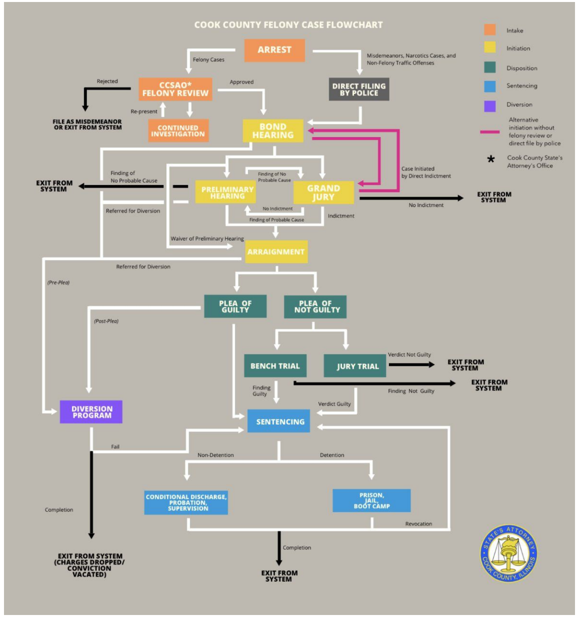
Analyzing Correlates to Prison Sentencing in Cook County

Akhil Ghosh & Anthony Orso

Sentencing Dataset

- Sentencing data from the Cook County data catalog
  - Part of a series of data that captures legal proceedings in Cook County
- 281K records spanning from 1912 to April 5, 2023
  - Vast majority of records after 2000 however
- Reflects the judgement from the courts on people who have been found guilty through disposition
- 44 different variables including demographics and arrest information



Data Clean-up

- Originally, 280 thousands records
  - Included non-primary charges and updated sentencing records
  - Decided to look at the primary and up-to-date record
- Decided to subset the records
  - Looked at race, gender, number of charges, age, received date, offense type, sentencing type, case id, case participant id, sentencing judge
- Binned offenses, and sentencing types
  - Offenses binned into 10 categories
  - Sentencing binned into probation, prison, supervision and death
    - Utilized Commitment Type if it was a PROMIS Conversion

```r
sentencing <- sentencing %>% mutate(offenses_binned =
  case_when(
    UPDATED_OFFENSE_CATEGORY %in% violent_categories ~ "Violent Crimes",
    UPDATED_OFFENSE_CATEGORY %in% nonviolent_categories ~ "Non-violent Crimes",
     UPDATED_OFFENSE_CATEGORY %in% sex_categories ~ "Sex Crimes",
    UPDATED_OFFENSE_CATEGORY %in% property_categories ~ "Property Crimes",
    UPDATED_OFFENSE_CATEGORY %in%  firearm_categories ~ "Firearm/Weapon-related Crimes",
    UPDATED_OFFENSE_CATEGORY %in% dui_categories ~ "DUI and License-related Crimes",
    UPDATED_OFFENSE_CATEGORY %in%  fraud_categories ~ "Fraud and Deception",
    UPDATED_OFFENSE_CATEGORY %in% legal_categories ~ "Legal process crimes",
    UPDATED_OFFENSE_CATEGORY %in% narc_categories ~ "Narcotics",
    UPDATED_OFFENSE_CATEGORY == "PROMIS Conversion" ~ NA_character_,
    TRUE ~ "Other Offenses"
  )
```

```r
sentencing <- sentencing %>%
  mutate(SENTENCE_TYPE = case_when(
    SENTENCE_TYPE == "Conversion" ~ COMMITMENT_TYPE,
    TRUE ~ SENTENCE_TYPE
  ),collapsed_sentence_type = case_when(
    SENTENCE_TYPE %in% c("Cook County Boot Camp","2nd Chance Probation",
                         "Conditional Discharge", "Conditional Release",
                         "Probation", "Probation Terminated Instanter",
                         "Probation Terminated Satisfactorily",
                         "Probation Terminated Unsatisfactorily") ~ "Probation",

    SENTENCE_TYPE %in% c("Jail", "Prison", "Illinois Department of Corrections") ~ "Prison",
    SENTENCE_TYPE == "Inpatient Mental Health Services" ~ "Mental Health Services",
    SENTENCE_TYPE %in% c("Natural Life", "Death") ~ "Death",
    SENTENCE_TYPE %in% c("Vocational Rehabilitation Impact Center(VRIC)","Supervision",
                         "Cook County Impact Incarceration Program") ~ "Supervision",
    TRUE ~ SENTENCE_TYPE
  ))
```

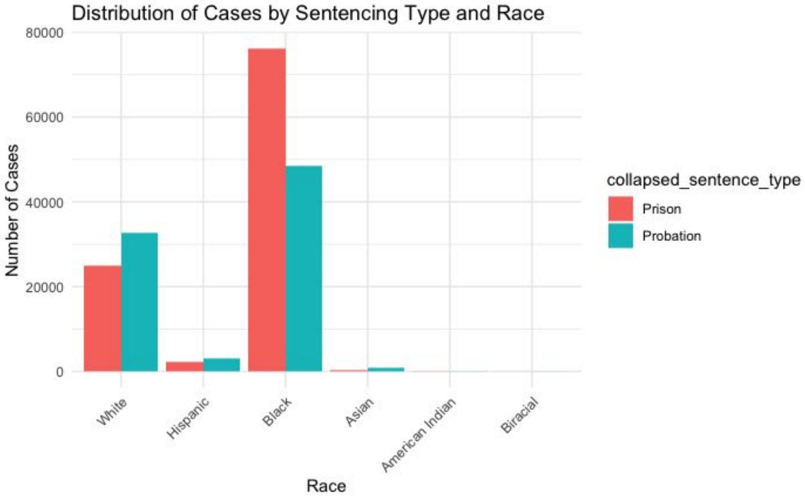
Data Clean-up

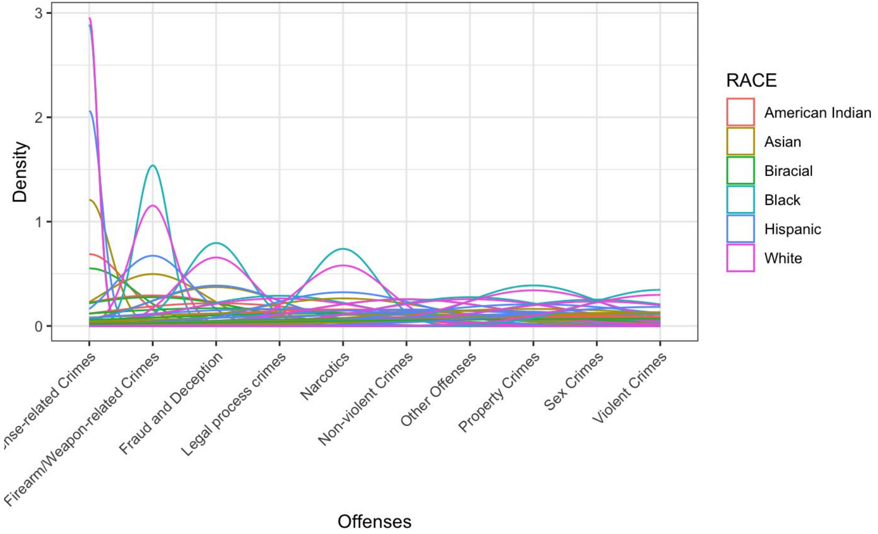
- Still had over 10k missing records for offense category due to PROMIS Conversions
- Decided to utilize disposition offense title to help fill in the missing information
- Was able to reduce the amount of missing information to only 138 records

```r
sentencing <- sentencing %>%
  mutate(offenses_binned =
           if_else((PRIMARY_CHARGE_FLAG == T & CURRENT_SENTENCE_FLAG == T & is.na(offenses_binned)),
                   case_when(
                     str_detect(DISPOSITION_CHARGED_OFFENSE_TITLE, violent_pattern_regex) ~ "Violent Crimes",
                     str_detect(DISPOSITION_CHARGED_OFFENSE_TITLE, str_c("SEX")) ~ "Sex Crimes",
                     str_detect(DISPOSITION_CHARGED_OFFENSE_TITLE, str_c("BAIL|DISORDERLY|PCS|THREATEN|RESIST")) ~ "Non-violent Crimes",
                     str_detect(DISPOSITION_CHARGED_OFFENSE_TITLE, str_c("UUW|FIREARM|WEAPON")) ~ "Firearm/Weapon-related Crimes",
                     str_detect(DISPOSITION_CHARGED_OFFENSE_TITLE, str_c("DUI|DRIVING|VEHICLES|DRIVING|BAC|UNDER INFLU")) ~ 'DUI and License-related Cri
                     str_detect(DISPOSITION_CHARGED_OFFENSE_TITLE, str_c("SUBSTANCE|POSS|MANU|DRUG|NARC")) ~ "Narcotics",
                     str_detect(DISPOSITION_CHARGED_OFFENSE_TITLE, str_c("THEFT|DISP MERCH|INVASION|DAMAGE PROP|TRESPASS|PROPERTY")) ~ "Property Crimes'
                     str_detect(DISPOSITION_CHARGED_OFFENSE_TITLE, "FORGERY|CREDIT CARD") ~ "Fraud and Deception",
                     T ~ offenses_binned),
                   offenses_binned)
  )
```

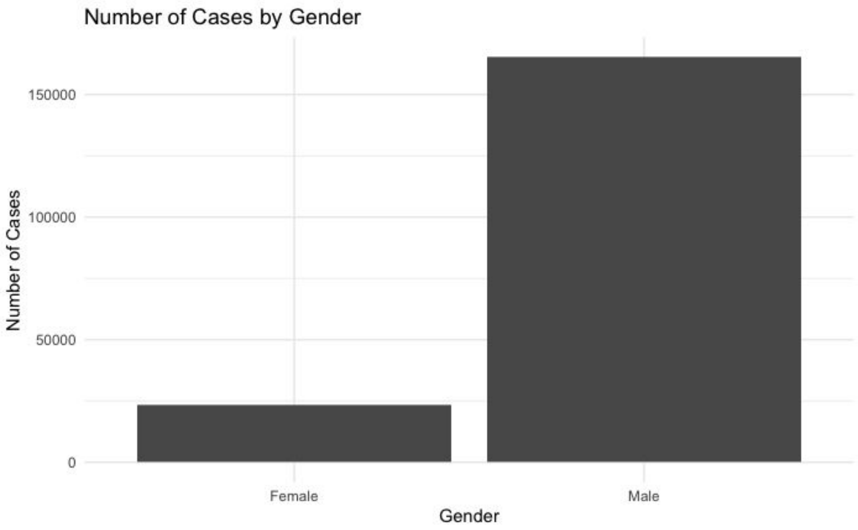
```
.
DUI and License-related Crimes  Firearm/Weapon-related Crimes           Fraud and Deception          Legal process crimes                     Narcotics
                         28495                          21166                          6626                           213                         53730
            Non-violent Crimes                Other Offenses               Property Crimes                    Sex Crimes               Violent Crimes
                          1092                           2957                         27491                          6251                         45152
```

EDA

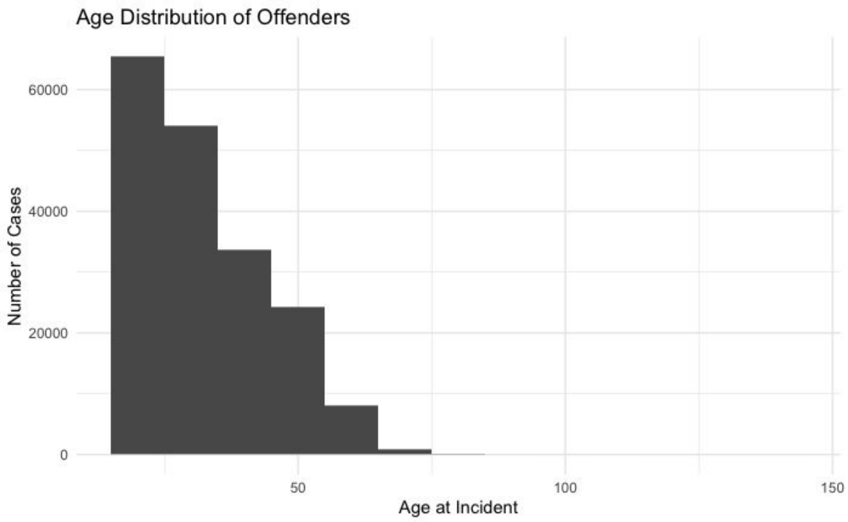

Distribution of Cases by Sentencing Type and Race
80000
60000
40000
20000
0
Number of Cases
White
Hispanic
Black
Asian
American Indian
Biracial
Race
collapsed_sentence_type
Prison
Probation


3
2
1
0
Density
Firearm/Weapon-related Crimes
Fraud and Deception
Legal process crimes
Narcotics
Non-violent Crimes
Other Offenses
Property Crimes
Sex Crimes
Violent Crimes
Offenses
RACE
American Indian
Asian
Biracial
Black
Hispanic
White

EDA



Number of Cases by Gender
Number of Cases
150000
100000
50000
0
Female
Male
Gender



Age Distribution of Offenders
Number of Cases
60000
40000
20000
0
50
100
150
Age at Incident

Model Building

Initial Model Considerations

- Based on our EDA, our initial hunch is to either do a multinomial logistic regression or mixed effects model
- A multinomial model would assess the impact of predictor variables on whether a convicted person receives supervision, probation, or a prison sentence
- A mixed effects model would test whether variables such as race, individual convictees, or judges have a degree of randomness and, thus, warrant a combined fixed and random effect

Mixed Effects Model

- Because of known discrepancies in criminal justice, we decided to first program the following model which test a binary response variable
  - Levels: Probation (0), Prison (1)
    - Baselines : Race = White, offenses = Violent Crimes, Gender = Female

```
fit_judges_glmer <- glmer(collapsed_sentence_type ~ RACE+ offenses_binned + GENDER + CHARGE_COUNT
                        + AGE_AT_INCIDENT + (1 | SENTENCE_JUDGE),
                        family = binomial(), data = data)
```

Results

```
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod']
 Family: binomial  ( logit )
Formula: collapsed_sentence_type ~ RACE + offenses_binned + GENDER + CHARGE_COUNT +      AGE_AT_INCIDENT + (1 | SENTENCE_JUDGE)
   Data: data

      AIC       BIC    logLik  deviance  df.resid 
 236355.4  236547.9 -118158.7  236317.4    186136 

Scaled residuals: 
    Min      1Q  Median      3Q     Max 
-4.8499 -0.8118 -0.5553  0.9470  3.3959 

Random effects:
 Groups         Name        Variance Std.Dev.
 SENTENCE_JUDGE (Intercept) 0.1466   0.3828  
Number of obs: 186155, groups:  SENTENCE_JUDGE, 326

Fixed effects:
                                               Estimate Std. Error z value Pr(>|z|)    
(Intercept)                                   1.4642899  0.0392862  37.272  < 2e-16 ***
RACEHispanic                                 -0.0219133  0.0307070  -0.714 0.475459    
RACEBlack                                    -0.7080767  0.0117964 -60.025  < 2e-16 ***
RACEAsian                                     0.3181619  0.0663357   4.796 1.62e-06 ***
RACEAmerican Indian                          -0.5564122  0.2295448  -2.424 0.015351 *  
RACEBiracial                                  0.2647861  0.4193918   0.631 0.527807    
offenses_binnedNon-violent Crimes             0.0540457  0.0680921   0.794 0.427361    
offenses_binnedFirearm/Weapon-related Crimes  0.0726967  0.0187593   3.875 0.000107 ***
offenses_binnedDUI and License-related Crimes 0.8145959  0.0171451  47.512  < 2e-16 ***
offenses_binnedFraud and Deception            1.0787885  0.0300616  35.886  < 2e-16 ***
offenses_binnedLegal process crimes           0.4056273  0.1489266   2.724 0.006456 ** 
offenses_binnedProperty Crimes                0.3597074  0.0171296  20.999  < 2e-16 ***
offenses_binnedOther Offenses                 0.9883373  0.0425461  23.230  < 2e-16 ***
offenses_binnedSex Crimes                     0.0216900  0.0305878   0.709 0.478258    
offenses_binnedNarcotics                      0.6584009  0.0148189  44.430  < 2e-16 ***
GENDERMale                                   -1.1156376  0.0159486 -69.952  < 2e-16 ***
CHARGE_COUNT                                  0.0299129  0.0070569   4.239 2.25e-05 ***
AGE_AT_INCIDENT                              -0.0167123  0.0004438 -37.660  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Evaluation

- Weak standard deviation for the random effect
  - Judges are passing sentences pretty uniformly

- Some interesting insights about individual judges though

	(Intercept) <dbl>
Carl Anthony Walker	1.135859471
Carl B Boyd	0.673181391
Carmen Aguilar	1.611404039
Carol A Kipperman	-1.265461766
Carol M Howard	-0.071502051
Carol Pearce McCarthy	0.170353060
Casandra Lewis	1.403062255
Catherine Marie Haberkorn	-0.361851447
Charles P Burns	-0.928237196
Cheryl D Cesario	0.248253733

31-40 of 326 rows

Description: df [326 x 1]

	(Intercept) <dbl>
Arthur F Hill	-1.030455843

```
sen2 %>% filter(SENTENCE_JUDGE == 'Carmen  Aguilar' ) %>%
  select(collapsed_sentence_type) %>%
  pull(collapsed_sentence_type) %>% table()
```

```
.
  Prison Probation
      78       266
```

```
sen2 %>% filter(SENTENCE_JUDGE == 'Arthur F Hill' ) %>%
  select(collapsed_sentence_type) %>%
  pull(collapsed_sentence_type) %>% table()
```

```
.
  Prison Probation
    1625      1001
```

Multinomial Logistic Regression w/ Race

- Goal is to account for the difference between supervision, probation, and prison sentences given the previous model's explanatory variables plus race
- We chose nominal over ordinal because it's difficult to discretize the difference between supervision, probation, and prison
- Is supervision 0, probation 4, and prison 100? Murky waters.
- Odds ratios between the levels makes more sense
- 1: Prison, 2: Probation, Baseline: Supervision

multi = vglm(collapsed_sentence_type ~ as.factor(offenses_binned) + as.factor(GENDER) + CHARGE_COUNT+ AGE_AT_INCIDENT + as.factor(RACE), family = multinomial, data = forVGLM)

```
Call:
vglm(formula = collapsed_sentence_type ~ as.factor(offenses_binned) +
    as.factor(GENDER) + CHARGE_COUNT + AGE_AT_INCIDENT + as.factor(RACE),
    family = multinomial, data = forVGLM)

Coefficients:
                                                        Estimate Std. Error z value             Pr(>|z|)
(Intercept):1                                           0.259826   0.644433   0.403             0.686811
(Intercept):2                                           1.166687   0.644530   1.810             0.070275 .
as.factor(offenses_binned)DUI and License-related Crimes:1  0.085010   0.203467   0.418             0.676088
as.factor(offenses_binned)DUI and License-related Crimes:2  0.731948   0.204224   3.584             0.000338 ***
as.factor(offenses_binned)Firearm/Weapon-related Crimes:1  -0.447206   0.202974  -2.203             0.027576 *
as.factor(offenses_binned)Firearm/Weapon-related Crimes:2  -0.522457   0.203922  -2.562             0.010406 *
as.factor(offenses_binned)Fraud and Deception:1         0.056471   0.221883   0.255             0.799104
as.factor(offenses_binned)Fraud and Deception:2         1.035458   0.221907   4.666       0.000003068433 ***
as.factor(offenses_binned)Legal process crimes:1       -0.329699   0.442632  -0.745             0.456356
as.factor(offenses_binned)Legal process crimes:2        0.012437   0.439415   0.028             0.977420
as.factor(offenses_binned)Narcotics:1                   1.258289   0.207467   6.065       0.000000001319 ***
as.factor(offenses_binned)Narcotics:2                   1.734198   0.208270   8.327 < 0.0000000000000002 ***
as.factor(offenses_binned)Property Crimes:1             0.313530   0.203624   1.540             0.123622
as.factor(offenses_binned)Property Crimes:2             0.562539   0.204417   2.752             0.005925 **
as.factor(offenses_binned)Sex Crimes:1                  1.723165   0.274453   6.279       0.000000000342 ***
as.factor(offenses_binned)Sex Crimes:2                  1.609832   0.275280   5.848       0.000000004976 ***
as.factor(offenses_binned)Violent Crimes:1              0.694992   0.202783   3.427             0.000610 ***
as.factor(offenses_binned)Violent Crimes:2              0.564627   0.203655   2.772             0.005563 **
as.factor(GENDER)Male:1                                 1.667628   0.047600  35.034 < 0.0000000000000002 ***
as.factor(GENDER)Male:2                                 0.554328   0.046723  11.864 < 0.0000000000000002 ***
CHARGE_COUNT:1                                          0.030134   0.028767   1.048             0.294867
CHARGE_COUNT:2                                          0.057874   0.028708   2.016             0.043804 *
AGE_AT_INCIDENT:1                                       0.025656   0.001982  12.945 < 0.0000000000000002 ***
AGE_AT_INCIDENT:2                                       0.008551   0.001983   4.311       0.000016231547 ***
as.factor(RACE)Asian:1                                 -0.851530   0.627878  -1.356             0.175034
as.factor(RACE)Asian:2                                  0.056065   0.626687   0.089             0.928715
as.factor(RACE)Biracial:1                              -0.599728   1.231979  -0.487             0.626400
as.factor(RACE)Biracial:2                               0.580636   1.195939   0.486             0.627318
as.factor(RACE)Black:1                                  1.095674   0.607487   1.804             0.071291 .
as.factor(RACE)Black:2                                  0.972840   0.607414   1.602             0.109242
as.factor(RACE)Hispanic:1                               0.049130   0.614621   0.080             0.936288
as.factor(RACE)Hispanic:2                               0.681418   0.614403   1.109             0.267399
as.factor(RACE)White:1                                  0.029462   0.607579   0.048             0.961325
as.factor(RACE)White:2                                  0.630818   0.607489   1.038             0.299083
---
```

Multinomial Logistic Regression w/ Race Evaluation

- No race parameters achieve statistical significance across both levels, which begs the question whether the other statistically insignificant parameters have high p-values because of the effect of race parameters
- Based on this data set, we conclude race is not a strong enough predictor for sentencing
- Future consideration: latent variables that describe race such as law enforcement agency or district where the crime was committed could be used

Multinomial Logistic Regression w/o Race

```
Call:
vglm(formula = collapsed_sentence_type ~ as.factor(offenses_binned) +
    as.factor(GENDER) + CHARGE_COUNT + AGE_AT_INCIDENT, family = multinomial,
    data = forVGLM)

Coefficients:
                                                      Estimate Std. Error z value              Pr(>|z|)
(Intercept):1                                         0.774774   0.214741   3.608              0.000309 ***
(Intercept):2                                         1.959141   0.215639   9.085 < 0.0000000000000002 ***
as.factor(offenses_binned)DUI and License-related Crimes:1  0.055244   0.203008   0.272              0.785525
as.factor(offenses_binned)DUI and License-related Crimes:2  0.718436   0.204145   3.519              0.000433 ***
as.factor(offenses_binned)Firearm/Weapon-related Crimes:1 -0.099799   0.202041  -0.494              0.621339
as.factor(offenses_binned)Firearm/Weapon-related Crimes:2 -0.421404   0.203387  -2.072              0.038272 *
as.factor(offenses_binned)Fraud and Deception:1       0.312678   0.221116   1.414              0.157336
as.factor(offenses_binned)Fraud and Deception:2       1.105677   0.221530   4.991    0.0000006004106459 ***
as.factor(offenses_binned)Legal process crimes:1     -0.207476   0.440793  -0.471              0.637863
as.factor(offenses_binned)Legal process crimes:2      0.035705   0.439070   0.081              0.935188
as.factor(offenses_binned)Narcotics:1                 1.565658   0.206720   7.574    0.0000000000000362 ***
as.factor(offenses_binned)Narcotics:2                 1.816690   0.207892   8.739 < 0.0000000000000002 ***
as.factor(offenses_binned)Property Crimes:1           0.469823   0.203056   2.314              0.020681 *
as.factor(offenses_binned)Property Crimes:2           0.594715   0.204232   2.912              0.003592 **
as.factor(offenses_binned)Sex Crimes:1                1.814065   0.274069   6.619    0.0000000000361609 ***
as.factor(offenses_binned)Sex Crimes:2                1.632676   0.275217   5.932    0.0000000029866872 ***
as.factor(offenses_binned)Violent Crimes:1            0.873807   0.202233   4.321    0.0000155469028357 ***
as.factor(offenses_binned)Violent Crimes:2            0.606325   0.203480   2.980              0.002885 **
as.factor(GENDER)Male:1                               1.640396   0.047401  34.607 < 0.0000000000000002 ***
as.factor(GENDER)Male:2                               0.541416   0.046650  11.606 < 0.0000000000000002 ***
CHARGE_COUNT:1                                        0.031624   0.028739   1.100              0.271162
CHARGE_COUNT:2                                        0.059268   0.028723   2.063              0.039075 *
AGE_AT_INCIDENT:1                                     0.025235   0.001955  12.906 < 0.0000000000000002 ***
AGE_AT_INCIDENT:2                                     0.007725   0.001959   3.944    0.0000802868844053 ***
```

Model Parameter Interpretations

- Overall, the odds of observing probation or prison over supervision for most offenses increases compared to non-violent crime
- In particular, sexual assault and narcotics multiply the odds ratio by the greatest factor, especially the odds ratio for prison over supervision
- Interesting enough, firearms related offenses decreased the odds for both probation and prison over supervision compared to non-violent crimes
- The multiplicative impact of male gender relative to female gender was positive (especially for prison compared to supervision), suggesting men go to prison over receiving supervision at a much higher rate
    - Makes sense based on EDA where men disproportionately commit violent/drug/sex offenses

Model Parameter Interpretations

- The odds of observing probation over supervision multiplies by a factor of 1.06106 for a one unit increase in charge count
  - 10 charges = 3.27x odds(probation)/odds(supervision)
- The odds of observing prison over supervision multiples by a factor of 1.025556 for a one-unit increase in age
  - Given fixed levels, the odds of a 30-year-old receiving prison is 28% higher than a 20-year-old
  - Younger has more leeway because “oh they’re just a dumb kid who made a mistake”

Model Significance

- Likelihood Ratio Test
- Test statistic: 18512 on 32 df for chi-squared distribution
- P-value: < 2.2 x 10^-16
- **Conclusion: We reject the null hypothesis that all Betas = 0 and conclude at least one parameter is statistically significant**

Prediction

```
> new
  offenses_binned GENDER CHARGE_COUNT AGE_AT_INCIDENT
1      Sex Crimes   Male            1              20
2      Sex Crimes Female            1              20
> predict(multi2,new,type='response')
     Prison Probation Supervision
1 0.6000554 0.3948333 0.005111336
2 0.3312750 0.6541722 0.014552813
```

Questions?