Background

Access to safe water is a fundamental human right, yet many residents across the United States lack the plumbing infrastructure necessary to obtain it. Past research from the U.S. Geological Survey (USGS) Social and Economic Drivers program has identified several social vulnerability determinants associated with water insecurity, but it has not analyzed how multiple such factors can interact to influence plumbing insecurity. Among the frequently studied variables are education and race/ethnicity, which raises the question: *to what extent is plumbing insecurity across U.S. counties associated with educational attainment levels and ethnic composition?*

The U.S. Census Bureau's American Community Survey (ACS) annually publishes demographic, social, economic, and housing data for the U.S. population, enabling analysis of this association at the county level. This study hypothesizes that if a county has higher proportions of Hispanic/Latino residents and lower educational attainment, then it is likely to experience higher levels of plumbing insecurity.

Results

Ablation Study

Fig. 1. R^2 of model variants

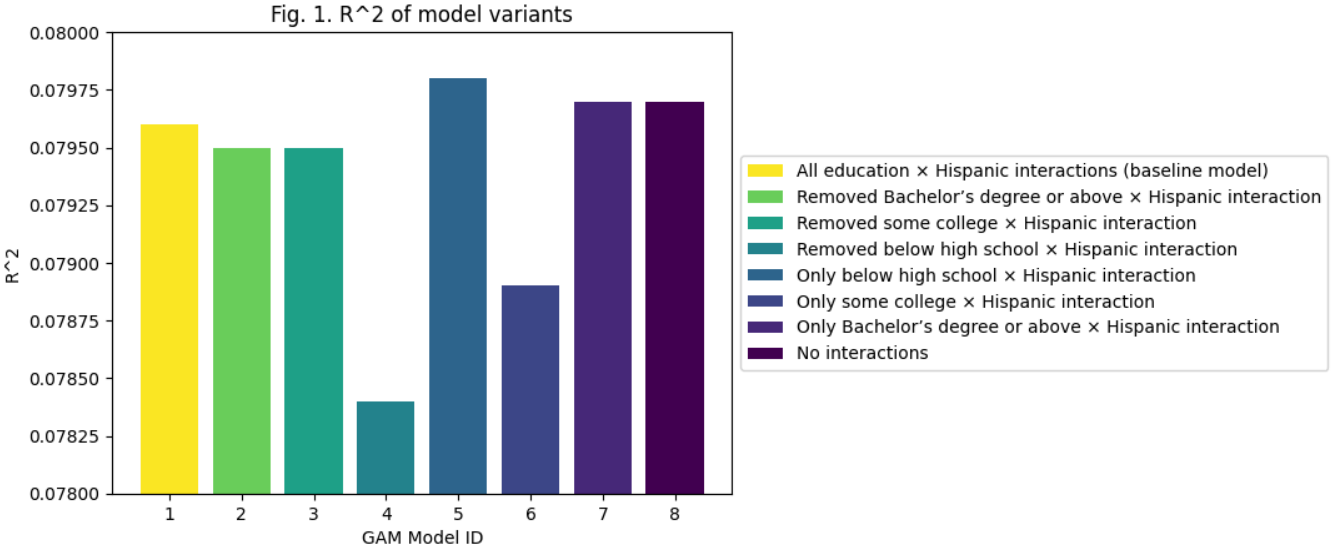

Figure 1 compares the R^2 predictive performance across all model variants. Differences between model variants and the baseline are minimal ($|\Delta R^2| < 0.02$).

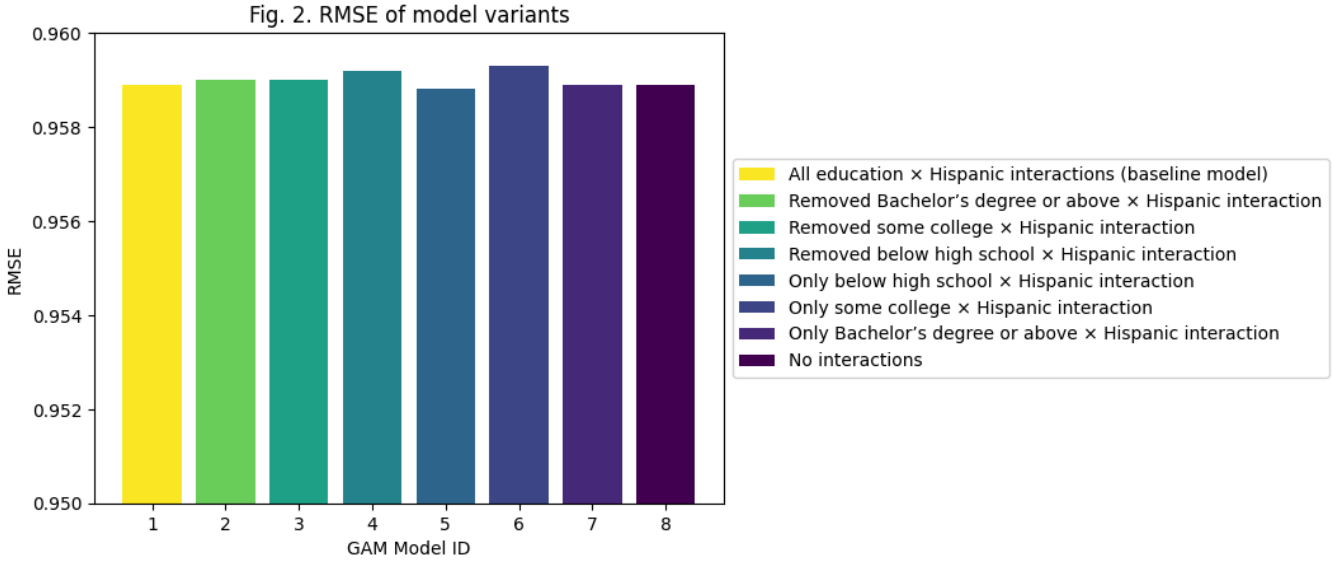

Figure 2 compares the RMSE error across all model variants. Differences between model variants and the baseline are minimal ($|\Delta RMSE| < 0.001$).

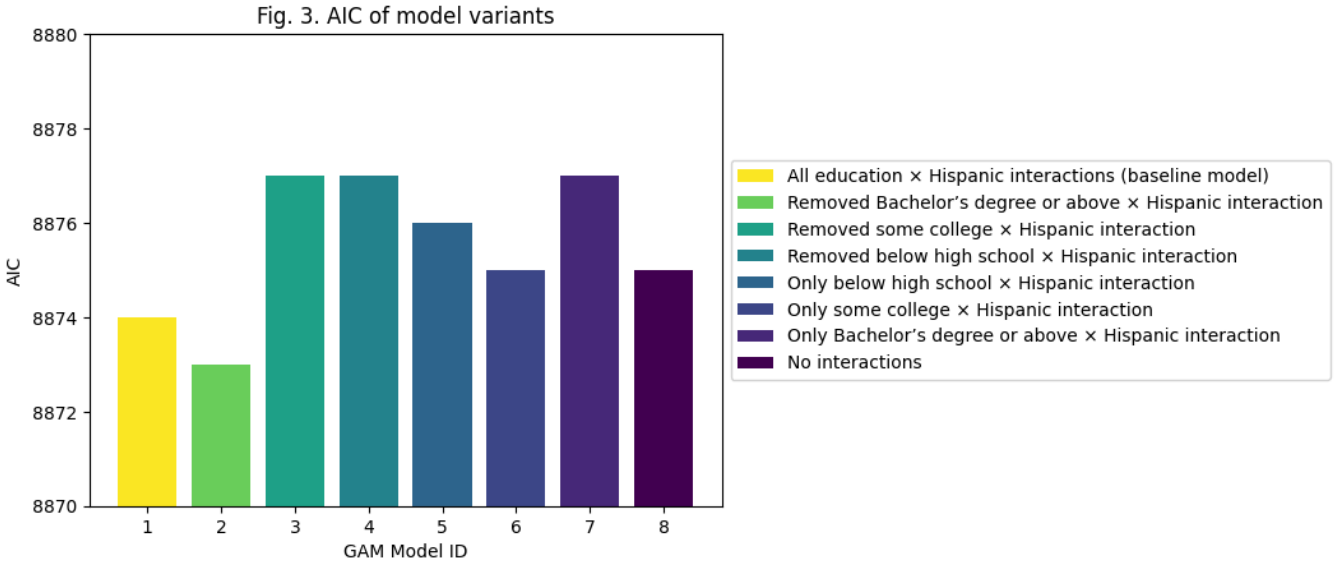

Figure 3 compares the AIC goodness-of-fit across all model variants. Differences between model variants and the baseline are minimal ($|\Delta AIC| < 5$).

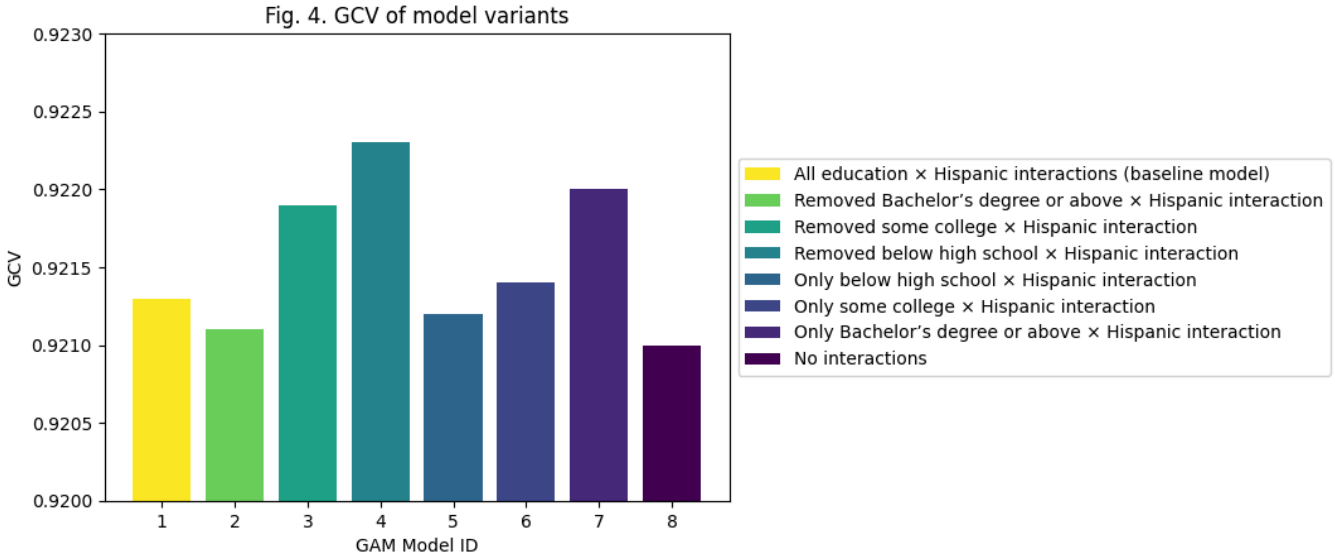

Figure 4 compares the GCV error across all model variants. Differences between model variants and the baseline are minimal ($|\Delta GCV| < 0.002$).

Partial Dependence Plots (PDPs)

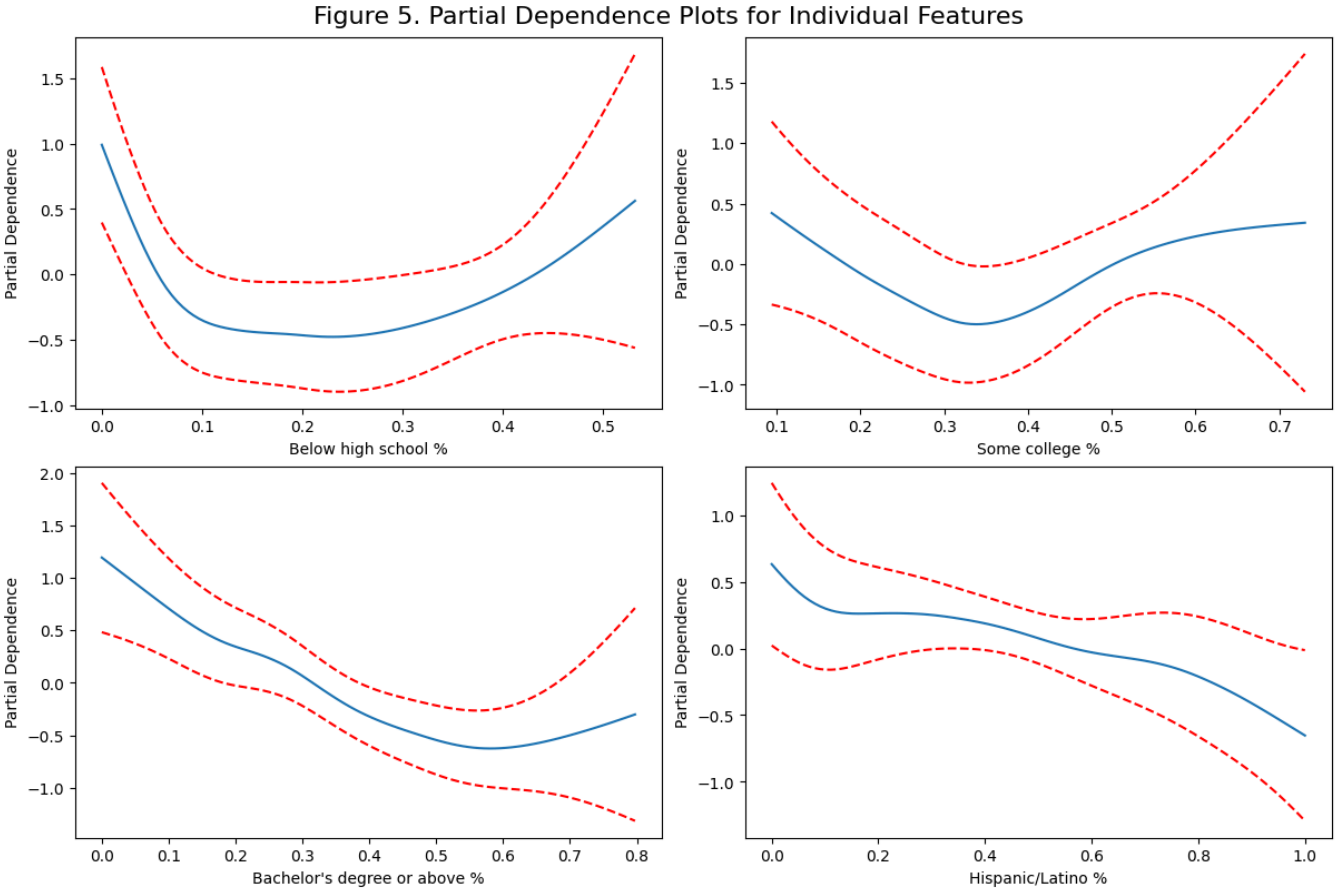

Figure 5 presents the partial dependence plots for individual features, highlighting how each feature independently influences the model's predictions. The positive partial

dependence across a substantial portion of each feature's observed range suggests these features generally contribute to predicting plumbing insecurity.

Figure 6. Partial Dependence Surfaces for Interactions

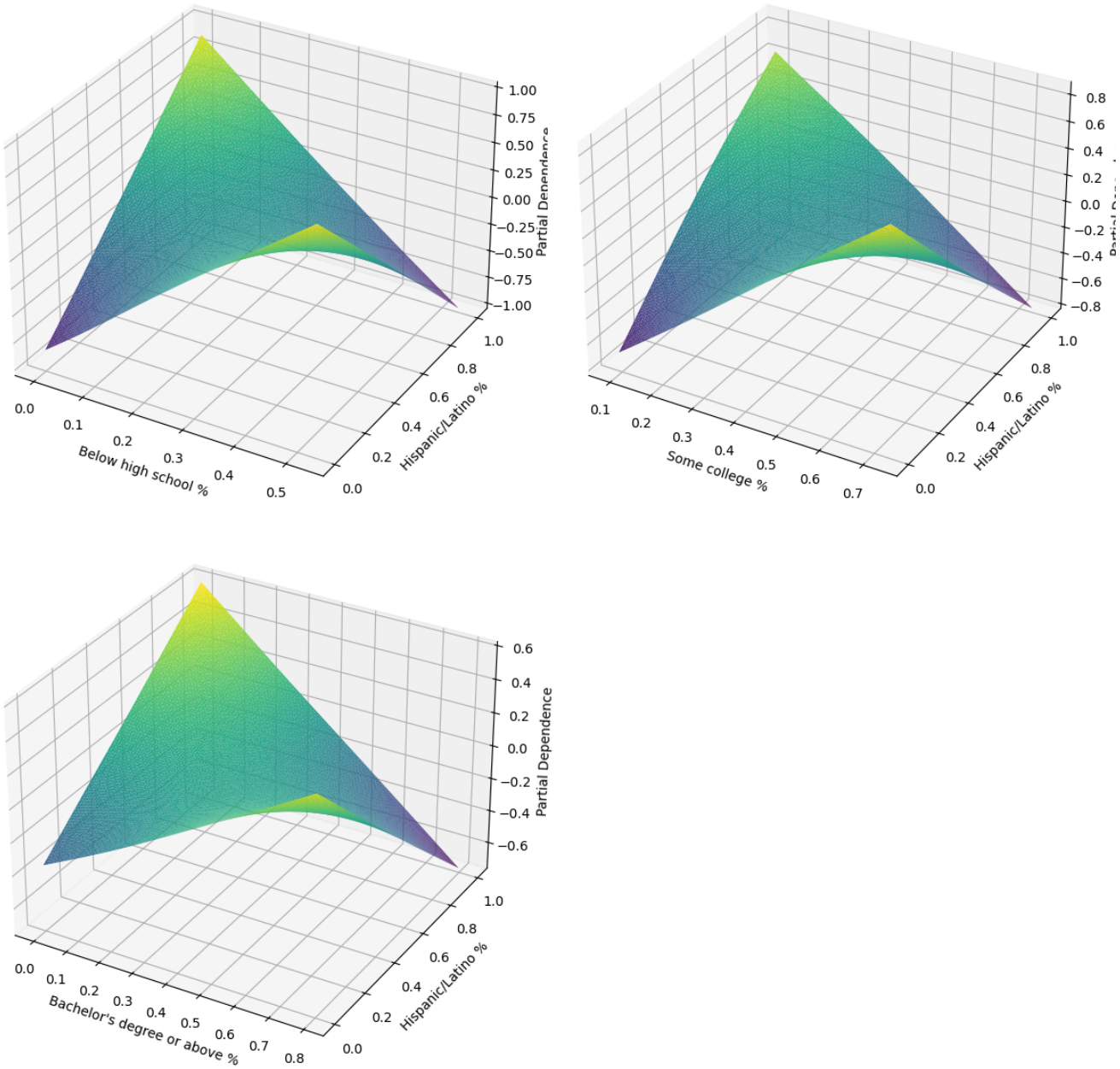

Figure 6 presents the partial dependence surfaces for feature interactions. Peak contributions occur when one feature is at its minimum and the other at its maximum, suggesting that each feature contributes more individually than together. Additionally, the range of partial dependence is lower for the interactions compared to the individual features, indicating a smaller predictive contribution.

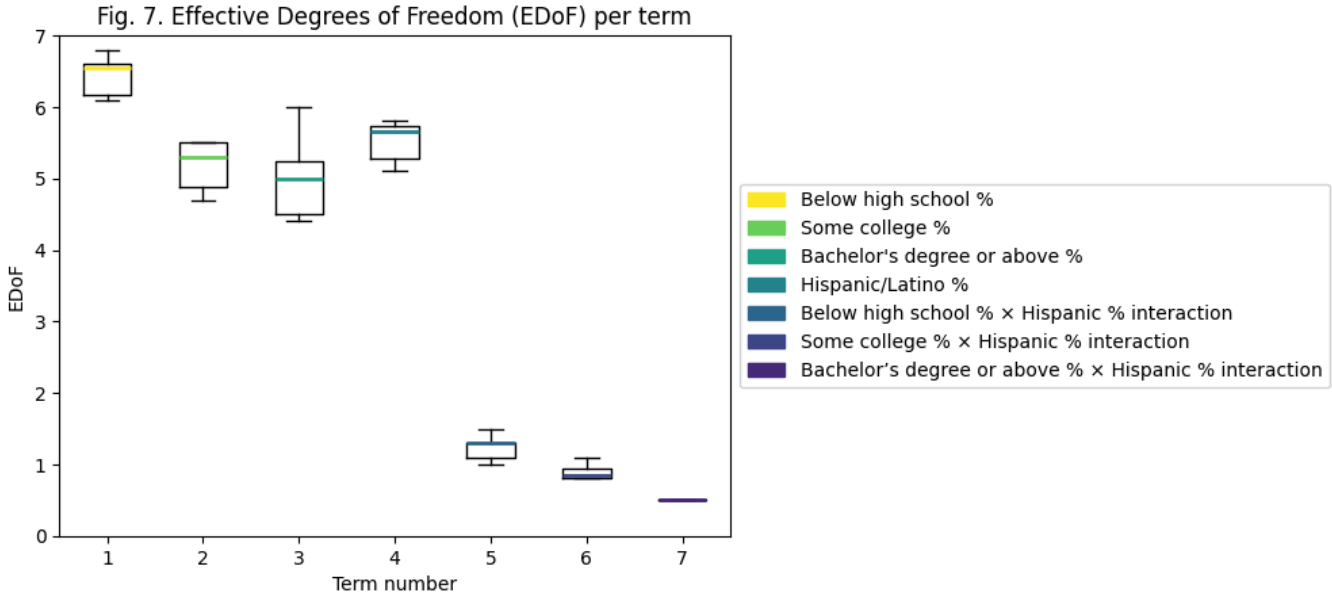

Figure 7 compares the effective degrees of freedom (EDoF) per term. Individual feature terms have EDoFs between 4 and 7, indicating that the model fits nuanced, nonlinear relationships for them. In contrast, interaction terms have EDoFs between 0 and 2, indicating that the model fits them with less complexity, possibly due to weaker or less consistent patterns in the data.

Discussion

This study investigated the extent to which plumbing insecurity is associated with educational attainment and ethnic composition across U.S. counties. The results indicate that each risk factor contributes individually to the prediction of plumbing insecurity. However, their interaction does not substantially improve predictive performance. Therefore, for modeling purposes, education and ethnicity provide more value when considered independently. These findings contradict the original hypothesis, which anticipated a stronger combined effect of the two factors on plumbing insecurity.

The scope of this study was limited as it only analyzed two of the top social vulnerability indicators. As such, the results should not be generalized to interactions of any factors. Future research should examine a broader range of variables to assess whether any combination of risk factors (beyond education and ethnicity) is a stronger predictor of plumbing insecurity.

Code and Data Availability

The Jupyter notebook and dataset used can be found at https://github.com/eDoggo3779/Water_Insecurity.

The dataset was compiled from the National Historical Geographic Information System (NHGIS), https://www.nhgis.org/.

Acknowledgements

This study would not have been possible without the support of a few key individuals. I would like to thank The Coding School for training in data science techniques and exposure to various domains; my teaching assistant, Kayvan Jalali, for his technical guidance and consistently thoughtful responses to my questions or requests; and my mentor, Dr. Tara Flaugher, for her persistent patience and nuanced views, which helped shape the presentation and language of this work.