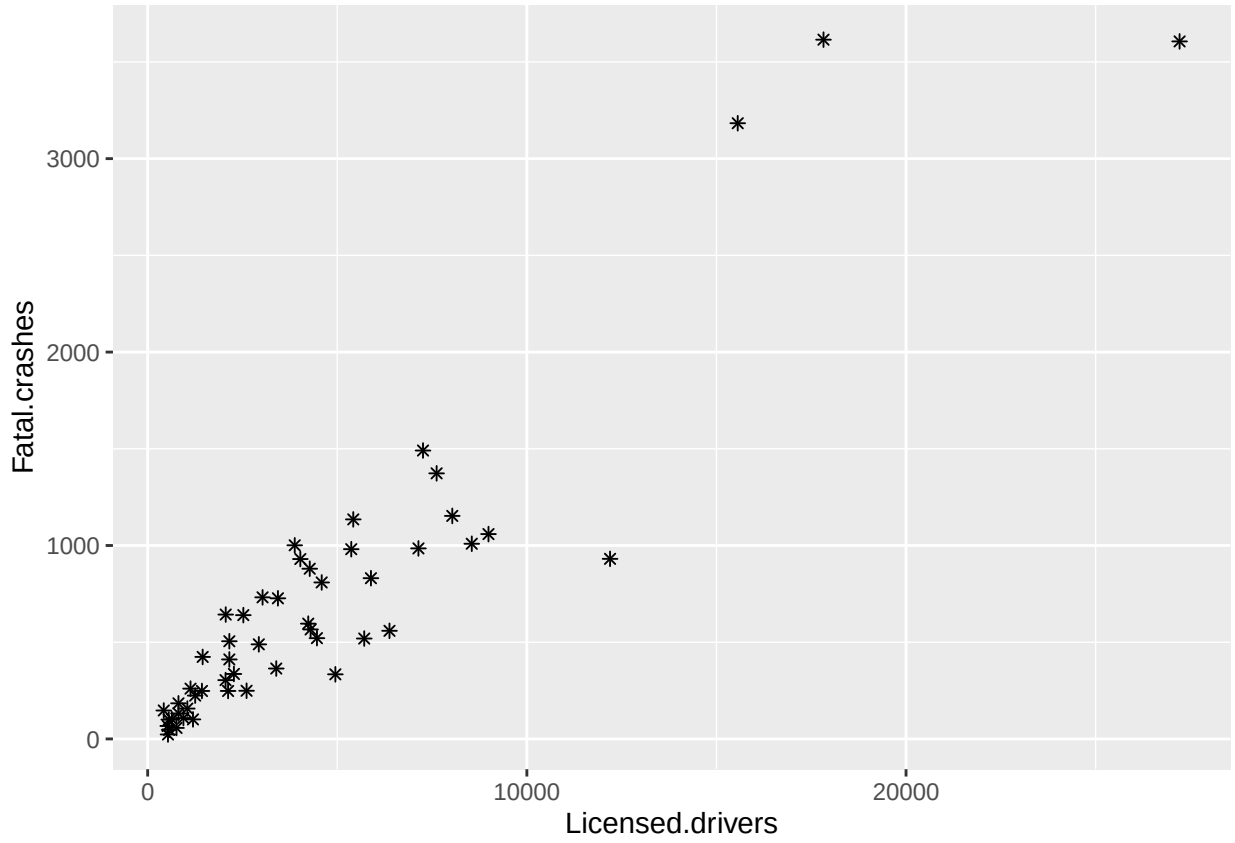
# Homework 5

Will Scheib

## Problem 1

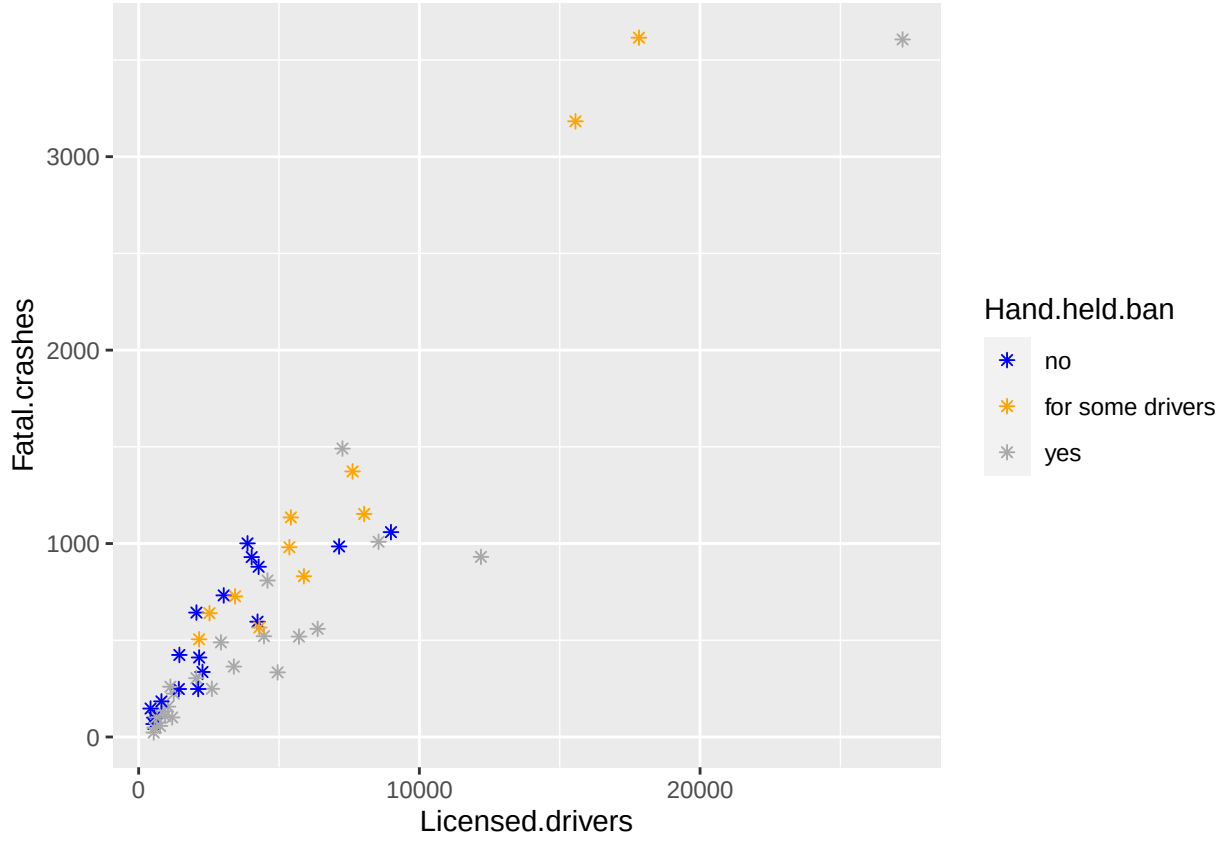
### Part a

```
state_crashes <- read.csv("data/state crashes.csv")

ggplot(state_crashes, aes(x=Licensed.drivers, y=Fatal.crashes)) +
  geom_point(shape=8)
```



### Part b

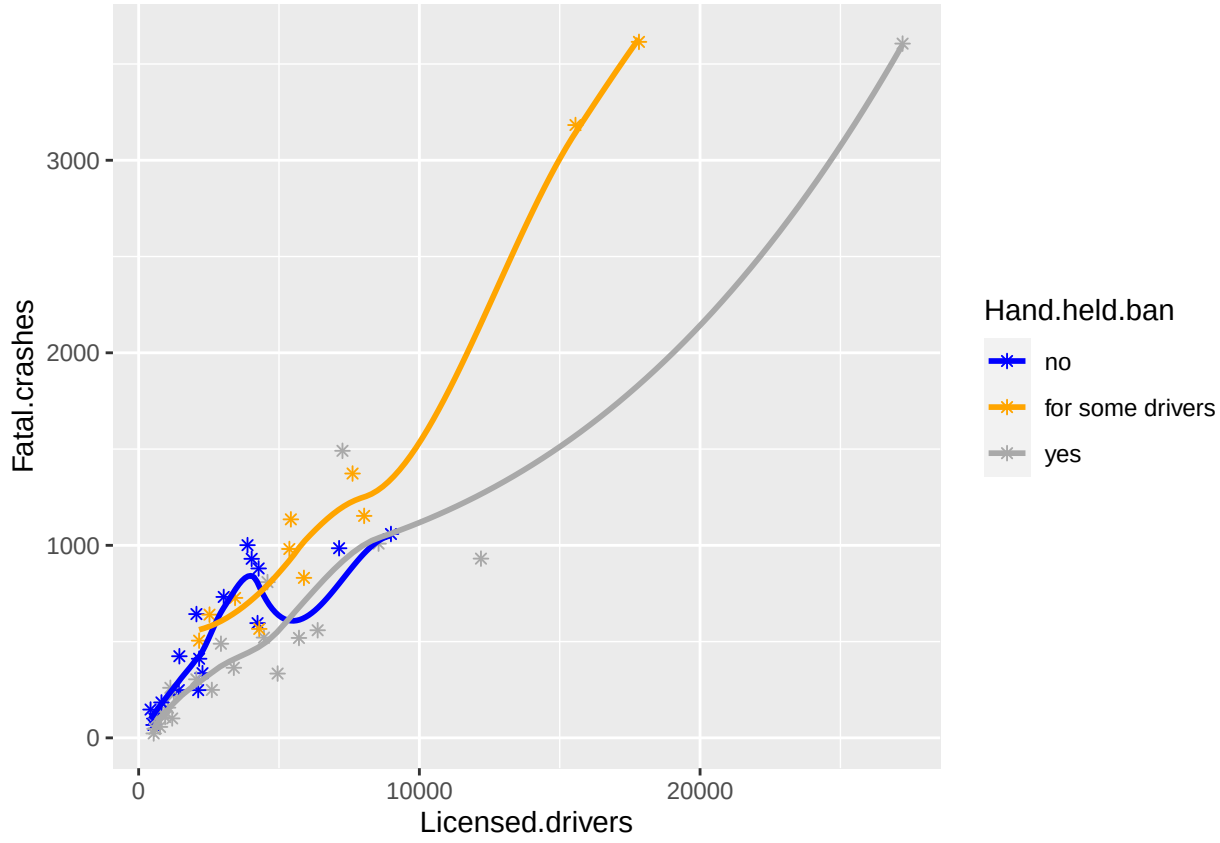
```
state_crashes$Hand.held.ban <- factor(
  state_crashes$Hand.held.ban,
  labels=c("no", "for some drivers", "yes")
)

ggplot(
    state_crashes,
    aes(x=Licensed.drivers, y=Fatal.crashes, color=Hand.held.ban)
) +
  geom_point(shape=8) +
  scale_color_manual(values=c("blue", "orange", "dark gray"))
```



### Part c

```
ggplot(
    state_crashes,
    aes(x=Licensed.drivers, y=Fatal.crashes, color=Hand.held.ban)
) +
  geom_point(shape=8) +
  scale_color_manual(values=c("blue", "orange", "dark gray")) +
  geom_smooth(se=FALSE)
```

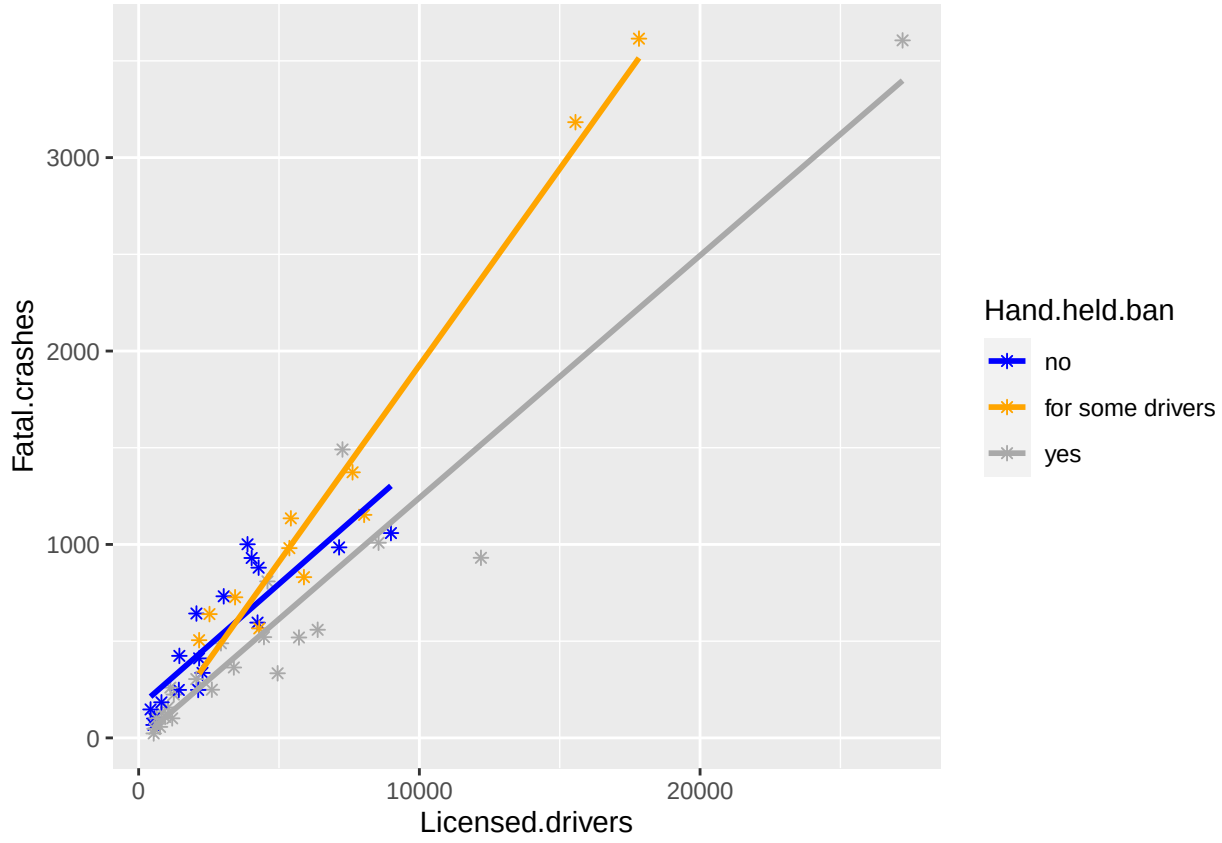
```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



### Part d

```
ggplot(
    state_crashes,
    aes(x=Licensed.drivers, y=Fatal.crashes, color=Hand.held.ban)
) +
  geom_point(shape=8) +
  scale_color_manual(values=c("blue", "orange", "dark gray")) +
  geom_smooth(method=lm, se=FALSE)
```

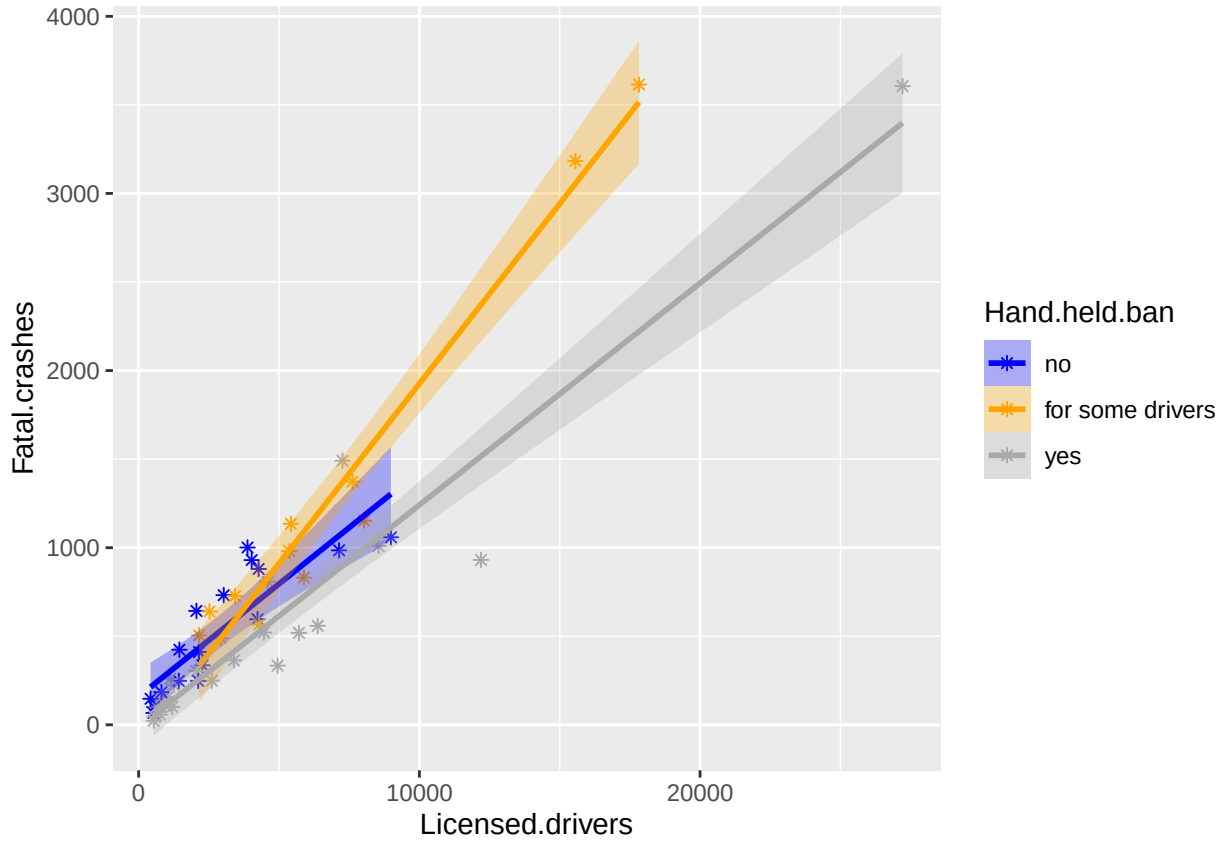
```
## `geom_smooth()` using formula 'y ~ x'
```



### Part e

```
ggplot(
    state_crashes,
    aes(
      x=Licensed.drivers, y=Fatal.crashes,
      color=Hand.held.ban, fill=Hand.held.ban
    )
) +
  geom_point(shape=8) +
  scale_color_manual(values=c("blue", "orange", "dark gray")) +
  scale_fill_manual(values=c("blue", "orange", "dark gray")) +
  geom_smooth(method=lm, alpha=0.3)
```

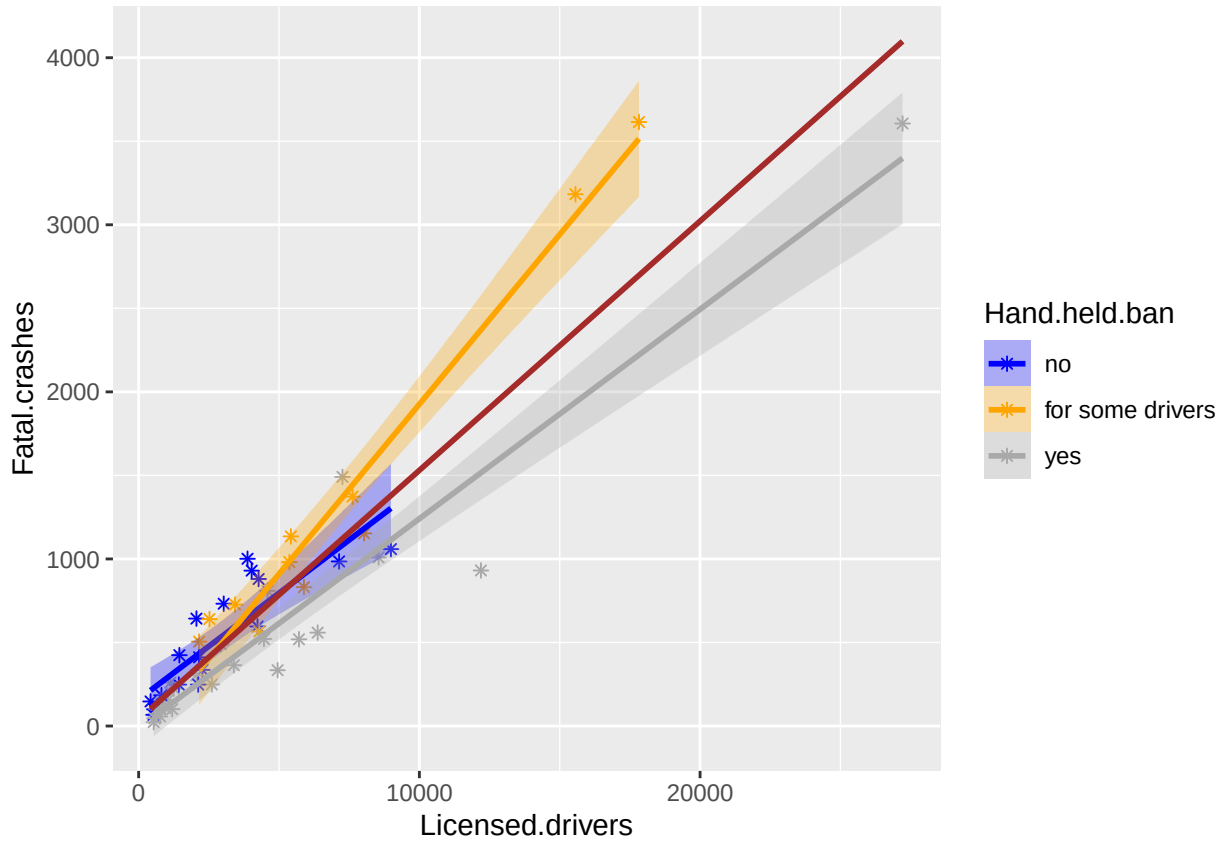
```
## `geom_smooth()` using formula 'y ~ x'
```



### Part f

```
ggplot(
    state_crashes,
    aes(
      x=Licensed.drivers, y=Fatal.crashes
    )
) +
  geom_point(aes(color=Hand.held.ban), shape=8) +
  scale_color_manual(values=c("blue", "orange", "dark gray")) +
  scale_fill_manual(values=c("blue", "orange", "dark gray")) +
  geom_smooth(
    aes(color=Hand.held.ban, fill=Hand.held.ban),
    method=lm,
    alpha=0.3
  ) +
  geom_smooth(color="brown", method=lm, se=FALSE)
```

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



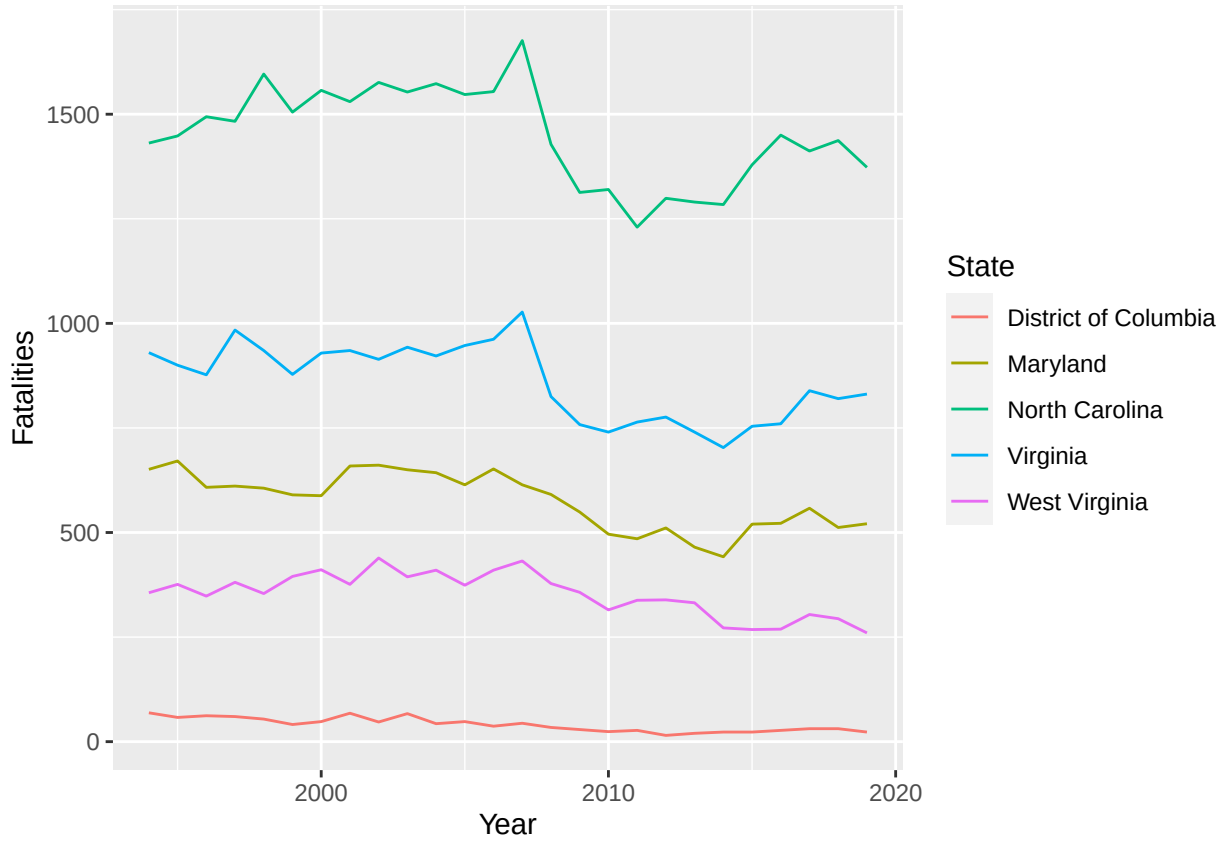
### Part g

As there are more licensed drivers, it is likely that there will be more fatal crashes. Also, states in which there are laws banning handheld devices for some drivers tend to have more accidents when there are many licensed drivers.

## Problem 2

### Part a

```
fatalities <- read.csv("data/fatalities.csv")

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

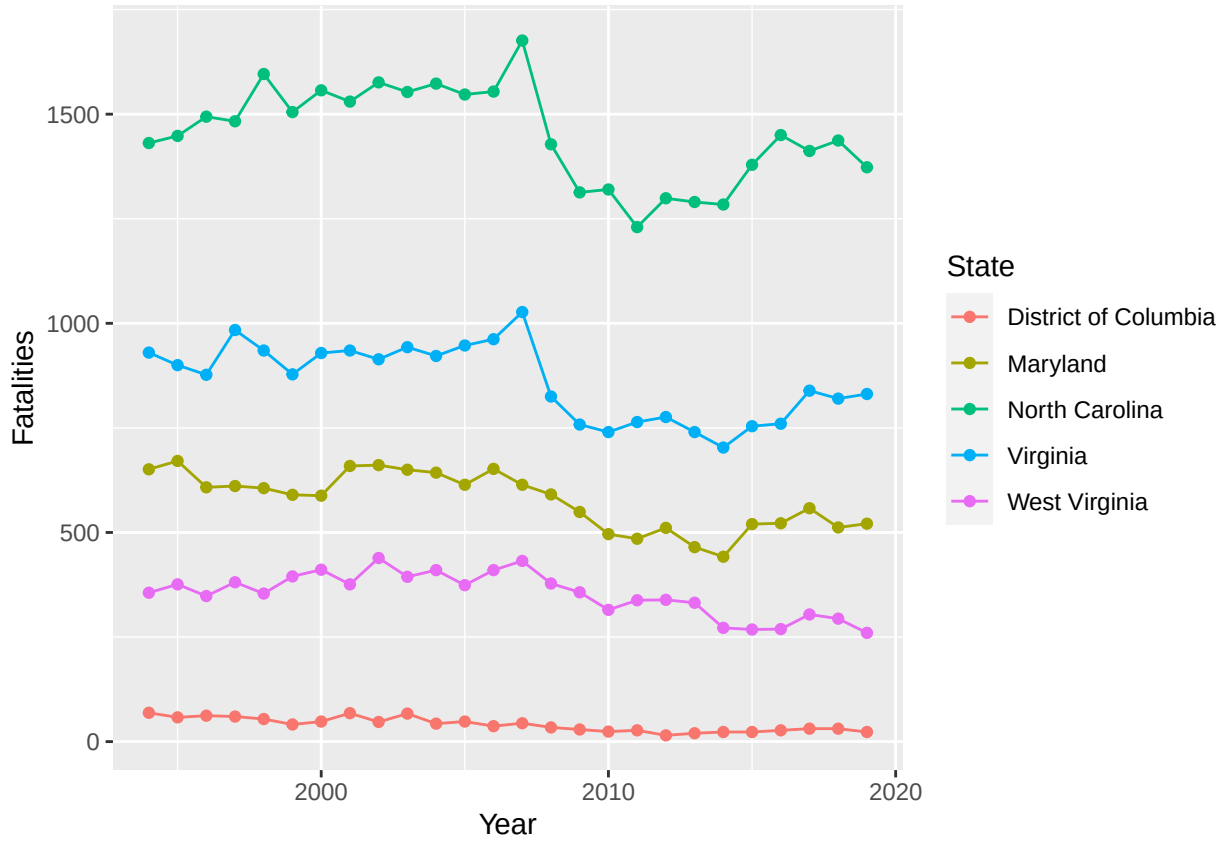



### Part b

The states do have similar trends, with a spike around 2007 and a general downward trend.

### Part c

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

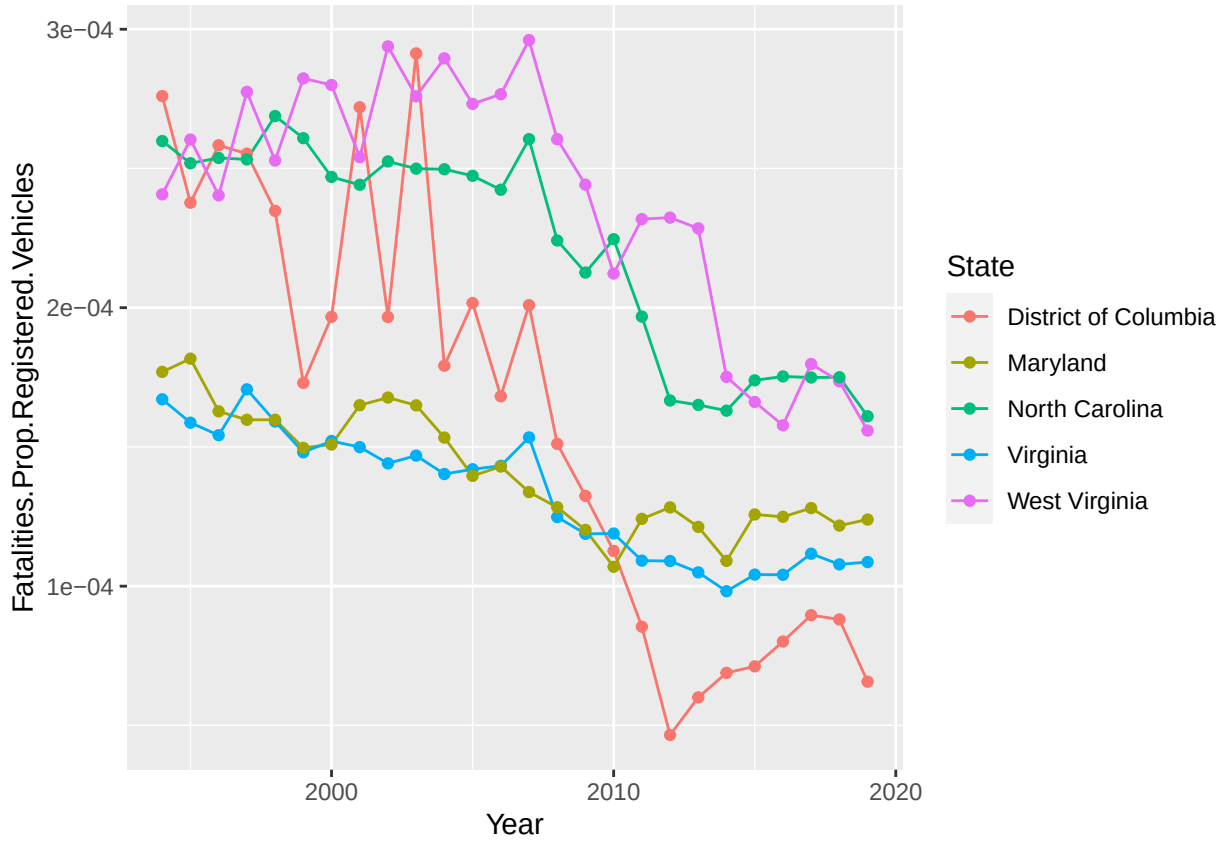



### Part d

It is easier to distinguish particular values in addition to general trends.

### Part e

```
fatalities <- fatalities %>%
  mutate(Fatalities.Prop.Registered.Vehicles=Fatalities/(Registered.Vehicles*10^3))

ggplot(fatalities, aes(x=Year, y=Fatalities.Prop.Registered.Vehicles, color=State)) +
  geom_line() +
  geom_point()
```




### Part f

Yes, because fatalities in DC seem to be dropping faster than elsewhere.

### Part g

```
fatalities <- fatalities %>%
  mutate(Fatalities.Prop.Miles.Driven=Fatalities/(Vehicle.Miles*10^9))

ggplot(fatalities, aes(x=Year, y=Fatalities.Prop.Miles.Driven, color=State)) +
  geom_line() +
  geom_point()
```


### Part h

The lines are all closer together, maybe because fatal accidents are better related to miles driven than to number of registered vehicles.

### Part i

```
fatalities <- fatalities %>%
  mutate(Fatalities.Prop.Miles.Driven=Fatalities/(Vehicle.Miles*10^9))

ggplot(fatalities, aes(x=Year, y=Fatalities.Prop.Miles.Driven, color=State)) +
  geom_line() +
  geom_point(aes(shape=State))
```


### Part j

No, because the line color makes it pretty easy to distinguish between states already.