1. Propensity Scores
   1. According to Clarke, Kenkel, & Rueda (2011), overadjusting for irrelevant covariates can create problems for interpreting propensity scores. Accordingly, I decided to use a stripped-down propensity model based on a subset of the variables. Output and a density plot of propensity estimates stratified by outcome are below.  
        
      Call:

glm(formula = qsmk ~ age + factor(sex) + factor(race), family = "binomial",

data = nhefs.final)

Deviance Residuals:

Min 1Q Median 3Q Max

-1.0824 -0.8079 -0.7051 1.3598 2.0738

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) -1.851517 0.226972 -8.157 3.42e-16 \*\*\*

age 0.022555 0.004698 4.801 1.58e-06 \*\*\*

factor(sex)1 -0.273412 0.112734 -2.425 0.0153 \*

factor(race)1 -0.465416 0.182533 -2.550 0.0108 \*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

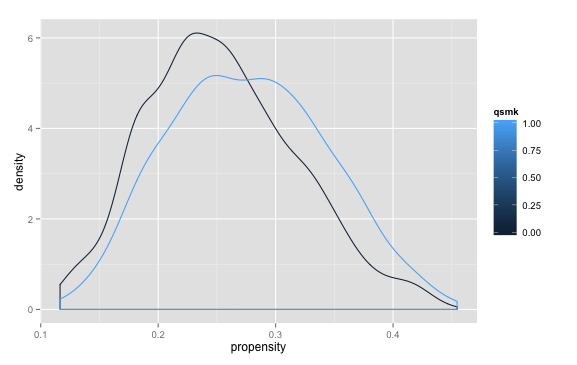
(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1925.3 on 1678 degrees of freedom

Residual deviance: 1887.7 on 1675 degrees of freedom

AIC: 1895.7

Number of Fisher Scoring iterations: 4



* 1. Stuff
  2. Propensity score estimates that predict the outcome “too perfectly” will have poor overlap, that is, the cells containing unusual outcomes for a particular set of predictors will be extremely small. Because of this, the estimates for the expected counterfactual given that set of covariates will be unreliable and misleading.

1. Project idea