# Estimating Causal Effects using Difference in Difference

Code

A particular state raised its cap on weekly earnings that were covered by worker's compensation. We want to know if this new policy caused workers to spend more time unemployed

Hide

```
library(tidyverse)  # ggplot(), %>%, mutate(), and friends

injury <- read_csv("https://raw.githubusercontent.com/vntkumar8/musical-spoon/main/injury_data.csv")
```

- duration (main response variable): Duration of unemployment benefits, measured in weeks
- **log_duration**: Logged version of duratation (log(duration))
- after_1980: Indicator variable marking if the observation happened before (0) or after (1) the policy change in 1980. This is our time (or **before/after** variable)
- highearn: Indicator variable marking if the observation is a low (0) or high (1) earner. This is our group (or **treatment/control**) variable

# Exploratory data analysis

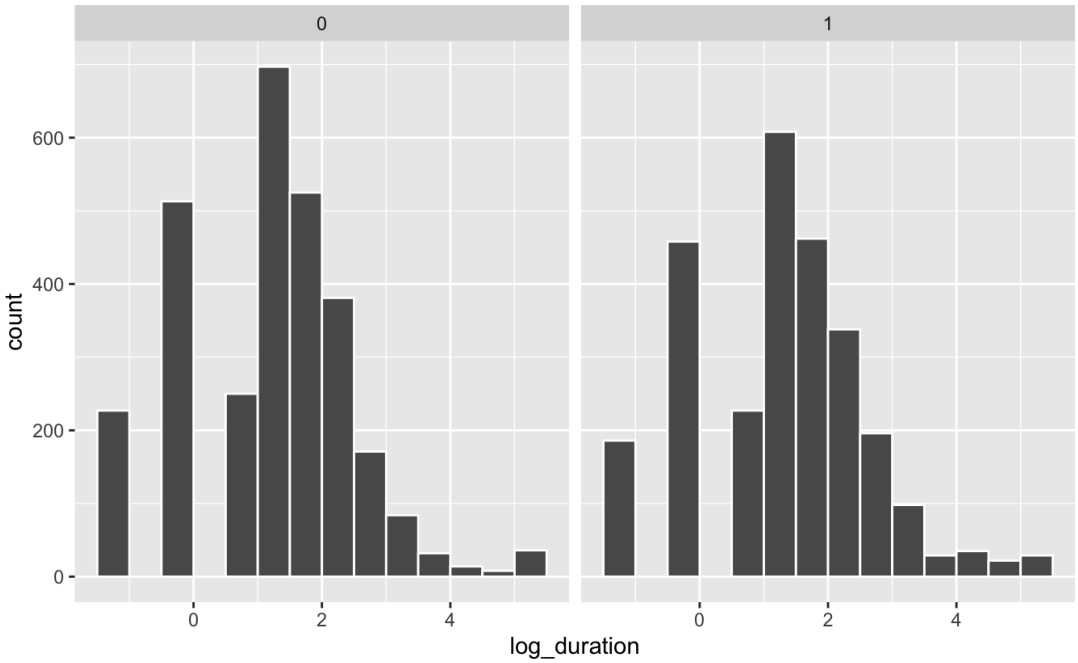
Look at the distribution of unemployment benefits across high and low earners (our control and treatment groups)

Hide

```
ggplot(data = injury, aes(x = duration)) +
  # binwidth = 8 makes each column represent 2 months (8 weeks)
  # boundary = 0 make it so the 0-8 bar starts at 0 and isn't -4 to 4
  geom_histogram(binwidth = 8, color = "white", boundary = 0) +
  facet_wrap(vars(highearn))
```

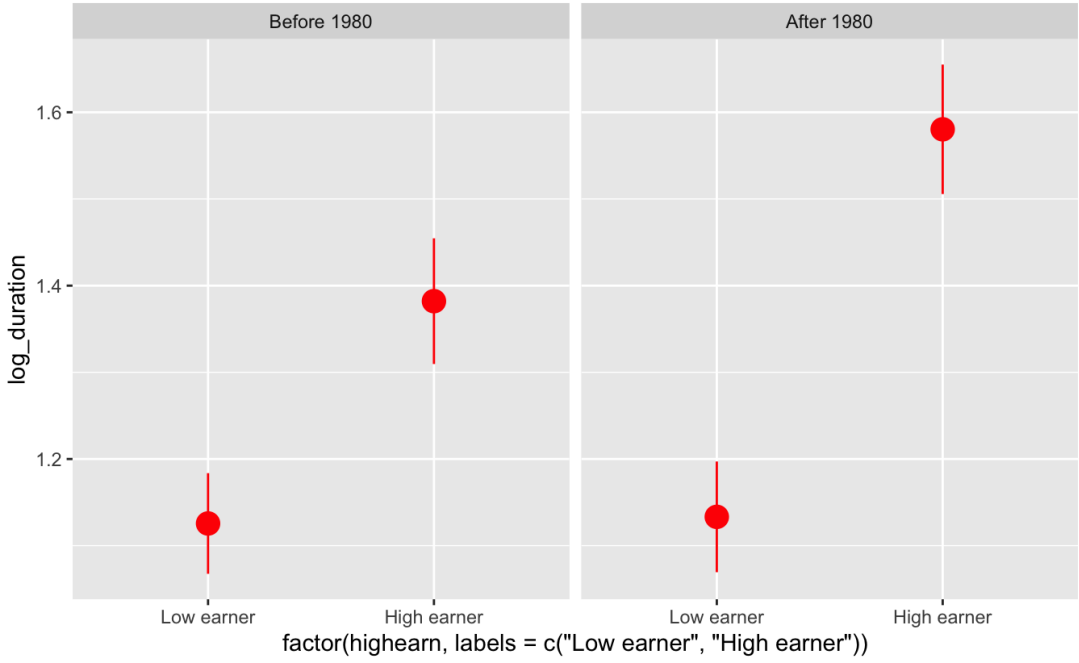
Hide

```
ggplot(data = injury, mapping = aes(x = log_duration)) +
  geom_histogram(binwidth = 0.5, color = "white", boundary = 0) +
  facet_wrap(vars(after_1980))
```




Hide

```
ggplot(injury, aes(x = factor(highearn,labels = c("Low earner", "High earner")), y = log_duration)) +
  stat_summary(geom = "pointrange", size = 1, color = "red",
               fun.data = "mean_se", fun.args = list(mult = 1.96)) +
  facet_wrap(vars(factor(after_1980,labels = c("Before 1980", "After 1980"))))
```




Hide

```
NA
NA
```

# Diff-in-Diff

| | Before 1980 | After 1980 | Delta |
|---|---|---|---|
| Low Earners | A | B | B-A |
| High Earners | C | D | D-C |
| Delta | C-A | D-B | (D-C)-(B-A) |

# Regression Analysis

$$\log(\text{duration}) = \beta_0 + \beta_1\ \text{highearn} + \beta_2\ \text{after\_1980} + \beta_3\ (\text{highearn} \times \text{after\_1980}) + \epsilon$$

Hide

```
model <- lm(log_duration~highearn+after_1980+highearn*after_1980,data=injury)
summary(model)
```

```

Call:
lm(formula = log_duration ~ highearn + after_1980 + highearn *
    after_1980, data = injury)

Residuals:
    Min      1Q  Median      3Q     Max
-2.9666 -0.8872  0.0042  0.8126  4.0784

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)       1.125615   0.030737  36.621  < 2e-16 ***
highearn          0.256479   0.047446   5.406 6.72e-08 ***
after_1980        0.007657   0.044717   0.171  0.86404
highearn:after_1980 0.190601   0.068509   2.782  0.00542 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.269 on 5622 degrees of freedom
Multiple R-squared:  0.02066,   Adjusted R-squared:  0.02014
F-statistic: 39.54 on 3 and 5622 DF,  p-value: < 2.2e-16
```

## Interpretation:

log(Y) = Intercept + B1 * X + Error

"One unit increase in IV is associated with a ( `B1 * 100` ) percent increase in DV."

Increase of minimum wage is increasing the unemployment duration of high earners by 19%