Data Visualization

Johnathan Hsu

12/4/2017

You will be able to find my graphics in the corresponding folder along with the description paragraphs. Code below are 2 .gifs that will be used at my presentation, you can also find the code for the caseloads by different types of cases.

The reason why I am limiting the race of my sample to only black and white defendants is because they account for 1.1 out of 1.2 million observations. Additionally, I chose to use age as my dependent variable because it is the best way to represent the distribution of the punitiveness of crime over a life course; the age-crime curve is also supported by the literature in criminology. The total prison time that the defendant is sentenced to is a good representation of the outcome because it incorporates different forms of judicial discretion that results in the outcome of the case.

The reason why I chose to incorporate the different types of crimes in the time series plot is because it shows us the degree which drug cases is driving prosecution rates. It is also interesting that until recently when it was surpassed by immigration cases, it has been quite steady. This will be used to justify my study and emphasis on drug related crimes.

```r
library(tidyverse)
library(ggplot2)
library(readr)
library(gganimate)
library(ggthemes)
library(viridis)

data <- read.csv("~/monitoring-federal-criminal-sentences/clean_data/merged_data/96-15.csv")

data <- data %>%
  filter(TOTPRISN < 1000) %>%
  filter(!is.na(TYPE)) %>%
  filter(!is.na(EDUCATN))

# Creating labels
race_labels <- c('1' = "White Defendants",
                 '2' = "Black Defendants")

case_labels <- c("0" = "Other Cases",
                 "1" = "Drug Crimes",
                 "2" = "Firearms",
                 "3" = "Immigration",
                 "4" = "Violent Crimes")

education_labels <- c("0" = "No Schooling",
                      "1" = "Below High School",
                      "2" = "High School",
                      "3" = "Special Degree",
                      "4" = "College Graduate",
                      "5" = "Advanced/Professional Degree")

mytheme <- theme(text = element_text(family = "serif", colour = "red1"),
                 plot.background = element_rect(fill = "oldlace"),
```

```r
                  legend.background = element_rect(colour = "black"),
                  legend.position = "bottom",
                  panel.border = element_rect(colour = "black", fill = NA))
education_gg <- ggplot(data, aes(x = AGE,
                                 y = TOTPRISN,
                                 fram = as.factor(YEAR),
                                 color = XCRHISSR)) +
                    labs(title = "Federal Criminal Sentences by Education",
                    y = "Total Prison Time Sentenced",
                    x = "Age", color = "Race") +
                    guides(guide_legend(title = "Race")) +
                    scale_color_manual(labels = c("White", "Black"),
                                       values = c("blue", "red")) +
                    coord_cartesian(xlim = c(19,53),
                                    ylim = c(0, 150)) +
                    facet_grid(.~ as.factor(EDUCATN),
                               labeller = as_labeller(education_labels)) +
                    geom_smooth(method = 'loess')

education_gg <- education_gg + mytheme

education_anime <- gganimate(education_gg, interval = 0.4, fps = 4, saver = "gif")
gganimate_save(offense_anime, "education.gif", interval = 0.4, fps = 4, saver = "gif")

offense_gg <- ggplot(data, aes(x = AGE, y = TOTPRISN, frame = as.factor(YEAR),
                               color = as.factor(MONRACE))) +
                    labs(title = "Federal Criminal Sentences by Offense Type",
                        y = "Total Prison Time Sentenced",
                        x = "Age",
                        color = "Race") +
                    guides(guide_legend(title = "Race")) +
                    scale_color_manual(labels = c("White", "Black"),
                                       values = c("blue", "red")) +
                    coord_cartesian(xlim = c(19,53), ylim = c(0, 150)) +
                    facet_grid(.~ TYPE, labeller = as_labeller(case_labels)) +
                    geom_smooth(method = 'loess')

offense_gg <- offense_gg + mytheme
offense_anime <- gganimate(offense_gg, interval = 0.4, fps = 4, saver = "gif")
gganimate_save(offense_anime, "offense.gif", interval = 0.4, fps = 4, saver = "gif")

density_gg <- ggplot(data, aes(x = AGE , y = TOTPRISN, frame = as.factor(YEAR),
                               color = as.factor(MONRACE))) +
                    labs(color='Race') +
                    labs(title = "Federal Criminal Sentences Drug Comparison",
                         y = "Total Prison Time Sentenced",
                         x = "Age") +
                        facet_grid(.~as.factor(DRUG), labeller = as_labeller(drug_labels)) +
                        scale_color_manual(labels = c("White", "Black"), values = c("blue", "red")) +
                        coord_cartesian(xlim = c(20,50), ylim = c(0,150)) +
                        geom_smooth(method = "loess")

density_gg <- density_gg + mytheme
```

```r
density_anime <- gganimate(density_gg, interval = 0.4, fps = 10, saver = "gif")
gganimate_save(density_anime, "drug.gif", interval = 0.4, fps = 4, saver = "gif")

# below is code to construct the plot of federal caseload over 2 decades by the type of the case.

drug <- data %>%
  filter(TYPE == 1)

firearms <- data %>%
  filter(TYPE == 2)

immigration <- data %>%
  filter(TYPE == 3)

violent <- data %>%
  filter(TYPE == 4)

drug <- as_tibble(table(drug$YEAR))
drug$Var1 <- as.numeric(drug$Var1)

firearms <- as_tibble(table(firearms$YEAR))
firearms$Var1 <- as.numeric(firearms$Var1)

immigration <- as_tibble(table(immigration$YEAR))
immigration$Var1 <- as.numeric(immigration$Var1)

violent <- as_tibble(table(violent$YEAR))
violent$Var1 <- as.numeric(violent$Var1)

federal_incarceration <- ggplot() +
  geom_line(data = caseload, aes(Var1, n, color = "All Cases"), size = 1.5) +
  geom_line(data = drug, aes(Var1, n, color = "Drug Crimes"), size = 1.5) +
  geom_line(data = immigration, aes(Var1, n, color = "Immmigration"), size = 1.5) geom_line(data = firea
  geom_line(data = violent, aes(Var1, n, color = "Violent Crimes"), size = 1.5) +
  labs(title = "Federal Criminal Caseload Time Series",
      y = "Frequency",
      x = "Year", color = NULL) +
  coord_cartesian(xlim = c(1996, 2015)) +
  mytheme +
  scale_color_brewer(palette = "Set1") +
  theme(plot.background = element_rect(size=1,linetype="solid",color="black"))

federal_incarceration
```