Multivariate Exploratory Data Analysis

Exoneree Project

Alejandro Hernandez

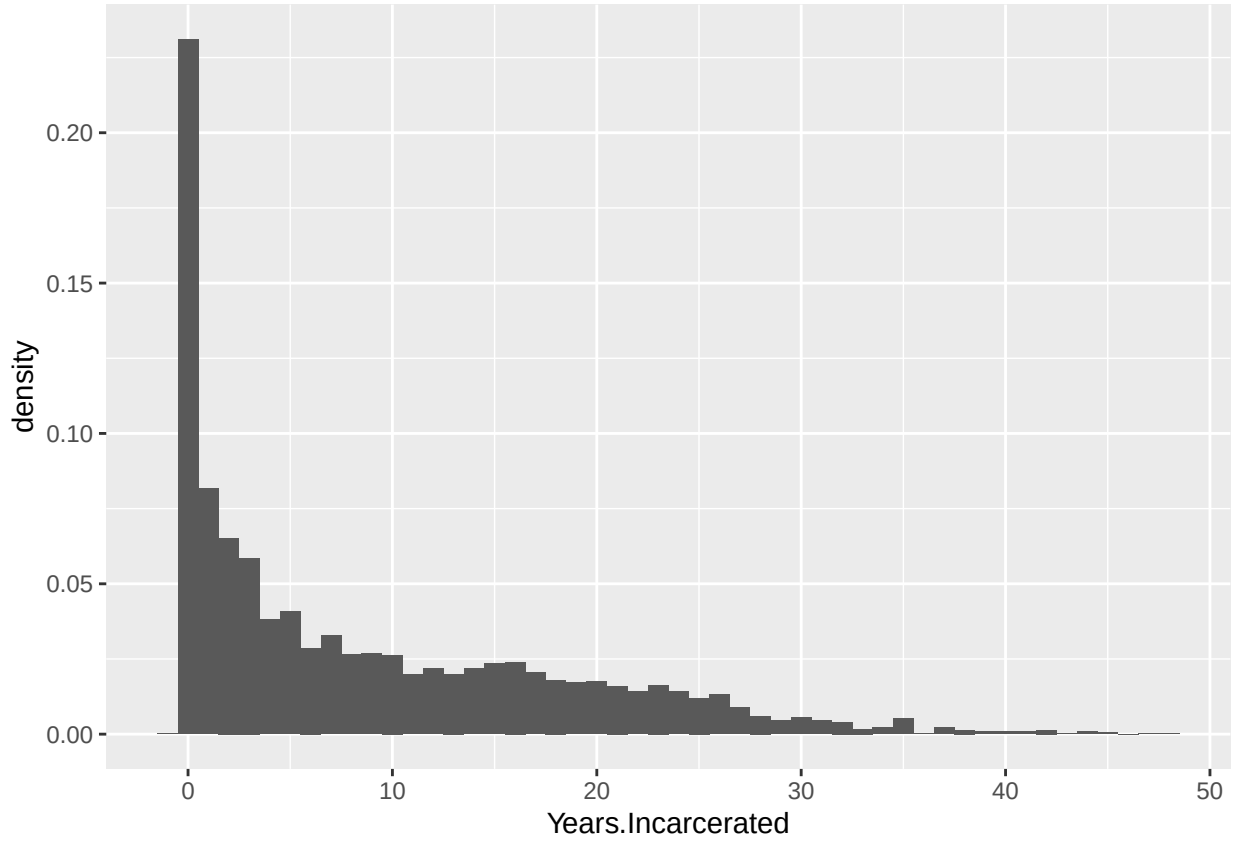
2024-07-16

```
rm(list = ls()) # clear workspace

# load data
exonerees <- read.csv("data/exonerees-may2024.csv")
```

Profiles of Age

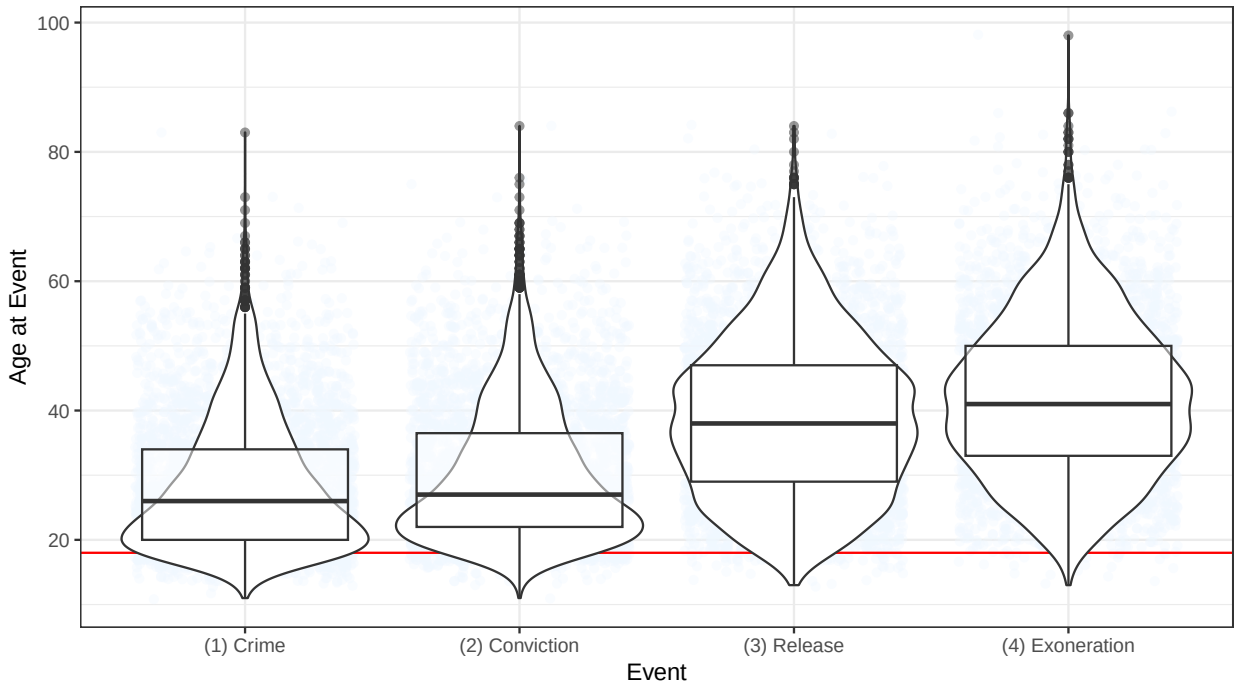
```
exonerees %>%
  ggplot(aes(x = Years.Incarcerated, y = ..density..)) +
  geom_histogram(bins = 50)
```

```
## Warning: The dot-dot notation (‘..density..‘) was deprecated in ggplot2 3.4.0.
## i Please use ‘after_stat(density)‘ instead.
## This warning is displayed once every 8 hours.
## Call ‘lifecycle::last_lifecycle_warnings()‘ to see where this warning was
## generated.
```

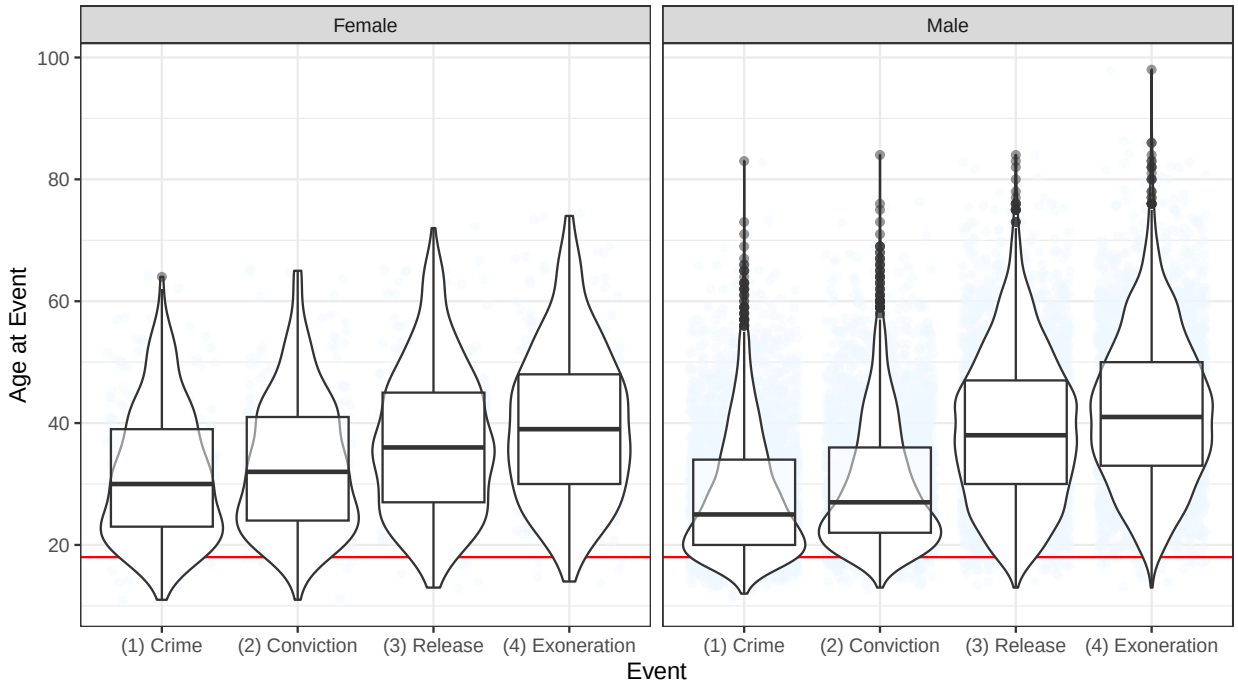



```
gg_age <- exonerees %>%
  filter(!is.na(Age.at.Crime)) %>%
  ggplot() +
  # jitter points (these must be overlayed by the horizontal line)
  geom_jitter(aes(y = Age.at.Crime, x = "(1) Crime"),
              width = 0.4, color = "aliceblue", alpha = 0.4) +
  geom_jitter(aes(y = Age.at.Conviction, x = "(2) Conviction"),
              width = 0.4, color = "aliceblue", alpha = 0.4) +
  geom_jitter(aes(y = Age.at.Release, x = "(3) Release"),
              width = 0.4, color = "aliceblue", alpha = 0.4) +
  geom_jitter(aes(y = Age.at.Exoneration, x = "(4) Exoneration"),
              width = 0.4, color = "aliceblue", alpha = 0.4) +
  # marker at 18 year
  geom_hline(yintercept = 18, color = "red") +
  # age at crime
  geom_violin(aes(y = Age.at.Crime, x = "(1) Crime")) +
  geom_boxplot(aes(y = Age.at.Crime, x = "(1) Crime"), alpha = 0.5) +
  # age at conviction
  geom_violin(aes(y = Age.at.Conviction, x = "(2) Conviction")) +
  geom_boxplot(aes(y = Age.at.Conviction, x = "(2) Conviction"), alpha = 0.5) +
  # age at release
  geom_violin(aes(y = Age.at.Release, x = "(3) Release")) +
  geom_boxplot(aes(y = Age.at.Release, x = "(3) Release"), alpha = 0.5) +
  # age at exoneration
  geom_violin(aes(y = Age.at.Exoneration, x = "(4) Exoneration")) +
  geom_boxplot(aes(y = Age.at.Exoneration, x = "(4) Exoneration"), alpha = 0.5) +
```

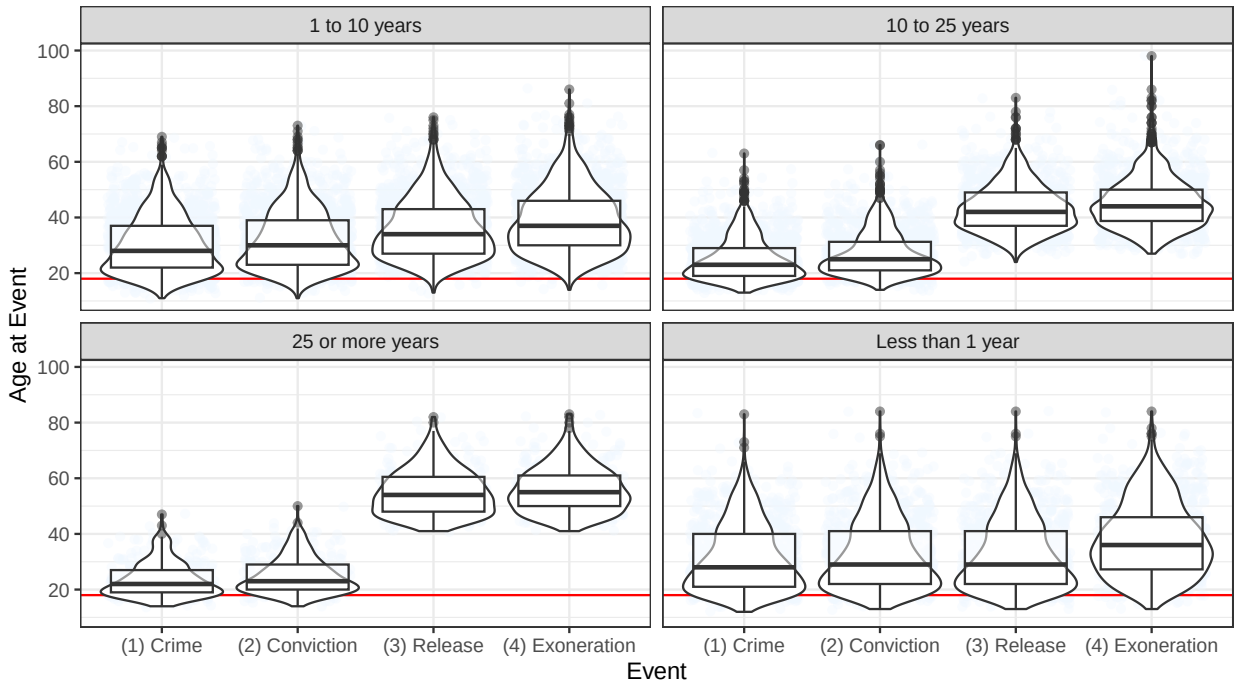
```
  # plot specifics
  ylab("Age at Event") + xlab("Event") +
  scale_y_continuous(breaks = seq(0, 100, 20)) +
  theme_bw()

gg_age
```




```
gg_age + facet_wrap(vars(Sex))
```

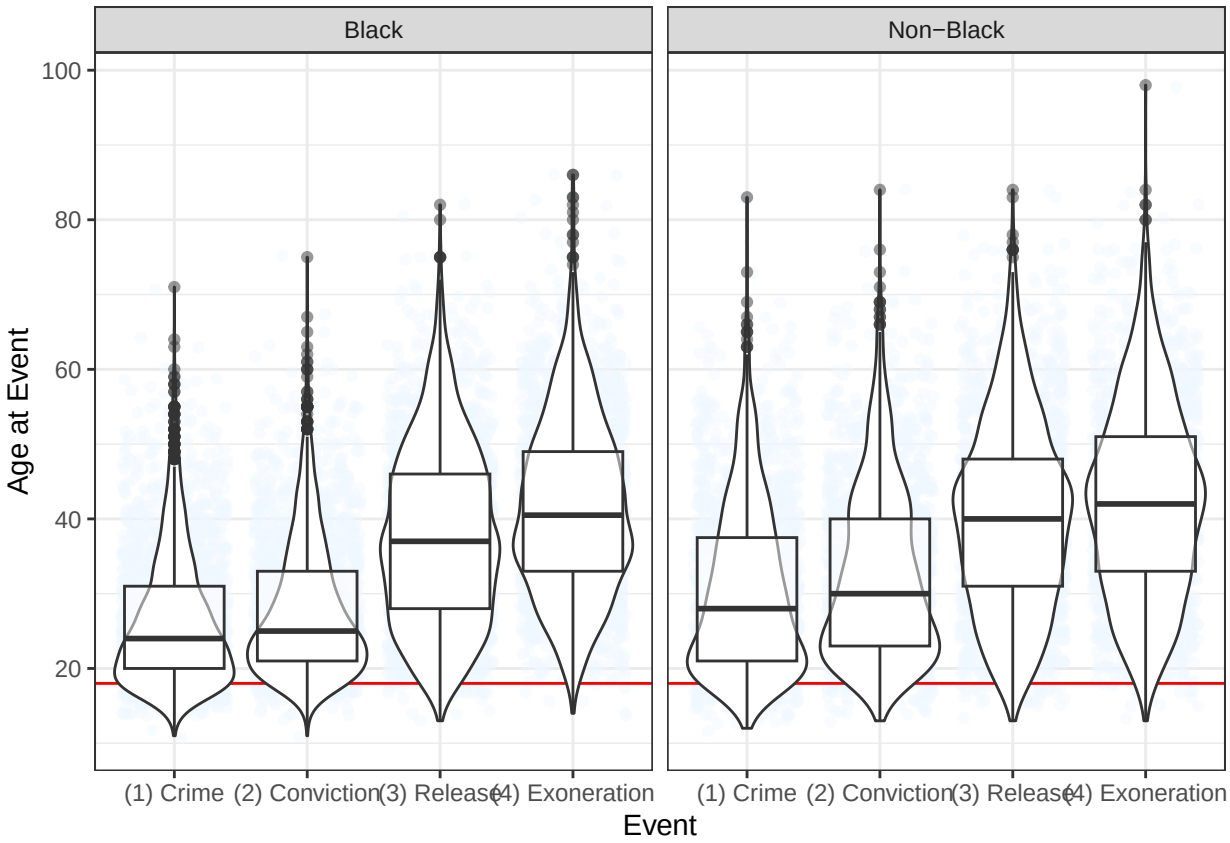



```
gg_age + facet_wrap(vars(Time.Incarcerated))
```

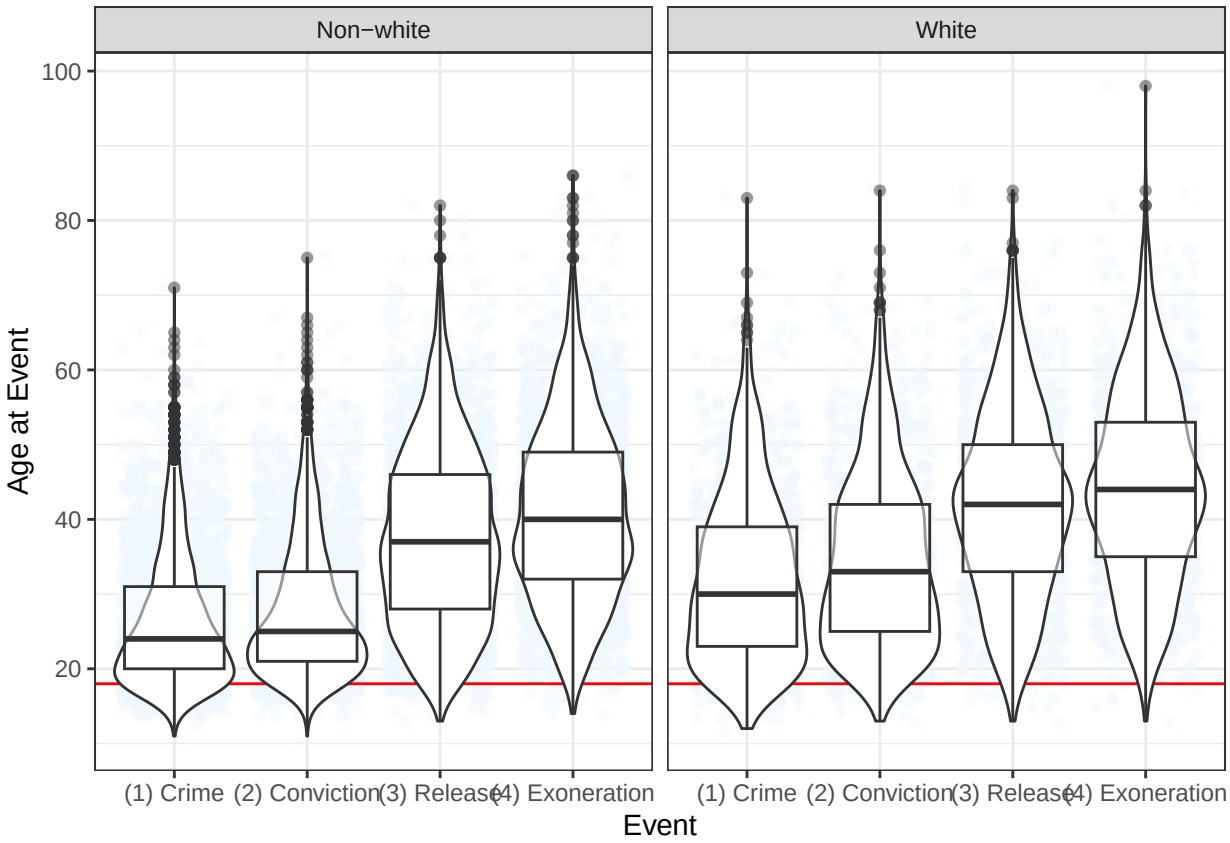



```
gg_age <- exonerees %>%
  filter(!is.na(Age.at.Crime) & !is.na(Race)) %>%
  ggplot() +
  # jitter points (these must be overlayed by the horizontal line)
  geom_jitter(aes(y = Age.at.Crime, x = "(1) Crime"),
```

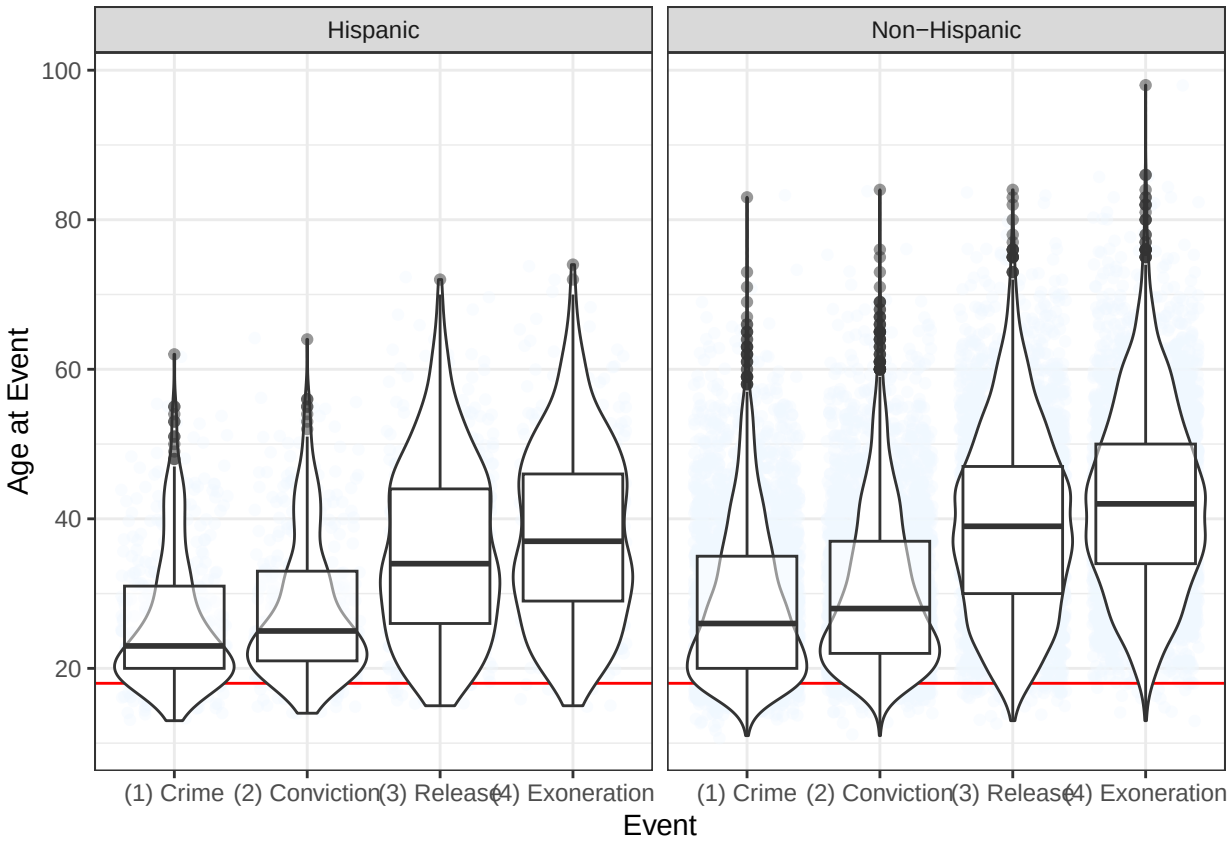
```
              width = 0.4, color = "aliceblue", alpha = 0.4) +
  geom_jitter(aes(y = Age.at.Conviction, x = "(2) Conviction"),
              width = 0.4, color = "aliceblue", alpha = 0.4) +
  geom_jitter(aes(y = Age.at.Release, x = "(3) Release"),
              width = 0.4, color = "aliceblue", alpha = 0.4) +
  geom_jitter(aes(y = Age.at.Exoneration, x = "(4) Exoneration"),
              width = 0.4, color = "aliceblue", alpha = 0.4) +
  # marker at 18 year
  geom_hline(yintercept = 18, color = "red") +
  # age at crime
  geom_violin(aes(y = Age.at.Crime, x = "(1) Crime")) +
  geom_boxplot(aes(y = Age.at.Crime, x = "(1) Crime"), alpha = 0.5) +
  # age at conviction
  geom_violin(aes(y = Age.at.Conviction, x = "(2) Conviction")) +
  geom_boxplot(aes(y = Age.at.Conviction, x = "(2) Conviction"), alpha = 0.5) +
  # age at release
  geom_violin(aes(y = Age.at.Release, x = "(3) Release")) +
  geom_boxplot(aes(y = Age.at.Release, x = "(3) Release"), alpha = 0.5) +
  # age at exoneration
  geom_violin(aes(y = Age.at.Exoneration, x = "(4) Exoneration")) +
  geom_boxplot(aes(y = Age.at.Exoneration, x = "(4) Exoneration"), alpha = 0.5) +
  # plot specifics
  ylab("Age at Event") + xlab("Event") +
  scale_y_continuous(breaks = seq(0, 100, 20)) +
  theme_bw()

gg_age + facet_wrap(vars(Race = ifelse(Race == "Black", "Black", "Non-Black")))
```




```
gg_age + facet_wrap(vars(Race = ifelse(Race == "White", "White", "Non-white")))
```



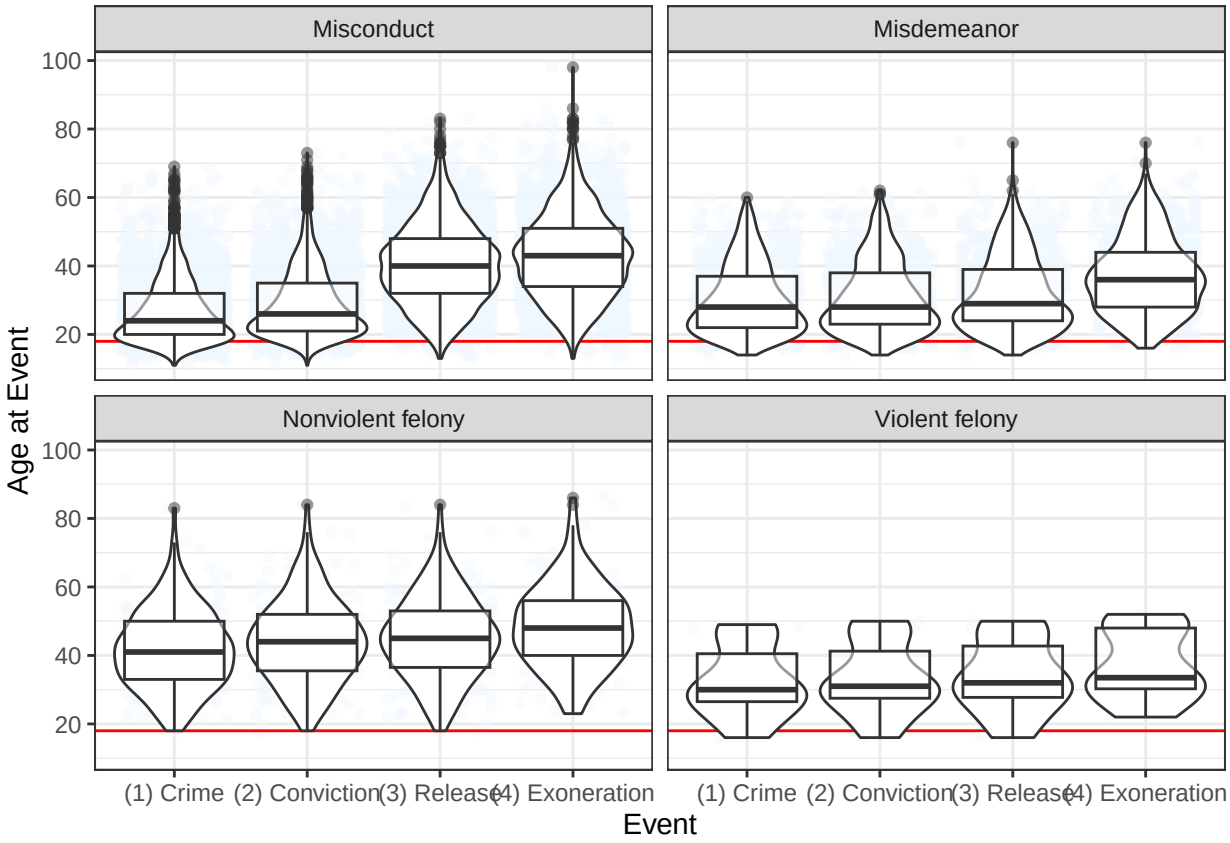

```
gg_age + facet_wrap(vars(Race = ifelse(Race == "Hispanic", "Hispanic", "Non-Hispanic")))
```




```
# gg_age + facet_wrap(vars(Race = ifelse(Race == "Asian", "Asian", "Non-Asian")))
# gg_age + facet_wrap(vars(Race = ifelse(Race == "Native American", "Native American", "Non-Native Amer
# gg_age + facet_wrap(vars(Race = ifelse(Race == "Other", "Other", "Non-Other")))
```

```
gg_age <- exonerees %>%
  filter(!is.na(Age.at.Crime) # &
           # Race %in% c("White", "Black", "Hispanic") &
           # Crime.Category %in% c("Violent felony", "Misdemeanor")
         ) %>%
  ggplot() +
  # jitter points (these must be overlayed by the horizontal line)
  geom_jitter(aes(y = Age.at.Crime, x = "(1) Crime"),
              width = 0.4, color = "aliceblue", alpha = 0.4) +
  geom_jitter(aes(y = Age.at.Conviction, x = "(2) Conviction"),
              width = 0.4, color = "aliceblue", alpha = 0.4) +
  geom_jitter(aes(y = Age.at.Release, x = "(3) Release"),
              width = 0.4, color = "aliceblue", alpha = 0.4) +
  geom_jitter(aes(y = Age.at.Exoneration, x = "(4) Exoneration"),
              width = 0.4, color = "aliceblue", alpha = 0.4) +
  # marker at 18 year
  geom_hline(yintercept = 18, color = "red") +
  # age at crime
  geom_violin(aes(y = Age.at.Crime, x = "(1) Crime")) +
  geom_boxplot(aes(y = Age.at.Crime, x = "(1) Crime"), alpha = 0.5) +
  # age at conviction
```

```
  geom_violin(aes(y = Age.at.Conviction, x = "(2) Conviction")) +
  geom_boxplot(aes(y = Age.at.Conviction, x = "(2) Conviction"), alpha = 0.5) +
  # age at release
  geom_violin(aes(y = Age.at.Release, x = "(3) Release")) +
  geom_boxplot(aes(y = Age.at.Release, x = "(3) Release"), alpha = 0.5) +
  # age at exoneration
  geom_violin(aes(y = Age.at.Exoneration, x = "(4) Exoneration")) +
  geom_boxplot(aes(y = Age.at.Exoneration, x = "(4) Exoneration"), alpha = 0.5) +
  # plot specifics
  ylab("Age at Event") + xlab("Event") +
  scale_y_continuous(breaks = seq(0, 100, 20)) +
  theme_bw()

gg_age + facet_wrap(vars(Crime.Category))
```




```
gg_age <- exonerees %>%
  filter(!is.na(Age.at.Crime) & Race %in% c("Black", "White")) %>%
  ggplot() +
  # jitter points (these must be overlayed by the horizontal line)
  geom_jitter(aes(y = Age.at.Crime, x = "(1) Crime"),
              width = 0.4, color = "aliceblue", alpha = 0.4) +
  geom_jitter(aes(y = Age.at.Conviction, x = "(2) Conviction"),
              width = 0.4, color = "aliceblue", alpha = 0.4) +
  geom_jitter(aes(y = Age.at.Release, x = "(3) Release"),
              width = 0.4, color = "aliceblue", alpha = 0.4) +
```

```
  geom_jitter(aes(y = Age.at.Exoneration, x = "(4) Exoneration"),
              width = 0.4, color = "aliceblue", alpha = 0.4) +
  # marker at 18 year
  geom_hline(yintercept = 18, color = "red") +
  # age at crime
  geom_violin(aes(y = Age.at.Crime, x = "(1) Crime")) +
  geom_boxplot(aes(y = Age.at.Crime, x = "(1) Crime"), alpha = 0.5) +
  # age at conviction
  geom_violin(aes(y = Age.at.Conviction, x = "(2) Conviction")) +
  geom_boxplot(aes(y = Age.at.Conviction, x = "(2) Conviction"), alpha = 0.5) +
  # age at release
  geom_violin(aes(y = Age.at.Release, x = "(3) Release")) +
  geom_boxplot(aes(y = Age.at.Release, x = "(3) Release"), alpha = 0.5) +
  # age at exoneration
  geom_violin(aes(y = Age.at.Exoneration, x = "(4) Exoneration")) +
  geom_boxplot(aes(y = Age.at.Exoneration, x = "(4) Exoneration"), alpha = 0.5) +
  # plot specifics
  ylab("Age at Event") + xlab("Event") +
  scale_y_continuous(breaks = seq(0, 100, 20)) +
  theme_bw()

gg_age + facet_wrap(vars(Race))
```