```r
# read data
covidDat<-read.csv("./data/RaceDataEntryCRDT.csv")
popDat<-read.csv("./data/StatePopulation.csv")

library(tidyverse)

# transforming character values of number into numeric values
df<-popDat%>%mutate(Total.population=as.numeric(gsub(",","",Total.population)),
  Hispanic..of.any.race.=as.numeric(gsub(",","",Hispanic..of.any.race.)),
  Non.Hispanic.White=as.numeric(gsub(",","",Non.Hispanic.White)),
  Non.Hispanic.Black=as.numeric(gsub(",","",Non.Hispanic.Black)),
  Non.Hispanic.Asian=as.numeric(gsub(",","",Non.Hispanic.Asian)),
  Non.Hispanic.American.Indian=as.numeric(gsub(",","",
                                           Non.Hispanic.American.Indian)))
names(df)<-c("state","pop","Hispanic","White","Black","Asian","NHAI")

# transform population data from wide to long format
popDat_long <- gather(df, key=race, racepop, 3:7, factor_key=TRUE)



# add state column to population data
popDat_long$state<-as.character(levels(popDat_long$state))[popDat_long$state]

# remove "District of Columbia"
popDat_long<-popDat_long[-which(popDat_long$state=="District of Columbia"),]

# change state name from abbreviation to full name
for(i in 1:length(popDat_long$state)){
popDat_long$state[i]<-state.abb[which(state.name == popDat_long$state[i])]
}

names(popDat_long)<-c("State","Pop","Race","Racepop")


## covid data

# transform NA into 0
covidDat[is.na(covidDat)] <- 0

# remove rows with 0
coviddf<-covidDat%>%
  filter(Cases_Total!="")%>%
  filter(Cases_Total!="0")%>%
  filter(Cases_White!="0")%>%
  filter(Cases_White!="")%>%
  mutate(Cases_Total=as.numeric(Cases_Total),
         Cases_White=as.numeric(Cases_White))%>%
  select(Date,State,Deaths_White,Deaths_Black,Deaths_LatinX,
         Deaths_Asian,Deaths_AIAN,Deaths_NHPI,Deaths_Multiracial)

library(dplyr)
names(coviddf)<-c("Date","State","White","Black","Hispanic",
```

```r
                 "Asian","NHAI","NHPI","multi")

# sum death numbers over all the period
coviddf_sum<-coviddf[,-1] %>% group_by(State) %>% summarise_each(funs(sum))

# transform covid data from wide to long
coviddf_long<-gather(coviddf_sum,key=Race,death_cases,2:8)

# merge covid data and population data by state and race
combdf<-left_join(coviddf_long,popDat_long,by=c("State","Race"))

# remove "DC" and "GU", since data for those two states is incomplete.
combdf2<-combdf%>%filter(State!="DC")
combdf2<-combdf2%>%filter(State!="GU")

# calculate death rate by dividing death number by population for each race.
finaldf<-combdf2%>%
  mutate(death_rate=death_cases/Racepop,race_rate=Racepop/Pop)

# indicate observations (states) which is 1.33 times higher than the rate of the race for the state.
# we will highlight the states which is indicated here
finaldf$ind<-NA
for(i in 1:length(finaldf$ind)){
finaldf$ind[i]<-
  ifelse(i %in% which(finaldf$death_rate>=(1.33*finaldf$race_rate)),"1","0")
}
finaldf$ToHighlight<-finaldf$ind

# Is at least 33% higher than the Census Percentage of Population.


# generate barchart for white group
ggplot()+
  geom_bar(aes(x=State,y=death_rate, fill = ToHighlight),
           data=finaldf[which(finaldf$Race=="White"),],stat="identity")+
  scale_fill_manual( values = c( "1"="tomato", "0"="gray" ), guide = FALSE )+
  theme_bw()+
  theme(axis.text=element_text(size=6,angle=90),
        plot.title = element_text(hjust = 0.5))+
  labs(title="Death Rate of covid by state, Race = White",
       x="State",y="Death Rate")

# generate barchart for Black group
ggplot()+
  geom_bar(aes(x=State,y=death_rate, fill = ToHighlight),
           data=finaldf[which(finaldf$Race=="Black"),],
           stat="identity")+
  scale_fill_manual( values = c( "1"="tomato", "0"="gray" ), guide = FALSE )+
  theme_bw()+
  theme(axis.text=element_text(size=6,angle=90),
        plot.title = element_text(hjust = 0.5))+
  labs(title="Death Rate of covid by state, Race = Black",
       x="State",y="Death Rate")
```

```r
# generate barchart for Asian group
ggplot()+
  geom_bar(aes(x=State,y=death_rate, fill = ToHighlight),
           data=finaldf[which(finaldf$Race=="Asian"),],
           stat="identity")+
  scale_fill_manual( values = c( "1"="tomato", "0"="gray" ), guide = FALSE )+
  theme_bw()+
  theme(axis.text=element_text(size=6,angle=90),
        plot.title = element_text(hjust = 0.5))+
  labs(title="Death Rate of covid by state, Race = Asian",
       x="State",y="Death Rate")

# generate barchart for Hispanic group
ggplot()+
  geom_bar(aes(x=State,y=death_rate, fill = ToHighlight),
           data=finaldf[which(finaldf$Race=="Hispanic"),],stat="identity")+
  scale_fill_manual( values = c( "1"="tomato", "0"="gray" ), guide = FALSE )+
  theme_bw()+
  theme(axis.text=element_text(size=6,angle=90),
        plot.title = element_text(hjust = 0.5))+
  labs(title="Death Rate of covid by state, Race = Hispanic",
       x="State",y="Death Rate")

# generate barchart for NHAI group
ggplot()+
  geom_bar(aes(x=State,y=death_rate, fill = ToHighlight),
           data=finaldf[which(finaldf$Race=="NHAI"),],stat="identity")+
  scale_fill_manual( values = c( "1"="tomato", "0"="gray" ),
                     guide = FALSE )+
  theme_bw()+
  theme(axis.text=element_text(size=6,angle=90),
        plot.title = element_text(hjust = 0.5))+
  labs(title="Death Rate of covid by state, Race = NHAI",
       x="State",y="Death Rate")
```