Team 27
Data Science For All / Women's Summit Fall 2021
Correlation One

Understanding the factors influencing COVID-19 vaccination uptake rates in the US

Davina Mellows
davina.mellows@warwick.ac.uk

Francesca Iovu
iovufrancesca11@gmail.com

Merve Bektas
merve.bektas@redbull.com

Rali Dimitrova
ralicavdimitrova@gmail.com

Tanya Poppe
tanyapoppe@gmail.com

Varsha Ramineni
varsharamineni@gmail.com

Abstract

COVID-19 vaccines have been proven to minimise morbidity, mortality and thereby the ongoing global devastation experienced throughout the pandemic. Vaccine hesitancy is a barrier to global health, protection of vulnerable populations and to pandemic recovery. This project assessed demographic, socioeconomic, health, political and pandemic impact factors in US counties, also their ability to predict the proportion of the population over 12 years of age who were fully vaccinated. Political climate had the greatest predictive value, where Republican counties were at the greatest risk of low vaccination rates. Other predictive factors include estimated vaccine hesitancy and concern, the proportion of non-Hispanic Black population and a high pandemic vulnerability index, which increase the risk of severe COVID-19 infection and death. We conclude that community engagement with trusted leaders is essential for vaccine acceptance and in order to maximise vaccine uptake, public health information should be decoupled from political beliefs or interests. Resources should be directed to communities who are more likely to be vaccine hesitant due to historical mistreatment by the health system in the US.

Contents

1 Introduction

1.1 Problem Overview

The COVID-19 pandemic has had devastating effects on people all around the world, with over 242 million cases and almost 5 million deaths [8]. The impact of COVID-19 infections and the pandemic response on both quality of life and economies [2, 13] motivated concentrated efforts to develop effective vaccines to ameliorate disease burden. Global access to COVID-19 vaccines through effective and equitable distribution must be a key policy priority to achieve herd immunity and end the global pandemic [19]. A key determinant of successful vaccine deployment is the trust in offered vaccines and the institutions which approve and administer them. Timely vaccine deployment is essential to minimise the risk of more infectious and deadly variants which threaten to slow the global fight against COVID-19 [16].

1.2 Specific Issue

Vaccine hesitancy is a primary challenge in the deployment of COVID-19 vaccines and poses one of the most important public health issues in 2020/2021 [7]. America's continuing scepticism of COVID-19 vaccines now makes it an outlier among other developed western countries [6] where US media influence may affect global vaccine hesitancy. The hesitancy present in the US has complex drivers which are not well understood, especially among underserved segments of the population [9]. This project focused on the factors which affected US county vaccination rates.

1.3 Our Approach

We aimed to identify factors which were predictive of COVID-19 vaccination rates. We examined demographic, socioeconomic, health, COVID-19 severity, and political factors of US counties. We modelled how these predicted the proportion of the population over 12 years of age who were fully vaccinated.

2 Data

All datasets required details to be available at the US County level. The most recent data available for each source were taken. Where the same data was duplicated in multiple data-sets, we prioritised the original source over data included in any secondary analysis eg. a vulnerability index calculated from other data. Our main data sources include the Centre for Disease Control (CDC), 2019 US Census and various survey data. For further details on our data sources refer to the Appendix A and our shared GitHub repository [5].

2.1 Centres for Disease Control

Data were downloaded on the 4th of October 2021 and includes estimates until the 3rd of October.

2.1.1 Target features of interest

- % Population over 12 years of age (yo+) fully vaccinated

- % Population over 18 years of age fully vaccinated
- % Population over 65 years of age fully vaccinated

2.1.2 COVID-19 severity

- COVID-19 Cases (total number until 3rd of October 2020)
- COVID-19 Deaths (total number until 3rd of October 2020)

2.1.3 Social Vulnerability

Social Vulnerability Index (SVI) incorporates a number of factors including economic data, education, family characteristics, housing language ability, ethnicity and vehicle access.

- SVI (continuous): from 0 (least vulnerable) to 1 (most vulnerable)
- SVI (category): Very Low (0.0-0.19), Low (0.20-0.39); Moderate (0.40-0.59); High (0.60-0.79); Very High (0.80-1.0).

2.1.4 Level of concern for vaccination rollout

The Surgo Covid-19 Vaccine Coverage Index (CVAC) captures supply and demand related challenges that may hinder rapid, widespread COVID-19 vaccine coverage in US counties. This was done through five specific themes: historic undervaccination, sociodemographic barriers, resource-constrained healthcare system, healthcare accessibility barriers, and irregular care-seeking behaviors.

- CVAC level of concern for vaccination rollout (continuous): from 0 (lowest concern) to 1 (highest concern).
- CVAC level of concern for vaccination rollout (category): Very Low (0.0-0.19); Low (0.20-0.39); Moderate (0.40-0.59); High (0.60-0.79); Very High (0.80-1.0).

2.1.5 Vaccination Hesitancy

The CDC estimates of vaccination hesitancy were estimated using the the U.S. Census Bureau's Household Pulse Survey and the Census Bureau's 2019 American Community Survey 1-year Public Use Microdata Sample. These include the question: *"Once a vaccine to prevent Covid-19 is available to you, would you get a vaccine?"* and given the following options: 1) "Definitely get a vaccine"; 2) "Probably get a vaccine"; 3) "Probably not get a vaccine"; 4) "Definitely not get a vaccine."

- Estimated hesitant: Estimate of percentage of adults who describe themselves as "Probably not" or "Definitely not" going to get a COVID-19 vaccine once one is available to them
- Estimated hesitant or unsure: Estimate of percentage of adults who describe themselves as "Unsure", "Probably not", or "Definitely not" going to get a COVID-19 vaccine once one is available to them
- Estimated strongly hesitant: Estimate of percentage of adults who describe themselves as "Definitely not" going to get a COVID-19 vaccine once one is available to them

2.1.6 Underlying medical conditions associated with high risk of severe COVID-19

- % Obesity
- % Cardiovascular disease
- % Chronic obstructive pulmonary disorder (COPD)
- % Chronic kidney disease (CKD)
- % Diabetes
- % Any of the above
- % Smoking

2.2 Texas Department of State health Services

- % Population over 12 years of age fully vaccinated
- % Population over 65 years of age fully vaccinated

2.3 Area Health Resources Files

- % Individuals 18-64 years with no health insurance (estimates from 2017)
- % Education, Health Care, Social Assistance Workers (estimates 2010-2017)
- Average household size and type (estimates 2010)
- % Individuals 25+ with no high school diploma (estimates 2017)
- % Individuals living in poverty (estimates 2013-2017)
- Sex (% male)

2.4 US Census 2019

- Population density
- Urban-rural code - six levels of the metropolitan to rural distribution of counties which holds information about the access to care often centralised in large metropoles.

2.5 American Community Survey 2019

- % Hispanic
- % Non-Hispanic American Indian/Alaska Native
- % Non-Hispanic Black
- % Non-Hispanic Native Hawaiian/Pacific Islander
- % Non-Hispanic White

2.6 Election 2020

- % Votes for Democratic candidate
- % Votes for Republican candidate
- % Point difference between Democratic and Republican candidate
- % Population that voted

2.7 Pandemic Vulnerability

The Pandemic Vulnerability Index (PVI) is derived as a linear combination of 20 factors from four major domains: Infection Rate, Population Concentration, Intervention Measures, and Health & Environment. This is a score that measures a county's current vulnerability.

- PVI (continuous): from 0 (lowest venerability) and 1 (highest vulnerability). Updated daily, estimate take for the 3rd of October.
- Premature Death - Years of potential life lost before age 75 per 100,000 population (age-adjusted) based on 2016-2018 National Centre for Health Statistics - Mortality Files. This is a broad measure of health and conditions that have been associated with more severe outcomes from COVID-19 infection.
- Testing - Population divided by tests performed with greater numbers indicating lower testing rate. Lower testing rate expected to result in increased undetected infection (Updated daily, estimate take for the 3rd of October).
- Traffic - Average traffic volume per meter of major roadways in the county from 2018 EPA EJSCREEN, with greater traffic volume expected to increase the spread of infection.
- Day Time Population Density - Estimated daytime population with greater daytime expected to increase the spread of infection.

3 Data preprocessing

3.1 Vaccination rate

To provide fuller coverage of the US counties, we merged the CDC vaccination data and the Texas vaccination data. Data on the percent of US population older than 12 were not available for Idaho, Hawaii and several counties in California; data on vaccination rate of US population older than 18 are not recorded for Texas, Hawaii and several counties in California; data on vaccination rate of US population older than 65 were not available for Hawaii and several counties in California. Vaccination data were also available for Puerto Rico and Guam. Vaccination rates for the three groupings of the population are depicted in Figure 1.

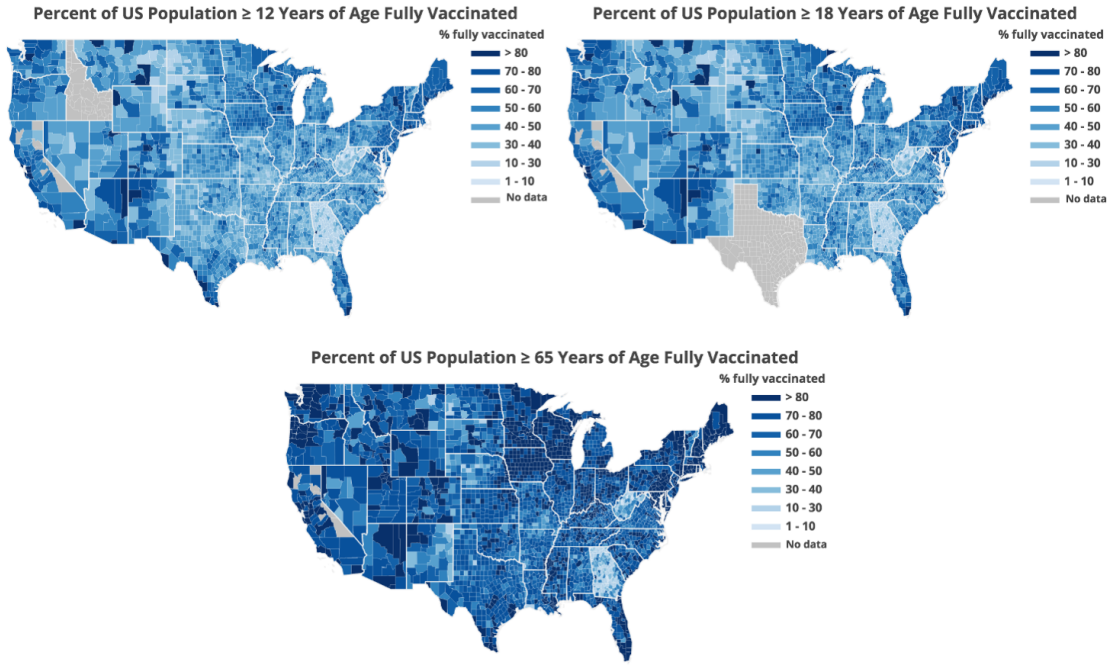

Figure 1: Percent of the US population fully vaccinated. In grey, counties with missing data. Darker blue indicates higher vaccination rate. Sources: CDC Covid-19 tracker and Texas Gov.

3.2 COVID-19 cases and deaths

When examining the number of confirmed COVID-19 cases and deaths, we noticed that the rates in several counties were 0 (primarily counties in Utah). Given the length of the pandemic and the infection rate of COVID-19, it is highly unlikely that there were 0 cases since March 2020. Therefore, these values were assigned to missing. While death rates of 0 are more plausible, we decided to also assign these values to missing, following the logic that it is highly likely that information for both cases and death were not provided.

Data on COVID-19 cases and deaths were available as a number per county, with larger counties (e.g. Los Angeles) having a significantly larger case load. We therefore calculated the case/death proportion for each county using the 2019 US Census population estimates (latest available). COVID-19 cases and deaths as a proportion of the population on a county level are visualised in Figure 2.

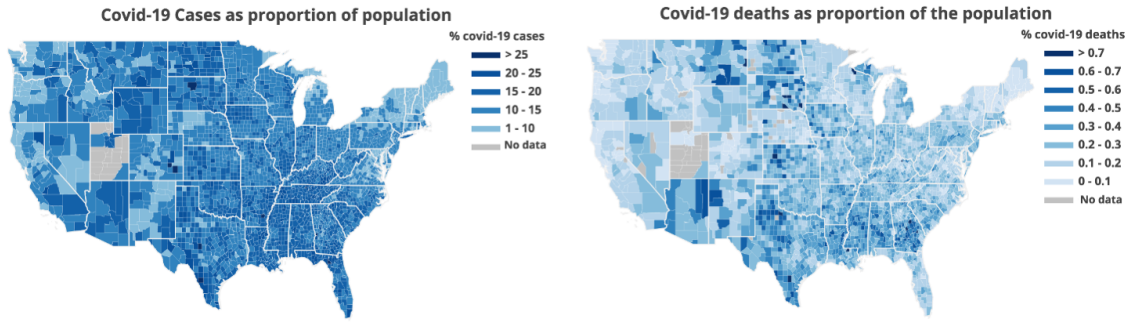

Figure 2: COVID-19 cases and deaths as proportion of the population in the US. In grey, counties with missing data. Darker blue indicates higher rates. Sources: CDC COVID-19 tracker.

3.3 Dataset merging

Datasets were merged using the FIPS codes for each county, available as a variable in each dataset.

4 Exploratory data analysis

We first explored the distribution of all features and their association (Spearman's ρ) with our target variables (fully vaccinated individuals aged 12yo+, 18yo+, 65yo+). Then, we selected features of interest and dealt with missing values.

4.1 COVID-19 cases and death rate

We observed a moderate negative association between vaccination rate (12yo+, 18yo+, 65yo+) and confirmed COVID-19 cases and deaths per county ($-0.09 < \rho < -0.30$), with highest correlation estimates for the 18yo+ group (cases: $\rho = -0.23$; death: $\rho = -0.29$) and the 12yo+ group (cases $\rho = -0.21$; death $\rho = -0.30$). This suggests that in counties with high vaccination rate, there were lower COVID-19 cases and death.

4.2 Socio-economic factors

Counties with a lower proportion of individuals without health insurance had a higher proportion of fully vaccinated 12yo+ ($\rho = -0.38$), 18yo+ ($\rho = -0.40$) and 65yo+ ($\rho = -0.40$) (Figure 3). The proportion of fully vaccinated was overall lower in counties where there was a higher rate of individuals without a high school diploma (12yo+ $\rho = -0.37$; 18yo+ $\rho = -0.39$; 65yo+ $\rho = -0.37$) (Figure 3). Furthermore, counties with a high % of the population living in poverty had lower rates of 12yo+ ($\rho = -0.35$), 18yo+ ($\rho = -0.37$) and 65yo+ fully vaccinated ($\rho = -0.33$) (Figure 3). Vaccination rates were also higher in more populous counties (12yo+ $\rho = 0.32$; 18yo+ $\rho = 0.30$; 65yo+ $\rho = 0.34$). We found no association between vaccination rates and household size or type.

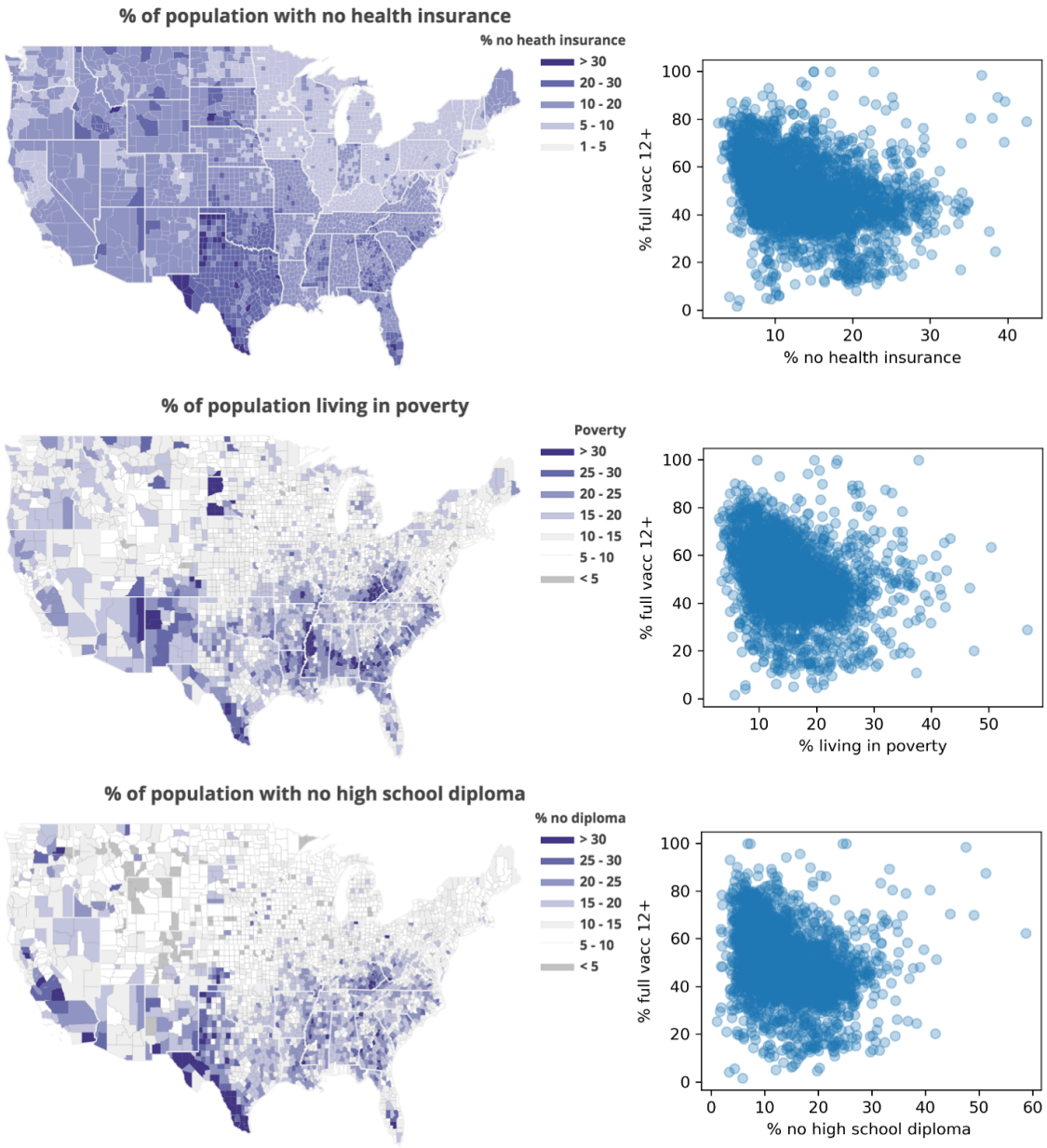

Figure 3: Proportion of population with no health insurance (top), living in poverty (mid), with no high school diploma (bottom). Darker colour indicate higher proportion. Association between these variables and proportion of fully vaccinated individuals aged 12 and over are also shown. Sources: HRSA and COVID-Health-Disparities dataset.

Overall, there was a negative correlation between Social Vulnerability index (SVI) and the

proportion of fully vaccinated individuals, $\rho = -0.20$, $\rho = -0.22$, $\rho = -0.22$ for 12yo+, 18yo+ and 65yo+ groups, respectively (Figure 4).

4.3 Underlying medical conditions

The prevalence of underlying medical conditions ranged between counties from 22% to 66% and the most common condition was obesity at 35 ± 4.5% (mean ± standard deviation). The prevalence of any condition was highly correlated with the prevalence of individual underlying medical conditions: obesity ($\rho = 0.89$); diabetes ($\rho = 0.84$); heart disease ($\rho = 0.78$); chronic obstructive pulmonary disorder ($\rho = 0.81$); chronic kidney disease ($\rho = 0.73$). We observed moderate negative correlations between the prevalence of any underlying conditions and full vaccination rates (12yo+ $\rho = -0.40$; 18yo+ $\rho = -0.39$; 65yo+ $\rho = -0.29$). Surprisingly, this showed that counties with higher incidences of underlying medical conditions had lower vaccination rates.

4.4 Vaccination Hesitancy

Estimated vaccination hesitancy showed a moderate negative correlation with the vaccine rates (Figure 4). For instance the Spearman ρ of the 'Hesitant' or 'Unsure' with the proportion of fully vaccinated individuals aged 12yo+, 18yo+ and 65yo+ were -0.37, -0.37 and -0.25, respectively. The levels of concern for vaccination roll out (CVAC) also showed a negative association with the vaccination rate (12yo+: $\rho = -0.38$; 18yo+: $\rho = -0.40$; 65yo+: $\rho = -0.36$) with highest rates concentrated in the southern and western regions of the US (Figure 4).

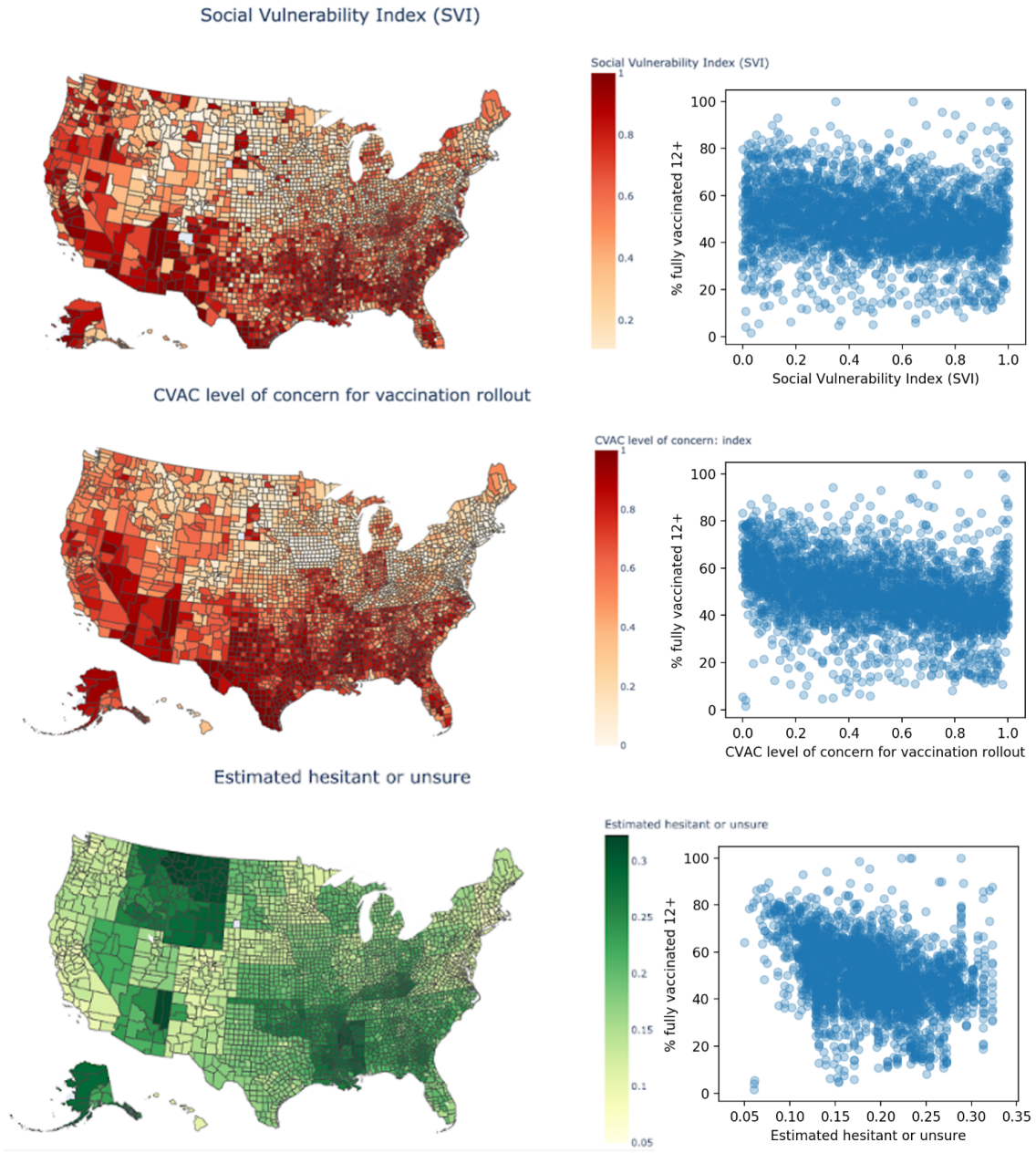

Figure 4: Social Vulnerability, Level of Concern over vaccination roll out and Vaccination hesitancy across the US. Scatter plots depicting the association between these features and the proportion of fully vaccinated individuals aged 12 and older are also shown. Sources: CDC vaccination hesitancy dataset.

4.5 Pandemic Vulnerability Index

We observed a negative association between PVI and proportion of fully vaccinated individuals (12yo+: $\rho = -0.21$; 18yo+: $\rho = -0.26$; 65yo+: $\rho = -0.11$), suggesting that there is lower vaccination rates in counties with higher PVI.

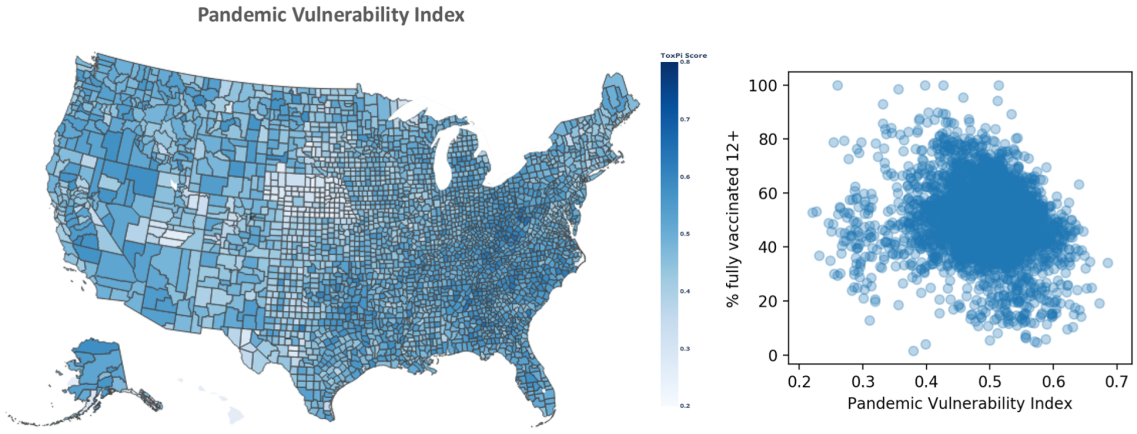

Figure 5: Pandemic Vulnerability Index across the US. Darker blue indicates higher pandemic vulnerability index. Sources: COVID-19 Pandemic Vulnerability Index (PVI).

4.6 Election Results 2020

We observed a strong association between elections results and vaccination rates 6. The proportion of the population above age 12 that were fully vaccinated showed a negative correlation with % republican votes ($\rho = -0.56$) and a positive correlation with the % democratic votes ($\rho = 0.56$).

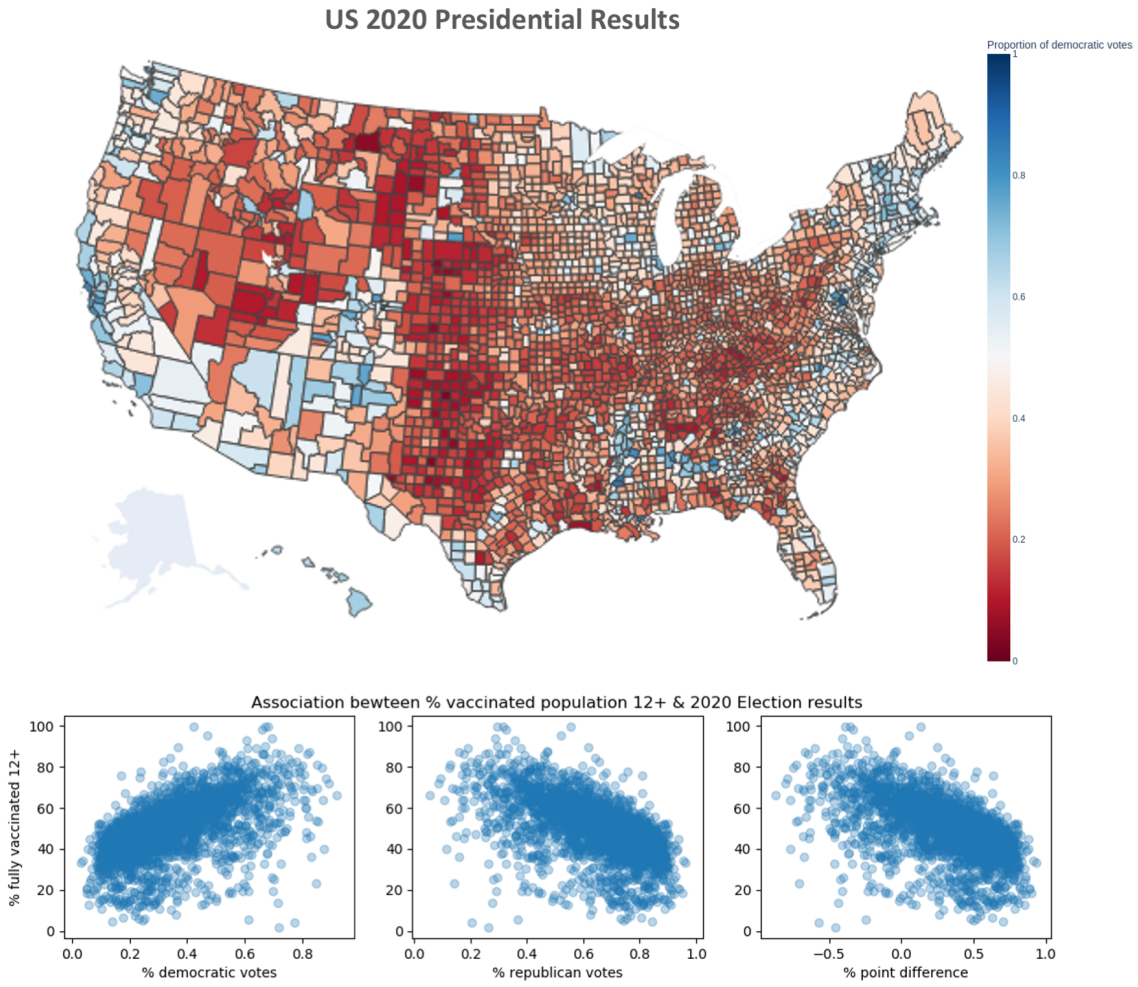

Figure 6: 2020 Presidential election results across the US. Blue depicts higher % democratic votes, red depicts higher % republican votes. Sources: New York Times, Fox News, Politico.

4.7 Feature Selection and Exclusion

Our target variable was chosen to be the proportion of fully vaccinated individuals older than 12. While we had more complete information about vaccination rates in the 65yo+ population (Figure 1), we chose to look at 12yo+ as it covers a bigger proportion of the population, had more complete data than the 18yo+ category and we observed the same association between these two variables and the rest of the features.

We also decided to exclude counties in Puerto Rico, Guam and Alaska, as many features of interest were not available for these states/territories, including but not limited to SVI, hesitancy and 2020 election results.

Features excluded due to no association with our target variable included:

- Household type
- Household size
- Education, Health Care, Social Assistance Workers (%)
- Sex (% male)
- % Population over 65 years of age

Other variables were decided as redundant due to very high correlations ($\rho > 0.8$) with other features. These include:

- Estimated hesitant and Estimated strongly hesitant were dropped in favour of Estimated hesitant or unsure
- % Point difference and % GOP were dropped in favour of % DEM and % voted
- % Obesity, % heart disease, % COPD, % diabetes, % CKD in favour of % any of these underlying conditions

The following two categorical features were dropped in favour of their continuous counterparts:

- SVI Category
- CVAC level of concern for vaccination rollout Category

Figure 7 shows a heat-map of the correlations between the target variable and all features.

4.8 Missing data imputation

Following feature selection, we explored for missing data in our final dataset. SVI information was missing for one county in New Mexico and this was populated with the mean SVI for New Mexico state. Data were missing for 24 counties for % Covid-19 cases, 47 counties for % death rate and 53 counties for Premature Death. Instead of removing these counties, we chose to impute them using Random Forest Regression. For each of these features, we split the data into training (counties with available data) and test (counties with missing data) sets. We trained the model on the training data using features, which showed a correlation of $\rho > 0.3$ with the target variable.

Our final dataset included 3053 counties in the mainland US with available data on the proportion of fully vaccinated individuals aged 12+ and 24 continuous features (Figure 7)

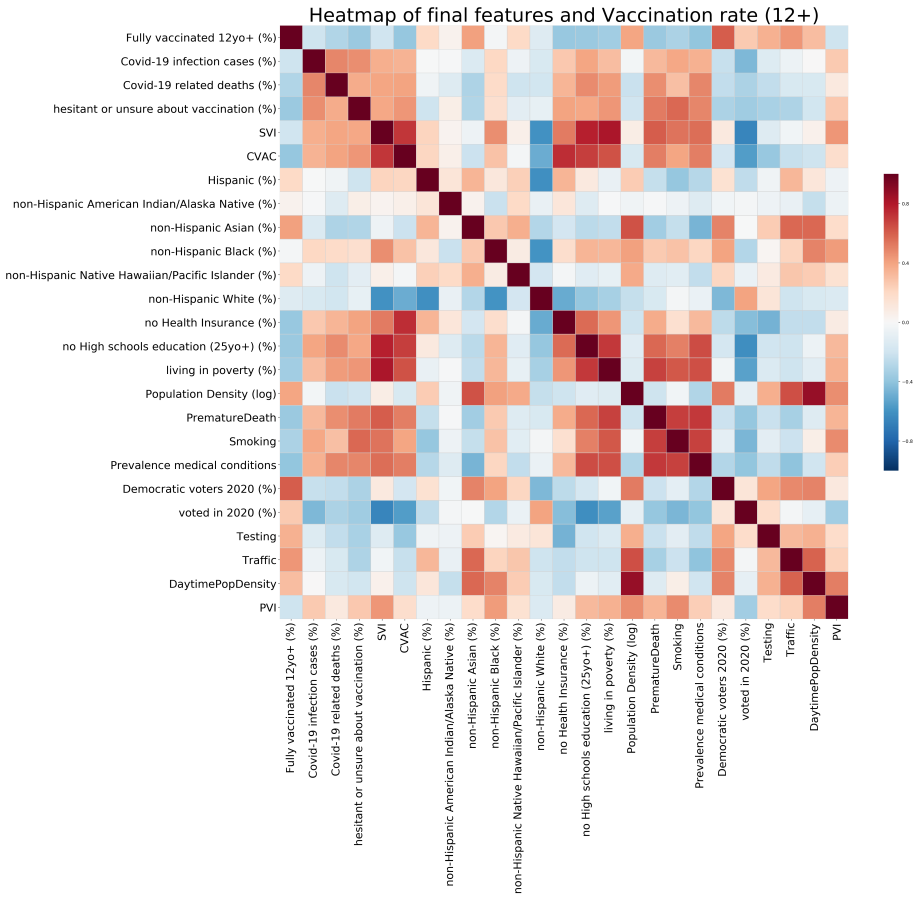

Figure 7: Heatmap of the correlations (Spearman ρ) between % fully vaccinated individuals 12 or older and all features in the final dataset. Positive associations are shown in red, while negative in blue.

5 Modelling

5.1 Train, validation and test sets

To tune models' hyperparameters and assess models' generalisation to unseen data, we split the data into train, test and validation samples. 20% of the data, or 611 counties, were set aside as a hold-out test sample. Data of the rest of the 2442 counties, were split into training (80%, 1953 counties) and validation (20%, 489 counties) samples.

5.2 Model selection

To predict vaccination rate in the US, we explored several possible models. These included Regularised Ridge Regression (`Ridge()`) and the following decision tree-based ensemble methods:

Random Forest (`RandomForestRegressor()`), Extra Tree (`ExtraTreeClassifier()`), AdaBoost (`AdaBoostRegressor()`) and Extreme Gradient Boosting Regression (`XGBRegressor()`), as implemented in `scikit-learn` (XGBRegressor from `xgboost`).

We optimised the hyperparameters of the models using 5-fold Cross Validation (CV) on the training data. This was implemented with scikit-learn `GridSeachCV` using mean absolute error (MAE) as a loss function. Model performance was evaluated using the MAE, root mean squared error (RMSE) and Spearman's ρ between the observed and the predicted vaccination rates in the validation sample. Finally, the model with lowest MAE and RMSE was chosen and applied to the hold-out test data.

5.3 Model performance

Best prediction on the validation set was achieved using `ExtraTreesRegressor` (optimised parameters: n_estimators = 500, max_features = "auto", max_depth = 15, min_samples_leaf = 3) with MAE of 5.93% and RMSE of 8.60 % and correlation between observed and predicted vaccination rate of $\rho = 0.80$. `RandomForestRegressor` (n_estimators = 500, max_features = "sqrt", max_depth = 20, min_samples_leaf = 1) predicted vaccination rate with MAE of 6.09%, RMSE % 8.69 and $\rho = 0.79$; `XGBRegressor` (n_estimators = 500, learning_rate = 0.01, max_depth = 4, early_stopping_rounds = 4) with MAE of 6.38% and RMSE of 8.87 %, $\rho = 0.78$; `Ridge` (alpha = 11) with MAE of 6.76%, RMSE of 9.78% and $\rho = 0.75$. Worst performance was observed using `AdaBoostRegressor` (n_estimators = 1000, learning_rate = 0.01) with MAE of 7.29 %, RMSE of 10.01% and $\rho = 0.73$.

5.4 Final model

Once the best performing model was selected (`ExtraTreesRegressor`), we trained the model on the combined train and validation samples to then predict the test (hold-out) data. Vaccination rates were predicted with MAE of 6.17% and RMSE of 8.53%, with a correlation between observed and predicted rates of $\rho = 0.81$. As expected, higher errors were observed in counties with very low or very high vaccination rates.

6 Results

To identify the most influential variables predicting the vaccination rate, we explored the top ten most important features using `permutation_importance`. We see the decrease in the model score when a single feature value is randomly shuffled (Figure 8 Left). Consistent with our EDA, 2020 Election results, Pandemic Vulnerability Index, non-Hispanic Black (%), vaccine hesitancy, rates of health insurance and concern over vaccine roll-out were among the most important features for predicting the vaccination rate. Other features included non-Hispanic Black (%), non-Hispanic American Indian/Alaska Native (%), population density, Smoking and testing.

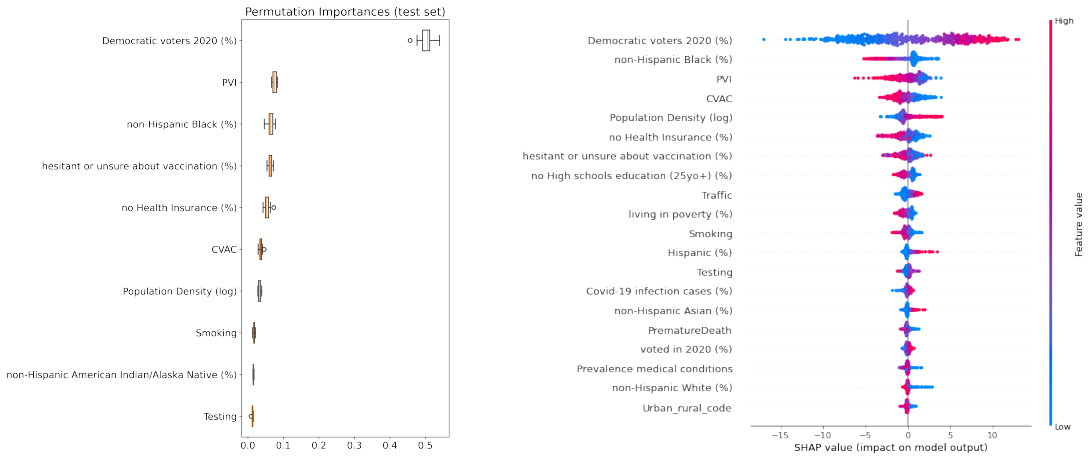

Figure 8: **Left:** Box-plots for the 10 features with highest permutation importance extracted from our final model on hold-out test set. **Right:** SHAP summary plot: This shows the SHAP values on the x-axis. Values on the left represent the observations that shift the predicted value in the negative direction while the points on the right contribute to shifting the prediction in a positive direction. All the features are on the left y-axis.

To better understand the predictions and explore the impact of each feature we examined the SHapley Additive exPlanations (SHAP Values). Figure 8 (Right) gives us a birds eye view of SHAP feature importance and what is driving it. We see that the feature related to the 2020 election results has the highest SHAP feature importance; low percentage of Democratic voters reduces the predicted vaccination rate whereas high percentage of Democratic voter increases the rate. Figure 9 shows a clear cut off point: if the percentage of democratic voters is above roughly 50% then this increases the predicted vaccination rate.

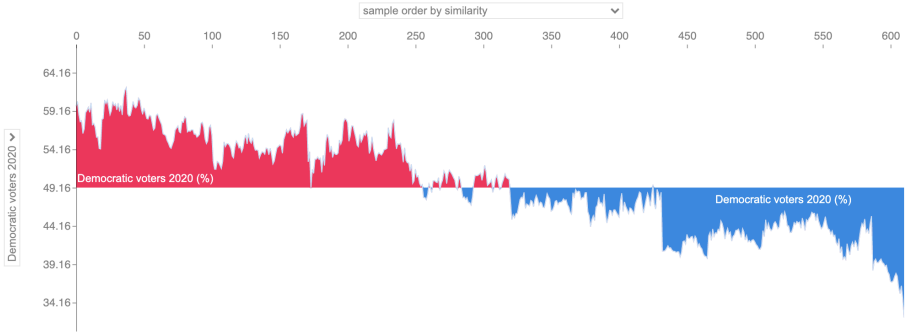

Figure 9: SHAP force plot: Each position on the x-axis is an instance of the data clustered by explanation similarity. The y axis is the variable % Democratic votes. Red SHAP values increase the prediction, blue values decrease it.

The partial dependence plots in Figure 10 show the marginal effect one or two features have on the predicted vaccination rate. We see an approximately linear and positive trend between the 2020 election results and the effect on predicted vaccination rate (SHAP Value). The variables Pandemic Vulnerability Index (PVI), Concern for vaccination roll-out and percentage non-Hispanic Black (%) show a negative trend. Figure 10 (Bottom left) shows that counties with low non-Hispanic Black population are more likely to have higher vaccination rates if they have high percentage of Democratic votes. Figure10 (Bottom right) shows that counties with high CVAC are more likely to have higher vaccination rates if they have a low non-Hispanic Black population.

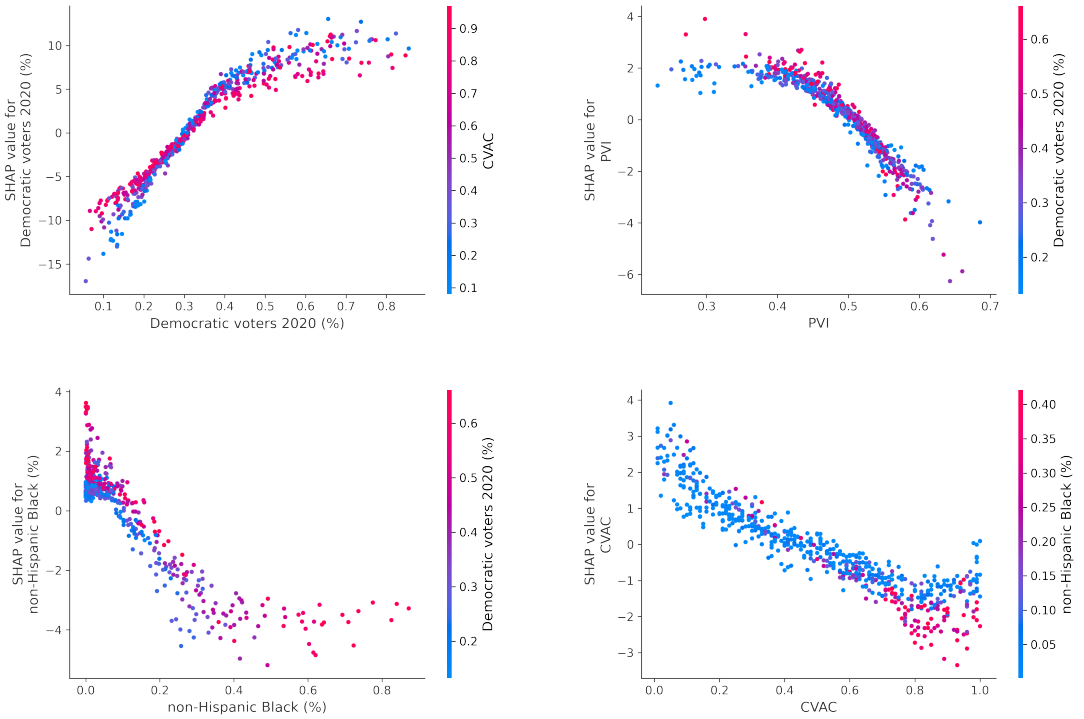

Figure 10: SHAP dependence plots: SHAP values plotted again variables Democratic votes 2020 (%), PVI, non-Hispanic Black (%) and CVAC. Each observation is colour coded by the another variable (that has greatest interaction with given variable) scale given on right hand side of each.

7 Conclusions

Vaccination roll-out in the US is influenced by multiple factors including socioeconomic, political climate and reported vaccination hesitancy and concern. Our analysis showed that the 2020 presidential election results were the dominant driver of COVID-19 vaccination rates in the US. Indeed, COVID-19 vaccines have been politicised and used for ideological interests [1]. Similarly, in France the acceptation of a COVID-19 vaccination was found dependent on voting at the first round of the 2017 presidential election [12]. This argues that politicisation of the COVID-19 vaccine undermines

people's confidence in them, and that attitudes towards the vaccine are integrated into the societal and cultural divide that political ideologies create. Trusted community leaders hold responsibility in public perception, particularly at times of fear associated with the pandemic. Indeed, Portugal, one of the countries with highest vaccination rate, [4] is a good example. While there existed initial hesitation about the COVID-19 vaccines and some misinformation circulating on social media, vaccination roll-out has been hugely successful due to the choice of leaders to "detach" the vaccination campaign from politics [17].

Our results suggest that the government should direct more resources to promote COVID-19 vaccination among African American and American Native communities. These communities have had disproportionately higher COVID-19 disease burden and COVID-19 related death rates [10, 3], with data indicating that African Americans are twice less likely to receive COVID-19 vaccination compared to non-Hispanic Whites [11]. The vaccination hesitancy in these communities is believed to stem from historical mistreatment by the health care system leading to a significant mistrust in these institutions [18]. While a range of strategies to address the vaccine hesitancy in these communities have been outlined [15], building policies based on these recommendations is yet to come. We also observed that the CVAC level of concern for vaccination roll-out measure, which represents historic under vaccination and socioeconomic barriers, was also an important predictor of the COVID-19 vaccination uptake. This suggests that the historical perceptions and barriers present in vulnerable communities continue to influence present and future public health projects and that targeted community engagement is necessary.

In early October 2021 (just after the start of this project) a number of states, cities and private companies across the US started to enforce COVID-19 vaccination mandates to ensure a safe workplace, with some companies offering incentives and other adopting the 'No jab, no job' policy [14]. While the mandate is an effective way to motivate some hesitant individuals to get vaccinated, it could also create a further divide between groups as well as between political parties. This could also increase the mistrust in the government, especially in groups that are already hesitant about vaccination due to political reasons or those groups that have been historically mistreated by the health care system, including African-Americans among other minority communities [15]. These measures will no doubt influence vaccine roll-out in the months to follow and should be considered in future models trying to understand which factors influence vaccine choice.

Comparing the perception and uptake of the COVID-19 vaccine to other widespread vaccination schemes (e.g. measles) would further identify what the barriers of vaccine uptake are unique to the current pandemic. This could inform future early public health and pandemic messaging. Furthermore, our analysis focused on the US, however conducting analysis for different countries may illuminate common barriers and perceptions, as well as compare the effectiveness of various vaccination campaigns across the globe to inform future policies.

To conclude, our work argues that in order to maximise vaccine uptake, leaders across the ideological and political spectrum should work together to promote vaccinations. Public health information should be decoupled from political beliefs or interests and instead attached to scientific knowledge and evidence. Resources should be directed to communities who are more likely to be vaccine hesitant due to historical mistreatment by the health system.

Appendix A: Data Sources

COVID 19 Vaccinations in the United States by County - CDC
COVID-19 Vaccine in Texas Dashboard - Texas Gov
List of US FIPS codes by county - Wikipedia
COVID-19 Data Repository by the Center for Systems Science and Engineering - Github
County Population Totals - US Census
COVID Health Disparity - Github
Prevalence of Selected Underlying Medical Conditions - CDC
Urban or Rural county - CDC
Vaccine Hesitancy for COVID 19 - CDC
US Election Results - Github
Pandemic Vulnerability - Github

References

[1] Ali Haif Abbas. Politicizing covid-19 vaccines in the press: A critical discourse analysis. *International Journal for the Semiotics of Law-Revue internationale de Sémiotique juridique*, pages 1–19, 2021.

[2] José Manuel Aburto, Jonas Schöley, Ilya Kashnitsky, Luyin Zhang, Charles Rahal, Trifon I Missov, Melinda C Mills, Jennifer B Dowd, and Ridhi Kashyap. Quantifying impacts of the covid-19 pandemic through life-expectancy losses: a population-level study of 29 countries. *International Journal of Epidemiology*, 2021.

[3] David Blumenthal, Elizabeth J Fowler, Melinda Abrams, and Sara R Collins. Covid-19—implications for the health care system, 2020.

[4] Our World In Data. Coronavirus (covid-19) vaccinations. https://ourworldindata.org/covid-vaccinations, 2021.

[5] Rali Dimitrova, Merve Bektas, Francesca Iovu, Davina Mellows, Tanya Poppe, and Varsha Ramineni. Team 27 github. https://github.com/ralidimitrova/DS4A_team27, 2021.

[6] Economist. America has remained unusually vaccine sceptical. *The Economist*, Sept 2021. URL https://www.economist.com/graphic-detail/2021/09/06/america-has-remained-unusually-vaccine-sceptical.

[7] Ariel Fridman, Rachel Gershon, and Ayelet Gneezy. Covid-19 and vaccine hesitancy: A longitudinal study. *PloS one*, 16(4):e0250123, 2021.

[8] World Health Organization: Geneva. Who covid-19 dashboard, 2020. URL https://covid19.who.int/.

[9] Sara Melotte and Mayank Kejriwal. Predicting zip code-level vaccine hesitancy in us metropolitan areas using machine learning models on public tweets. *arXiv preprint arXiv:2108.01699*, 2021.

[10] Nana-Yaa Misa, Berenice Perez, Kellie Basham, Essence Fisher-Hobson, Brittany Butler, Kolette King, Douglas AE White, and Erik S Anderson. Racial/ethnic disparities in covid-19 disease burden & mortality among emergency department patients in a safety net health system. *The American journal of emergency medicine*, 45:451–457, 2021.

[11] N Ndugga, O Pham, L Hill, S Artiga, and S Mengistu. Latest data on covid-19 vaccinations race/ethnicity. *Kais Family Found*, 2021.

[12] Patrick Peretti-Watel, Valérie Seror, Sébastien Cortaredona, Odile Launay, Jocelyn Raude, Pierrea Verger, Lisa Fressard, François Beck, Stéphane Legleye, Olivier l'Haridon, et al. A future vaccination campaign against covid-19 at risk of vaccine hesitancy and politicisation. *The Lancet Infectious Diseases*, 20(7):769–770, 2020.

[13] Julian Reif, Hanke Heun-Johnson, Bryan Tysinger, and Darius Lakdawalla. Measuring the covid-19 mortality burden in the united states: A microsimulation study. *Annals of internal medicine*, 2020.

[14] Reuters. U.s. workers face job losses as covid-19 vaccine mandates kick in. https://www.reuters.com/world/us/us-workers-face-layoffs-us-covid-19-vaccine-mandates-kick-2021-10-19/, 2021.

[15] EA Rusoja and BA Thomas. The covid-19 pandemic, black mistrust, and a path forward. *EClinicalMedicine*, 35, 2021.

[16] Pratha Sah, Thomas N Vilches, Seyed M Moghadas, Meagan C Fitzpatrick, Burton H Singer, Peter J Hotez, and Alison P Galvani. Accelerated vaccine rollout is imperative to mitigate highly transmissible covid-19 variants. *EClinicalMedicine*, 35:100865, 2021.

[17] New York Times. In portugal, there is virtually no one left to vaccinate. https://www.nytimes.com/2021/10/01/world/europe/portugal-vaccination-rate.html, 2021.

[18] David R Williams, Harold W Neighbors, and James S Jackson. Racial/ethnic discrimination and health: Findings from community studies. *American journal of public health*, 93(2):200–208, 2003.

[19] Olivier J Wouters, Kenneth C Shadlen, Maximilian Salcher-Konrad, Andrew J Pollard, Heidi J Larson, Yot Teerawattananon, and Mark Jit. Challenges in ensuring global access to covid-19 vaccines: production, affordability, allocation, and deployment. *The Lancet*, 2021.