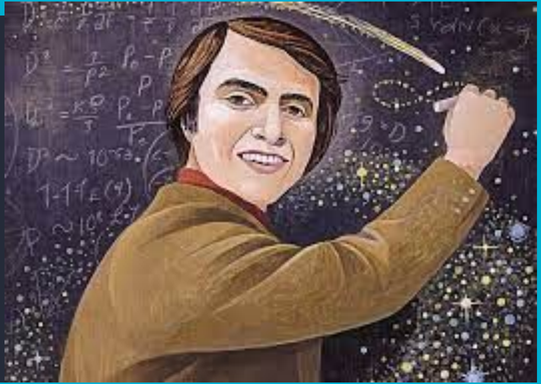
Segmenting Audiences for Campaign Communications



Cluster Analysis of Americans’ Views on Science

PROBLEM STATEMENT:

A **public health organization** is planning a public education **campaign** to promote scientific understanding and support for **scientific and technological research** and advancement in the US.

The first step will be to gain insight into different **audience segments** to understand how attitudes toward a range of topics in science vary, to inform **messaging strategy**.

GOVERNANCE
ADVOCACY
INFORMATION
CAPACITY
PROTECTION
PREVENTION
PROMOTION

Research Questions

- How do audiences **cluster** based on their beliefs about science?

- How do remaining attitudinal and self-reported behavioral variables coalesce to represent **latent constructs**?

- How well do these constructs, combined with demographic features, **predict audience segment**?

Methods

- **Clustering**: The goal of kmeans cluster analysis is to find optimal groupings for the observations in a data set. Ideally the observations grouped into a cluster are highly similar to one another, and each cluster is unique from the other clusters.

- **Decomposition**: The goal of factor analysis is to identify overlap in, or correlation between variables and create a new set of factors that represent the same latent traits as the original variables, but using fewer dimensions and reducing redundancy.

- **Classification**: The primary goal of classification will be to gain insight into the key markets and the audience segments (clusters), including mediating and moderating features (interactions). Since interpretation is the priority, the models for comparison will be logistic regression with regularization using ridge, lasso and elasticnet.

The Data

Pew Research Center

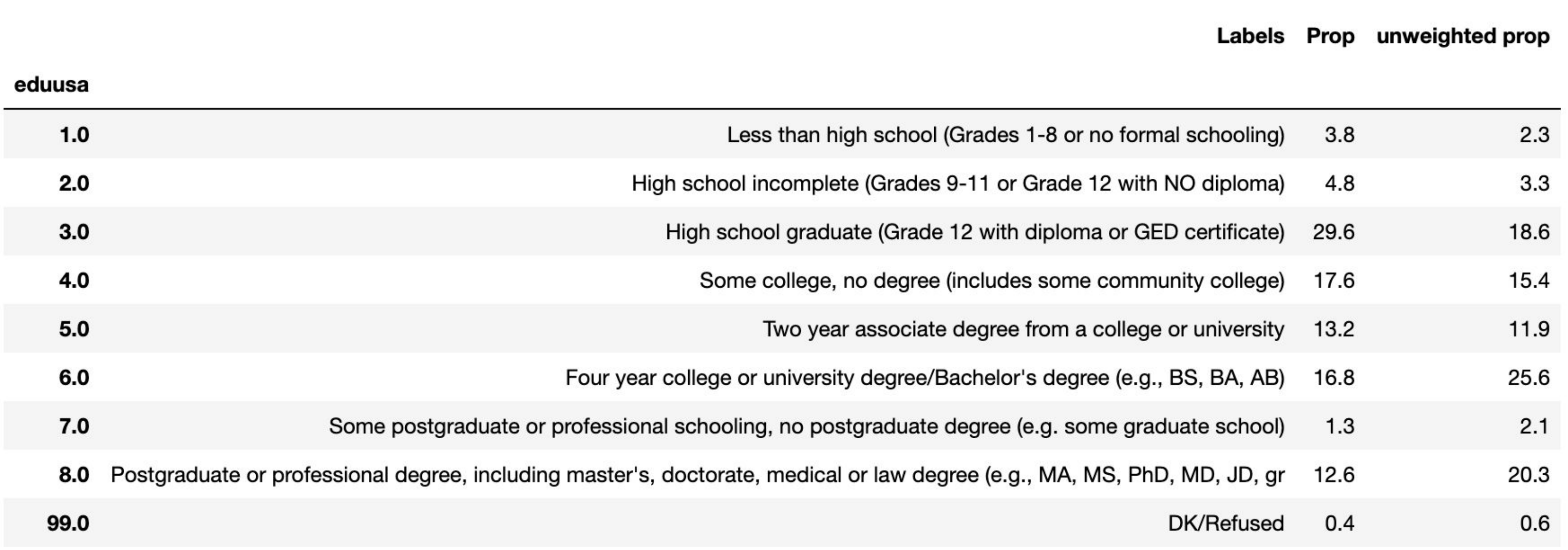
Pew Research Center conducted an International Science Survey in 20 different countries between October 1, 2019 to March 15, 2020.

The dataset contains 244 features (including a survey weight column), and 32,330 observations. The survey sample is intended to be representative of the adult population in each of the countries, and so the survey weight variable must be used when analyzing responses and making inferences about the general population.

62 variables comprise the attitudinal and self-reported behavioral measures that will be used for the unsupervised learning models. These features are categorical, with a range of 2-4 answer options. I have recoded these features to be ordinal, standardized between -1 and 1, where 1 is the 'affirmative' response and -1 is the 'opposing' response. Missing values, coded as 'don't know' or 'refused', were recoded to the mid-point value of 0.

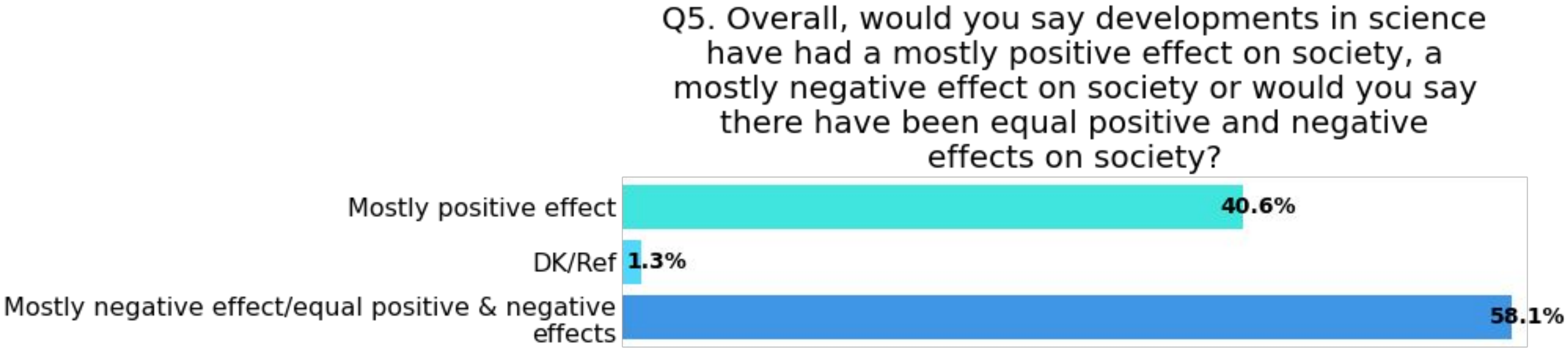
The remaining variables are demographic. These will not be used in the unsupervised modeling phase, but will be incorporated into the supervised learning modeling step.

The Data - Sample Weights

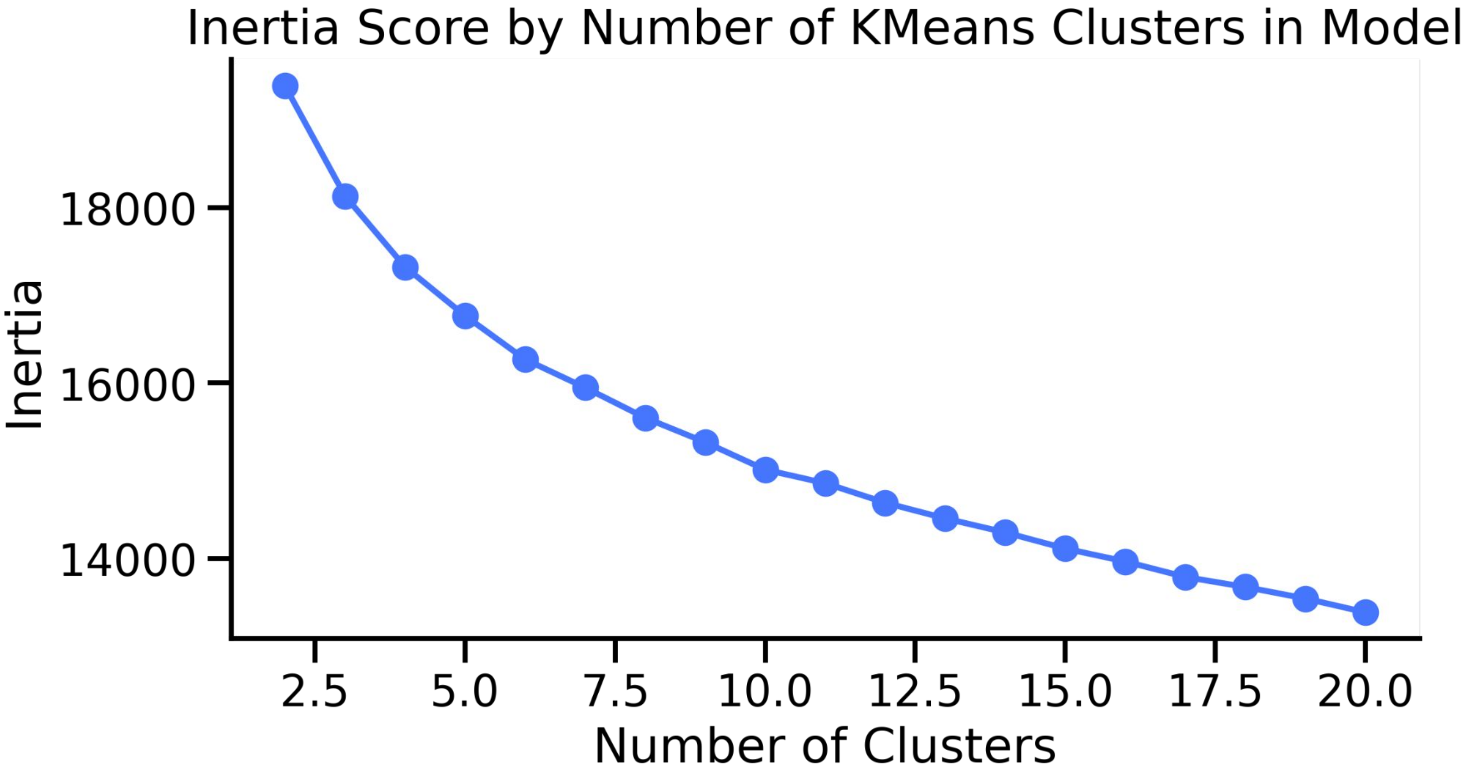


eduusa	Labels	Prop	unweighted prop
1.0	Less than high school (Grades 1-8 or no formal schooling)	3.8	2.3
2.0	High school incomplete (Grades 9-11 or Grade 12 with NO diploma)	4.8	3.3
3.0	High school graduate (Grade 12 with diploma or GED certificate)	29.6	18.6
4.0	Some college, no degree (includes some community college)	17.6	15.4
5.0	Two year associate degree from a college or university	13.2	11.9
6.0	Four year college or university degree/Bachelor's degree (e.g., BS, BA, AB)	16.8	25.6
7.0	Some postgraduate or professional schooling, no postgraduate degree (e.g. some graduate school)	1.3	2.1
8.0	Postgraduate or professional degree, including master's, doctorate, medical or law degree (e.g., MA, MS, PhD, MD, JD, gr	12.6	20.3
99.0	DK/Refused	0.4	0.6

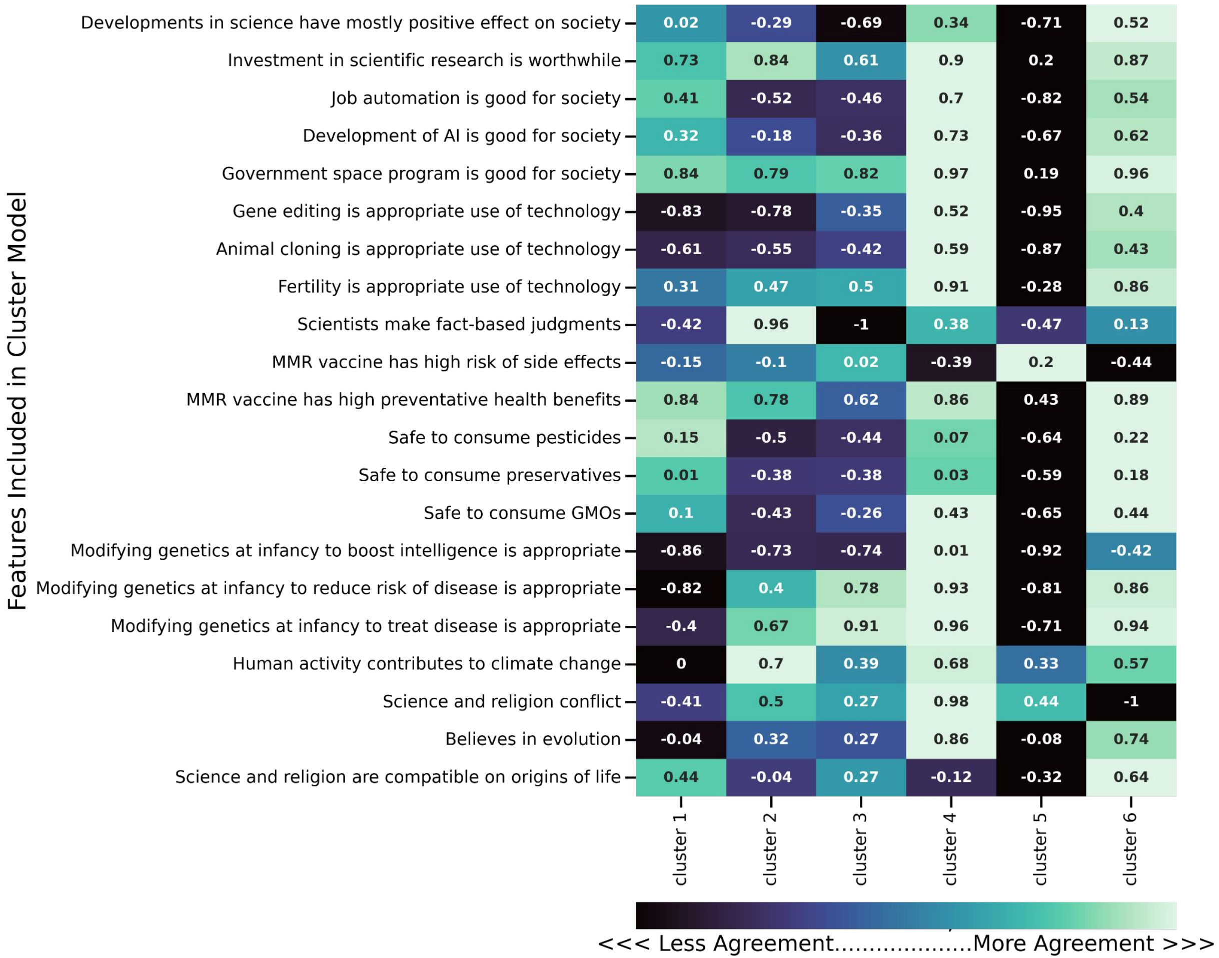
Cluster Analysis - Inputs



Cluster Analysis - Inertia

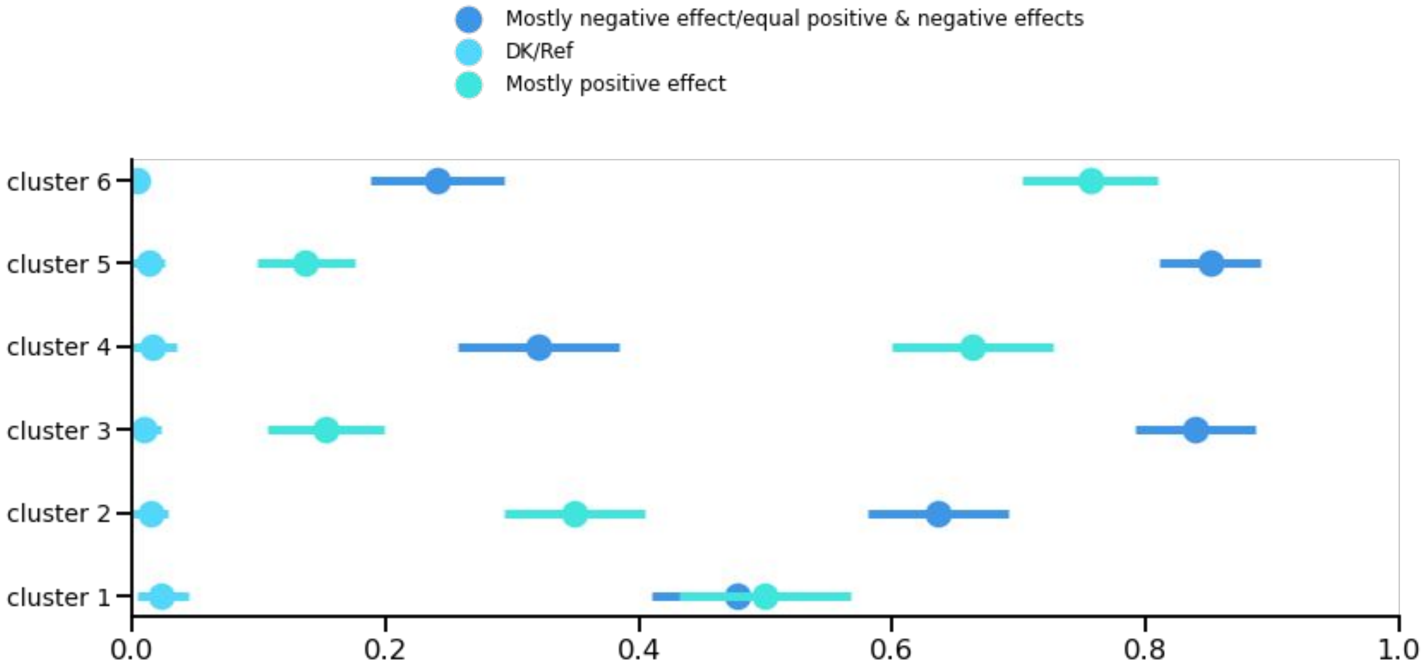


Cluster Averages for Features in the Cluster Model
clusters: k-means, 6 clusters



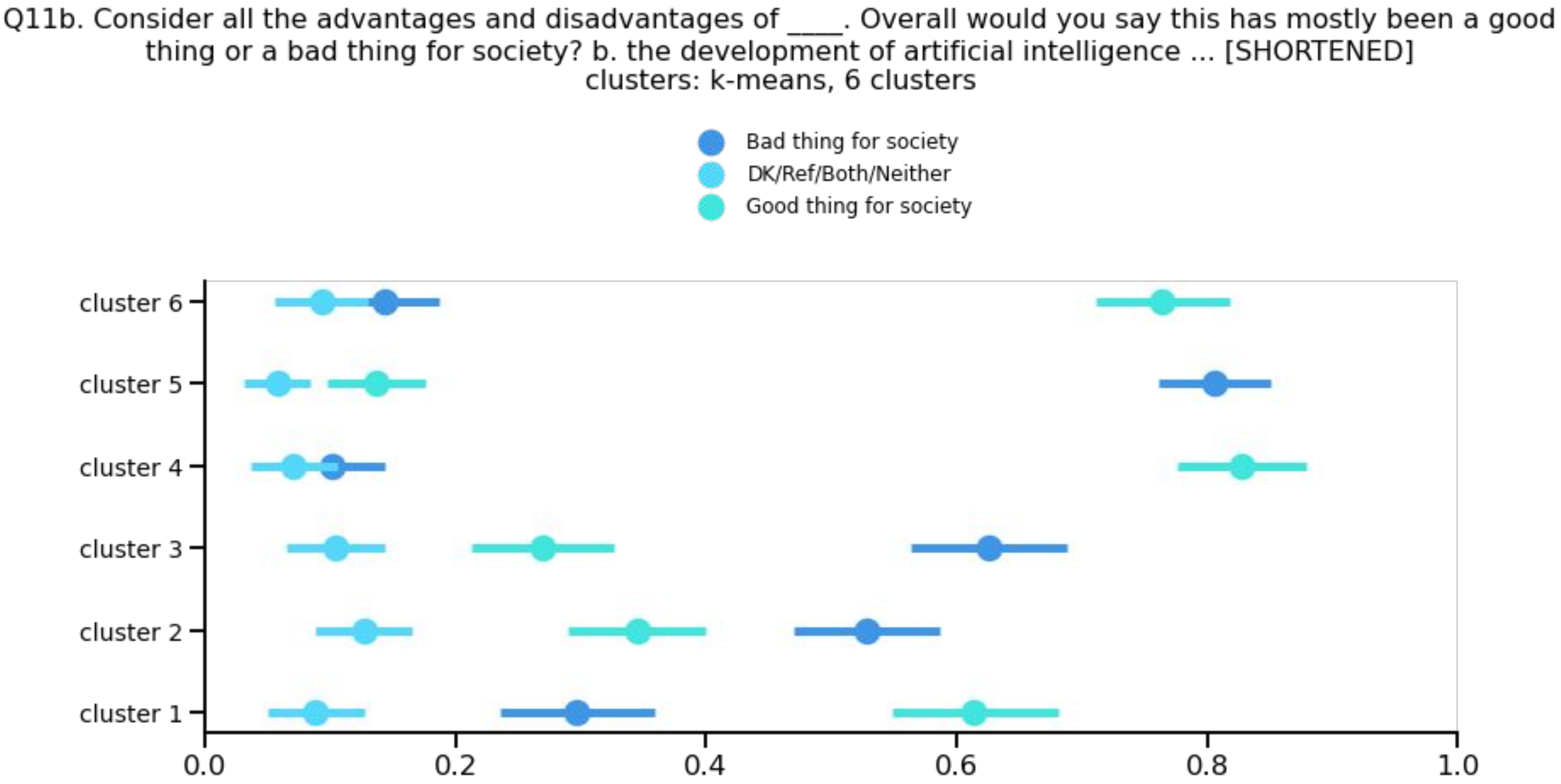
	cluster 1	cluster 2	cluster 3	cluster 4	cluster 5	cluster 6
Developments in science have mostly positive effect on society	0.02	-0.29	-0.69	0.34	-0.71	0.52
Investment in scientific research is worthwhile	0.73	0.84	0.61	0.9	0.2	0.87
Job automation is good for society	0.41	-0.52	-0.46	0.7	-0.82	0.54
Development of AI is good for society	0.32	-0.18	-0.36	0.73	-0.67	0.62
Government space program is good for society	0.84	0.79	0.82	0.97	0.19	0.96
Gene editing is appropriate use of technology	-0.83	-0.78	-0.35	0.52	-0.95	0.4
Animal cloning is appropriate use of technology	-0.61	-0.55	-0.42	0.59	-0.87	0.43
Fertility is appropriate use of technology	0.31	0.47	0.5	0.91	-0.28	0.86
Scientists make fact-based judgments	-0.42	0.96	-1	0.38	-0.47	0.13
MMR vaccine has high risk of side effects	-0.15	-0.1	0.02	-0.39	0.2	-0.44
MMR vaccine has high preventative health benefits	0.84	0.78	0.62	0.86	0.43	0.89
Safe to consume pesticides	0.15	-0.5	-0.44	0.07	-0.64	0.22
Safe to consume preservatives	0.01	-0.38	-0.38	0.03	-0.59	0.18
Safe to consume GMOs	0.1	-0.43	-0.26	0.43	-0.65	0.44
Modifying genetics at infancy to boost intelligence is appropriate	-0.86	-0.73	-0.74	0.01	-0.92	-0.42
Modifying genetics at infancy to reduce risk of disease is appropriate	-0.82	0.4	0.78	0.93	-0.81	0.86
Modifying genetics at infancy to treat disease is appropriate	-0.4	0.67	0.91	0.96	-0.71	0.94
Human activity contributes to climate change	0	0.7	0.39	0.68	0.33	0.57
Science and religion conflict	-0.41	0.5	0.27	0.98	0.44	-1
Believes in evolution	-0.04	0.32	0.27	0.86	-0.08	0.74
Science and religion are compatible on origins of life	0.44	-0.04	0.27	-0.12	-0.32	0.64

Cluster Analysis - Inputs by Cluster

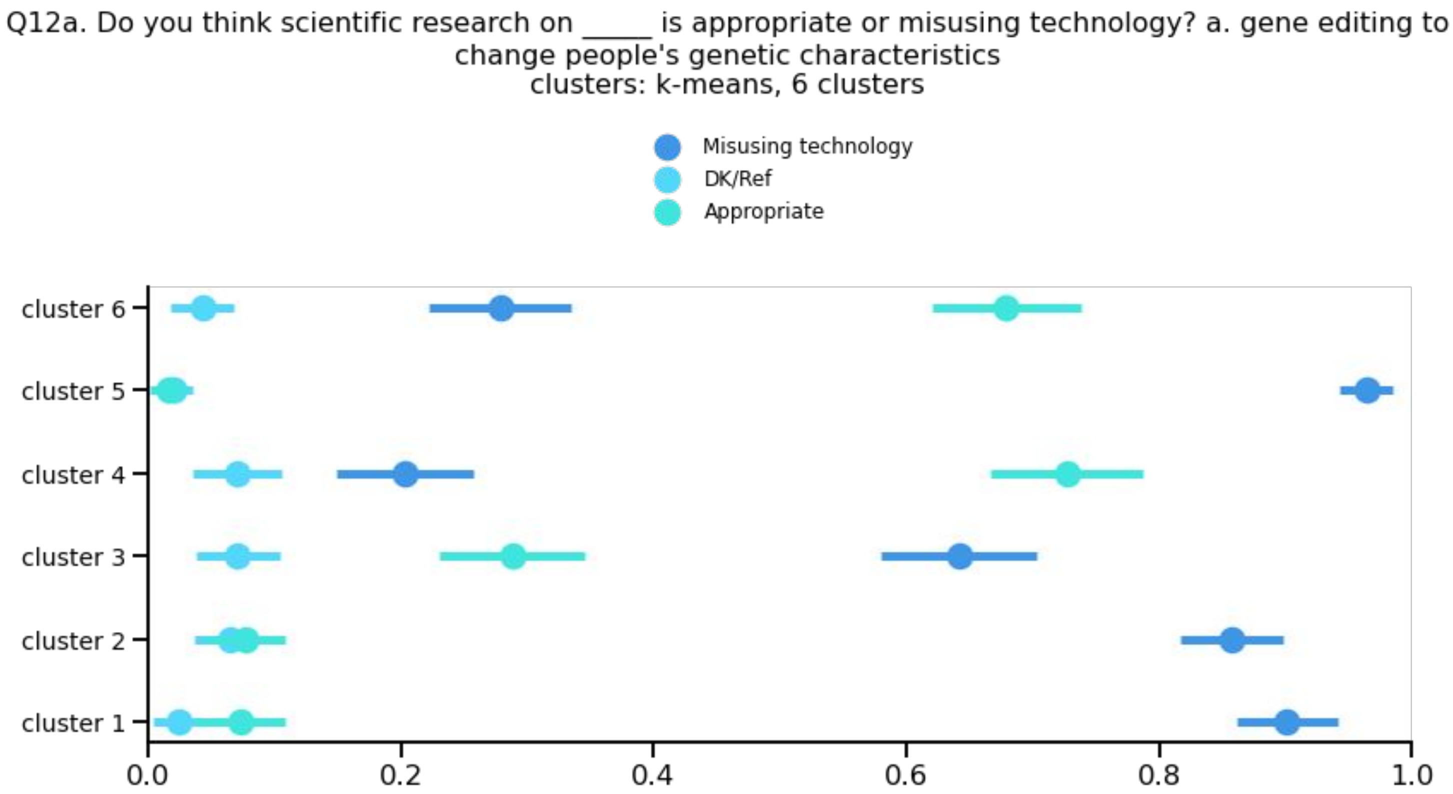


Q5. Overall, would you say developments in science have had a mostly positive effect on society, a mostly negative effect on society or would you say there have been equal positive and negative effects on society?
clusters: k-means, 6 clusters

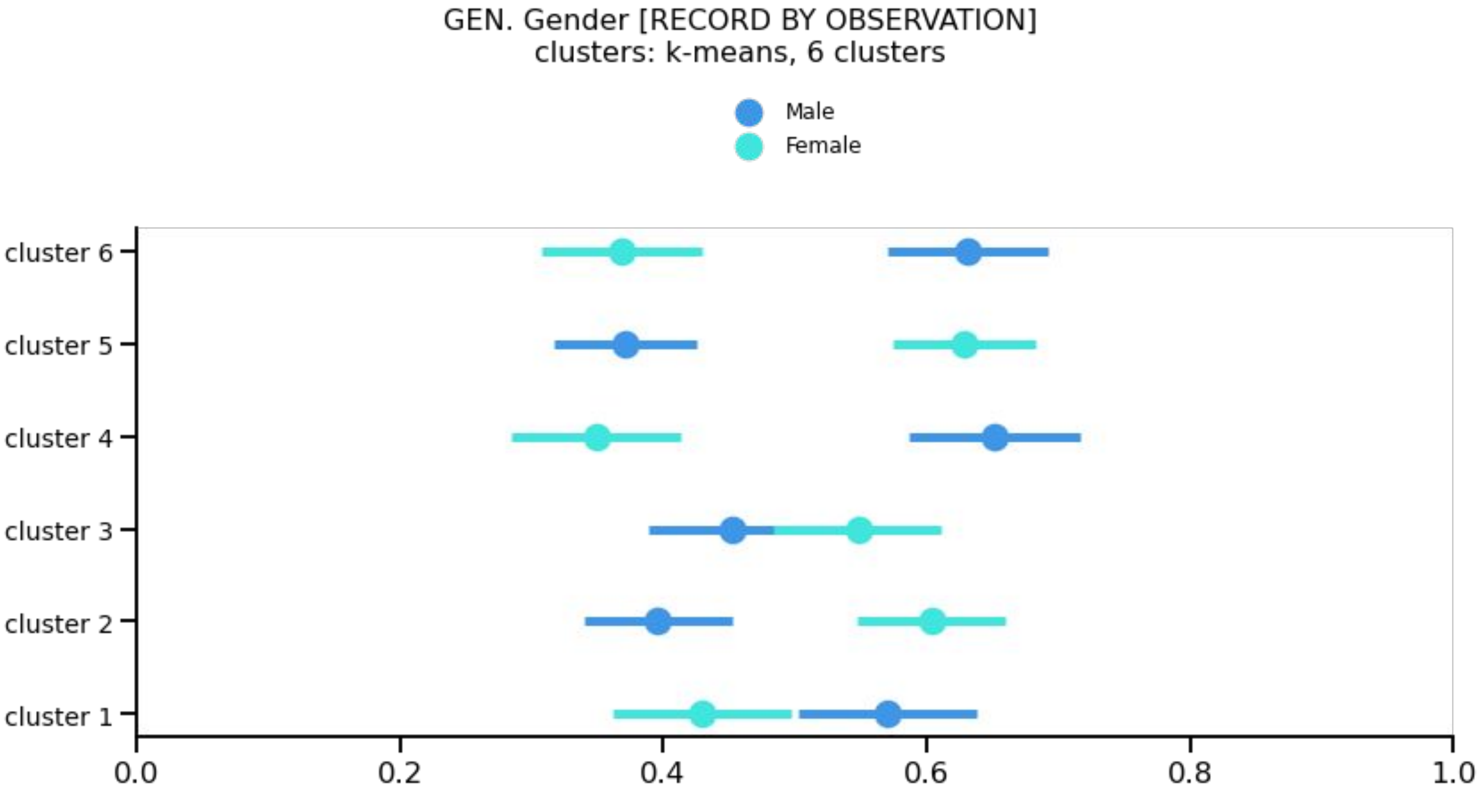
Cluster Analysis - Inputs by Cluster



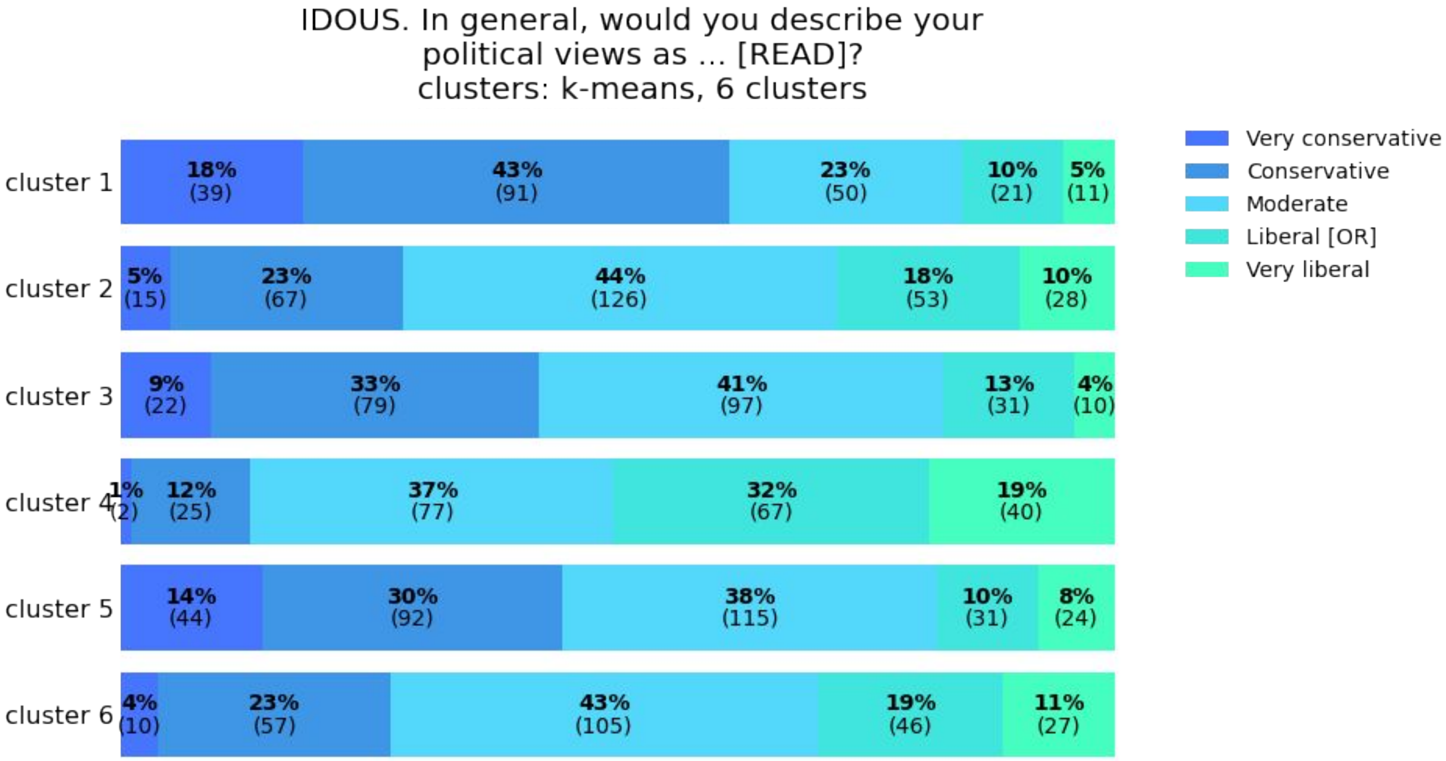
Cluster Analysis - Inputs by Cluster



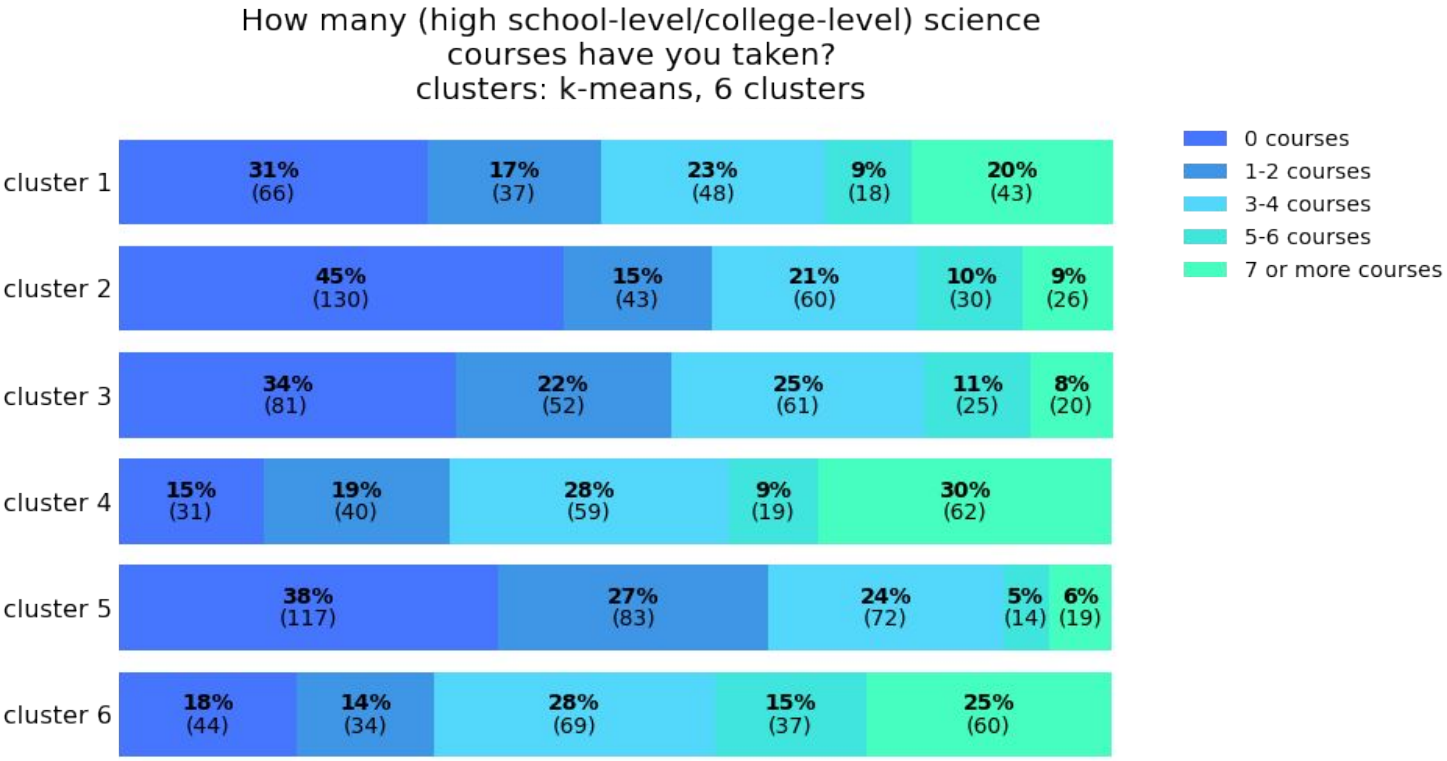
Cluster Analysis - External Features by Cluster



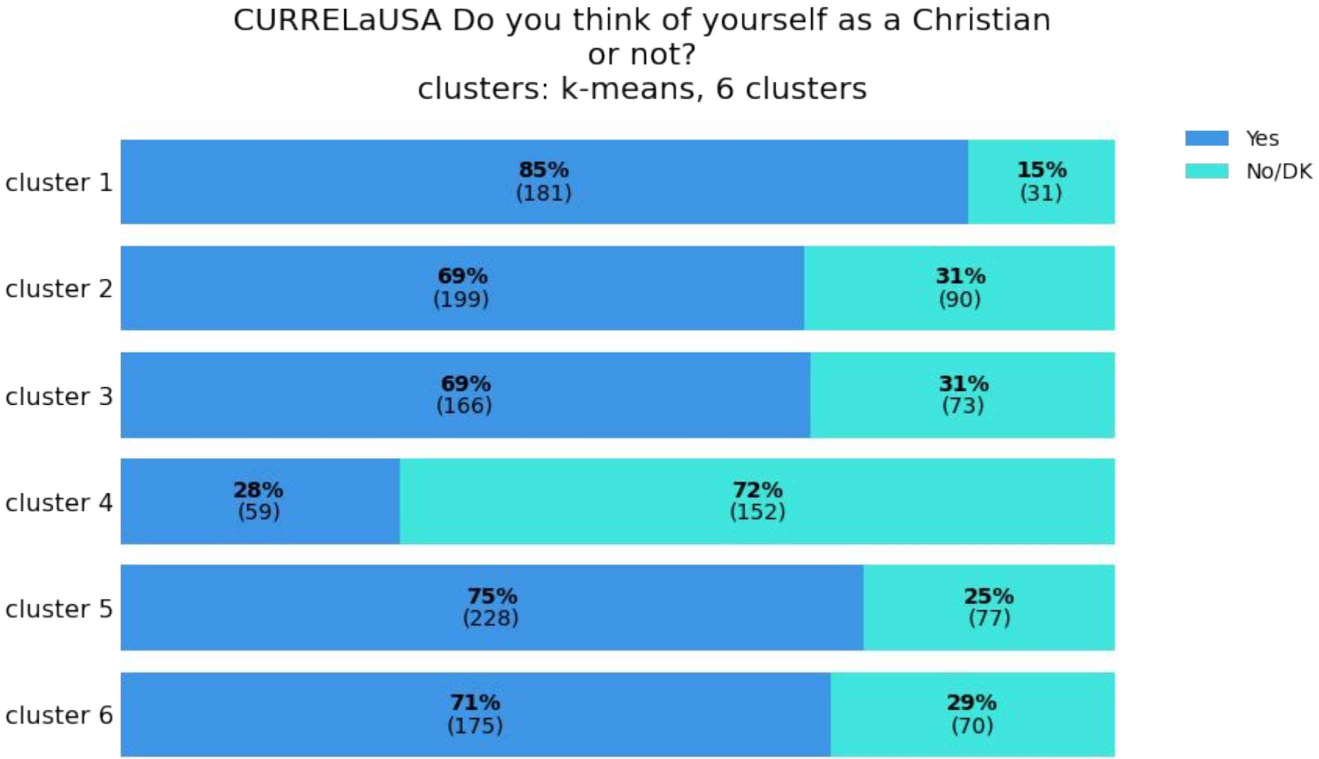
Cluster Analysis - External Features by Cluster



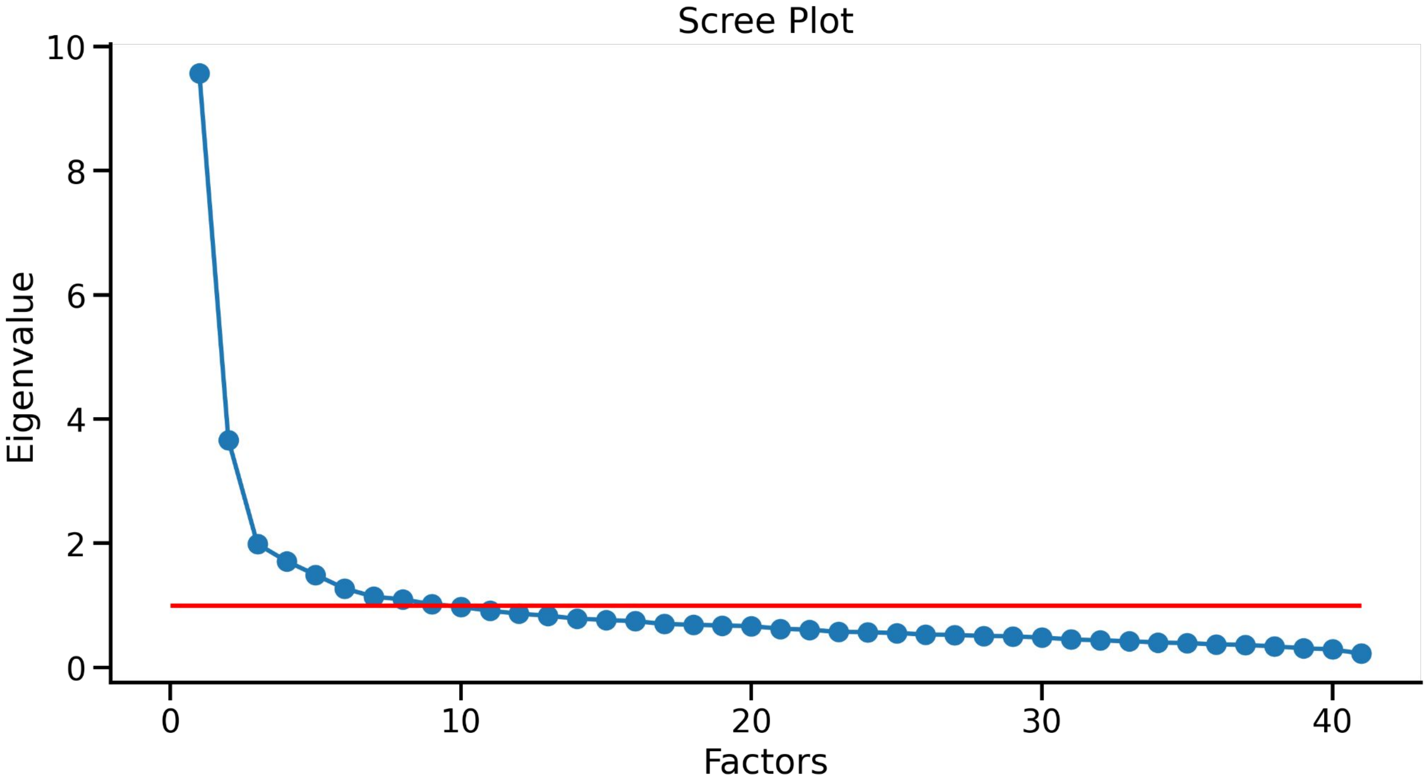
Cluster Analysis - External Features by Cluster



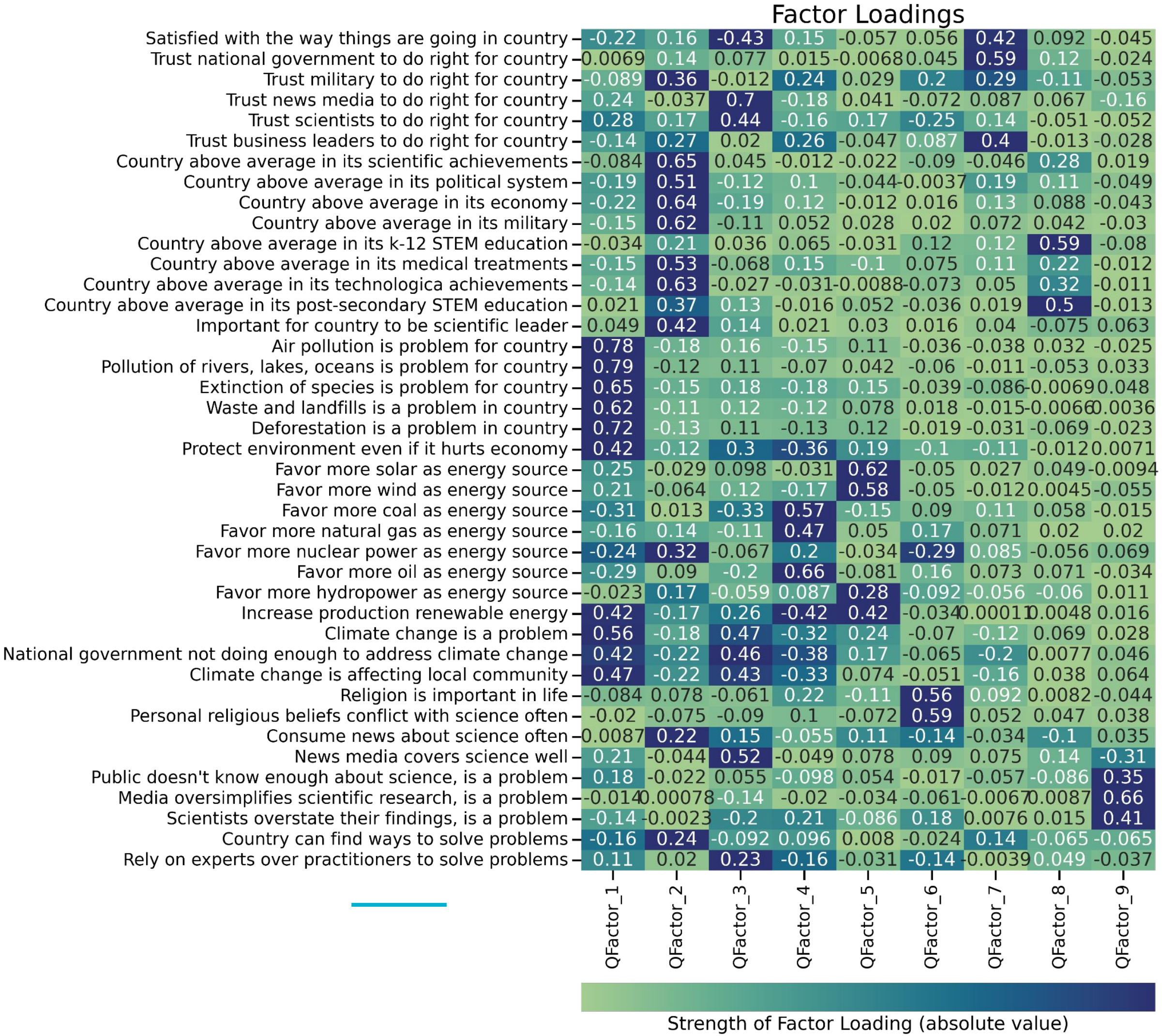
Cluster Analysis - External Features by Cluster



Factor Analysis - Scree Plot

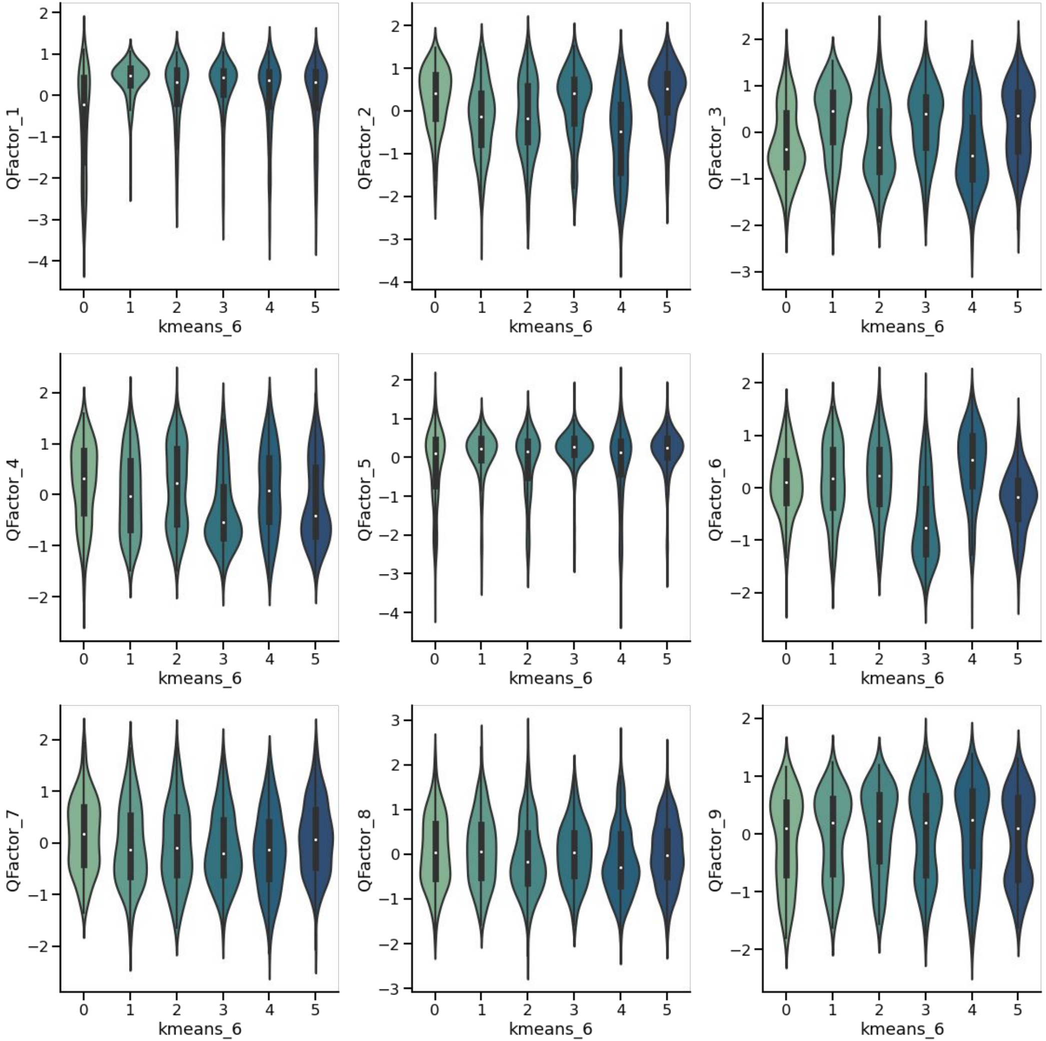


Factor Analysis Heatmap

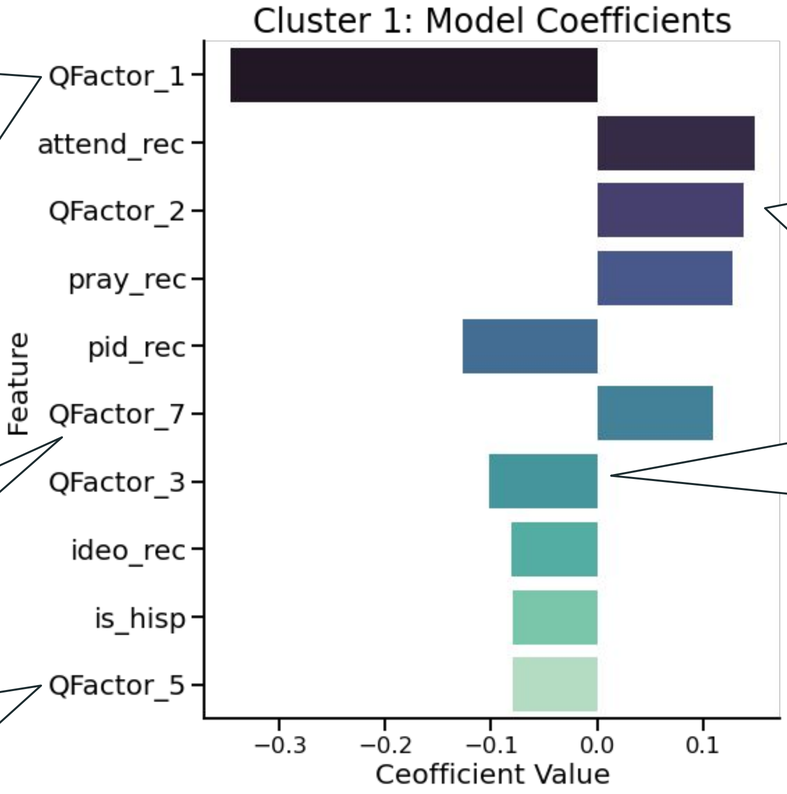


Factor Loadings

	QFactor_1	QFactor_2	QFactor_3	QFactor_4	QFactor_5	QFactor_6	QFactor_7	QFactor_8	QFactor_9
Satisfied with the way things are going in country	-0.22	0.16	-0.43	0.15	-0.057	0.056	0.42	0.092	-0.045
Trust national government to do right for country	0.0069	0.14	0.077	0.015	-0.0068	0.045	0.59	0.12	-0.024
Trust military to do right for country	-0.089	0.36	-0.012	0.24	0.029	0.2	0.29	-0.11	-0.053
Trust news media to do right for country	0.24	-0.037	0.7	-0.18	0.041	-0.072	0.087	0.067	-0.16
Trust scientists to do right for country	0.28	0.17	0.44	-0.16	0.17	-0.25	0.14	-0.051	-0.052
Trust business leaders to do right for country	-0.14	0.27	0.02	0.26	-0.047	0.087	0.4	-0.013	-0.028
Country above average in its scientific achievements	-0.084	0.65	0.045	-0.012	-0.022	-0.09	-0.046	0.28	0.019
Country above average in its political system	-0.19	0.51	-0.12	0.1	-0.044	-0.0037	0.19	0.11	-0.049
Country above average in its economy	-0.22	0.64	-0.19	0.12	-0.012	0.016	0.13	0.088	-0.043
Country above average in its military	-0.15	0.62	-0.11	0.052	0.028	0.02	0.072	0.042	-0.03
Country above average in its k-12 STEM education	-0.034	0.21	0.036	0.065	-0.031	0.12	0.12	0.59	-0.08
Country above average in its medical treatments	-0.15	0.53	-0.068	0.15	-0.1	0.075	0.11	0.22	-0.012
Country above average in its technologica achievements	-0.14	0.63	-0.027	-0.031	-0.0088	-0.073	0.05	0.32	-0.011
Country above average in its post-secondary STEM education	0.021	0.37	0.13	-0.016	0.052	-0.036	0.019	0.5	-0.013
Important for country to be scientific leader	0.049	0.42	0.14	0.021	0.03	0.016	0.04	-0.075	0.063
Air pollution is problem for country	0.78	-0.18	0.16	-0.15	0.11	-0.036	-0.038	0.032	-0.025
Pollution of rivers, lakes, oceans is problem for country	0.79	-0.12	0.11	-0.07	0.042	-0.06	-0.011	-0.053	0.033
Extinction of species is problem for country	0.65	-0.15	0.18	-0.18	0.15	-0.039	-0.086	-0.0069	0.048
Waste and landfills is a problem in country	0.62	-0.11	0.12	-0.12	0.078	0.018	-0.015	-0.0066	0.0036
Deforestation is a problem in country	0.72	-0.13	0.11	-0.13	0.12	-0.019	-0.031	-0.069	-0.023
Protect environment even if it hurts economy	0.42	-0.12	0.3	-0.36	0.19	-0.1	-0.11	-0.012	0.0071
Favor more solar as energy source	0.25	-0.029	0.098	-0.031	0.62	-0.05	0.027	0.049	-0.0094
Favor more wind as energy source	0.21	-0.064	0.12	-0.17	0.58	-0.05	-0.012	0.0045	-0.055
Favor more coal as energy source	-0.31	0.013	-0.33	0.57	-0.15	0.09	0.11	0.058	-0.015
Favor more natural gas as energy source	-0.16	0.14	-0.11	0.47	0.05	0.17	0.071	0.02	0.02
Favor more nuclear power as energy source	-0.24	0.32	-0.067	0.2	-0.034	-0.29	0.085	-0.056	0.069
Favor more oil as energy source	-0.29	0.09	-0.2	0.66	-0.081	0.16	0.073	0.071	-0.034
Favor more hydropower as energy source	-0.023	0.17	-0.059	0.087	0.28	-0.092	-0.056	-0.06	0.011
Increase production renewable energy	0.42	-0.17	0.26	-0.42	0.42	-0.034	0.00011	0.0048	0.016
Climate change is a problem	0.56	-0.18	0.47	-0.32	0.24	-0.07	-0.12	0.069	0.028
National government not doing enough to address climate change	0.42	-0.22	0.46	-0.38	0.17	-0.065	-0.2	0.0077	0.046
Climate change is affecting local community	0.47	-0.22	0.43	-0.33	0.074	-0.051	-0.16	0.038	0.064
Religion is important in life	-0.084	0.078	-0.061	0.22	-0.11	0.56	0.092	0.0082	-0.044
Personal religious beliefs conflict with science often	-0.02	-0.075	-0.09	0.1	-0.072	0.59	0.052	0.047	0.038
Consume news about science often	0.0087	0.22	0.15	-0.055	0.11	-0.14	-0.034	-0.1	0.035
News media covers science well	0.21	-0.044	0.52	-0.049	0.078	0.09	0.075	0.14	-0.31
Public doesn't know enough about science, is a problem	0.18	-0.022	0.055	-0.098	0.054	-0.017	-0.057	-0.086	0.35
Media oversimplifies scientific research, is a problem	-0.014	0.00078	-0.14	-0.02	-0.034	-0.061	-0.0067	0.0087	0.66
Scientists overstate their findings, is a problem	-0.14	-0.0023	-0.2	0.21	-0.086	0.18	0.0076	0.015	0.41
Country can find ways to solve problems	-0.16	0.24	-0.092	0.096	0.008	-0.024	0.14	-0.065	-0.065
Rely on experts over practitioners to solve problems	0.11	0.02	0.23	-0.16	-0.031	-0.14	-0.0039	0.049	-0.037


QFactor_1
2
1
0
−1
−2
−3
−4
0
1
2
3
4
5
kmeans_6
QFactor_2
2
1
0
−1
−2
−3
−4
0
1
2
3
4
5
kmeans_6
QFactor_3
2
1
0
−1
−2
−3
0
1
2
3
4
5
kmeans_6
QFactor_4
2
1
0
−1
−2
0
1
2
3
4
5
kmeans_6
QFactor_5
2
1
0
−1
−2
−3
−4
0
1
2
3
4
5
kmeans_6
QFactor_6
2
1
0
−1
−2
0
1
2
3
4
5
kmeans_6
QFactor_7
2
1
0
−1
−2
0
1
2
3
4
5
kmeans_6
QFactor_8
3
2
1
0
−1
−2
−3
0
1
2
3
4
5
kmeans_6
QFactor_9
2
1
0
−1
−2
0
1
2
3
4
5
kmeans_6

Multinomial Logistic Regression w/ Ridge Reg



Environmental Issues
+Air pollution is problem
+Pollution of rivers, lakes, oceans is problem
+Extinction of species is problem
+Waste and landfills is a problem
+Deforestation is a problem
+Protect environment even if it hurts economy
+Increase production renewable energy
+Climate change is a problem
+Nat’l gov not doing enough on climate change
+Climate change is affecting local community

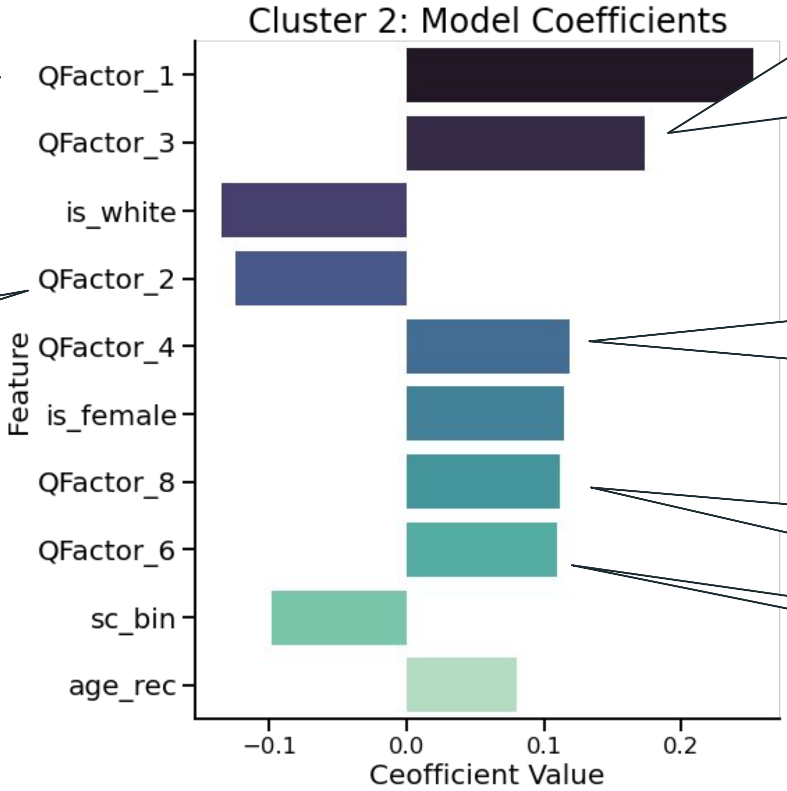
Satisfied & Trust in Gov’t, Business
+Satisfied w/ way things are going in country
+Trust nat’l government to do right for country
+Trust business leaders to do right for country

Favor Renewable Energy
+Favor more solar as energy source
+Favor more wind as energy source
+Favor more hydropower as energy source
+Increase production renewable energy

Patriotic Sentiment
+Trust military to do right for country
+Country above avg in scientific achievements
+Country above average in its political system
+Country above average in its economy
+Country above average in its military
+Country above average in medical treatments
+Country above average in tech achievements
+Important for country to be scientific leader
+Favor more nuclear power as energy source
+Consume news about science often
+Country can find ways to solve problems

Dissatisfaction but Trust in Media, Experts
–Satisfied w/ way things are going in country
+Trust news media to do right for country
+Trust scientists to do right for country
+Climate change is a problem
+Nat’l gov not doing enough on climate change
+Climate change is affecting local community
+News media covers science well
+Rely on experts over practitioners to solve problems

Multinomial Logistic Regression w/ Ridge Reg



Environmental Issues
+Air pollution is problem
+Pollution of rivers, lakes, oceans is problem
+Extinction of species is problem
+Waste and landfills is a problem
+Deforestation is a problem
+Protect environment even if it hurts economy
+Increase production renewable energy
+Climate change is a problem
+Nat'l gov not doing enough on climate change
+Climate change is affecting local community

Patriotic Sentiment
+Trust military to do right for country
+Country above avg in scientific achievements
+Country above average in its political system
+Country above average in its economy
+Country above average in its military
+Country above average in medical treatments
+Country above average in tech achievements
+Important for country to be scientific leader
+Favor more nuclear power as energy source
+Consume news about science often
+Country can find ways to solve problems

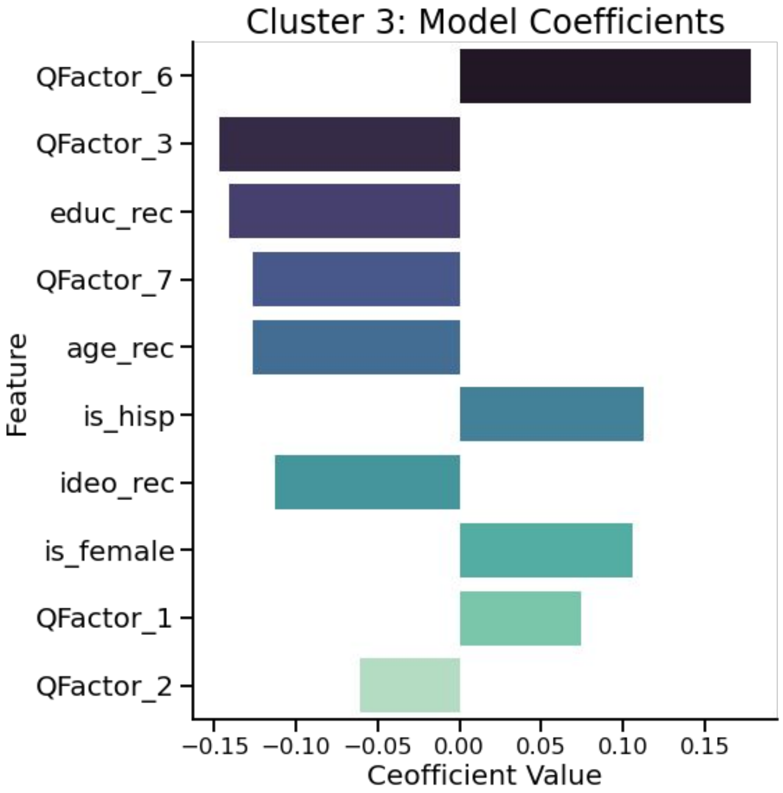
Dissatisfaction but Trust in Media, Experts
–Satisfied w/ way things are going in country
+Trust news media to do right for country
+Trust scientists to do right for country
+Climate change is a problem
+Nat'l gov not doing enough on climate change
+Climate change is affecting local community
+News media covers science well
+Rely on experts over practitioners to solve problems

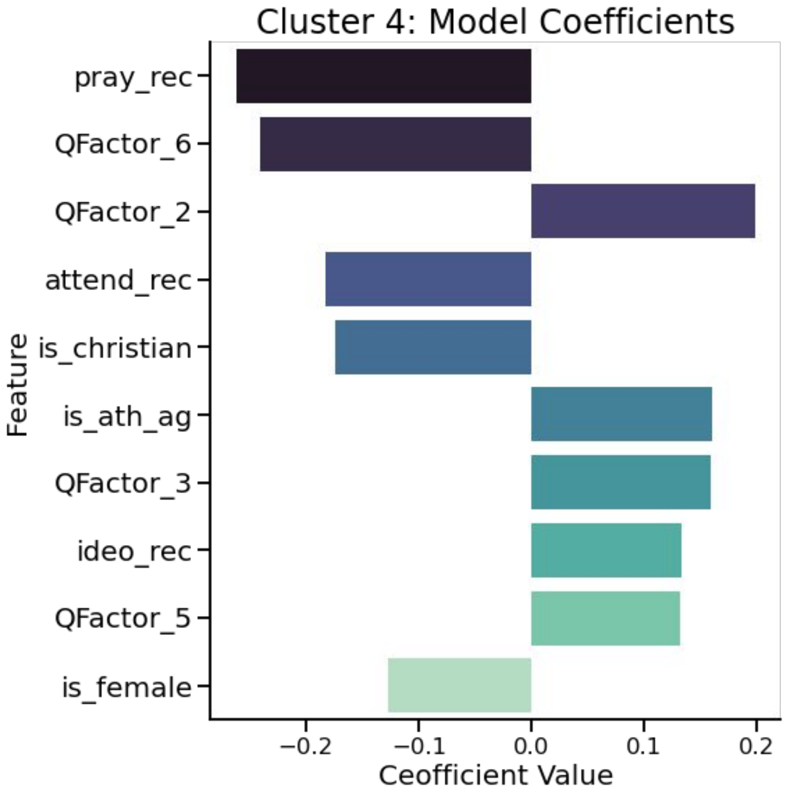
Favor Fossil Fuels not Renewables
+Favor more coal as energy source
+Favor more natural gas as energy source
+Favor more oil as energy source
–Increase production renewable energy

Positive on STEM Education in Country
+Country above avg in its k-12 STEM education
+Country above avg in post-secondary STEM education

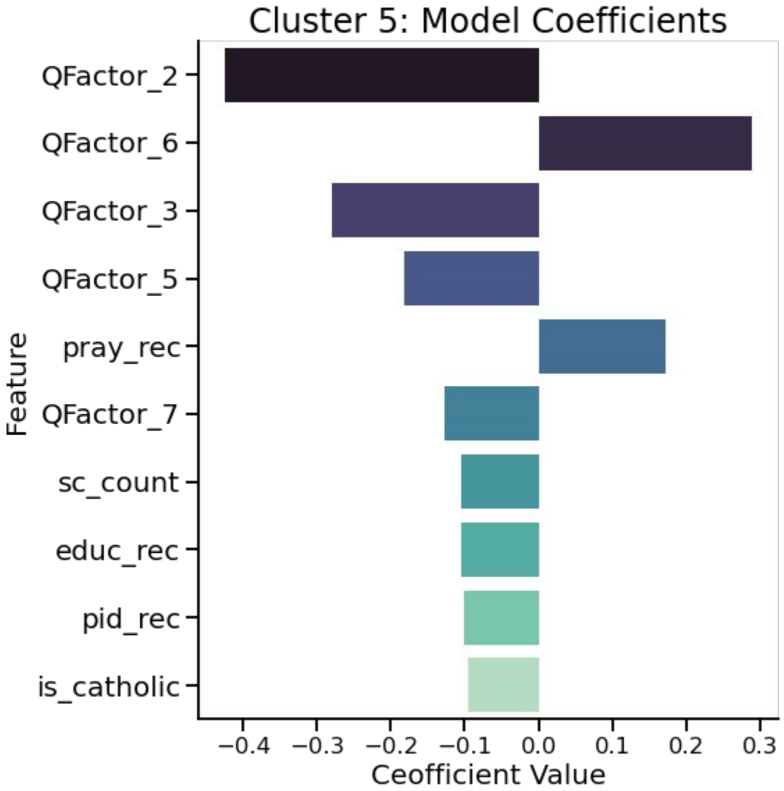
Religion is Important, Conflicts w/ Science
–Favor more nuclear power as energy source
+Religion is important in life
+Pers religious beliefs conflict w science often

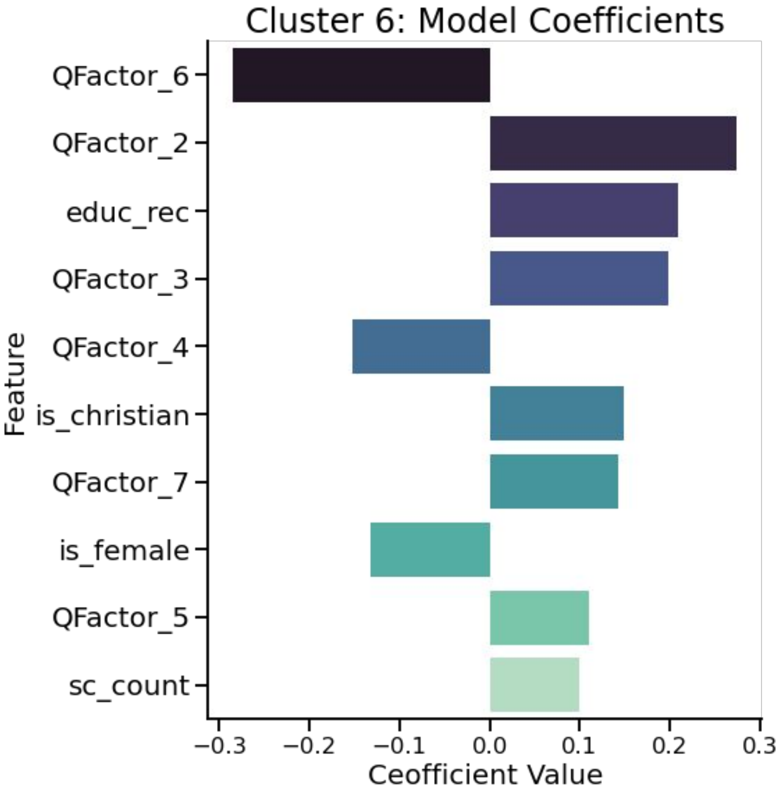
Multinomial Logistic Regression w/ Ridge Reg





Multinomial Logistic Regression w/ Ridge Reg





Recommendations/Next Steps

- **More data, more features.** Consumer and behavioral data for message targeting.

- **Qualitative research**. Focus groups could provide rich insights into subgroups.

- **Message testing.** Optimize messages to audience segment.