COVID-19 Vaccine Distribution Plan via Machine Learning Methods

Adam Brooks Riley Bergendahl Mehmet Gumus Jack (Kuangwen) Ren

Summary

As of today, **COVID-19** has caused over ***500,000*** (and rising) deaths in the United States and has caused many to lose their jobs, making serious social and economic impact on the global community. Vaccines, now more than ever, are the key to stopping the pandemic, and **vaccine strategy** studies, especially with limited vaccine supplies, not only concerns the government policy makers but also the public. Currently, each state in the United States develops its own vaccine distribution program, based on **CDC guidelines**, that focus on front-line workers, essential workers, and high-age population. However, with limited vaccine supplies and order of vaccines on a first-come first-serve basis, states that have a higher high-risk population may not get enough vaccines while some lower risk states may have more than needed.

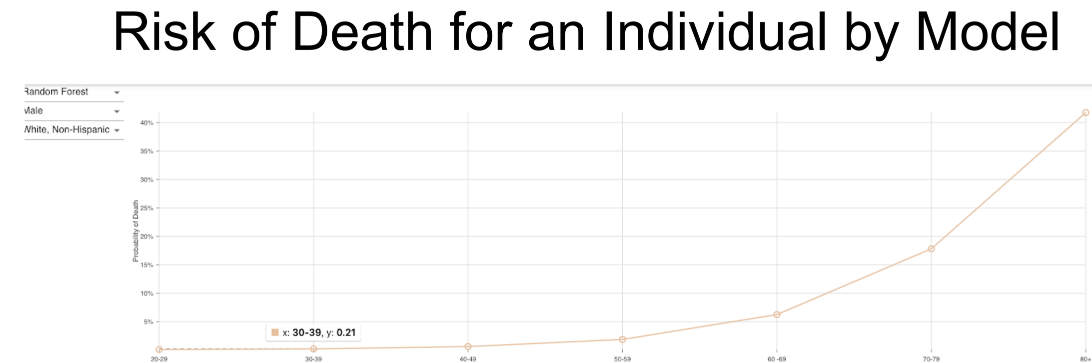
Our optimized **vaccine distribution** method minimizes the total number of deaths by stratifying the population by risk, determined by historical **COVID-19 death rates** among different demographic groups. Using **Machine Learning** and demographic data for each state, we determine the risk to that state and project the number of deaths in an unvaccinated population to allocate the existing vaccine supply accordingly. Our **Vaccine Distribution** method **reduces death rates** by about ***2.7%*** overall.

Assessing Modeling Techniques

The method first applies **Logistic Regression**, **Random Forest**, and **Naive Bayes** models to estimate the **probability of death** of each group combination of sex, ethnicity, and age-group.

Model	AUC	Testing Accuracy
Random Forest	0.9185	95.91%
Logistic Regression	0.9155	95.45%
Naive Bayes	0.9095	95.49%

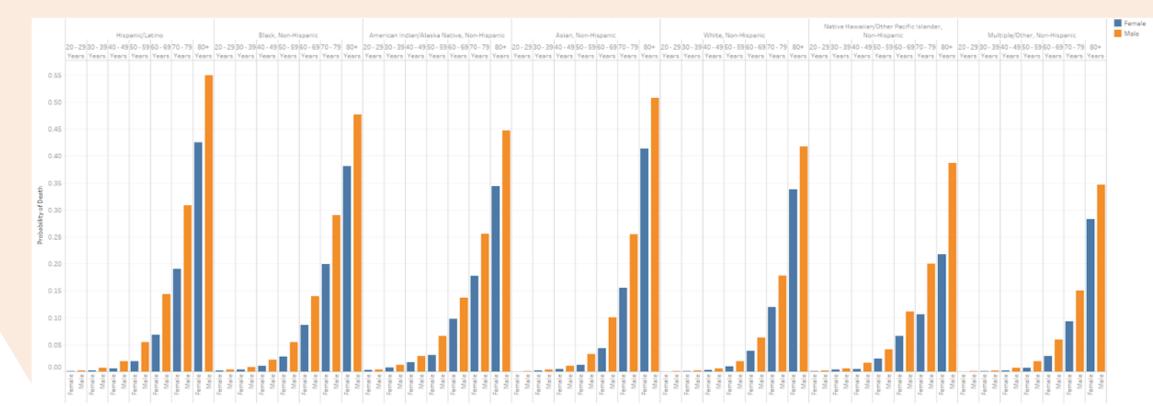
Risk of Death for an Individual by Model



*Using each modeling technique, we can see the **probability of death** per demographic. **Random Forest** produces the best model and is used moving forward.*

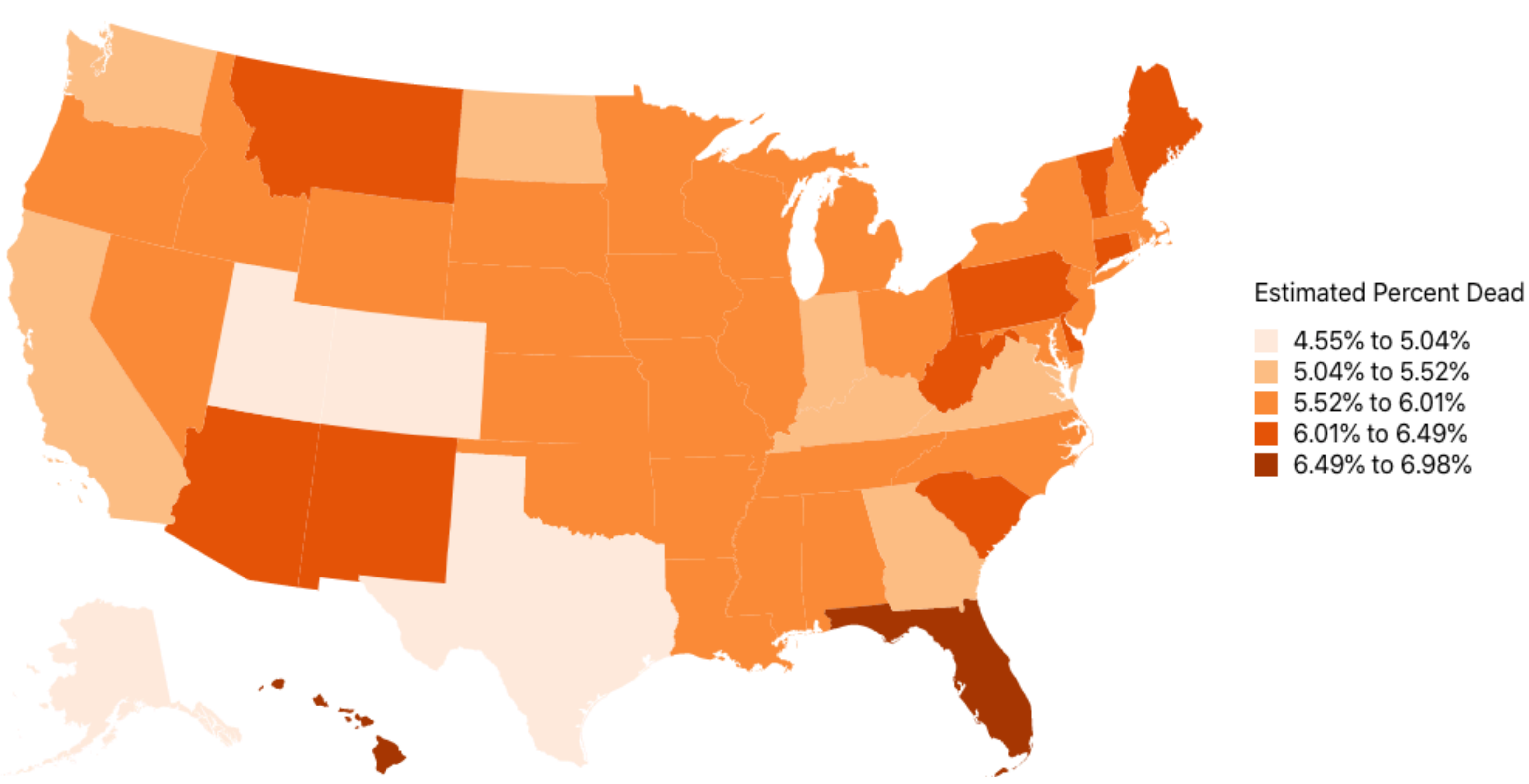
Modeling Probability of Death by Demographic Groups

Death Probability by Demographic Group



Each **Random Forest** tree predicts the **death probability** of a demographic group as the fraction of samples of the deaths in the **leaf node** in which the demographic group is located. The **Random Forest** yields a single probability value for each demographic group by averaging the predicted **death probabilities** of all trees.

State Priority Groups to Receive Vaccine



Data

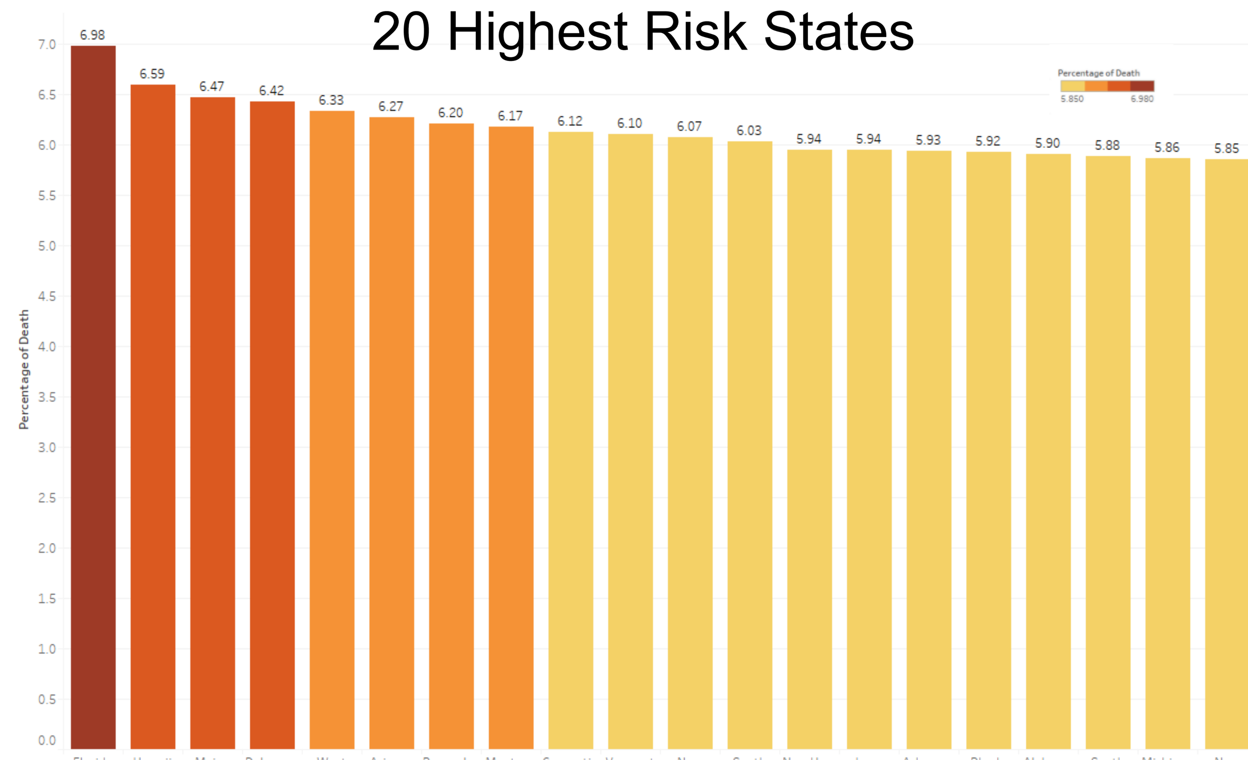
Name	Source	Obtained by	Size	# of Records	Temporal?
SC-EST2019-ALLDATA6	United States Census	Downloaded	18.4 MB	234.8K	Yes
COVID-19 Case Surveillance Public Use Data	CDC	API	2.7 GB	22.5M	Yes
Provisional COVID-19 Death Counts by Sex, Age, and State	CDC	API	9.4 MB	52.3K	Yes
US State Vaccinations	Our World in Data	Downloaded	828 KB	14K	Yes

Results

Based on the **proposed state priorities** (left) the cumulative number of **COVID-19** deaths remains less than the actual cumulative number of **COVID-19** deaths starting from week 3. Between the dates Jan 23, 2021 and Feb 06, 2021, there were ***57,348*** **COVID-19 deaths**. If the vaccines were distributed as we proposed, then there would be approximately ***53,785*** **COVID-19 deaths** between these dates. From Jan 23, 2021 to Apr 10, 2021, ***112,293*** people lost their lives from **COVID-19** in the US. The **Random Forest** model estimates that this number would be ***109,276*** under the proposed vaccine allocation plan. This estimation shows that if the vaccines were distributed to the states according to their priority level we found, then there would be a decrease in the number of **COVID-19** deaths. This approach shows ***3017*** less deaths as of Apr 10, 2021, which is about ***2.7%*** more effective compared to the current approach.

Assessing the Risk of Each State by Demographic Composition

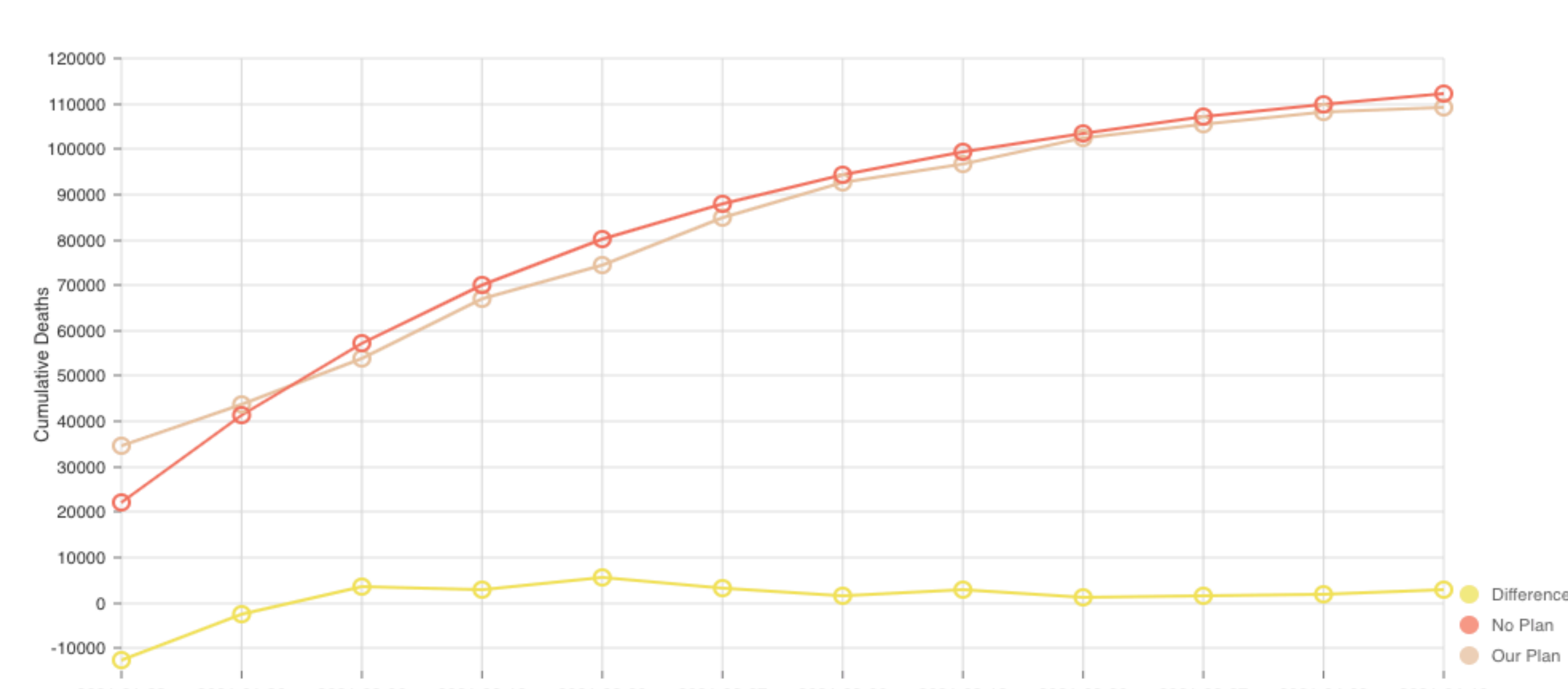
20 Highest Risk States



The **expected percentage of death** of each demographic group is calculated by multiplying the population in each group with its associated **probability of death**. Each state is ranked by **expected percentage of death** to determine the state's risk. At this stage, the results show the expected **probability of death** or **state risk** without any vaccination and/or mitigation methods like social distancing or mandatory mask wearing.

Evaluation

Effectiveness of Distribution Plan



Our proposed approach prioritizes states that have the highest risk, defined as the highest estimated number of deaths. The **Random Forest** model is applied to examine the relationship between the vaccine allocated in each state under the current allocation program using the daily vaccine distributed (cumulative) in the United States and the actual cumulative number of **COVID-19** deaths. The features are defined as the number of vaccinations distributed to each state, and the response as cumulative **COVID-19** deaths. Each row will represent a week of data and its ending date will be indicated by the date column. The number of vaccines allocated to each state is based on the total amount of vaccines distributed in the United States multiplied the weighted average of the estimated number of deaths in each state.