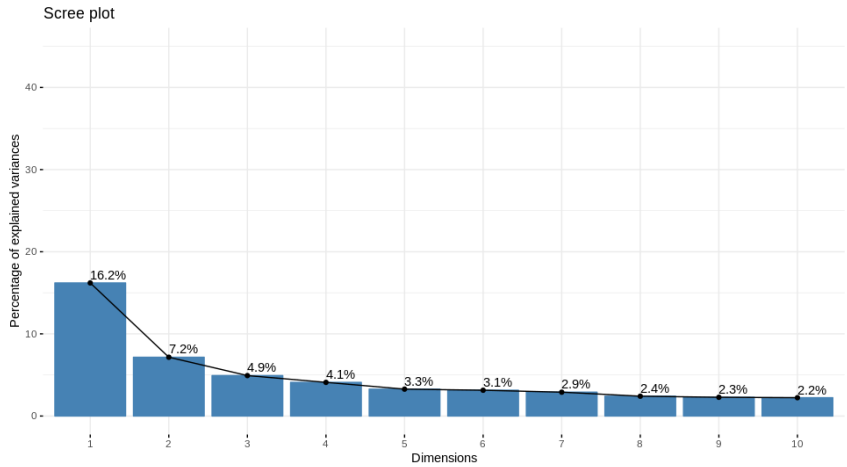


Dimensionality reduction

- PCA is not quite theoretically sound due to the fact that we are dealing with categorical variables and it requires scaled data
- Instead, I use multiple correspondence analysis (MCA)
  - Performs singular value decomposition on the indicator matrix of the data (instead of the covariance matrix like in PCA)
  - In essence, makes dummy variables out of the original variables
  - Could have achieved a similar result by binarizing the data by hand and running PCA on it

Dimensionality reduction (cont.)





Model proposition

- First dimension explains the person's religious feelings (`inner`)
  - V150 and V151 ask the subject their opinion on the purpose of religion (existential, uphold traditions, do good things)
  - V73 asks if it is important to spoil oneself
  - V211 asks if they are proud of their nationality
- Second dimension deals more with how the person feels about people of different religions (`outer`)
  - V156 asks the subject if people of other religions are just as moral as their religion's members
  - V153 deals with whether religion is superior to science when they conflict
  - V154 questions whether the only acceptable religion is one's own

Regression results

	Dependent variable:	
	`inner`	`outer`
	OLS	*instrumental variable*
	(1)	(2)
age	0.004^{***}	
	(0.0001)	
sex	0.007^{*}	
	(0.004)	
educ	0.048^{***}	
	(0.001)	
inc_ineq		-0.029^{***}
		(0.007)
gov_resp		0.006^{***}
		(0.002)
unreg_comp		-0.008^{***}
		(0.001)
Constant	-0.448^{***}	0.161^{***}
	(0.010)	(0.033)
Observations	55,140	55,140
R^2	0.063	-0.027
Adjusted R^2	0.063	-0.027
Residual Std. Error (df = 55136)	0.465	0.397
F Statistic	$1{,}243.375^{***}$ (df = 3; 55136)	
Note:		*p<0.1; **p<0.05; ***p<0.01

Big takeaways

- Data exploration showed some small clusters, but on the whole shows that religious denomination was a poor way to cluster the responses
- MCA was unable to sufficiently explain the variance in the subjects' responses
- Coefficients are all very significant, but are also very small and do not explain the highest-variance components of the religious questions

Future work

- Find a better algorithm that can create better components
- Work with different variables