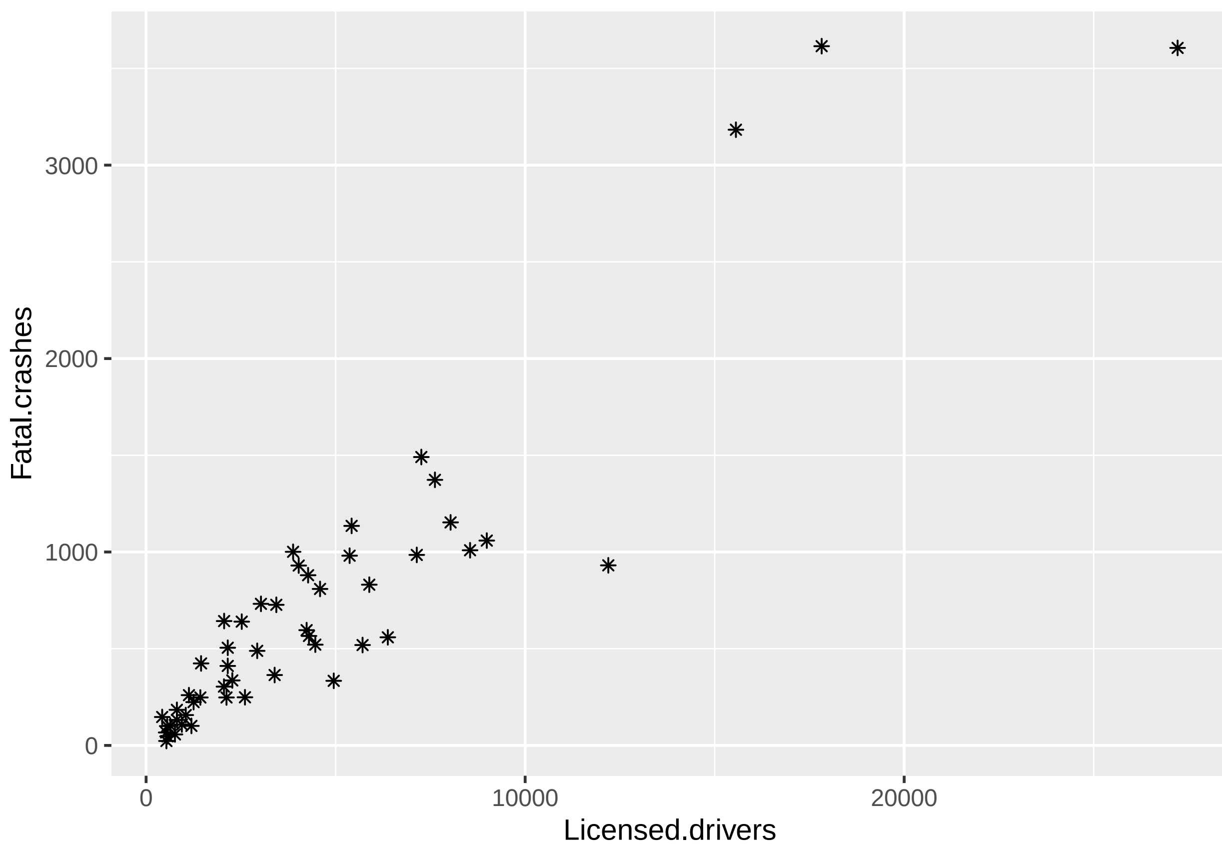
Homework 5

Noah McIntire

Problem 1

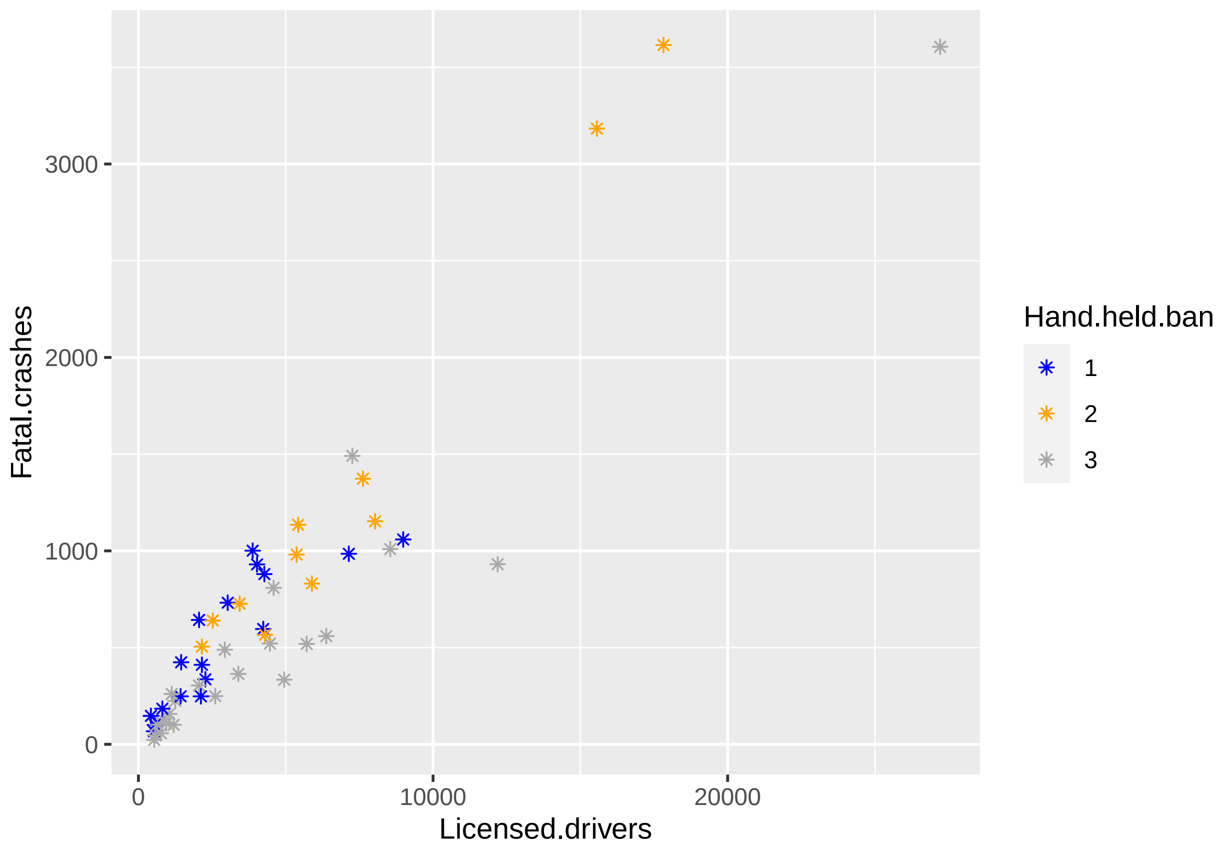
a)

```
library(ggplot2)
crash<-read.csv("/Users/noahmcintire/Desktop/STAT 3080/state crashes.csv")
plot1<-ggplot(crash, aes(x=Licensed.drivers, y=Fatal.crashes)) + geom_point(shape=8)
plot1
```



b)

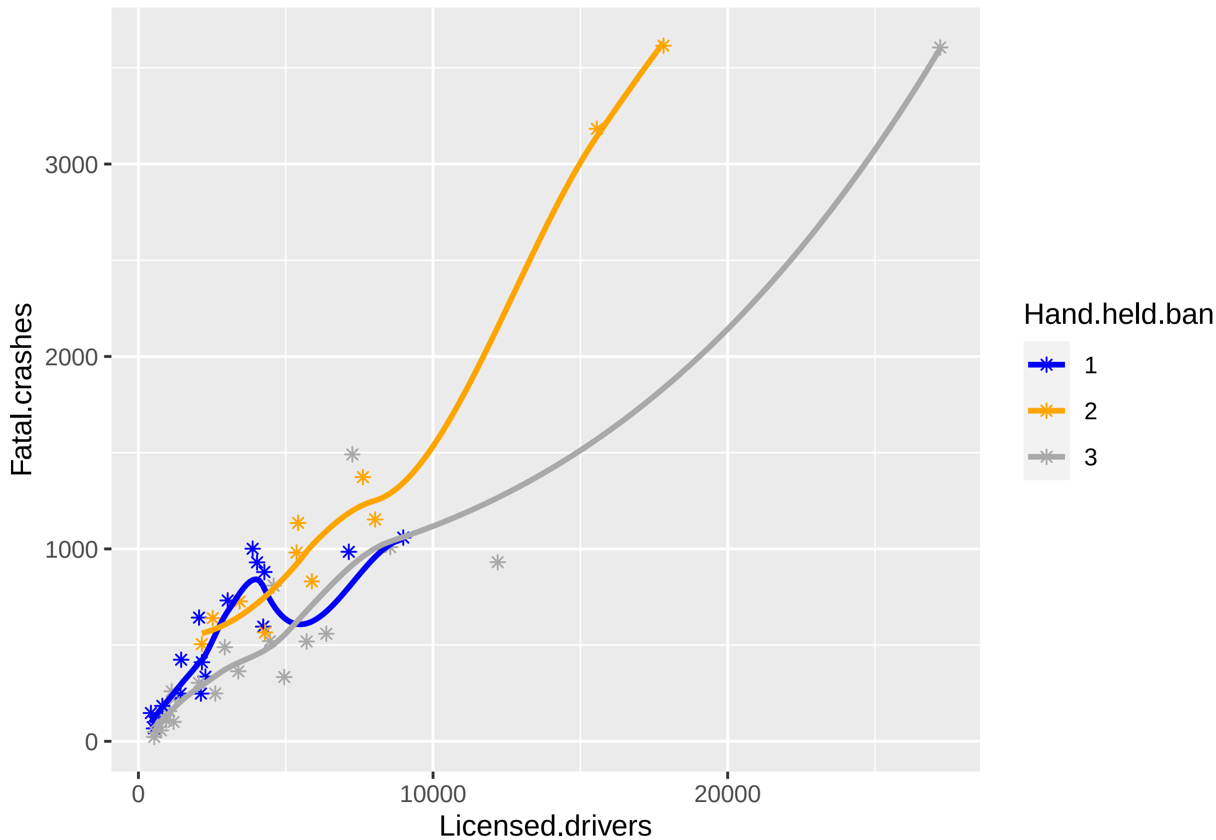
```
crash$Hand.held.ban<- as.factor(crash$Hand.held.ban)
plot2<-ggplot(crash, aes(x=Licensed.drivers, y=Fatal.crashes, color=Hand.held.ban)) + ge
plot2
```



c)

```
plot3<- plot2 + geom_smooth(se = F)
plot3
```

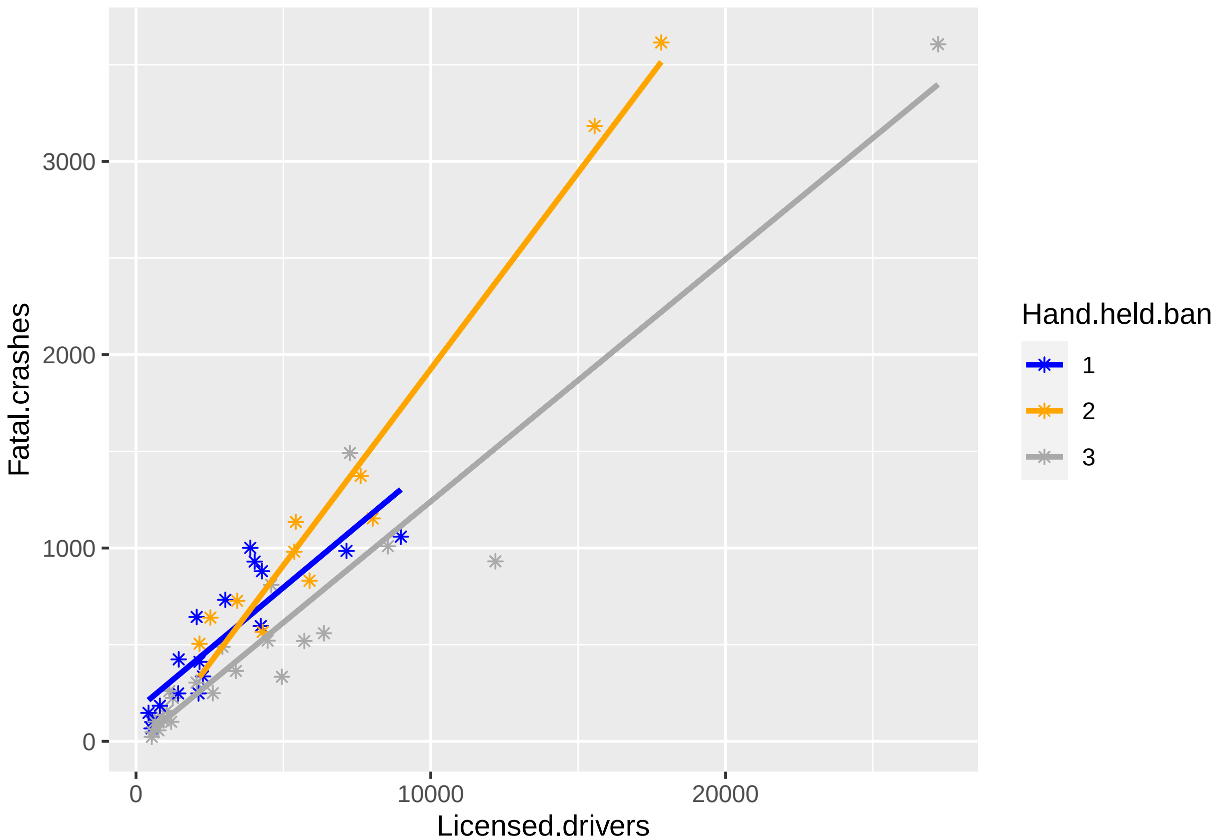
```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



d)

```
plot4<- plot2 + geom_smooth(method=lm,se = F)
plot4
```

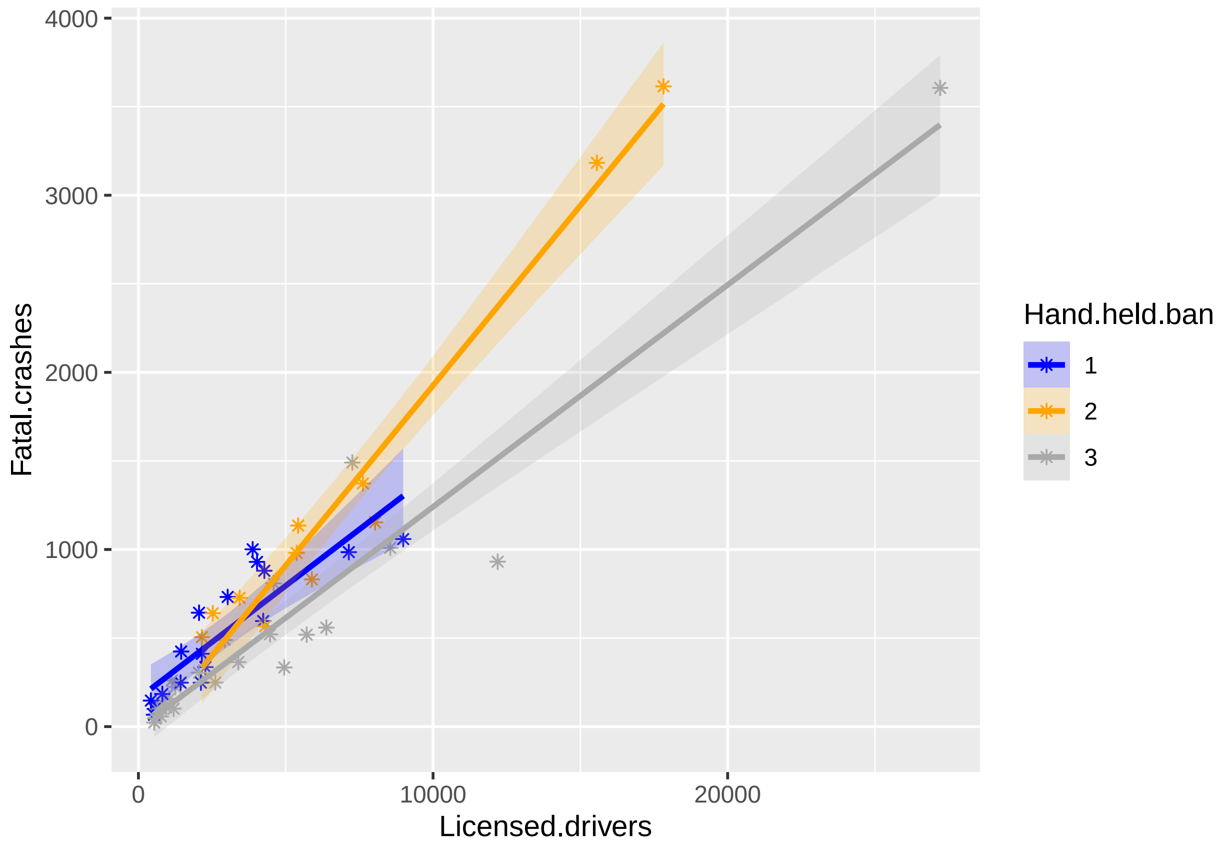
```
## `geom_smooth()` using formula 'y ~ x'
```



e)

```
plot5 <-ggplot(crash, aes(x=Licensed.drivers, y=Fatal.crashes, color=Hand.held.ban)) + g
plot5
```

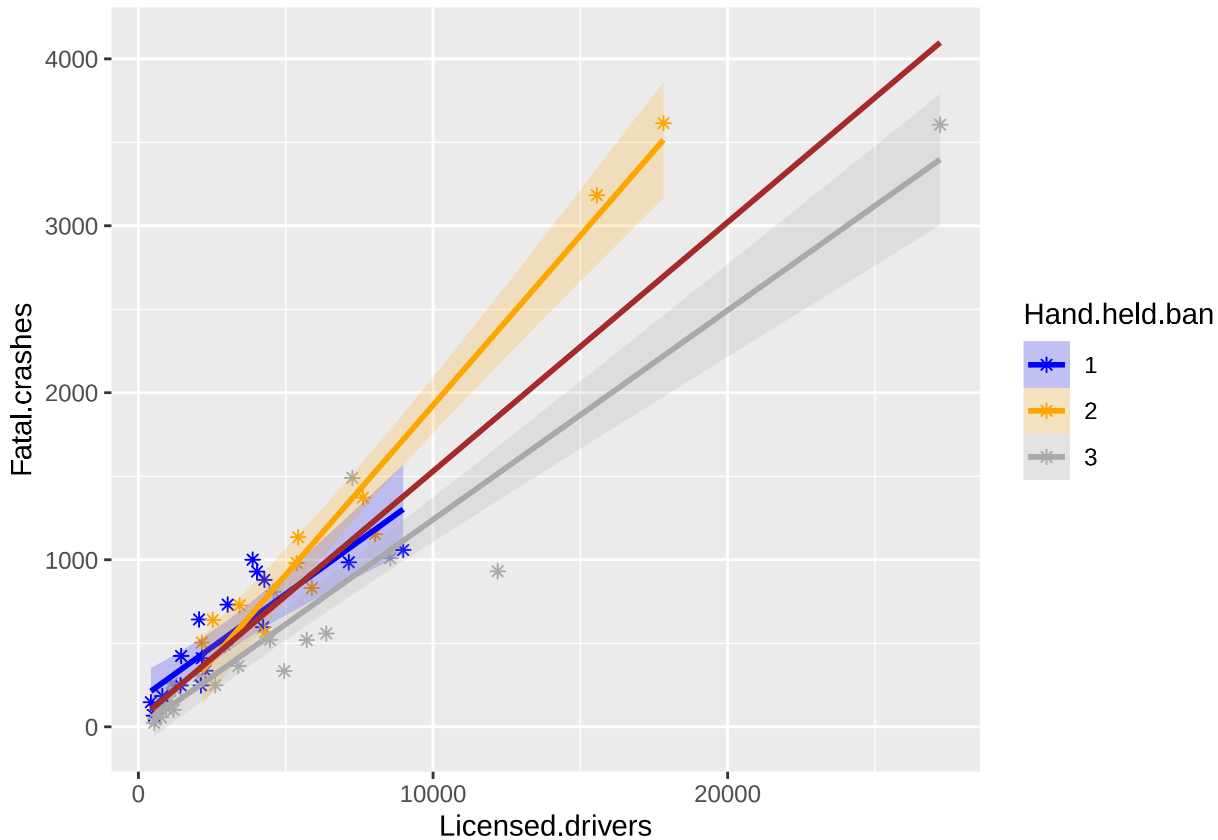
```
## `geom_smooth()` using formula 'y ~ x'
```



f)

```
plot6 <-plot5 + geom_smooth(method=lm, color="brown", se=F)
plot6
```

```
## `geom_smooth()` using formula 'y ~ x'
## `geom_smooth()` using formula 'y ~ x'
```



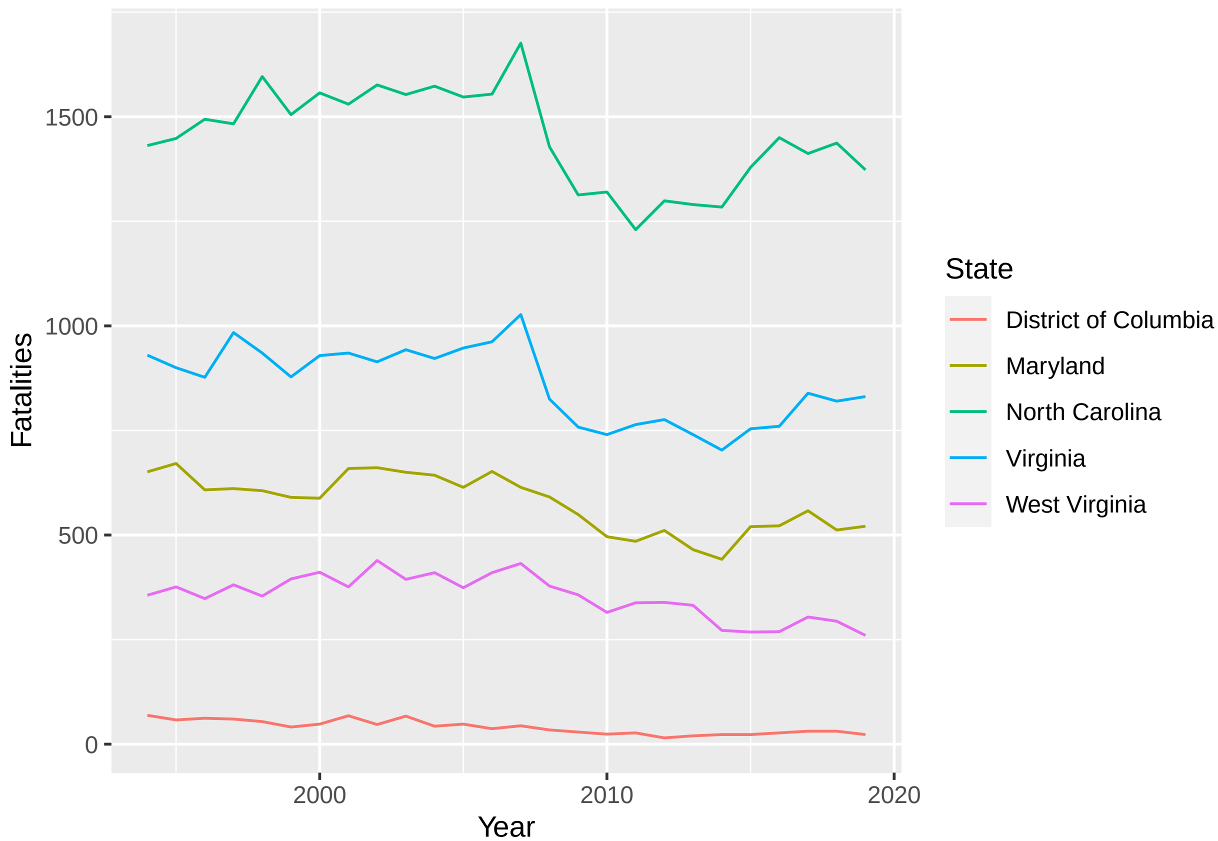
g)

Based on the graph, we can see that there is a strong, positive, linear relationship between the number of fatal accidents and the number of licensed drivers in each state. We also can say that having a hand held electronic device ban will likely cause a decrease the amount of fatal accidents within the state. However, an image does not provide any form of statistical significance, but instead helps to provide a useful visualization.

Problem 2

a)

```
fatal<-read.csv("/Users/noahmcintire/Desktop/STAT 3080/fatalities.csv")
plot7<- ggplot(fatal, aes(x=Year, y=Fatalities, color=State)) + geom_line()
plot7
```

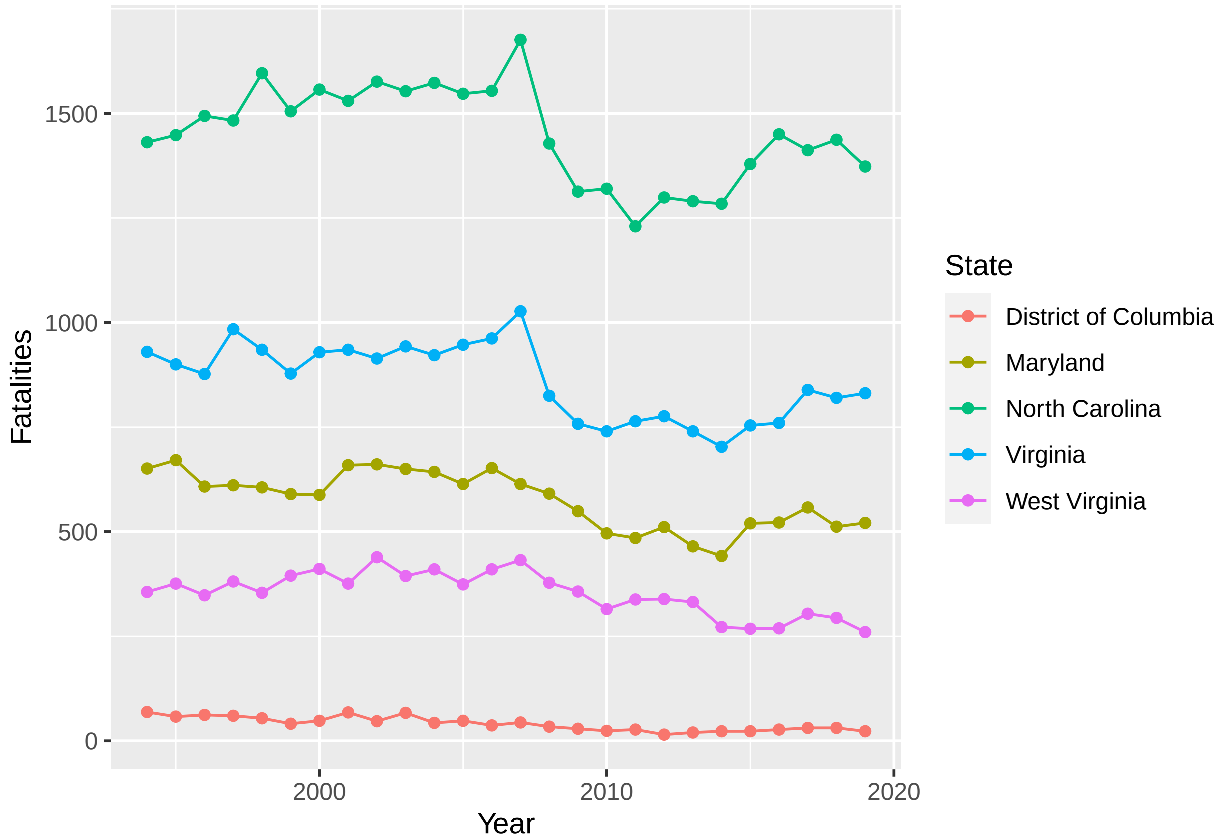


b)

Looking at the whole graph, it seems that the number of fatalities in each state has gone down over time. Additionally, each state's population is reflected in the graph, as the state with the largest population has the most fatalities, while the states that are smaller have less and less road fatalities as the population decreases.

c)

```
plot8<-plot7 + geom_point()
plot8
```

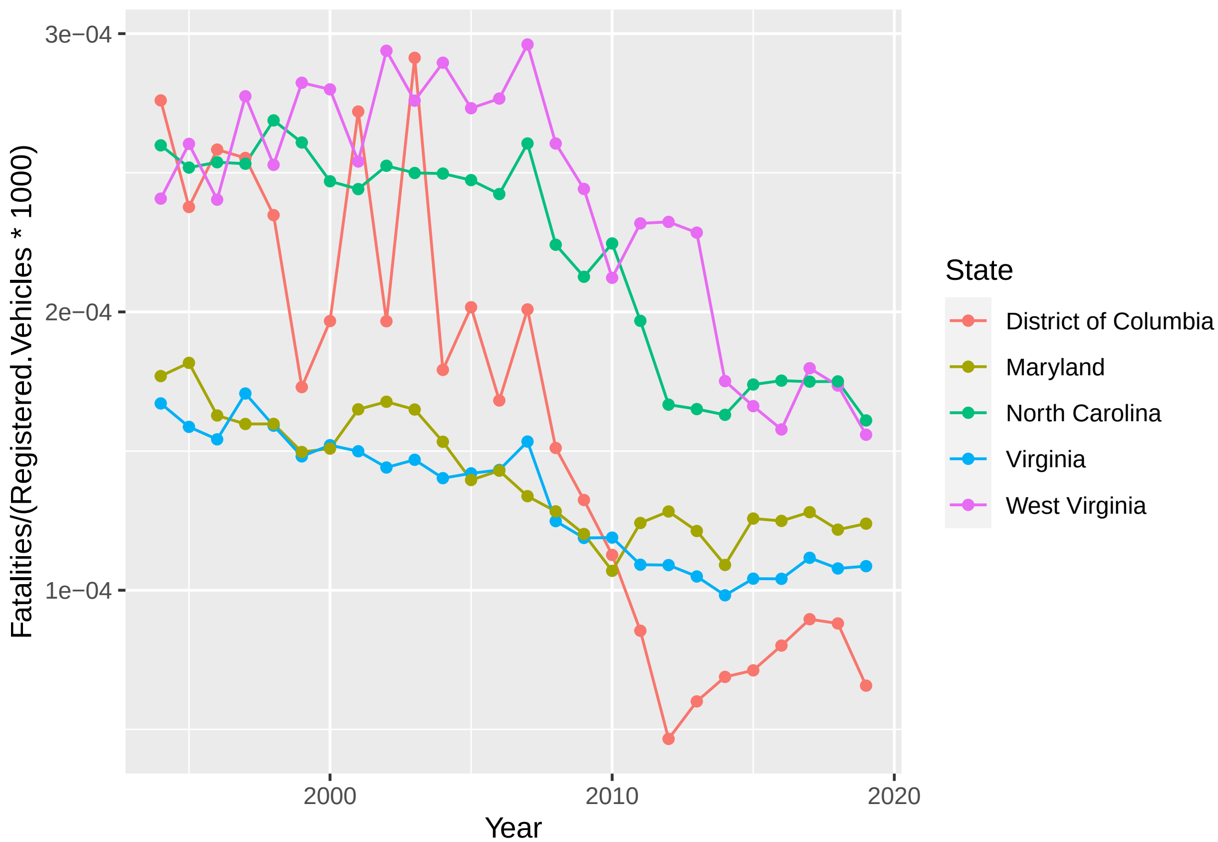


d)

The plot in part c allows us to see each point of data that was recorded, making it easier to visualize what happened between each year and compare individual years among each state.

e)

```
plot9<-ggplot(fatal, aes(x=Year, y=Fatalities/(Registered.Vehicles *1000), color=State )
plot9
```

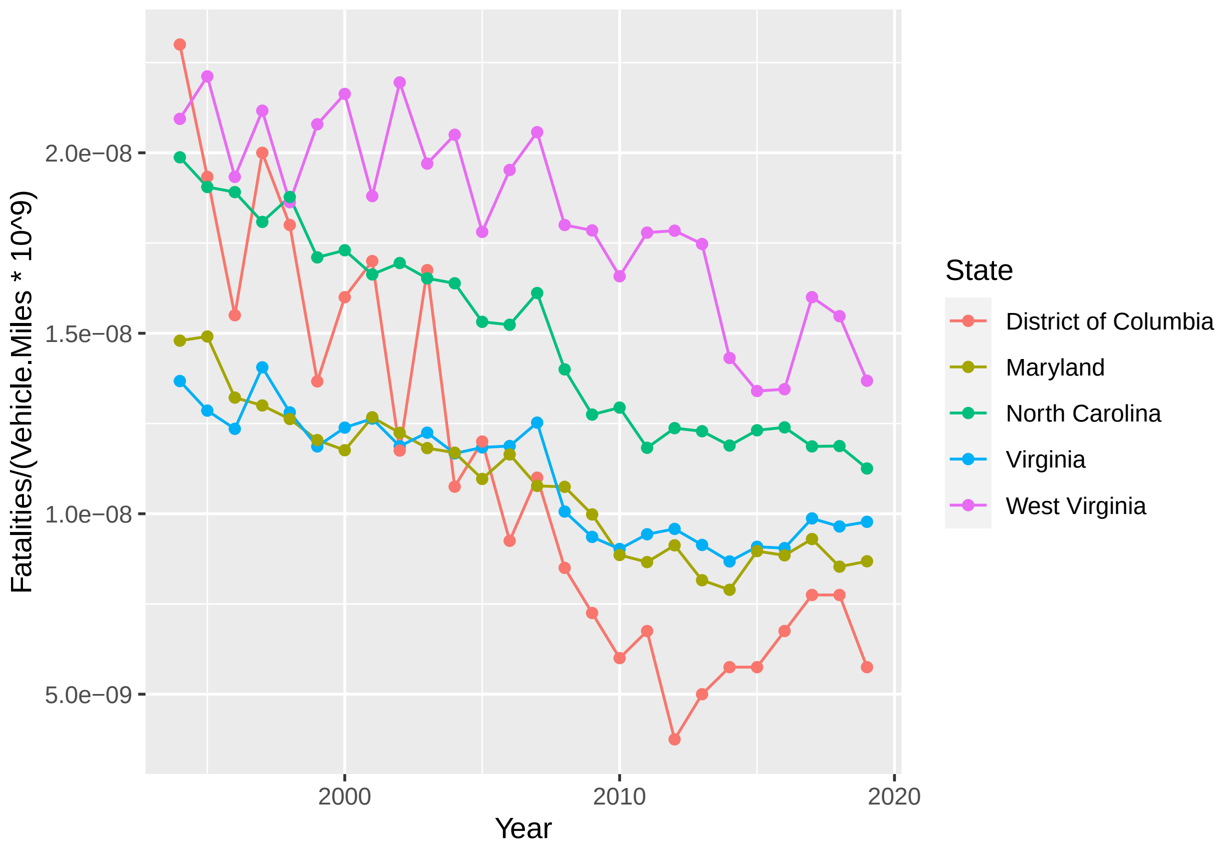


f)

By including the number of registered vehicles within the ratio, we are able to more easily compare each state's rate of fatalities. This still supports my previous statement in part b that all fatalities in car accidents have fallen in each state, especially in proportion to number of vehicles registered. Hoever, we can see that certain states have improved much more drastically in comparison to number of registered vehicles, such as DC and West Virginia.

g)

```
plot10<-ggplot(fatal, aes(x=Year, y=Fatalities/(Vehicle.Miles *10^9), color=State )) + g
plot10
```

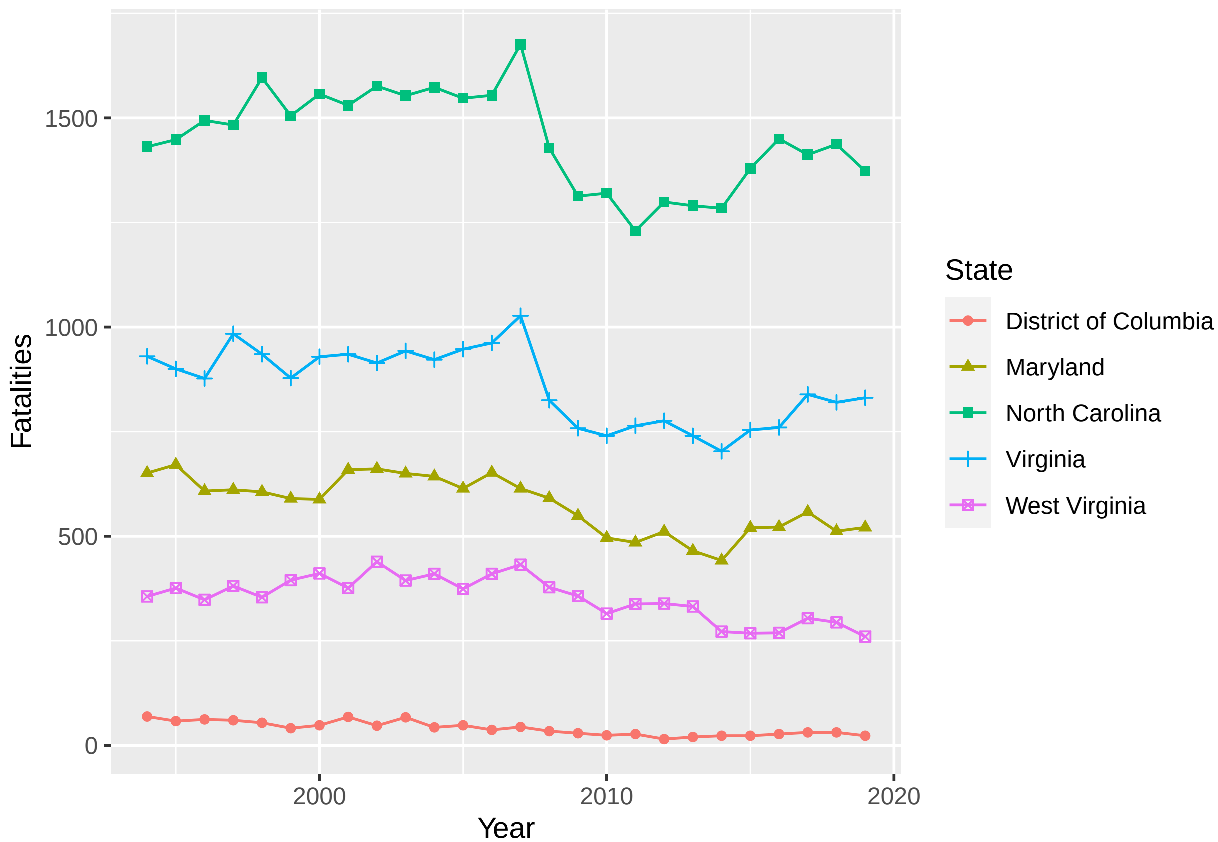


h)

One of the biggest differences by using miles instead of registered vehicles is that we can see that West Virginia has much more fatalities per mile driven then other states throughout most of the graph.

i)

```
plot11<-ggplot(fatal, aes(x=Year, y=Fatalities, color=State, shape=State)) + geom_line()
plot11
```



j)

Varying the point by state provided does not provide any significant benefit in communicating the information. Because each color distinguishes each state, and that both the line and point have the same color, there is no need to change the shape. It makes the graphic look less consistent and less appealing to the eye.