Challenges addressed

* Ratio of homeownership to racial representation in county population
* Inconsistent reporting of Race
* Inconsistent reporting of county names
* Hispanic as a race vs ethnicity
* Low population areas with missing or underrepresented demographics
* “Hotspots” in the data across several metrics
* ”Other” category multi-faceted
* 2+ races vs. inclusion-based estimate of population
* Strong variable correlations
* Many possible interaction effects

Completed

* Decent model for White population
* Ok model for Black population
* Not great models for Asian or Hispanic populations
* Not enough data to do rural vs non rural
* Best current models seem to be with at least 500 folks of a race in that county

Steps for forward progress

* What interaction effects to add
* Ways to improve the model
  + Tuning - RR
    - By trees
    - By variable selection
  + More variables – DL A-N, AP N-Z
    - HO
      * Size of household
      * N homes
      * N homes owned
      * N building permits
    - Wellbeing
      * Poverty
    - Wealth distribution
* Possible to create single model with interaction effects?
  + Single model with RACE as a factor and 500 cut-off sits at about 68%, likely favors populations better represented in split model since they have more data