Race-Specific Risk Factors for Homeownership Disparity in the United States: Supplemental Materials

Rachel Richardson[1,*], Damon Leach[1], Anastasiya Prymolenna[2], David Degnan[1], Natalie Winans[1], and Lisa Bramer[1]

[1]*902 Battelle Boulevard, Richland, WA, 99354, Biological Sciences Division, Pacific Northwest National Laboratory, USA*
[2]*902 Battelle Boulevard, Richland, WA, 99354, Environmental Molecular Sciences Division, Pacific Northwest National Laboratory, USA*

Supplemental Materials

Tables
- Table 1: Descriptions of variables
- Table 2: Division county percentages per Race dataset
- Table 3: Number of counties in each dataset

Plots
- Figure 1: Dataset timeline
- Figure 2: Important variables for White models with and without outliers
- Figure 3: Correlation between predictor variables
- Figure 4: Model performance on 10% holdout data

*Corresponding author. Email: rachel.richardson@pnnl.gov.

Table 1: List of variables used in modeling and corresponding descriptions. Variables with high importance in at least one model indicated with bold text. Importance was calculated with the *randomForest* R package using a permutation method in which larger increases in MSE indicate variables more important to the overall model performance. All variables were averaged to the county level.

Variable	Dataset	Description
annualTot	Job availability	Average number of total jobs
com10less	Unequal commute times	Proportion of people with job access with a commute time of less than 10 minutes
com10to29	Unequal commute times	Proportion of people with job access with a commute time of 10 to 29 minutes
com30to59	Unequal commute times	Proportion of people with job access with a commute time of 30 to 59 minutes
com60plus	Unequal commute times	Proportion of people with job access with a commute time of 60 minutes or more
costBurd	Homeownership demographics	Share of homeowners with a mortgage who are cost-burdened
costBurdSev	Homeownership demographics	Share of homeowners with a mortgage who are severely cost-burdened spending 50 percent or more of annual income on housing costs
costIncRatio	Homeownership demographics	Median monthly ownership cost to income ratio
division	US Census Bureau Regions and Divisions with State FIPS Codes	Each of the four census Regions is divided into two or more census Divisions
earn100ami	Homeownership demographics	Number of homeowners earning less than 100 percent of area median income
earn150ami	Homeownership demographics	Number of homeowners earning less than 150 percent of area median income
enc	US Census Bureau Regions and Divisions with State FIPS Codes	Factor variable for counties in East North Central Division of the US
esc	US Census Bureau Regions and Divisions with State FIPS Codes	Factor variable for counties in East South Central Division of the US
gqest	Annual Resident Population Estimates	Group Quarters total population estimate
HO.over.Race	Annual County Resident Population	The ratio of racial or ethnic homeownership representation out of total homeowners to the racial or ethnic 19 population representation out of total 19 total population
hoCost	Homeownership demographics	Median monthly homeownership costs
hsRate	Education	Proportion of individuals with at least a high school degree

income	Homeownership demographics	Median yearly income of homeowners calculated from medowncostincomeratio medownercost
job14to29	Job availability	Proportion of jobs that are available for ages 14 to 19
job30to54	Job availability	Proportion of jobs that are available for ages 30 to 54
job55	Job availability	Proportion of jobs that are available for ages 55 and older
jobAAPI	Job availability	Proportion of jobs that are available for those who are pacific islander
jobAgr	Job availability	Proportion of jobs that are available for the agriculture industry
jobAsian	Job availability	Proportion of jobs that are available for those of Asian descent
jobAssc	Job availability	Proportion of jobs that are available for those with some college education
jobBach	Job availability	Proportion of jobs that are available for those with a bachelors degree or higher
jobBlack	Job availability	Proportion of jobs that are available for those who are Black
jobConst	Job availability	Proportion of jobs that are available for the construction industry
jobEdu	Job availability	Proportion of jobs that are available for the education industry
jobEntmt	Job availability	Proportion of jobs that are available for the entertainment industry
jobFemale	Job availability	Proportion of jobs that are available for those who identify as female
jobFood	Job availability	Proportion of jobs that are available for the food services industry
jobHC	Job availability	Proportion of jobs that are available for the health care industry
jobHisp	Job availability	Proportion of jobs that are available for those who are Hispanic
jobHS	Job availability	Proportion of jobs that are available for those with a high school education
jobIncomeH	Job availability	Proportion of jobs that are available with monthly earnings more than $3333
jobIncomeL	Job availability	Proportion of jobs that are available with monthly earnings less than $1250
jobIncomeM	Job availability	Proportion of jobs that are available with monthly earnings from $1250 to $3333
jobInfo	Job availability	Proportion of jobs that are available for the information industry
jobIns	Job availability	Proportion of jobs that are available for the insurance industry

jobLessHS	Job availability	Proportion of jobs that are available for those with less than a high school education
jobMale	Job availability	Proportion of jobs that are available for those who identify as male
jobManuf	Job availability	Proportion of jobs that are available for the manufacturing industry
jobMgmt	Job availability	Proportion of jobs that are available for the management industry
jobMining	Job availability	Proportion of jobs that are available for the mining industry
jobMRacial	Job availability	Proportion of jobs that are available for those who are multiracial
jobNative	Job availability	Proportion of jobs that are available for those who are Native American
jobNotHisp	Job availability	Proportion of jobs that are available for those who are not Hispanic
jobOther	Job availability	Proportion of jobs that are available for the other services industry
jobPubAd	Job availability	Proportion of jobs that are available for the public administration industry
jobRealEst	Job availability	Proportion of jobs that are available for the real estate industry
jobRetail	Job availability	Proportion of jobs that are available for the retail industry
jobScience	Job availability	Proportion of jobs that are available for the science industry
jobTransp	Job availability	Proportion of jobs that are available for the transportation industry
jobUtilities	Job availability	Proportion of jobs that are available for the utilities industry
jobWhite	Job availability	Proportion of jobs that are available for those who are White
jobWholeS	Job availability	Proportion of jobs that are available for the wholesale industry
jobWM	Job availability	Proportion of jobs that are available for the waste management industry
midatl	US Census Bureau Regions and Divisions with State FIPS Codes	Factor variable for counties in Middle Atlantic Division of the US
midwest	US Census Bureau Regions and Divisions with State FIPS Codes	Factor variable for counties in Midwest Region of the US
mort100ami	Homeownership demographics	Number of homeowners with an outstanding mortgage earning less than 100 percent of area-medianincome

mort150ami	Homeownership demographics	Number of homeowners with an outstanding mortgage earning less than 150% of areamedianincome
mountain	US Census Bureau Regions and Divisions with State FIPS Codes	Factor variable for counties in Mountain Division of the US
neweng	US Census Bureau Regions and Divisions with State FIPS Codes	Factor variable for counties in New England Division of the US
northeast	US Census Bureau Regions and Divisions with State FIPS Codes	Factor variable for counties in Northeast Region of the US
nw100.150ami	Homeownership demographics	Percentage of non-White homeowners earning between 100% and 150% of area median income
pacific	US Census Bureau Regions and Divisions with State FIPS Codes	Factor variable for counties in Pacific Division of the US
perDiem	DOD per diem rates	Per diem rates per county
popDensity	Average Household Size and Population Density by County	Number of people per square kilometer
rBirth	Annual Resident Population Estimates	Birth rate
rDeath	Annual Resident Population Estimates	Death rate
rDomestic	Annual Resident Population Estimates	Net domestic migration rate
region	US Census Bureau Regions and Divisions with State FIPS Codes	Four census regions encompassing each quadrant of the United States West Midwest South Northeast
rForeclose	Homeownership demographics	Urban predicted foreclosure rate metric
rIntl	Annual Resident Population Estimates	Net international migration rate
rMigration	Annual Resident Population Estimates	Net migration rate
rNaturalInc	Annual Resident Population Estimates	Natural increase rate
south	US Census Bureau Regions and Divisions with State FIPS Codes	Factor variable for counties in South Region of the US
southatl	US Census Bureau Regions and Divisions with State FIPS Codes	Factor variable for counties in South Atlantic Division of the US
w100.150ami	Homeownership demographics	Percentage of White homeowners earning between 100% and 150% of area median income

west	US Census Bureau Regions and Divisions with State FIPS Codes	Factor variable for counties in West Region of the US
wnc	US Census Bureau Regions and Divisions with State FIPS Codes	Factor variable for counties in West North Central Division of the US
wsc	US Census Bureau Regions and Divisions with State FIPS Codes	Factor variable for counties in West South Central Division of the US

Table 2: Percentage of counties in each holdout division for each model dataset. Note that only counties with a race population of $\geqslant 500$ were included per race model.

Division	Asian	Black	Hispanic	White
East North Central	15.03%	14.45%	15.13%	17.87%
East South Central	7.91%	13.51%	9.45%	12.38%
Middle Atlantic	10.96%	8.74%	8.56%	7.82%
Mountain	7.23%	4.62%	7.87%	6.43%
New England	4.75%	2.81%	3.22%	3.44%
Pacific	10.51%	6.07%	7.39%	5.99%
South Atlantic	22.03%	25.29%	21.83%	21.08%
West North Central	9.83%	8.74%	10.68%	11.21%
West South Central	11.75%	15.75%	15.88%	13.76%

Table 3: Number of counties recorded in datasets before processing.

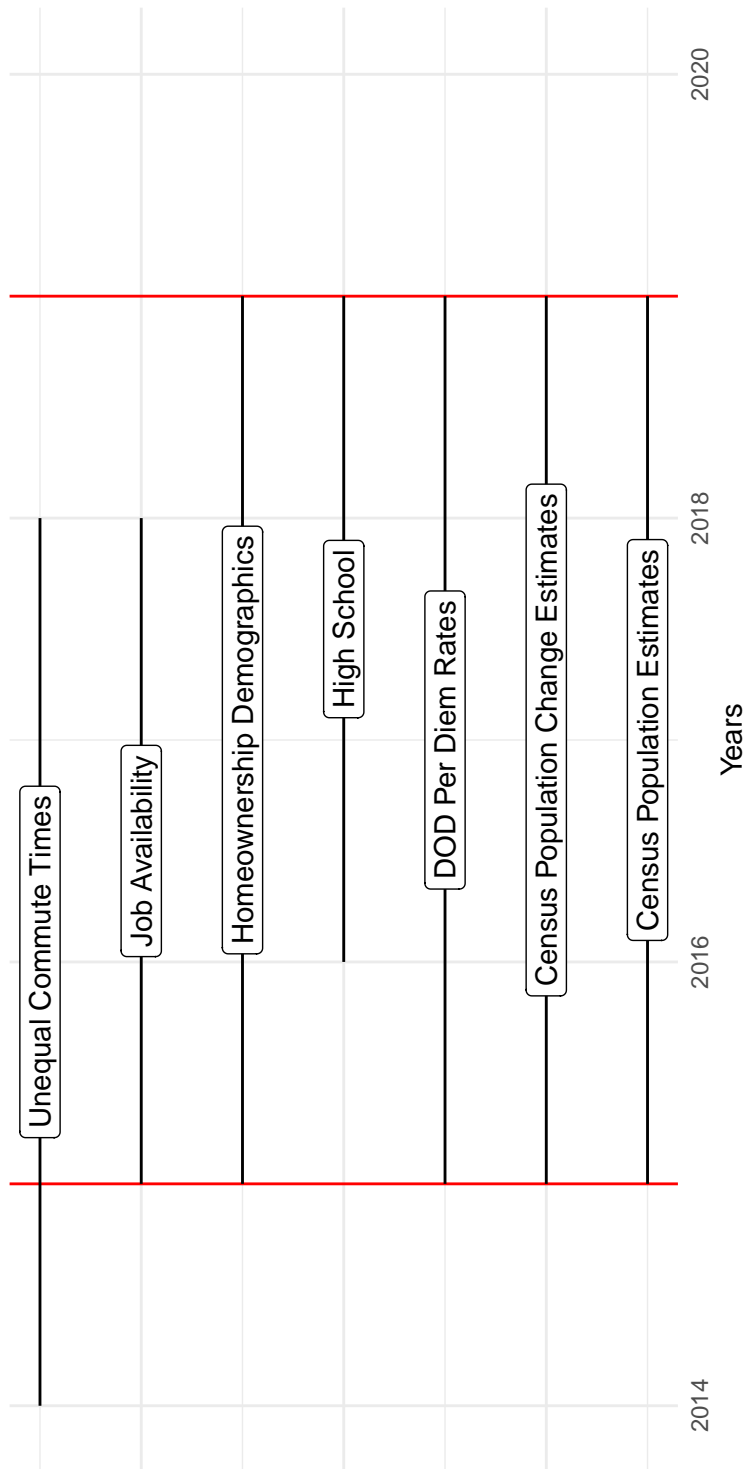| Dataset	Number of counties
Census population estimates	3,142
Census county influx	3,193
Per diem rates	3,067
High school graduation	3,151
Homeownership demographics	3,141
Job availability demographics	3,142
Unequal commute demographics	1,913

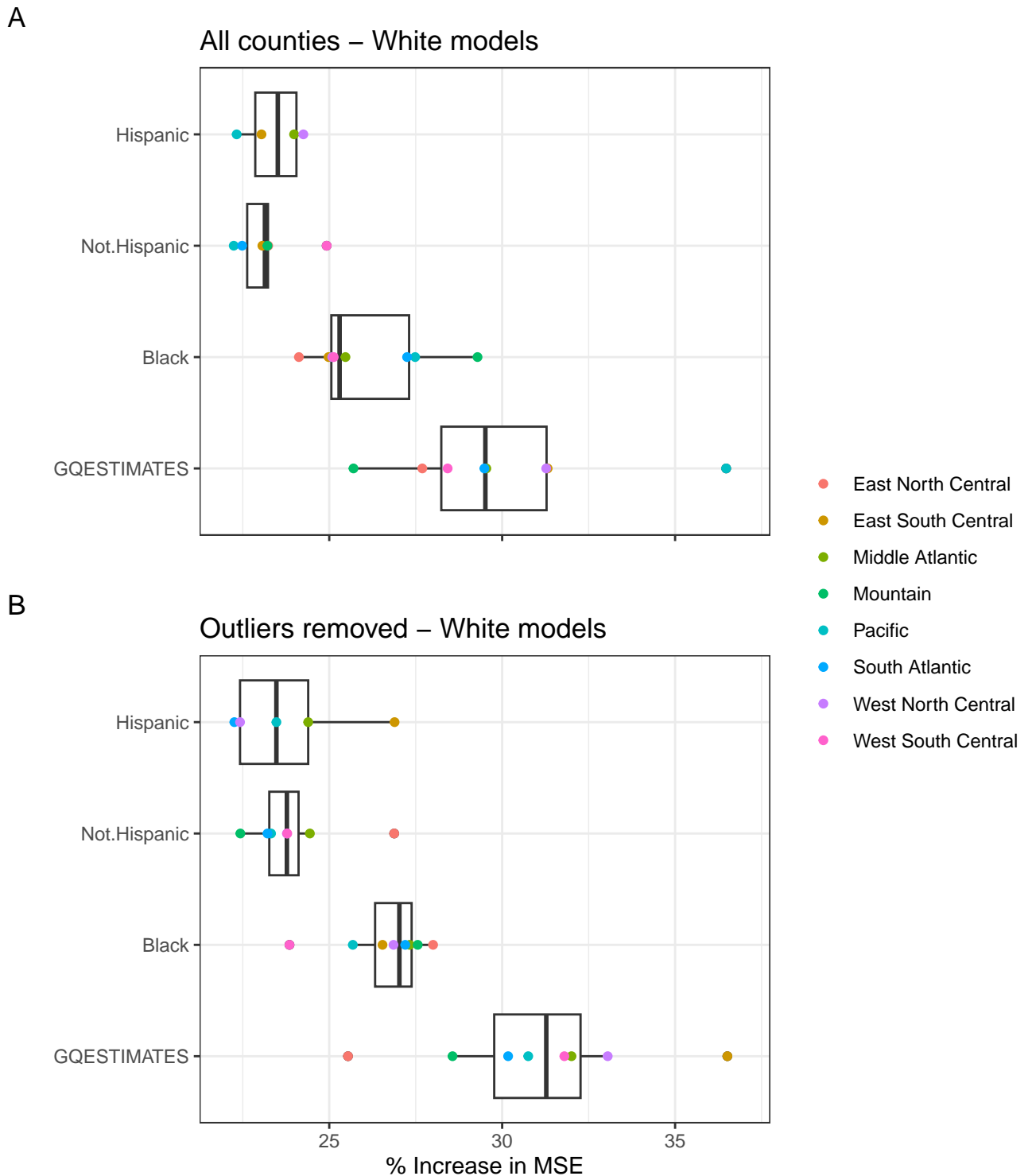Figure 1: Timelines of collected datasets after selecting for years within the 2015 to 2019 window (red lines).

Figure 2: Percent increase in MSE for select important variables in White models before (A) and after (B) removal of potential outlier counties with White population proportion > 0.90 and $HEI < 0.50$. Percent increase in MSE was calculated with the randomForest R package using a permutation method in which larger increases in MSE indicate variables more important to the overall model performance.
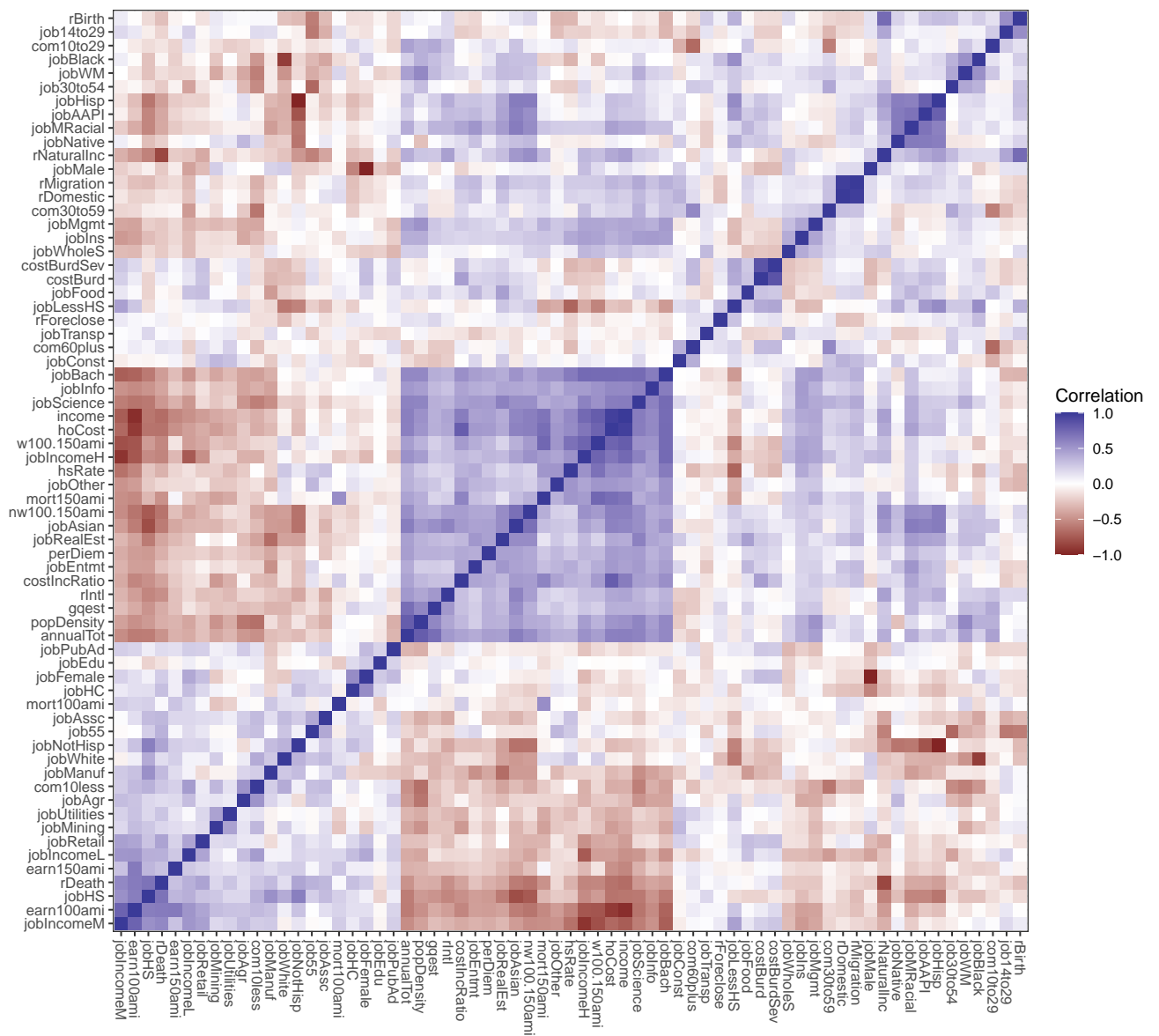
Figure 3: Spearman's correlations across numeric predictor variables in model. Order of variables was determined by complete-linkage (agglomerative) hierarchical clustering performed on Euclidean distances between correlations across features.
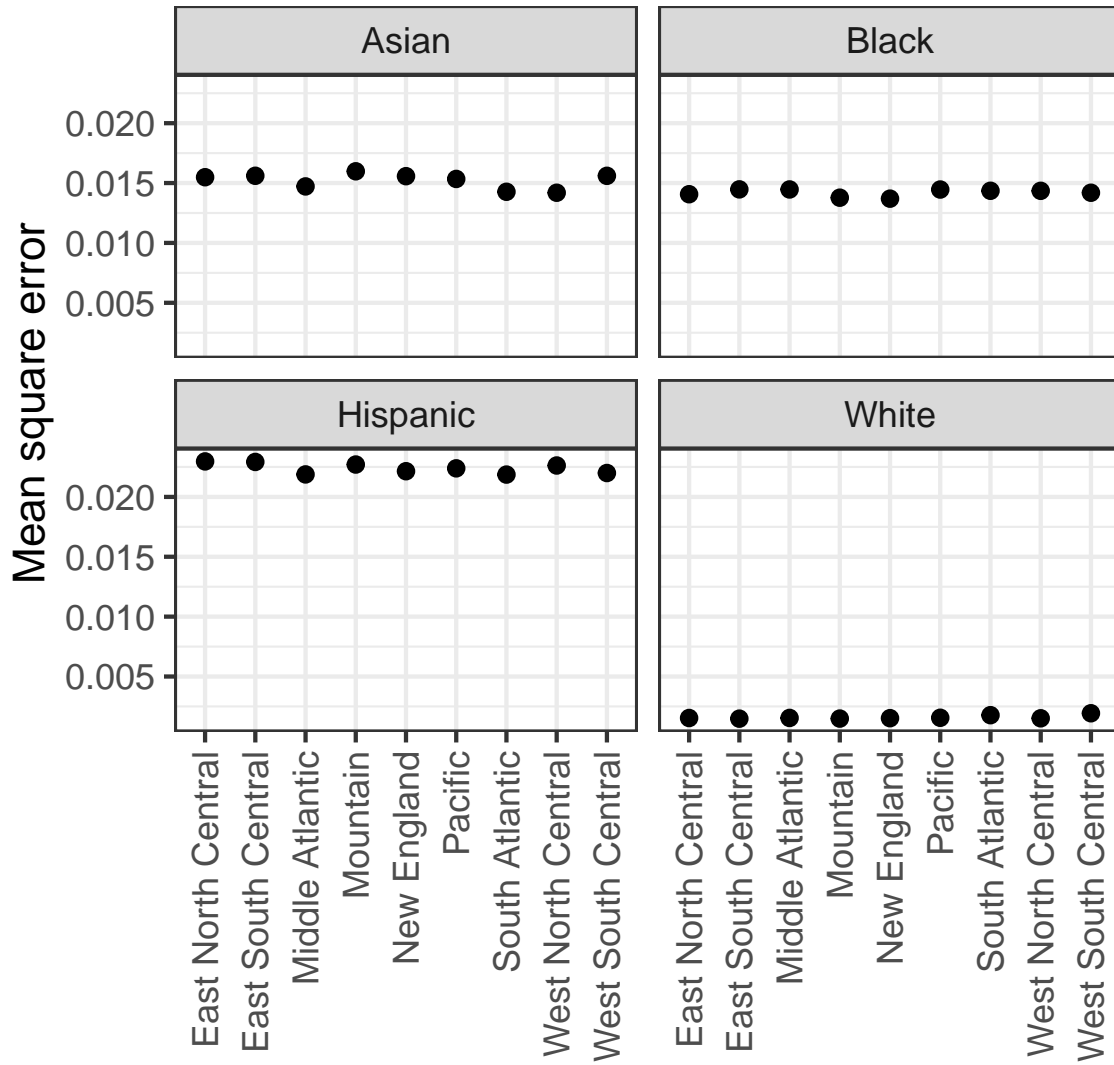
Figure 4: Performance of cross-validation models by holdout region on 10% holdout data.