

Predicting COVID-19 Vaccine Uptake in the United States

By Princess Allotey¹, Robin Armstrong², Erik Bergland³, Eddie Mitchell⁴, and Nikki Wang⁵

¹Carnegie Mellon University, Department of Mathematical Sciences ²Cornell University, Department of Mathematics ³Brown University, Department of Applied Mathematics

⁴University of Tennessee-Knoxville, Department of Mathematics ⁵Johns Hopkins University, Department of Mathematics

Johns Hopkins University | SNF Agora Institute | Baltimore, MD

Introduction

Vaccines play a pivotal role in mitigating the harm from COVID-19. It is therefore critical to understand the factors that influence one's decision to take the vaccine. Previous work¹ has shown that party alignment correlates strongly with willingness to get vaccinated. In this project, we explore the extent to which other variables are predictive of vaccine uptake.

Research Questions

1. What are the most influential factors leading to individuals in the U.S. choosing not to get vaccinated for COVID-19, beyond party vote?
2. Based on our findings, what kind of policies/actions would we recommend to encourage more individuals to choose to get the COVID-19 vaccine?

Datasets and Methods

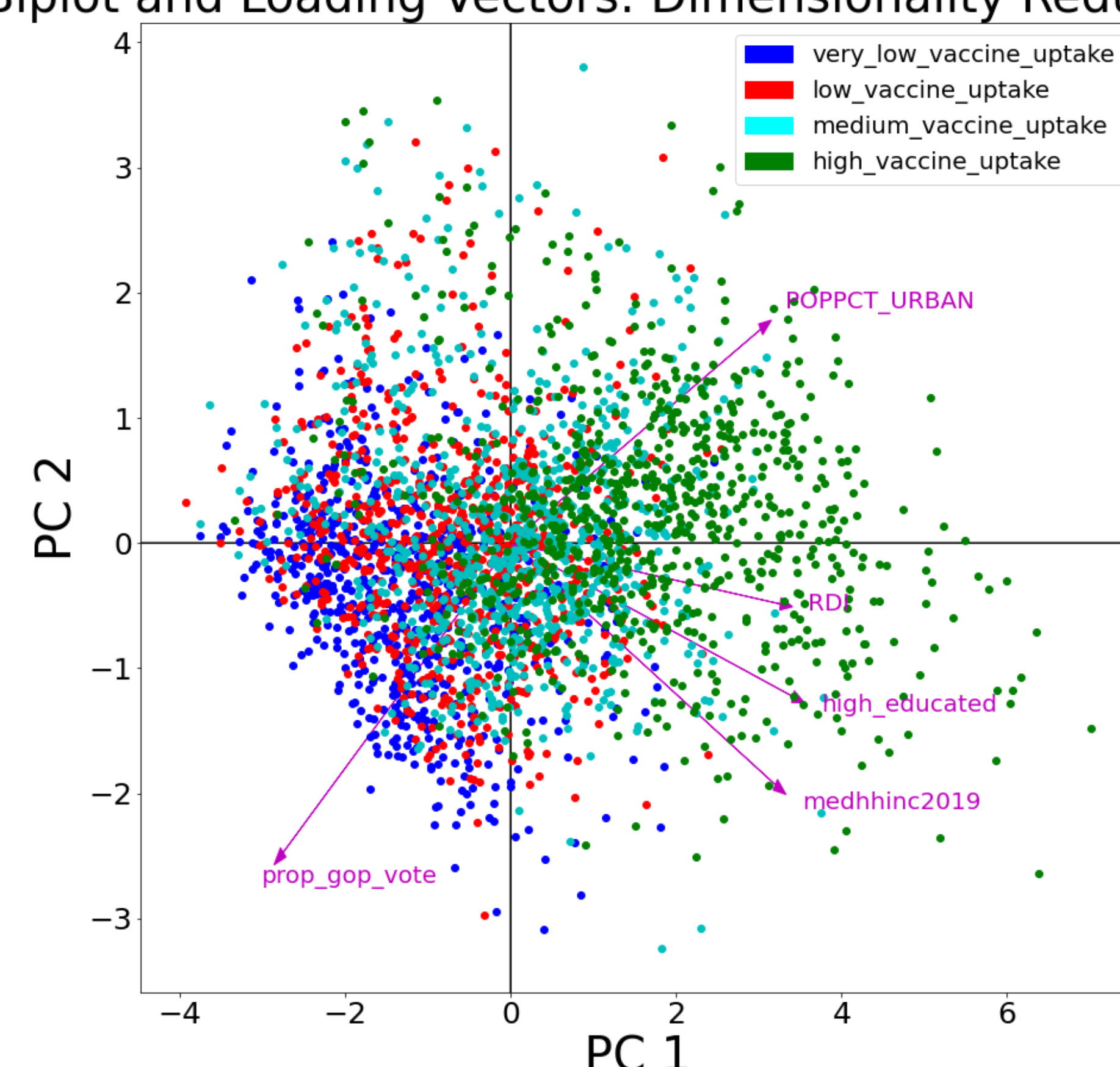
Datasets used:

- Inherited project data: county-level social, demographic, and vaccine-related information.
- CoVaxxy²: Twitter-based social media misinformation data.

Methods employed:

- For exploratory analysis, we made use of PCA and correlation matrices.
- To predict vaccine uptake, we implemented both multinomial logistic regression and random forest.
- We also used both k-nearest neighbors and k-means clustering to predict quartiles of vaccine uptake.

Biplot and Loading Vectors: Dimensionality Reduction



Results

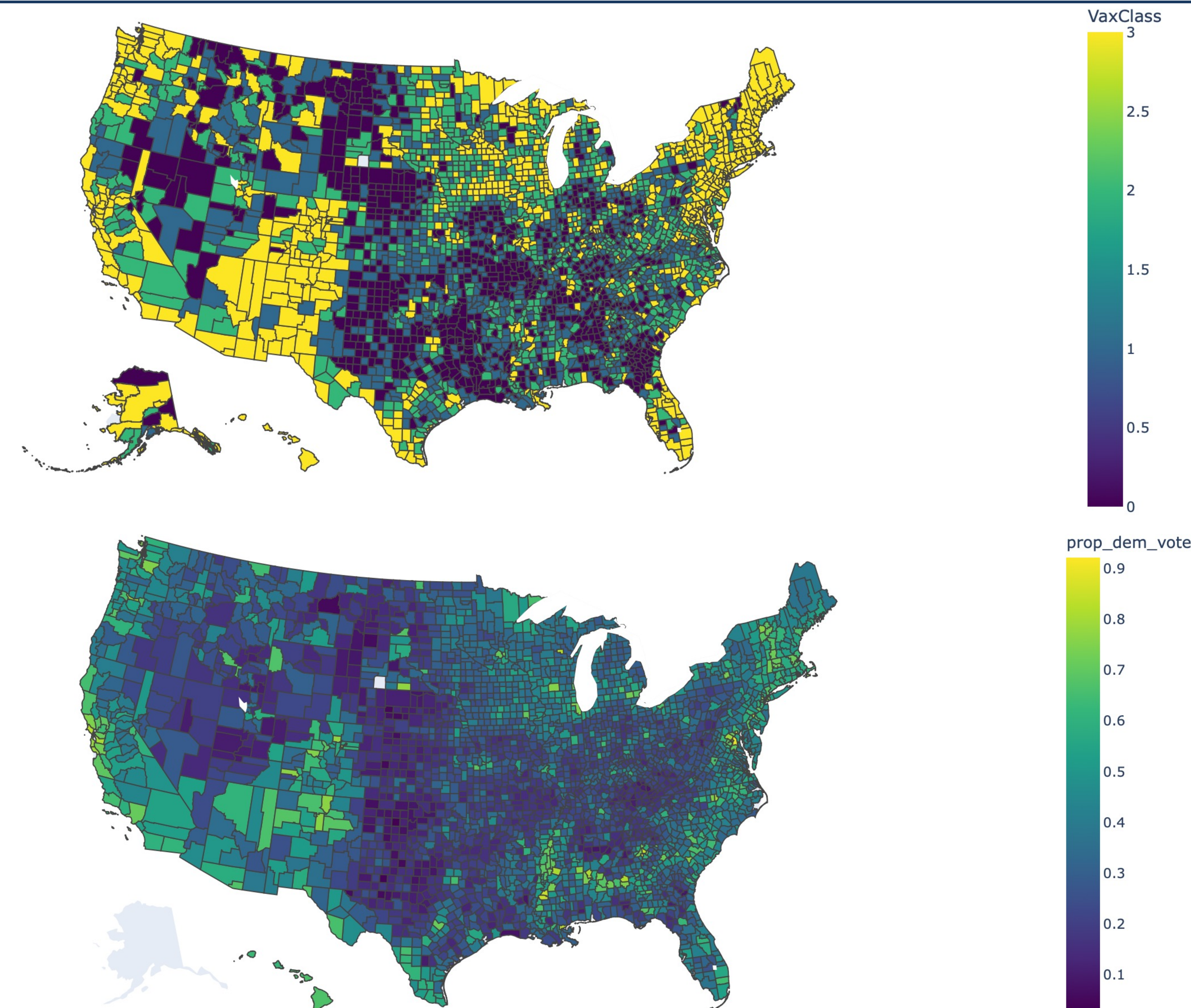


Figure 1—Party Affiliation and Vaccine Uptake

(top) Vaccination uptake broken down into quartiles. Higher index of quartile corresponds to higher vaccination rate.
(bottom) Proportion of county that voted democrat in the 2020 presidential election.

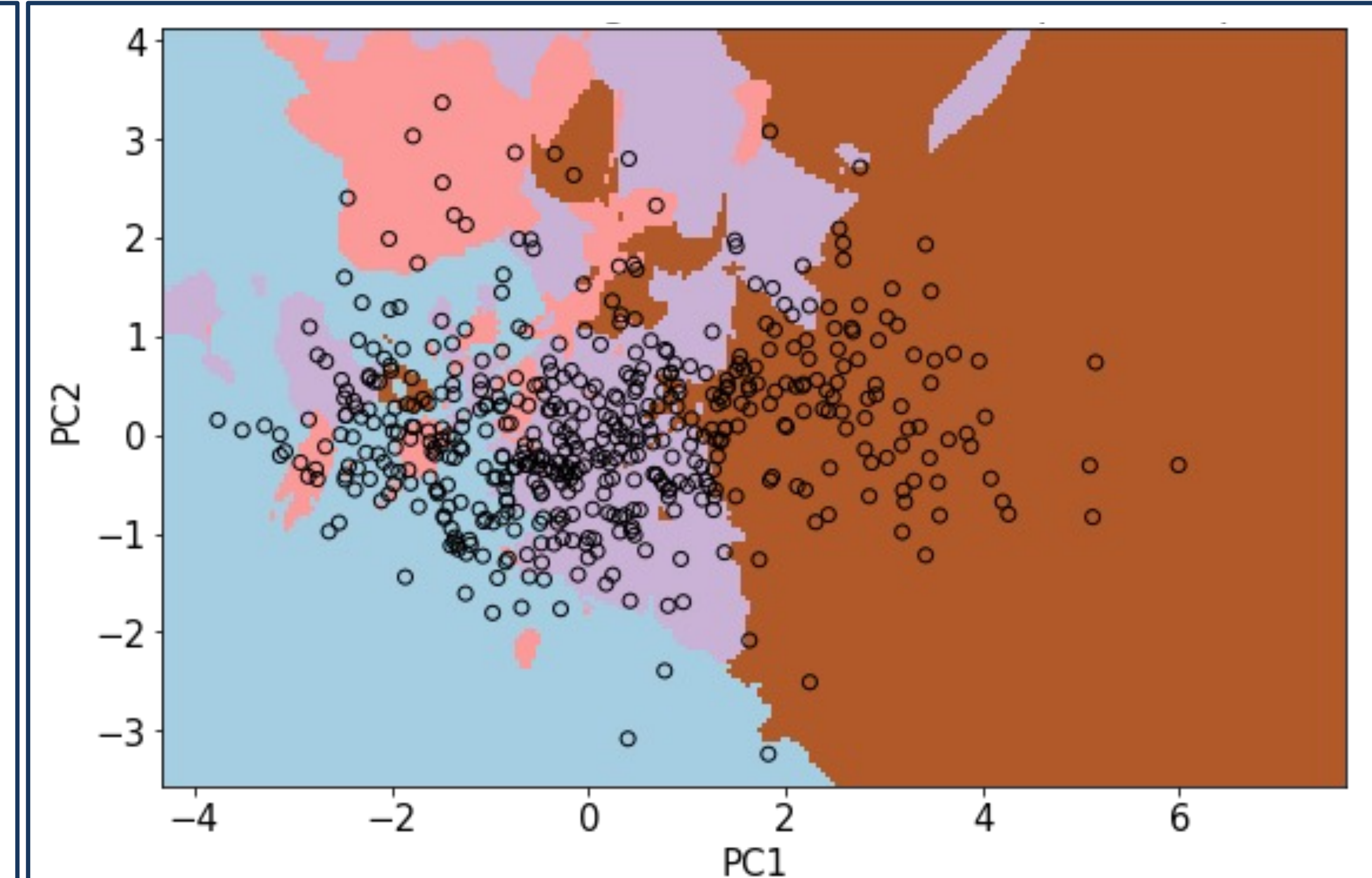


Figure 4—Decision Boundary for k-NN classifier

Classifying vaccination quartile based on 5 covariates. For visualization purposes, a two-dimensional cross section of decision boundary is shown.

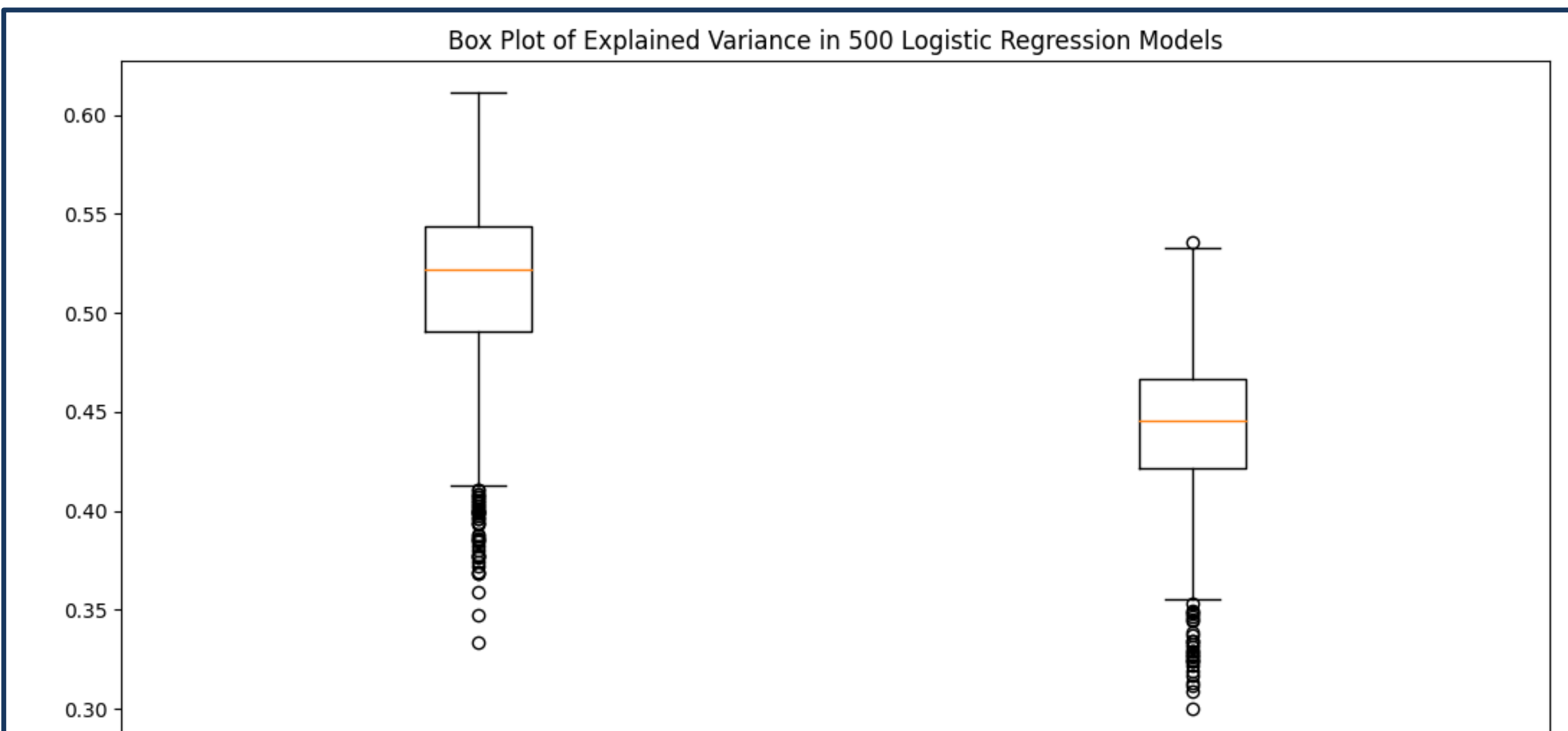


Figure 5 – Logistic Regression Models

Left box plot corresponds to explained variance using logistic regression on an expanded variable set. Right boxplot is logistic regression only taking into account the proportion of GOP voters in a county.

Conclusion & Acknowledgements

Our analysis demonstrates that while the inclusion of other social and demographic variables improves model accuracy, performance still largely hinges on party affiliation. On this basis, it appears that efforts to reduce partisan polarization remain the most effective approach to increasing vaccine uptake.

This project was supported by JHU AMS and the SNF Agora Institute through the Data Science for Democracy workshop.

References

- ¹ The relationship between vaccination rates and COVID-19 cases and deaths in the USA , E. Dong & L. Gardner., JHU CSSE, 2021
- ² CoVaxxy: A Collection of English-Language Twitter Posts About COVID-19 Vaccines, DeVerna, M. et al., Proceedings of the International AAAI Conference on Web and Social Media, (2021)
- ³ Scikit-learn: Machine Learning in Python, Pedregosa et al., JMLR 12, pp. 2825-2830, 2011.

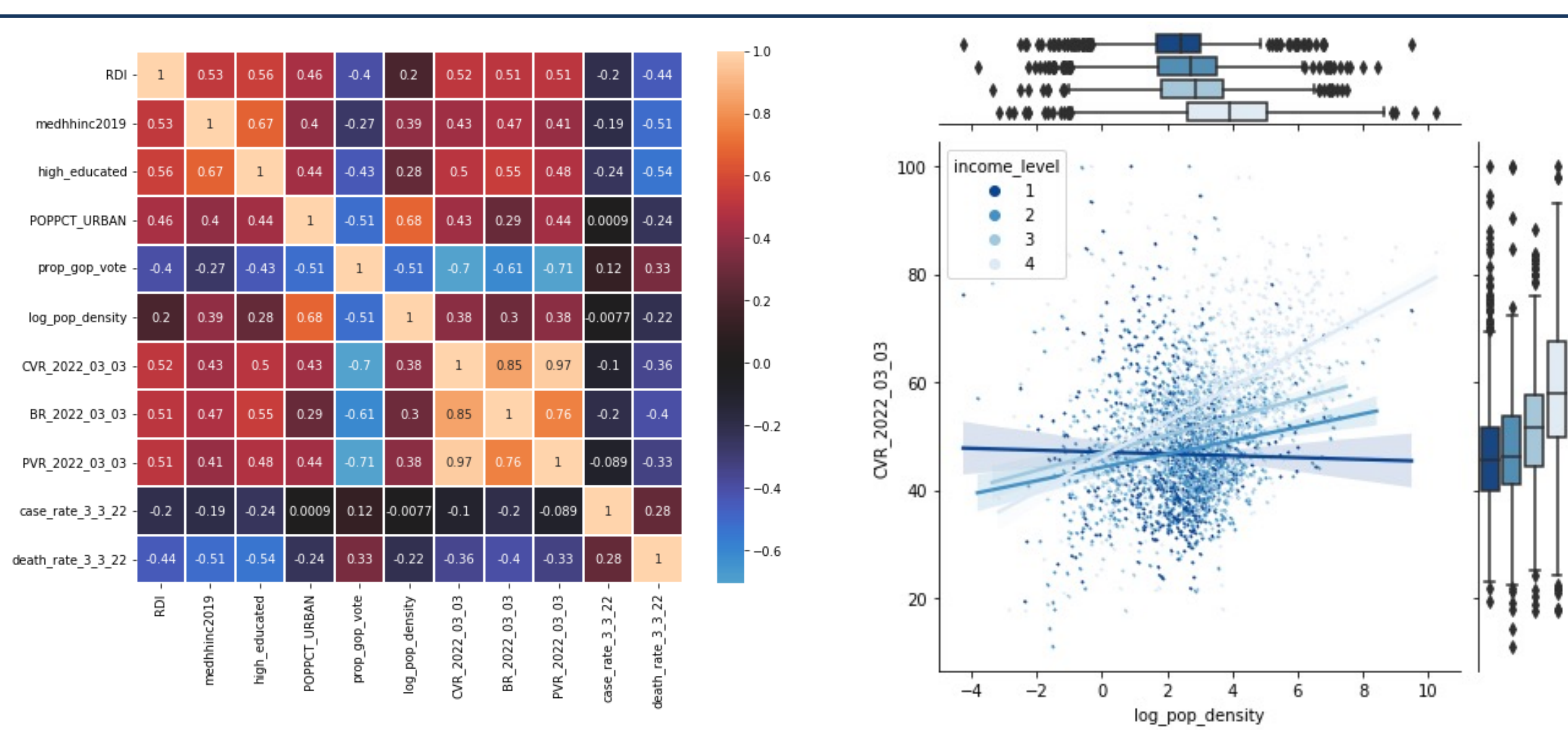


Figure 2—Exploratory Data Analysis

(left) Correlation matrix of selected data. (right) Joint plot of population density and complete vaccine rate.