

# Short-Report-3

Jeanny Zhang, Nghi Lam

November 7, 2021

## Introduction

In 1986, the Supreme Court ruled in the *Batson v. Kentucky* case that it is unconstitutional to exclude people from a jury based on their race or gender. However, the APM analysis of the juror selection phase in the Fifth Court District of Mississippi revealed wide racial disparities in how prosecutors and defense attorneys use their peremptory strikes. The data set used in this study was distilled from the data sets from the APM Reports podcast “In the Dark”, which contains court records on 2295 potential jurors for 89 jury trials in Mississippi that District Attorney Doug Evans had prosecuted between 1992 and 2017.

The goals of this analysis are (1) verify the findings of the APM methodology report that claims that whether a jury has prior familiarity with the victim or the witness (`know_vic` and `know_wit`), has family or close friend that has been the victim of a crime (`fam_crime_victim`), had been the victim of a crime (`crime_victim`), had prior information on the case (`prior_info`), and gender (`gender`) have no significant effect on whether they would be struck from jury duty, after accounting for whether the juror has been accused of being involved in criminal activity (`accused`), is black (`is_black`), has family or close friend who was accused of being involved in criminal activity (`fam_accused`), expressed reservations about imposing the death penalty (`death_hesitation`), has prior familiarity with the defendant (`know_def`), was the same race as the defendant (`same_race`), and has family or close friend in law enforcement (`fam_law_enforcement`); (2) determine whether the defendant’s race (`def_race`) is a better predictor than the juror and defendant being of the same race (`same_race`); and (3) if there is significant interaction between the juror’s race (`is_black`) and any of the 6 significant predictors mentioned in (1).

## Results

### Verify the findings of the APM methodology report

Our data set contains racial information on 2295 potential jurors, their relationship to the trial and judicial system, and whether they were struck by the state. All variables are categorical. The initial EDA of the data set reveals a higher proportion of black juror getting struck by the state (around 50%) than the white juror (around 12%), and the probability of black juror getting struck by the state is approximately 53% while that of white juror getting struck by the state is about 11%.

In order to find a good logistic binary regression model to verify the APM study findings, we first fitted a “rich” model with all 13 variables. According to the model summary, only 8 of the predictors were statistically significant, including `accused`, `is_black`, `fam_accused`, `death_hesitation`, `know_def`, `same_race`, `fam_law_enforcement`, and `fam_crime_victim`. The model satisfied all model assumptions, and no influential case was observed. The VIFs of the model did not suggest any collinearity between variables. A model with the 8 statistically significant variables was fitted, and the term `fam_crime_victim` was not statistically significant anymore.

Next, ANOVA tests were conducted. The term `fam_crime_victim` was confirmed to be statistically insignificant after doing an ANOVA test comparing a model with the other seven predictors and the “rich” model. Finally, the 7 predictors we selected for our regression model are `accused`, `is_black`, `fam_accused`, `death_hesitation`, `know_def`, `same_race`, and `fam_law_enforcement`.

The residual plots and the case influence statistics suggested that the model met the assumptions and there was no influential case. The largest Cook’s distance was approximately 0.03, which was too small to be considered influential. Additionally, removing this case did not affect the model significantly. The VIFs of the model did not suggest any collinearity between variables. The binary logistic model we found for modeling the likelihood of a juror being struck by the state is shown below.

	Estimate	SE	z value	p value
Intercept	-2.4307	0.1012	-24.017	<2e-16
accusedTRUE	2.5128	0.5455	4.606	4.10e-06
is_blackTRUE	1.8972	0.1411	13.443	<2e-16
fam_accusedTRUE	1.8476	0.162	11.402	<2e-16
death_hesitationTRUE	1.8243	0.5916	3.084	0.002044
know_defTRUE	1.3257	0.2233	5.937	2.91e-09
same_raceTRUE	0.3603	0.1399	2.575	0.010036
fam_law_enforcementTRUE	-0.5627	0.1622	-3.468	0.000524

Table 1. Model estimates, standard error, z- and p-values of the probability of a juror being struck by the states on `accused`, `is_black`, `fam_accused`, `death_hesitation`, `know_def`, `same_race`, and `fam_law_enforcement`.

Our logistic regression model matched the model proposed in the APM methodology paper. All the terms, coefficients, odds ratios and their confidence intervals were the same between the two models.

#### Explore whether `same_race` is a better predictor than `defendant_race`

The question of whether `defendant_race` was a better predictor than `same_race` in the model was raised. A model with the selected 7 predictors was fitted again except that `same_race` was replaced with `defendant_race`. According to the model summary, the p-values of all levels of `defendant_race` were larger than 0.05 which indicate that they are statistically insignificant. This was enough evidence to remove `defendant_race` from our model, and `same_race` is a better predictor than `defendant_race`.

#### Interactions between a juror’s race and the other predictors

Next, we determined if there was any interaction between a juror’s race (`is_black`) and the other 6 predictors in our model. We first did an EDA by plotting boxplots of `struck_state` against `is_black` faceted with each of the other 6 predictors. Although the EDA showed that the effect of each predictor on whether the juror got struck by the state did not seem to differ between races, there were some situations where the gap in the proportions of jurors getting struck between the two races seems too wide, which were worth further investigation.

Models with interaction terms were created for all 6 of the predictors. The statistically significant interaction terms were `is_black × fam_accused` and `is_black × know_def`. The outliers of the resulting interaction model were checked, and there were no influential cases. The VIFs of the model suggest a slight collinearity between `know_def` and `is_black × know_def`, but it was not a big concern in this case. The following barplots display how the effect of juror race on the odds of getting struck by the state was affected by `fam_accused` and `know_def`.



Figure 1. The effect of juror’s race on the odds of getting struck by the state was affected by `fam_accused` and `know_def`

Our final model for the APM study is:

$$\begin{aligned} \text{logit}(\hat{\text{struck\_state}}) = & -2.327 + 2.566\text{accused} + 1.737\text{is\_black} + 1.573\text{fam\_accused} + 1.824\text{death\_hesitation} \\ & + 0.687\text{now\_def} + 0.317\text{same\_race} - 0.548\text{fam\_law\_enforcement} \\ & + 0.812\text{is\_black} \times \text{fam\_accused} + 1.162\text{is\_black} \times \text{know\_def} \end{aligned}$$

## Discussion

In conclusion, our first analysis verified the findings of the APM methodology report that claims that whether a jury has prior familiarity with the victim or the witness, has family or close friend that has been the victim of a crime, had been the victim of a crime, had prior information on the case, and gender have no significant effect on whether they would be struck from jury duty, after accounting for whether the juror has been accused of being involved in criminal activity, is black, has family or close friend who was accused of being involved in criminal activity, expressed reservations about imposing the death penalty, has prior familiarity with the defendant, is the same race as the defendant, and has family or close friend in law enforcement. Interestingly, our second analysis suggested that the juror’s race is a better predictor for whether they were struck than the juror and defendant being of the same race. Finally, our third analysis revealed that the interactions between the juror’s race and whether the juror has family or close friend who was accused of being involved in criminal activity, has prior familiarity with the defendant, and being of the same race as the defendant have a significant effect on whether they were struck by the state from jury duty.

Based on our final model, the effect of being black on the odds of getting struck by the state is dependent on `fam_accused` and `know_def`. The juror being black is associated with a  $e^{1.737+0.812\text{fam\_accused}+1.162\text{know\_def}}$  multiplicative change in the odds of getting struck by the state after accounting for all other predictors. If the juror has no family or close friend who was accused of crime, has no prior familiarity with the defendant, and is not the same race as the defendant, the juror being black is associated with a 468% increase in the odds of getting struck by the state, holding all other predictors fixed (95% CI 325% to 663%). It is also noteworthy that the term `know_def` is not statistically significant in our final model, meaning that the effect of `know_def` on the odds of getting struck is only significant for the black jurors but not the white jurors.

One of the limitations of the data set is the thirteen jurors with unknown race who were eventually categorized as white. If any of these jurors were actually black and the data point is an influential outlier, there might be a problem in our model.

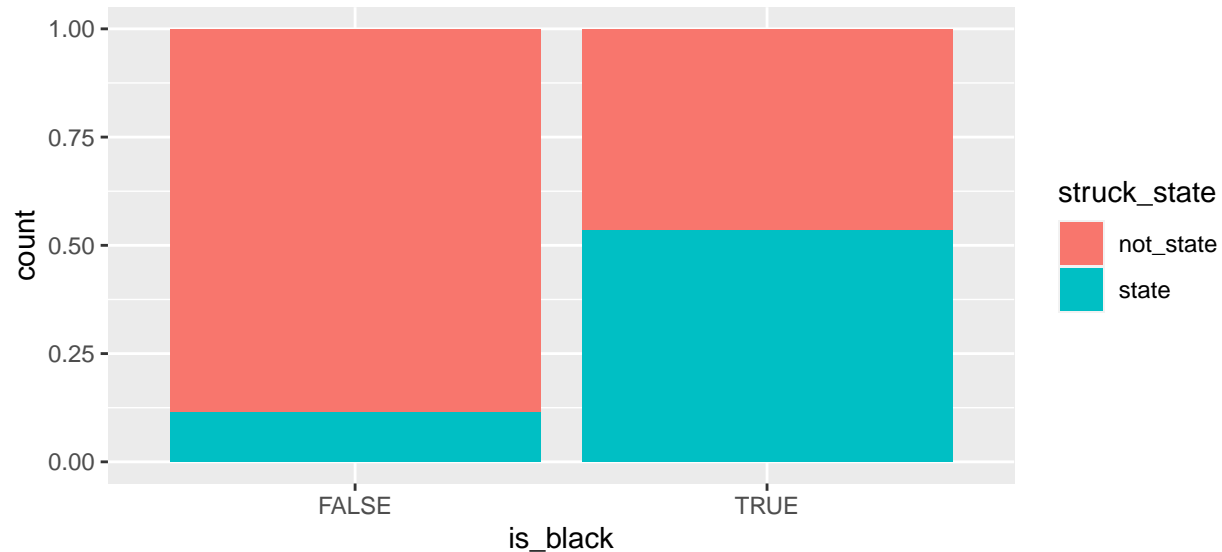
```
# load apm data frame
apm <- read.csv("http://math.carleton.edu/kstclair/data/APM_245report3.csv")
# EDA
summary(apm)
```

```
tapply(apm$struck_state_bin, apm$is_black, summary)
```

4

```
##
## $`TRUE`
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0000 0.0000  1.0000  0.5344  1.0000  1.0000
```

```
ggplot(apm, aes(x=is_black, fill = struck_state)) +
  geom_bar(position = "fill")
```



```
# model with variables from group A
```

```
apm_a_predictors <- glm(struck_state_bin~accused+is_black+fam_accused+death_hesitation+know_def+same_race+
  fam_law_enforcement, data = apm)
summary(apm_a_predictors)
```

```
##
## Call:
## glm(formula = struck_state_bin ~ accused + is_black + fam_accused +
##      death_hesitation + know_def + same_race + fam_law_enforcement,
##      family = "binomial", data = apm)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.4693  -0.4874  -0.4107  -0.3127   2.4667
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -2.4307     0.1012  -24.017  < 2e-16 ***
## accusedTRUE      2.5128     0.5455   4.606 4.10e-06 ***
## is_blackTRUE     1.8972     0.1411  13.443  < 2e-16 ***
## fam_accusedTRUE  1.8476     0.1620  11.402  < 2e-16 ***
## death_hesitationTRUE 1.8243     0.5916   3.084 0.002044 **
## know_defTRUE     1.3257     0.2233   5.937 2.91e-09 ***
## same_raceTRUE    0.3603     0.1399   2.575 0.010036 *
## fam_law_enforcementTRUE -0.5627     0.1622  -3.468 0.000524 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
```

```

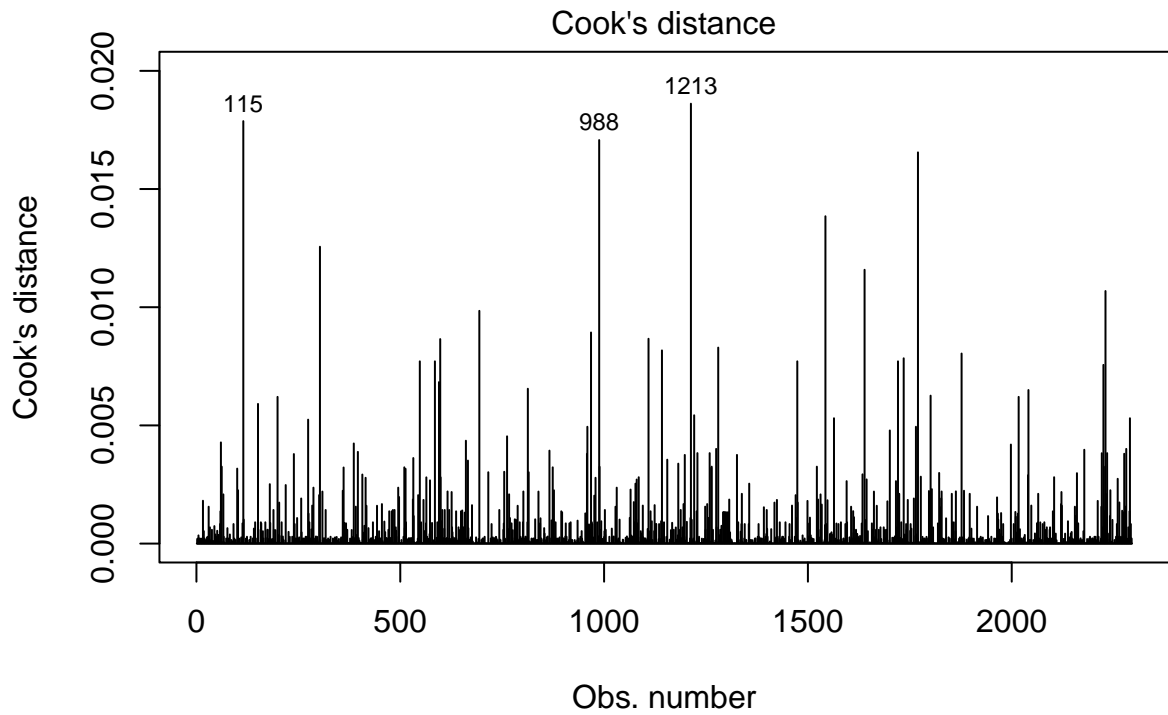
##      Null deviance: 2579.5   on 2294   degrees of freedom
## Residual deviance: 1887.6   on 2287   degrees of freedom
## AIC: 1903.6
##
## Number of Fisher Scoring iterations: 5
# model with all variables
apm_all_predictors <- glm(struck_state_bin~accused+is_black+fam_accused+death_hesitation+know_def+same_race+
summary(apm_all_predictors)

##
## Call:
## glm(formula = struck_state_bin ~ accused + is_black + fam_accused +
##      death_hesitation + know_def + same_race + fam_law_enforcement +
##      know_vic + know_wit + fam_crime_victim + crime_victim + prior_info +
##      gender, family = "binomial", data = apm)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.4903  -0.4904  -0.4069  -0.2553   2.4813
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -2.449611    0.120825 -20.274 < 2e-16 ***
## accusedTRUE       2.539562    0.545295  4.657 3.21e-06 ***
## is_blackTRUE      1.896292    0.143144 13.247 < 2e-16 ***
## fam_accusedTRUE   1.860810    0.164490 11.313 < 2e-16 ***
## death_hesitationTRUE 1.893782    0.595388  3.181 0.001469 **
## know_defTRUE      1.355257    0.229931  5.894 3.77e-09 ***
## same_raceTRUE     0.374871    0.140993  2.659 0.007842 **
## fam_law_enforcementTRUE -0.581547    0.164103 -3.544 0.000394 ***
## know_vicTRUE      0.293692    0.237816  1.235 0.216846
## know_witTRUE      -0.286390    0.231691 -1.236 0.216426
## fam_crime_victimTRUE 0.541522    0.274917  1.970 0.048865 *
## crime_victimTRUE  0.009626    0.310544  0.031 0.975272
## prior_infoTRUE    -0.183799    0.194147 -0.947 0.343793
## genderMale        0.017205    0.121666  0.141 0.887545
## genderUnknown     0.022099    0.369050  0.060 0.952251
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 2579.5   on 2294   degrees of freedom
## Residual deviance: 1880.7   on 2280   degrees of freedom
## AIC: 1910.7
##
## Number of Fisher Scoring iterations: 5
# model with just statistically significant variables
apm_sig_predictors <- glm(struck_state_bin~accused+is_black+fam_accused+death_hesitation+know_def+same_race+
summary(apm_sig_predictors)

##
## Call:
## glm(formula = struck_state_bin ~ accused + is_black + fam_accused +

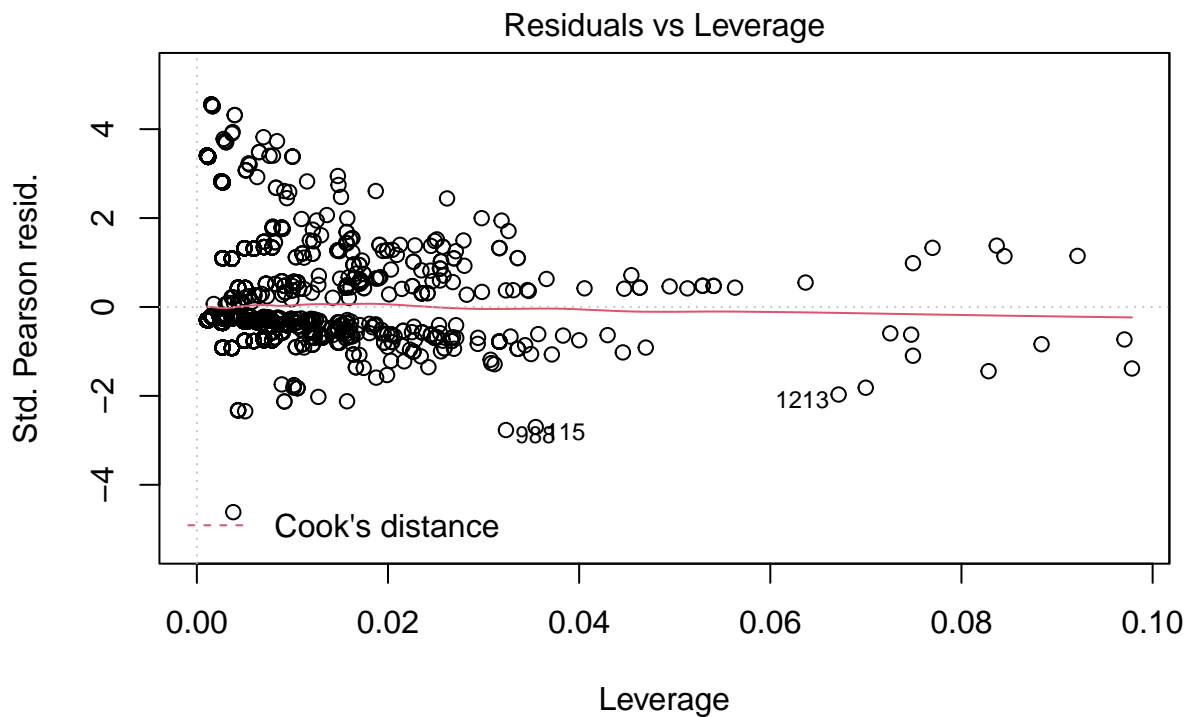
```

```
##      death_hesitation + know_def + same_race + fam_law_enforcement +
##      fam_crime_victim, family = "binomial", data = apm)
##
## Deviance Residuals:
##      Min        1Q      Median        3Q        Max
## -2.4615  -0.4839  -0.4054  -0.3040   2.4890
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -2.4574     0.1027  -23.927 < 2e-16 ***
## accusedTRUE        2.4994     0.5431   4.602 4.19e-06 ***
## is_blackTRUE       1.9037     0.1413  13.476 < 2e-16 ***
## fam_accusedTRUE    1.8373     0.1621  11.336 < 2e-16 ***
## death_hesitationTRUE 1.8487     0.5938   3.113 0.001851 **
## know_defTRUE       1.3247     0.2244   5.904 3.54e-09 ***
## same_raceTRUE      0.3717     0.1401   2.653 0.007976 **
## fam_law_enforcementTRUE -0.5939     0.1633  -3.637 0.000276 ***
## fam_crime_victimTRUE 0.5304     0.2724   1.947 0.051543 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 2579.5  on 2294  degrees of freedom
## Residual deviance: 1883.9  on 2286  degrees of freedom
## AIC: 1901.9
##
## Number of Fisher Scoring iterations: 5
# check outliers
plot(apm_all_predictors, which=4)
```



```
glm(struck_state_bin ~ accused + is_black + fam_accused + death_hesitation ...
```

```
plot(apm_all_predictors, which=5)
```



```
glm(struck_state_bin ~ accused + is_black + fam_accused + death_hesitation ...
```

```
# slice out case 1213
apm_all_predictors_aug <- augment(apm_all_predictors)
apm_all_predictors_aug %>% slice(1213)
```



```

## # A tibble: 1 x 20
##   struck_state_bin accused is_black fam_accused death_hesitation know_def
##   <int> <lgl> <lgl> <lgl> <lgl> <lgl>
## 1      0 TRUE  FALSE  FALSE  FALSE  TRUE
## # ... with 14 more variables: same_race <lgl>, fam_law_enforcement <lgl>,
## #   know_vic <lgl>, know_wit <lgl>, fam_crime_victim <lgl>, crime_victim <lgl>,
## #   prior_info <lgl>, gender <chr>, .fitted <dbl>, .resid <dbl>,
## #   .std.resid <dbl>, .hat <dbl>, .sigma <dbl>, .cooks_d <dbl>

# refit model without case 1213
apm_all_predictors_no_1213 <- glm(struck_state_bin~accused+is_black+fam_accused+death_hesitation+know_d
summary(apm_all_predictors_no_1213)

##
## Call:
## glm(formula = struck_state_bin ~ accused + is_black + fam_accused +
##   death_hesitation + know_def + same_race + fam_law_enforcement +
##   know_vic + know_wit + fam_crime_victim + crime_victim + prior_info +
##   gender, family = "binomial", data = apm, subset = -c(1213))
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.4975  -0.4889  -0.4065  -0.2524   2.4856
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -2.45208    0.12095  -20.274 < 2e-16 ***
## accusedTRUE      2.79510    0.59570   4.692 2.70e-06 ***
## is_blackTRUE     1.89568    0.14329  13.230 < 2e-16 ***
## fam_accusedTRUE  1.85779    0.16454  11.290 < 2e-16 ***
## death_hesitationTRUE 1.88089    0.59591   3.156 0.00160 **
## know_defTRUE     1.38400    0.23078   5.997 2.01e-09 ***
## same_raceTRUE    0.36961    0.14110   2.619 0.00881 **
## fam_law_enforcementTRUE -0.59037    0.16444  -3.590 0.00033 ***
## know_vicTRUE     0.32142    0.23747   1.354 0.17588
## know_witTRUE    -0.25890    0.23119  -1.120 0.26278
## fam_crime_victimTRUE 0.53336    0.27539   1.937 0.05277 .
## crime_victimTRUE  0.00524    0.31087   0.017 0.98655
## prior_infoTRUE   -0.17453    0.19418  -0.899 0.36874
## genderMale       0.01864    0.12175   0.153 0.87831
## genderUnknown    0.02576    0.36901   0.070 0.94436
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 2578.9  on 2293  degrees of freedom
## Residual deviance: 1877.3  on 2279  degrees of freedom
## AIC: 1907.3
##
## Number of Fisher Scoring iterations: 5

# check collinearity
vif(apm_all_predictors)

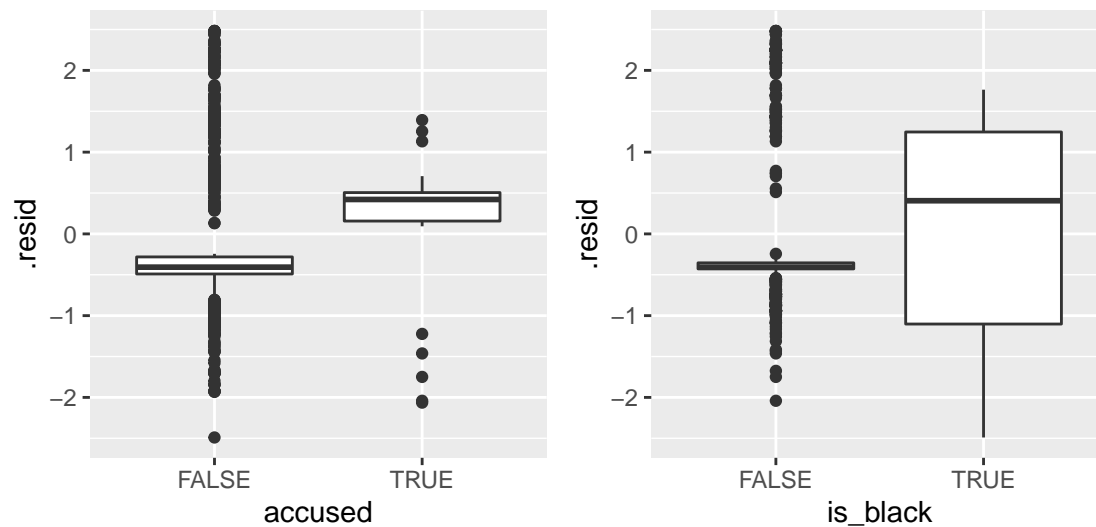
##              GVIF Df GVIF^(1/(2*Df))

```

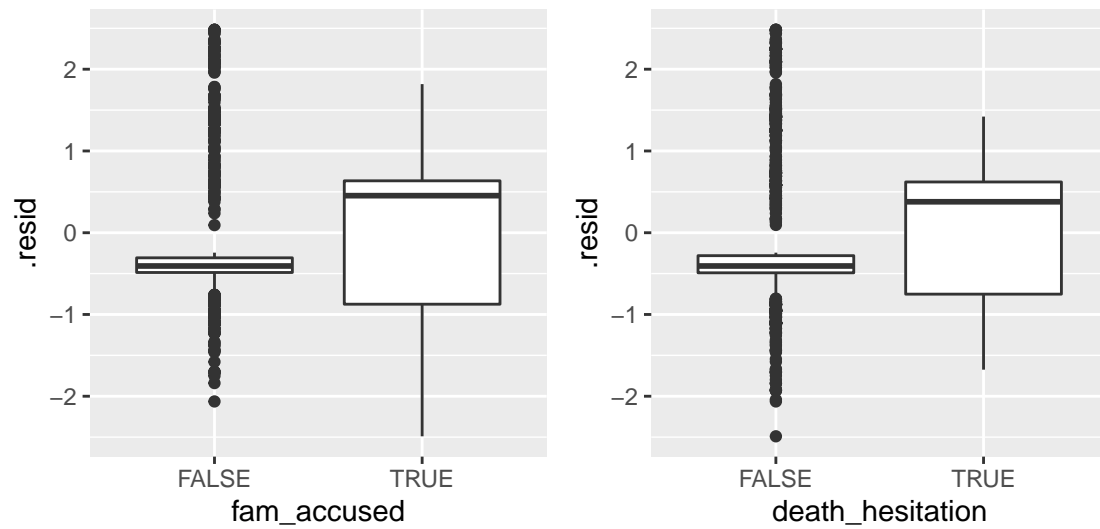
```
## accused          1.008066  1      1.004025
## is_black         1.510719  1      1.229113
## fam_accused      1.077727  1      1.038136
## death_hesitation 1.034042  1      1.016878
## know_def         1.067495  1      1.033197
## same_race        1.468510  1      1.211821
## fam_law_enforcement 1.057932  1      1.028558
## know_vic         1.198553  1      1.094784
## know_wit         1.063579  1      1.031300
## fam_crime_victim  1.035862  1      1.017773
## crime_victim     1.015931  1      1.007934
## prior_info       1.223739  1      1.106228
## gender           1.042612  2      1.010487
```

```
# check residuals
```

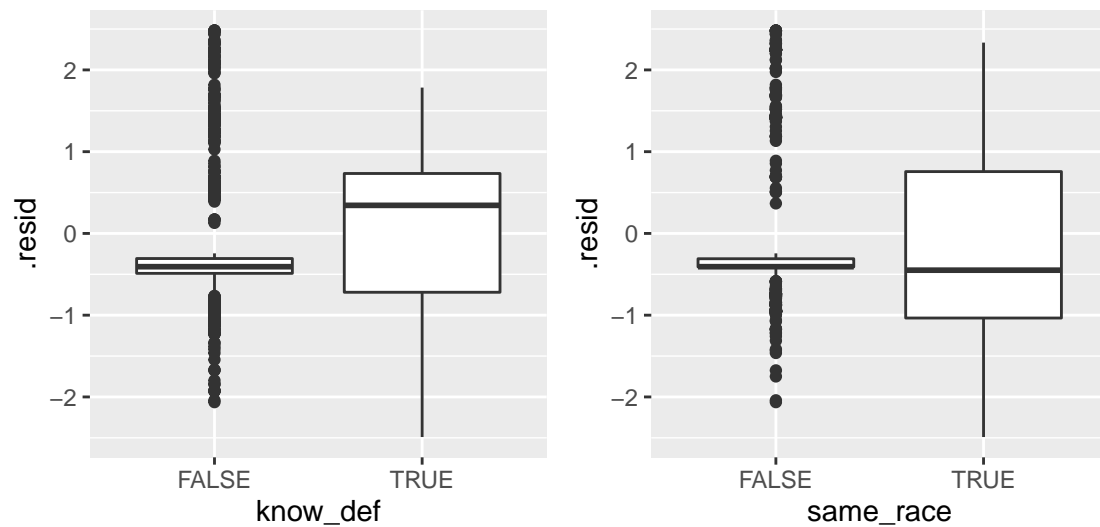
```
resid1 <- ggplot(apm_all_predictors_aug, aes(x=accused, y=.resid)) + geom_boxplot()
resid2 <- ggplot(apm_all_predictors_aug, aes(x=is_black, y=.resid)) + geom_boxplot()
resid1+resid2
```



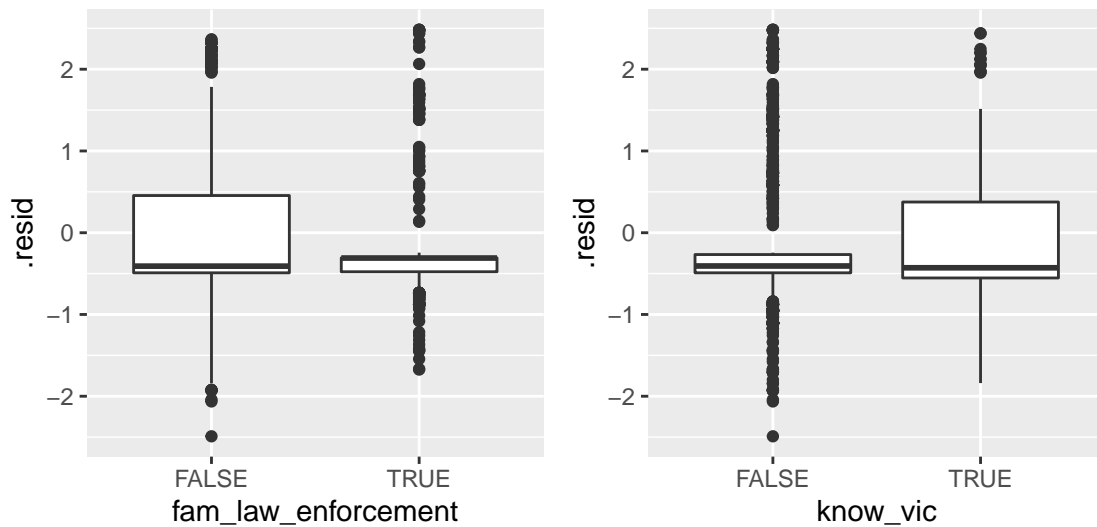
```
resid3 <- ggplot(apm_all_predictors_aug, aes(x=fam_accused, y=.resid)) + geom_boxplot()
resid4 <- ggplot(apm_all_predictors_aug, aes(x=death_hesitation, y=.resid)) + geom_boxplot()
resid3+resid4
```



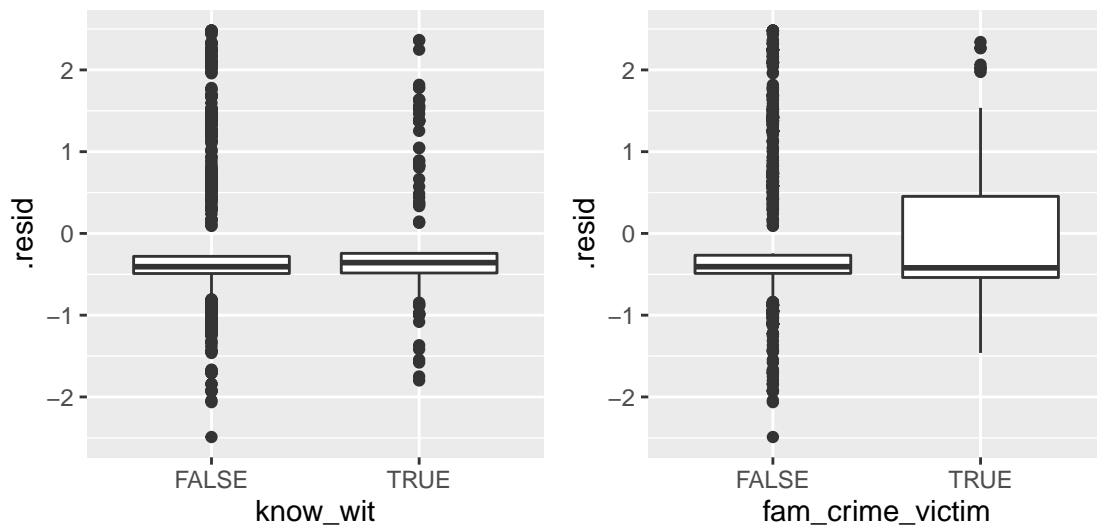
```
resid5 <- ggplot(apm_all_predictors_aug, aes(x=know_def, y=.resid)) + geom_boxplot()
resid6 <- ggplot(apm_all_predictors_aug, aes(x=same_race, y=.resid)) + geom_boxplot()
resid5+resid6
```



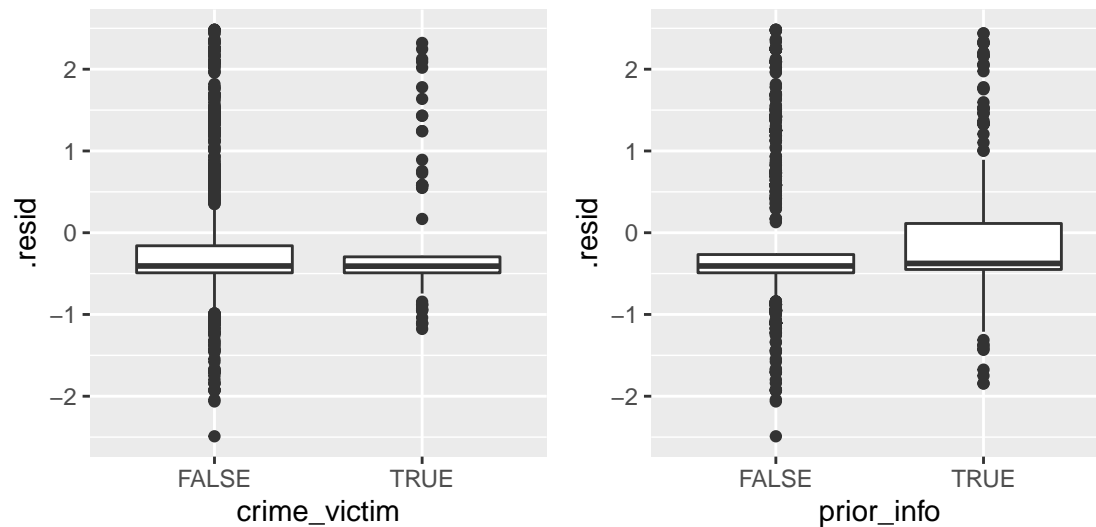
```
resid7 <- ggplot(apm_all_predictors_aug, aes(x=fam_law_enforcement, y=.resid)) + geom_boxplot()
resid8 <- ggplot(apm_all_predictors_aug, aes(x=know_vic, y=.resid)) + geom_boxplot()
resid7+resid8
```



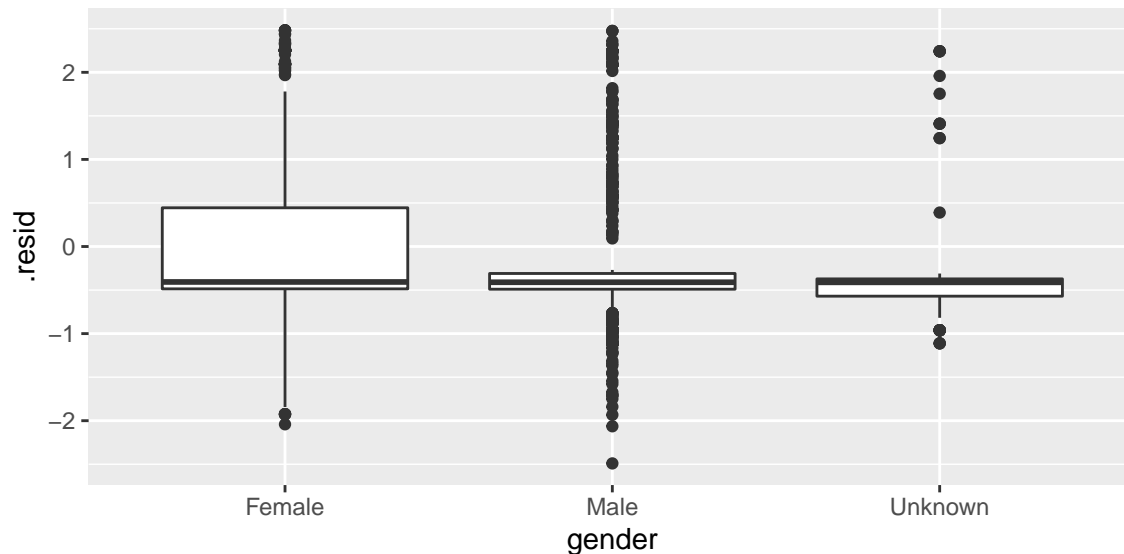
```
resid9 <- ggplot(apm_all_predictors_aug, aes(x=know_wit, y=.resid)) + geom_boxplot()
resid10 <- ggplot(apm_all_predictors_aug, aes(x=fam_crime_victim, y=.resid)) + geom_boxplot()
resid9+resid10
```



```
resid11 <- ggplot(apm_all_predictors_aug, aes(x=crime_victim, y=.resid)) + geom_boxplot()
resid12 <- ggplot(apm_all_predictors_aug, aes(x=prior_info, y=.resid)) + geom_boxplot()
resid11+resid12
```



```
resid13 <- ggplot(apm_all_predictors_aug, aes(x=gender, y=.resid)) + geom_boxplot()
resid13
```



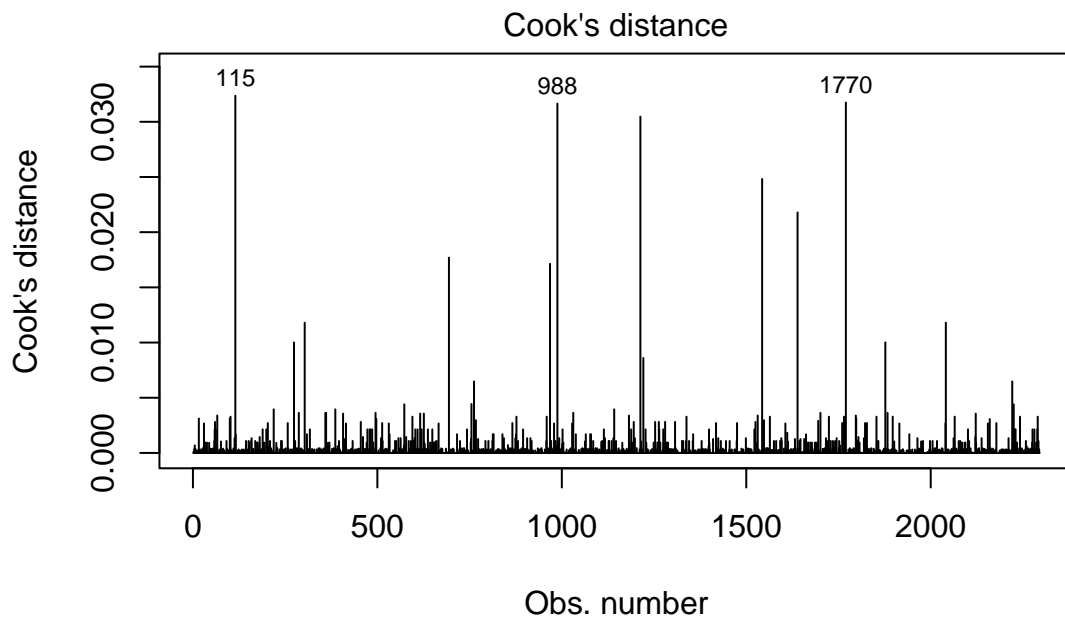
```
# anova tests
anova(apm_sig_predictors, apm_all_predictors, test="Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: struck_state_bin ~ accused + is_black + fam_accused + death_hesitation +
##   know_def + same_race + fam_law_enforcement + fam_crime_victim
## Model 2: struck_state_bin ~ accused + is_black + fam_accused + death_hesitation +
##   know_def + same_race + fam_law_enforcement + know_vic + know_wit +
##   fam_crime_victim + crime_victim + prior_info + gender
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1      2286      1883.9
## 2      2280      1880.7  6   3.2579  0.7758
```

```
anova(apm_a_predictors, apm_all_predictors, test="Chisq")
```

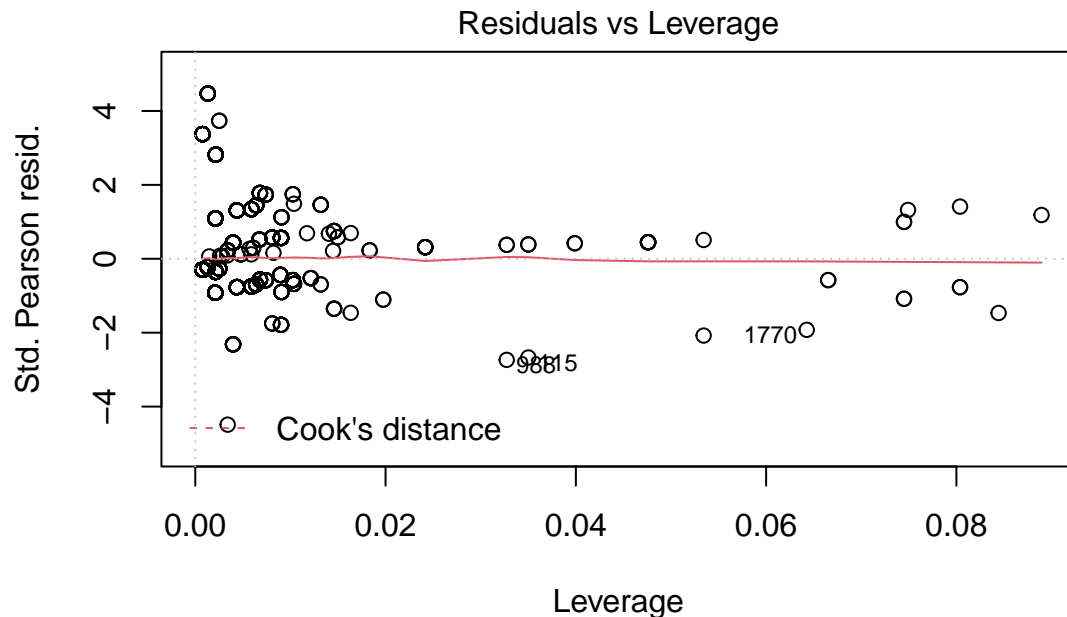
```
## Analysis of Deviance Table
```

```
##
## Model 1: struck_state_bin ~ accused + is_black + fam_accused + death_hesitation +
##   know_def + same_race + fam_law_enforcement
## Model 2: struck_state_bin ~ accused + is_black + fam_accused + death_hesitation +
##   know_def + same_race + fam_law_enforcement + know_vic + know_wit +
##   fam_crime_victim + crime_victim + prior_info + gender
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1      2287      1887.5
## 2      2280      1880.7  7   6.8792  0.4416
# check outliers for the smaller model
plot(apm_a_predictors, which=4)
```



```
glm(struck_state_bin ~ accused + is_black + fam_accused + death_hesitation
```

```
plot(apm_a_predictors, which=5)
```



```
glm(struck_state_bin ~ accused + is_black + fam_accused + death_hesitation
```

```
# slice out case 1213
```

```
apm_a_predictors_aug <- augment(apm_a_predictors)
```

```
apm_a_predictors_aug %>% slice(115)
```

```
## # A tibble: 1 x 14
```

```
##   struck_state_bin accused is_black fam_accused death_hesitation know_def
```

```
##   <int> <lgl> <lgl> <lgl> <lgl> <lgl>
```

```
## 1 0 TRUE FALSE TRUE FALSE FALSE
```

```
## # ... with 8 more variables: same_race <lgl>, fam_law_enforcement <lgl>,
```

```
## # .fitted <dbl>, .resid <dbl>, .std.resid <dbl>, .hat <dbl>, .sigma <dbl>,
```

```
## # .cooksad <dbl>
```

```
# remove case 115
```

```
apm_a_predictors_no_115 <- glm(struck_state_bin~accused+is_black+fam_accused+death_hesitation+know_def+
```

```
summary(apm_a_predictors_no_115)
```

```
##
```

```
## Call:
```

```
## glm(formula = struck_state_bin ~ accused + is_black + fam_accused +
```

```
## death_hesitation + know_def + same_race + fam_law_enforcement,
```

```
## family = "binomial", data = apm, subset = -c(115))
```

```
##
```

```
## Deviance Residuals:
```

```
##   Min      1Q   Median      3Q      Max
```

```
## -2.4746 -0.4870 -0.4106 -0.3117  2.4693
```

```
##
```

```
## Coefficients:
```

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

```
## (Intercept)      -2.4308     0.1013  -23.995 < 2e-16 ***
```

```
## accusedTRUE       2.7979     0.5949   4.703 2.56e-06 ***
```

```
## is_blackTRUE      1.8950     0.1413  13.410 < 2e-16 ***
```

```
## fam_accusedTRUE   1.8677     0.1625  11.497 < 2e-16 ***
```

```
## death_hesitationTRUE 1.8243     0.5918   3.083 0.002052 **
```

```
## know_defTRUE      1.3234     0.2237   5.917 3.28e-09 ***
```

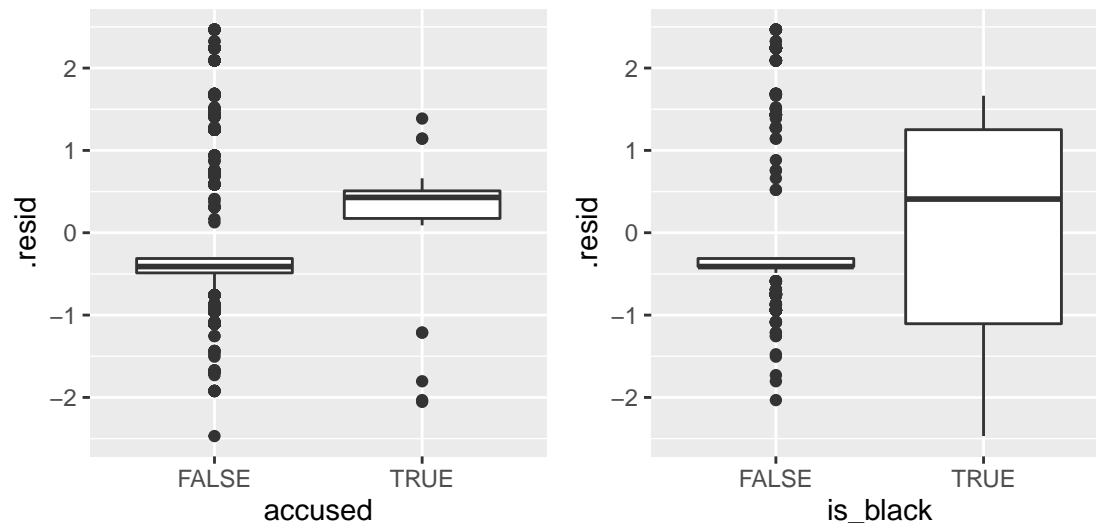
```
## same_raceTRUE          0.3586      0.1401    2.560 0.010468 *
## fam_law_enforcementTRUE -0.5695      0.1625   -3.503 0.000459 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 2578.9  on 2293  degrees of freedom
## Residual deviance: 1883.2  on 2286  degrees of freedom
## AIC: 1899.2
##
## Number of Fisher Scoring iterations: 5
```

```
# check collinearity
vif(apm_a_predictors)
```

```
##          accused          is_black          fam_accused          death_hesitation
##          1.001851          1.474367          1.052252          1.001176
##          know_def          same_race fam_law_enforcement
##          1.020771          1.452398          1.032267
```

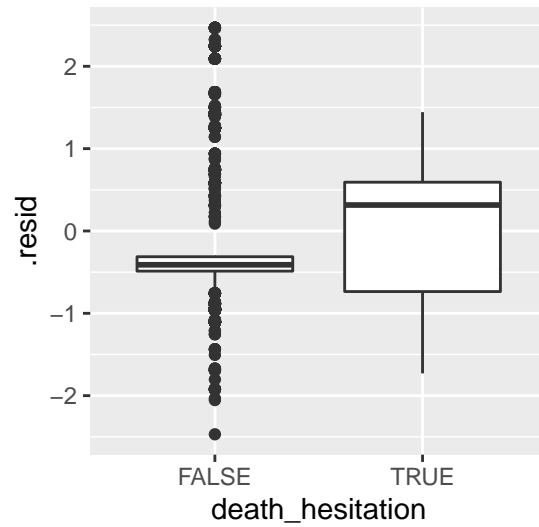
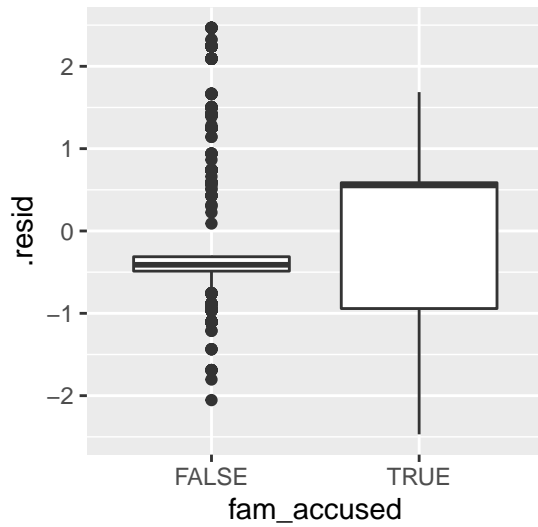
```
# check residuals
```

```
a_resid1 <- ggplot(apm_a_predictors_aug, aes(x=accused, y=.resid)) + geom_boxplot()
a_resid2 <- ggplot(apm_a_predictors_aug, aes(x=is_black, y=.resid)) + geom_boxplot()
a_resid1+a_resid2
```

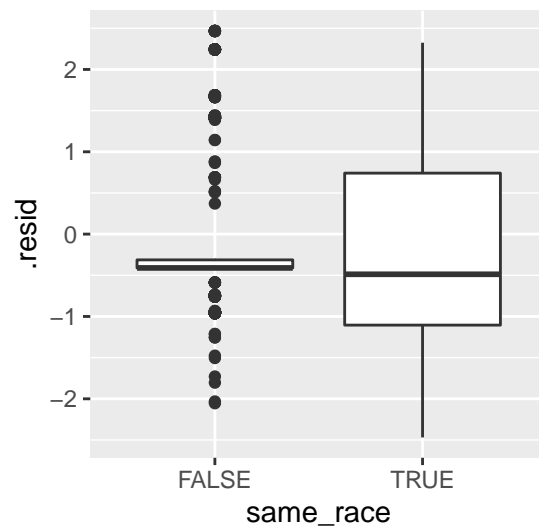
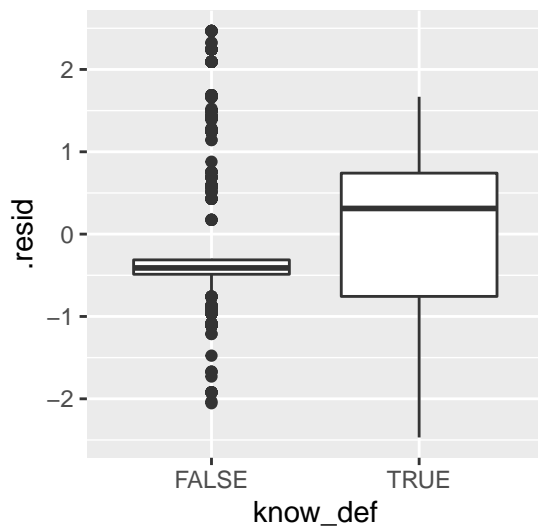


```
a_resid3 <- ggplot(apm_a_predictors_aug, aes(x=fam_accused, y=.resid)) + geom_boxplot()
a_resid4 <- ggplot(apm_a_predictors_aug, aes(x=death_hesitation, y=.resid)) + geom_boxplot()
a_resid3+a_resid4
```

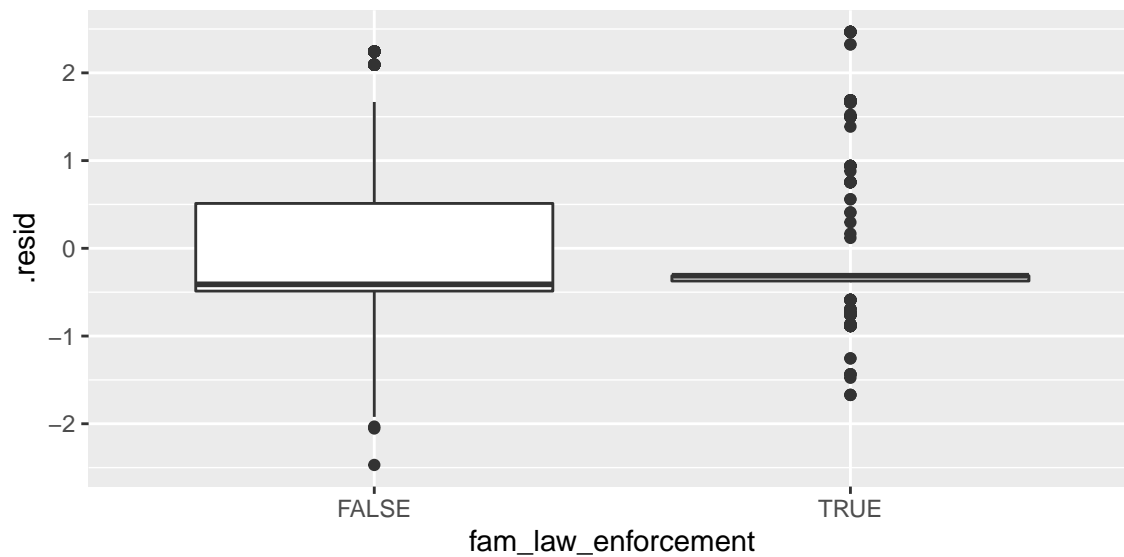




```
a_resid5 <- ggplot(apm_a_predictors_aug, aes(x=know_def, y=.resid)) + geom_boxplot()
a_resid6 <- ggplot(apm_a_predictors_aug, aes(x=same_race, y=.resid)) + geom_boxplot()
a_resid5+a_resid6
```



```
a_resid7 <- ggplot(apm_a_predictors_aug, aes(x=fam_law_enforcement, y=.resid)) + geom_boxplot()
a_resid7
```



```
# fit the model from methodology
summary(apm_a_predictors)
```

```
##
## Call:
## glm(formula = struck_state_bin ~ accused + is_black + fam_accused +
##      death_hesitation + know_def + same_race + fam_law_enforcement,
##      family = "binomial", data = apm)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.4693  -0.4874  -0.4107  -0.3127   2.4667
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -2.4307     0.1012 -24.017  < 2e-16 ***
## accusedTRUE      2.5128     0.5455   4.606 4.10e-06 ***
## is_blackTRUE     1.8972     0.1411  13.443  < 2e-16 ***
## fam_accusedTRUE  1.8476     0.1620  11.402  < 2e-16 ***
## death_hesitationTRUE 1.8243     0.5916   3.084 0.002044 **
## know_defTRUE     1.3257     0.2233   5.937 2.91e-09 ***
## same_raceTRUE    0.3603     0.1399   2.575 0.010036 *
## fam_law_enforcementTRUE -0.5627     0.1622  -3.468 0.000524 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 2579.5  on 2294  degrees of freedom
## Residual deviance: 1887.6  on 2287  degrees of freedom
## AIC: 1903.6
##
## Number of Fisher Scoring iterations: 5
tidy(apm_a_predictors, conf.int = TRUE)

## # A tibble: 8 x 7
```

##	term	estimate	std.error	statistic	p.value	conf.low	conf.high
##	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
## 1	(Intercept)	-2.43	0.101	-24.0	1.87e-127	-2.63	-2.24
## 2	accusedTRUE	2.51	0.545	4.61	4.10e- 6	1.52	3.69
## 3	is_blackTRUE	1.90	0.141	13.4	3.39e- 41	1.62	2.18
## 4	fam_accusedTRUE	1.85	0.162	11.4	4.07e- 30	1.53	2.17
## 5	death_hesitationTRUE	1.82	0.592	3.08	2.04e- 3	0.702	3.04
## 6	know_defTRUE	1.33	0.223	5.94	2.91e- 9	0.891	1.77
## 7	same_raceTRUE	0.360	0.140	2.57	1.00e- 2	0.0845	0.633
## 8	fam_law_enforcementTRUE	-0.563	0.162	-3.47	5.24e- 4	-0.887	-0.250

```
# accused
exp(2.5127795)
```

```
## [1] 12.33918
exp(1.52055098)
```

```
## [1] 4.574745
exp(3.6911472)
```

```
## [1] 40.09081
# is_black
exp(1.8971633)
```

```
## [1] 6.666955
exp(1.62292260)
```

```
## [1] 5.06788
exp(2.1764695)
```

```
## [1] 8.815129
# fam_accused
exp(1.8475979)
```

```
## [1] 6.344561
exp(1.53208764)
```

```
## [1] 4.627828
exp(2.1677100)
```

```
## [1] 8.738251
# death_hesitation
exp(1.8243438)
```

```
## [1] 6.198726
exp(0.70207619)
```

```
## [1] 2.017938
exp(3.0405746)
```

```
## [1] 20.91726
```

```

# know_def
exp(1.3256965)

## [1] 3.764807
exp(0.89080433)

## [1] 2.437089
exp(1.7671475)

## [1] 5.854131
# same_race
exp(0.3602561)

## [1] 1.433697
exp(0.08451597)

## [1] 1.08819
exp(0.6333615)

## [1] 1.883933
# fam_law_enforcement
exp(-0.5626784)

## [1] 0.5696812
exp(-0.88650361)

## [1] 0.4120941
exp(-0.2498507)

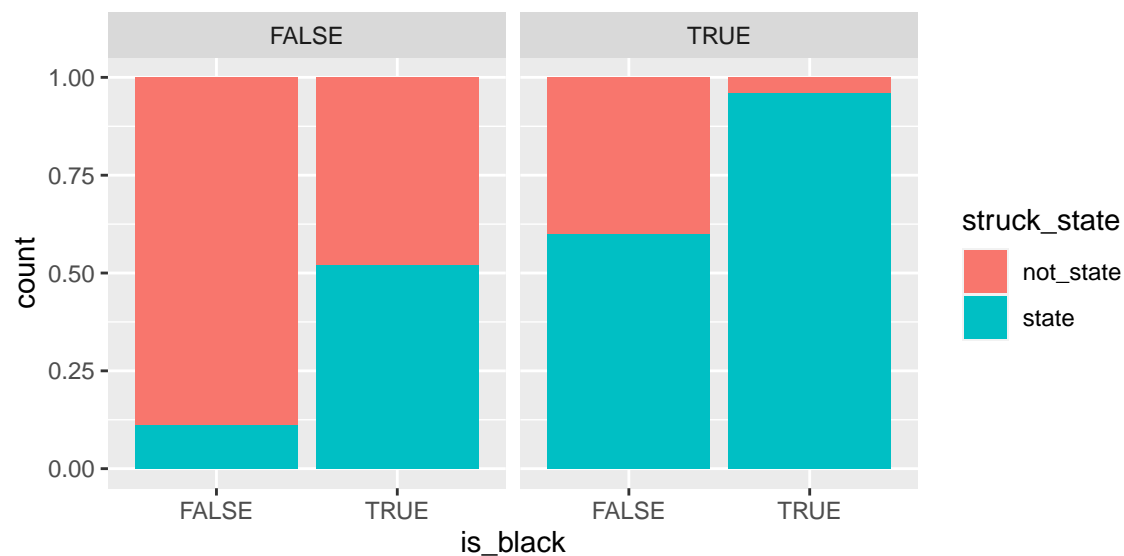
## [1] 0.7789171
# defendant_race vs. same_race
apm_a_predictors2 <- glm(struck_state_bin~accused+is_black+fam_accused+death_hesitation+know_def+defendant_race,
summary(apm_a_predictors2)

##
## Call:
## glm(formula = struck_state_bin ~ accused + is_black + fam_accused +
##      death_hesitation + know_def + defendant_race + fam_law_enforcement,
##      family = "binomial", data = apm)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.4549  -0.4557  -0.4075  -0.3108   2.4716
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -1.7172     0.5368  -3.199  0.001379 **
## accusedTRUE         2.5092     0.5471   4.586 4.51e-06 ***
## is_blackTRUE        2.1154     0.1186  17.837 < 2e-16 ***
## fam_accusedTRUE     1.8738     0.1619  11.576 < 2e-16 ***
## death_hesitationTRUE 1.9026     0.5857   3.248 0.001161 **
## know_defTRUE        1.4205     0.2209   6.429 1.28e-10 ***

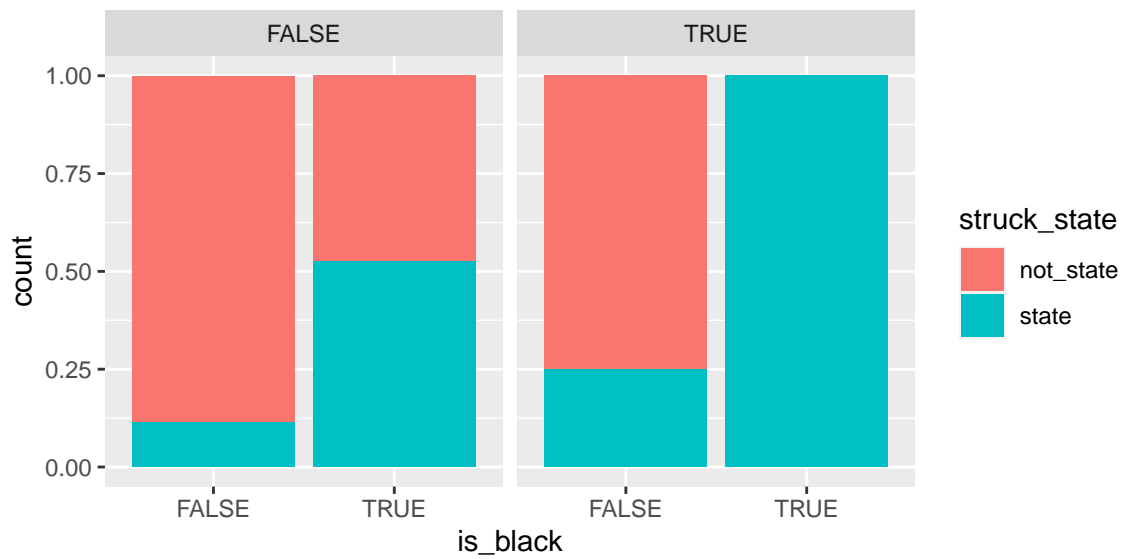
```

```
## defendant_raceBlack      -0.7297      0.5402   -1.351 0.176801
## defendant_raceUnknown    -0.1306      0.6575   -0.199 0.842523
## defendant_raceWhite      -0.4955      0.5497   -0.901 0.367357
## fam_law_enforcementTRUE  -0.5594      0.1618   -3.458 0.000545 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 2579.5  on 2294  degrees of freedom
## Residual deviance: 1888.1  on 2285  degrees of freedom
## AIC: 1908.1
##
## Number of Fisher Scoring iterations: 5
# defedant_race is not significant
```

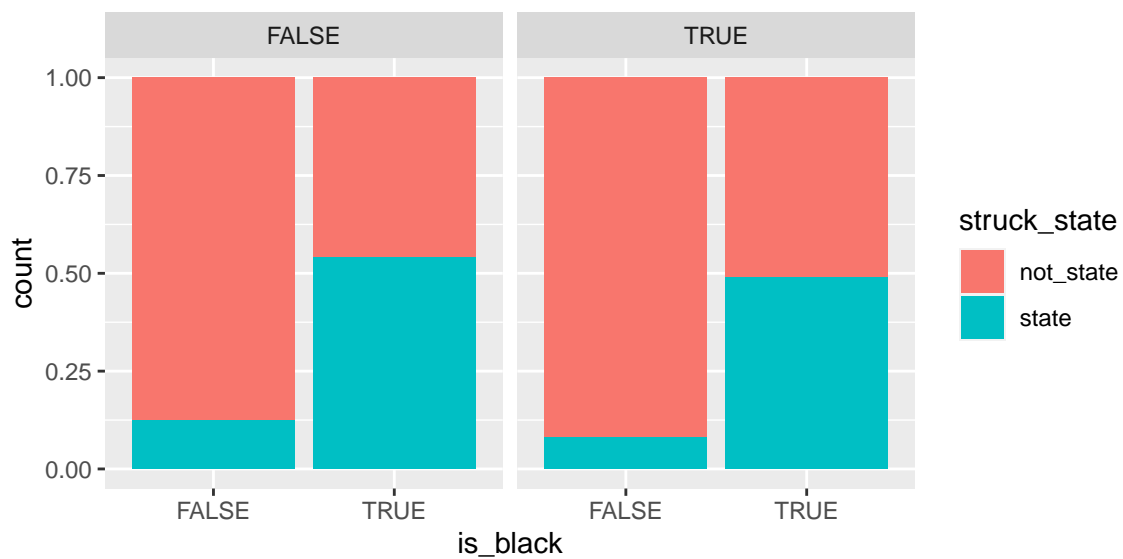
```
# interaction between is_black and other predictors
ggplot(apm, aes(x=is_black, fill=struck_state)) +
  geom_bar(position="fill") +
  facet_wrap(~accused)
```



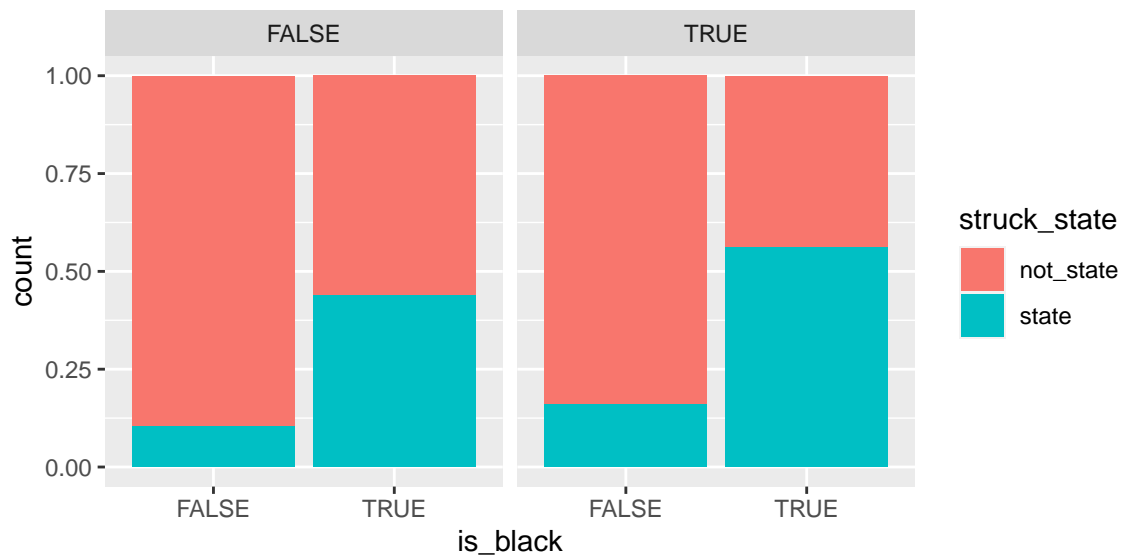
```
ggplot(apm, aes(x=is_black, fill=struck_state)) +
  geom_bar(position="fill") +
  facet_wrap(~death_hesitation)
```



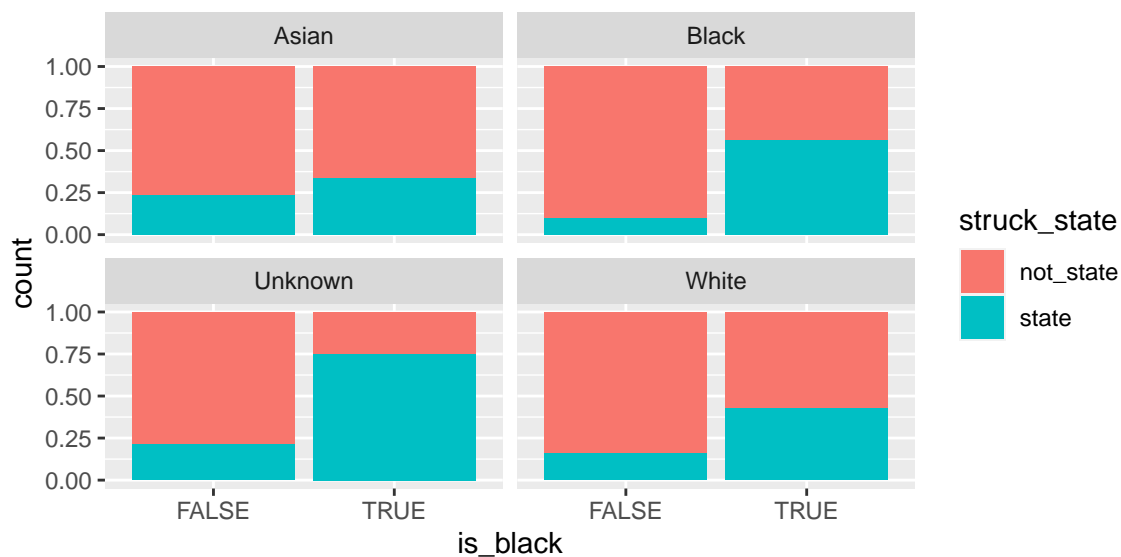
```
ggplot(apm, aes(x=is_black, fill=struck_state)) +
  geom_bar(position="fill") +
  facet_wrap(~fam_law_enforcement)
```



```
ggplot(apm, aes(x=is_black, fill=struck_state)) +
  geom_bar(position="fill") +
  facet_wrap(~same_race)
```



```
ggplot(apm, aes(x=is_black, fill=struck_state)) +
  geom_bar(position="fill") +
  facet_wrap(~defendant_race)
```



```
p1 <- ggplot(apm, aes(x=is_black, fill=struck_state)) +
  geom_bar(position="fill") +
  facet_wrap(~fam_accused) +
  labs(y="Proportion",
       x="Family or close friend had been accused of crime",
       title="Juror is black") +
  scale_fill_manual(
    values=c("skyblue", "blue"),
    name=" ",
    labels=c("Not struck by state", "Struck by state")
  )
```

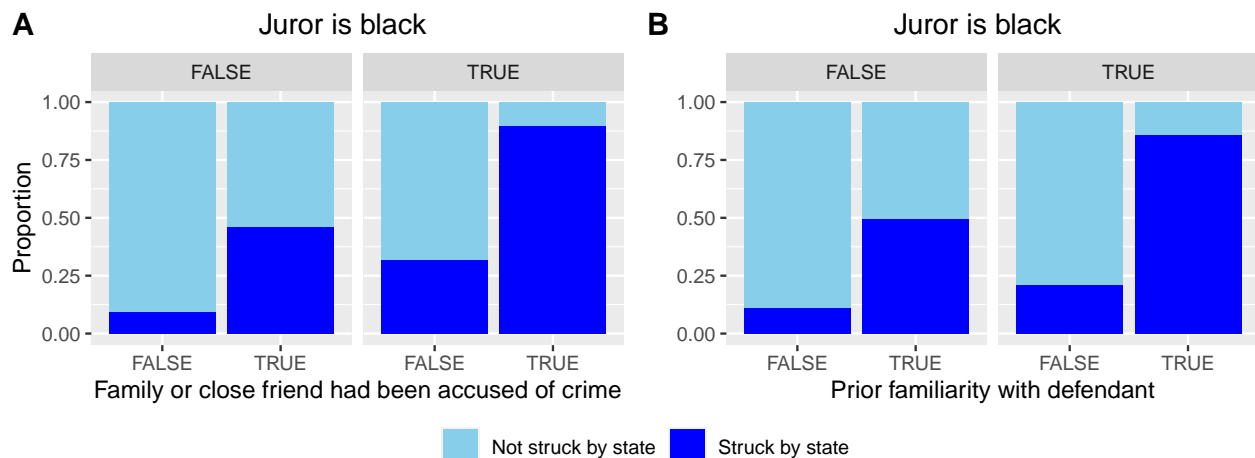
```
p2 <- ggplot(apm, aes(x=is_black, fill=struck_state)) +
  geom_bar(position="fill") +
```

```

facet_wrap(~know_def) +
labs(y=" ",
      x="Prior familiarity with defendant",
      title="Juror is black") +
scale_fill_manual(
  values=c("skyblue", "blue"),
  name=" ",
  labels=c("Not struck by state", "Struck by state")
)

ggarrange(p1, p2, labels = c("A", "B"), common.legend = TRUE, legend = "bottom", ncol = 2, nrow = 1)

```



```

# explore interaction models
apm_a_predictors_inter_all <- glm(struck_state_bin~accused+is_black+fam_accused+death_hesitation+know_def,
summary(apm_a_predictors_inter_all)

```

```

##
## Call:
## glm(formula = struck_state_bin ~ accused + is_black + fam_accused +
##      death_hesitation + know_def + same_race + fam_law_enforcement +
##      accused * is_black + fam_accused * is_black + death_hesitation *
##      is_black + know_def * is_black + same_race * is_black + fam_law_enforcement *
##      is_black, family = "binomial", data = apm)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.8183  -0.5597  -0.4234  -0.3077   2.4795
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -2.3668     0.1131  -20.930  < 2e-16 ***
## accusedTRUE         2.2840     0.6982   3.271  0.00107 **
## is_blackTRUE        1.9653     0.2039   9.637  < 2e-16 ***
## fam_accusedTRUE     1.6165     0.2086   7.749 9.26e-15 ***
## death_hesitationTRUE 0.9133     0.8676   1.053  0.29252
## know_defTRUE        0.6858     0.3764   1.822  0.06841 .
## same_raceTRUE       0.5925     0.1927   3.075  0.00211 **
## fam_law_enforcementTRUE -0.6597     0.2271  -2.905  0.00367 **
## accusedTRUE:is_blackTRUE 0.8627     1.2492   0.691  0.48979

```



```
## is_blackTRUE:fam_accusedTRUE      0.8178      0.3742      2.185 0.02886 *
## is_blackTRUE:death_hesitationTRUE 14.1875    328.7281      0.043 0.96557
## is_blackTRUE:know_defTRUE          1.2222      0.5054      2.418 0.01560 *
## is_blackTRUE:same_raceTRUE         -0.5808      0.2745     -2.116 0.03435 *
## is_blackTRUE:fam_law_enforcementTRUE 0.2830      0.3350      0.845 0.39818
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
```

```
## (Dispersion parameter for binomial family taken to be 1)
```

```
##
```

```
##      Null deviance: 2579.5  on 2294  degrees of freedom
```

```
## Residual deviance: 1866.7  on 2281  degrees of freedom
```

```
## AIC: 1894.7
```

```
##
```

```
## Number of Fisher Scoring iterations: 14
```

```
# collinearity
```

```
vif(apm_a_predictors_inter_all)
```

```
##              accused              is_black
##          1.471378              3.034618
##          fam_accused          death_hesitation
##          1.553895              1.008118
##          know_def              same_race
##          2.299936              2.708702
##          fam_law_enforcement    accused:is_black
##          1.902615              1.460087
##          is_black:fam_accused    is_black:death_hesitation
##          1.538198              1.000007
##          is_black:know_def        is_black:same_race
##          2.397630              5.176079
##          is_black:fam_law_enforcement
##          2.007127
```

```
apm_a_predictors_inter1 <- glm(struck_state_bin~accused+is_black+fam_accused+death_hesitation+know_def+
```

```
summary(apm_a_predictors_inter1) # accused*is_black not significant
```

```
##
```

```
## Call:
```

```
## glm(formula = struck_state_bin ~ accused + is_black + fam_accused +
```

```
##      death_hesitation + know_def + same_race + fam_law_enforcement +
```

```
##      accused * is_black, family = "binomial", data = apm)
```

```
##
```

```
## Deviance Residuals:
```

```
##      Min       1Q   Median       3Q      Max
## -2.4697  -0.4879  -0.4118  -0.3132   2.4656
```

```
##
```

```
## Coefficients:
```

```
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -2.4250     0.1013  -23.931  < 2e-16 ***
## accusedTRUE         2.1515     0.7135   3.015 0.002566 **
## is_blackTRUE        1.8895     0.1415  13.357  < 2e-16 ***
## fam_accusedTRUE     1.8485     0.1618  11.422  < 2e-16 ***
## death_hesitationTRUE 1.8232     0.5912   3.084 0.002044 **
## know_defTRUE        1.3314     0.2232   5.966 2.43e-09 ***
```

```

## same_raceTRUE          0.3567      0.1401    2.547 0.010880 *
## fam_low_enforcementTRUE -0.5656      0.1624   -3.482 0.000499 ***
## accusedTRUE:is_blackTRUE  0.9353      1.2558    0.745 0.456377
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 2579.5  on 2294  degrees of freedom
## Residual deviance: 1886.9  on 2286  degrees of freedom
## AIC: 1904.9
##
## Number of Fisher Scoring iterations: 5
apm_a_predictors_inter2 <- glm(struck_state_bin~accused+is_black+fam_accused+death_hesitation+know_def+
summary(apm_a_predictors_inter2) # fam_accused*is_black significant

##
## Call:
## glm(formula = struck_state_bin ~ accused + is_black + fam_accused +
##      death_hesitation + know_def + same_race + fam_low_enforcement +
##      fam_accused * is_black, family = "binomial", data = apm)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.6432  -0.4955  -0.4224  -0.3243   2.4377
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -2.3716     0.1032  -22.983  < 2e-16 ***
## accusedTRUE        2.5158     0.5409   4.651 3.31e-06 ***
## is_blackTRUE       1.8089     0.1465  12.350  < 2e-16 ***
## fam_accusedTRUE    1.5838     0.2075   7.634 2.28e-14 ***
## death_hesitationTRUE 1.8442     0.5836   3.160 0.001576 **
## know_defTRUE       1.3313     0.2219   6.000 1.97e-09 ***
## same_raceTRUE      0.3361     0.1406   2.390 0.016826 *
## fam_low_enforcementTRUE -0.5470     0.1634  -3.347 0.000818 ***
## is_blackTRUE:fam_accusedTRUE 0.7739     0.3712   2.085 0.037092 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 2579.5  on 2294  degrees of freedom
## Residual deviance: 1882.9  on 2286  degrees of freedom
## AIC: 1900.9
##
## Number of Fisher Scoring iterations: 5
apm_a_predictors_inter3 <- glm(struck_state_bin~accused+is_black+fam_accused+death_hesitation+know_def+
summary(apm_a_predictors_inter3) # death_hesitation*is_black not significant

##
## Call:
## glm(formula = struck_state_bin ~ accused + is_black + fam_accused +

```

```

##      death_hesitation + know_def + same_race + fam_law_enforcement +
##      death_hesitation * is_black, family = "binomial", data = apm)
##
## Deviance Residuals:
##      Min        1Q      Median        3Q        Max
## -2.4643   -0.4877   -0.4125   -0.3157    2.4592
##
## Coefficients:
##                                Estimate Std. Error z value Pr(>|z|)
## (Intercept)                   -2.4212     0.1011 -23.954 < 2e-16 ***
## accusedTRUE                    2.5096     0.5448   4.606 4.10e-06 ***
## is_blackTRUE                   1.8836     0.1413  13.333 < 2e-16 ***
## fam_accusedTRUE                1.8520     0.1617  11.454 < 2e-16 ***
## death_hesitationTRUE           0.8810     0.8816   0.999 0.317650
## know_defTRUE                   1.3209     0.2234   5.914 3.35e-09 ***
## same_raceTRUE                  0.3520     0.1400   2.515 0.011915 *
## fam_law_enforcementTRUE        -0.5528     0.1620  -3.411 0.000646 ***
## is_blackTRUE:death_hesitationTRUE 14.1940  352.3902   0.040 0.967871
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 2579.5  on 2294  degrees of freedom
## Residual deviance: 1883.2  on 2286  degrees of freedom
## AIC: 1901.2
##
## Number of Fisher Scoring iterations: 14
apm_a_predictors_inter4 <- glm(struck_state_bin~accused+is_black+fam_accused+death_hesitation+know_def+
summary(apm_a_predictors_inter4) # know_def*is_black significant

##
## Call:
## glm(formula = struck_state_bin ~ accused + is_black + fam_accused +
##      death_hesitation + know_def + same_race + fam_law_enforcement +
##      know_def * is_black, family = "binomial", data = apm)
##
## Deviance Residuals:
##      Min        1Q      Median        3Q        Max
## -2.6382   -0.4932   -0.4189   -0.3188    2.4515
##
## Coefficients:
##                                Estimate Std. Error z value Pr(>|z|)
## (Intercept)                   -2.3892     0.1017 -23.486 < 2e-16 ***
## accusedTRUE                    2.5642     0.5444   4.710 2.47e-06 ***
## is_blackTRUE                   1.8330     0.1436  12.767 < 2e-16 ***
## fam_accusedTRUE                1.8469     0.1612  11.460 < 2e-16 ***
## death_hesitationTRUE           1.8017     0.5906   3.051 0.002283 **
## know_defTRUE                   0.6932     0.3800   1.824 0.068112 .
## same_raceTRUE                  0.3440     0.1403   2.451 0.014241 *
## fam_law_enforcementTRUE        -0.5649     0.1635  -3.454 0.000552 ***
## is_blackTRUE:know_defTRUE       1.1209     0.5042   2.223 0.026208 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 2579.5 on 2294 degrees of freedom
## Residual deviance: 1882.2 on 2286 degrees of freedom
## AIC: 1900.2
##
## Number of Fisher Scoring iterations: 5
apm_a_predictors_inter5 <- glm(struck_state_bin~accused+is_black+fam_accused+death_hesitation+know_def+
summary(apm_a_predictors_inter5) # same_race*is_black sort of significant

##
## Call:
## glm(formula = struck_state_bin ~ accused + is_black + fam_accused +
## death_hesitation + know_def + same_race + fam_law_enforcement +
## same_race * is_black, family = "binomial", data = apm)
##
## Deviance Residuals:
## Min 1Q Median 3Q Max
## -2.4667 -0.5325 -0.3968 -0.3028 2.4922
##
## Coefficients:
## Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.5020 0.1106 -22.620 < 2e-16 ***
## accusedTRUE 2.5514 0.5485 4.652 3.29e-06 ***
## is_blackTRUE 2.1474 0.1961 10.949 < 2e-16 ***
## fam_accusedTRUE 1.8746 0.1632 11.489 < 2e-16 ***
## death_hesitationTRUE 1.8631 0.5912 3.151 0.001626 **
## know_defTRUE 1.3535 0.2232 6.065 1.32e-09 ***
## same_raceTRUE 0.6204 0.1959 3.168 0.001537 **
## fam_law_enforcementTRUE -0.5576 0.1623 -3.436 0.000591 ***
## is_blackTRUE:same_raceTRUE -0.5006 0.2742 -1.826 0.067865 .
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 2579.5 on 2294 degrees of freedom
## Residual deviance: 1884.3 on 2286 degrees of freedom
## AIC: 1902.3
##
## Number of Fisher Scoring iterations: 5
apm_a_predictors_inter6 <- glm(struck_state_bin~accused+is_black+fam_accused+death_hesitation+know_def+
summary(apm_a_predictors_inter6) # fam_law_enforcement*is_black not significant

##
## Call:
## glm(formula = struck_state_bin ~ accused + is_black + fam_accused +
## death_hesitation + know_def + same_race + fam_law_enforcement +
## fam_law_enforcement * is_black, family = "binomial", data = apm)
##
## Deviance Residuals:
## Min 1Q Median 3Q Max
```

```
## -2.4632 -0.4933 -0.4169 -0.2888 2.5290
##
## Coefficients:
##
## Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.3991 0.1035 -23.182 < 2e-16 ***
## accusedTRUE 2.5026 0.5473 4.573 4.82e-06 ***
## is_blackTRUE 1.8401 0.1479 12.444 < 2e-16 ***
## fam_accusedTRUE 1.8646 0.1635 11.406 < 2e-16 ***
## death_hesitationTRUE 1.8552 0.5935 3.126 0.001774 **
## know_defTRUE 1.3249 0.2241 5.911 3.40e-09 ***
## same_raceTRUE 0.3539 0.1400 2.528 0.011462 *
## fam_law_enforcementTRUE -0.7570 0.2299 -3.292 0.000993 ***
## is_blackTRUE:fam_law_enforcementTRUE 0.4111 0.3286 1.251 0.210824
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 2579.5 on 2294 degrees of freedom
## Residual deviance: 1886.0 on 2286 degrees of freedom
## AIC: 1904
##
## Number of Fisher Scoring iterations: 5
```

```
apm_a_predictors_inter_test <- glm(struck_state_bin~accused+is_black+fam_accused+death_hesitation+know_def,
summary(apm_a_predictors_inter_test) # know_def is not significant
```

```
##
## Call:
## glm(formula = struck_state_bin ~ accused + is_black + fam_accused +
## death_hesitation + know_def + same_race + fam_law_enforcement +
## fam_accused * is_black + know_def * is_black + same_race *
## is_black, family = "binomial", data = apm)
##
## Deviance Residuals:
## Min 1Q Median 3Q Max
## -2.8337 -0.5547 -0.4168 -0.3208 2.4464
##
## Coefficients:
##
## Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.4000 0.1118 -21.466 < 2e-16 ***
## accusedTRUE 2.6147 0.5421 4.823 1.41e-06 ***
## is_blackTRUE 2.0197 0.1999 10.104 < 2e-16 ***
## fam_accusedTRUE 1.5927 0.2075 7.676 1.64e-14 ***
## death_hesitationTRUE 1.8686 0.5809 3.216 0.00130 **
## know_defTRUE 0.6687 0.3768 1.774 0.07600 .
## same_raceTRUE 0.6059 0.1926 3.145 0.00166 **
## fam_law_enforcementTRUE -0.5411 0.1650 -3.279 0.00104 **
## is_blackTRUE:fam_accusedTRUE 0.8364 0.3727 2.244 0.02482 *
## is_blackTRUE:know_defTRUE 1.2486 0.5049 2.473 0.01339 *
## is_blackTRUE:same_raceTRUE -0.5752 0.2744 -2.096 0.03609 *
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
```

```

##
## Null deviance: 2579.5 on 2294 degrees of freedom
## Residual deviance: 1872.7 on 2284 degrees of freedom
## AIC: 1894.7
##
## Number of Fisher Scoring iterations: 5
# anova test, probably not valid
anova(apm_a_predictors_inter_test, apm_a_predictors_inter_all, test="Chisq")

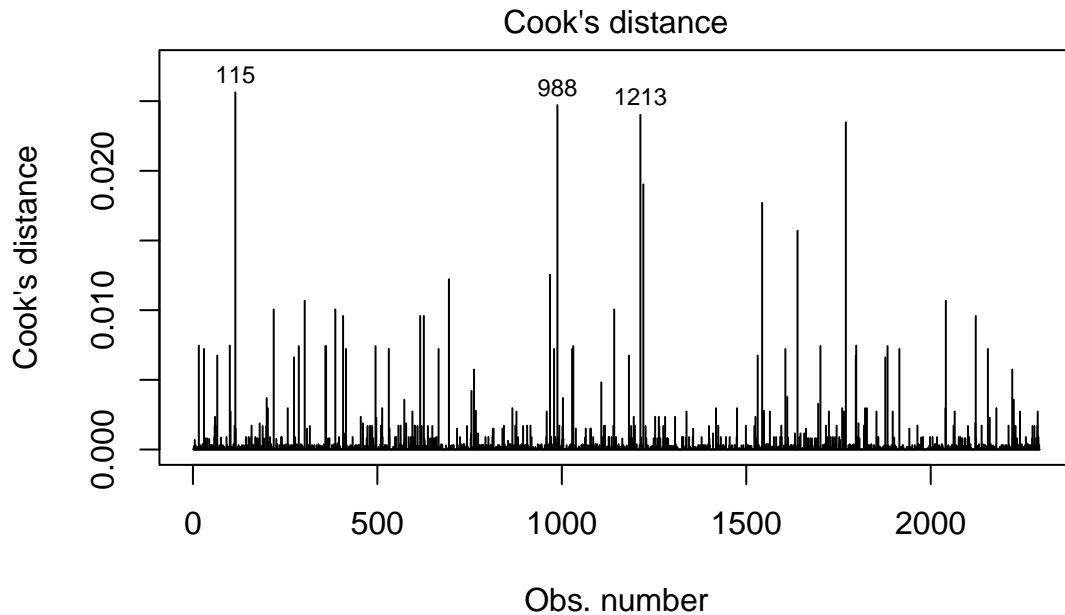
## Analysis of Deviance Table
##
## Model 1: struck_state_bin ~ accused + is_black + fam_accused + death_hesitation +
## know_def + same_race + fam_law_enforcement + fam_accused *
## is_black + know_def * is_black + same_race * is_black
## Model 2: struck_state_bin ~ accused + is_black + fam_accused + death_hesitation +
## know_def + same_race + fam_law_enforcement + accused * is_black +
## fam_accused * is_black + death_hesitation * is_black + know_def *
## is_black + same_race * is_black + fam_law_enforcement * is_black
## Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1 2284 1872.7
## 2 2281 1866.7 3 5.9842 0.1124
# final model
apm_a_predictors_inter_sig <- glm(struck_state_bin~accused+is_black+fam_accused+death_hesitation+know_d
summary(apm_a_predictors_inter_sig)

##
## Call:
## glm(formula = struck_state_bin ~ accused + is_black + fam_accused +
## death_hesitation + know_def + same_race + fam_law_enforcement +
## fam_accused * is_black + know_def * is_black, family = "binomial",
## data = apm)
##
## Deviance Residuals:
## Min 1Q Median 3Q Max
## -2.8213 -0.5017 -0.4316 -0.3313 2.4206
##
## Coefficients:
## Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.3265 0.1036 -22.450 < 2e-16 ***
## accusedTRUE 2.5660 0.5393 4.758 1.96e-06 ***
## is_blackTRUE 1.7373 0.1492 11.642 < 2e-16 ***
## fam_accusedTRUE 1.5726 0.2062 7.628 2.38e-14 ***
## death_hesitationTRUE 1.8235 0.5820 3.133 0.001728 **
## know_defTRUE 0.6865 0.3749 1.831 0.067063 .
## same_raceTRUE 0.3174 0.1411 2.249 0.024504 *
## fam_law_enforcementTRUE -0.5483 0.1649 -3.326 0.000882 ***
## is_blackTRUE:fam_accusedTRUE 0.8118 0.3715 2.185 0.028870 *
## is_blackTRUE:know_defTRUE 1.1620 0.5015 2.317 0.020511 *
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 2579.5 on 2294 degrees of freedom

```

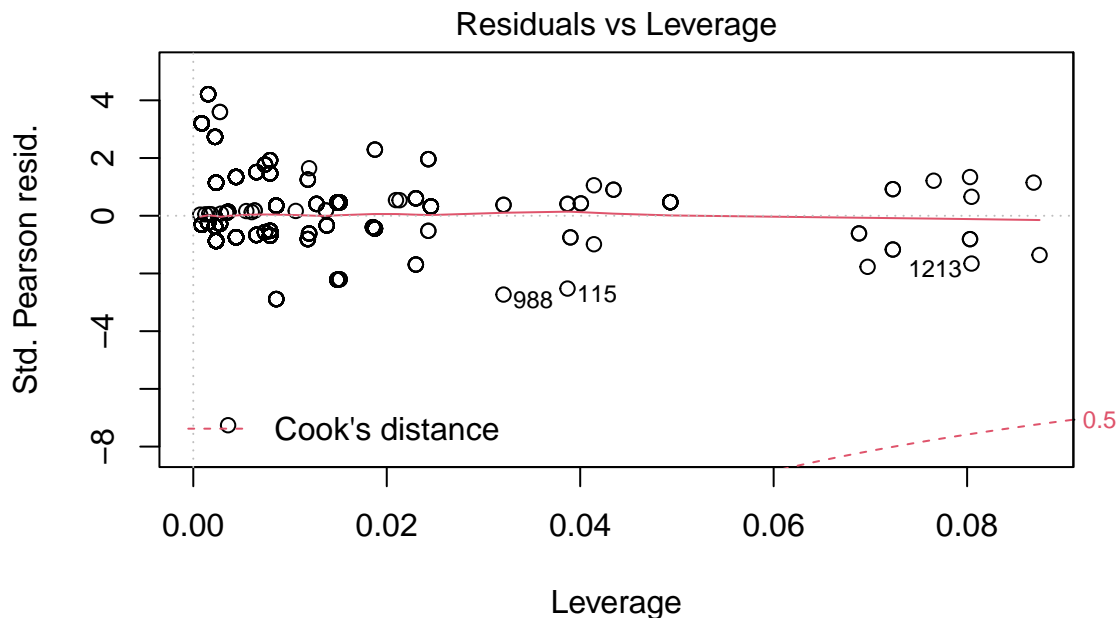
```
## Residual deviance: 1877.0 on 2285 degrees of freedom
## AIC: 1897
##
## Number of Fisher Scoring iterations: 5
```

```
# check outliers for interaction model
plot(apm_a_predictors_inter_sig, which=4)
```



```
glm(struck_state_bin ~ accused + is_black + fam_accused + death_hesitation
```

```
plot(apm_a_predictors_inter_sig, which=5)
```



```
glm(struck_state_bin ~ accused + is_black + fam_accused + death_hesitation
```

```
# slice out case 988
apm_a_predictors_inter_sig_aug <- augment(apm_a_predictors_inter_sig)
apm_a_predictors_inter_sig_aug %>% slice(115)
```

```
## # A tibble: 1 x 14
##   struck_state_bin accused is_black fam_accused death_hesitation know_def
##           <int> <lgl>   <lgl>     <lgl>         <lgl>         <lgl>
## 1             0 TRUE    FALSE     TRUE         FALSE         FALSE
## # ... with 8 more variables: same_race <lgl>, fam_law_enforcement <lgl>,
## #   .fitted <dbl>, .resid <dbl>, .std.resid <dbl>, .hat <dbl>, .sigma <dbl>,
## #   .cooksdi <dbl>

# remove case 988
apm_a_predictors_inter_sig_no_115 <- glm(struck_state_bin~accused+is_black+fam_accused+death_hesitation+
summary(apm_a_predictors_inter_sig_no_115)

##
## Call:
## glm(formula = struck_state_bin ~ accused + is_black + fam_accused +
##   death_hesitation + know_def + same_race + fam_law_enforcement +
##   fam_accused * is_black + know_def * is_black, family = "binomial",
##   data = apm, subset = -c(115))
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.8224  -0.5011  -0.4313  -0.3299   2.4241
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -2.3282     0.1038 -22.436 < 2e-16 ***
## accusedTRUE         2.8410     0.5880   4.832 1.35e-06 ***
## is_blackTRUE        1.7377     0.1494  11.631 < 2e-16 ***
## fam_accusedTRUE     1.6039     0.2060   7.785 6.97e-15 ***
## death_hesitationTRUE 1.8223     0.5824   3.129 0.001755 **
## know_defTRUE        0.6703     0.3779   1.774 0.076090 .
## same_raceTRUE       0.3166     0.1413   2.241 0.025039 *
## fam_law_enforcementTRUE -0.5555     0.1652  -3.364 0.000769 ***
## is_blackTRUE:fam_accusedTRUE 0.7828     0.3714   2.108 0.035065 *
## is_blackTRUE:know_defTRUE 1.1809     0.5038   2.344 0.019091 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 2578.9  on 2293  degrees of freedom
## Residual deviance: 1872.8  on 2284  degrees of freedom
## AIC: 1892.8
##
## Number of Fisher Scoring iterations: 5

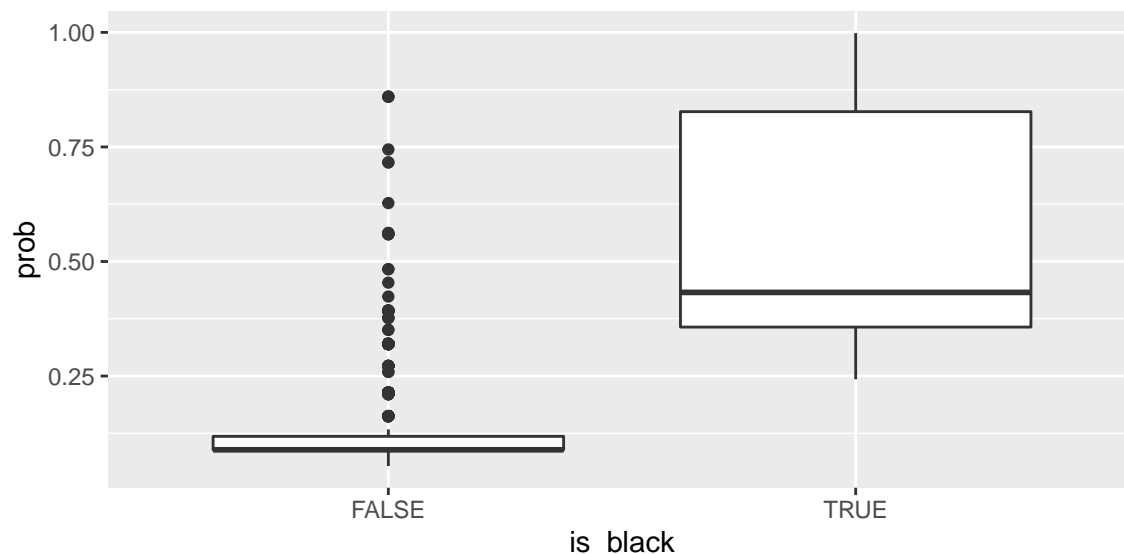
# collinearity
vif(apm_a_predictors_inter_sig)

##           accused           is_black           fam_accused
##      1.005285         1.629873         1.519757
## death_hesitation know_def           same_race
##      1.001559         2.282398         1.457806
## fam_law_enforcement is_black:fam_accused is_black:know_def
##      1.030875         1.511836         2.359504
```



```
# plot probability of getting struck against is_black
apm_prob <- apm_a_predictors_inter_sig_aug %>%
  mutate(
    prob = exp(.fitted)/(1+exp(.fitted))
  )

ggplot(apm_prob, aes(x=is_black, y=prob)) +
  geom_boxplot()
```



```
# interpret effects
tidy(apm_a_predictors_inter_sig, conf.int = TRUE)
```

```
## # A tibble: 10 x 7
##   term                estimate std.error statistic  p.value conf.low conf.high
##   <chr>                <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>
## 1 (Intercept)         -2.33     0.104   -22.4  1.29e-111 -2.53    -2.13
## 2 accusedTRUE          2.57     0.539    4.76  1.96e- 6  1.58     3.73
## 3 is_blackTRUE          1.74     0.149   11.6  2.51e- 31  1.45     2.03
## 4 fam_accusedTRUE       1.57     0.206    7.63  2.38e- 14  1.16     1.97
## 5 death_hesitationTR~   1.82     0.582    3.13  1.73e- 3  0.714    3.02
## 6 know_defTRUE          0.687    0.375    1.83  6.71e- 2 -0.0955  1.39
## 7 same_raceTRUE         0.317    0.141    2.25  2.45e- 2  0.0391  0.593
## 8 fam_law_enforcemen~  -0.548    0.165   -3.33  8.82e- 4 -0.878   -0.231
## 9 is_blackTRUE:fam_a~   0.812    0.371    2.19  2.89e- 2  0.109    1.57
## 10 is_blackTRUE:know_~  1.16     0.502    2.32  2.05e- 2  0.215    2.19
```

```
exp(1.7373028) - 1
```

```
## [1] 4.681997
```

```
exp(1.44691177) - 1
```

```
## [1] 3.249969
```

```
exp(2.0322287) -1
```

```
## [1] 6.631075
```