

Exploring the ProPublica COMPAS data

Loading and filtering data

```
library(dplyr)
library(ggplot2)
library(grid)
library(gridExtra)
library(tidyverse)
```

After loading the raw data, we see the following number of rows:

```
raw_data <- read.csv("~/My Drive/tex-documents/working-papers/algo-fairness/algo-fairness-m/rutgers-pre")
nrow(raw_data)
```

```
## [1] 7214
```

By filtering the data a bit, the number of rows becomes:

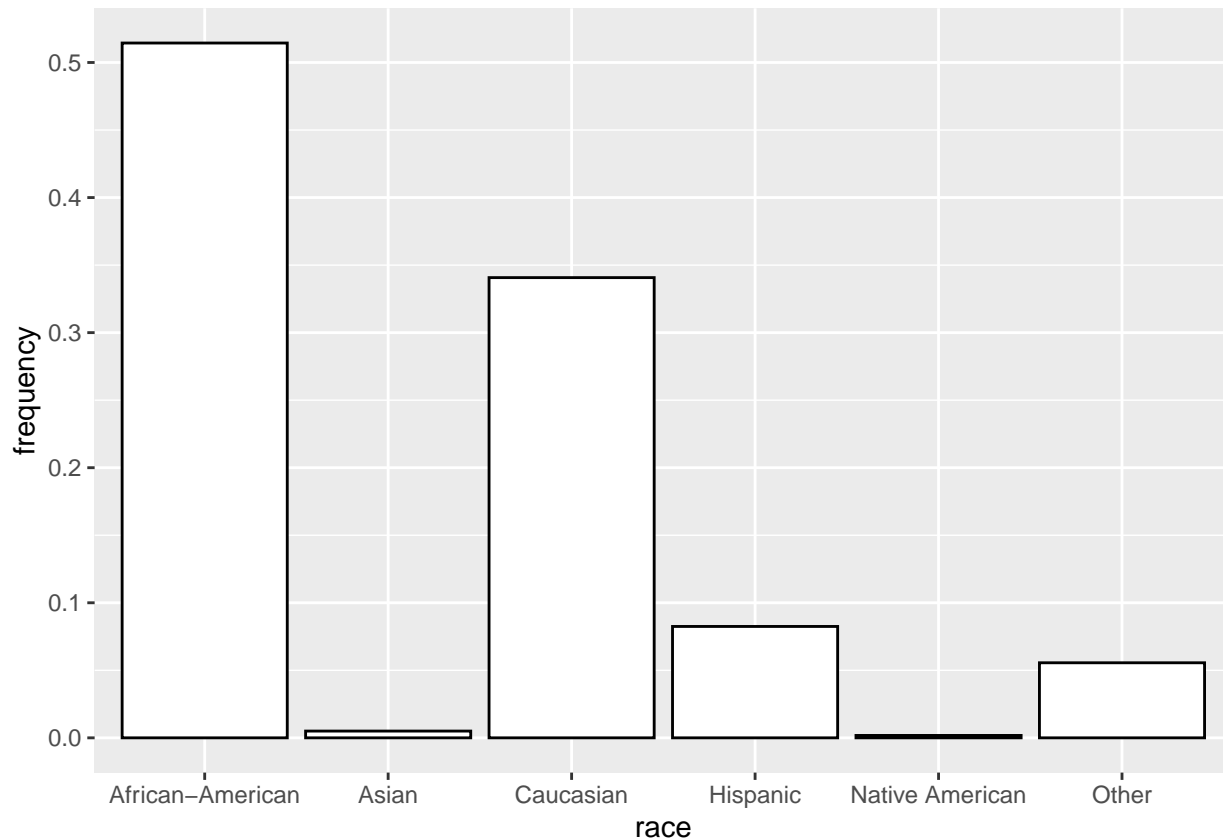
```
df <- dplyr::select(raw_data, age, c_charge_degree, race, age_cat, score_text, sex, priors_count,
                    days_b_screening_arrest, decile_score, is_recid, two_year_recid, c_jail_in, c_jail_out) %>%
  filter(days_b_screening_arrest <= 30) %>%
  filter(days_b_screening_arrest >= -30) %>%
  filter(is_recid != -1) %>%
  filter(c_charge_degree != "0") %>%
  filter(score_text != 'N/A')
nrow(df)
```

```
## [1] 6172
```

Race breakdown

The database is mostly constituted of blacks, whites and Hispanics. Asians and native-Americans are a small percentage.

```
race <- ggplot(df, aes(x=race)) +
  # geom_histogram(aes(y=..count../sum(..count..)), colour="black", fill="grey100") +
  geom_bar(aes(y=..count../sum(..count..)), colour="black", fill="grey100") +
  xlab("race") +
  ylab("frequency")
race
```



Recidivism

The recidivism rate is different across races. It is higher among blacks (almost 50%), while it is comparably lower among whites and Hispanics (about 40%) and even lower among Asians (about 25%).

Recidivism is defined as a rearrest for a criminal offense that occurred within two years after the COMPAS risk assessment took place. Arrest is therefore used as a proxy for criminal activity.

```
pblack <- ggplot(data=filter(df, race == "African-American"), aes(x=two_year_recid)) +
  geom_histogram(aes(y=..count../sum(..count..)), binwidth = 1, colour="black", fill="grey45") +
  ylab("frequency") +
  ylim(0, 1) +
  scale_x_continuous(breaks = seq(0, 1, by = 1)) +
  theme(plot.title = element_text(hjust = 0.5)) +
  theme(text=element_text(size = 15))

pwhite <- ggplot(data=filter(df, race == "Caucasian"), aes(x=two_year_recid)) +
  geom_histogram(aes(y=..count../sum(..count..)), binwidth = 1, colour="black", fill="grey100") +
  ylab("frequency") +
  ylim(0, 1) +
  scale_x_continuous(breaks = seq(0, 1, by = 1)) +
  theme(plot.title = element_text(hjust = 0.5)) +
  theme(text=element_text(size = 15))

phispanic <- ggplot(data=filter(df, race == "Hispanic"), aes(x=two_year_recid)) +
  geom_histogram(aes(y=..count../sum(..count..)), binwidth = 1, colour="black", fill="grey80") +
  ylab("frequency") +
```

```

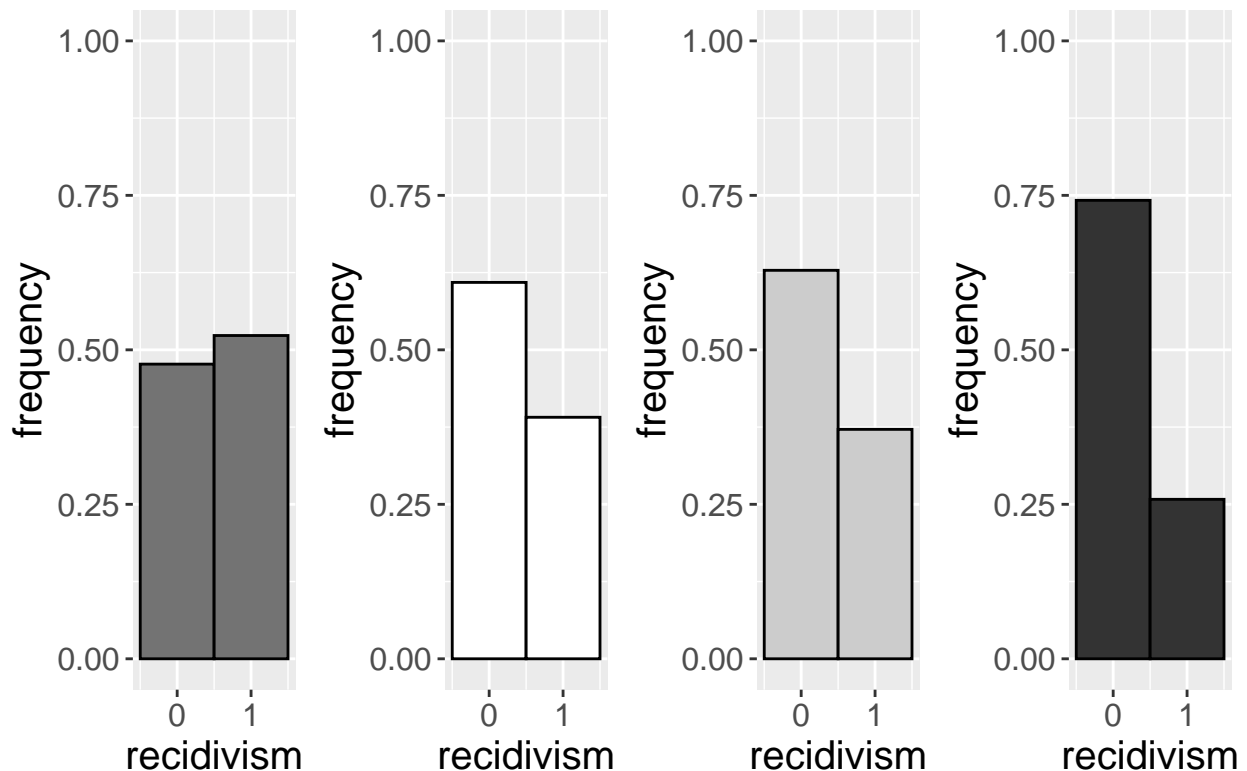
ylim(0, 1) +
scale_x_continuous(breaks = seq(0, 1, by = 1)) +
theme(plot.title = element_text(hjust = 0.5)) +
theme(text=element_text(size = 15))

pasian <- ggplot(data=filter(df, race == "Asian"), aes(x=two_year_recid)) +
geom_histogram(aes(y=..count../sum(..count..)), binwidth = 1, colour="black", fill="grey20")
ylab("frequency") +
ylim(0, 1) +
scale_x_continuous(breaks = seq(0, 1, by = 1)) +
theme(plot.title = element_text(hjust = 0.5)) +
theme(text=element_text(size = 15))

grid.arrange(pblack, pwhite, phispanic, pasian, ncol = 4, top=textGrob("Recidivism: Black, White, Hispanic, Asian"))

```

Recidivism: Black, White, Hispanic, Asia



Risk scores distribution by race

COMPAS assigns each individual a score between 1 (very low risk of recidivism) and 10 (high risk of recidivism). We can plot the distribution of these scores by race. The scores of blacks are evenly distributed across all values from 1 to 10. The scores of Hispanics tend to be more concentrated towards lower values. This trend is even more apparent for whites.

```

pblack_s <- ggplot(data=filter(df, race == "African-American"), aes(x=decile_score)) +
geom_histogram(aes(y=..count../sum(..count..)), binwidth = 1, colour="black", fill="grey45")
ylab("frequency") +
ylim(0, 0.3) +

```

```

scale_x_continuous(breaks = seq(0, 10, by = 1)) +
theme(plot.title = element_text(hjust = 0.5)) +
theme(text=element_text(size = 15))

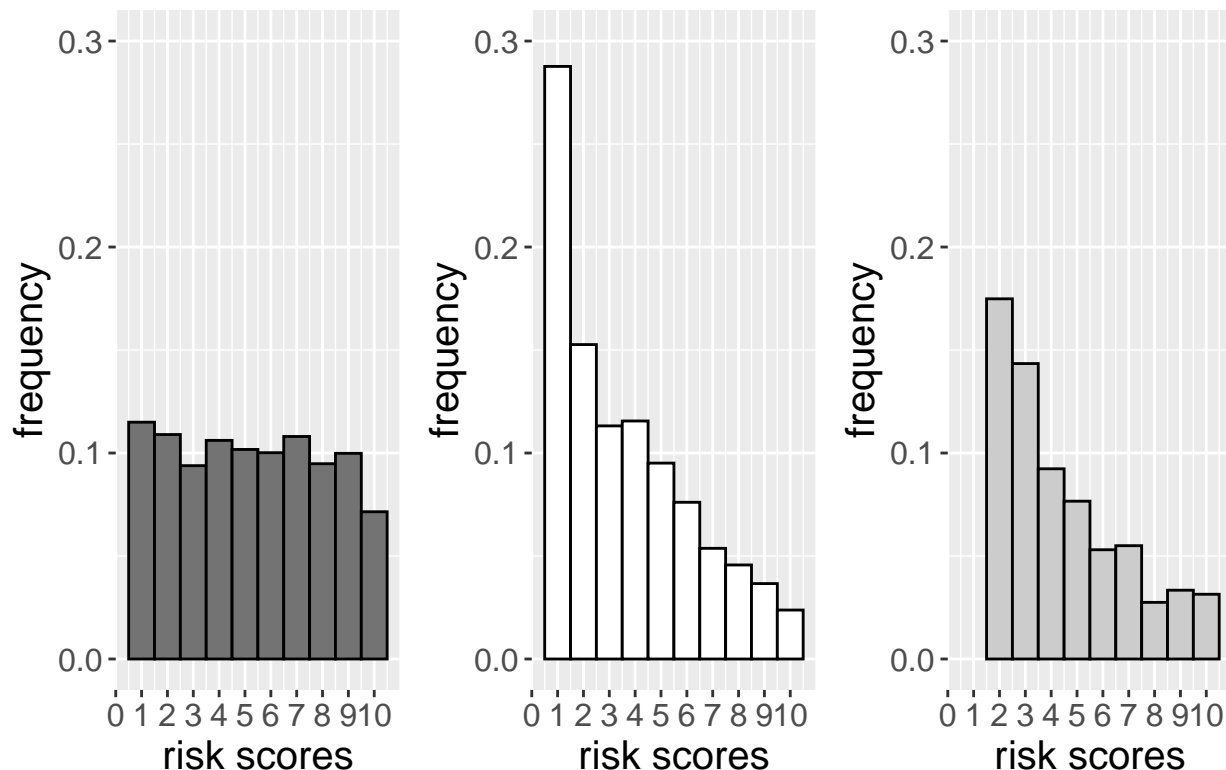
pwhite_s <- ggplot(data=filter(df, race == "Caucasian"), aes(x=decile_score)) +
  geom_histogram(aes(y=..count../sum(..count..)), binwidth = 1, colour="black", fill="grey100") +
  ylab("frequency") +
  ylim(0, 0.3) +
  scale_x_continuous(breaks = seq(0, 10, by = 1)) +
  theme(plot.title = element_text(hjust = 0.5)) +
  theme(text=element_text(size = 15))

phispanic_s <- ggplot(data=filter(df, race == "Hispanic"), aes(x=decile_score)) +
  geom_histogram(aes(y=..count../sum(..count..)), binwidth = 1, colour="black", fill="grey80") +
  ylab("frequency") +
  ylim(0, 0.3) +
  scale_x_continuous(breaks = seq(0, 10, by = 1)) +
  theme(plot.title = element_text(hjust = 0.5)) +
  theme(text=element_text(size = 15))

grid.arrange(pblack_s, pwhite_s, phispanic_s, ncol = 3, top=textGrob("Scores: Black, White, Hispanic",

```

Scores: Black, White, Hispanic

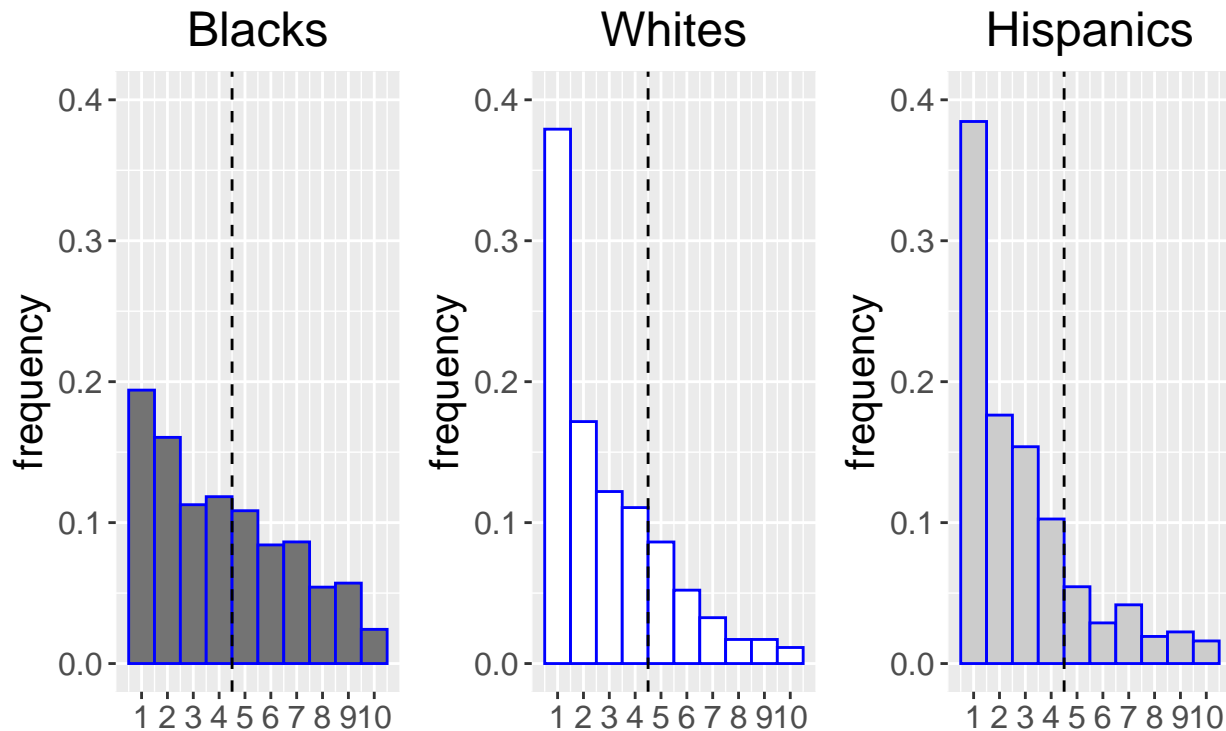


Risk scores distribution by race (non re-offenders only)

Let's focus on those individuals in the database for whom no criminal activity (i.e. arrest) was recorded within a period of two years. Call them non re-offenders. The distribution of the scores for these individuals

should be different, mostly concentrated toward lower values. If they did not re-offend, COMPAS should have been able to predict that and assign them lower scores. This is the case for the three racial groups – a sign that COMPAS is, to some extent, tracking future criminal behavior (i.e. arrest). However, the scores are more clearly concentrated towards lower values for whites and Hispanics, and much less clearly for blacks.

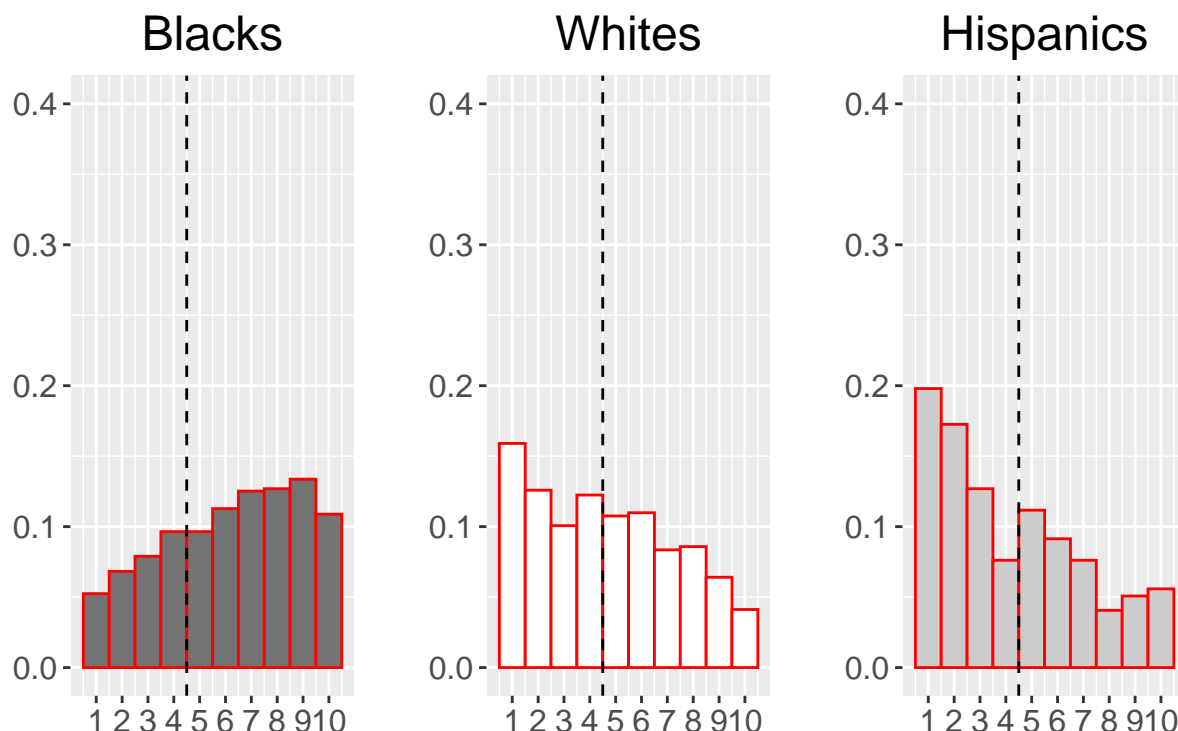
Scores for reoffenders



Risk scores distribution by race (re-offenders only)

Let's focus on those individuals in the database for whom criminal activity (i.e. arrest) was recorded within a period of two years. Call them re-offenders. The distribution of the scores for these individuals should be concentrated toward higher values. If they did re-offend, COMPAS should have been able to predict that and assign them higher scores. This is the case for blacks, but not for whites and Hispanics – a sign that COMPAS is, to some extent, tracking future criminal behavior (i.e. arrest) for blacks, but much less so for whites and Hispanics.

Scores for reoffenders



Racial Disparities in false positives classification

The false positives rates is higher for blacks compared to whites and Hispanics. COMPAS does not return a yes/no decision, but simply a risk score between 1 and 10. We can force a yes/no decision by stipulating that anyone with a score of at least 5 is classified as a re-offenders by COMPAS. So, a false positive occurs when someone who is not a re-offenders is classified as such by COMPAS. The false positive classification rate is defined as the following fraction: $\frac{\#[\text{individual is classified as reoffender but is not}]}{\#[\text{individual is not a re-offender}]}$.

This is the conditional probability $Pr(\text{individual is classified as reoffender} \mid \text{individual is not a reoffender})$.

The false positive classification rate is 41% for blacks, 21% for whites and 18% for Hispanics. There is a significant difference between blacks, on one hand, and whites and Hispanics on the other hand. The disparity does not go away if the threshold is increased to at least 7 instead of 5. The false positive classification rate becomes 21% for blacks, 8% for whites and 10% for Hispanics.

```
black_I <- subset(df, race == "African-American" & is_recid==0)
black_I_Cg <- subset(df, race == "African-American" & is_recid==0 & decile_score>4)

fp_black <- nrow(black_I_Cg)/nrow(black_I)

print(paste0("Compas false positive rate for blacks: ", fp_black))

## [1] "Compas false positive rate for blacks: 0.414407988587732"

white_I <- subset(df, race == "Caucasian" & is_recid==0)
white_I_Cg <- subset(df, race == "Caucasian" & is_recid==0 & decile_score>4)

fp_white <- nrow(white_I_Cg)/nrow(white_I)
```

```

print(paste0("Compas false positive rate for Whites: ", fp_white))

## [1] "Compas false positive rate for Whites: 0.216436126932465"

hispanic_I <- subset(df, race == "Hispanic" & is_recid == 0)
hispanic_I_Cg <- subset(df, race == "Hispanic" & is_recid == 0 & decile_score > 4)

fp_hispanic <- nrow(hispanic_I_Cg)/nrow(hispanic_I)

print(paste0("Compas false positive rate for Hispanics: ", fp_hispanic))

## [1] "Compas false positive rate for Hispanics: 0.182692307692308"

black_I <- subset(df, race == "African-American" & is_recid == 0)
black_I_Cg <- subset(df, race == "African-American" & is_recid == 0 & decile_score > 6)

fp_black <- nrow(black_I_Cg)/nrow(black_I)

print(paste0("Compas false positive rate for blacks (>6): ", fp_black))

## [1] "Compas false positive rate for blacks (>6): 0.221825962910128"

white_I <- subset(df, race == "Caucasian" & is_recid == 0)
white_I_Cg <- subset(df, race == "Caucasian" & is_recid == 0 & decile_score > 6)

fp_white <- nrow(white_I_Cg)/nrow(white_I)

print(paste0("Compas false positive rate for Whites (>6): ", fp_white))

## [1] "Compas false positive rate for Whites (>6): 0.0781122864117168"

hispanic_I <- subset(df, race == "Hispanic" & is_recid == 0)
hispanic_I_Cg <- subset(df, race == "Hispanic" & is_recid == 0 & decile_score > 6)

fp_hispanic <- nrow(hispanic_I_Cg)/nrow(hispanic_I)

print(paste0("Compas false positive rate for Hispanics (>6): ", fp_hispanic))

## [1] "Compas false positive rate for Hispanics (>6): 0.0993589743589744"

```

Age

One might think that the disparity is attributable to age differences between races. Blacks tend to be younger than whites and Hispanics. Younger people tend to commit more crime. The score distributions for blacks and whites become more similar if we compare individual of the same age, say younger than 25. The disparities in false positive classification rate shrink significantly (though does not disappear entirely) if we compare whites and blacks of the same age group, say younger than 25. Interestingly, the false positive classification rates goes up significantly for both groups, to 60% for young blacks and 48% for young whites.

```

black_age <- ggplot(data=filter(df, race == "African-American"), aes(x=age_cat)) +
  # geom_histogram(aes(y=..count../sum(..count..)), colour="black", fill="grey45") +
  geom_bar(aes(y=..count../sum(..count..)), colour="black", fill="grey45") +
  xlab("age") +
  ylab("frequency") +
  ylim(0, 1) +
  theme(plot.title = element_text(hjust = 0.5)) +

```

```

    theme(text=element_text(size = 20)) +
    theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
    scale_x_discrete(limits = c("Less than 25", "25 - 45", "Greater than 45"))

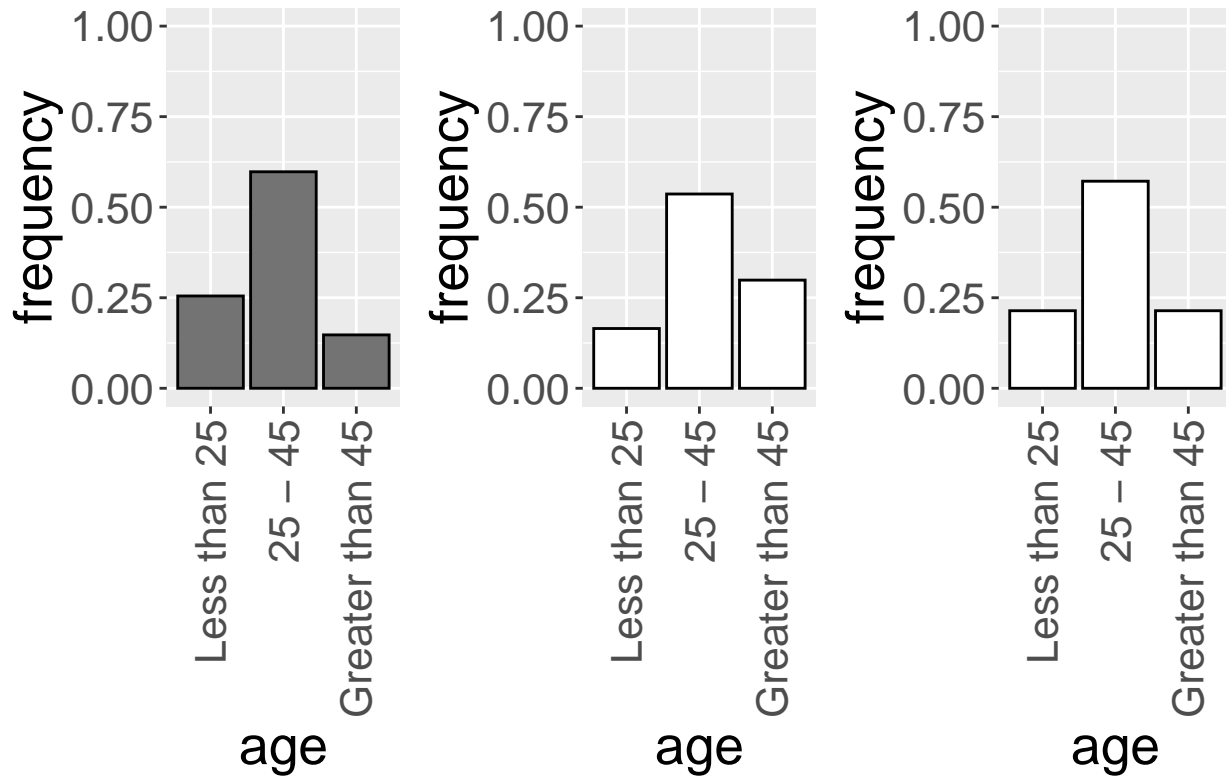
white_age <- ggplot(data=filter(df, race == "Caucasian"), aes(x=age_cat)) +
  # geom_histogram(aes(y=..count../sum(..count..)), colour="black", fill="grey100") +
  geom_bar(aes(y=..count../sum(..count..)), colour="black", fill="grey100") +
  xlab("age") +
  ylab("frequency") +
  ylim(0, 1) +
  theme(plot.title = element_text(hjust = 0.5)) +
  theme(text=element_text(size = 20)) +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
  scale_x_discrete(limits = c("Less than 25", "25 - 45", "Greater than 45"))

hispanic_age <- ggplot(data=filter(df, race == "Hispanic"), aes(x=age_cat)) +
  # geom_histogram(aes(y=..count../sum(..count..)), colour="black", fill="grey80") +
  geom_bar(aes(y=..count../sum(..count..)), colour="black", fill="grey100") +
  xlab("age") +
  ylab("frequency") +
  ylim(0, 1) +
  theme(plot.title = element_text(hjust = 0.5)) +
  theme(text=element_text(size = 20)) +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
  scale_x_discrete(limits = c("Less than 25", "25 - 45", "Greater than 45"))

grid.arrange(black_age, white_age, hispanic_age, ncol = 3, top=textGrob("Age: Black, White, Hispanic",

```

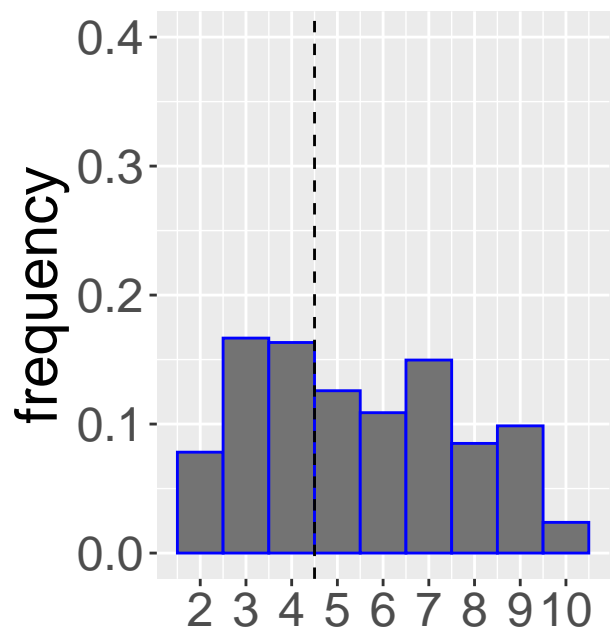
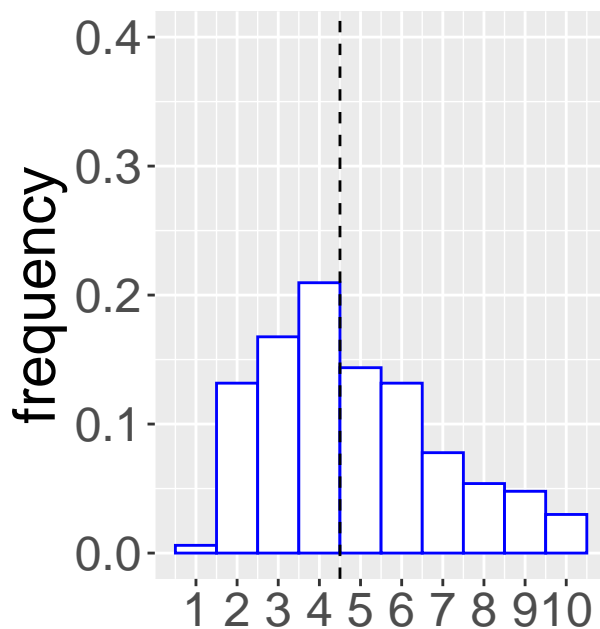

Age: Black, White, Hispanic



on-reoffenders: Scores by Race and Age

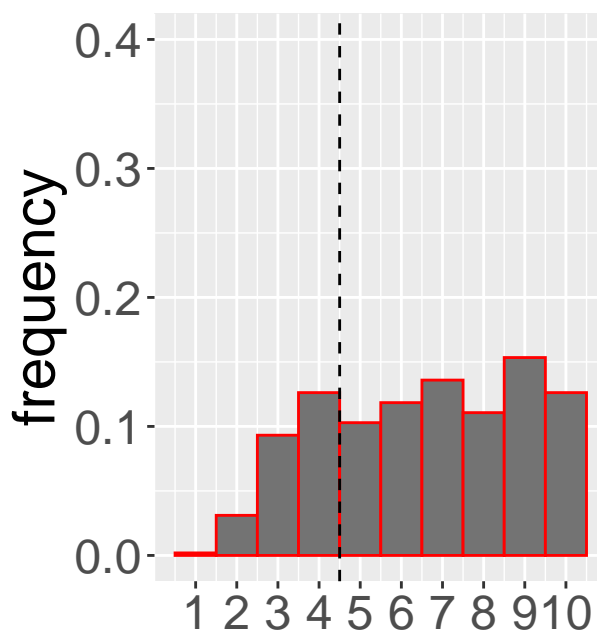
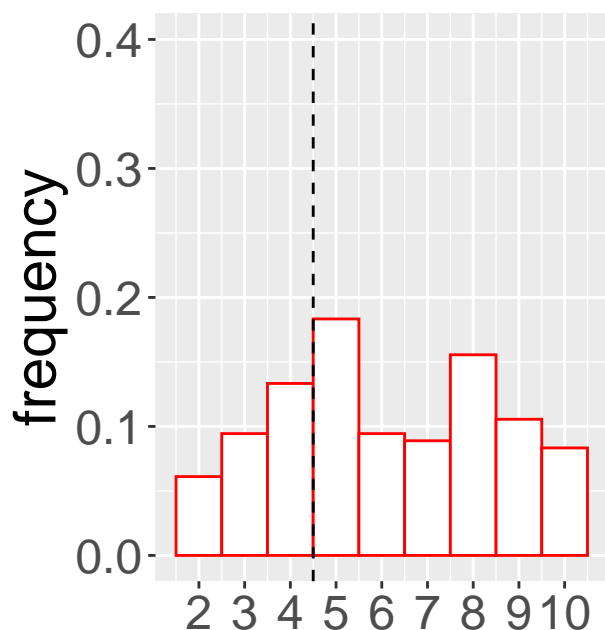
White and < 25

Black and < 25



Reoffenders: Scores by Race and Age

White and < 25 Black and < 25



```
black_young_I <- subset(df, race == "African-American" & is_recid == 0 & age_cat == "Less than 25")
black_young_I_Cg <- subset(df, race == "African-American" & is_recid == 0 & decile_score > 4 & age_cat == "Less than 25")

nrow(black_young_I)

## [1] 294

nrow(black_young_I_Cg)

## [1] 174

fp_young_black <- nrow(black_young_I_Cg)/nrow(black_young_I)

print(paste0("COMPAS false positive rate for blacks (< 25): ", fp_young_black))

## [1] "COMPAS false positive rate for blacks (< 25): 0.591836734693878"

white_young_I <- subset(df, race == "Caucasian" & is_recid == 0 & age_cat == "Less than 25")
white_young_I_Cg <- subset(df, race == "Caucasian" & is_recid == 0 & decile_score > 4 & age_cat == "Less than 25")

nrow(white_young_I)

## [1] 167

nrow(white_young_I_Cg)

## [1] 81

fp_young_white <- nrow(white_young_I_Cg)/nrow(white_young_I)

print(paste0("COMPAS false positive rate for whites (< 25): ", fp_young_white))

## [1] "COMPAS false positive rate for whites (< 25): 0.48502994011976"
```