# HW 03

Matt Scheffel

2022-09-12

```
library("tidyverse")
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.3.6     v purrr   0.3.4
## v tibble  3.1.8     v dplyr   1.0.9
## v tidyr   1.2.0     v stringr 1.4.1
## v readr   2.1.2     v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
killings.df<-read.csv("PoliceKillings.csv", header=TRUE)
```

## Question 1

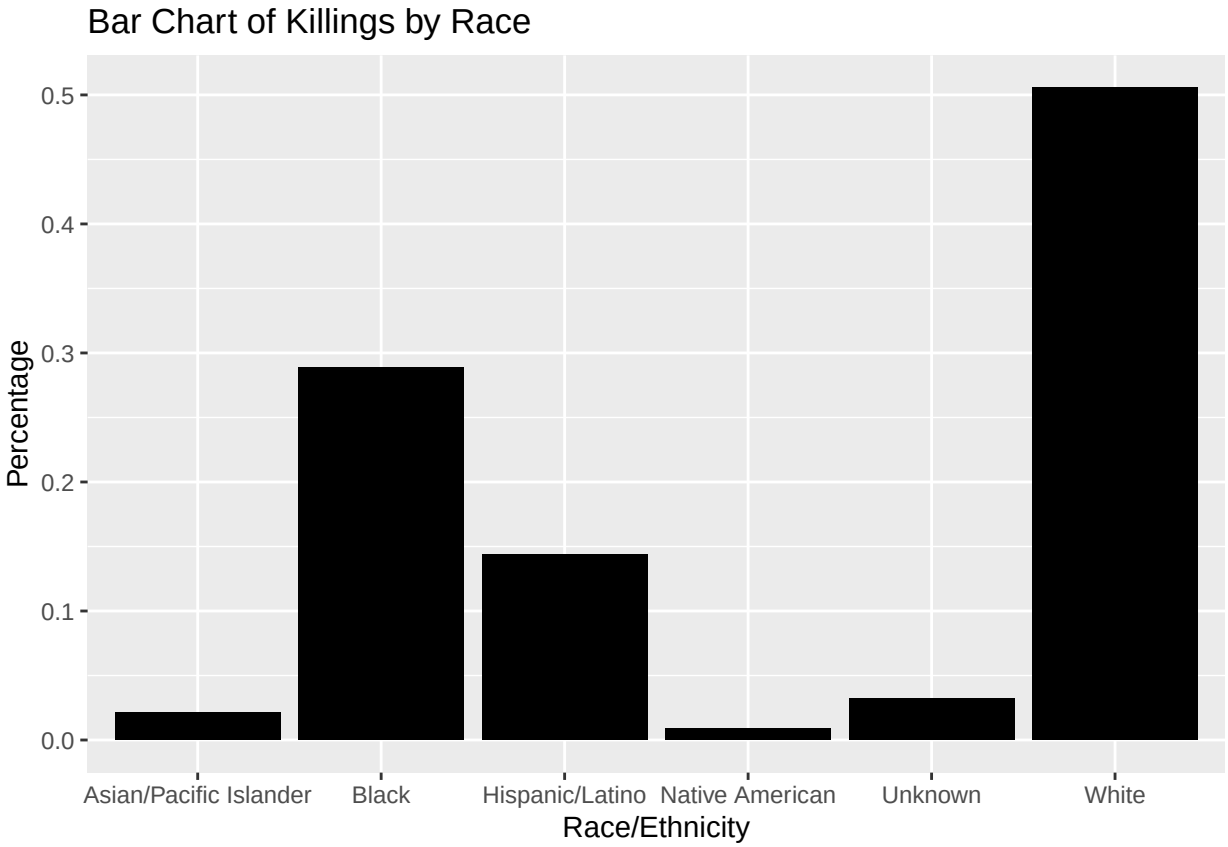
## Q1A

```
percentkillings<-killings.df%>%
  filter(!is.na(raceethnicity)) %>%
  group_by(raceethnicity)%>%
  summarize(Counts=n())%>%
  mutate(Percent=Counts/sum(Counts))

mytable<-table(killings.df$raceethnicity)
round(prop.table(mytable)*100, 2)
```

```
##
## Asian/Pacific Islander                 Black        Hispanic/Latino
##                   2.14                 28.91                   14.35
##        Native American               Unknown                   White
##                   0.86                  3.21                   50.54
```

```
percentkillings%>%
  ggplot(aes(x=raceethnicity, y=Percent))+
  geom_bar(fill="black", stat="identity")+
  labs(x="Race/Ethnicity", y ="Percentage",
       title="Bar Chart of Killings by Race")
```




Comment: According to the US census data, people who identify as white make up 75.8% of the US population, people who identify as Black make up 13.6%, and people who identify as Hispanic/Latino make up 18.9%. However, here we can see that the Black population makes up for almost 28.9% of the police killings in our dataset, while the white population makes up just over 50%. Essentially, our data shows us that for some of the races/ethnicities, the amount of deaths are statistically disproportionate in comparison to their share of the population.

# Q1B

```
age.num<-(as.numeric(killings.df$age))
```

```
## Warning: NAs introduced by coercion
```

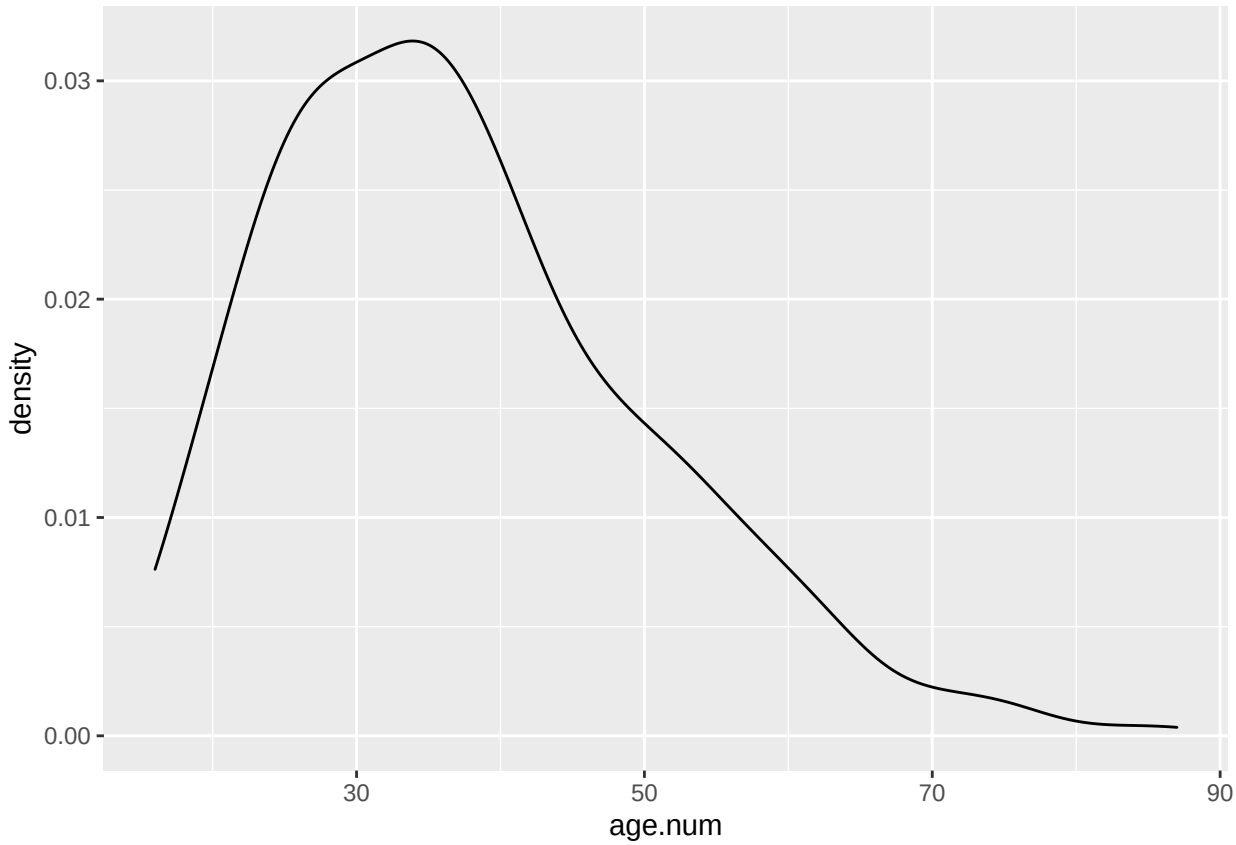
```
killings.df<-data.frame(killings.df, age.num)
is.numeric(killings.df$age.num)
```

```
## [1] TRUE
```

# Q1C

```
ggplot(killings.df,aes(x=age.num))+
  geom_density()
```

```
## Warning: Removed 4 rows containing non-finite values (stat_density).
```

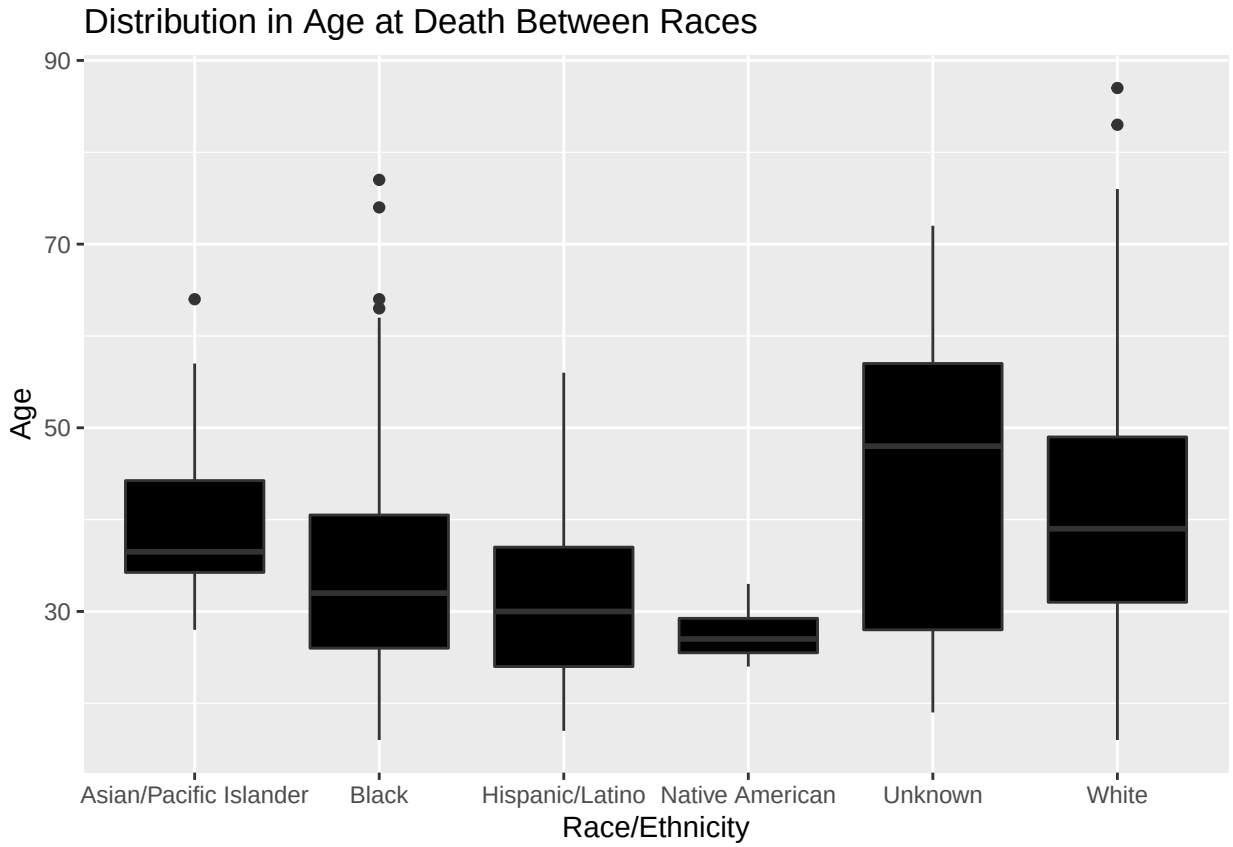



Comments: This density plot shows that the majority of police killings occur while victims are young to middle aged, with the early 30's being the most densely populated at around 30%. As the population ages, there a far fewer deaths at the hands of police according to this data, dropping off towards 0% in the 80's.

## Q1D

```
ggplot(killings.df, aes(x=raceethnicity, y=age.num))+
  geom_boxplot(fill="black")+
  labs(x="Race/Ethnicity", y="Age",
       title="Distribution in Age at Death Between Races")
```

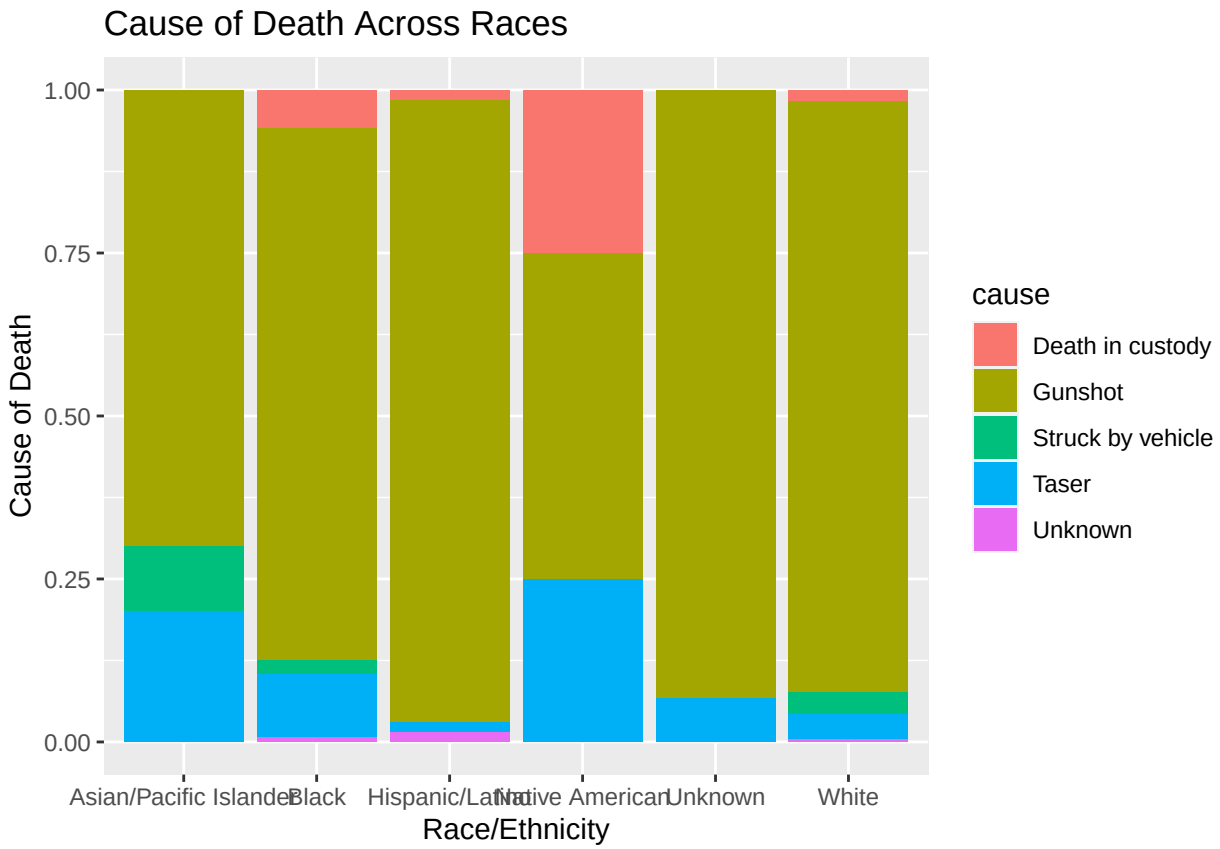
```
## Warning: Removed 4 rows containing non-finite values (stat_boxplot).
```




Comments:This boxplot shows the distance in age at death (due to police killings) across different race/ethnicities. We can see that "White" has the largest distance among the "whiskers", however a few other groups also have high outliers, particularly the Black population. Native Americans are concentrated within the shortest distance (and youngest median age). "Unknown" has the highest median age.

## Q1E

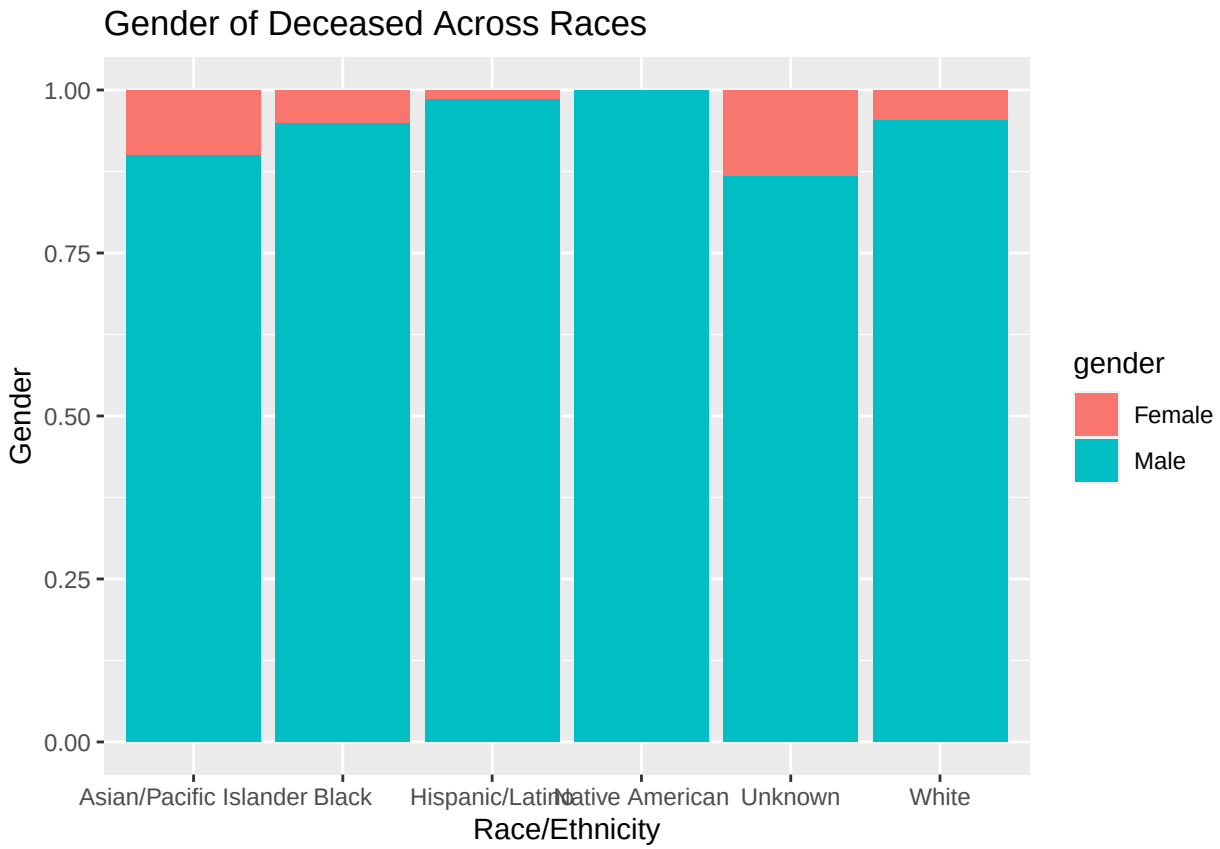
```
ggplot(killings.df, aes(x=raceethnicity, fill=cause))+
  geom_bar(position= "fill")+
  labs(x="Race/Ethnicity", y="Cause of Death",
       title="Cause of Death Across Races")
```




Comments: This (fill) bar chart shows the specific cause of death among different races/ethnicities. We see that a larger proportion of Blacks and Native Americans die in police custody. We also can see that Asian/Pacific Islanders have the largest proportion of deaths from being struck by a vehicle. Most causes of death are from gunshots. From this visualization, I would say that the cause of death does not appear to be independent from race/ethnicity for certain variables (especially tasers, vehicles, and deaths in custody.) for certain races, the method of death appears disproportionate and thus not independent. However, death from gunshot does appear to be independent as it is the leading cause of death for every race/ethnicity.

## Q1F

```
ggplot(killings.df, aes(x=raceethnicity, fill=gender))+
  geom_bar(position= "fill")+
  labs(x="Race/Ethnicity", y="Gender",
       title="Gender of Deceased Across Races")
```




Comments: This visualization shows the gender of the victims across race. All around, we can see that females make up a much smaller proportion of deaths from police killings than males. After that, we can see that women from the Unknown and Asian/PI populations have the highest proportions of deaths, respectively.

## Question 2

```
statecovid.df<-read.csv("stateCovid.csv", header=TRUE)
statepop.df<-read.csv("State_pop_election.csv", header=TRUE)
```

## Q2A

```
statecovid.df<-rename(statecovid.df, State = state)
statecovid.df<-statecovid.df%>%
  select(-X)
newdata<-merge(statecovid.df,statepop.df, by="State")
newdata<-newdata[order(newdata$State),]
head(newdata)
```

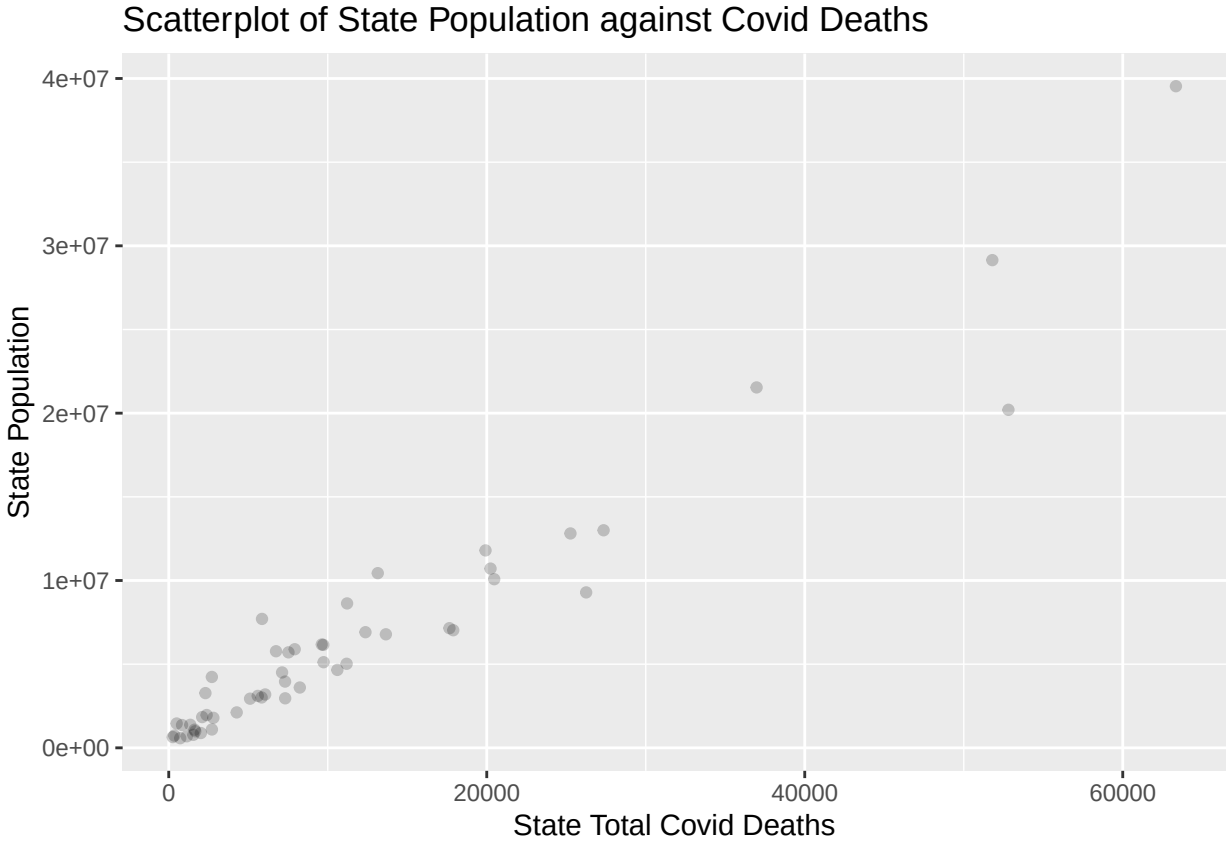
```
##        State state_total_cases state_total_deaths state.rate Population
## 1    Alabama            545028              11188      2.05%    5024279
## 2     Alaska             69826                352      0.50%     733391
## 3    Arizona            882691              17653      2.00%    7151502
## 4   Arkansas            341889               5842      1.71%    3011524
## 5 California           3793055              63345      1.67%   39538223
## 6   Colorado            547961               6746      1.23%    5773714
```

```
##   Election
## 1    Trump
## 2    Trump
## 3    Biden
## 4    Trump
## 5    Biden
## 6    Biden
```

## Q2B

```
ggplot(newdata, aes(x=state_total_deaths,y=Population))+
  geom_point(alpha=0.2)+
  labs(x="State Total Covid Deaths", y="State Population",
       title="Scatterplot of State Population against Covid Deaths")
```

```
## Warning: Removed 1 rows containing missing values (geom_point).
```




Comments: This visualization shows a scatterplot of State Population against Covid Deaths. There is a relatively easy pattern to follow with the scatterplot: as state population increases, the total number of Covid deaths seem to rise. There is likely a correlation between state population and state Covid deaths. Logically, this makes sense, as a larger population means there are more people to contract and spread the disease (and potentially die from it).

## Fin