Deus Ex Machina: Using an Algorithm to Create Religious Categories

Abstract:

For decades one of the most difficult problems facing scholars of religion is how to classify individuals into simplified, yet meaningful categories. Over time a number of classification schemes have been proposed with RELTRAD being the most popular of these methods. However, in recent years the explosion of machine learning has allowed researchers the ability to use algorithms that have no pre-conceived notions or inherent biases to help illuminate problems in the social sciences. Using the technique of k-means clustering to sort individuals into six religious categories, this note offers a slightly different way for scholars to think about religious classification. In the end, the algorithm sees five religious categories, and believes that the distinction between mainline Protestants and Catholics as largely non-existent, while evangelical Protestants are a much more heterogeneous group than has previously been described.

Classifying Religion

As long as social scientists have tried to understand the impact that religion has daily life, they have struggled with how to classify the religious experience. The earliest work in the field had recognized that religiosity means something vastly different based on the type of religion being observed. For instance, Emile Durkheim used the variation in suicide among European nations to argue that there was something inherently different about Protestant, Catholic, and Jewish religiosity (Durkheim 1897). Durkheim's contemporary, Max Weber, pushed this distinction further by arguing that a specific subset of Protestant Christianity, Calvinism, drove adherents to embrace capitalism for a number of theological reasons. This ethic of work was not as apparent among non-Calvinist Protestants (Weber 2002). In both cases, Durkheim and Weber argue classifying individuals based on their religious tradition or theological outlook was crucial to understanding the nuances of the religious experience.

A few decades later social scientists began to take up this problem in earnest. In his canonical work on tolerance in the United States, Samuel Stouffer understood this difference as largely one of geography. He divided Protestants and Catholics up into northern and southern varieties, noting that Protestants who lived in the South attended church more frequently than their Northern counterparts (Stouffer 1955). While Stouffer's demarcation was described as geographic, it is likely the first attempt at delineating what would be later described as mainline and evangelical Protestants.

In the middle part of the 20th Century classifying religion lurched forward in a disjointed and unfocused manner. other scholars tacitly embraced this denominational family approach and expanded on its usage to sort religious traditions. For example, some used inter-church associations as a deciding factor (Jacquet 1991), others used survey responses to theology questions from both clergy (Hadden 1969) and laity (Hunter 1981), while still others utilized the official church documents regarding orthodox beliefs (Melton and Geisendorfer 1977).

Yet, other scholars pushed back against this religious tradition approach and argued instead for a continuum that would place each denominational family on a tripartite scale of liberal – moderate – fundamentalist (Smith 1990). This approach (called the FUND classification) was touted to be highly

parsimonious, however, several denominations were placed in an "excluded" category as they were not readily classified into one of the three categories in FUND. However, soon after it's publication other members of the scholarly community argued that FUND is imprecise and overly reductive as would be required when placing all religious traditions in three categories (Kellstedt et al. 1996). In response, these scholars have proposed to combined the FUND categories with the denominational family approach to create 18 individual subgroups. In this conception, each religious tradition would be divided into traditionalist, centrist, and modernist Catholics or evangelicals, for example (Green 2004).

However, all those approaches largely faded from prominence upon the development of the RELTRAD classification scheme which was published by Steensland et al. (2000). This scheme struck a middle ground between the three simplified categories of FUND and the 18 subgroups proposed by John Green to create seven religious traditions based on denominations. These seven denominations (evangelical Protestant, mainline Protestant, black Protestant, Catholic, Jewish, other religion, and no religion) were novel in that they divided Protestant Christianity based on both theological and racial lines. The RELTRAD approach has been widely embraced by social scientists (Dougherty, Johnson, and Polson 2007; Frendreis and Tatalovich 2011), having been cited over 1,000 times in scholarly literature from a variety of sub-fields.

Other scholars have noted some shortcomings in the RELTRAD approach. For example, RELTRAD uses racial filters in some instances along with church attend requirements for non-denominational Protestants, something that conflates religious belong with religious behavior (Hackett and Lindsay 2008). Others have noted that some of the syntax provided by the RELTRAD authors to code the General Social Survey contained errors which could lead to different statistical results (Stetzer and Burge 2016).

Despite these reservations, RELTRAD has remained largely unchallenged as the de facto scheme to classify religious traditions in the social science literature. However, in the last few years, statistical analysis has grown exponentially in its sophistication and accessibility. Most notably, machine learning has provided a way for researchers to revisit many of the assumptions made by prior scholars to see if artificially intelligent algorithms see the lines of demarcation in the same way that scholars do. One particularly useful machine

learning technique, k-means clustering, allows a researcher to feed a number of response items from a large number of respondents into an algorithm which clusters individuals into specific groups that make the most sense from a statistical standpoint. Using the k-means clustering technique, this work explores the seven groups that are created as part of RELTRAD to see if machine learning generates religious typology in the same way that the academic community views this classification problem.

K-means Clustering

K-means clustering is a type of unsupervised machine learning, meaning that the algorithm is not given any type of classification markers but must instead infer its own categorization scheme given the variables included in the dataset (Basu, Banerjee, and Mooney 2002). The algorithm is given the number of clusters to generate by the user then iterates over each observation in the data as a means to center each respondent as closely as possible to the middle of each cluster so that "the within-cluster sum of squares is minimized" (Hartigan and Wong 1979, 100). The end result is each individual in the dataset is sorted in one of the clusters that were created and each centroid has a mean computed for each of the variables that are included in the training data (Trevino 2016).

There are many cases where k-means clustering can be especially helpful, most notably in the business world. For example, retailers use k-means to divided its customer base into different segments which can then be targeted with promotions or coupons to drive up sales (Kuo, Ho, and Hu 2002). Over companies have used k-means clustering to generate recommender systems for e-commerce websites based on prior purchase history and other demographic variables (Kim and Ahn 2008). Other statisticians have used k-means clustering to help the sensors in a smartphone determine what type of physical activity is occurring as a means to track exercise by the phone's userbase (Yang 2009).

Despite the overall utility of this classification technique, the use of k-means clustering is just beginning in the social sciences. For instance, researchers in education used k-means to divided community college students into five different categories which can then be utilized by academic support professionals to target services with a goal of increasing graduation rates (Bahr, Bielby, and House 2011). However, its utility in the study of religion has yet to be explored. K-means is well suited to the task of classifying people into

groups as it designed to find points of demarcation between users who may have a great deal of similarity. This makes it ideal for testing our previous understanding of religious classification.

Data and Measures

In order to create an ideal comparison between the clusters created by k-means and the groups generated by the RELTRAD typology, the General Social Survey was chosen as a data source. When Steensland et al (2000) published their article justifying the use of RELTRAD they included syntax for recreating the typology in the General Social Survey for a number of statistical analysis programs. This syntax has been reviewed and amended over time (Stetzer and Burge 2016) but stands as a well-suited reference for comparing any new religious classification system. While RELTRAD coded respondents can be created over four decades of the GSS, the meaning of religious traditions has obviously shifted over time and to minimize this reality the dataset was reduced to contain respondents from 2008 to 2016. This dataset contains 11,446 respondents, a number that is more than sufficient to generate meaningful clusters.

In total, fourteen variables were chosen as a means to undertake the clustering process. These variables can basically be divided up into demographic characteristics and religious/political variables. Those items that measuring demographics were: age, gender, a dichotomous variable for white respondents, a dichotomous variable for African-American respondents, education, and income. The religious/political variables include church attendance, a respondent's view of the Bible broken into three dichotomous variables, a measure of support for gay marriage, support for abortion rights, a measure of political tolerance, and a measure of party identification.[1] Before these variables can be utilized they must be placed on the same numerical scale, so that the algorithm does not place more weight on those variables with a wide range of values. Most notably each of the variables included (often called features) must be scaled in a similar fashion(Visalakshi and Thangavel 2009). In this case, each continuous variable was scaled from 0 to 1 to match the dichotomous nature of many of the items that were included. In addition, k-means clustering cannot work with data that has missing values and they were excluded from the analysis (Li et al. 2004). As a result of this process, the final dataset contained 9,998 respondents.

---

[1] The full variable wording is available in the appendix. Recoding syntax is available at BLINDED

Determining the Number of Clusters

One of the shortcomings of k-means clustering is that it does not prescribe to the user the correct number of clusters to specify. Instead, that choice is left largely to the discretion of the analyst. There has been a considerable debate about which method is the most appropriate to aid in making this decision (Ketchen and Shook 1996; Sugar and James 2003), however, the oldest and easily interpretable approach is called the "elbow method" (Thorndike 1953). A model is estimated which determines the level of variance explained as one adds additional clusters to the analysis. This model was estimated for the data being examined with results being displayed in Figure 1.

INSERT FIGURE 1 HERE

As is apparent from this graphic, a great deal of variance is reduced by adding the first three clusters (as measured by the within groups sum of squares). This variance is still reduced a significant degree until one arrives at six clusters. At this point the line changes to nearly ninety degrees creating the "elbow" which gives this technique its name. One can also note that it is possible to move past six clusters, as there is a reasonable amount of variance reduction moving from six to eight total clusters, however this decision is considered a subjective one by researchers (Dutta et al. 2011) and to ease the simplicity of interpretation this clustering algorithm will use six clusters to separate individuals into groups.

Cluster Creation

Utilizing the 2008-2016 GSS sample of 9,998 respondents, the k-means clustering algorithm was conducted using six clusters as the optimal solution. This analysis was computed using the k-means function including with the R statistical software package. By default this function uses is the Hartigan and Wong algorithm, which is very similar to the MacQueen algorithm (MacQueen and others 1967) but is much faster at generating clusters and iterating over large datasets. The clusters were creating using 1,000 iterations and a number of models were specified using different starting points in the data to ensure replicability of the outcome. The results of one of those models is displayed in Table 1 below with the other models only differing a small amount from these results. Note for each cluster the mean value for each variable is reported, which aids in interpretation.

INSERT TABLE 1 HERE

One of the important considerations of clustering analysis is that the algorithm provides no rationale for the formation of each of the clusters, instead, that justification for each must be inferred by the analyst. In this case, it would be helpful to make some general comments about the nature of the clusters. First, the sizes of the clusters all range from approximately 1000-2000 respondents, which indicates that including additional clusters could likely make the small clusters even smaller and the larger clusters might make up a disproportionate size. It is also valuable to note which variables do not differ significantly between the six clusters. For instance, age, income, and education have very little variation across all clusters. In fact, for each of these demographic factors never vary more than 11% from the lowest value to the highest value across all six clusters. The other variables are of much greater consequence.

Cluster #1 and Cluster #5 are almost exactly alike. Both contain samples that are predominantly white (86%), with ages and incomes that are indistinguishable. Also, these two clusters attend church at essentially the same rate and both clusters contain nearly all biblical literalists (94% and 98% respectively). On social issues, they are very conservative, but this is especially the case for opposition to abortion rights. In terms of party identification, these clusters are the most conservative. The only difference here is that Cluster #1 is entirely male, while Cluster #2 is entirely female. It seems that the most clear inference is that Cluster #1 is white, male, evangelicals, while Cluster #2 is white, female, evangelicals. It is worthwhile to consider that while the algorithm splits these seemingly two identical groups on gender lines, that is the only difference between the two. If RELTRAD comports with this algorithm it should sort many evangelicals into Cluster #1 and Cluster #5.

Cluster #2 and Cluster #6 are also similar in many of the ways that the prior two clusters share commonalities. For both clusters income, race, education, and church attendance are statistically indistinguishable. In addition, both clusters are almost completely filled with people who believe that the bible is the inspired word of God but should not be taken literally. However, moving into the area of politics some divergence is revealed. There are vast differences in the tolerance level, with Cluster #2 being the least tolerant of all the clusters, while Cluster #6 is the most tolerant. On both gay marriage and abortion, Cluster

#2 is extremely conservative. In fact, Cluster #2 is the most conservative group on social issues in the entire sample, this is despite the fact that no one in this cluster believes that the bible is literally true. In contrast, while Cluster #6 also contains no biblical literalists, their support for gay marriage and abortion is the highest of any of the seven groups. This becomes even more puzzling when considering that Cluster #2 is only 8% more conservative from a party identification standpoint than their counterparts in Cluster #6. It would be accurate to describe these two clusters as mainline Protestants or Catholics when considering their theological orientation, but Cluster #2 would be socially conservative mainline Protestants or Catholics, while Cluster #6 would be socially liberal Christians. It's also possible that some evangelical Protestants who are more moderate in their theology, but more conservative in their politics might find their way into Cluster #2.

Cluster #3 was clearly created by the algorithm as a catch-all for those in the sample who believe that the Bible is a book of fables (the most liberal option in the GSS). In fact, nearly 99% of all individuals who have this perception of the bible were sorted into Cluster #3. Looking at the other variables, this cluster attends church at a rate that is 20% lower than any other group in the analysis. Regarding politics, this group is the second most Democratic in the party identification, while their support of gay marriage and abortion rights are the highest of any cluster in the sample. It is apparent from these indicators that this cluster is compromised of non-church going individuals, or those who do not come from a Christian background and therefore have a more liberal view of the Bible. This cluster should be comprised largely of those who indicate "no faith" or "other faith" in the RELTRAD classification scheme.

Finally, Cluster #4 diverges significantly on one crucial variable: racial composition. In the other five clusters, they range from 86% to 91% white, while Cluster #4 has zero white respondents. Instead, this group is 99% African-American. This cluster has relatively high church attendance rates, but their belief in the bible is distributed relatively equally across the three response items. This group shows moderate levels of tolerance, as well as moderate support for gay marriage and abortion rights. However, despite the fact that their position on social issues places them in the middle of the pack, their party identification is much more Democrat-leaning than any of the other clusters. This result clearly describes the contours of Black Protestants as described by RELTRAD.

To take a step back, the k-means clustering algorithm appears to see the religious landscape in a very similar way as the RELTRAD classification scheme. Clusters #1 and #5 describe all the touchstones of white, evangelical ideology and theology, with high levels of religiosity and more conservative views on social issues. On the other hand, Clusters #2 and #6 are indicative of two types of theological moderates: those who are liberally politically and those who are more conservative. This should be where many mainline Protestants and Catholics from RELTRAD find themselves. Cluster #3 is composed of those who are both theologically liberal and religiously inactive, while also showing a great deal of support for liberalized social policy. RELTRAD nones should be overwhelmingly in this category. Finally, Cluster #4 is almost entirely African Americans. As such, there is a diversity of viewpoints on matters of theology, but despite the fact that this group is the most Democrat in party identification, it is not the most liberal on social issues. Taken together, these results somewhat comport with the general conception of RELTRAD's seven categories, although the clustering algorithm essentially sees five groups as Clusters #1 and #5 could likely be collapsed if their only difference is gender. However, the true test for RELTRAD is if each of the categories it creates falls into the correct clusters that were just described.

RELTRAD in Clusters

As previously described, Clusters #1 and #5 contained many of the trademark characteristics of evangelicals: high levels of church attendance along with a literal view of the bible. In addition, these clusters were socially conservative on gay marriage and abortion. Figure 2 below displays how RELTRAD evangelicals are sorted and the results are somewhat mixed. While Clusters #1 and #5 do contain the highest percentage of evangelicals, just 50% of all evangelicals fall in those two clusters. Clusters #2 which contained theologically moderate, but socially conservative respondents captured about 20% of evangelicals, while Cluster #6 was another 14%. This provides a mixed bag. While many evangelicals are sorted into the typical evangelical clusters, nearly a third fall into clusters that look much more like their mainline or Catholic counterparts.

INSERT FIGURE 2 HERE

Catholic and mainline Protestants are included together, and their results are displayed in Figure 3. It is striking how the algorithm sorts these two groups into relatively the same proportions across the six clusters. Here, the results comport much more closely with the basic parameters of RELTRAD. As previously described, Clusters #2 and #6 represent the mainline approach theologically with Cluster #2 being socially liberal and Cluster #6 being socially conservative. Those two clusters make up nearly 60% of both mainline Protestants and Catholics. It is worthwhile to note that mainliners were slightly more likely to be in the socially liberal cluster than Catholics, but the difference is small (~3%). It is also worth considering that the next most frequent clusters (#3 and #6) have a completely different composition. Cluster #3 are non-church going theological liberals, while #5 is orthodox social conservatives.

INSERT FIGURE 3 HERE

The results for Black Protestants are unsurprising (the histogram is available in the Appendix). As described, Cluster #4 was 99% African American and the algorithm put 96% of Black Protestants into this cluster. It is clear that machine learning sees race as being a primary line of demarcation and that is born out in these results. Sorting Jewish respondents puts 52% in Cluster #3 which is the non-church attending and socially liberal group. A third of the remaining Jewish individuals were placed in Clusters #2 and #6, which consist of theological moderates who are either liberal or conservative on social issues respectively (these results are visualized in the Appendix, as well).

INSERT FIGURE 4 HERE

Turning our attention to those who are of other faith in RELTRAD and those who express no faith, the results are fascinating. Cluster #3, which is those who rarely attend church and have a low view of the Bible is nearly 55% of those who have no faith which is not surprising. What is somewhat counterintuitive is that nearly a quarter of those of no faith fall into the moderate religion clusters of #2 and #6. Turning to those of "other faith," the results are mixed. In fact, for this RELTRAD group the results are distributed more evenly across the six clusters than all the others. While the largest share is Cluster #3 at 28%, twenty percent are in Clusters #2 and #6. It seems that the cluster analysis does not know exactly how to sort this group, which is consistent with RETLRAD's "catch-all" philosophy toward the other faith category.

Discussion

  To view these results holistically, there are some areas in which machine learning sees religious traditions in the same way as RELTRAD, but these two approaches diverge in some respects. Clearly, there needs to be a sorting that occurs on racial lines, that fact is made clear in that 96% of black Protestants were placed in the predominantly African-American cluster. In addition, there needs to be a classification for those who have a liberal view of the bible and never attend religious services. These two categories are captured well in RELTRAD and also in this cluster analysis. In addition, Jewish respondent are clustered in a way that is consistent with their perceived political and religious outlooks.

  On the other hand, RELTRAD makes clear lines of demarcation around evangelicals, mainliners, and Catholics that are not as evident in the k-means clustering exercise. Consider the fact that only 50% of RELTRAD evangelicals are placed in the clearly evangelical clusters, it support the fact that this category is much more heterogeneous than the singular category of evangelical Protestants. When considering the more moderate clusters, social science needs to provide flexibility for those who are theologically moderate but could either be socially conservative or socially liberal. In addition, the cluster analysis provides support for the reality that there is quite a bit of consistency between Catholics and mainline Protestants. While obviously, their religious traditions differ, many take the same view of the Bible while being split in the same proportions on political matters.

  Finally, students of American religion need to seriously reconsider those who fall into the "no faith" and "other faith" category of RELTRAD. According to these results, while about half of those with no faith fall into the cluster of religious liberalism, that means that the half do not. In fact, a quarter of these individuals were sorted into the moderate Clusters #2 and #5. Also, a real reconsideration of the "other faith" group is justified from these results. This group is an amorphous, inconsistent category that contains no theoretical or methodological consistency.

  Obviously, machine learning algorithms are not a one size fits all tool that will solve all the problems that social scientists face. The kinds of problems that they can address and the solutions that they can generate are quite limited at this stage. However, they offer scholars something that has never been available

before: a truly objective, bias-less approach to some of the questions that have bedeviled us for decades.

When they are used in concert with strong theoretical underpinnings and solid research designs that can offer

new and interesting insight into the social world.

Bibliography

Bahr, Peter Riley, Rob Bielby, and Emily House. 2011. "The Use of Cluster Analysis in Typological Research on Community College Students." *New Directions for Institutional Research* 2011 (S1):67–81. https://doi.org/10.1002/ir.417.

Basu, Sugato, Arindam Banerjee, and Raymond Mooney. 2002. "Semi-Supervised Clustering by Seeding." In *In Proceedings of 19th International Conference on Machine Learning (ICML-2002.*

Dougherty, Kevin D., Byron R. Johnson, and Edward C. Polson. 2007. "Recovering the Lost: Remeasuring U.S. Religious Affiliation." *Journal for the Scientific Study of Religion* 46 (4):483–99.

Durkheim, Emile. 1897. *Le Suicide: Étude de Sociologie*. F. Alcan.

Dutta, Haimonti, Rebecca J Passonneau, Austin Lee, Axinia Radeva, Boyi Xie, David L Waltz, and Barbara Taranto. 2011. "Learning Parameters of the K-Means Algorithm From Subjective Human Annotation." In *FLAIRS Conference*.

Frendreis, John, and Raymond Tatalovich. 2011. "Validating Protestant Denominational Classifications Using the Chaves Inerrancy Scale." *Politics and Religion* 4 (2):355–65. https://doi.org/10.1017/S1755048311000150.

Green, J.C. 2004. *The American Religious Landscape and Political Attitudes: A Baseline for 2004*. Washington, DC: Pew Forum on Religion and Public Life.

Hackett, C., and D.M. Lindsay. 2008. "Measuring Evangelicalism: Consequences of Different Operationalization Strategies." *Journal for the Scientific Study of Religion* 47 (3):499–514.

Hadden, Jeffrey K. 1969. *The Gathering Storm in the Churches*. [1st ed.] edition. Doubleday.

Hartigan, John A, and Manchek A Wong. 1979. "Algorithm AS 136: A k-Means Clustering Algorithm." *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 28 (1):100–108.

Hunter, James Davison. 1981. "Operationalizing Evangelicalism: A Review, Critique & Proposal." *Sociology of Religion* 42 (4):363–72. https://doi.org/10.2307/3711547.

Jacquet, Constant H. 1991. *Yearbook of American and Canadian Churches*. Abingdon Press.

Kellstedt, Lyman A., John C. Green, James L. Guth, and Corwin E. Smidt. 1996. "Grasping the Essentials: The Social Embodiment of Religion and Political Behavior." *Religion and the Culture Wars: Dispatches from the Front*, 174–192.

Ketchen, David J., and Christopher L. Shook. 1996. "The Application of Cluster Analysis in Strategic Management Research: An Analysis and Critique." *Strategic Management Journal* 17 (6):441–58. https://doi.org/10.1002/(SICI)1097-0266(199606)17:6<441::AID-SMJ819>3.0.CO;2-G.

Kim, Kyoung-jae, and Hyunchul Ahn. 2008. "A Recommender System Using GA K-Means Clustering in an Online Shopping Market." *Expert Systems with Applications* 34 (2):1200–1209. https://doi.org/10.1016/j.eswa.2006.12.025.

Kuo, RJ, LM Ho, and Clark M Hu. 2002. "Integration of Self-Organizing Feature Map and K-Means Algorithm for Market Segmentation." *Computers & Operations Research* 29 (11):1475–93.

Li, Dan, Jitender Deogun, William Spaulding, and Bill Shuart. 2004. "Towards Missing Data Imputation: A Study of Fuzzy k-Means Clustering Method." In *Rough Sets and Current Trends in Computing*, 3066:573–79.

MacQueen, James, and others. 1967. "Some Methods for Classification and Analysis of Multivariate Observations." In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1:281–97.

Melton, J Gordon, and James V Geisendorfer. 1977. *A Directory of Religious Bodies in the United States*. Vol. 91. Scholarly Title.

Smith, T.W. 1990. "Classifying Protestant Denominations." *Review of Religious Research*, 225–245.

Steensland, B., J.Z. Park, M.D. Regnerus, and L.D. Robinson. 2000. "Measure of American Religion: Toward Improving the State of the Art, The." *Social Forces* 79:291.

Stetzer, Ed, and Ryan P. Burge. 2016. "Reltrad Coding Problems and a New Repository." *Politics and Religion* 9 (1):187–90. https://doi.org/10.1017/S1755048315000929.

Stouffer, S.A. 1955. "Communism, Conformity, and Civil Liberties: A Cross-Section of the Nation Speaks Its Mind." *Garden City, NY: Doubleday, 1955*.

Sugar, Catherine A, and Gareth M James. 2003. "Finding the Number of Clusters in a Dataset: An Information-Theoretic Approach." *Journal of the American Statistical Association* 98 (463):750–63.

Thorndike, Robert L. 1953. "Who Belongs in the Family?" *Psychometrika* 18 (4):267–76. https://doi.org/10.1007/BF02289263.

Trevino, Andrea. 2016. "Introduction to K-Means Clustering." December 6, 2016. https://www.datascience.com/blog/k-means-clustering.

Visalakshi, N Karthikeyani, and K Thangavel. 2009. "Impact of Normalization in Distributed K-Means Clustering." *International Journal of Soft Computing* 4 (4):168–72.

Weber, M. 2002. *The Protestant Ethic and the "Spirit" of Capitalism and Other Writings*. Penguin Group USA.

Yang, Jun. 2009. "Toward Physical Activity Diary: Motion Recognition Using Simple Acceleration Features with Mobile Phones." In *Proceedings of the 1st International Workshop on Interactive Multimedia for Consumer Electronics*, 1–10.

Table 1

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
Age	0.57	0.56	0.52	0.5	0.57	0.52
Male	1	0.48	0.56	0.38	0	0.42
Income	0.9	0.93	0.93	0.84	0.88	0.94
White	0.86	0.91	0.88	0	0.86	0.9
Black	0	0	0.02	0.99	0	0.01
Education	0.62	0.69	0.73	0.65	0.63	0.72
Church Attendance	0.56	0.45	0.16	0.54	0.62	0.37
Literal	0.94	0	0	0.54	0.98	0
Inspired	0	0.99	0	0.34	0	0.97
Fables	0	0	0.99	0.09	0	0
Tolerance	0.36	0.18	0.56	0.38	0.36	0.81
Pro-Gay Marriage	0.27	0.11	0.55	0.36	0.31	0.79
Pro-Choice	0.11	0	0.47	0.25	0.12	0.62
Republican ID	0.57	0.53	0.37	0.21	0.54	0.45
N	1046	1967	2040	1520	1414	2011

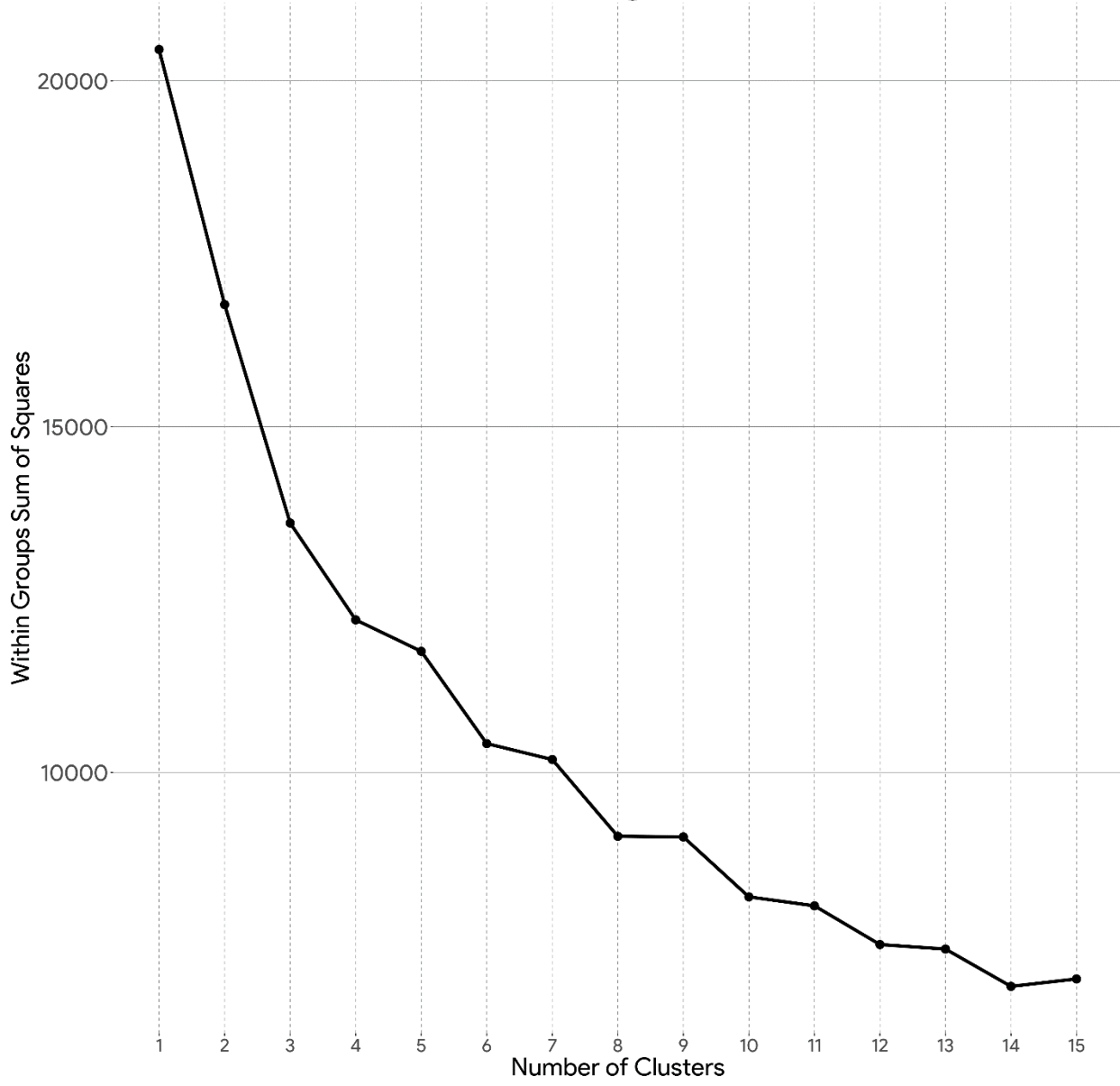Figure 1

Elbow Method for Choosing Number of Clusters
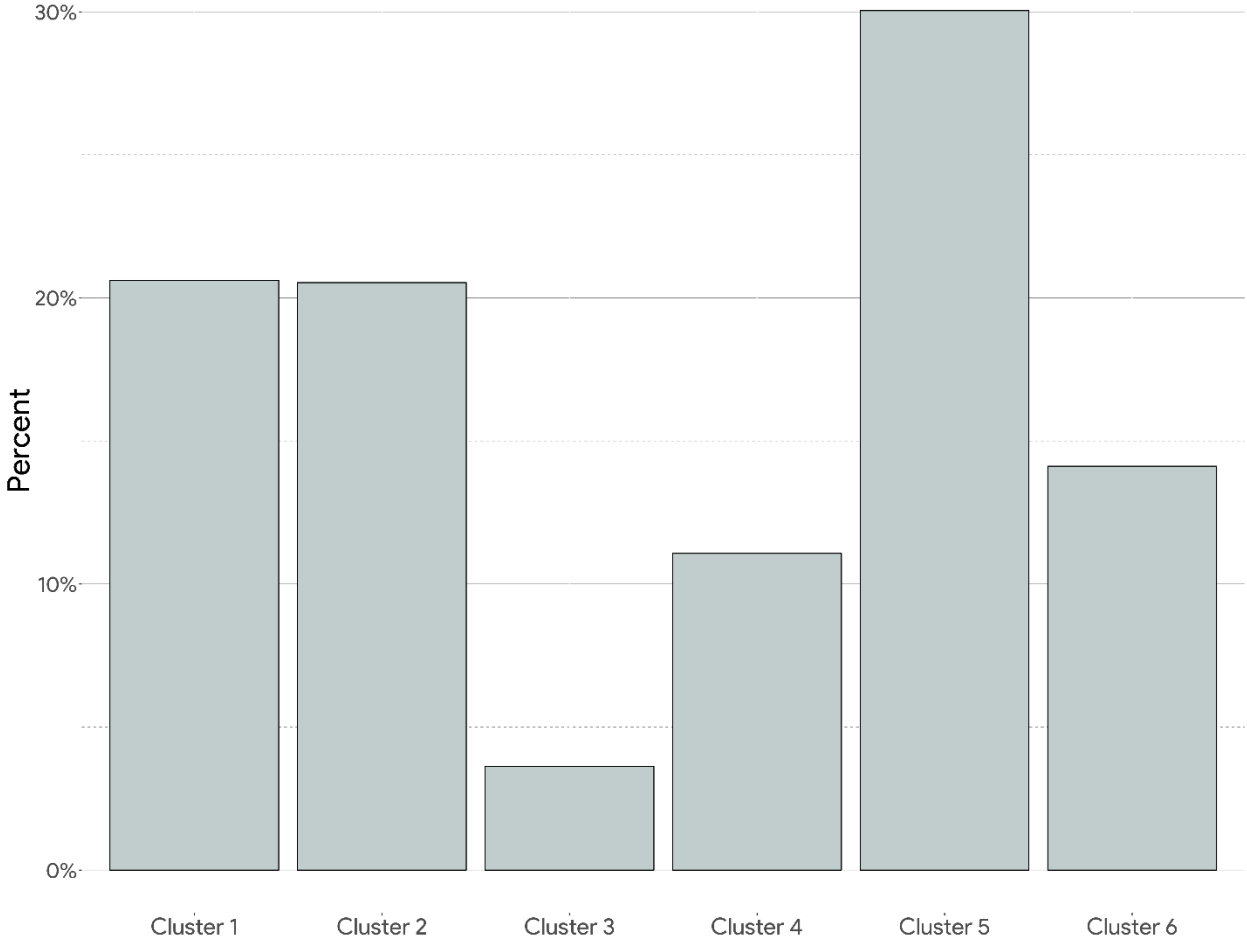
Figure 2

Evangelical Protestants

Figure 3
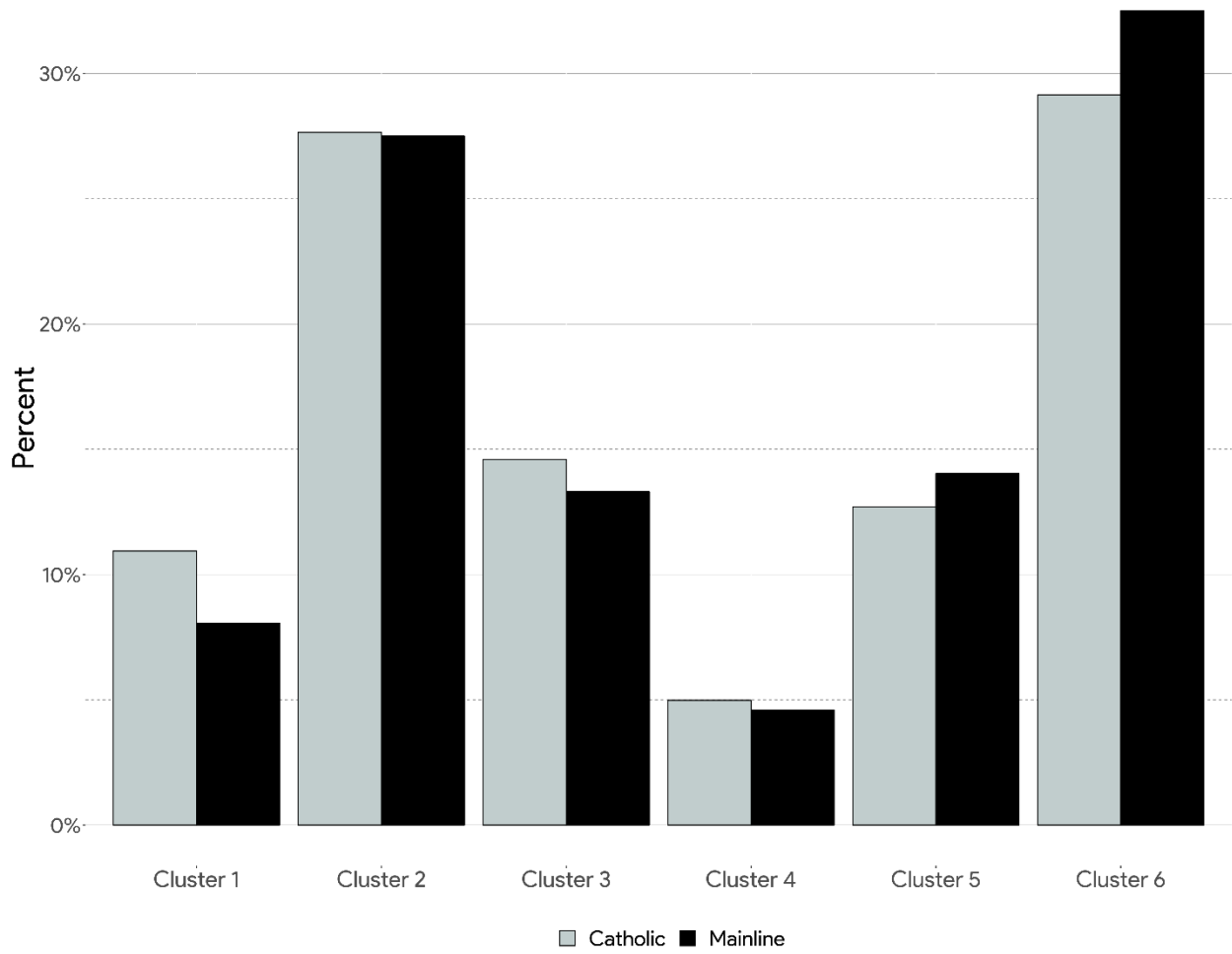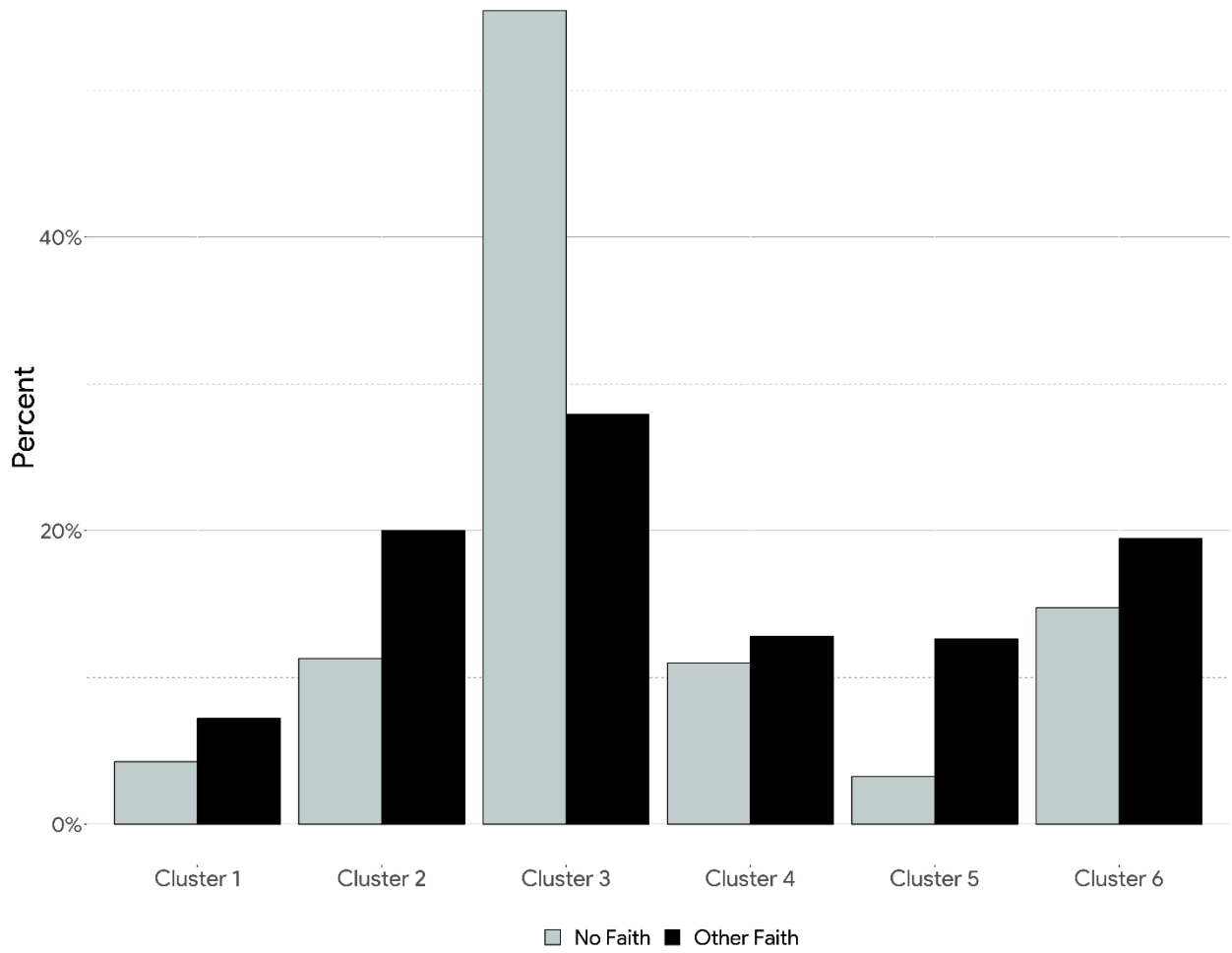
Catholics and Mainline Protestants

Figure 4

No Faith and Other Faith

Appendix

Variable Coding for the Regression Analysis:

Age:

2017 – birth_yr

Male:

"Are you male or female?"

1 = Male; Female =0; else = NA

Income:

"In which of these groups did your total family income, from all sources, fall last year before taxes, that is?"

1= Less than 1k; 2 = 1k -2k; 3 = 2k-3k; 4 = 3k-4k; 5= 4k – 5k; 6 = 5k – 6k; 7 = 7k-8k; 8=8k-10k; 9 = 10k – 15k; 11 = 20k – 25k; 12 = 25k+; else = NA

White:

"What racial or ethnic group best describes you?"

1= White; else = 0

Black:

"What racial or ethnic group best describes you?"

1= Black; else = 0

Education:

"What is the highest grade in elementary school or high school that finished and got credit for"

1-20 were coded as they appeared in the codebook; else = NA

Church Attendance:

"Aside from weddings and funerals, how often do you attend religious services?

0 = Never; 1 = Less than once a year; 2 = Once a year; 3 = Several times a year ; 4 = Once a month; 5 = 2-3x a month; 6 = Nearly every week; 7 = Every week; 8 = More than once a week ; else = NA

View of the Bible:

"Which of these statements comes closest to describing your feelings about the Bible?"

Word of God: Literalist =1; else =0
Inspired Word of God: Inspired =1; else =0
Book of Fables: Fables =1; else =0

<u>Tolerance:</u>

The GSS asks respondents about five groups (atheists, homosexuals, racists, militarists, and communists) in three scenarios (place a book in the local library, teach at a local school, give a public speech). These fifteen items were summed whereby higher values indicates a greater level of tolerance.

<u>Pro-Gay Marriage:</u>

"Homosexual couples should have the right to marry one another."

1 = Strongly Disagree; 2 = Disagree; 3 = Neither agree or disagree; 4 = Agree; 5 = Strongly Agree; else = NA

<u>Pro-Choice:</u>

"Please tell me whether or not you think it should be possible for a pregnant woman to obtain a legal abortion if…The woman wants it for any reason?"

Yes =1; else =0

<u>Republican ID:</u>

"Generally speaking, do you usually think of yourself as a Republican, Democrat, Independent, or what?"

0 = Strong Democrat, 1 = Not Strong Democrat, 2 = Independent, Near Democrat, 3 = Independent, 4= Independent, Near Republican, 5 = Not Strong Republican, 6 = Strong Republican, else = NA

Each variable was normalized so that the maximum value was 1 and the minimum value was zero as a means to aid interpretation.

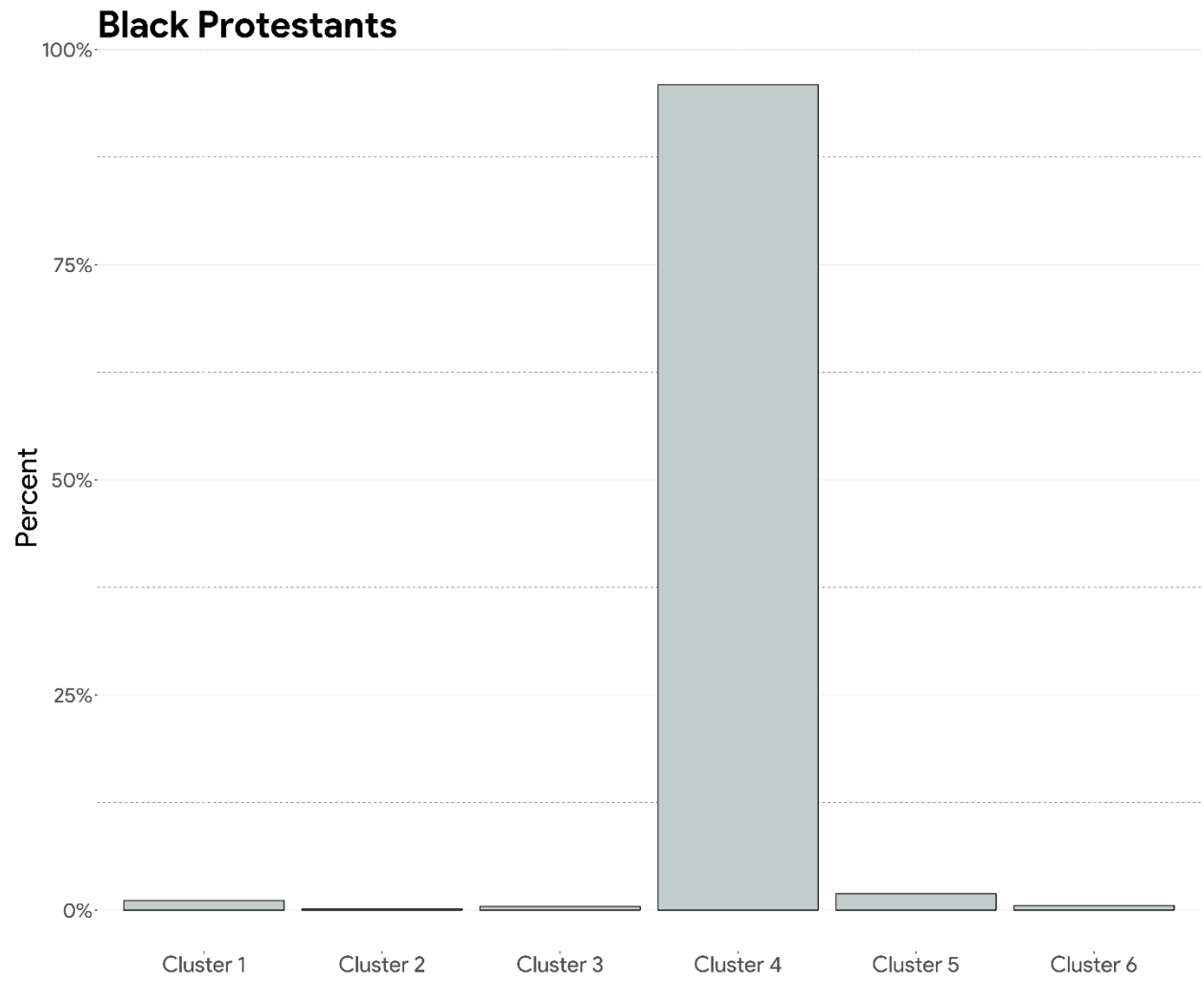**Black Protestants**

Figure A2

Jewish