Homework 3

Group Leopard
George Linder
Eduardo Schiappa Pietra
Vidvat Ramachandran
Dror Shvadron

9/26/2019

Contents

Introduction

America does a poor job tracking and accounting for its unsolved homicides. Every year, at least 5,000 killers get away with murder. The rate at which police clear homicides through arrest has declined over the years until, today, about a third go unsolved. The Murder Accountability Project is a nonprofit group organized in 2015 and dedicated to educate Americans on the importance of accurately accounting for unsolved homicides within the United States.

Packages

```
library(tidyverse)
library(readr)
library(knitr)
require(patchwork)
library(ggmap)
library(maps)
library(mapdata)
library(stringr)
```

Data

```
murder <- read_csv("https://www.dropbox.com/s/wzp6o78lcosp3ux/map.csv?dl=1")
names(murder) <- tolower(names(murder))
```

Tasks

Task 1

The following questions/tasks will require you to manipulate and summarise the data. Be sure to consult the data dictionary so you understand the variables and their set of possible values.

Each answer should include a one to two sentence write up. Your code output should only contain the necessary rows and variables from the data frame to answer the question or complete the task.

2017 NC Counties

How many distinct counties in North Carolina had a recorded homicide in 2017?

```
NC_counties <- murder %>% select(cntyfips, state, year) %>%
  filter(state == 'North Carolina', year == 2017) %>%
  group_by(cntyfips)%>%
  unique() %>%
  select(cntyfips) %>%
  nrow()
```

We have 73 rows of data in our output for 2017 homicides by county in North Carolina.

NC Year-Month Homicide Combinations

Which year and month combinations had the three most homicides in North Carolina from 2013 to 2017?

```
top_months <- murder %>%
  select(state, year, month) %>%
  filter(state == 'North Carolina', year %in% 2013:2017) %>%
  group_by(year, month) %>%
  summarise(count = n()) %>%
  arrange(desc(count)) %>%
  head(3)

kable(top_months)
```

year	month	count
2017	August	70
2016	March	64
2017	April	63

August 2017 had the most homicides (70). March, 2016 had 64 homicides and April, 2017 had 63 homicides.

Top States with Solved Cross-Racial Murders

What were the top three states that had the most homicides by "Murder and non-negligent manslaughter" that were solved and crossed racial lines, i.e., where the offender and victim were of different race? Include the counts in your answer.

```
topracial <- murder %>%
  select(state, solved, homicide, vicrace, offrace) %>%
```

```r
  filter(solved == 'Yes', homicide == 'Murder and non-negligent manslaughter', vicrace != offrace) %>%
  group_by(state) %>%
  summarise(count = n()) %>%
  arrange(desc(count)) %>%
  slice(1:3)

kable(topracial)
```

state	count
California	12916
Texas	7273
New York	6617

The top 3 states that had solved homicides that crossed racial lines were California, Texas, and New York. California had 12,916 instances, Texas had 7,273 and New York had 6,617.

Unsolved Homicides in California

For the state of California, what is the relative frequency of unsolved homicides for each year in the data set?

```r
total_homicides <- murder %>%
  select(state, solved, year) %>%
  filter(state == "California") %>%
  group_by(year) %>%
  summarise(count = n())

unsolved_homicides <- murder %>%
  select(state, solved, year) %>%
  filter(state == "California", solved == 'No') %>%
  group_by(year) %>%
  summarise(count = n())

frequency <- data.frame(unsolved_homicides$year, unsolved_homicides$count / total_homicides$count)

names(frequency) <- c('year', 'unsolved_freq')
frequency <- round(frequency, digits = 3)
#frequency

ggplot(data = frequency, mapping = aes(x = year, y = unsolved_freq)) +
  geom_line() +
  labs(title = "Relative frequency of unsolved homicides in California")+
  xlab("Year")+
  ylab("Frequency")
```
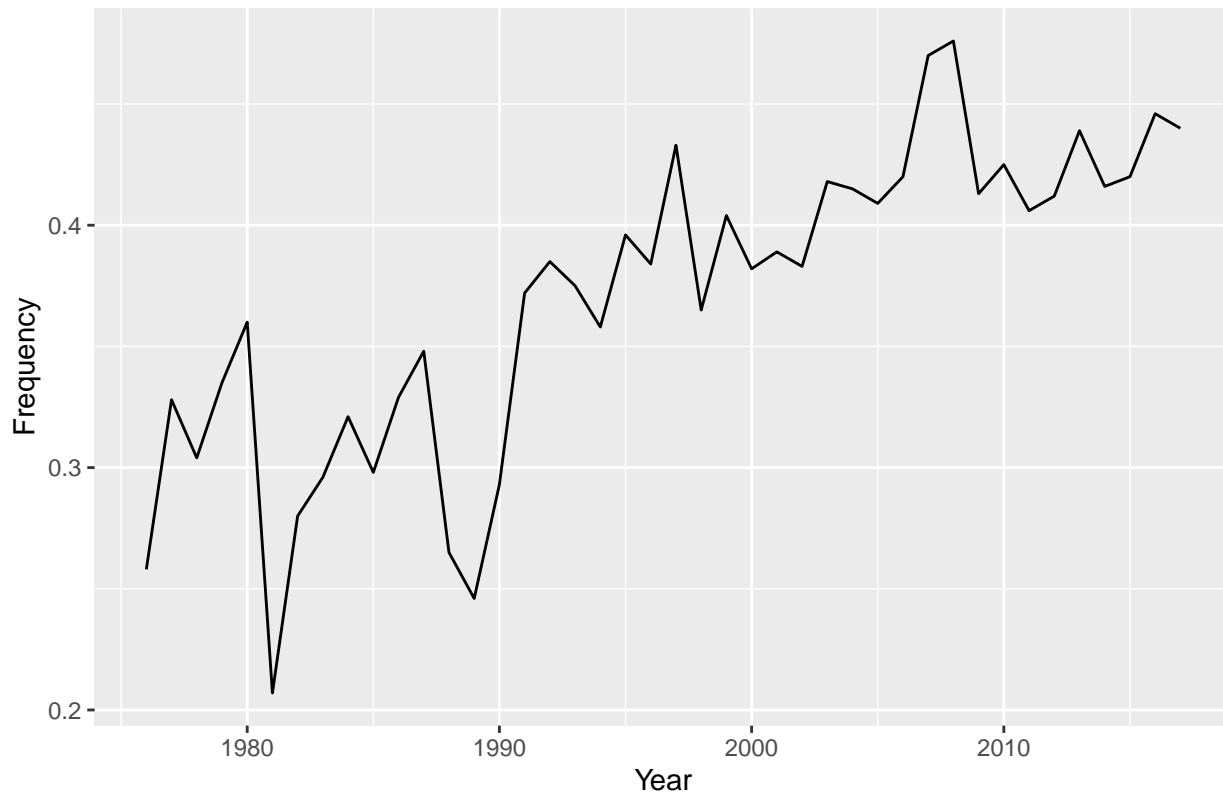
Relative frequency of unsolved homicides in California

Define a variable age_gap as the offender's age minus the victim's age. For each offender sex level (not Unknowns), what is the median age gap, 10th percentile age gap, and 90th percentile age gap? Your output should be in a single data frame.

Age Gap in Victim and Offender's Age

```r
agegap <- murder %>%
  filter(offsex != "Unknown", (vicage>0&vicage<100), (offage>0&offage<100)) %>%
  select(vicage, offage, offsex) %>%
  mutate(age_gap = offage - vicage) %>%
  filter(offsex != 'Unknown') %>%
  group_by(offsex) %>%
  summarise(median = median(age_gap), pct_10 = quantile(age_gap, .10), pct_90 = quantile(age_gap, .90))

kable(agegap)
```

offsex	median	pct_10	pct_90
Female	-2	-23	21
Male	0	-24	16

States with Largest Change in Homicides

Which five states had the largest change in the number of homicides by "Murder and non-negligent manslaughter" in the 10 years before and after the 1994 crime bill? Consider 1985 - 1994 and 1995 - 2004.

```
before_bill <- murder %>%
  select(state, homicide, year) %>%
  filter(year %in% seq(1985, 1994, 1),
         homicide == 'Murder and non-negligent manslaughter') %>%
  group_by(state) %>%
  summarise(count_before_bill = n())

after_bill <- murder %>%
  select(state, homicide, year) %>%
  filter(year %in% seq(1995, 2004, 1),
         homicide == 'Murder and non-negligent manslaughter') %>%
  group_by(state) %>%
  summarise(count_after_bill = n())

combined <- merge(before_bill, after_bill, on = 'state')

combined <- combined %>%
            mutate(change = abs(count_after_bill - count_before_bill)) %>%
            select(state, change) %>%
            arrange(desc(change)) %>%
            slice(1:5)

kable(combined)
```

state	change
New York	10386
California	9992
Texas	8177
Michigan	3144
Florida	2380

We took the absolute value of the difference before and after the bill was passed because there were states where homicides were reduced significantly after the bill and there were also states like Florida where homicides increased after the bill was passed. Another way to look at this problem would be to look at the five states where homicides were reduced the most after the bill was passed.

Task 2

Gun policy in California

For this task we chose to investigate the effect of gun control policy in California. We saw that California had one of the largest changes in homicides in the 10 years before and after the 1994 crime bill (the Federal Assault Weapons Ban and Violent Crime Control and Law Enforcement Act). We decided to take a closer look into California around this time period because is a state that currently has some of the strictest gun laws. The early 90s were a tumultuous time in California following the '101 California Massacre' that helped set the '94 crime bills in action. We used population data to help visualize the relative frequency of homicides involving guns in counties across this state.

First, we needed annual population data at the county level. Our source for the data is the National Cancer Institute, which maintains a clean database of population estimates for the US. (Source:https://seer.cancer.gov/popdata/)

```r
# NOT RUN (save running time)
# PopData <- readr::read_fwf("Data/us.1969_2017.19ages.adjusted.txt.gz", fwf_widths(c(4,2,2,3,2,1,1,1,2
#                            c("year", "state", "stateFIPS", "countyFIPS", "registry", "race", "origin"
#                            cols(stateFIPS = col_integer(),pop = col_integer()))
#
# #for later
# PopData1 <- PopData %>%
#   group_by(year, state, stateFIPS, countyFIPS) %>%
#   summarise(pop=sum(pop))
#
# CalPop <-  PopData %>%
#   filter(state == "CA") %>%
#   group_by(year) %>%
#   summarise(pop=sum(pop))

#save(PopData1, CalPop, file = "Data/processed.Rdata")
load("Data/processed.Rdata")

gun_types <-
  c(
    'Firearm, type not stated',
    'Handgun - pistol, revolver, etc',
    'Other gun',
    'Rifle',
    'Shotgun'
  )

california <- murder %>%
  filter(state == 'California',
         weapon %in% gun_types)

# Split county name and 'county' and change to_lower
california <-  california %>%
  separate(
    col = cntyfips,
    into = c("subregion", "remaining"),
    sep = ","
  ) %>%
  mutate(subregion = tolower(subregion))

# Major weapon policy in 1994
california <- california %>%
  mutate(preBan = year < 1993)
```

Gun murders per year in California 1976-2017
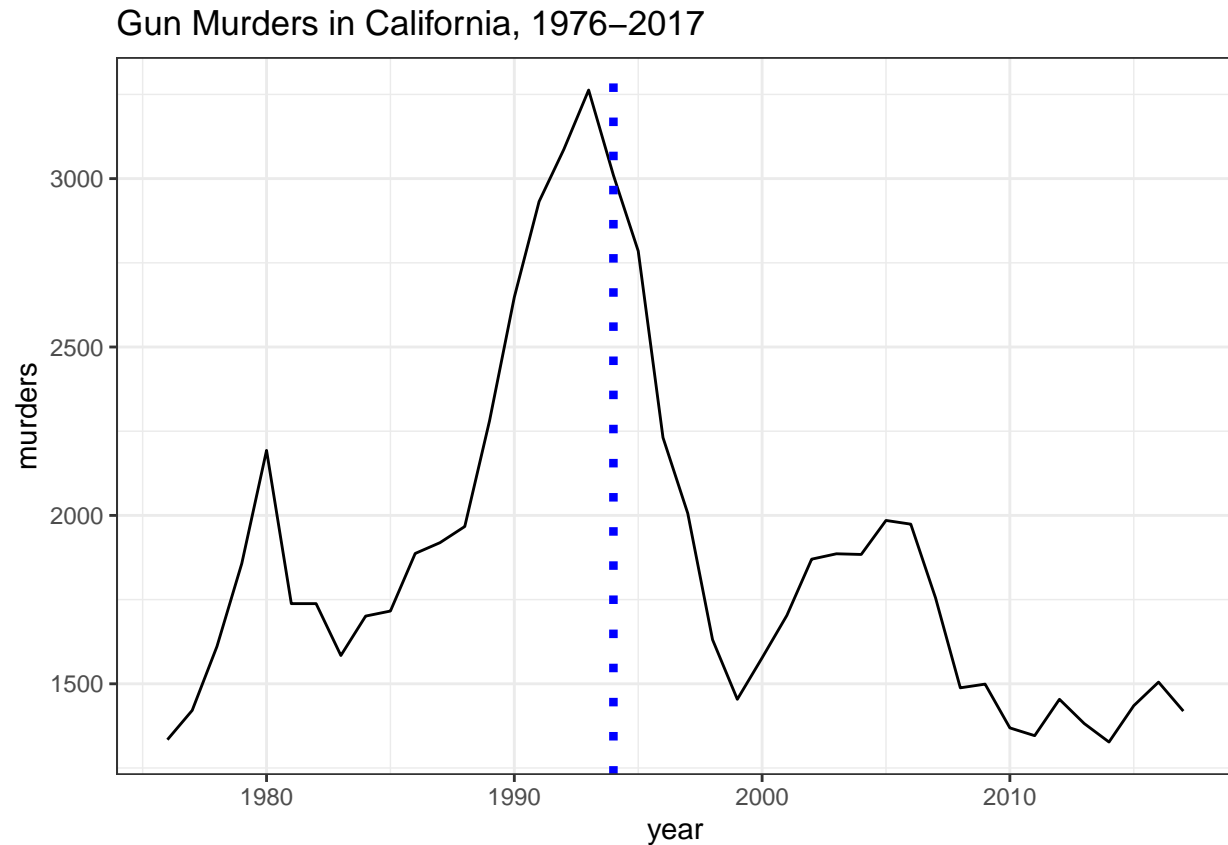
```r
cal_annual <- california %>%
  group_by(year) %>%
  summarize(murders = n())

g1 <- ggplot(cal_annual, aes(x = year, y = murders)) +
    geom_line() +
    theme_bw() +
```

```
    labs(title = "Gun Murders in California, 1976-2017")+
    geom_vline(xintercept = 1994, linetype="dotted",
               color = "blue", size=1.5)
```

g1

Gun Murders in California, 1976–2017

Gun murders as percent of population

```
cal_annual <- cal_annual %>%
  left_join(CalPop) %>%
  mutate(pct = murders/pop*10000)

g2 <- ggplot(cal_annual, aes(x = year, y = pct)) +
  geom_line()+
  theme_bw() +
  labs(title = "Gun Murders in California (from Population), 1976-2017")+
  ylab(label = "Murders, per 10K")+
  geom_vline(xintercept = 1994, linetype="dotted",
             color = "blue", size=1.5)
```

g2

Gun Murders in California (from Population), 1976–2017
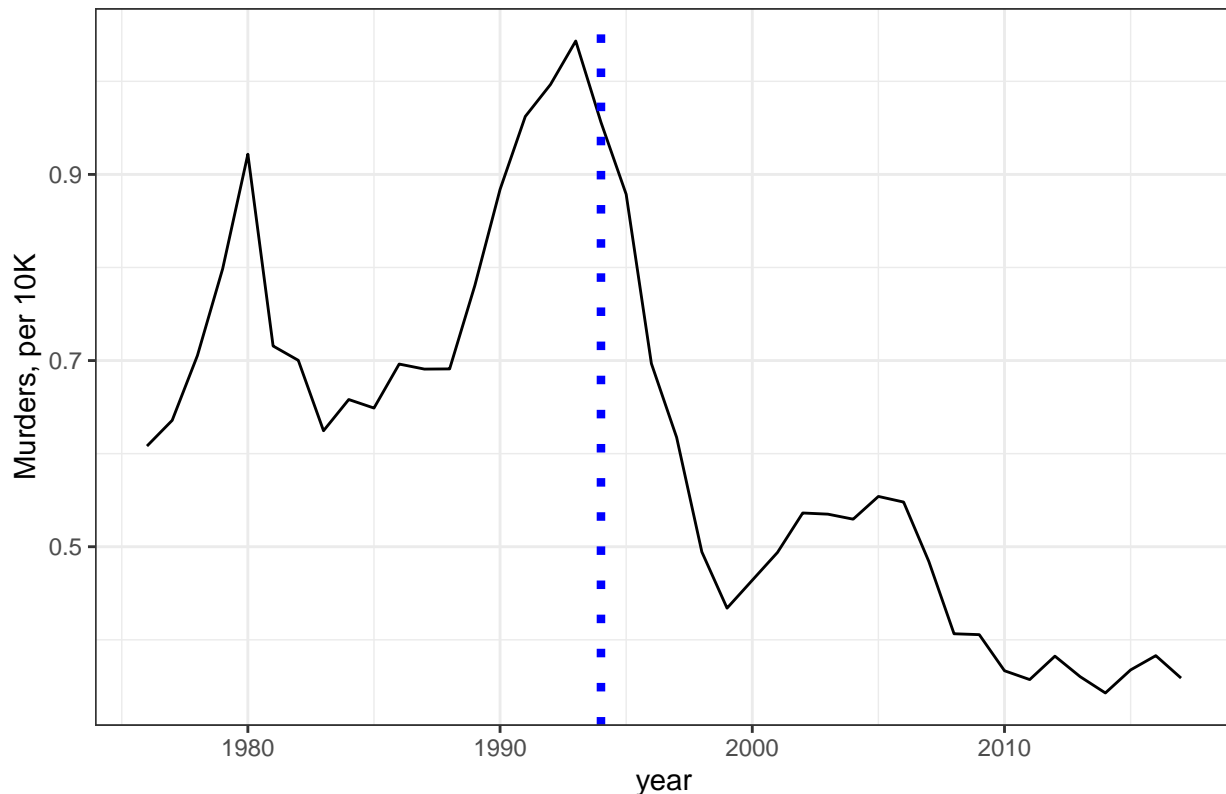
Gun murders by county

```r
california <- murder %>%
               filter(state == 'California')
gun_types <- c('Firearm, type not stated', 'Handgun - pistol, revolver, etc', 'Other gun', 'Rifle', 'Sho
california <- california %>%
               filter(weapon %in% gun_types)
# Split county name and 'county' and change to_lower
california <-  california %>%
  separate(col = cntyfips, into = c("subregion", "remaining"), sep = ",") %>%
  mutate(subregion = tolower(subregion))
```

The map plots here show that most counties had a decrease in gun-related crimes relative to the population, there are just a few counties where the such crimes increased. A couple of counties had too few data or data was not available and those have been greyed out in the plots. The lack of data for those counties also meant that their averages were not reliable for our analysis. We can see from the lightening of the maps that the 1994 crime bill did help reduce gun-related crimes.

```r
#D. Kahle and H. Wickham. ggmap: Spatial Visualization with ggplot2.
#  The R Journal, 5(1), 144-161. URL
#  http://journal.r-project.org/archive/2013-1/kahle-wickham.pdf
states <- map_data("state")
ca_map <- subset(states, region == "california")
counties <- map_data("county")
ca_county <- subset(counties, region == "california")
ditch_the_axes <- theme(
  axis.text = element_blank(),
```

```r
    axis.line = element_blank(),
    axis.ticks = element_blank(),
    panel.border = element_blank(),
    panel.grid = element_blank(),
    axis.title = element_blank()
  )
basemap_cali <- ggplot(data = ca_map, mapping = aes(x = long, y = lat, group = group)) +
  coord_fixed(1.3) +
  geom_polygon(color = "black", fill = "gray") +
  geom_polygon(data = ca_county, fill = NA, color = "white") +
  geom_polygon(color = "black", fill = NA) + theme_bw() + ditch_the_axes


cal_dat <- PopData1 %>%
  filter(year %in% 1976:2017,
         state == "CA")

cal_dat$countyFIPS = ceiling( as.numeric(cal_dat$countyFIPS)/2 )

cal_dat <- cal_dat %>%
  rename(abr = state)

# Extract county nos
california <- california %>%
  mutate(countyFIPS = as.numeric(str_extract(ori, "[:digit:]{3}")) )

cali_pre94 <- california %>%
  filter(year <= 1994) %>%
  group_by(state, year, subregion, countyFIPS) %>%
  summarise(count = n()) %>%
  filter( ! (subregion == "san francisco" & countyFIPS == 1) ) %>%
  merge(cal_dat) %>%
  mutate(death_per100k = 100000*(count/pop) )
cali_post94 <- california %>%
  filter(year > 1994) %>%
  group_by(state, year, subregion, countyFIPS) %>%
  summarise(count = n()) %>%
  filter( ! (subregion == "san francisco" & countyFIPS == 1) & countyFIPS %in% 1:58  ) %>%
  merge(cal_dat) %>%
  mutate(death_per100k = 100000*(count/pop) )

cpre <- cali_pre94 %>%
  group_by(subregion) %>%
  summarise(avg_death_per100k = mean(death_per100k))
cpost <- cali_post94 %>%
  group_by(subregion) %>%
  summarise(avg_death_per100k = mean(death_per100k))

pre94_map <- merge(cpre, ca_county)
post94_map <- merge(cpost, ca_county)

m1 <- basemap_cali + geom_polygon(data = pre94_map, aes(fill = avg_death_per100k) ) +
```

```
    scale_fill_gradient(low = "white", high = "tomato4", limits = c(0.8,20)) +
    labs(fill = "Homicides per 100,000")
m2 <- basemap_cali + geom_polygon(data = post94_map, aes(fill = avg_death_per100k ) ) +
    scale_fill_gradient(low = "white", high = "tomato4", limits = c(0.8,20)) +
    labs(fill = "Homicides per 100,000")
m1/m2 + plot_annotation(title = "Average of Gun-related Homicides per 100,000 people in California \nbet
```

Average of Gun–related Homicides per 100,000 people in California between the years 1976 and 1994(top) and 1995 and 2017(bottom)
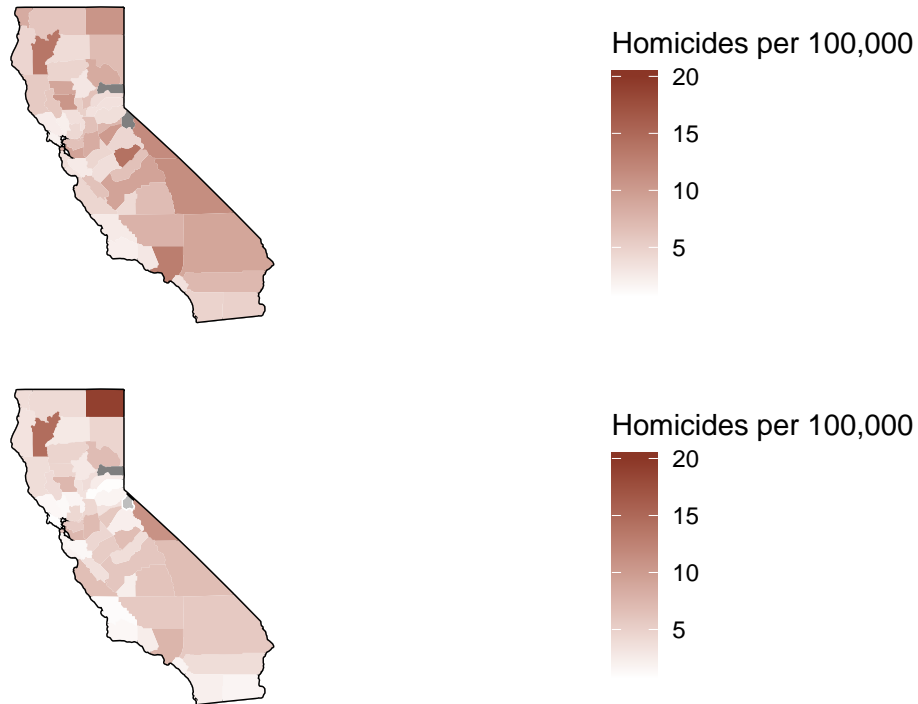
We are interested in exploring how the changes in gun laws undertaken in California are related to the trend in murders with fire arms observed throughout the years and between counties.

The next graph shows the evolution of the distribution of murders between counties in California. There is one boxplot for each year in the data. The colours of the boxplots reference the periods between major changes in California's gun legislation. There black line plot connects the means of the number of murders between counties across time.

We have 6 relevant periods that we want to analize: 1) 1976 - 1989: Period before any of the laws analized became active. We call it the "looselaws" period. 2) 1990 - 1992: Period following the Roberti-Roos Assault Weapons Control Act of 1989 (AWCA). We call it the "b_AWCA_effect" period. 3) 1993 - 1999: Period following the Federal Assault Weapons Ban. We call it the "c_federal_ban_era". 4) 2000 - 2012: Period following the amendment of the Roberti-Roos Act's. We call it the "d_amended_AWCA" period. 5) 2013 - 2015: Period following the passing of 10 New California Gun Control Laws. We call it the "e_post_control_laws". 6) 2016 - 2017: Period following the passing of 6 new gun control bills and additional restrictions. We call it the "gunpocalypse" period.

```
gun_types <- c('Firearm, type not stated', 'Handgun - pistol, revolver, etc', 'Other gun', 'Rifle', 'Sho
california <- murder %>%
  filter(state=="California", weapon==gun_types) %>%
  select(cntyfips, state, solved, year, month, weapon) %>%
  group_by(year,cntyfips) %>% mutate(count_county=n())
```

```r
#We create identifiers for the periods previous and after relevant legislation
california <- california %>% mutate(gun_era=ifelse(
  year %in% 1976:1989, 'a_looselaws', ifelse(
    year %in% 1990:1992, 'b_AWCA_effect', ifelse(
      year %in% 1993:1999, 'c_federal_ban_era', ifelse(
        year %in% 2000:2012, 'd_amended_AWCA', ifelse(
          year %in% 2013:2015,'e_post_control_laws', 'f_gunpocalypse'))))))
#Boxplot of number of murders between California's counties differentiating by " gun eras"
california <- california %>% group_by(year) %>% mutate(means=mean(count_county))
ggplot(california)+
  geom_boxplot(aes(x=year, y=count_county, group=year, colour=gun_era))+
  geom_line( aes(x=year, y=means))+
  ggtitle("Distribution of the number of murders within California's counties")
```


Distribution of the number of murders within California's counties