Investigating Racial Profiling

MCMC Modeling for the NYC Stop-and-Frisk Policy

Chen Shi, Belle Xu

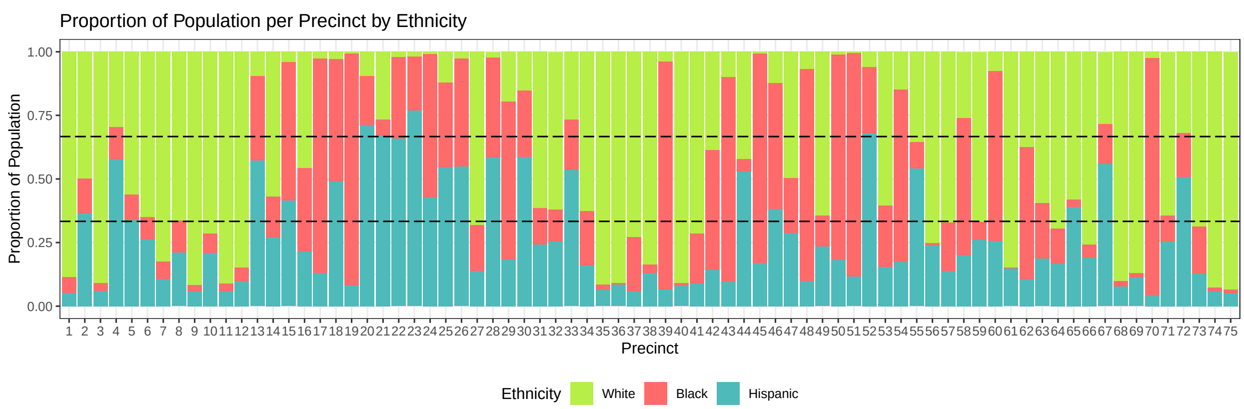
4/21/2021

```
stop <- read.table("stop-and-frisk.dat", header = TRUE)
```

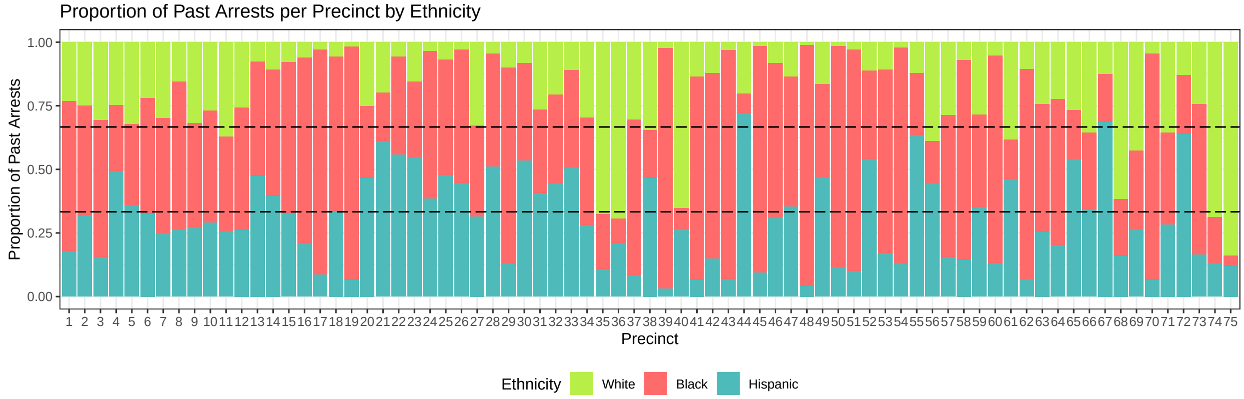
Exploratory Data Analysis

```
# Proportion of stops for each ethnicity in each precinct
stop %>%
  mutate(eth = factor(eth, levels = c("3", "1", "2")),
         precinct = factor(precinct)) %>%
  group_by(eth, precinct) %>%
  mutate(stop_eth = sum(stops)) %>%
  ggplot(mapping = aes(x = precinct, y = stop_eth, fill = eth)) +
  geom_bar(position = "fill", stat = "identity") +
  geom_hline(yintercept = 1/3, linetype=5) +
  geom_hline(yintercept = 2/3, linetype=5) +
  labs(y = "Proportion of stops", x = "Precinct",
       title = "Proportion of Stops per Precinct by Ethnicity")+
  theme_bw() +
  theme(legend.position = "bottom") +
  scale_fill_discrete(name = "Ethnicity",
                      labels=c("White", "Black", "Hispanic"),
                      type = c("#b7ee47","#ff6b6b","#4ebaba"))
```

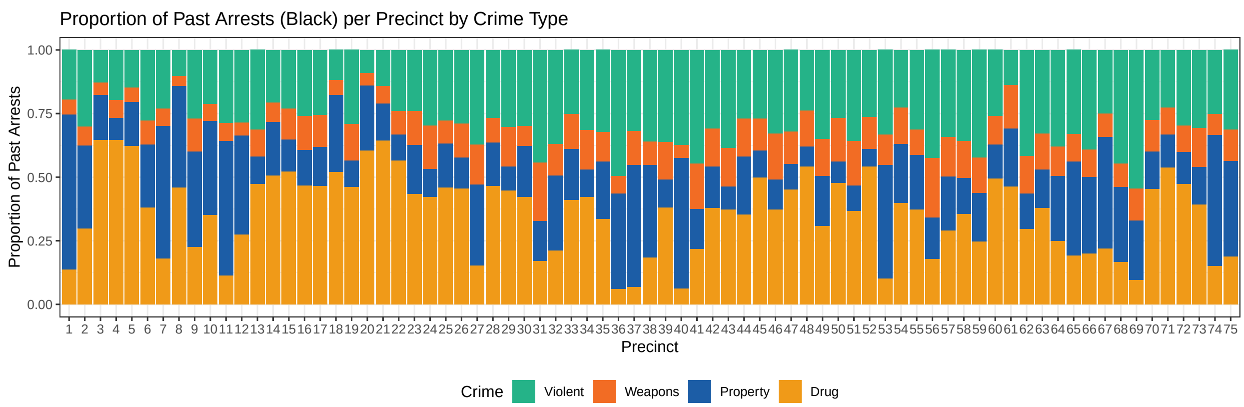
```
stop %>%
  mutate(eth = factor(eth, levels = c("3", "1", "2")),
         precinct = factor(precinct)) %>%
  group_by(eth, precinct) %>%
  ggplot(mapping = aes(x = precinct, y = pop, fill = eth)) +
  geom_bar(position = "fill", stat = "identity") +
  geom_hline(yintercept = 1/3, linetype=5) +
  geom_hline(yintercept = 2/3, linetype=5) +
  labs(title = "Proportion of Population per Precinct by Ethnicity",
         y = "Proportion of Population", x = "Precinct") +
  theme_bw() +
  theme(legend.position = "bottom") +
  scale_fill_discrete(name = "Ethnicity",
                      labels=c("White", "Black", "Hispanic"),
                      type = c("#b7ee47","#ff6b6b","#4ebaba"))
```




```
stop %>%
  mutate(eth = factor(eth, levels = c("3", "1", "2")),
         precinct = factor(precinct)) %>%
  group_by(eth, precinct) %>%
  mutate(crime_eth = sum(past.arrests)) %>%
  ggplot(mapping = aes(x = precinct, y = crime_eth, fill = eth)) +
  geom_bar(position = "fill", stat = "identity") +
  geom_hline(yintercept = 1/3, linetype=5) +
  geom_hline(yintercept = 2/3, linetype=5) +
  labs(title = "Proportion of Past Arrests per Precinct by Ethnicity",
       y = "Proportion of Past Arrests", x = "Precinct") +
  theme_bw() +
  theme(legend.position = "bottom") +
  scale_fill_discrete(name = "Ethnicity",
                      labels=c("White", "Black", "Hispanic"),
                      type = c("#b7ee47","#ff6b6b","#4ebaba"))
```

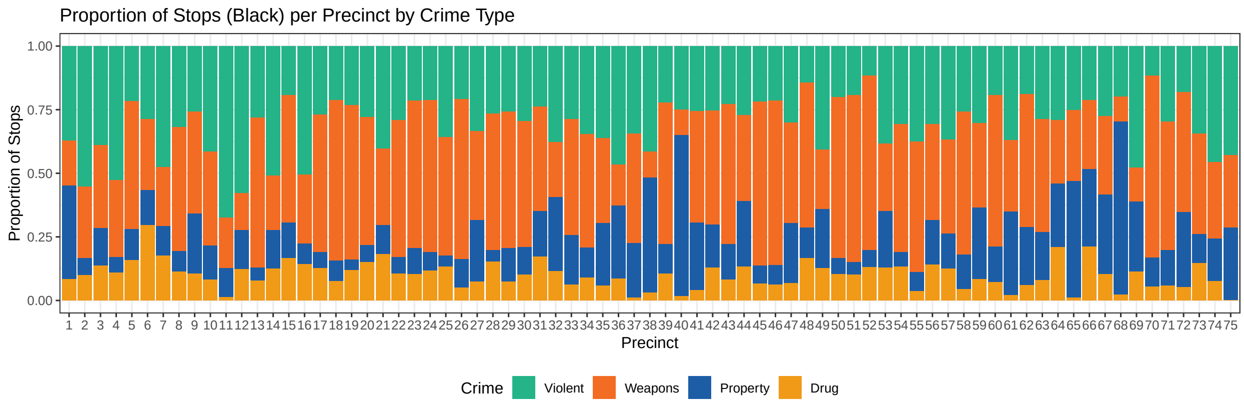



```
stop %>%
  mutate(eth = factor(eth, levels = c("3", "1", "2")),
         precinct = factor(precinct),
         crime = factor(crime)) %>%
  filter(eth == "1") %>%
  ggplot(mapping = aes(x = precinct, y = past.arrests, fill = crime)) +
  geom_bar(position = "fill", stat = "identity") +
  labs(title = "Proportion of Past Arrests (Black) per Precinct by Crime Type",
       y = "Proportion of Past Arrests", x = "Precinct") +
  theme_bw() +
  theme(legend.position = "bottom") +
  scale_fill_discrete(name = "Crime",
                      labels=c("Violent", "Weapons", "Property", "Drug"),
                      type = c("#25B388","#F26C24","#1C5DA6", "#F09A18"))
```

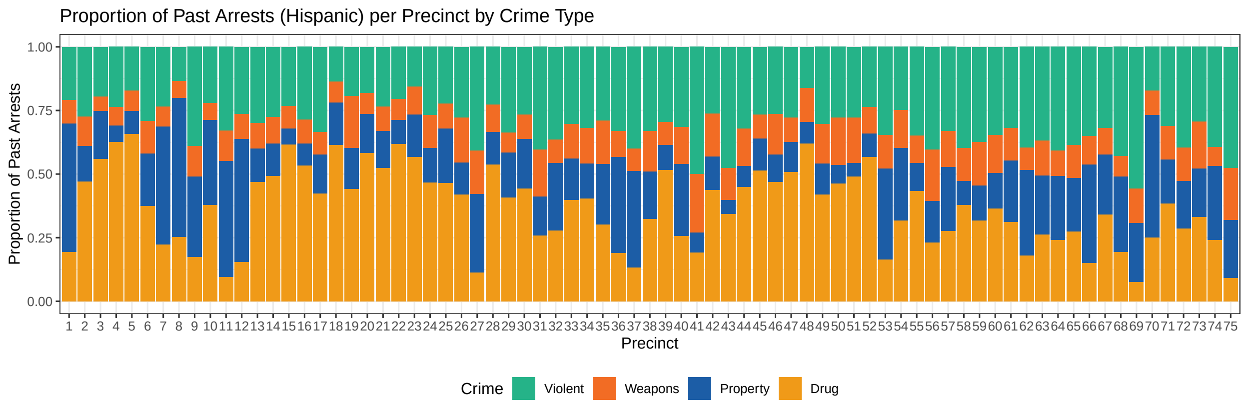



```
stop %>%
  mutate(eth = factor(eth, levels = c("3", "1", "2")),
         precinct = factor(precinct),
         crime = factor(crime)) %>%
  filter(eth == "1") %>%
  ggplot(mapping = aes(x = precinct, y = stops, fill = crime)) +
  geom_bar(position = "fill", stat = "identity") +
  labs(title = "Proportion of Stops (Black) per Precinct by Crime Type",
       y = "Proportion of Stops", x = "Precinct") +
  theme_bw() +
```

```
  theme(legend.position = "bottom") +
  scale_fill_discrete(name = "Crime",
                      labels=c("Violent", "Weapons", "Property", "Drug"),
                      type = c("#25B388","#F26C24","#1C5DA6", "#F09A18"))
```

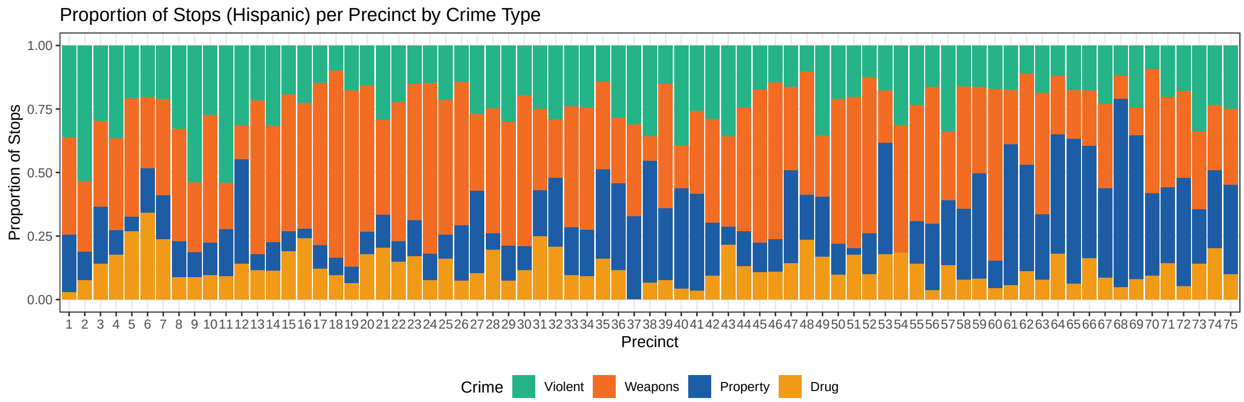



```
stop %>%
  mutate(eth = factor(eth, levels = c("3", "1", "2")),
         precinct = factor(precinct),
         crime = factor(crime)) %>%
  filter(eth == "2") %>%
  ggplot(mapping = aes(x = precinct, y = past.arrests, fill = crime)) +
  geom_bar(position = "fill", stat = "identity") +
  labs(title = "Proportion of Past Arrests (Hispanic) per Precinct by Crime Type",
       y = "Proportion of Past Arrests", x = "Precinct") +
  theme_bw() +
  theme(legend.position = "bottom") +
  scale_fill_discrete(name = "Crime",
                      labels=c("Violent", "Weapons", "Property", "Drug"),
                      type = c("#25B388","#F26C24","#1C5DA6", "#F09A18"))
```

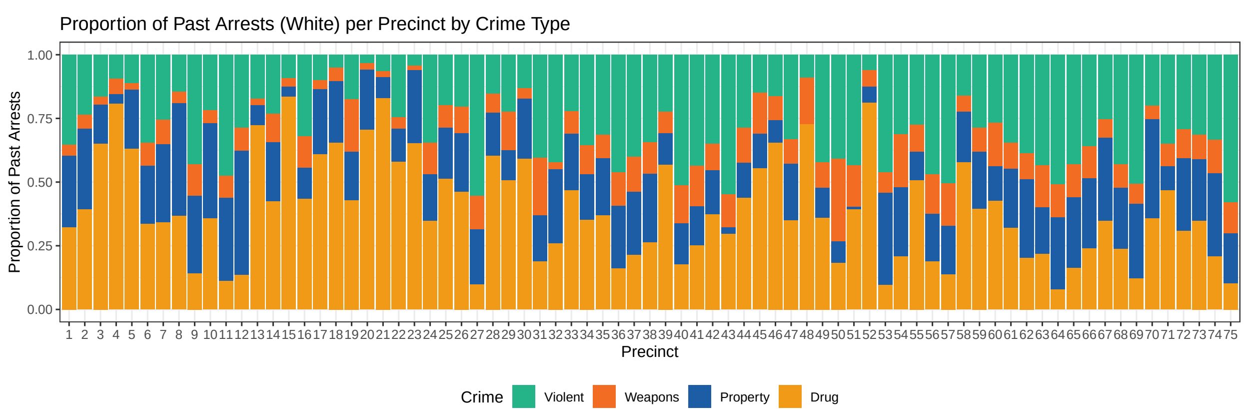



```
stop %>%
  mutate(eth = factor(eth, levels = c("3", "1", "2")),
         precinct = factor(precinct),
         crime = factor(crime)) %>%
```

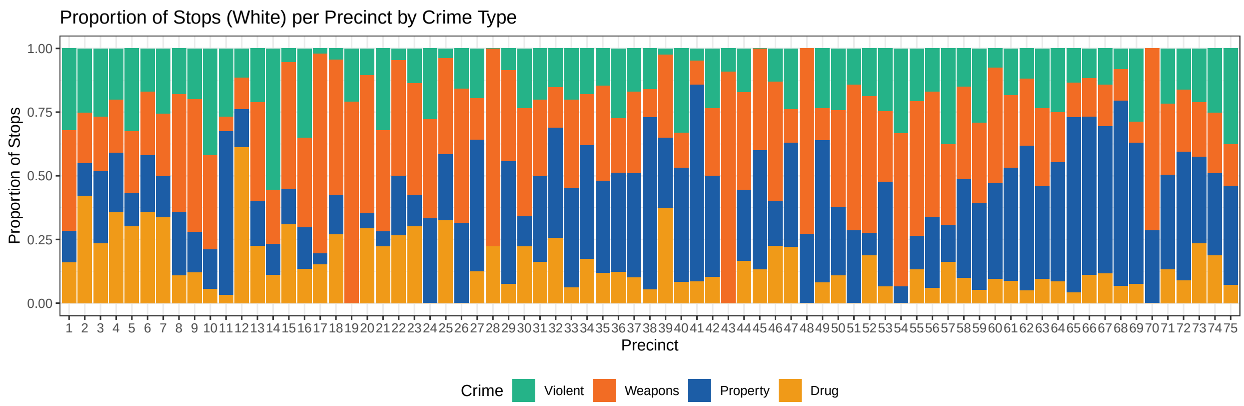
```
  filter(eth == "2") %>%
  ggplot(mapping = aes(x = precinct, y = stops, fill = crime)) +
  geom_bar(position = "fill", stat = "identity") +
  labs(title = "Proportion of Stops (Hispanic) per Precinct by Crime Type",
       y = "Proportion of Stops", x = "Precinct") +
  theme_bw() +
  theme(legend.position = "bottom") +
  scale_fill_discrete(name = "Crime",
                      labels=c("Violent", "Weapons", "Property", "Drug"),
                      type = c("#25B388","#F26C24","#1C5DA6", "#F09A18"))
```



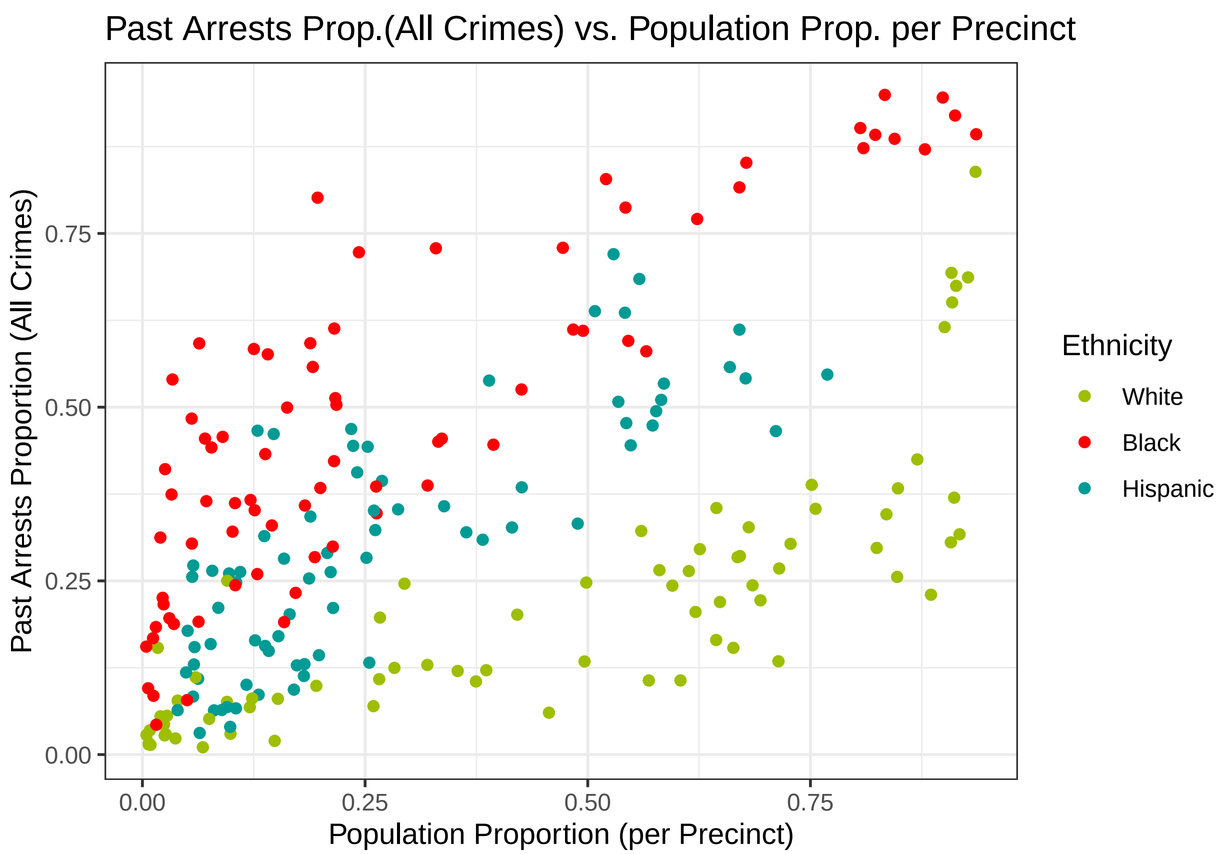

```
stop %>%
  mutate(eth = factor(eth, levels = c("3", "1", "2")),
         precinct = factor(precinct),
         crime = factor(crime)) %>%
  filter(eth == "3") %>%
  ggplot(mapping = aes(x = precinct, y = past.arrests, fill = crime)) +
  geom_bar(position = "fill", stat = "identity") +
  labs(title = "Proportion of Past Arrests (White) per Precinct by Crime Type",
       y = "Proportion of Past Arrests", x = "Precinct") +
  theme_bw() +
  theme(legend.position = "bottom") +
  scale_fill_discrete(name = "Crime",
                      labels=c("Violent", "Weapons", "Property", "Drug"),
                      type = c("#25B388","#F26C24","#1C5DA6", "#F09A18"))
```



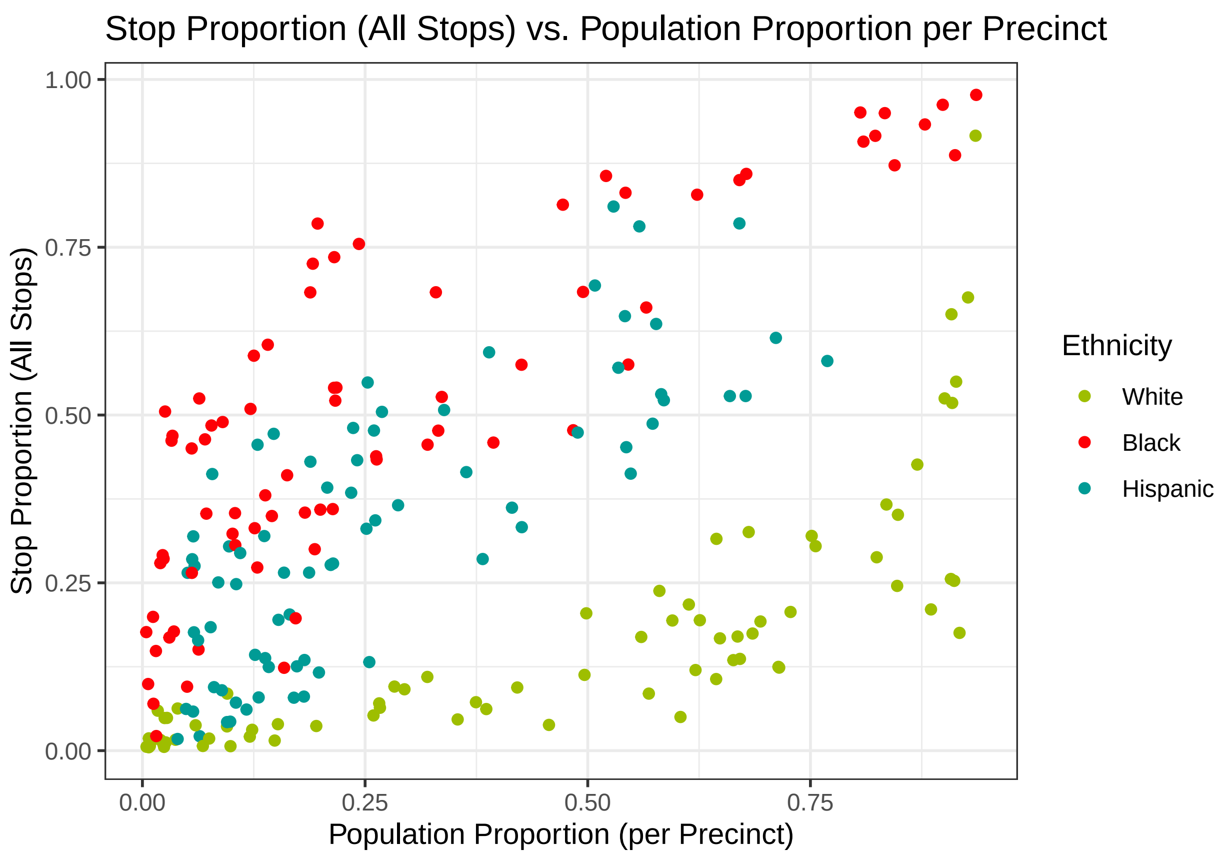
```
stop %>%
  mutate(eth = factor(eth, levels = c("3", "1", "2")),
         precinct = factor(precinct),
         crime = factor(crime)) %>%
  filter(eth == "3") %>%
  ggplot(mapping = aes(x = precinct, y = stops, fill = crime)) +
  geom_bar(position = "fill", stat = "identity") +
  labs(title = "Proportion of Stops (White) per Precinct by Crime Type",
       y = "Proportion of Stops", x = "Precinct") +
  theme_bw() +
  theme(legend.position = "bottom") +
  scale_fill_discrete(name = "Crime",
                      labels=c("Violent", "Weapons", "Property", "Drug"),
                      type = c("#25B388","#F26C24","#1C5DA6", "#F09A18"))
```




```
stop %>%
  mutate(eth = factor(eth, levels = c("3", "1", "2")),
         precinct = factor(precinct),
         crime = factor(crime)) %>%
  group_by(precinct) %>%
  mutate(total_pop = sum(pop) / 4,
         pop_prop = pop / total_pop,
         total_arrest = sum(past.arrests),
         crime_prop = past.arrests / total_arrest) %>%
  group_by(eth, precinct) %>%
  mutate(eth_crime = sum(crime_prop)) %>%
  filter(crime == 1) %>%
  ggplot(mapping = aes(x = pop_prop, y = eth_crime, color = eth)) +
  geom_point() + theme_bw()+
  labs(title="Past Arrests Prop.(All Crimes) vs. Population Prop. per Precinct",
       x = "Population Proportion (per Precinct)",
       y = "Past Arrests Proportion (All Crimes)") +
  scale_colour_discrete(name = "Ethnicity",
                        labels=c("White", "Black", "Hispanic"),
                        type = c("#9EBE00","#FD0006","#009B95"))
```




```
stop %>%
  mutate(eth = factor(eth, levels = c("3", "1", "2")),
         precinct = factor(precinct),
         crime = factor(crime)) %>%
  group_by(precinct) %>%
  mutate(total_pop = sum(pop) / 4,
         pop_prop = pop / total_pop,
         total_stop = sum(stops),
         stop_prop = stops / total_stop) %>%
  group_by(eth, precinct) %>%
  mutate(eth_stop = sum(stop_prop)) %>%
  filter(crime == 1) %>%
  ggplot(mapping = aes(x = pop_prop, y = eth_stop, color = eth)) +
  geom_point() + theme_bw()+
  labs(title = "Stop Proportion (All Stops) vs. Population Proportion per Precinct",
       x = "Population Proportion (per Precinct)", y = "Stop Proportion (All Stops)") +
  scale_colour_discrete(name = "Ethnicity",
                        labels=c("White", "Black", "Hispanic"),
                        type = c("#9EBE00","#FD0006","#009B95"))
```




```r
## Data Cleaning
total_crime <- stop %>%
  mutate(eth = factor(eth, levels = c("3", "1", "2")),
         precinct = factor(precinct),
         crime = factor(crime)) %>%
  group_by(precinct) %>%
  mutate(total_pop = sum(pop) / 4,
         pop_prop = pop / total_pop,
         total_arrest = sum(past.arrests),
         crime_prop = past.arrests / total_arrest) %>%
  mutate(past.arrests = if_else(past.arrests!= 0, past.arrests, as.integer(1)))
```

```r
mean(total_crime$stops)
```

```
## [1] 146.0222
```

```r
var(total_crime$stops)
```

```
## [1] 47254.93
```

Modeling

```r
model.1 <- stan_glm(data = total_crime,
                    formula = stops ~ eth * (pop_prop + crime) + pop_prop * crime,
                    family = neg_binomial_2(link = "log"),
                    offset = log(past.arrests),
                    seed = 360,
                    prior = cauchy(0, 2.5),
                    prior_intercept = cauchy(0, 2.5),
                    refresh = 0,
                   diagnostic_file = file.path(tempdir(), "glm1.csv"))
```

```r
model.2 <- stan_glm(data = total_crime,
                    formula = stops ~ eth * (crime_prop + pop_prop + crime) +
                      pop_prop * crime_prop + pop_prop * crime +
                      crime_prop * crime,
                    family = neg_binomial_2(link = "log"),
                    prior = cauchy(0, 2.5),
                    prior_intercept = cauchy(0, 2.5),
                    seed = 360,
                    refresh = 0,
                   diagnostic_file = file.path(tempdir(), "glm2.csv"))
```

```r
model.3 <- stan_glm(data = total_crime,
                    formula = stops ~ log(past.arrests)+
                      eth * (pop_prop + crime) + pop_prop * crime,
                    family = neg_binomial_2(link = "log"),
                    prior = cauchy(0, 2.5),
                    prior_intercept = cauchy(0, 2.5),
                    seed = 360,
                    refresh = 0,
                   diagnostic_file = file.path(tempdir(), "glm3.csv"))
```

```r
rstanarm::loo_compare(loo(model.1), loo(model.2), loo(model.3))
```

```
##         elpd_diff se_diff
## model.3   0.0       0.0
## model.2 -43.9      14.5
## model.1 -81.1      15.6
```

```r
model.3.full <- stan_glm(data = total_crime,
                    formula = stops ~  log(past.arrests)+
                      eth * (pop_prop + crime) + pop_prop * crime+
                      eth*log(past.arrests) + pop_prop*log(past.arrests) +
                      crime*log(past.arrests),
                    family = neg_binomial_2(link = "log"),
                    prior = cauchy(0, 2.5),
                    prior_intercept = cauchy(0, 2.5),
                    seed = 360,
                    refresh = 0,
                   diagnostic_file = file.path(tempdir(), "glm3full.csv"))
```

```r
bayesfactor_models(model.3, model.3.full, denominator = model.3.full)
```

```
## Warning: Bayes factors might not be precise.
## For precise Bayes factors, it is recommended sampling at least 40,000 posterior samples.
```

```
## Computation of Bayes factors: estimating marginal likelihood, please wait...
```

```
## Bayes Factors for Model Comparison
##
##     Model                                                              BF
## [1] log(past.arrests) + eth * (pop_prop + crime) + pop_prop * crime < 0.001
##
## * Against Denominator: [2] log(past.arrests) + eth * (pop_prop + crime) + pop_prop * crime + eth * lo
## *   Bayes Factor Type: marginal likelihoods (bridgesampling)
```

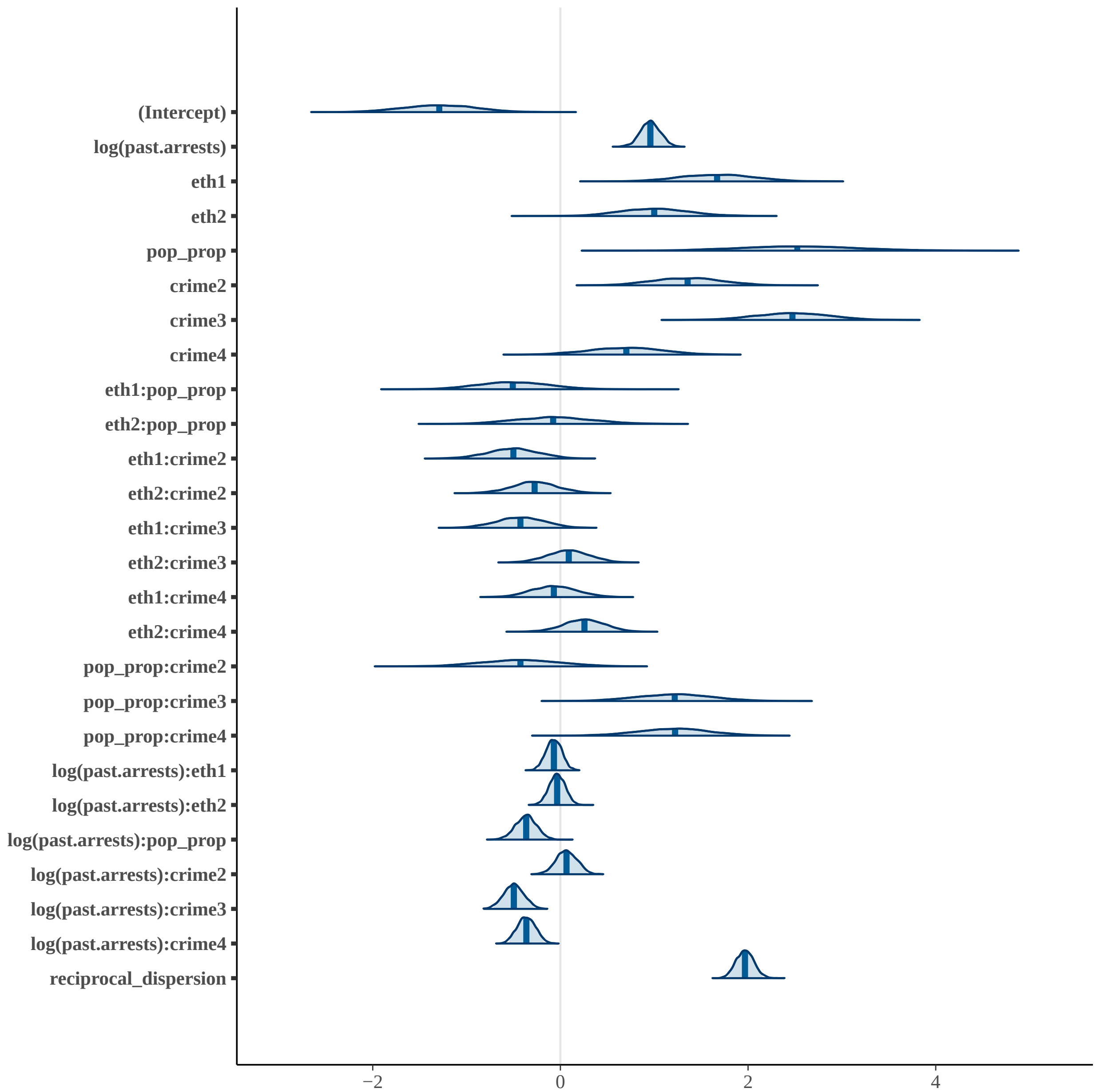
```r
rstanarm::loo_compare(loo(model.3), loo(model.3.full))
```

```
##              elpd_diff se_diff
## model.3.full   0.0       0.0
## model.3      -23.2       9.7
```

See diagnostic plots in the pdf file Diagnostic-Plots within this GitHub Repository.

Results

```r
mcmc_areas(as.matrix(model.3.full), prob = 0.95, prob_outer = 1)
```




```r
round(coef(model.3.full), 3)
```

```
##              (Intercept)        log(past.arrests)
##                   -1.291                    0.959
##                     eth1                     eth2
##                    1.671                    1.000
##                 pop_prop                   crime2
##                    2.525                    1.356
##                   crime3                   crime4
##                    2.473                    0.703
##            eth1:pop_prop            eth2:pop_prop
##                   -0.507                   -0.077
##              eth1:crime2              eth2:crime2
```

```
##                      -0.501                       -0.275
##                  eth1:crime3                  eth2:crime3
##                      -0.426                        0.089
##                  eth1:crime4                  eth2:crime4
##                      -0.070                        0.257
##              pop_prop:crime2              pop_prop:crime3
##                      -0.426                        1.218
##              pop_prop:crime4       log(past.arrests):eth1
##                       1.222                       -0.069
##       log(past.arrests):eth2 log(past.arrests):pop_prop
##                      -0.035                       -0.365
##     log(past.arrests):crime2     log(past.arrests):crime3
##                       0.065                       -0.496
##     log(past.arrests):crime4
##                      -0.363
```

```r
round(posterior_interval(model.3.full, prob = 0.95), 3)
```

```
##                              2.5%  97.5%
## (Intercept)                -2.045 -0.538
## log(past.arrests)           0.751  1.165
## eth1                        0.911  2.437
## eth2                        0.295  1.755
## pop_prop                    1.339  3.809
## crime2                      0.659  2.056
## crime3                      1.709  3.232
## crime4                     -0.034  1.432
## eth1:pop_prop              -1.208  0.243
## eth2:pop_prop              -0.843  0.749
## eth1:crime2                -1.051  0.013
## eth2:crime2                -0.744  0.205
## eth1:crime3                -0.920  0.054
## eth2:crime3                -0.347  0.511
## eth1:crime4                -0.532  0.414
## eth2:crime4                -0.179  0.679
## pop_prop:crime2            -1.240  0.412
## pop_prop:crime3             0.432  2.000
## pop_prop:crime4             0.423  1.960
## log(past.arrests):eth1     -0.234  0.089
## log(past.arrests):eth2     -0.199  0.127
## log(past.arrests):pop_prop -0.591 -0.154
## log(past.arrests):crime2   -0.154  0.282
## log(past.arrests):crime3   -0.709 -0.286
## log(past.arrests):crime4   -0.547 -0.177
## reciprocal_dispersion       1.790  2.156
```