COVID-19 Vaccine Accessibility & Discrimination

Spatial analysis on COVID-19 vaccine site availability and the demographic factors of US counties

WELLESLEY

JJ Sung ’22 Data Science Major Capstone

Research Questions

- Which regions in the United States have the highest/lowest accessibility to COVID-19 vaccines?
- What are the demographic factors (e.g. race, ethnicity, gender, income levels) that impact the vaccine accessibility of different counties in the US?
- Is there an interaction between the demographic factors of each county?

Data Acquisition

I downloaded the county demographic information and COVID-19 vaccination site location datasets from US Census Bureau and GIS Hub, respectively. All datasets are acquired on December 29th, 2021.

Data Exploration

I trimmed the original dataset – I excluded missing entries from the vaccination site dataset and Hawaii and Alaska sites, focusing on the mainland states.

In the end, there were 3,142 counties and 36,230 vaccination sites reflected in the analysis.

I aggregated the number of vaccination sites by county and created a dataset that holistically covers county characteristics. The demographic variables are total population, percentage of females, percentage of people of color, and median household income.

Models

I started with the first-order model which included all 4 independent variables. There is no multicollinearity found among the variables (*VIF* < 2). I also built a full interaction model which included all 4 independent variables and the possible pairs. I conducted variable selection processes using AIC and BIC as the criteria for both first-order and interaction models.

The AIC and BIC selection of the first-order model was identical to the full model. The AIC and BIC selection for the interaction model, however, excluded some of the pairs of variables. The final selection of the interaction model indicated that the following pairs of variables show significant interaction:

- Median Household Income and Total Population
- Median Household Income and Female Population
- Median Household Income and POC Population
- Total Population and POC Population
- Total Population and Female Population

- Cross-validation

	RMSE	R-squared	MAE
Full First-order	10.86324	0.892488	4.635927
AIC/BIC Interaction	10.21472	0.9151525	4.388303

- Final Model

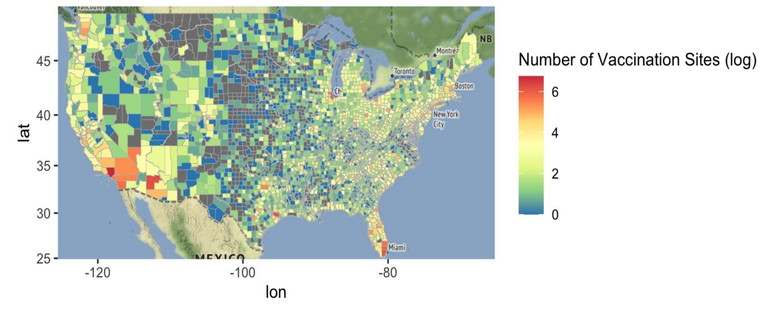
Final Model		
(Intercept)	5.894	**
med_income	-0.0002	***
total_popn	-0.00007	**
percent_POC	-4.784	***
percent_Female	-10.23	**
med_income:total_popn	-2.088×10^{-11}	***
med_income:percent_POC	0.0001	***
med_income:percent_Female	0.0004	***
total_popn:percent_POC	-4.307×10^{-6}	***
total_popn:percent_Female	0.00002	***
Signif. codes	0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1	

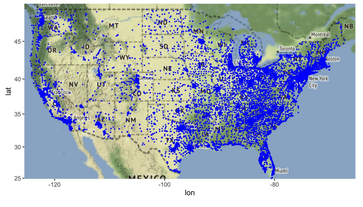
Discussion

The states that include the highest number of vaccination sites are California, Illinois, and Arizona, while the states with the lowest number of vaccination sites are Kentucky and Alabama.

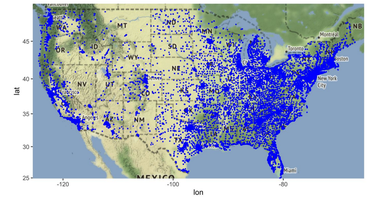
All the demographic variables were correlated with vaccine availability and did not show discrimination patterns for minority populations including female and BIPOC population. There was a significant interaction among population demographics and wealth measures.

The flaw in the vaccination site data is that not all counties are equally represented, as the volunteers who are responsible for certain counties are not randomly selected and understaffed. Therefore, Not all 3,142 counties considered in the analysis were reflected in the dataset.





05.19.21 Dataset



12.29.21 Dataset