Reason vs. Belief: How Humans Respond to Conflicts between Science and Religion

Victor Chen, Carlie McCleary, Frank Staszak, and Joe Houck

Research Problem Statement

In this project, we analyze how a person's demographic, professional, and socioeconomic backgrounds and behavioral traits may predict how (s)he responds to conflicts between science and religious belief. It is based on a large-scale global survey dataset, collected by the World Values Survey between 1981 and 2020.

Research Problem Statement

World Values Survey Wave 7: 2017-2020

Q15.- Important child qualities: religious faith

	TOTAL	ISO 3166-1 numeric country code				
		China	France	Germany	Egypt	United States
Important	20.7%	1.1%	8.9%	10.8%	81.7%	32.1%
Not mentioned	79.0%	98.4%	90.8%	88.9%	18.2%	67.7%
Don´t know	0.1%	-	0.3%	0.2%	0.1%	0.0%
No answer	0.2%	0.5%	-	0.0%	-	0.1%
(N)	(10,890)	(3,036)	(1,880)	(2,178)	(1,200)	(2,596)

Data

Data is collected from the World Values Survey, which contains over 500 survey questions and answers collected from almost half a million citizens around the world since 1981.

Specific Outcome of Interest - Respondent's answer to survey question

Please tell us if you strongly agree, agree, disagree, or strongly disagree with the following statements: Whenever science and religion conflict, religion is always right

1 Strongly agree

2 Agree

3 Disagree

4 Strongly disagree

-1 Don´t know

-2 No answer

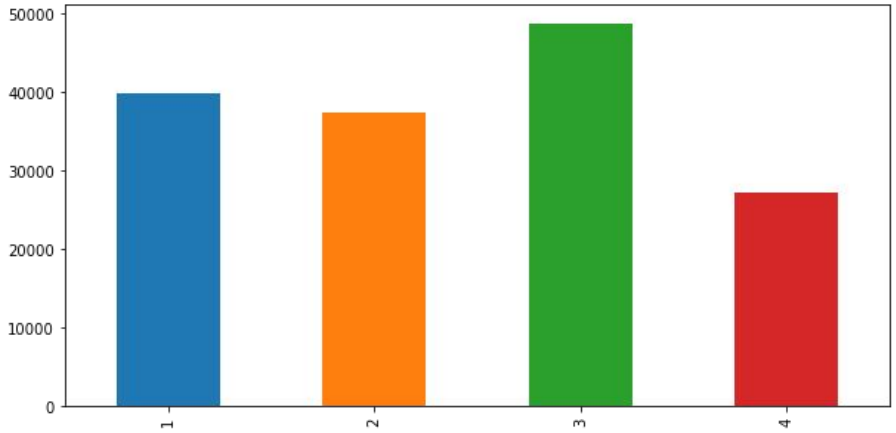
-3 Not applicable

-4 Not asked

-5 Missing;Unknown

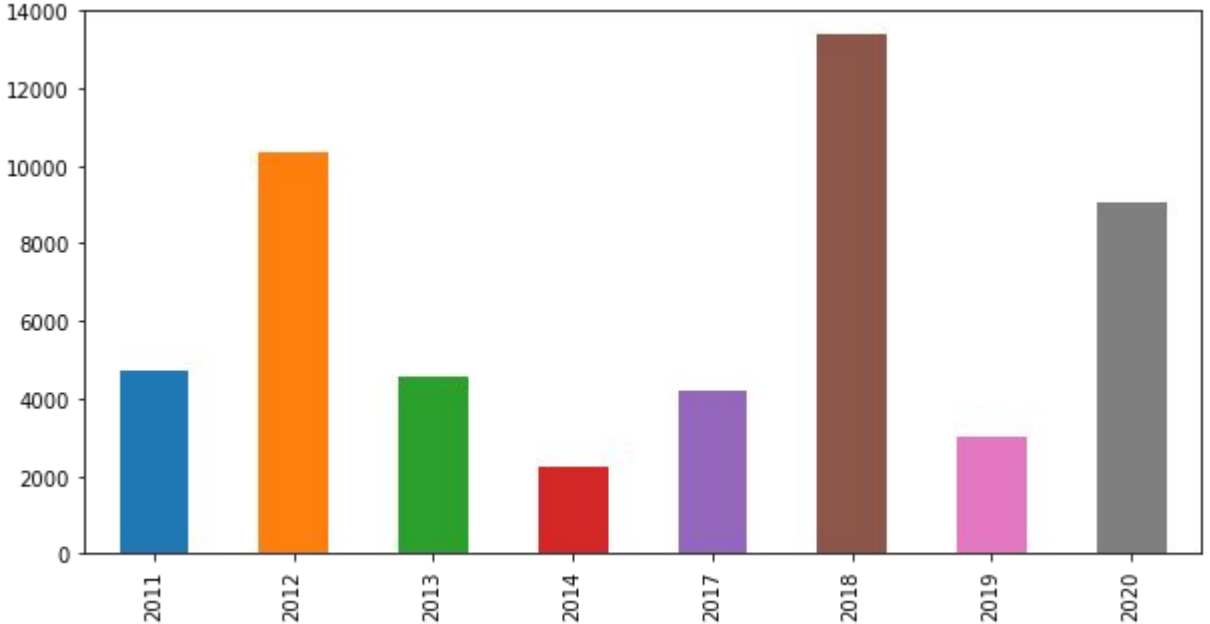
Descriptive statistics of numeric variables

	count	mean	std	min	25%	50%	75%	max
Religion_over_science	51,488.000	-0.000	1.000	-1.422	-0.453	0.516	0.516	1.485



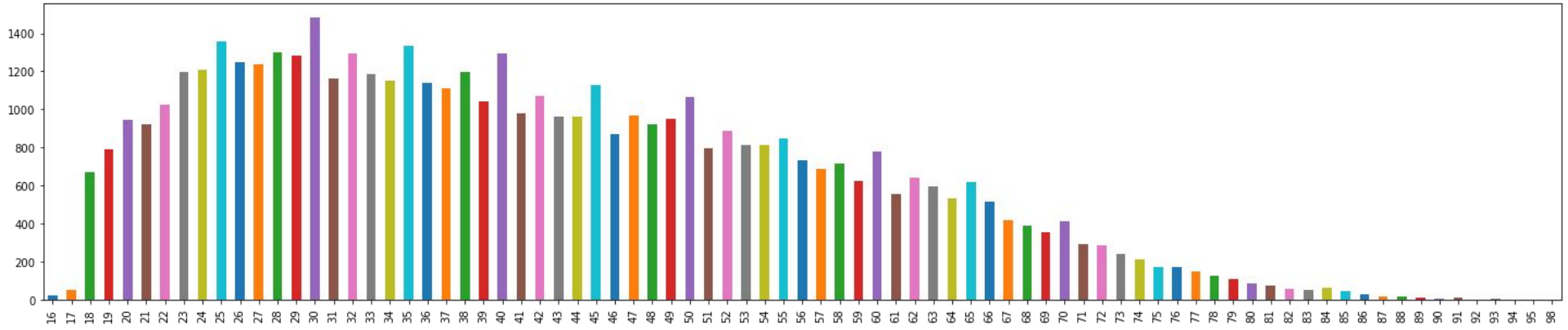
Preliminary analysis

Sample distribution by Year of interviews



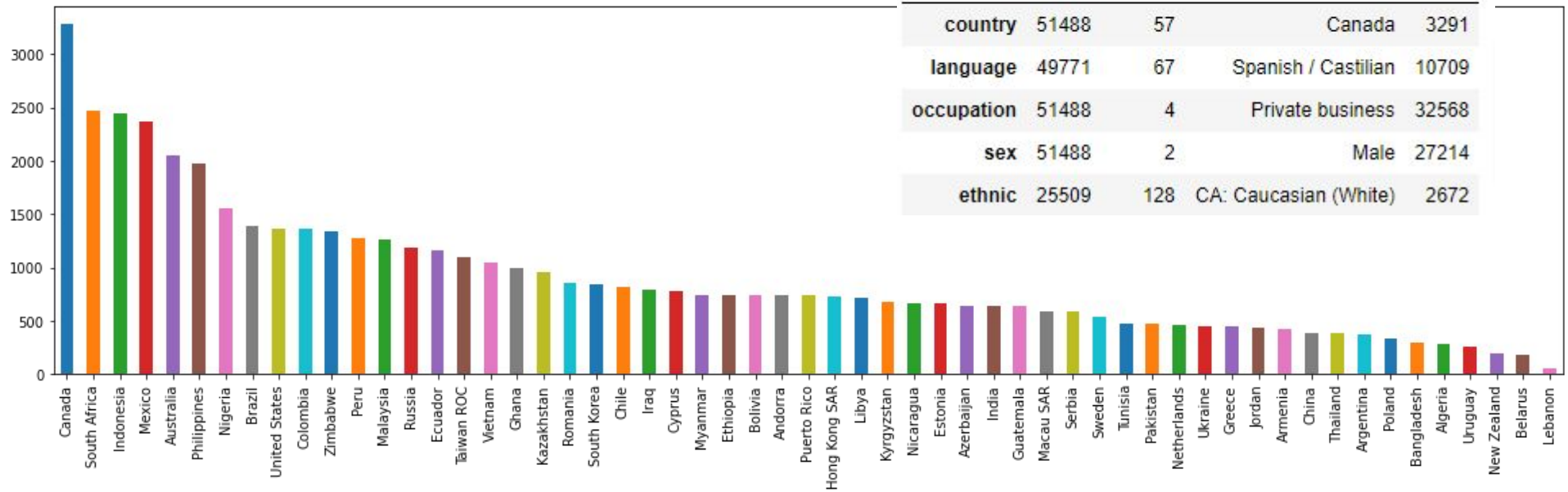
Preliminary analysis (cont'd)

Sample distribution by Age at the time of interview



Preliminary analysis (cont'd)

Sample distribution by Country



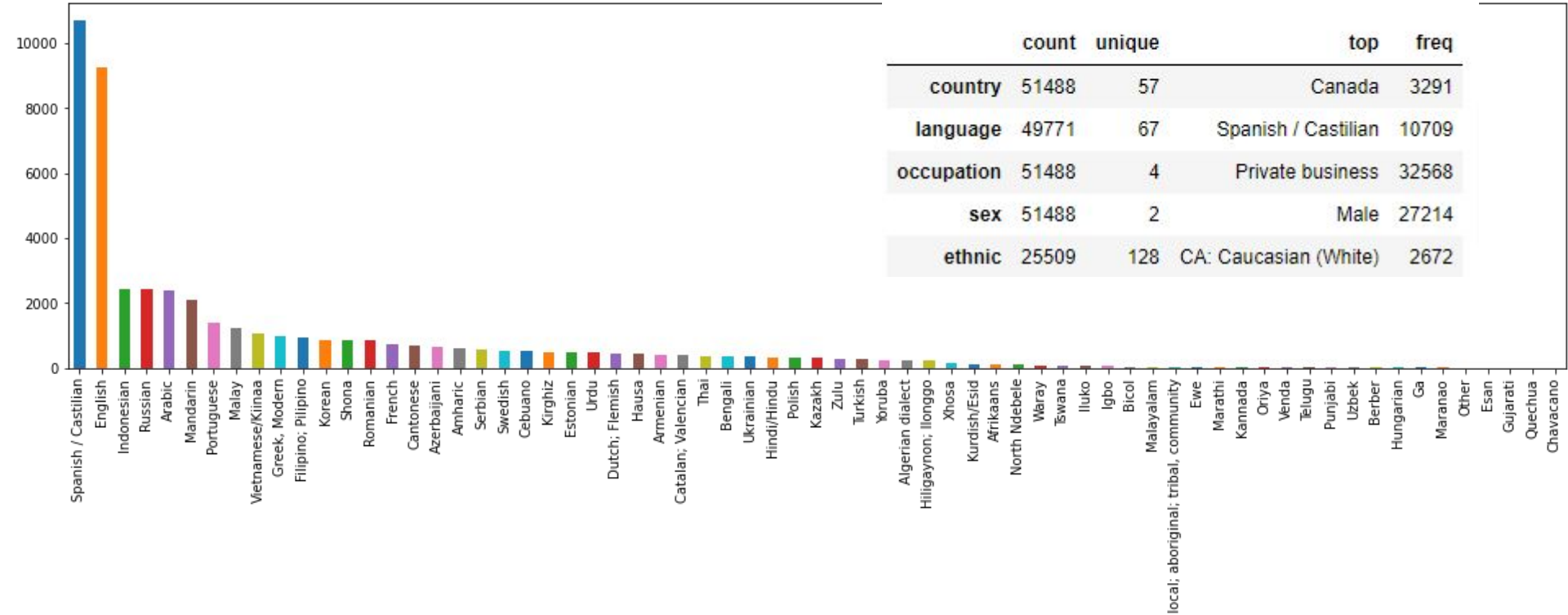
Descriptive statistics of categorical variables

---

	count	unique	top	freq
country	51488	57	Canada	3291
language	49771	67	Spanish / Castilian	10709
occupation	51488	4	Private business	32568
sex	51488	2	Male	27214
ethnic	25509	128	CA: Caucasian (White)	2672

Preliminary analysis (cont'd)

Sample distribution by Language



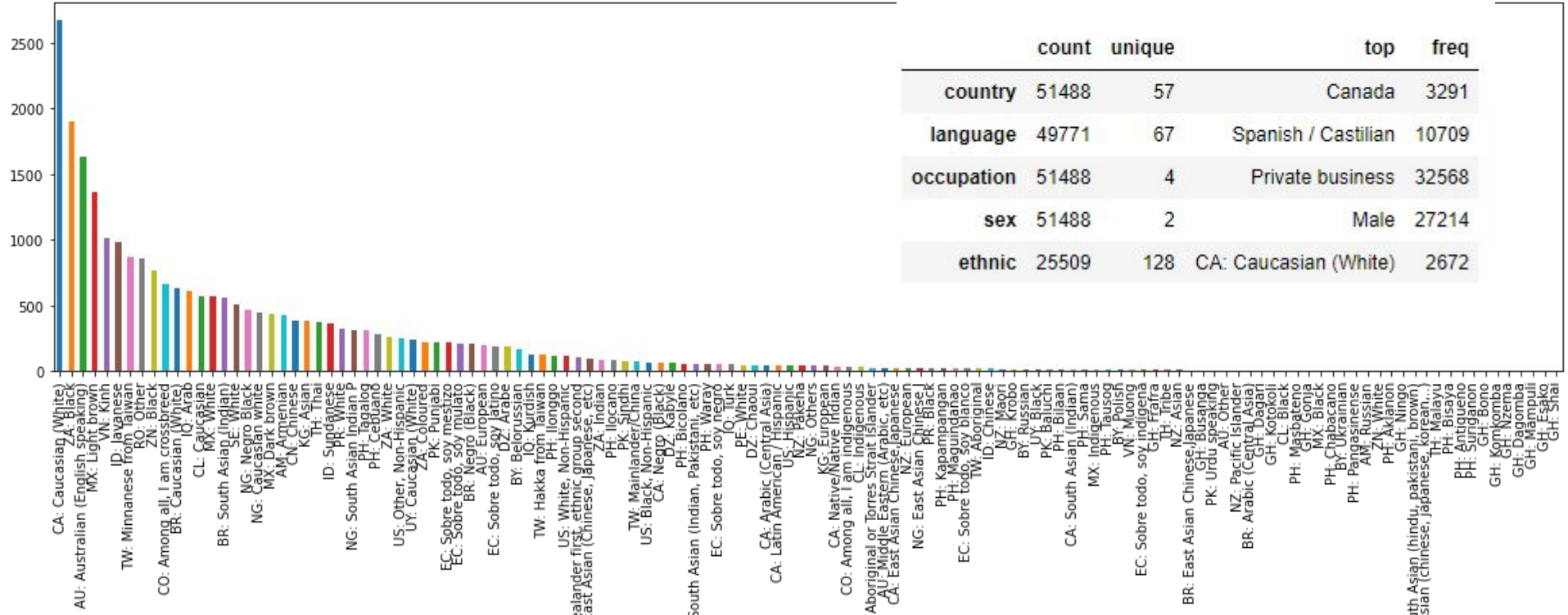
Descriptive statistics of categorical variables

---

	count	unique	top	freq
country	51488	57	Canada	3291
language	49771	67	Spanish / Castilian	10709
occupation	51488	4	Private business	32568
sex	51488	2	Male	27214
ethnic	25509	128	CA: Caucasian (White)	2672

Preliminary analysis (cont'd)

Sample distribution by Ethnicity

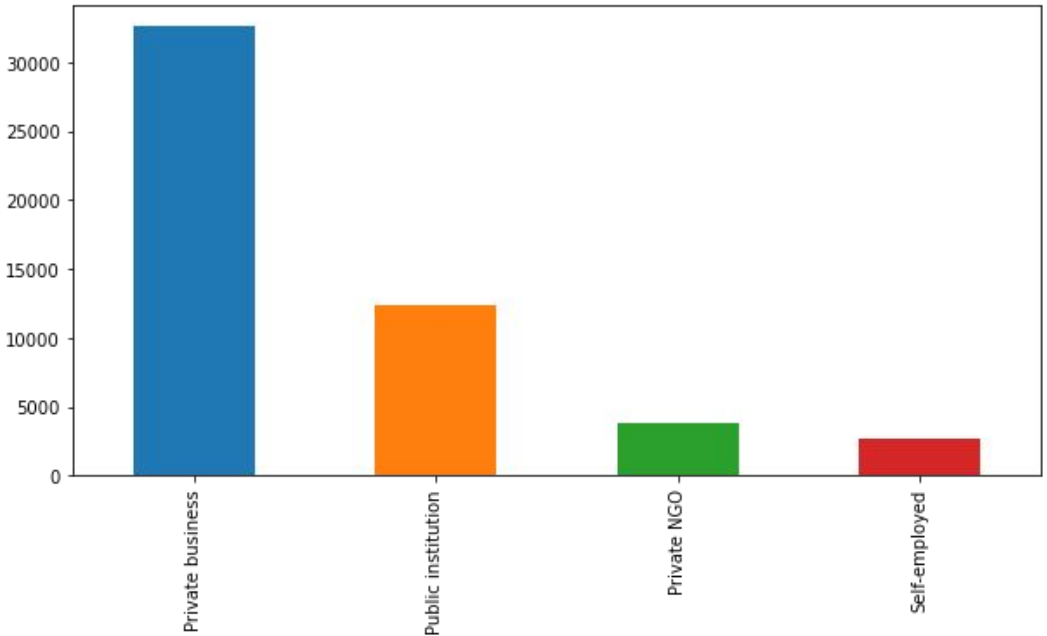


Descriptive statistics of categorical variables
---

	count	unique	top	freq
country	51488	57	Canada	3291
language	49771	67	Spanish / Castilian	10709
occupation	51488	4	Private business	32568
sex	51488	2	Male	27214
ethnic	25509	128	CA: Caucasian (White)	2672

Preliminary analysis (cont'd)

Sample distribution by Occupation



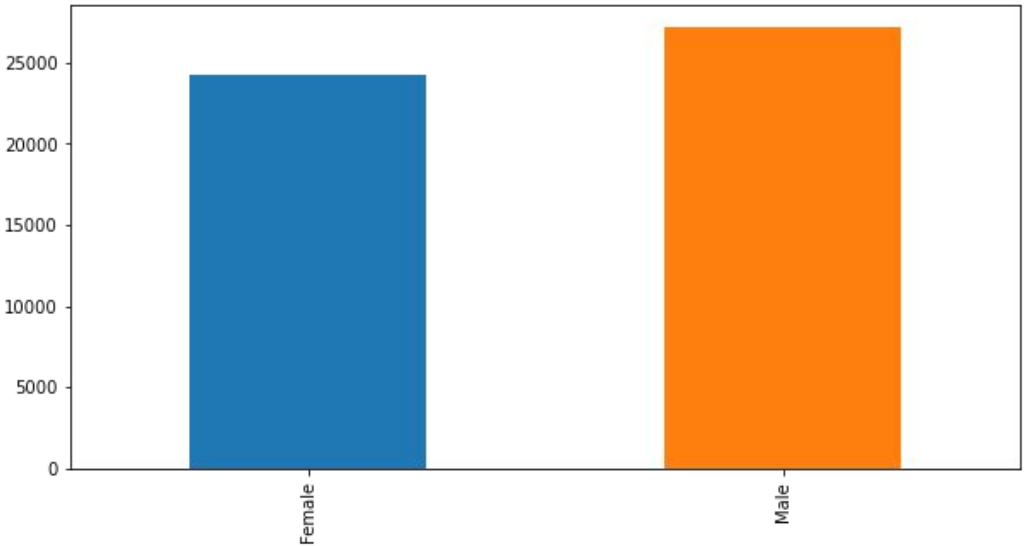
Descriptive statistics of categorical variables

---

	count	unique	top	freq
country	51488	57	Canada	3291
language	49771	67	Spanish / Castilian	10709
occupation	51488	4	Private business	32568
sex	51488	2	Male	27214
ethnic	25509	128	CA: Caucasian (White)	2672

Preliminary analysis (cont'd)

Sample distribution by Sex



Descriptive statistics of categorical variables

---

	count	unique	top	freq
country	51488	57	Canada	3291
language	49771	67	Spanish / Castilian	10709
occupation	51488	4	Private business	32568
sex	51488	2	Male	27214
ethnic	25509	128	CA: Caucasian (White)	2672

Feature selection

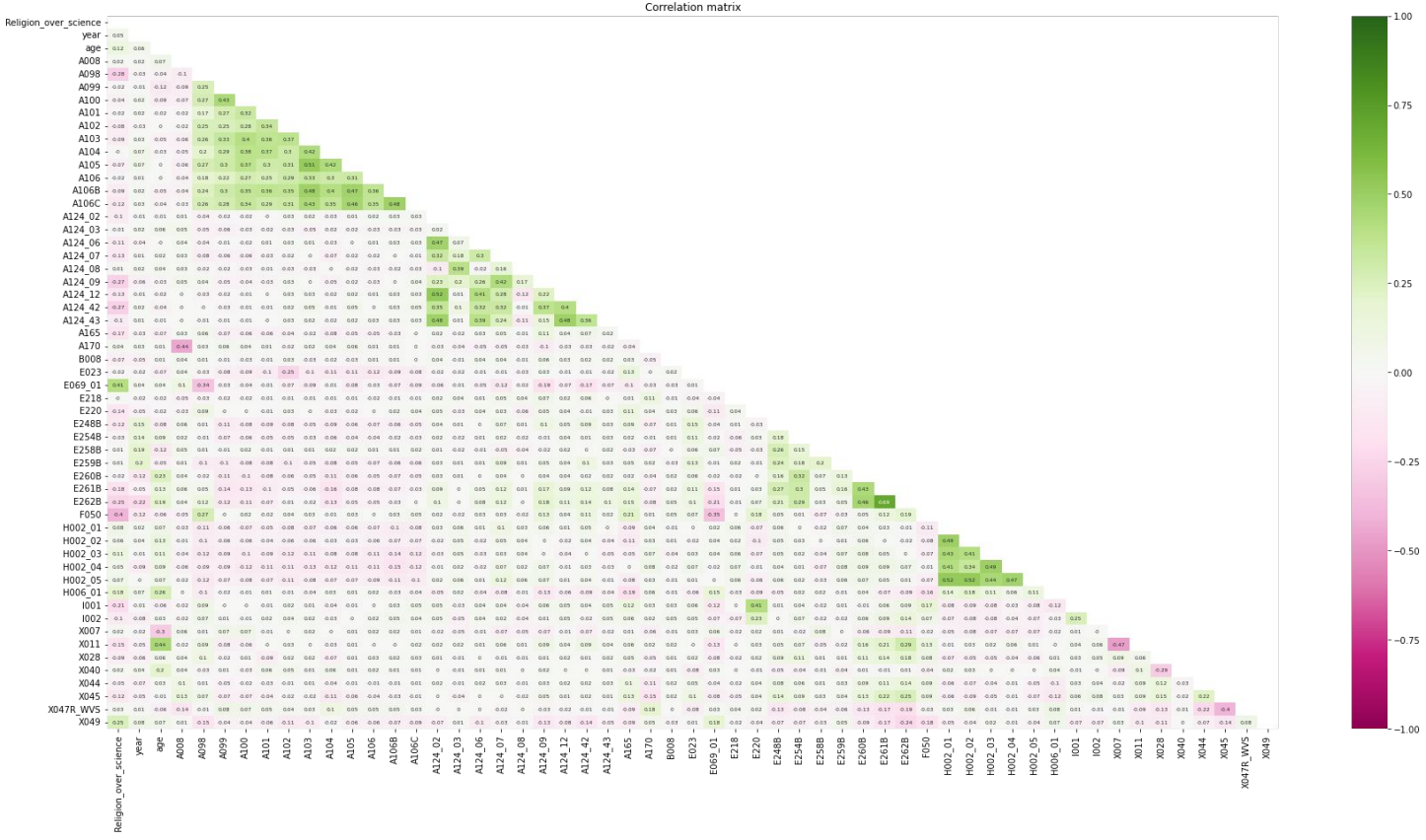
A group voting based on common sense, removing those with too many missing values

```
----------------------------------------------------------------------------------------------------
Descriptions of all variable labels other than the renamed ones
----------------------------------------------------------------------------------------------------
```

[13]:

	variable_code	variable_question
0	A008	Taking all things together, would you say you are:
1	A098	Now I am going to read out a list of voluntary organizations; for each one, could you tell me whether you are a member, an active member, an inactive member or not a member of that type of organization? Church or religious organization
2	A099	Now I am going to read out a list of voluntary organizations; for each one, could you tell me whether you are a member, an active member, an inactive member or not a member of that type of organization? Sport or recreational organization, football/baseball/rugby team
3	A100	Now I am going to read out a list of voluntary organizations; for each one, could you tell me whether you are a member, an active member, an inactive member or not a member of that type of organization? Art, music or educational organization
4	A101	Now I am going to read out a list of voluntary organizations; for each one, could you tell me whether you are a member, an active member, an inactive member or not a member of that type of organization? Labour Union
5	A102	Now I am going to read out a list of voluntary organizations; for each one, could you tell me whether you are a member, an active member, an inactive member or not a member of that type of organization? Political party
6	A103	Now I am going to read out a list of voluntary organizations; for each one, could you tell me whether you are a member, an active member, an inactive member or not a member of that type of organization? Environmental organization
7	A104	Now I am going to read out a list of voluntary organizations; for each one, could you tell me whether you are a member, an active member, an inactive member or not a member of that type of organization? Professional association

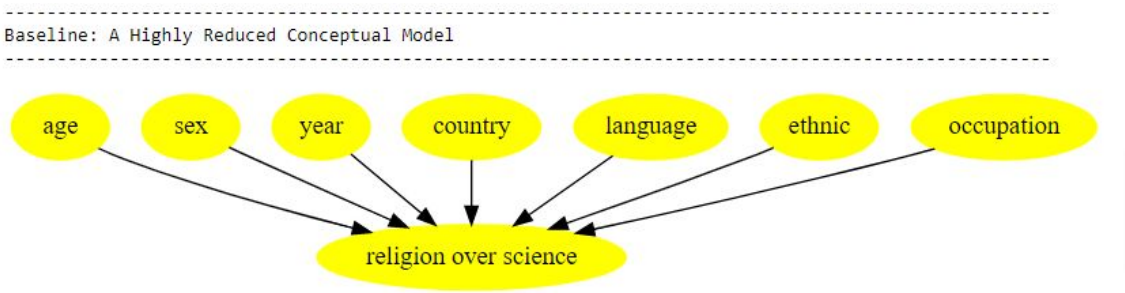
Feature selection



Machine Learning: Baseline model

Baseline Model: OLS of outcome, predicted by dummies of **age, sex, year, country, language, and occupation**.

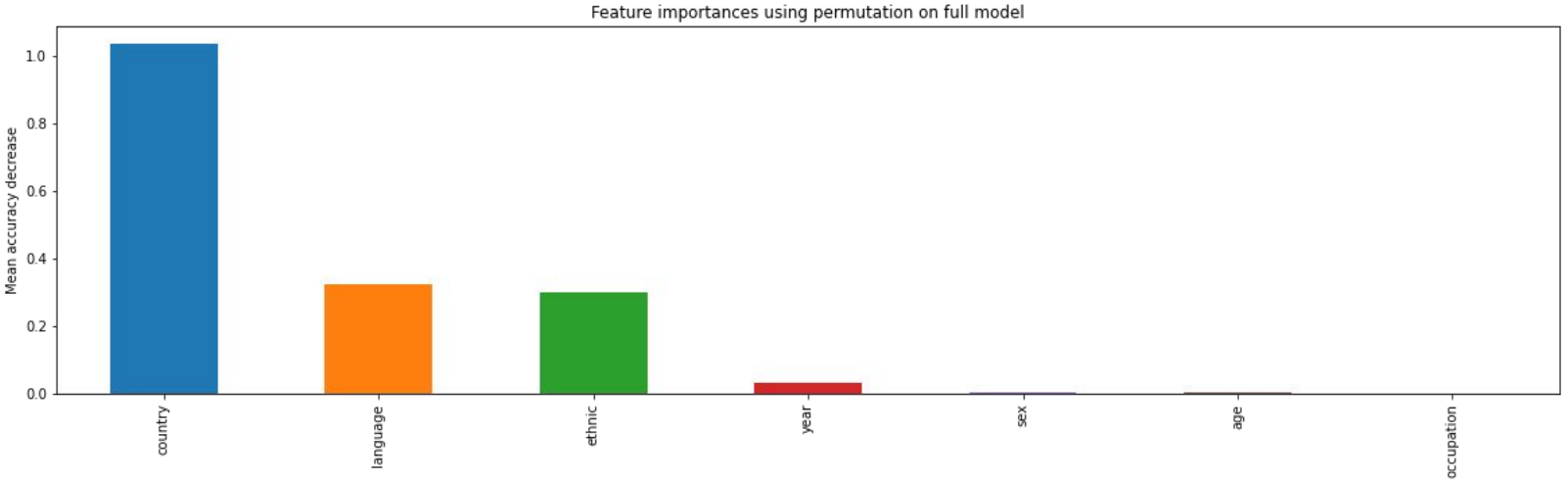
- R2 (train): 0.403
- R2 (test): 0.397
- Time (secs): 1.0629987716674805



Machine Learning: Baseline model

Baseline Model: OLS of outcome, predicted by dummies of age, sex, year, country, language, and occupation.

- R2 (train): 0.403
- R2 (test): 0.397
- Time (secs): 1.0629987716674805



Model 1: LASSO Regression

Justification for use of LASSO

- The World Values Survey data contains many dimensions, with high complexity.
  - LASSO uses L1 regularization as a method to minimize overfitting. Lasso regression lowers complexity and reduces noise in the model by shrinking coefficients, reducing the number of predictors selected.
- R-squared as a metric of evaluation.
  - R-squared provides a good measurement of how well the model fits the dependent variables based on the penalties enforced by Lasso regression.

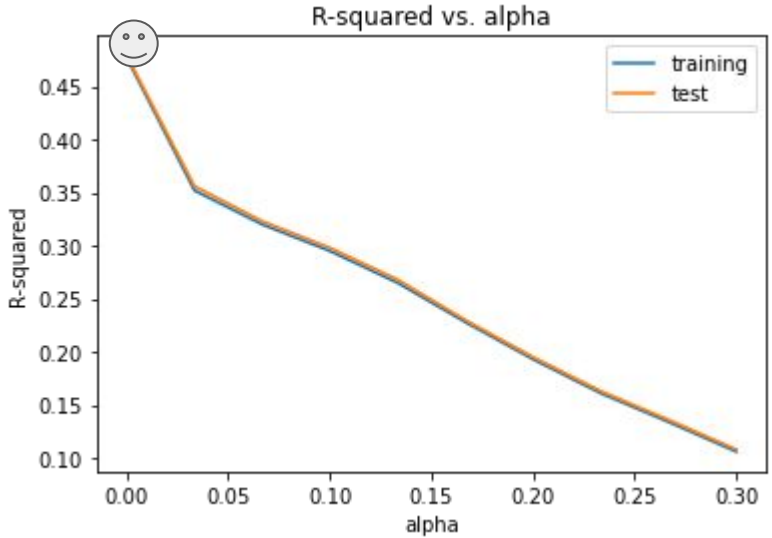
Challenges Encountered

- The penalties enforced by alpha hyperparameter tuning were high and reduced the data significantly, which reduced model complexity and helped identify significant features but at the cost of lowering prediction information.

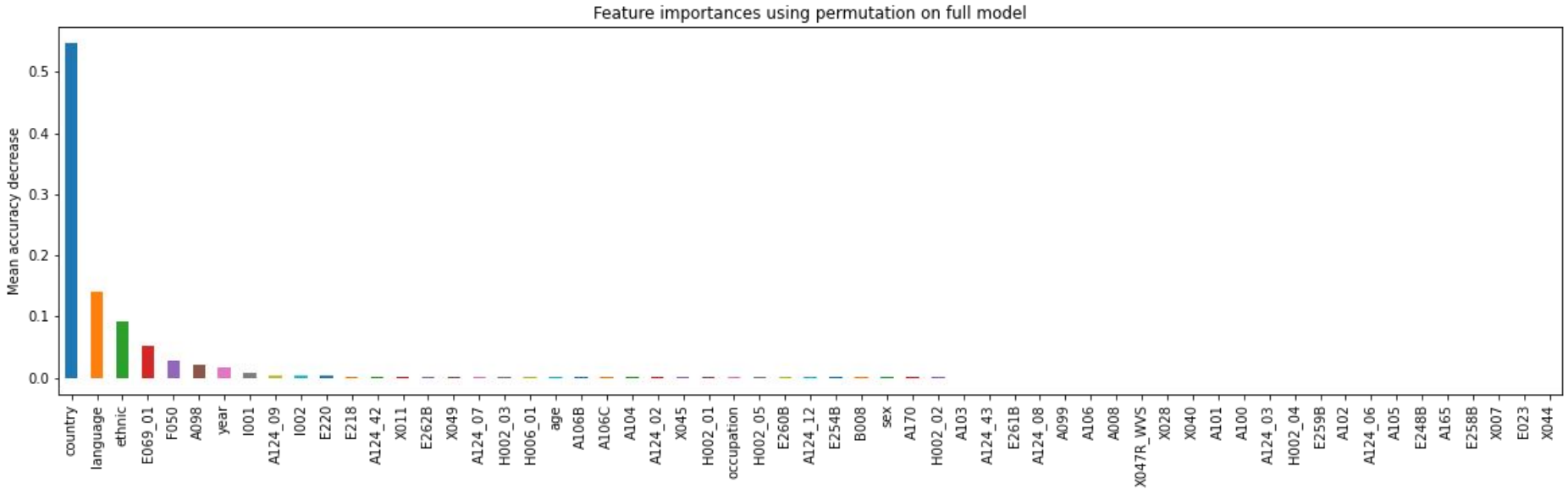
Model 1: LASSO Regression

	alpha	R2 (train)	R2 (test)	Time (sec)
0	0.0001	0.4770	0.4790	26.1797
1	0.0334	0.3520	0.3560	1.1230
2	0.0667	0.3200	0.3230	1.1189
3	0.1001	0.2950	0.2980	1.1590
4	0.1334	0.2650	0.2680	1.2110
5	0.1667	0.2280	0.2300	1.2220
6	0.2000	0.1930	0.1950	1.0670
7	0.2334	0.1610	0.1630	1.1150
8	0.2667	0.1340	0.1360	1.0440
9	0.3000	0.1060	0.1080	1.0209

At its peak, the regression modeled almost 48% of the outcomes.

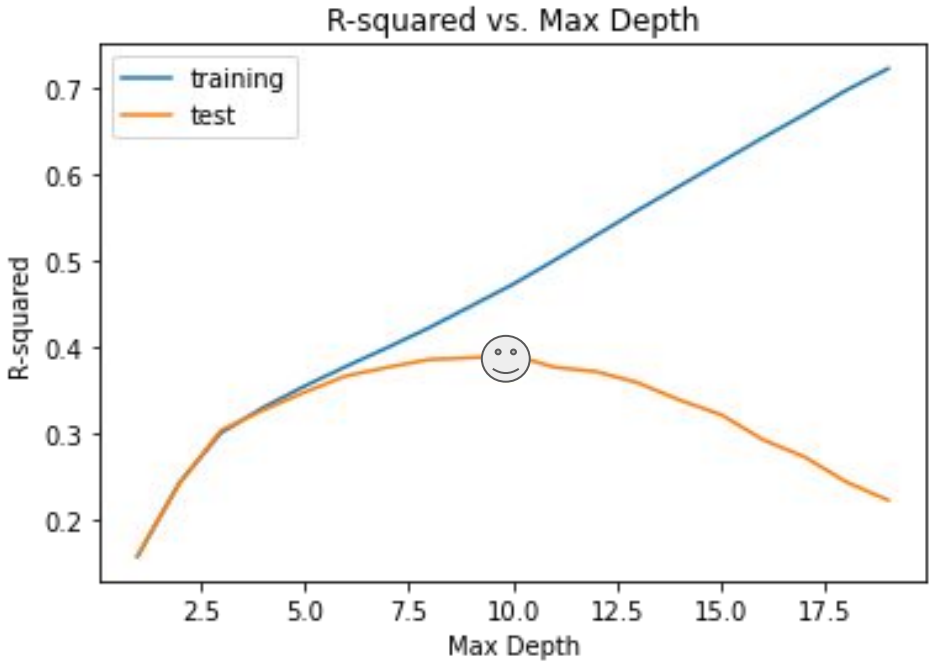


Model 1: LASSO Regression



Model 2: Decision Trees - Optimizing Parameters

	Max Depth	R2 (train)	R2 (test)	Time (sec)
0	1	0.1580	0.1590	0.7000
1	2	0.2430	0.2420	0.7320
2	3	0.3010	0.3040	0.7960
3	4	0.3300	0.3270	0.8911
4	5	0.3550	0.3480	1.0660
5	6	0.3780	0.3670	1.3060
6	7	0.4000	0.3770	1.6671
7	8	0.4230	0.3860	2.1129
8	9	0.4480	0.3880	2.7831
9	10	0.4730	0.3910	3.5100
10	11	0.5010	0.3770	4.3730
11	12	0.5300	0.3720	5.3019
12	13	0.5590	0.3590	6.6400
13	14	0.5870	0.3390	9.2246
14	15	0.6150	0.3220	9.5495
15	16	0.6430	0.2930	9.6465
16	17	0.6700	0.2730	10.9770
17	18	0.6980	0.2440	11.4315
18	19	0.7230	0.2230	11.3191

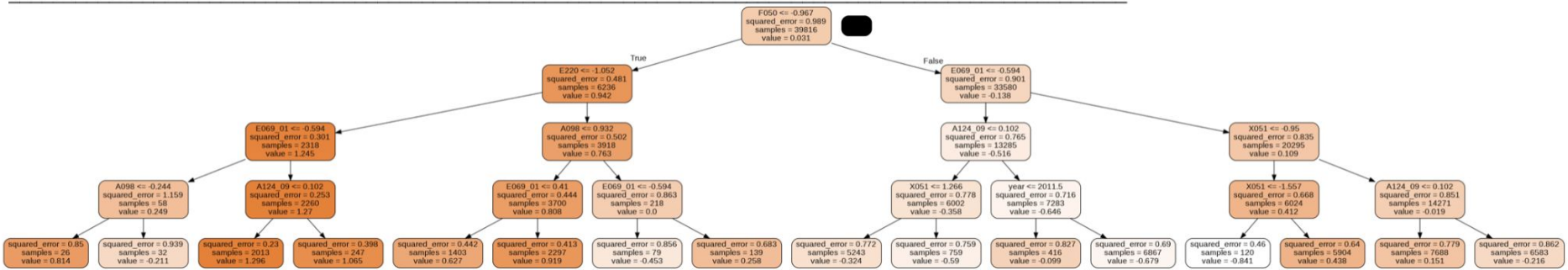


Model 2: Decision Tree - Results

R-squared value of 0.396 is about the same as the baseline score of 0.397 using linear regression

```
--------------------------------------------------------------------------------------------------------
Evaluating Model Results
--------------------------------------------------------------------------------------------------------
Model Score:  0.3956
Depth of tree:  9
Number of leaves:  372
Model parameters:  {'ccp_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': 9, 'max_features': None, 'max_leaf_nodes': None, 'min_impu
rity_decrease': 0.0, 'min_samples_leaf': 1, 'min_samples_split': 2, 'min_weight_fraction_leaf': 0.0, 'random_state': None, 'splitter': 'bes
t'}
Cross Validation Score:  [0.35927401 0.34935372 0.34673792 0.3526076  0.33471983 0.34297547
 0.38402932 0.37305161 0.36221815 0.39400221]
```

```
--------------------------------------------------------------------------------------------------------
Visualization of Decision Trees
--------------------------------------------------------------------------------------------------------
```



Decision Tree visuals allow for easy interpretation of how features influence prediction

Model 2: Decision Tree - What do the results tell us?

The features with the most importance were language, country, and the below respondent questions

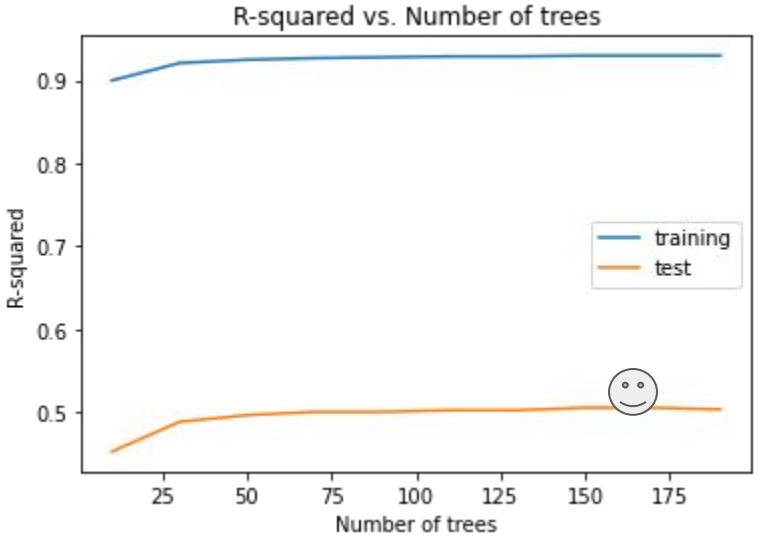
```
----------------------------------------------------------------------------------------------------
Top 5 most important features
----------------------------------------------------------------------------------------------------
language   0.1898
E069_01    0.1617
F050       0.1414
country    0.0494
A098       0.0446
dtype: float64
```

Note: descriptions for feature labels below

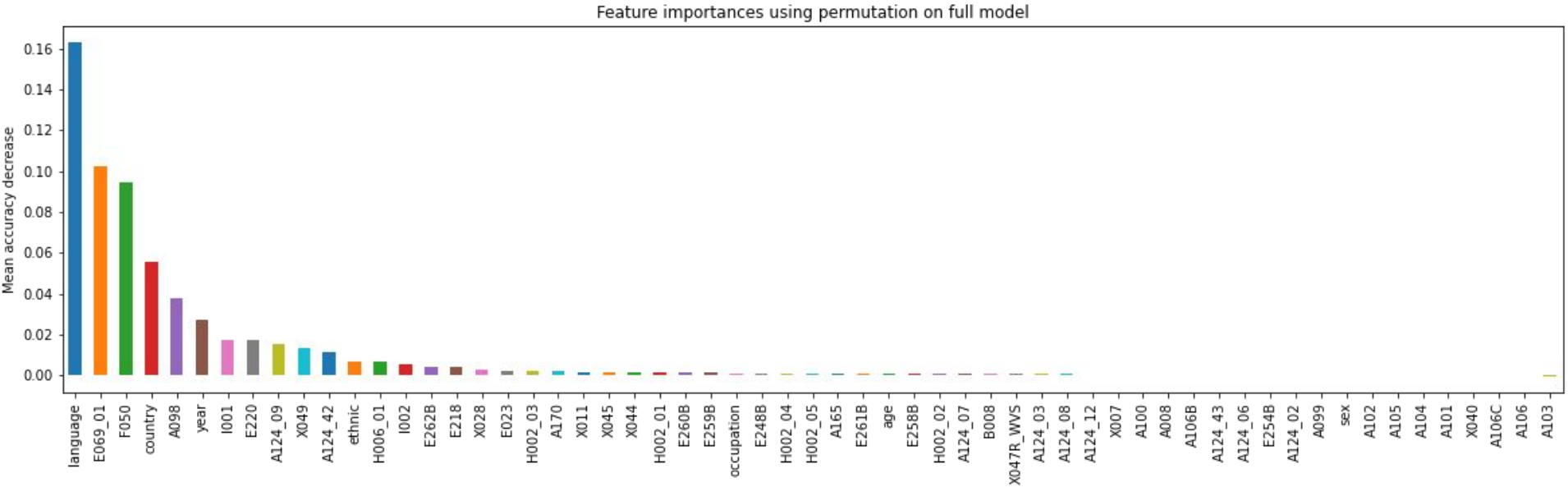
variable_code	variable_question
A098	Now I am going to read out a list of voluntary organizations; for each one, could you tell me whether you are a member, an active member, an inactive member or not a member of that type of organization? Church or religious organization
E069_01	I am going to name a number of organizations. For each one, could you tell me how much confidence you have in them: is it a great deal of confidence, quite a lot of confidence, not very much confidence or none at all? The Churches (mosque, temple etc.)
F050	In which of the following things do you believe, if you believe in any? God

Model 3: Random Forest

	Number of trees	R2 (train)	R2 (test)	Time (sec)
0	10	0.9000	0.4520	140.9888
1	30	0.9210	0.4880	477.5985
2	50	0.9250	0.4960	812.9001
3	70	0.9270	0.5000	1,037.0182
4	90	0.9280	0.5000	1,224.2471
5	110	0.9290	0.5020	1,405.0618
6	130	0.9290	0.5020	1,547.9552
7	150	0.9300	0.5050	1,760.1569
8	170	0.9300	0.5050	1,989.5166
9	190	0.9300	0.5030	2,198.9249



Model 3: Random Forest (cont'd)



Model 3: Random Forest (cont'd)

```
-----------------------------------------------------------------------------------------------------
Top 5 most important features
-----------------------------------------------------------------------------------------------------
language   0.1628
E069_01    0.1025
F050       0.0944
country    0.0554
A098       0.0381
dtype: float64
```

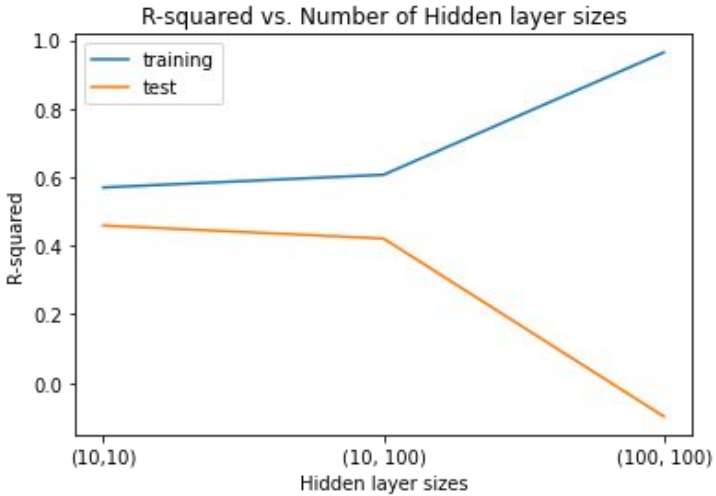
Note: descriptions for feature labels below

variable_code	variable_question
A098	Now I am going to read out a list of voluntary organizations; for each one, could you tell me whether you are a member, an active member, an inactive member or not a member of that type of organization? Church or religious organization
E069_01	I am going to name a number of organizations. For each one, could you tell me how much confidence you have in them: is it a great deal of confidence, quite a lot of confidence, not very much confidence or none at all? The Churches (mosque, temple etc.)
F050	In which of the following things do you believe, if you believe in any? God

Model 4: Neural Networks

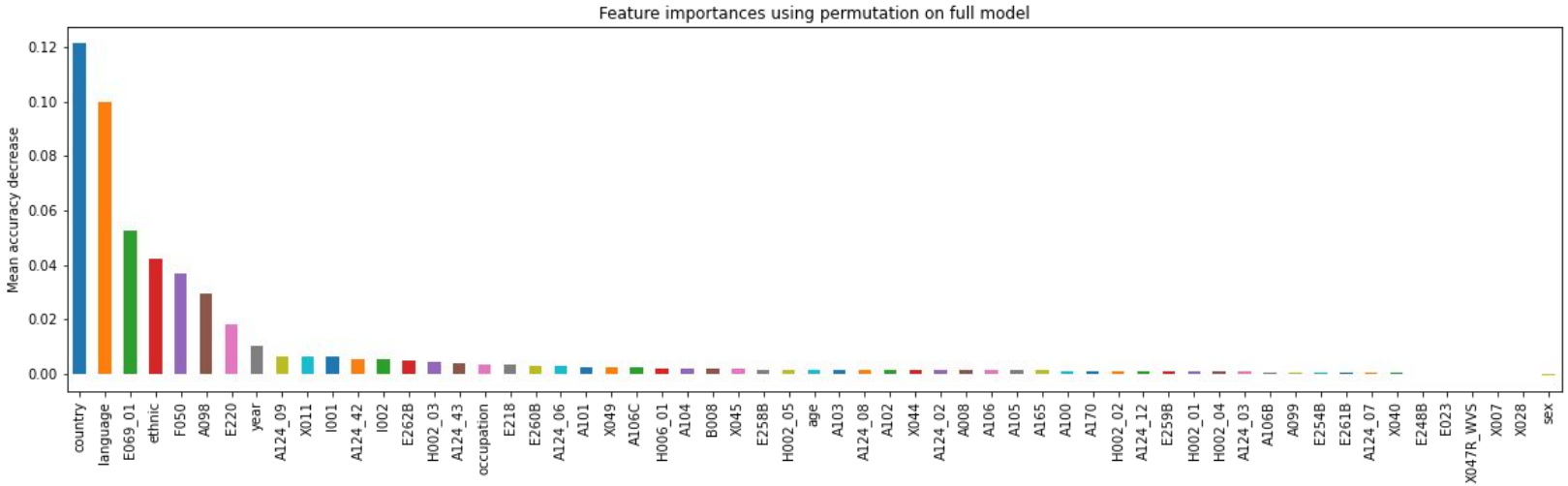
Leveraged MLP Regressor for the Neural Network model. The baseline model is based on pre-existing model parameters

	Hidden layer sizes	R2 (train)	R2 (test)	Time (sec)
0	(10, 10)	0.5700	0.4590	59.3333
1	(10, 100)	0.6070	0.4210	162.3520
2	(100, 100)	0.9640	-0.0980	291.1953



Model 4: Neural Networks

Leveraged MLP Regressor for the Neural Network model. The baseline model is based on pre-existing model parameters



Model 4: Neural Networks

Leveraged MLP Regressor for the Neural Network model. The baseline model is based on pre-existing model parameters

```
----------------------------------------------------------------------------------------------------
Top 5 most important features
----------------------------------------------------------------------------------------------------
country     0.1212
language    0.1000
E069_01     0.0524
ethnic      0.0424
F050        0.0369
dtype: float64
```

Note: descriptions for feature labels below

variable_code	variable_question
E069_01	I am going to name a number of organizations. For each one, could you tell me how much confidence you have in them: is it a great deal of confidence, quite a lot of confidence, not very much confidence or none at all? The Churches (mosque, temple etc.)
F050	In which of the following things do you believe, if you believe in any? God

Conclusion and Discussion

- Tradeoffs between accuracy and time efficiency
- Explainability is very important, e.g., decision tree image, feature importance
- Language and country are the most important predictors, more than how people believe/belong…

- What we would have done differently, if we had more time?