The University of Texas at Austin
McCombs School of Business

UNDERSTAND COVID-19 VACCINATION RATE

WHAT'S LIMITING US?

Jessie Lee, Ambikha Maharaj, Charan Musunuru, Karthick Ramasubramanian
TEAM 15 , The University of Texas at Austin

THE HEADLINE

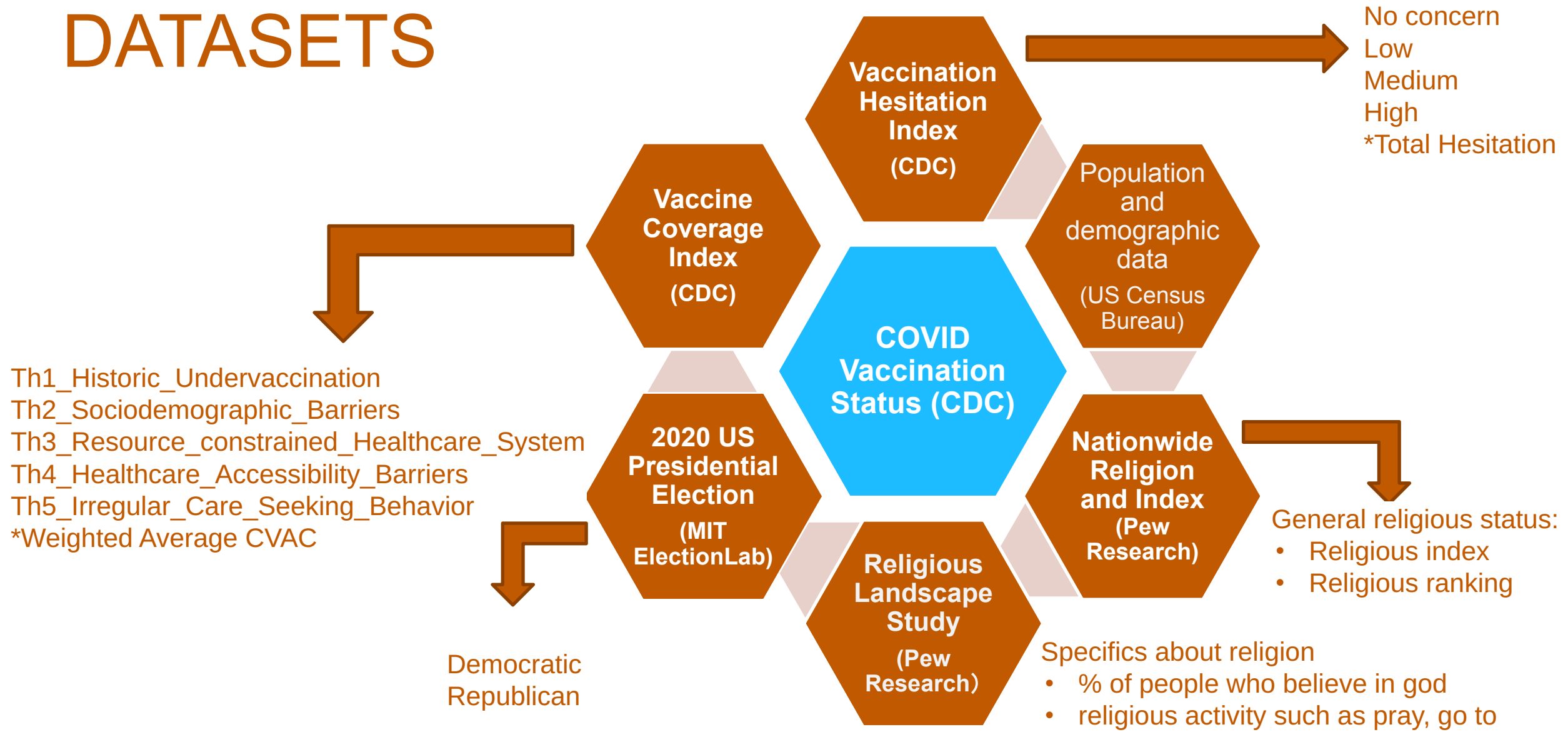
We are in a race against time to vaccinate Americans quickly, safely, and equitably.

- In February 2021, CDC study predicted the concerns in some states for vaccination roll out
- Now we know the vaccination rate at this stage, we can recalibrate ourselves to the most current data
- This exercise is to update our prediction model to fit the current vaccination result and understand the limiting factors that has high impact on the vaccination rate

OUTLINE

- Datasets Used
- Data Exploration
- Modeling
  - What are the Important Factors of Vaccination Progress (State level and County level)
- Conclusion

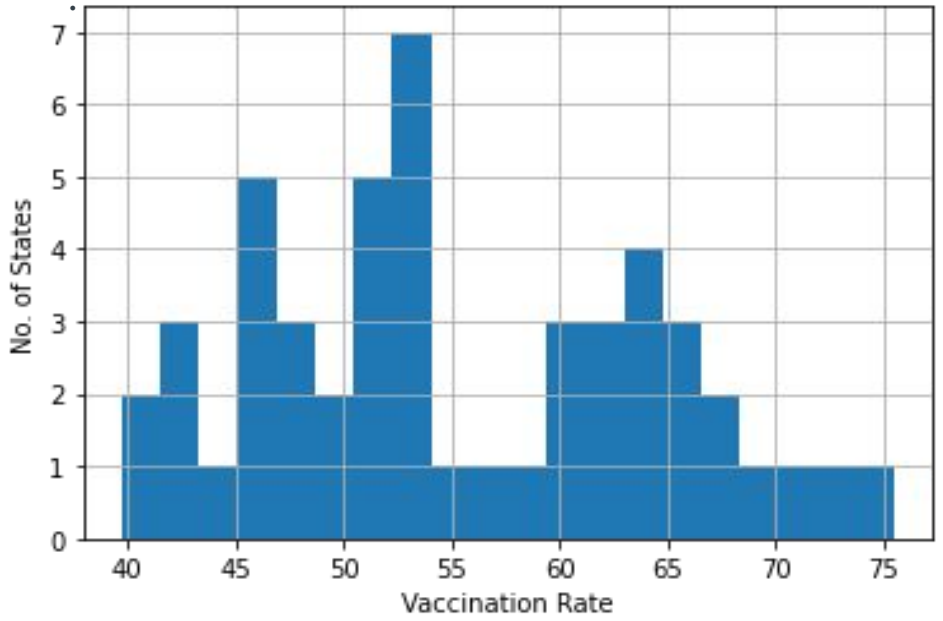
DATASETS


Vaccination Hesitation Index (CDC)
No concern
Low
Medium
High
*Total Hesitation
Vaccine Coverage Index (CDC)
Th1_Historic_Undervaccination
Th2_Sociodemographic_Barriers
Th3_Resource_constrained_Healthcare_System
Th4_Healthcare_Accessibility_Barriers
Th5_Irregular_Care_Seeking_Behavior
*Weighted Average CVAC
Population and demographic data (US Census Bureau)
COVID Vaccination Status (CDC)
2020 US Presidential Election (MIT ElectionLab)
Democratic
Republican
Nationwide Religion and Index (Pew Research)
General religious status:
• Religious index
• Religious ranking
Religious Landscape Study (Pew Research)
Specifics about religion
• % of people who believe in god
• religious activity such as pray, go to church…

EXPLORATORY DATA ANALYSIS (EDA)

STATE VACCINATION RECORDS

- There is a huge difference in vaccination rates across US



Top 10 Vaccinated States

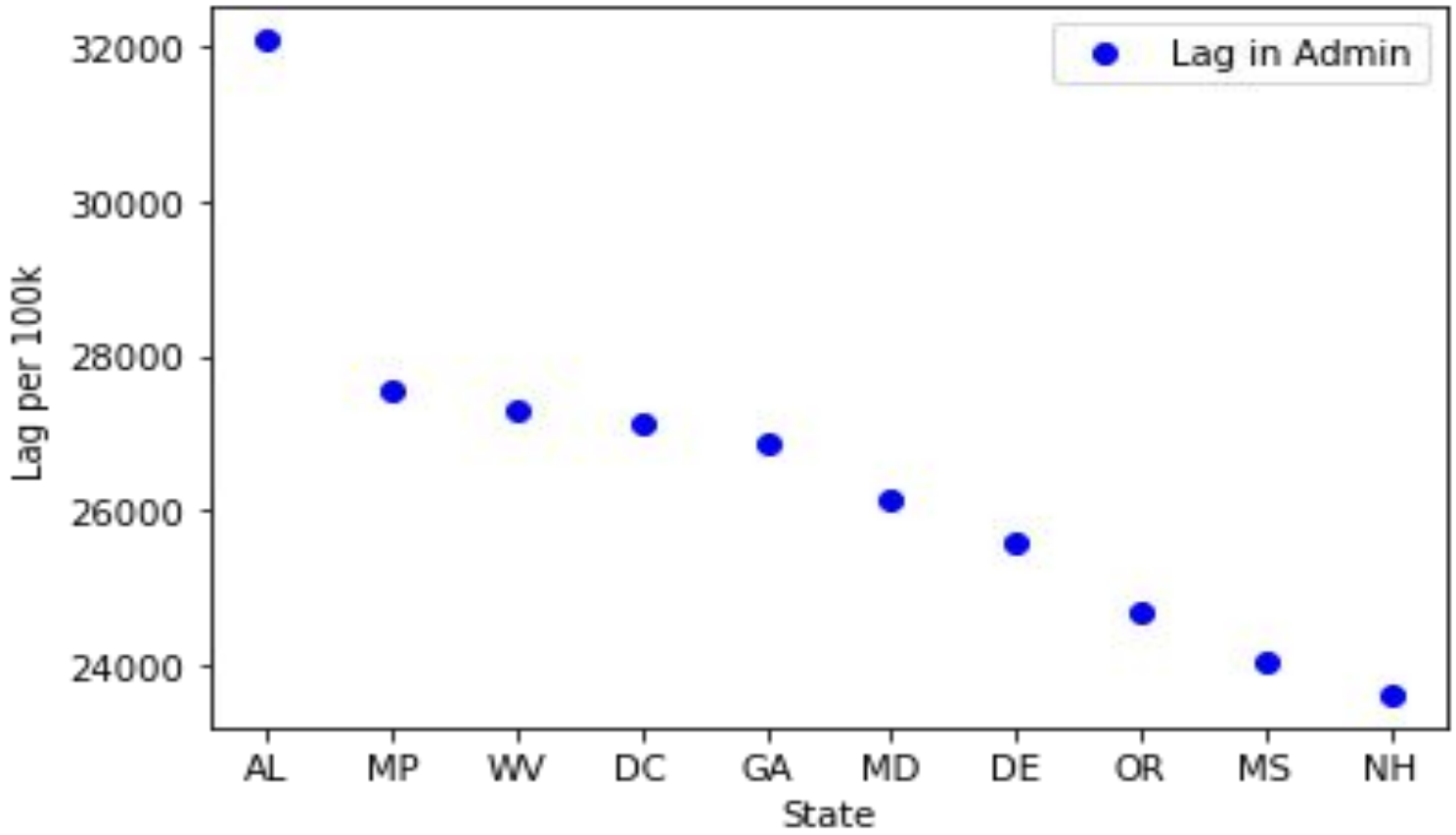
state	vac_rate
Vermont	75.5
Massachusetts	72.6
Hawaii	71.3
Connecticut	69.8
Maine	68.3
Rhode Island	67.3
New Jersey	65.8
Pennsylvania	65.4
New Mexico	65.3
California	64.7

Bottom 10 Vaccinated States

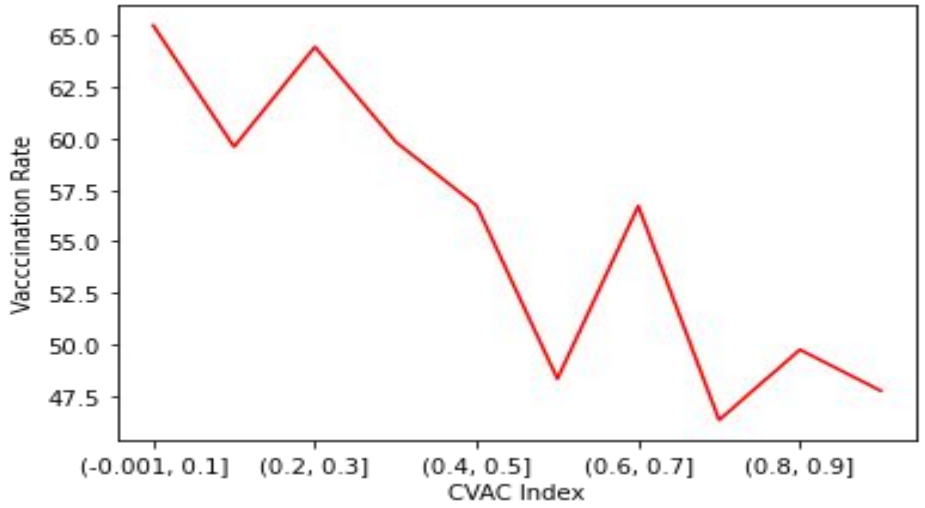
state	vac_rate
Mississippi	39.7
Idaho	41.1
Wyoming	41.6
Louisiana	42.1
Alabama	43.2
Tennessee	44.6
North Dakota	45.4
West Virginia	46.0
Georgia	46.2
South Carolina	46.6

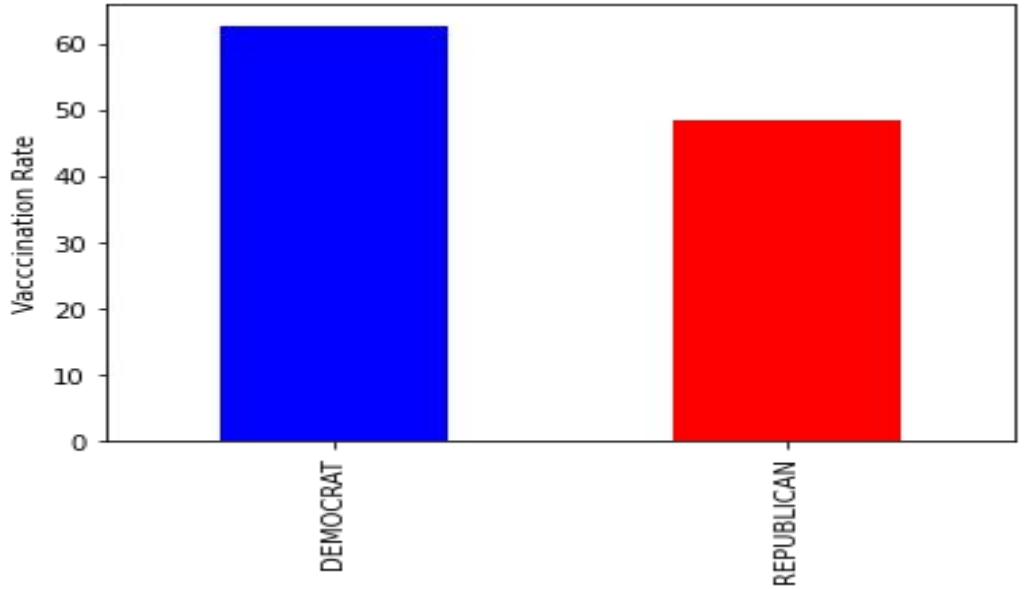
LAG IN DISTRIBUTION AND ADMINISTERED VACCINES

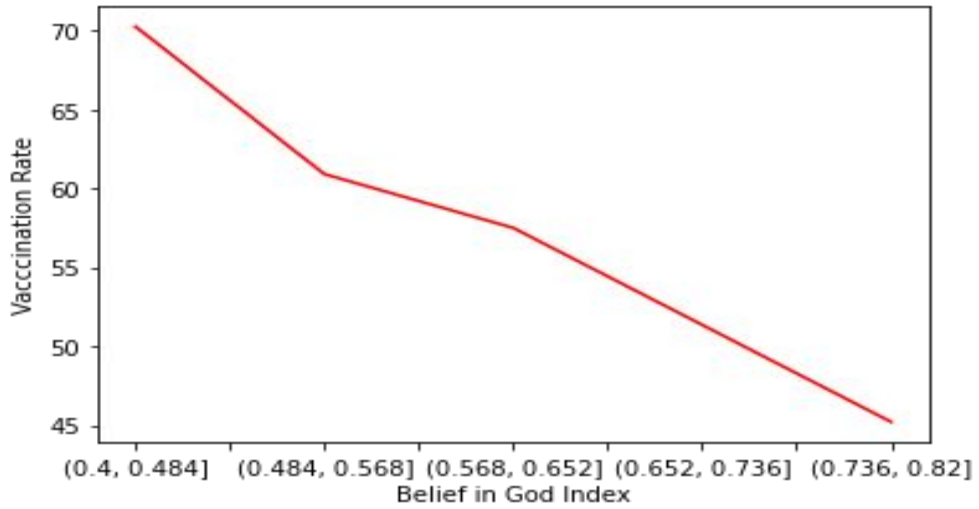
- Vaccines are lying dormant in undervaccinated states
- 4 out of bottom 10 states by vaccination rate have high lag.

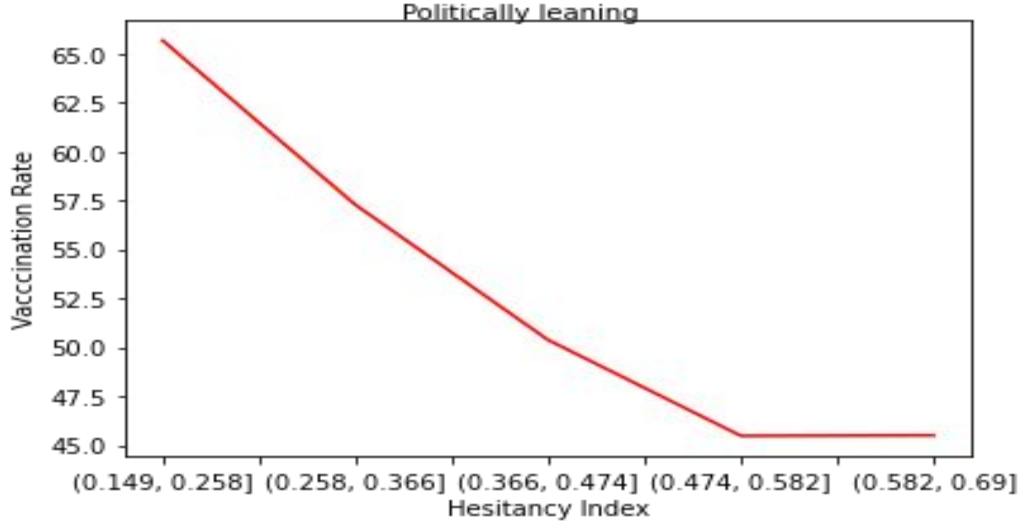


MORE EDA...


Vacccination Rate
65.0
62.5
60.0
57.5
55.0
52.5
50.0
47.5
(-0.001, 0.1]
(0.2, 0.3]
(0.4, 0.5]
(0.6, 0.7]
(0.8, 0.9]
CVAC Index

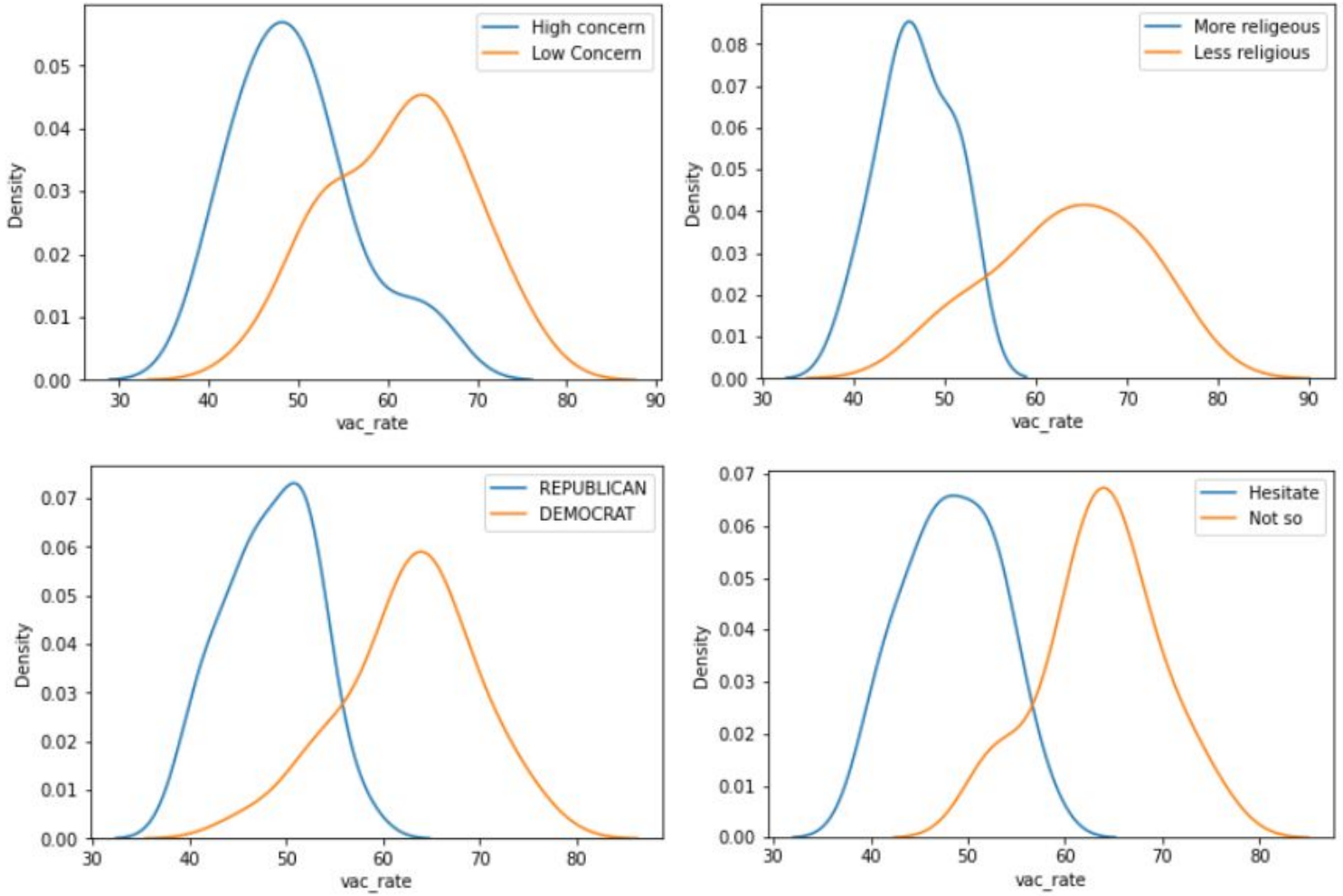

Vacccination Rate
60
50
40
30
20
10
0
DEMOCRAT
REPUBLICAN
Politically leaning


Vacccination Rate
70
65
60
55
50
45
(0.4, 0.484]
(0.484, 0.568]
(0.568, 0.652]
(0.652, 0.736]
(0.736, 0.82]
Belief in God Index


Vacccination Rate
65.0
62.5
60.0
57.5
55.0
52.5
50.0
47.5
45.0
(0.149, 0.258]
(0.258, 0.366]
(0.366, 0.474]
(0.474, 0.582]
(0.582, 0.69]
Hesitancy Index

WHAT MATTERS?

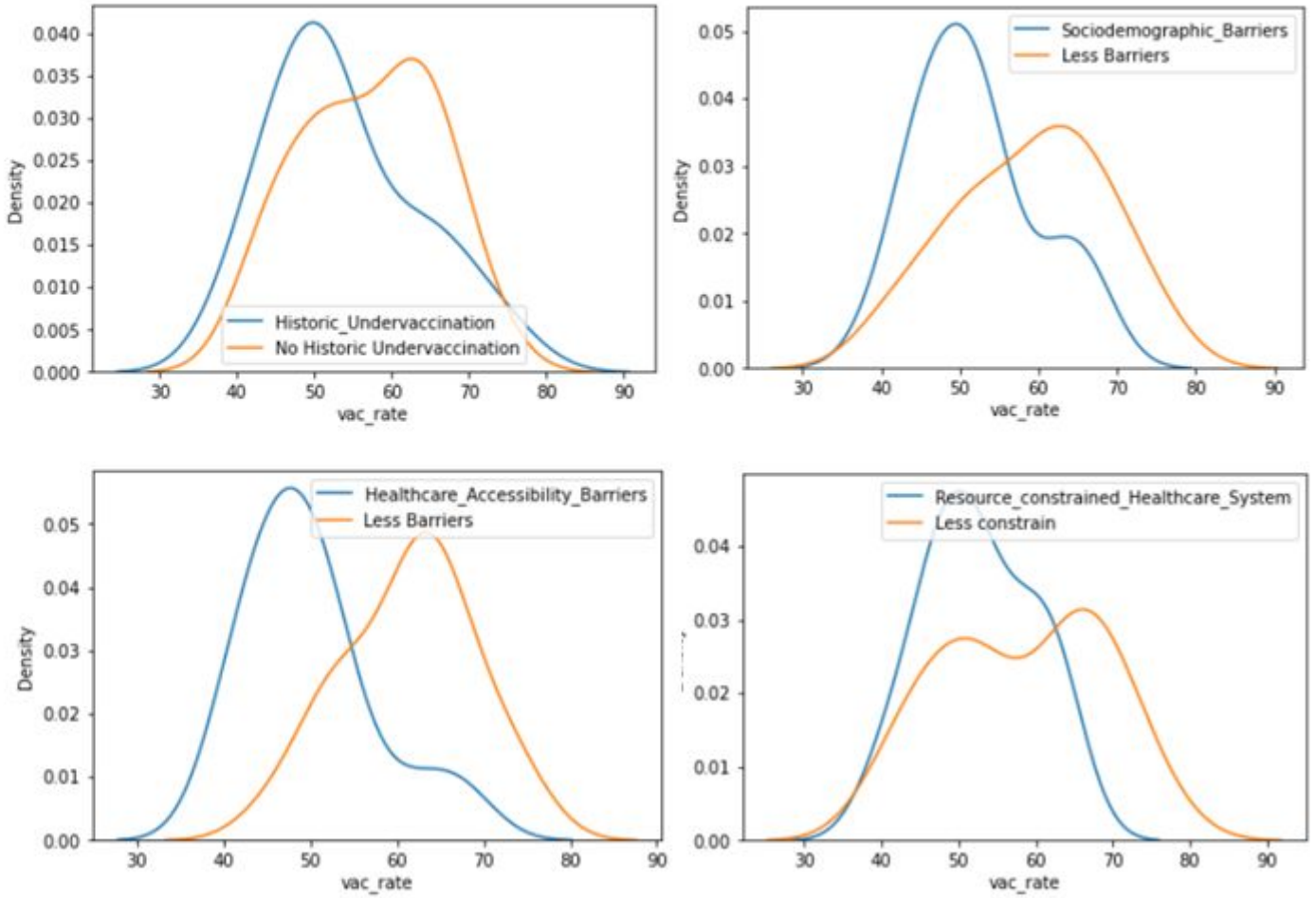
- Vaccination rate is more of a concern for states that:
  - Have over all poor coverage concerns from the medical facilities and historic vaccination concerns
  - Religious
  - Right leaning in political belief
  - More people said they are hesitate to get vaccine (so obvious here almost redundant!)



SOME MEDICAL CONCERNS

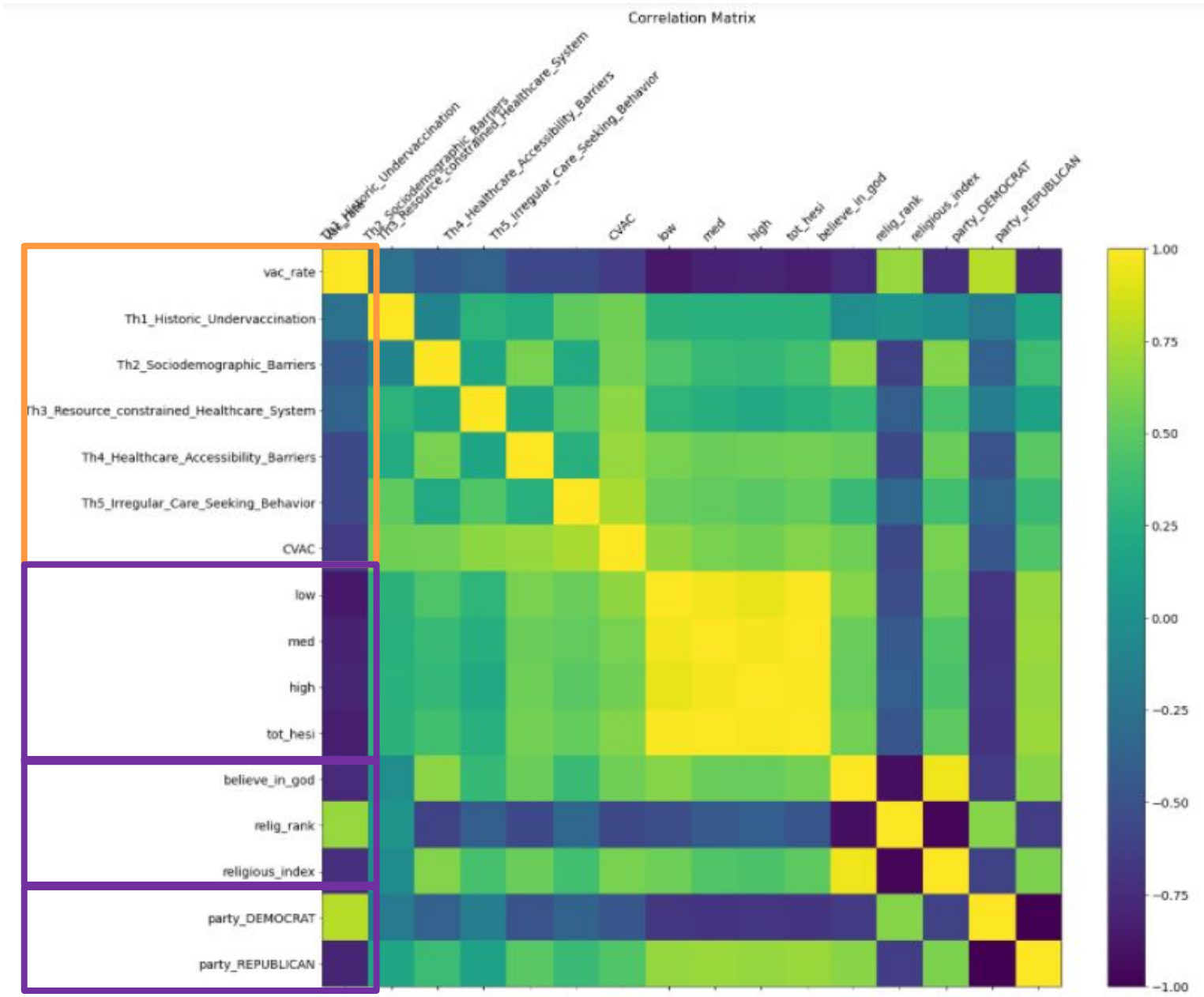
Some medical themes are more of a concern than others

- Less of a concern
  - Historic under vaccination
  - Sociodemographic barrier
  - Resource constraints
- More of a concern
  - Accessibility
  - Irregular care seeking behavior



CORRELATION

- 2 main predictor groups:
  - Medical themes
  - Behavior themes
    - Hesitation
    - Religion
    - Political belief



DATA MODELING AT STATE LEVEL AND COUNTY LEVEL

LINEAR REGRESSION

Use only medical themes

OLS Regression Results

Dep. Variable:	vac_rate	R-squared:	0.424
Model:	OLS	Adj. R-squared:	0.412
Method:	Least Squares	F-statistic:	[illegible]
Date:	Thu, 05 Aug 2021	Prob (F-statistic):	3.00e-07
Time:	10:03:23	Log-Likelihood:	-167.89
No. Observations:	50	AIC:	339.8
Df Residuals:	48	BIC:	343.6
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
[illegible]	65.8749	2.012	32.736	0.000	61.829	69.921
CVAC	-20.4754	3.442	-5.949	0.000	-27.396	-13.555

Omnibus:	0.010	Durbin-Watson:	1.527
Prob(Omnibus):	0.995	Jarque-Bera (JB):	0.157
Skew:	-0.019	Prob(JB):	0.925
Kurtosis:	2.729	Cond. No.	4.37

OLS Regression Results

Dep. Variable:	vac_rate	R-squared:	0.866
Model:	OLS	Adj. R-squared:	0.840
Method:	Least Squares	F-statistic:	33.12
Date:	Sun, 08 Aug 2021	Prob (F-statistic):	[illegible]
Time:	17:51:21	Log-Likelihood:	-131.45
No. Observations:	50	AIC:	280.9
Df Residuals:	41	BIC:	298.1
Df Model:	8		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	87.6989	6.147	14.267	0.000	75.285	100.113
Q("Th1_Historic_Undervaccination")	0.0579	2.374	0.024	0.981	-4.737	4.853
Q("Th2_Sociodemographic_Barriers")	4.4276	2.624	1.687	0.099	-0.871	9.727
Q("Th3_Resource_constrained_Healthcare_System")	-1.6051	2.131	-0.753	0.456	-5.909	2.699
Q("Th4_Healthcare_Accessibility_Barriers")	-1.3781	2.664	-0.517	0.608	-6.759	4.003
Q("Th5_Irregular_Care_Seeking_Behavior")	-3.8546	2.526	-1.526	0.135	-8.956	1.247
Q("tot_hesi")	26.0011	6.305	-4.124	0.000	-38.735	-13.267
Q("party_DEMOCRAT")	4.3877	1.652	2.656	0.011	1.051	7.725
Q("believe_in_god")	37.7872	9.715	-3.889	0.000	-57.408	-18.166

[illegible]	[illegible]	Durbin-Watson:	2.128
Prob(Omnibus):	0.721	Jarque-Bera (JB):	0.760
Skew:	-0.159	Prob(JB):	0.684
Kurtosis:	2.486	Cond. No.	38.3

All themes, R2=84%

Add ‘hesitation, R2 increases…

OLS Regression Results

Dep. Variable:	vac_rate	R-squared:	0.723
Model:	OLS	Adj. R-squared:	0.712
Method:	Least Squares	F-statistic:	61.45
Date:	Thu, 05 Aug 2021	Prob (F-statistic):	7.67e-14
Time:	10:03:44	Log-Likelihood:	-149.57
No. Observations:	50	AIC:	305.1
Df Residuals:	47	BIC:	310.9
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	75.4866	1.951	38.693	0.000	71.562	79.411
CVAC	-6.5703	3.102	-2.118	0.039	-12.810	-0.330
tot_hesi	-46.6893	6.551	-7.127	0.000	-59.868	-33.511

Omnibus:	1.643	Durbin-Watson:	1.975
Prob(Omnibus):	0.440	Jarque-Bera (JB):	1.144
Skew:	-0.035	Prob(JB):	0.564
Kurtosis:	2.262	Cond. No.	11.8

OLS Regression Results

Dep. Variable:	vac_rate	R-squared:	0.799
Model:	OLS	Adj. R-squared:	0.786
Method:	Least Squares	F-statistic:	[illegible]
Date:	Thu, 05 Aug 2021	Prob (F-statistic):	4.53e-16
Time:	15:31:25	Log-Likelihood:	-141.56
No. Observations:	50	AIC:	291.1
Df Residuals:	46	BIC:	298.8
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	65.5740	2.911	22.524	0.000	59.714	71.434
CVAC	-6.1621	2.673	-2.305	0.026	-11.542	-0.782
tot_hesi	-29.3764	7.005	-4.194	0.000	-43.476	-15.276
party_DEMOCRAT	7.0568	1.693	4.169	0.000	3.650	10.464

Omnibus:	0.849	Durbin-Watson:	2.543
Prob(Omnibus):	0.654	Jarque-Bera (JB):	0.891
Skew:	0.174	Prob(JB):	0.640
Kurtosis:	2.446	Cond. No.	16.1

Add political themes R2 increases further

LINEAR MODEL AT COUNTY LEVEL

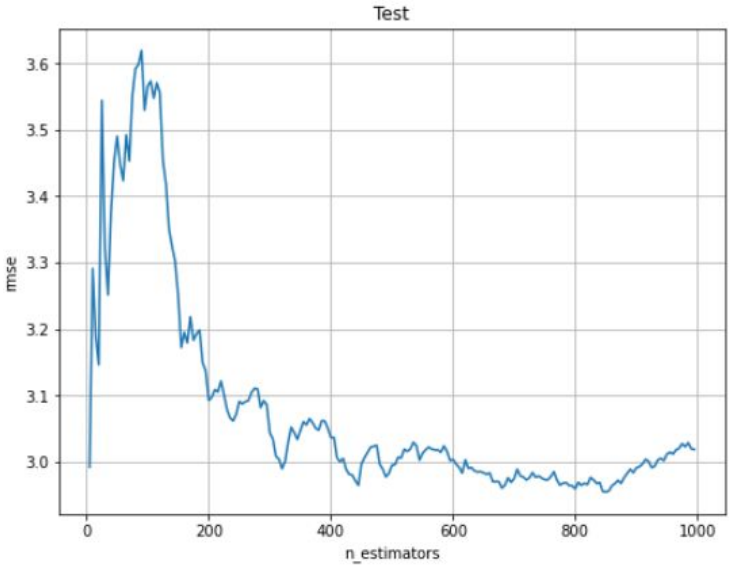
OLS Regression Results

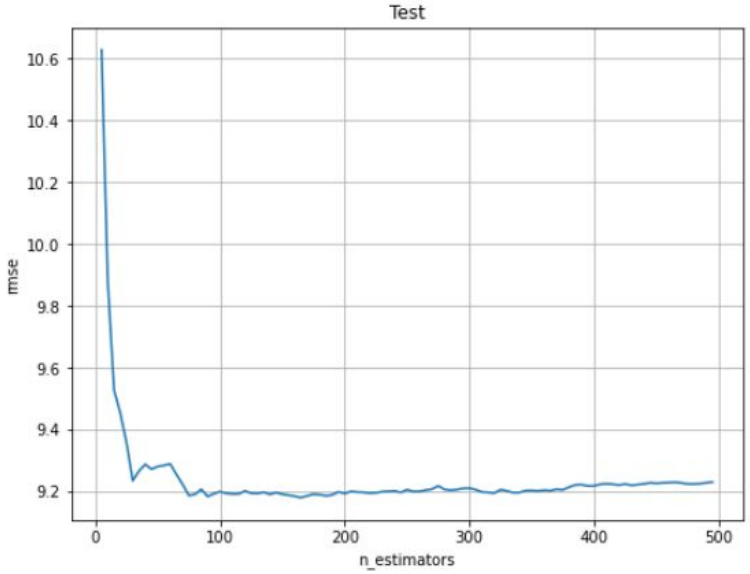
Dep. Variable:	vac_rate	R-squared:	0.304
Model:	OLS	Adj. R-squared:	0.302
Method:	Least Squares	F-statistic:	149.1
Date:	Sun, 08 Aug 2021	Prob (F-statistic):	1.24e-208
Time:	22:20:07	Log-Likelihood:	-10709.
No. Observations:	2744	AIC:	2.144e+04
Df Residuals:	2735	BIC:	2.149e+04
Df Model:	8		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	71.9380	2.160	33.312	0.000	67.704	76.173
Q("Th1_Historic_Undervaccination")	3.1128	0.897	3.469	0.001	1.353	4.872
Q("Th2_Sociodemographic_Barriers")	-6.4606	1.164	-5.553	0.000	-8.742	-4.179
Q("Th3_Resource_constrained_Healthcare_System")	-8.8037	0.897	-9.812	0.000	-10.563	-7.044
Q("Th4_Healthcare_Accessibility_Barriers")	-2.9567	1.285	-2.300	0.022	-5.477	-0.436
Q("Th5_Irregular_Care_Seeking_Behavior")	-3.6186	0.982	-3.683	0.000	-5.545	-1.692
Q("tot_hesi")	14.9817	2.310	6.487	0.000	10.453	19.510
Q("party")	12.0066	0.648	18.524	0.000	10.736	13.278
Q("believe_in_god")	-47.2805	3.851	-12.277	0.000	-54.832	-39.729

Omnibus:	234.423	Durbin-Watson:	1.119
Prob(Omnibus):	0.000	Jarque-Bera (JB):	419.064
Skew:	-0.599	Prob(JB):	1.00e-91
Kurtosis:	4.493	Cond. No.	33.1

RANDOM FORESTS





Parameters were optimized for model

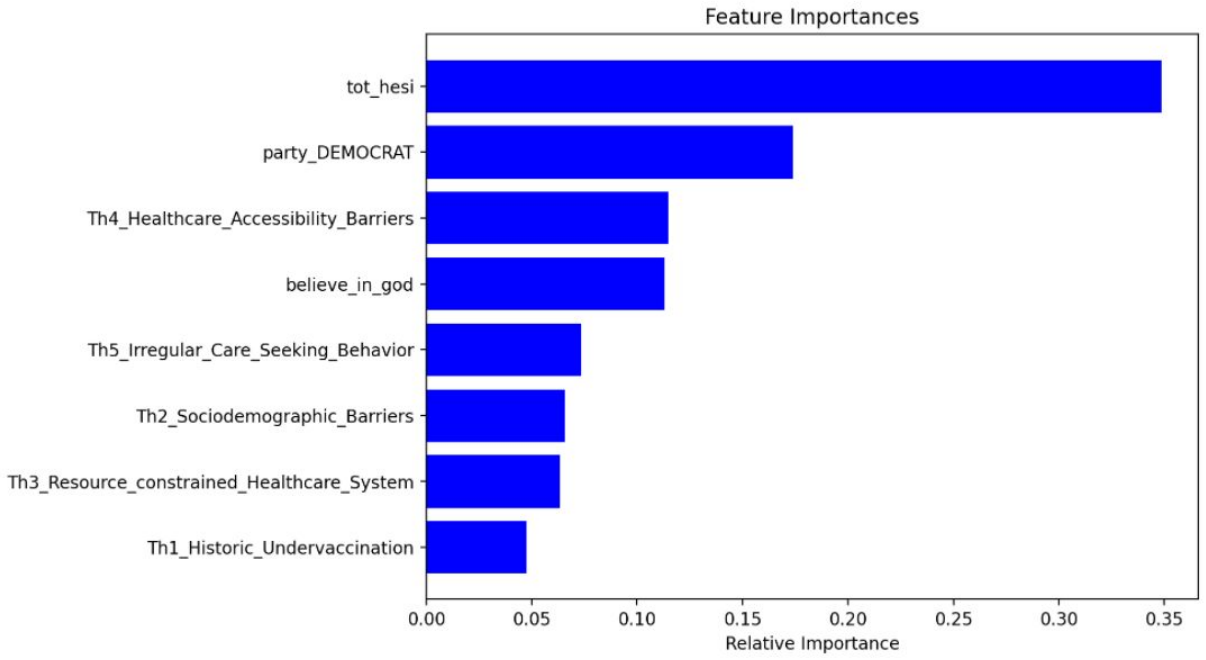
- n_estimators=850
- max_depth=6
- max_leaf_nodes=17

Parameters were optimized for model

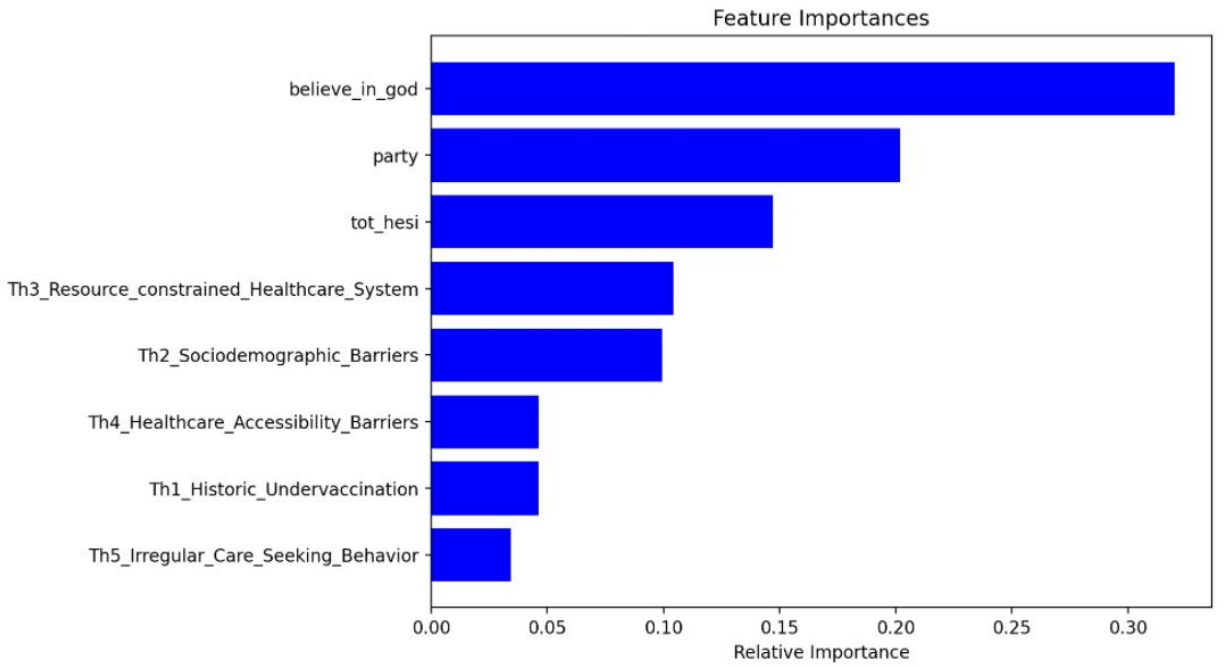
- n_estimators=165
- max_depth=25
- max_leaf_nodes=8

FEATURE IMPORTANCE

State Level (~50 samples)



County Level (~3100 samples)



Among all the features, religion, political views and hesitance are the most important factors related to vaccination rate

CONCLUSION

- Linear Regression and Random Forests were used to predict the vaccination based on 10 features

- Behavior related predictors are consistently more reliable than the CDC medical theme to predict the vaccination outcome

- These predictors include:
    - Hesitation
    - Religion
    - Political affiliation

Thank You!