

Race Paper Final

Ryan Larson - UMN

02/03/2021

Data Munging

Analysis with Misdemeanor, Gross Misdemeanor, and Felony Cases

```
#bar graph - median household income comes from ACS 2017 5-year estimates
adj <- data.frame(Race = c("Asian", "Black", "Hispanic", "Nat. Am.", "Other", "White"),
                  race_med_inc = c(71865, 32743, 46232, 36017, 42476, 69069),
                  ovr_med_inc = rep(65699, 6))

monsanc.short %>%
  select(race_impute, total_ff) %>%
  mutate(Race = case_when(
    race_impute=="asian"~"Asian",
    race_impute=="black"~"Black",
    race_impute=="hispanic"~"Hispanic",
    race_impute=="nat. am."~"Nat. Am.",
    race_impute=="other"~"Other",
    race_impute=="white"~"White"
  )) %>%
  filter(!is.na(Race)) %>%
  select(-race_impute) %>%
  left_join(adj, by = "Race") %>%
  mutate(race_adj = (ovr_med_inc/race_med_inc)) %>%
  group_by(Race) %>%
  summarize(
    `Raw_USD` = mean(total_ff, na.rm = T),
    `Adjusted_USD` = mean(total_ff*race_adj, na.rm = T),
    `Raw_sd` = sd(total_ff, na.rm = T),
    `Adjusted_sd` = sd(total_ff*race_adj, na.rm=T)
  ) %>%
  gather(key = "variable", value = "value", -Race) %>%
  separate(variable, into = c("adjusted", "stat"), sep = "_") %>%
  spread(key = "stat", value = "value") %>%
  mutate(Race = factor(Race, levels =
    c("Asian", "Hispanic", "White",
      "Other", "Nat. Am.", "Black"))),
```

```

  `Income Adjusted` = factor(adjusted, levels = c("Raw", "Adjusted")))) %>%
  select(-adjusted) %>%
ggplot()+
  geom_bar(aes(x=Race, y=USD, fill=`Income Adjusted`),
    stat="identity",
    position = position_dodge2())+
  geom_text(aes(x=Race, y=USD+25, group=`Income Adjusted`,
    label = round(USD,0)),
    position = position_dodge2(width = 1))+
  labs(title = "Figure 1: Fine/Fee Orders by Race")+
  scale_fill_manual(values = c("#00AFBB", "#FC4E07"))+
  theme_minimal()

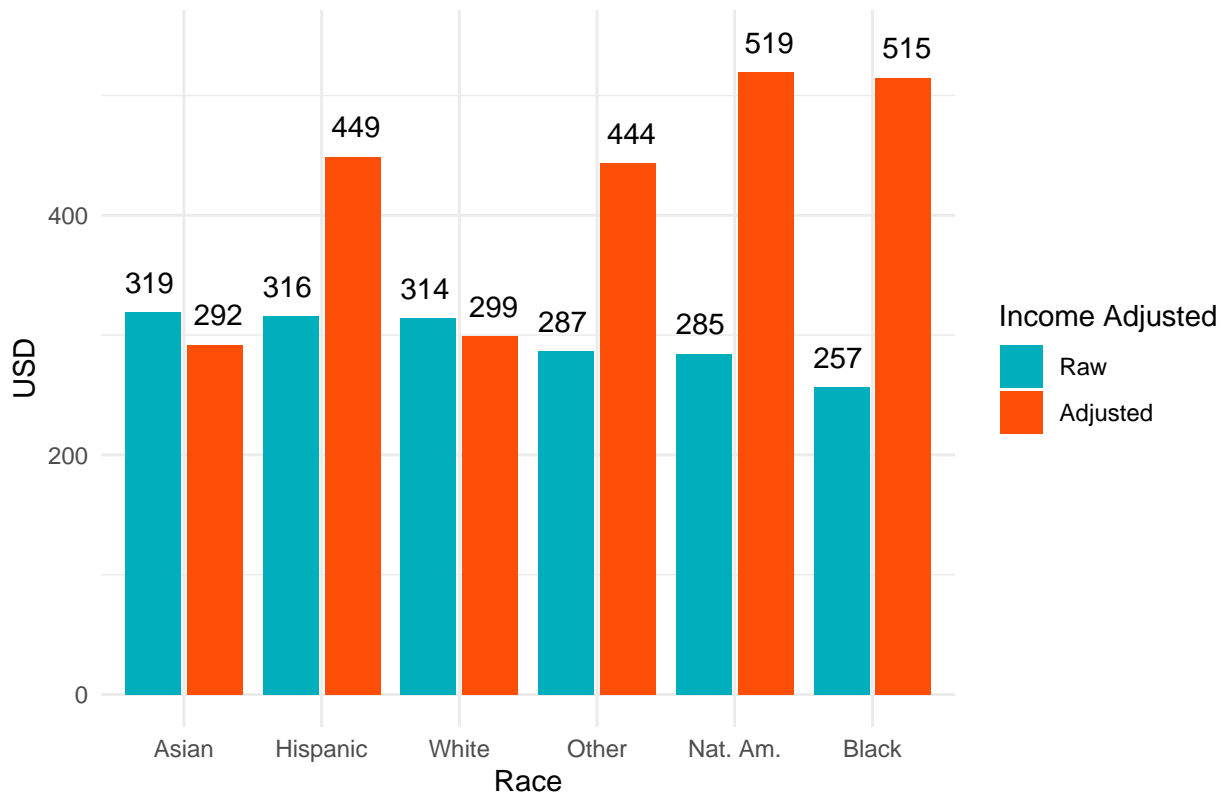
```

```

## Warning: Column `Race` joining character vector and factor, coercing into
## character vector

```

Figure 1: Fine/Fee Orders by Race



```

ggsave("figure_1.png", device = "png")

```

```

## Saving 6.5 x 4.5 in image

```

```

#confinement bar graph
monsanc.short %>%
  select(race_impute, conf_minus_stayed) %>%
  mutate(Race = case_when(
    race_impute=="asian"~"Asian",

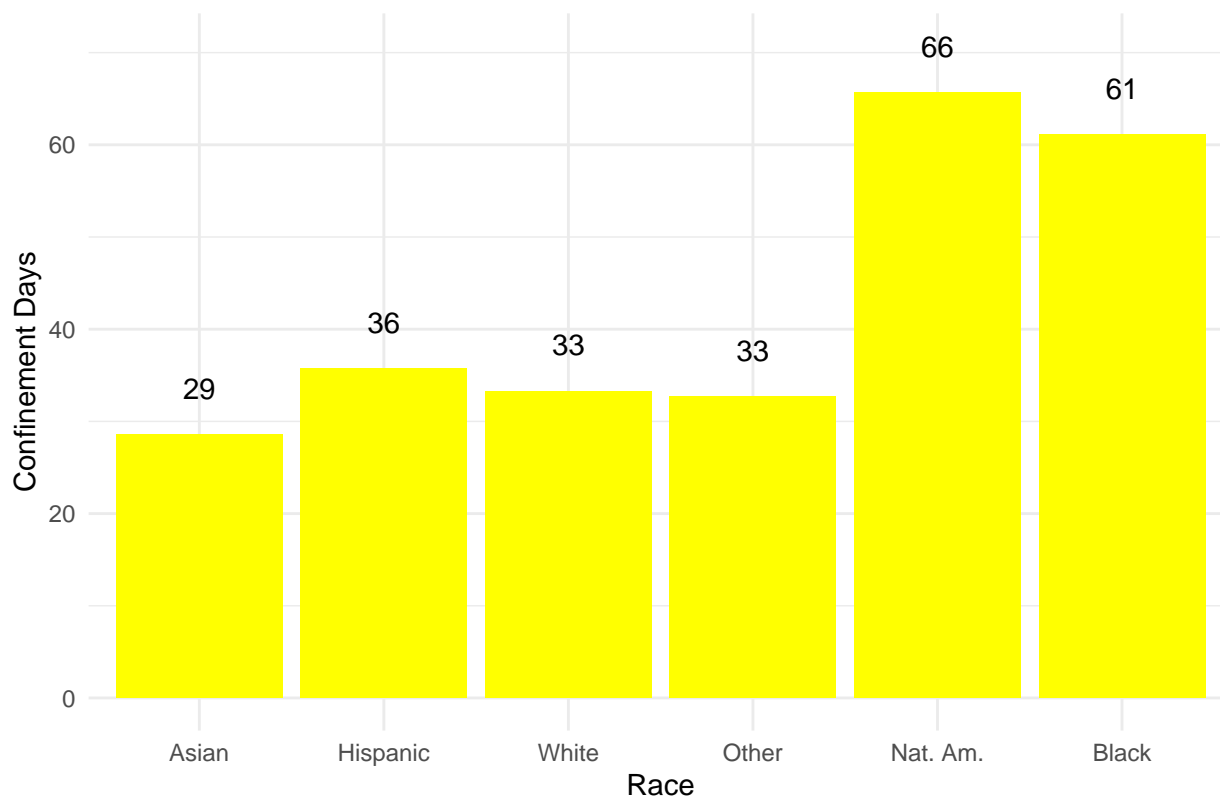
```

```

    race_impute=="black"~"Black",
    race_impute=="hispanic"~"Hispanic",
    race_impute=="nat. am."~"Nat. Am.",
    race_impute=="other"~"Other",
    race_impute=="white"~"White"
  )) %>%
  filter(!is.na(Race)) %>%
  select(-race_impute) %>%
  group_by(Race) %>%
  summarize(
    `Confinement Days` = mean(conf_minus_stayed, na.rm = T)
  ) %>%
  gather(key = "variable", value = "value", -Race) %>%
  separate(variable, into = "stat", sep = "_") %>%
  spread(key = "stat", value = "value") %>%
  mutate(Race = factor(Race, levels =
    c("Asian", "Hispanic", "White",
      "Other", "Nat. Am.", "Black")))
  ) %>%
  ggplot()+
  geom_bar(aes(x=Race, y=`Confinement Days`),
    fill = "yellow",
    stat="identity",
    position = position_dodge2())+
  geom_text(aes(x=Race, y=`Confinement Days`+5,
    label = round(`Confinement Days`,0)),
    position = position_dodge2(width = 1))+
  labs(title = "Figure 2: Confinement Length by Race")+
  theme_minimal()+
  theme(legend.position = "none")

```

Figure 2: Confinement Length by Race



```
ggsave("figure_2.png", device = "png")
```

```
## Saving 6.5 x 4.5 in image
```

```
#probation plot
```

```
monsanc.short %>%
```

```
  select(race_impute, prob_days) %>%
```

```
  mutate(Race = case_when(
```

```
    race_impute=="asian"~"Asian",
```

```
    race_impute=="black"~"Black",
```

```
    race_impute=="hispanic"~"Hispanic",
```

```
    race_impute=="nat. am."~"Nat. Am.",
```

```
    race_impute=="other"~"Other",
```

```
    race_impute=="white"~"White"
```

```
  )) %>%
```

```
  filter(!is.na(Race)) %>%
```

```
  select(-race_impute) %>%
```

```
  group_by(Race) %>%
```

```
  summarize(
```

```
    `Probation Days` = mean(prob_days, na.rm = T)
```

```
  ) %>%
```

```
  gather(key = "variable", value = "value", -Race) %>%
```

```
  separate(variable, into = "stat", sep = "_") %>%
```

```
  spread(key = "stat", value = "value") %>%
```

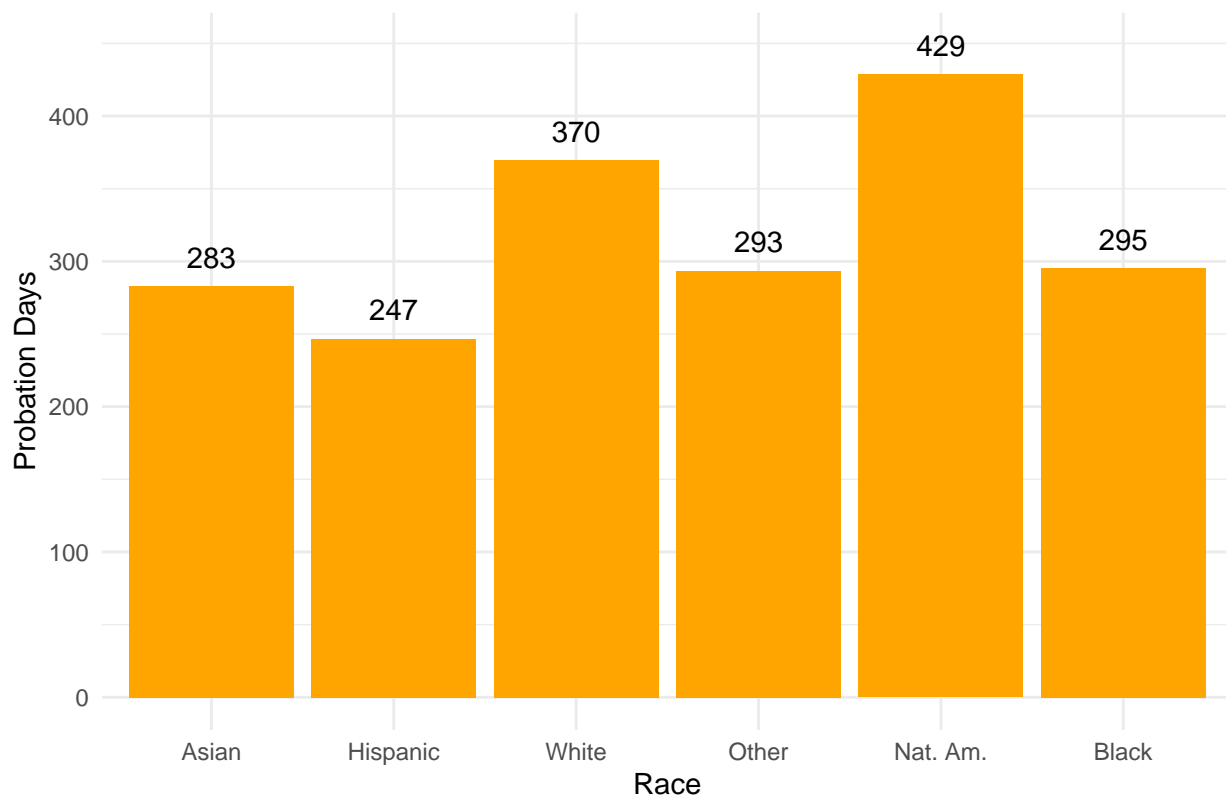
```
  mutate(Race = factor(Race, levels =
```

```

      c("Asian", "Hispanic", "White",
        "Other", "Nat. Am.", "Black"))
) %>%
ggplot()+
  geom_bar(aes(x=Race, y=`Probation Days`),
    fill = "orange",
    stat="identity",
    position = position_dodge2())+
  geom_text(aes(x=Race, y=`Probation Days`+20,
    label = round(`Probation Days`,0)),
    position = position_dodge2(width = 1))+
  labs(title = "Figure 3: Probation Length by Race")+
  theme_minimal()+
  theme(legend.position = "none")

```

Figure 3: Probation Length by Race



```

ggsave("figure_3.png", device = "png")

```

```

## Saving 6.5 x 4.5 in image

```

```

#Multivariate Regression - when all predictors are identical across models
#equivalent to separate OLS, but more efficient and takes into account error covariance

```

```

#multivariate regression

```

```

mv <- lm(cbind(log(total_ff+1), log(conf_minus_stayed+1), log(prob_days+1))~

```

```

        race_impute+gender_impute+log(age)+priors+perc_credit+trial_flag+charge_degree+
        charge_offense, data = monsanc.short)
summary(mv)

```

```

## Response log(total_ff + 1) :
##
## Call:
## lm(formula = `log(total_ff + 1)` ~ race_impute + gender_impute +
##     log(age) + priors + perc_credit + trial_flag + charge_degree +
##     charge_offense, data = monsanc.short)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5.872 -0.333  0.275  0.699 77.666
##
## Coefficients:
##
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      5.262e+00  1.270e-02  414.309 < 2e-16 ***
## race_imputeasian      6.054e-02  7.271e-03   8.325 < 2e-16 ***
## race_imputeblack     -4.299e-01  3.312e-03 -129.791 < 2e-16 ***
## race_imputehispanic    4.710e-02  4.770e-03   9.874 < 2e-16 ***
## race_imputemissing    -5.112e-02  3.001e-03  -17.035 < 2e-16 ***
## race_imputenat. am.   -2.449e-01  6.008e-03  -40.754 < 2e-16 ***
## race_imputeother     -1.516e-01  9.571e-03  -15.837 < 2e-16 ***
## gender_imputeM        3.017e-02  2.503e-03   12.054 < 2e-16 ***
## log(age)           -3.564e-02  3.301e-03  -10.799 < 2e-16 ***
## priors             -1.442e-02  1.459e-04  -98.871 < 2e-16 ***
## perc_credit        -2.093e-03  2.979e-05  -70.257 < 2e-16 ***
## trial_flagTRUE      -6.737e-01  1.188e-01   -5.673  1.4e-08 ***
## charge_degreeFelony  -4.577e-01  5.575e-03  -82.095 < 2e-16 ***
## charge_degreeGross Misdemeanor  1.178e-01  4.412e-03   26.703 < 2e-16 ***
## charge_offensealcohol/dui    5.081e-01  6.232e-03   81.543 < 2e-16 ***
## charge_offensedrug    2.863e-01  8.196e-03   34.935 < 2e-16 ***
## charge_offensehunt/fish    6.572e-03  1.027e-02    0.640    0.522
## charge_offenseother    2.922e-01  5.673e-03   51.503 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.327 on 1387817 degrees of freedom
## (117194 observations deleted due to missingness)
## Multiple R-squared:  0.05559,    Adjusted R-squared:  0.05558
## F-statistic: 4805 on 17 and 1387817 DF,  p-value: < 2.2e-16
##
##
## Response log(conf_minus_stayed + 1) :
##
## Call:

```

```

## lm(formula = `log(conf_minus_stayed + 1)` ~ race_impute + gender_impute +
##      log(age) + priors + perc_credit + trial_flag + charge_degree +
##      charge_offense, data = monsanc.short)
##
## Residuals:
##      Min      1Q  Median      3Q      Max
## -182.461  -0.336  -0.108   0.058   9.197
##
## Coefficients:
##
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      6.121e-01  9.887e-03   61.910 < 2e-16 ***
## race_imputeasian      2.767e-02  5.660e-03    4.888 1.02e-06 ***
## race_imputeblack      1.420e-01  2.578e-03   55.091 < 2e-16 ***
## race_impuتهispanic      4.273e-02  3.713e-03   11.508 < 2e-16 ***
## race_imputemissing     -7.607e-02  2.336e-03  -32.561 < 2e-16 ***
## race_imputenat. am.      4.302e-01  4.677e-03   91.985 < 2e-16 ***
## race_imputeother      4.647e-02  7.451e-03    6.238 4.45e-10 ***
## gender_imputeM      1.206e-01  1.948e-03   61.902 < 2e-16 ***
## log(age)          5.862e-02  2.569e-03   22.816 < 2e-16 ***
## priors            2.659e-02  1.135e-04  234.203 < 2e-16 ***
## perc_credit        5.010e-03  2.319e-05  216.025 < 2e-16 ***
## trial_flagTRUE      1.936e+00  9.245e-02   20.938 < 2e-16 ***
## charge_degreeFelony      3.088e+00  4.340e-03  711.572 < 2e-16 ***
## charge_degreeGross Misdemeanor 1.679e+00  3.435e-03  488.863 < 2e-16 ***
## charge_offensealcohol/dui     -5.642e-01  4.851e-03 -116.321 < 2e-16 ***
## charge_offensedrug     -9.580e-01  6.380e-03 -150.167 < 2e-16 ***
## charge_offensehunt/fish     -9.884e-01  7.992e-03 -123.677 < 2e-16 ***
## charge_offenseother     -8.881e-01  4.416e-03 -201.130 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.033 on 1387817 degrees of freedom
## (117194 observations deleted due to missingness)
## Multiple R-squared:  0.514, Adjusted R-squared:  0.514
## F-statistic: 8.635e+04 on 17 and 1387817 DF, p-value: < 2.2e-16
##
##
## Response log(prob_days + 1) :
##
## Call:
## lm(formula = `log(prob_days + 1)` ~ race_impute + gender_impute +
##      log(age) + priors + perc_credit + trial_flag + charge_degree +
##      charge_offense, data = monsanc.short)
##
## Residuals:
##      Min      1Q  Median      3Q      Max
## -14.1280  -1.8672  -0.6629   2.3790  11.1334
##

```

```
## Coefficients:
##
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    2.228e+00  2.431e-02   91.663 < 2e-16 ***
## race_imputeasian -1.801e-01  1.392e-02  -12.944 < 2e-16 ***
## race_imputeblack  1.525e-01  6.339e-03   24.049 < 2e-16 ***
## race_imputehispanic -4.194e-01  9.130e-03  -45.940 < 2e-16 ***
## race_imputemissing -1.477e+00  5.744e-03 -257.132 < 2e-16 ***
## race_imputenat. am.  2.698e-01  1.150e-02   23.461 < 2e-16 ***
## race_imputeother  6.778e-02  1.832e-02    3.700 0.000216 ***
## gender_imputeM    -2.940e-01  4.790e-03  -61.378 < 2e-16 ***
## log(age)         4.570e-01  6.317e-03   72.340 < 2e-16 ***
## priors           -1.260e-02  2.792e-04  -45.123 < 2e-16 ***
## perc_credit      5.006e-04  5.702e-05    8.779 < 2e-16 ***
## trial_flagTRUE   -1.921e+00  2.273e-01   -8.449 < 2e-16 ***
## charge_degreeFelony  2.255e+00  1.067e-02  211.366 < 2e-16 ***
## charge_degreeGross Misdemeanor 3.233e+00  8.445e-03  382.764 < 2e-16 ***
## charge_offensealcohol/dui  -3.591e-02  1.193e-02   -3.011 0.002605 **
## charge_offensedrug  -9.825e-01  1.569e-02  -62.637 < 2e-16 ***
## charge_offensehunt/fish -2.518e+00  1.965e-02 -128.138 < 2e-16 ***
## charge_offenseother -1.611e+00  1.086e-02 -148.399 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.54 on 1387817 degrees of freedom
## (117194 observations deleted due to missingness)
## Multiple R-squared:  0.2951, Adjusted R-squared:  0.2951
## F-statistic: 3.417e+04 on 17 and 1387817 DF, p-value: < 2.2e-16
```

```
#predictions for different hypothetical cases
```

```
library(emmeans)
```

```
pred <- emmeans(mv, ~race_impute|Type, mult.name = "Type",
                weights="proportional", type = "response") %>%
  as.data.frame() %>%
  mutate(Type = case_when(
    Type==1~"LFO",
    Type==2~"Confinement",
    Type==3~"Probation"
  ),
  race_impute = case_when(
    race_impute=="asian"~"Asian",
    race_impute=="black"~"Black",
    race_impute=="hispanic"~"Hispanic",
    race_impute=="nat. am."~"Nat. Am.",
    race_impute=="other"~"Other",
    race_impute=="white"~"White",
    race_impute=="missing"~"Missing")) %>%
  filter(race_impute!="Missing") %>%
```

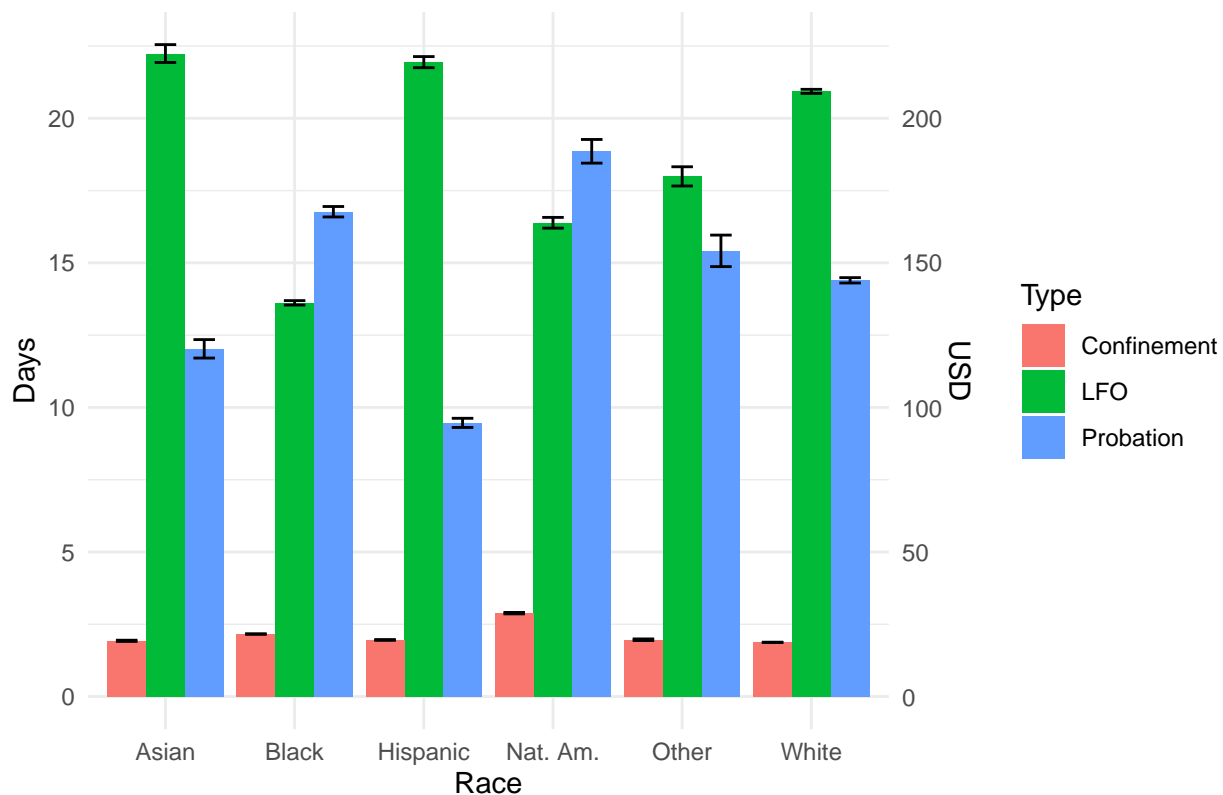


```
mutate(response = ifelse(Type=="LFO", response/10, response),
       lower.CL = ifelse(Type=="LFO", lower.CL/10, lower.CL),
       upper.CL = ifelse(Type=="LFO", upper.CL/10, upper.CL))
```

```
## Warning in (function (object, at, cov.reduce = mean, cov.keep = get_emm_option("cov.keep"),
## Auto-detection of the response transformation may be incorrect
```

```
ggplot(pred, aes(x=race_impute, y=response, fill=Type))+
  geom_bar(stat="identity",
          position = "dodge") +
  geom_errorbar(aes(min=lower.CL, ymax=upper.CL),
               width = .5,
               position = position_dodge(.9))+
  scale_y_continuous(name = "Days",
                    sec.axis = sec_axis( trans=~.*10, name="USD"))
)+
labs(title = "Figure 4: Estimated Marginal Punishment Packages",
     x = "Race")+
theme_minimal()
```

Figure 4: Estimated Marginal Punishment Packages



```
ggsave("figure_4.png", device = "png")
```

```
## Saving 6.5 x 4.5 in image
```

```

#descriptive stats for mv model
monsanc.short$miss <- seq_len(nrow(monsanc.short)) %in% na.action(mv)

#descriptive statistics for variables in analysis
ds <- monsanc.short %>%
  filter(miss==FALSE) %>%
  select(total_ff, conf_minus_stayed, prob_days,
         white,black,hispanic,asian,nativeam,other.race, race.miss,
         male,age,priors,trial_flag,
         felony,gross.mis,mis, petty.mis,
         violent,drug,alcohol.dui,other.offense,
         cap_ratio) %>%
  gather() %>%
  group_by(key) %>%
  summarize(
    mean = round(mean(value, na.rm=T),2),
    sd = round(sd(value, na.rm=T),2),
    min = min(value, na.rm=T),
    max = max(value, na.rm=T)
  )

print(ds, n = Inf)

```

```

## # A tibble: 23 x 5
##   key                mean      sd   min     max
##   <chr>             <dbl>  <dbl> <dbl>   <dbl>
## 1 age              33.0   12.4   15     115
## 2 alcohol.dui       0.15   0.36    0         1
## 3 asian             0.03   0.16    0         1
## 4 black             0.17   0.37    0         1
## 5 cap_ratio         0.75   0.15   0.2     1.15
## 6 conf_minus_stayed 31.7  318.    0    44087
## 7 drug              0.03   0.17    0         1
## 8 felony            0.06   0.24    0         1
## 9 gross.mis         0.09   0.28    0         1
## 10 hispanic         0.06   0.25    0         1
## 11 male             0.71   0.46    0         1
## 12 mis              0.85   0.35    0         1
## 13 nativeam         0.04   0.19    0         1
## 14 other.offense     0.75   0.43    0         1
## 15 other.race        0.01   0.12    0         1
## 16 petty.mis         0       0       0         0
## 17 priors           6.28   8.38    0        236
## 18 prob_days        235.   575.    0    133225
## 19 race.miss        0.24   0.43    0         1
## 20 total_ff        294.   211.    0        1000
## 21 trial_flag       0       0.01    0         1

```

```
## 22 violent          0.05  0.22  0      1
## 23 white            0.45  0.5   0      1

#FE instrumental Variables Regression
monsanc.short.iv <- monsanc.short %>% drop_na(total_ff_log,conf_minus_stayed_log,
black,hispanic,asian,nativeam,other.race,race.miss,male,age,
priors,perc_credit,trial_flag,felony,gross.mis,prison_flag, hunt.fish,
drug,alcohol. DUI,other.offense,filed_county,sentence_year,conf_minus_stayed_log)
```

#LFO

```
iv.conf <- ivreg(total_ff_log~conf_minus_stayed_log+
black+hispanic+asian+nativeam+other.race+race.miss+male+log(age)+
priors+perc_credit+trial_flag+felony+gross.mis+
drug+alcohol. DUI+hunt.fish+other.offense|
.-conf_minus_stayed_log+cap_ratio,
data = monsanc.short.iv)
```

```
summary(iv.conf, diagnostics=T)
```

```
##
## Call:
## ivreg(formula = total_ff_log ~ conf_minus_stayed_log + black +
##       hispanic + asian + nativeam + other.race + race.miss + male +
##       log(age) + priors + perc_credit + trial_flag + felony + gross.mis +
##       drug + alcohol. DUI + hunt.fish + other.offense | . - conf_minus_stayed_log +
##       cap_ratio, data = monsanc.short.iv)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1554.6024   -2.9144    -0.8905     0.7361    78.4913
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    10.824344   0.465081  23.274 < 2e-16 ***
## conf_minus_stayed_log -9.065604   0.743965 -12.186 < 2e-16 ***
## black           0.855485   0.108485   7.886 3.13e-15 ***
## hispanic        0.426813   0.045781   9.323 < 2e-16 ***
## asian           0.305024   0.054707   5.576 2.47e-08 ***
## nativeam        3.663259   0.324148  11.301 < 2e-16 ***
## other.race       0.247784   0.074792   3.313 0.000923 ***
## race.miss       -0.744350   0.060559 -12.291 < 2e-16 ***
## male            1.119053   0.091025  12.294 < 2e-16 ***
## log(age)         0.496963   0.049436  10.053 < 2e-16 ***
## priors           0.228987   0.020032  11.431 < 2e-16 ***
## perc_credit      0.042726   0.003681  11.606 < 2e-16 ***
## trial_flagTRUE   16.684883   1.651683  10.102 < 2e-16 ***
## felony           27.557763   2.299711  11.983 < 2e-16 ***
```

```
## gross.mis          15.334780    1.249210  12.276 < 2e-16 ***
## drug               -8.447721    0.719087 -11.748 < 2e-16 ***
## alcohol. DUI      -4.614090    0.422687 -10.916 < 2e-16 ***
## hunt.fish         -8.970016    0.740283 -12.117 < 2e-16 ***
## other.offense     -7.773046    0.663101 -11.722 < 2e-16 ***
##
## Diagnostic tests:
##               df1      df2 statistic p-value
## Weak instruments      1 1371732      145.3 <2e-16 ***
## Wu-Hausman           1 1371731      7089.2 <2e-16 ***
## Sargan                0      NA          NA      NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.255 on 1371732 degrees of freedom
## Multiple R-Squared:  -44.73, Adjusted R-squared:  -44.73
## Wald test: 102.3 on 18 and 1371732 DF, p-value: < 2.2e-16
```

```
#Probation
iv.prob <- ivreg(prob_days_log~conf_minus_stayed_log+
  black+hispanic+asian+nativeam+other.race+race.miss+male+log(age)+
  priors+perc_credit+trial_flag+felony+gross.mis+
  drug+alcohol. DUI+hunt.fish+other.offense|
  .-conf_minus_stayed_log+cap_ratio,
  data = monsanc.short.iv)

summary(iv.prob, diagnostics=T)
```

```
##
## Call:
## ivreg(formula = prob_days_log ~ conf_minus_stayed_log + black +
##       hispanic + asian + nativeam + other.race + race.miss + male +
##       log(age) + priors + perc_credit + trial_flag + felony + gross.mis +
##       drug + alcohol. DUI + hunt.fish + other.offense | . - conf_minus_stayed_log +
##       cap_ratio, data = monsanc.short.iv)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.0383  -1.7324  -0.6979   2.5065  104.3762
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.840934   0.133981  13.740 < 2e-16 ***
## conf_minus_stayed_log 0.662788   0.214322   3.092  0.00198 **
## black           0.061276   0.031253   1.961  0.04992 *
## hispanic       -0.450356   0.013189 -34.147 < 2e-16 ***
## asian          -0.197548   0.015760 -12.535 < 2e-16 ***
## nativeam       -0.040735   0.093381  -0.436  0.66268
```

```

## other.race          0.037055    0.021546    1.720    0.08547 .
## race.miss          -1.430342    0.017446   -81.987    < 2e-16 ***
## male               -0.374140    0.026222   -14.268    < 2e-16 ***
## log(age)           0.412569    0.014242    28.969    < 2e-16 ***
## priors             -0.030211    0.005771    -5.235    1.65e-07 ***
## perc_credit        -0.002754    0.001061    -2.596    0.00942 **
## trial_flagTRUE     -3.141821    0.475819    -6.603    4.03e-11 ***
## felony             0.204425    0.662503     0.309    0.75765
## gross.mis          2.119891    0.359874     5.891    3.85e-09 ***
## drug              -0.338087    0.207155    -1.632    0.10267
## alcohol.dui         0.342184    0.121768     2.810    0.00495 **
## hunt.fish          -1.860361    0.213262    -8.723    < 2e-16 ***
## other.offense      -1.017640    0.191027    -5.327    9.97e-08 ***
##
## Diagnostic tests:
##               df1      df2 statistic  p-value
## Weak instruments      1 1371732      145.3 < 2e-16 ***
## Wu-Hausman           1 1371731       15.1 0.000102 ***
## Sargan                0      NA         NA      NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.666 on 1371732 degrees of freedom
## Multiple R-Squared: 0.2242, Adjusted R-squared: 0.2241
## Wald test: 2.9e+04 on 18 and 1371732 DF, p-value: < 2.2e-16

```