# Difference-in-differences

In 1977, State A passed new legislation that required that all drivers and passengers wear seatbelts when driving. Their neighbor, State B, did not pass a similar law.

You have access to observational (i.e. non-experimental) data from both of these state agencies:

| Variable name | Description |
|---|---|
| `ID` | Unique row id |
| `state` | State A or State B |
| `year` | Year of observation |
| `death_rate` | Deaths per 100 crashes |

Your colleague attempted to measure the causal effect of this seatbelt law on crash fatalities using a difference-in-differences approach. They conducted some statistical analysis in R, but they forgot to interpret anything in the document, and now they've moved to a different office!

**Given the information provided below, interpret the results from this analysis, as well as any assumption checks or tests your colleague included. Did this law have an effect on the fatality rate? How much? Is it significant?**

```
library(tidyverse)
library(broom)

seatbelts <- read_csv("seatbelts.csv") %>%
  mutate(after_1977 = year > 1977)
```

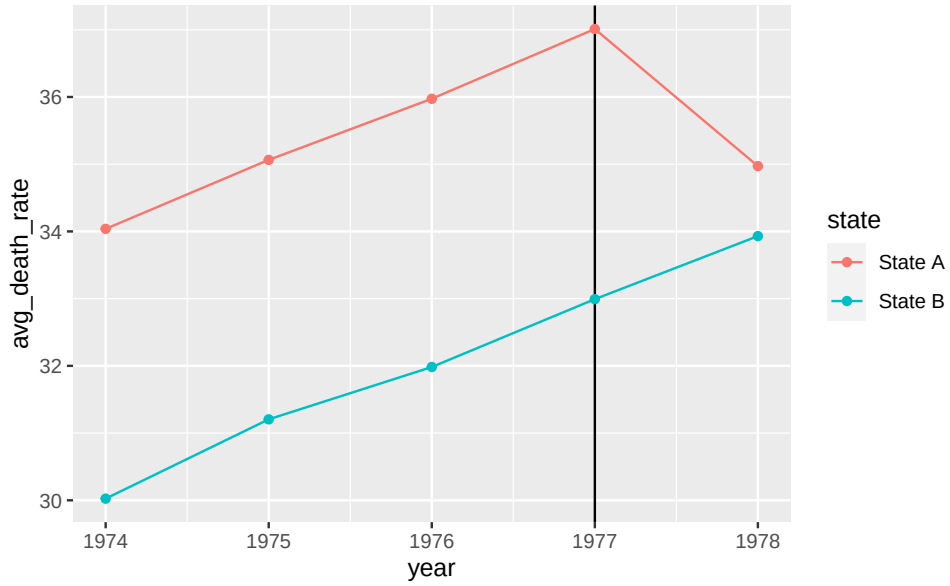
```
# See first few rows of data
head(seatbelts, n = 10)
```

| ID | state | year | death_rate | after_1977 |
|---|---|---|---|---|
| 0001 | State B | 1974 | 24.5 | FALSE |
| 0001 | State B | 1975 | 20.0 | FALSE |
| 0001 | State B | 1976 | 16.5 | FALSE |
| 0001 | State B | 1977 | 11.2 | FALSE |
| 0001 | State B | 1978 | 21.2 | TRUE |
| 0002 | State B | 1974 | 32.2 | FALSE |
| 0002 | State B | 1975 | 26.9 | FALSE |
| 0002 | State B | 1976 | 15.7 | FALSE |
| 0002 | State B | 1977 | 12.1 | FALSE |
| 0002 | State B | 1978 | 13.9 | TRUE |

↓ 1: What's going on here? ↓

```
seatbelts_by_year <- seatbelts %>%
  group_by(year, state) %>%
  summarize(avg_death_rate = mean(death_rate))

ggplot(seatbelts_by_year,
       aes(x = year, y = avg_death_rate, color = state)) +
  geom_vline(xintercept = 1977) +
  geom_line() +
  geom_point()
```




↓ 2: What's going on here? ↓

```
model1 <- lm(death_rate ~ after_1977 + state +
               after_1977 * state,
             data = seatbelts)
tidy(model1)
```

| term | estimate | std.error | statistic | p.value |
|---|---|---|---|---|
| (Intercept) | 35.52 | 0.097 | 366.00 | 0.000 |
| after_1977TRUE | -0.55 | 0.217 | -2.53 | 0.011 |
| stateState B | -3.97 | 0.137 | -29.01 | 0.000 |
| after_1977TRUE:stateState B | 2.93 | 0.306 | 9.57 | 0.000 |

```
glance(model1)
```

| r.squared | adj.r.squared | sigma | statistic | p.value | df | logLik | AIC | BIC | deviance | df.residual | nobs |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.034 | 0.034 | 9.68 | 297 | 0 | 3 | -92218 | 184446 | 184487 | 2341273 | 24996 | 25000 |