Data Science for Public Policy

Applied Introduction to R's Tidyverse

Lezhi Charles Cao - lc1267

```
library(tidyverse)
library(ipumsr)
library(srvyr)
library(haven)
library(ggplot2)
library(treemapify)
library(ggridges)
```

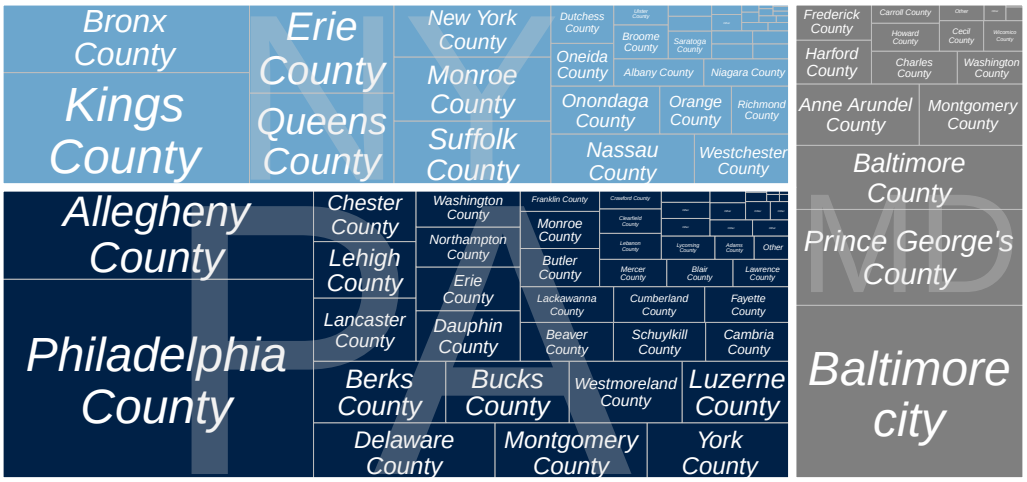
Visualization 01 (1 point)

```
gun_d <- read_csv("gun_deaths_us_1999_2019.csv")

gun.3states <- filter(
  gun_d,
  (State == "NY") | (State == "PA") | (State == "MD")
  )
gun.3states.d <- gun.3states %>%
  group_by(State, County) %>%
  summarize(total_deaths = sum(Deaths),
            .groups = "drop")

gun.3states.d <- gun.3states.d %>%
  group_by(State, County) %>%
  summarise(total_deaths = sum(total_deaths)) %>%
  ungroup() %>%
  mutate(County = if_else(total_deaths < 100,
                          "Other", County))  %>%
```

```
  group_by(State, County)
#counties with total gun death lower than 100 is combined into the
#other category

ggplot(gun.3states.d, aes(
  area = total_deaths,
  fill = State,
  label = County,
  subgroup = State)) +
  geom_treemap(show.legend = FALSE)+
  geom_treemap_subgroup_border(
    color = "white",
    show.legend = FALSE)+
  geom_treemap_text(
    fontface = "italic",
    color = "white",
    place = "center",
    min.size = 0.1,
    reflow = T)+
  geom_treemap_subgroup_text(
    place = "center",
    alpha = 0.3,
    color = "lightgrey",
    min.size =0,
    grow = T)+
  scale_fill_manual(
    values = c("NY" = "skyblue3",
               "PA" = "#002147",
               "NJ" = "darkgoldenrod"))+
  #color chosen from the different state colors
  labs(title= (paste ("Cumulative Gun Deaths by County for New York (NY), Pennsylvania (PA
       subtitle = "PA has the highest cumulative Gun Deaths by State and Individual County
       caption = "*Counties with cumulative gun deaths lower than 100 are merged into the
Data: Centers for Disease Control and Prevention") +
  theme(plot.title = element_text(size=11, face = "bold"),
        plot.subtitle = element_text(size=8.5, face = "italic"),
        plot.caption = element_text(size = 7))
```

Cumulative Gun Deaths by County for New York (NY), Pennsylvania (PA) an New Jersey (NJ): 1999 − 2019*

PA has the highest cumulative Gun Deaths by State and Individual County



*Counties with cumulative gun deaths lower than 100 are merged into the other category
Data: Centers for Disease Control and Prevention

Data URL: https://data.world/nkrishnaswami/gun-deaths-by-county/workspace/project-summary?agentid=nkrishnaswami&datasetid=gun-deaths-by-county

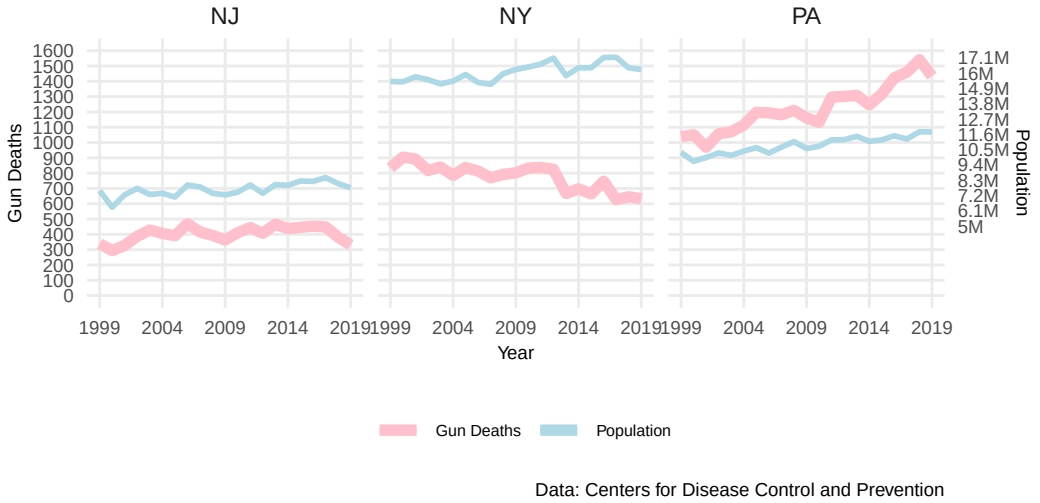
Visualization 02 (2 points)

```r
gun.states <-
  gun_d %>%
  filter(State == "PA" | State == "NY" | State == "NJ") %>%
  group_by(State,Year) %>%
  select(State,Year,Population,Deaths) %>%
  summarize(deaths=sum(Deaths), population = sum(Population))

ggplot(gun.states, aes(x = Year)) +
  geom_line(aes(y = deaths,
                color = "Gun Deaths"),
            size = 2) +
  geom_line(aes(y = population/11000, color = "Population"), size = 1) +
  scale_color_manual(
    values = c("Population" = "lightblue", "Gun Deaths" = "pink"),
    labels = c("Gun Deaths", "Population")
  ) +
  facet_wrap(~ State, ncol = 3) +
  labs(
    title = "Population and Gun Deaths for New Jersey (NJ), New Jersey (NJ) and \nPennslyv
    subtitle = "On average, NJ and NY's population and gun death have a similar trend, whi
    caption = "Data: Centers for Disease Control and Prevention") +
  theme_minimal() +
  scale_y_continuous(
    name = "Gun Deaths",
    limits = c(0, 1600),
    breaks = seq(0, 1600, 100),
    sec.axis = sec_axis(
      ~ . *11000,
      name = "Population",
      breaks = seq(5*10^6, 17600000, by = 1100000),
      labels = c("5M", "6.1M", "7.2M", "8.3M", "9.4M", "10.5M", "11.6M","12.7M", "13.8M",
    )
  )+
   scale_x_continuous(
    breaks = seq(1999, 2019, 5), labels = seq(1999, 2019, 5)
  )+
  theme(plot.title = element_text(size=11, face = "bold"),
        plot.subtitle = element_text(size=8.5, face = "italic"),
        plot.caption = element_text(size = 6.5),
```

```
        legend.title = element_blank(),
        legend.text = element_text(size = 6),
        legend.position = "bottom",
        axis.text = element_text(size=7),
        axis.title = element_text(size=7),
        panel.grid.minor.y = element_blank(),
        panel.grid.minor.x = element_blank())
```

Population and Gun Deaths for New Jersey (NJ), New Jersey (NJ) and Pennslyvania (PA): 1999 − 2019

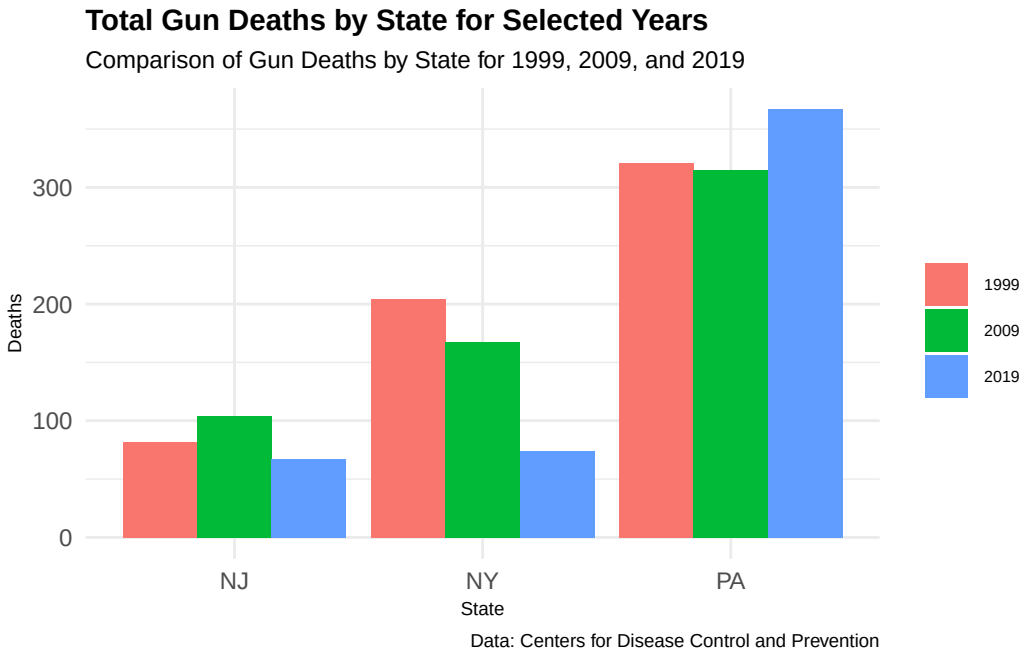
On average, NJ and NY's population and gun death have a similar trend, while PA's gun de start to deviate from population trend since 2014.



Data: Centers for Disease Control and Prevention

Visualization 03 (2 points)

```r
# Calculate 5-year intervals for the years 1999-2019
intervals <- seq(1999, 2019, by = 5)

# Create a dataset with the total gun deaths in each 5-year interval
gun_deaths_intervals <- gun_d %>%
  filter(State %in% c("NY", "PA", "NJ")) %>%
  mutate(Interval = cut(Year, breaks = intervals, labels = intervals[-1], include.lowest =
  group_by(State, Interval) %>%
  summarize(TotalDeaths = sum(Deaths))


library(dplyr)
library(ggplot2)

selected_years <- c(1999, 2009, 2019)
filtered_data <- gun_d %>%
  filter(State %in% c("NY", "PA", "NJ"), Year %in% selected_years)

# Bar Chart Comparing Total Gun Deaths for Selected States (1999, 2009, 2019)
state_chart <- ggplot(filtered_data, aes(x = State, y = Deaths, fill = as.factor(Year))) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(
    title = "Total Gun Deaths by State for Selected Years",
    subtitle = "Comparison of Gun Deaths by State for 1999, 2009, and 2019",
    caption = "Data: Centers for Disease Control and Prevention"
  ) +
  scale_fill_discrete(name = "Year") +
  theme_minimal() +
  theme(
    plot.title = element_text(size = 11, face = "bold"),
    plot.subtitle = element_text(size = 8.5),
    plot.caption = element_text(size = 6.5),
    legend.title = element_blank(),
    legend.text = element_text(size = 6),
    axis.title = element_text(size = 7)
  )

print(state_chart)
```


Total Gun Deaths by State for Selected Years
Comparison of Gun Deaths by State for 1999, 2009, and 2019
Deaths
300
200
100
0
NJ
NY
PA
State
1999
2009
2019
Data: Centers for Disease Control and Prevention

Exercise 04 (2 points)

```r
pie_chart <- ggplot(gun_deaths_intervals, aes(x = "", y = TotalDeaths, fill = as.factor(In
  geom_bar(stat = "identity") +
  coord_polar(theta = "y") +
  labs(
    title = "Distribution of Gun Deaths by 5-Year Intervals (1999-2019)",
    subtitle = "The Proportion of Gun Deaths in 5-Year Intervals",
    caption = "Data: Centers for Disease Control and Prevention"
  ) +
  scale_fill_discrete(name = "5-Year Interval") +
  theme_void() +
  theme(
    plot.title = element_text(size = 11, face = "bold"),
    plot.subtitle = element_text(size = 8.5),
    plot.caption = element_text(size = 6.5),
    legend.title = element_blank(),
    legend.text = element_text(size = 6)
  )

print(pie_chart)
```