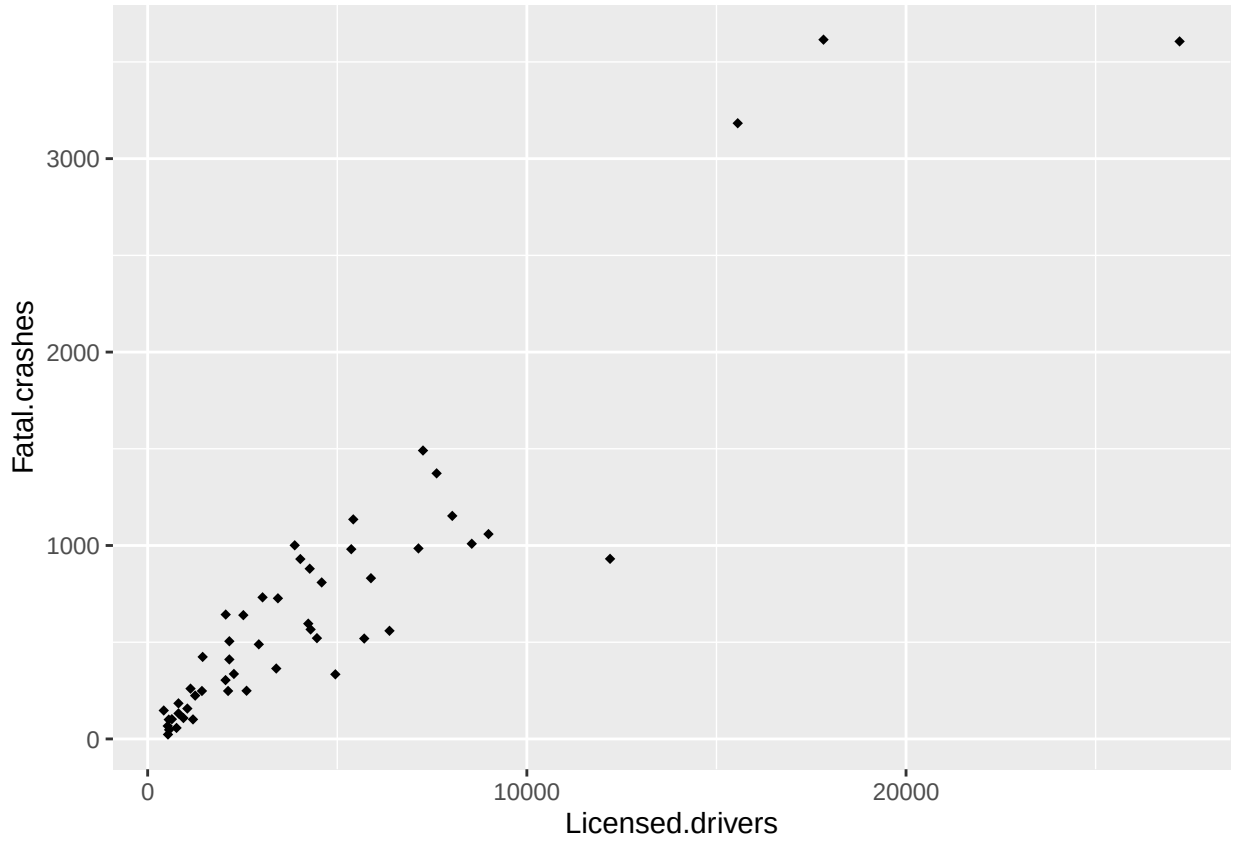
Homework 5

Carlos F Revilla (cfr5spw)

Problem 1

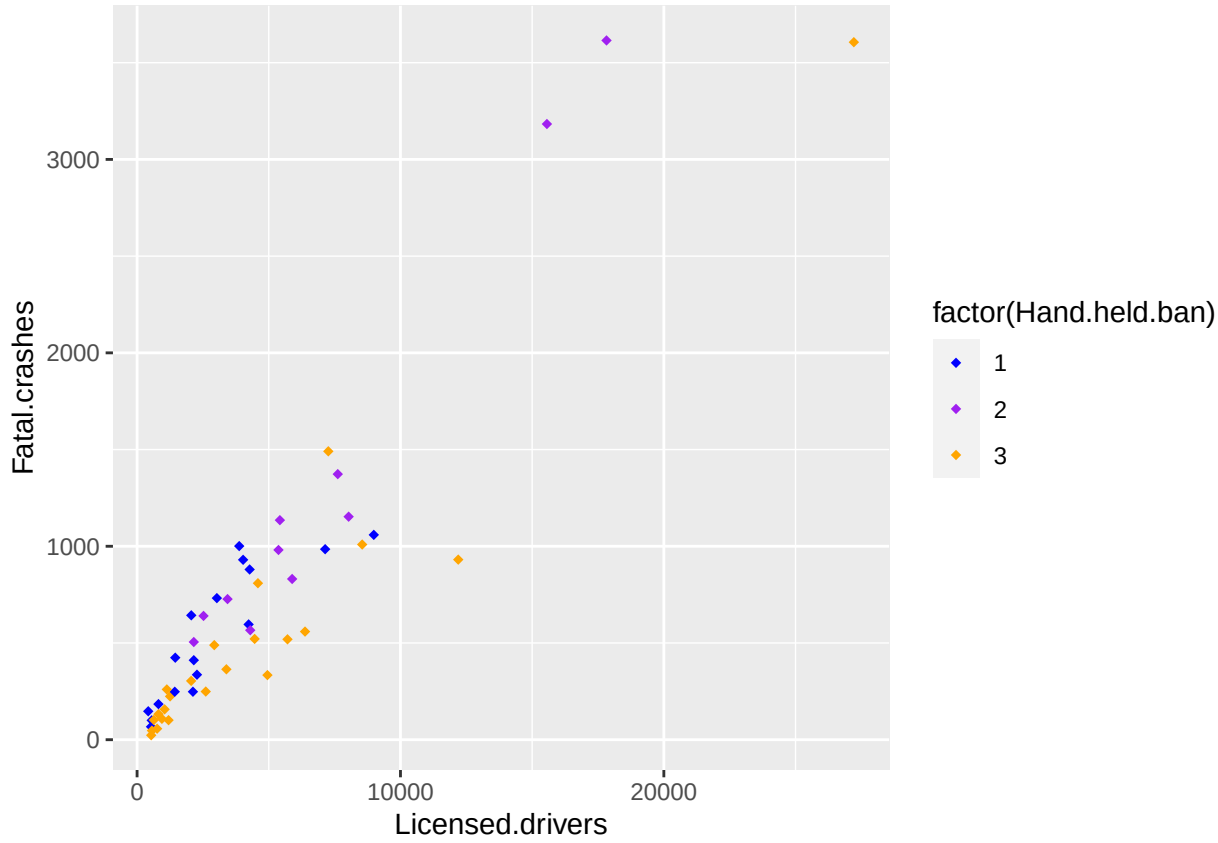
a)

```
data1<-read.csv("C:/Users/carlo/Desktop/TERM5/STAT 3080/state crashes.csv")
library(ggplot2)
print(ggplot(data1,mapping=aes(x = Licensed.drivers, y =Fatal.crashes))+geom_point(shape
```



b)

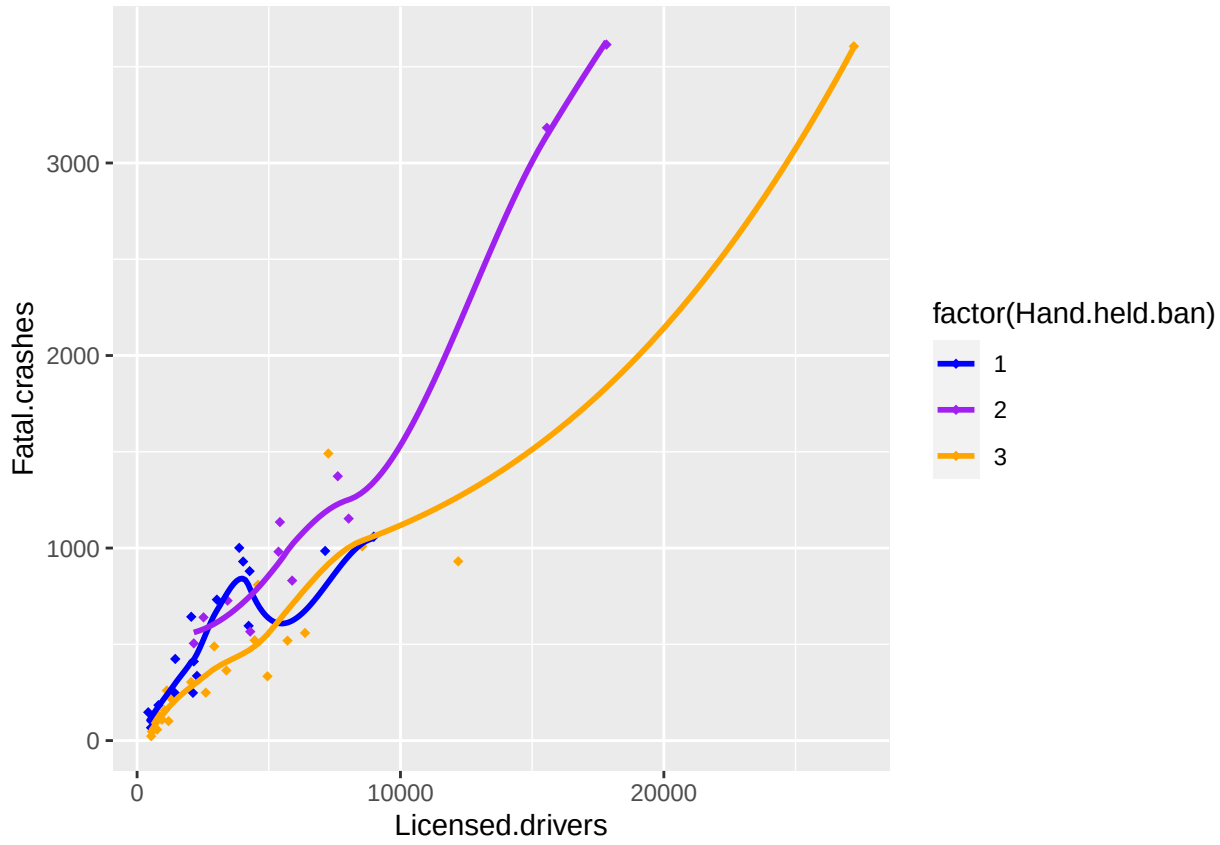
```
data1<-read.csv("state crashes.csv")
cols<-c("1"="blue","2"="purple","3"="orange")
ggplot(data1,aes(Licensed.drivers,Fatal.crashes, color = factor(Hand.held.ban)))+geom_po
```



c)

```
data1<-read.csv("state crashes.csv")
cols<-c("1"="blue","2"="purple","3"="orange")
ggplot(data1,aes(Licensed.drivers,Fatal.crashes, color = factor(Hand.held.ban)))+geom_po
```

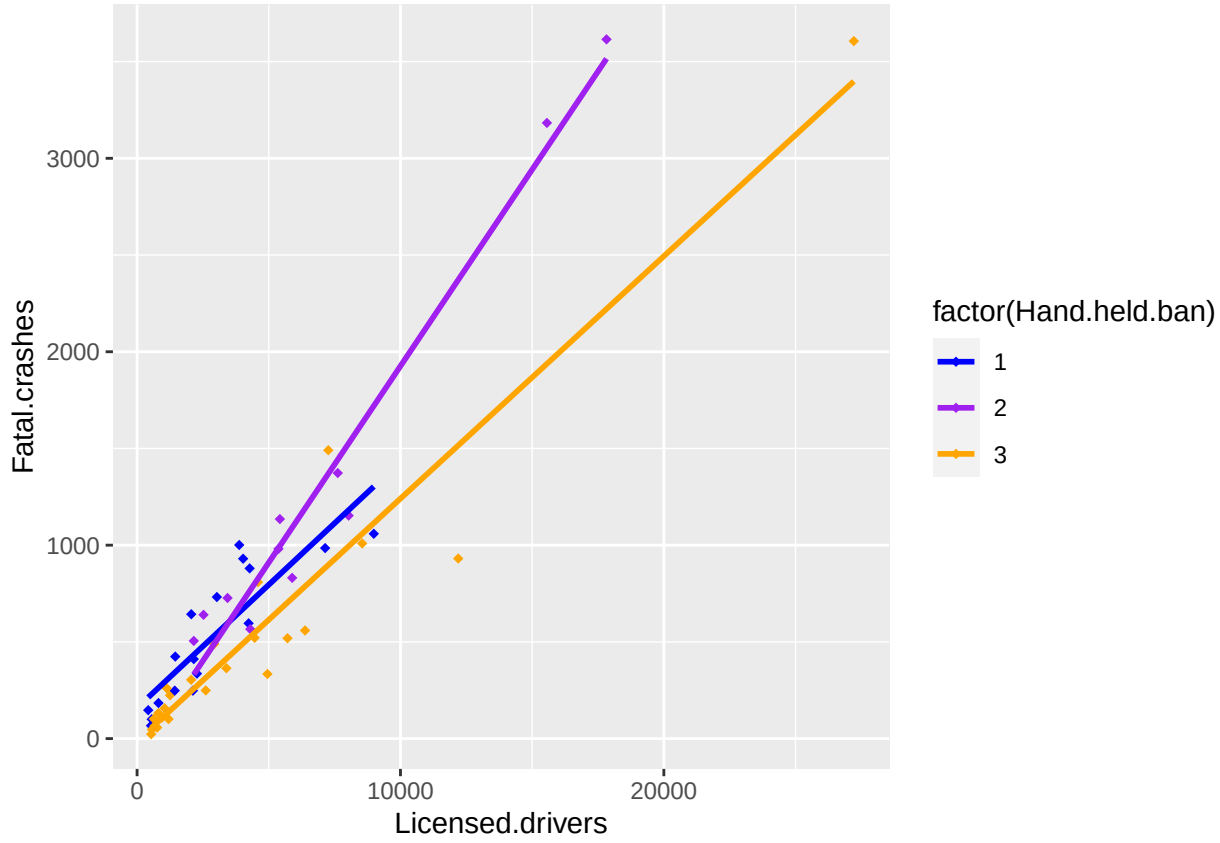
```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



d)

```
data1<-read.csv("state crashes.csv")
cols<-c("1"="blue","2"="purple","3"="orange")
ggplot(data1,aes(Licensed.drivers,Fatal.crashes, color = factor(Hand.held.ban)))+geom_po
```

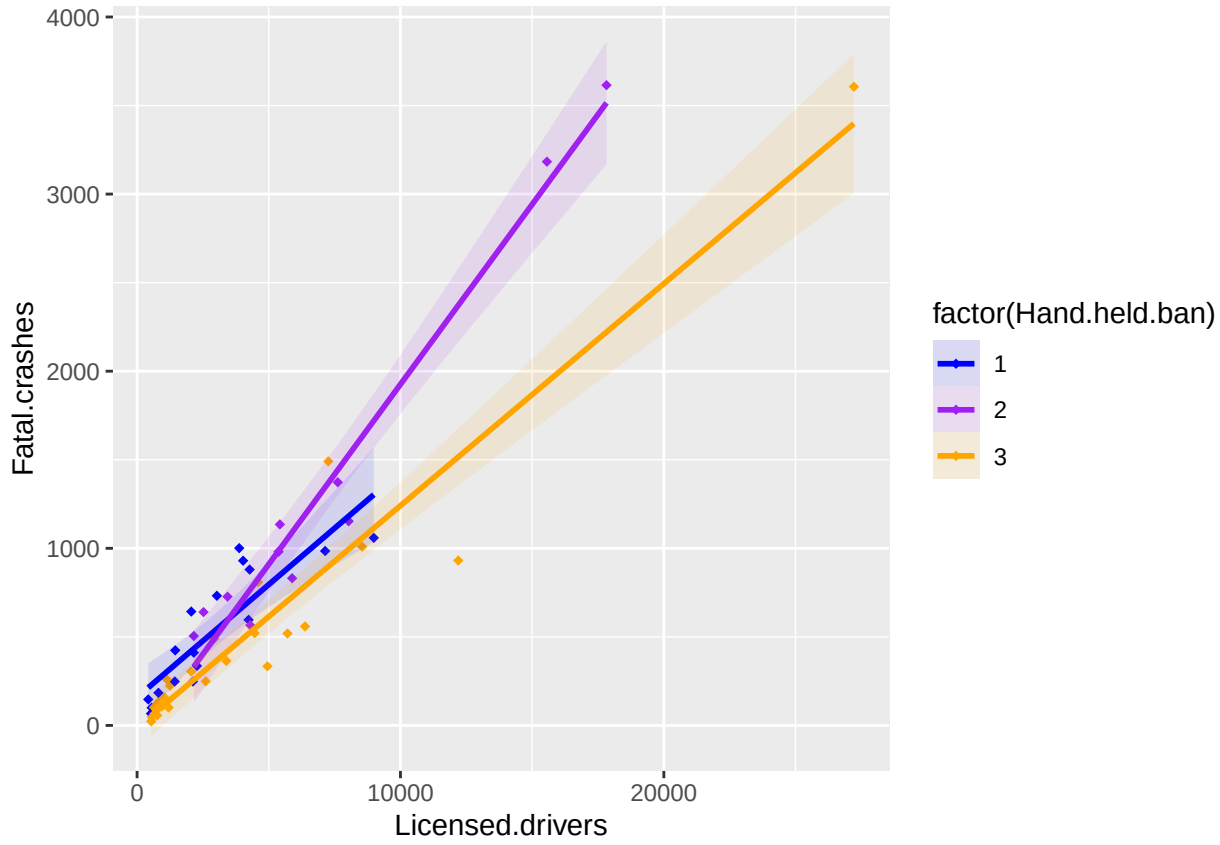
```
## `geom_smooth()` using formula 'y ~ x'
```



e)

```
data1<-read.csv("state crashes.csv")
cols<-c("1"="blue","2"="purple","3"="orange")
ggplot(data1,aes(Licensed.drivers,Fatal.crashes,fill = factor(Hand.held.ban) ,color = fa
```

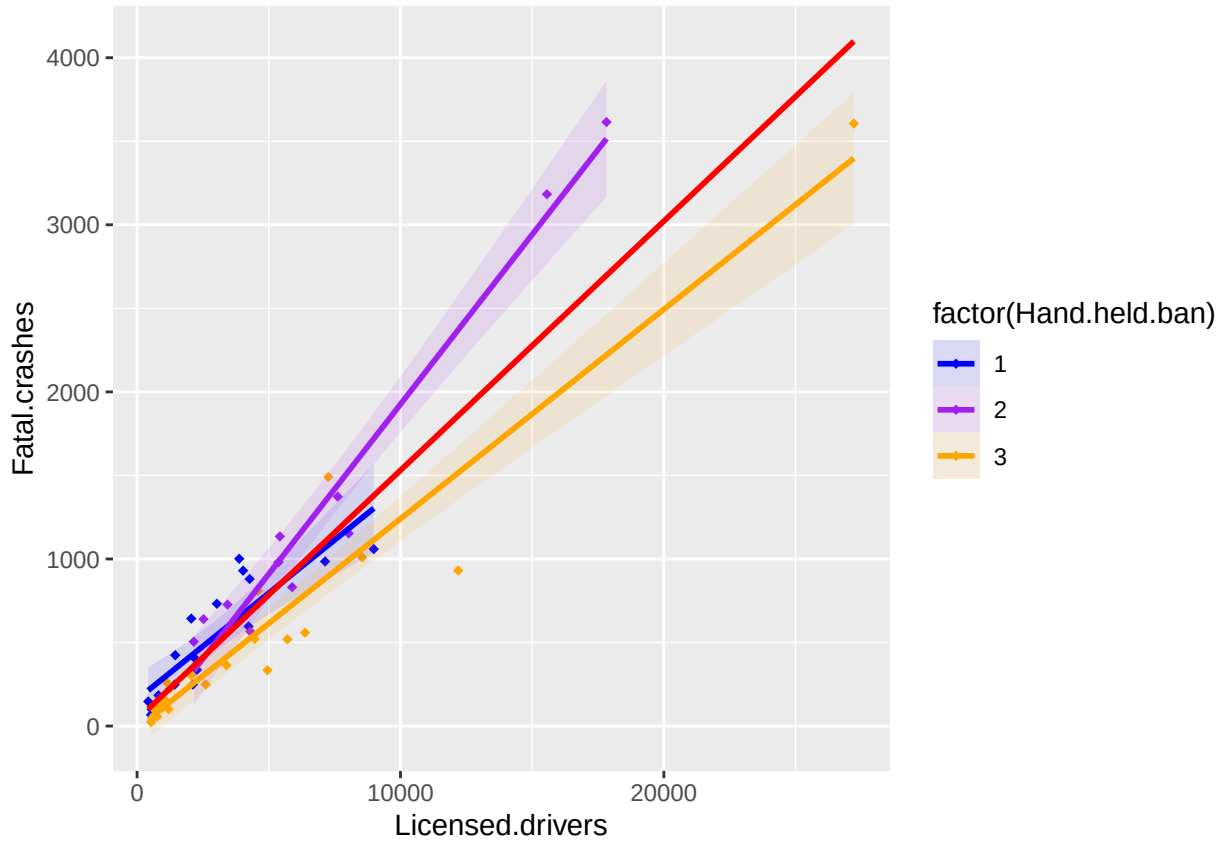
```
## `geom_smooth()` using formula 'y ~ x'
```



f)

```
data1<-read.csv("state crashes.csv")
cols<-c("1"="blue","2"="purple","3"="orange")
ggplot(data1,aes(Licensed.drivers,Fatal.crashes,fill = factor(Hand.held.ban) ,color = fa
```

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



g)

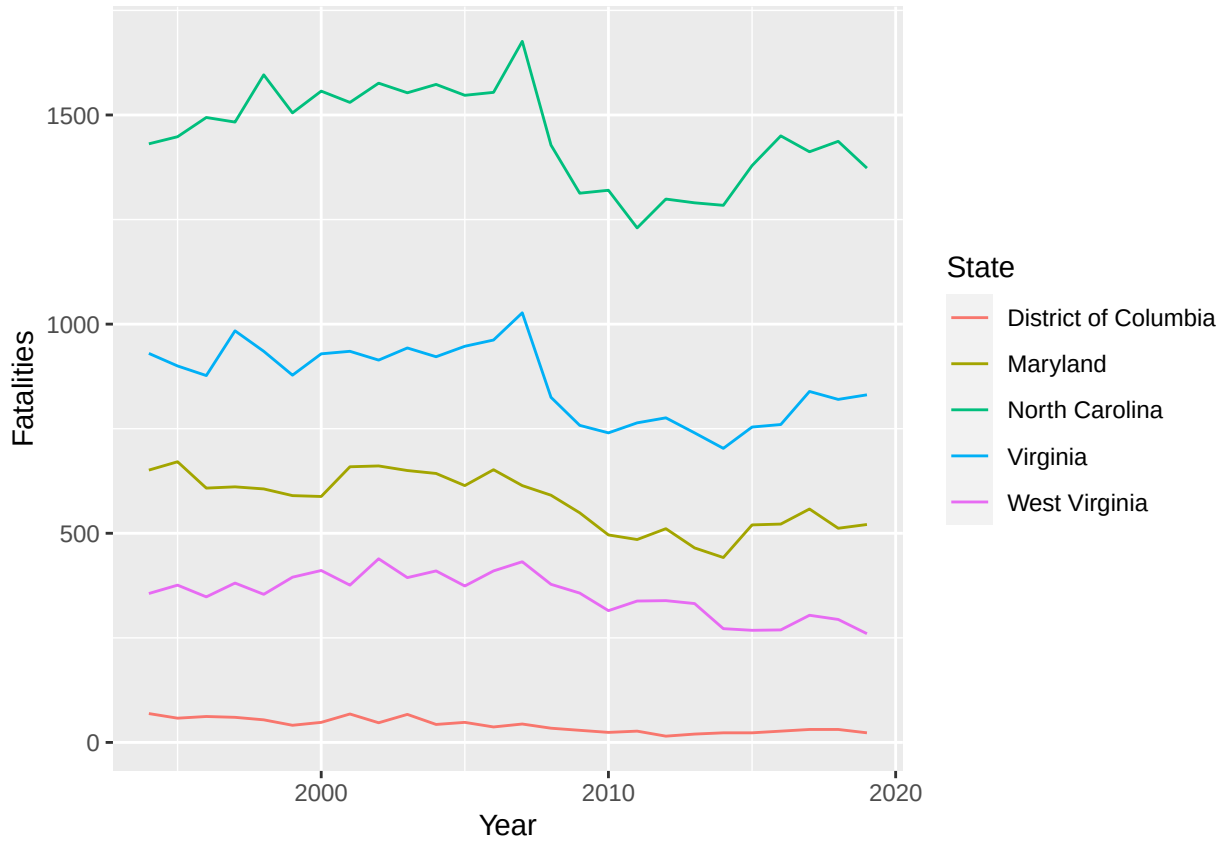
Conclusions: From the graph we can see a positive regression trend where in general, the states with more Licensed drivers tend to have more fatal car crashes.

Another interesting conclusion to point out is that in the States where hand-held electronic devices are banned the regression line of the data seems to be less so than in states that do not ban them (comparing 3(orange) with 1(blue)). Indicating that there are more fatal Crashes per Licensed drivers in (1) states than in (3) states

States where some drivers have a hand held electronic ban (purple)(2) seem to have less fatal crashes in states with less licenced drivers. However we look at states with more drivers that also have the some drivers hand held electronic bans(2) there seems to be a higher number of fatal crashes. ## Problem 2 ### a)

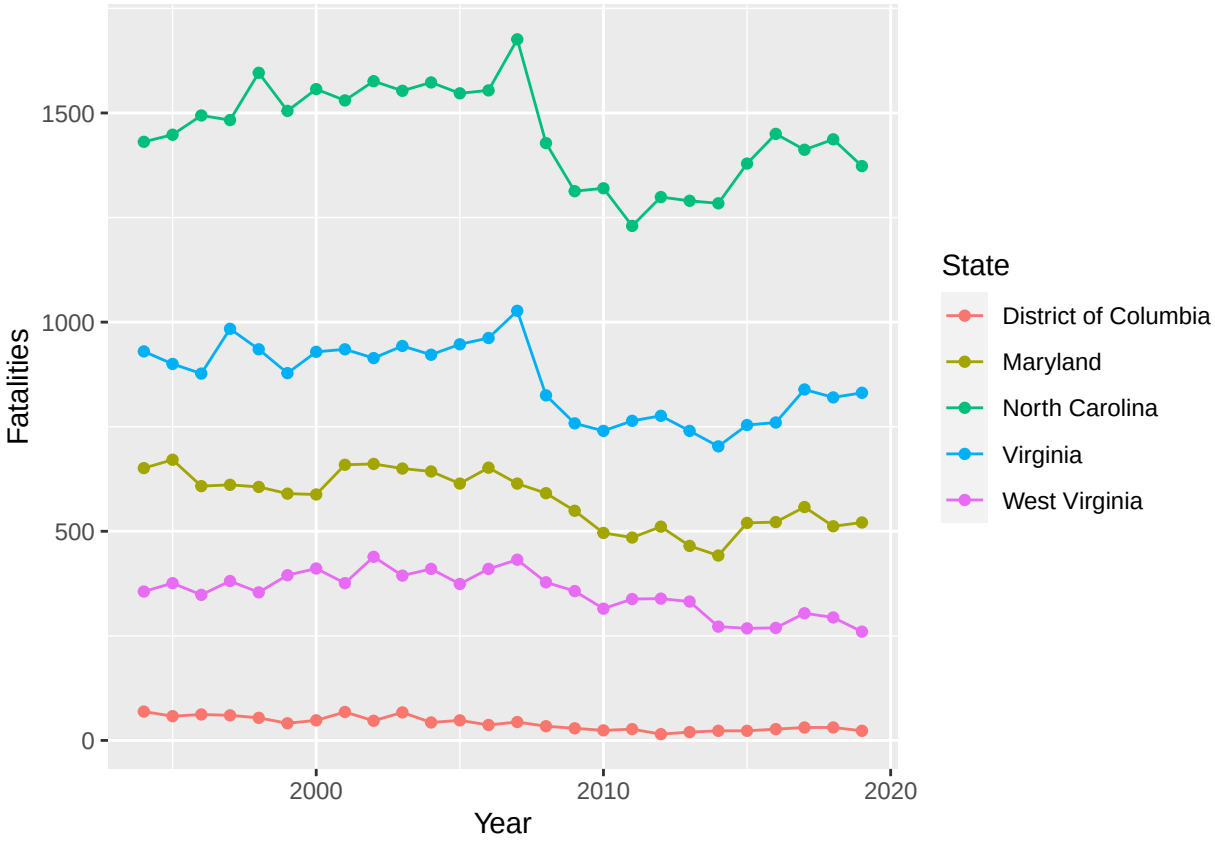
```
data2 <- read.csv("C:/Users/carlo/Desktop/TERM5/STAT 3080/fatalities.csv")

ggplot(data2,mapping=aes(x =Year,y=Fatalities,color=State))+geom_line()
```



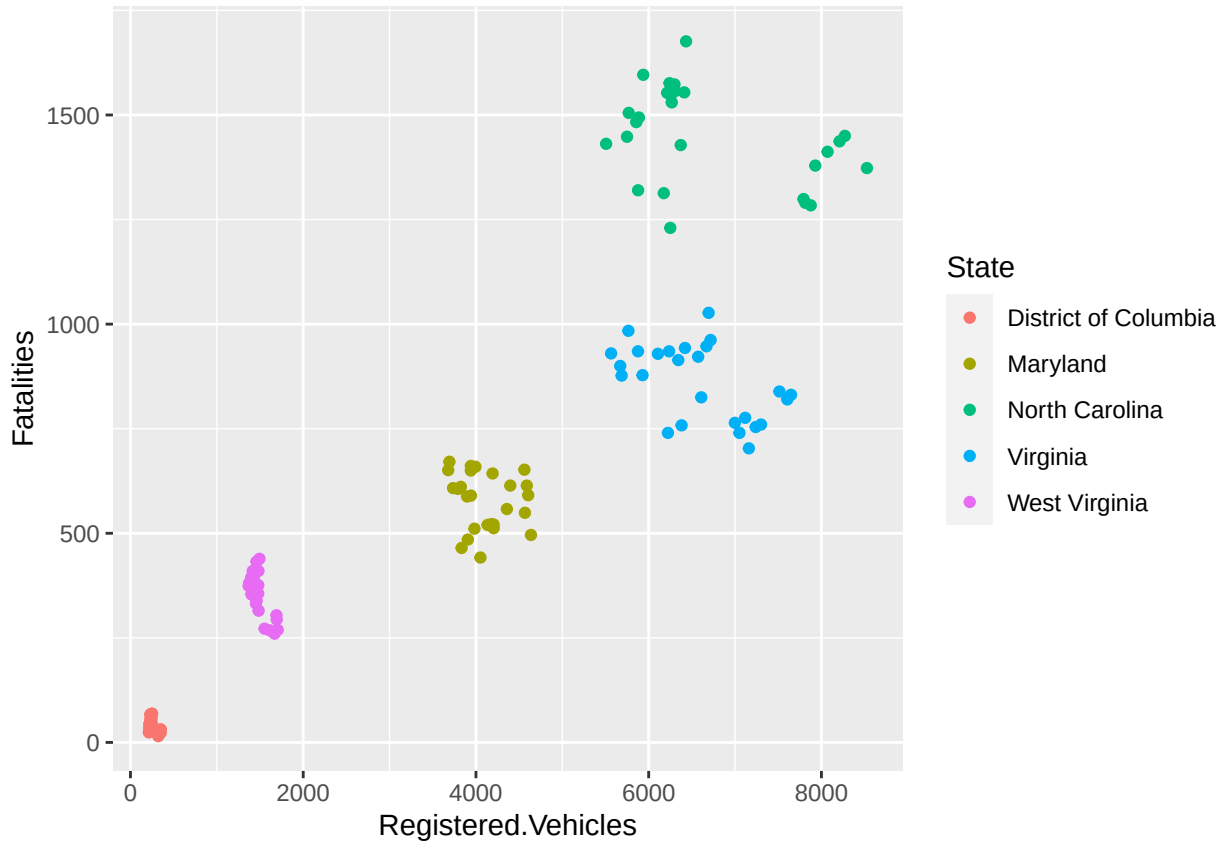
b) From the plot in part a) we can see that all of the states generally show the same trend. Also interesting to note that North Carolina and Virginia show a clear jump in the Number of fatalities in 2007. With North Carolina showing the most Fatalities and DC with the least. ### c)

```
ggplot(data2,mapping=aes(x =Year,y=Fatalities,color=State))+geom_line()+geom_point()
```



d) Plot created in part c) gives us the individual observations for each year by each state so we can clearly see how many observations we have in each state and how many fatalities for each year. ### e)

```
ggplot(data2,mapping=aes(x =Registered.Vehicles,y=Fatalities,color=State))+geom_point()+
```

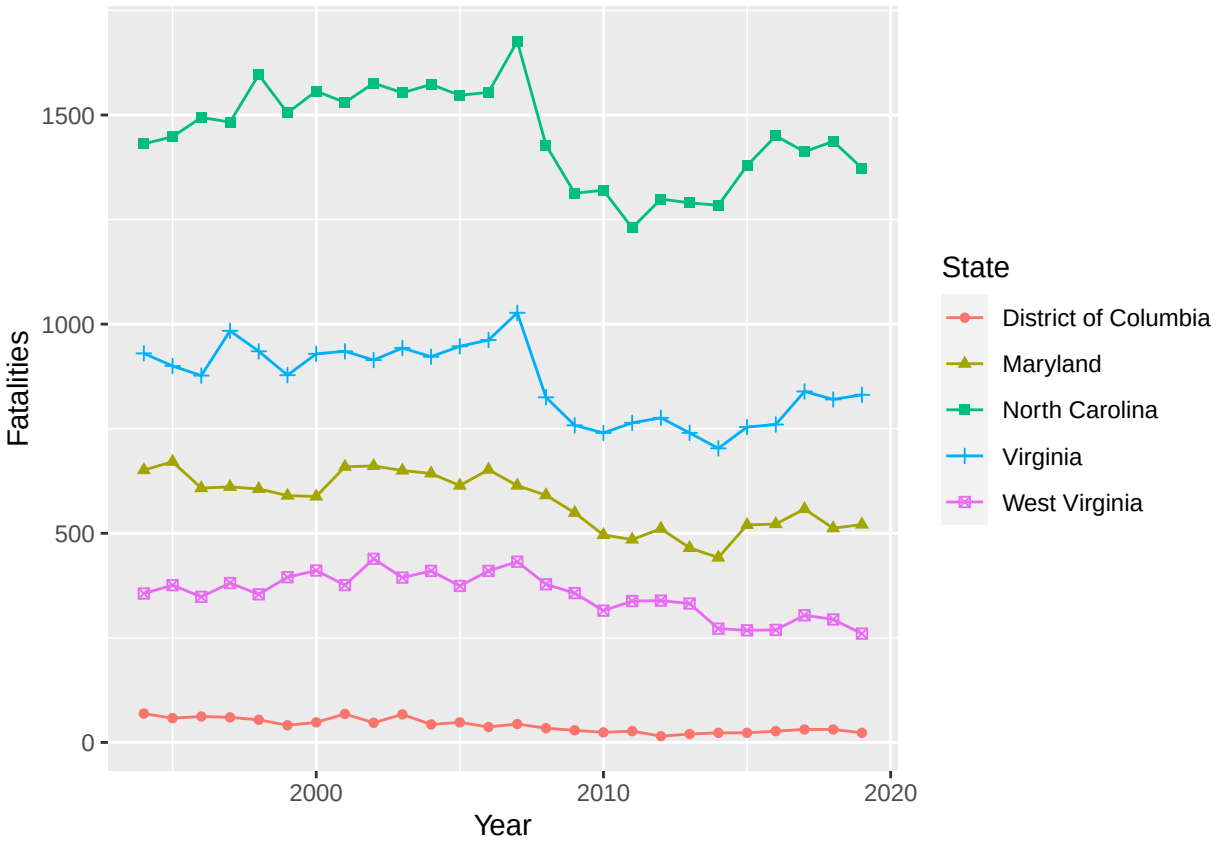


f) The plot in part e) shows that even though North Carolina and Virginia have a very similar ammount of Registered Vehicles over the years there are many more fatalities in north Carolina. So the trend does not stay constant like we said in part b) there seems to be more fatalities in North carolina than in virginia despite having a similar number of registered vehicles. ### g)

```
ggplot(data2,mapping=aes(x =Vehicle.Miles,y=Fatalities,color=State))+geom_point()
```

h) In plot e North Carolina and Virginia show similar registered vehicles with more fatalities in north carolina and in plot g northcarolina has more vehicle miles and more fatalities showing a more positive correlation. The main cause of this is most likely that even though there are similar ammount of registered vehicles in north carolina and virginia people tend to drive a lot more in north carolina and when your poplulation is driving more on average there will be greater chance of car acciedents and hence fatalities. ### i)

```
ggplot(data2,mapping=aes(x =Year,y=Fatalities,color=State))+geom_line()+geom_point(aes(s
```



j) Not necessarily because we already have the points separated by y Values and by colors so it is already pretty clear.