Homework 3

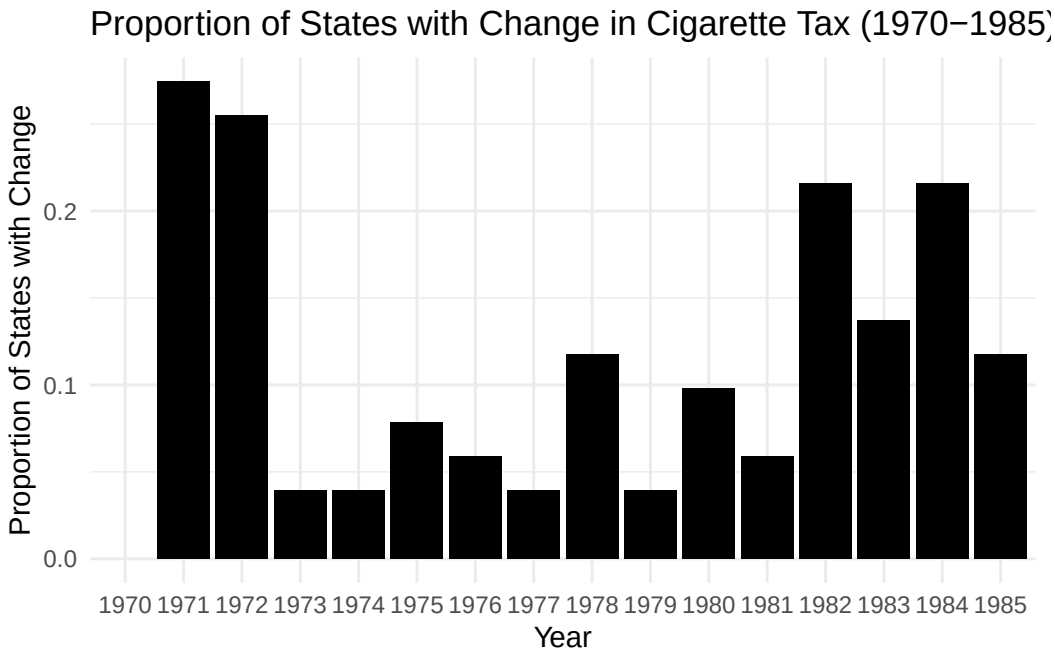
Research Methods, Spring 2024

Taeyoung Yim

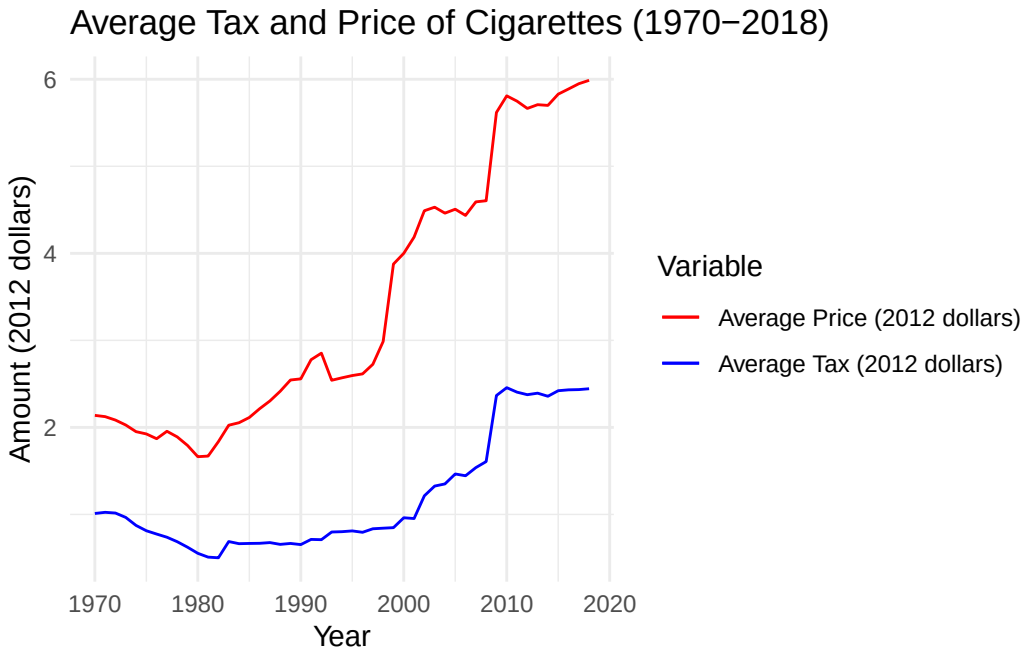
Summarize the Data

Answer the following based on the enrollment data: 1. Present a bar graph showing the proportion of states with a change in their cigarette tax in each year from 1970 to 1985.

```
Warning: Removed 1 rows containing missing values (`position_stack()`).
```

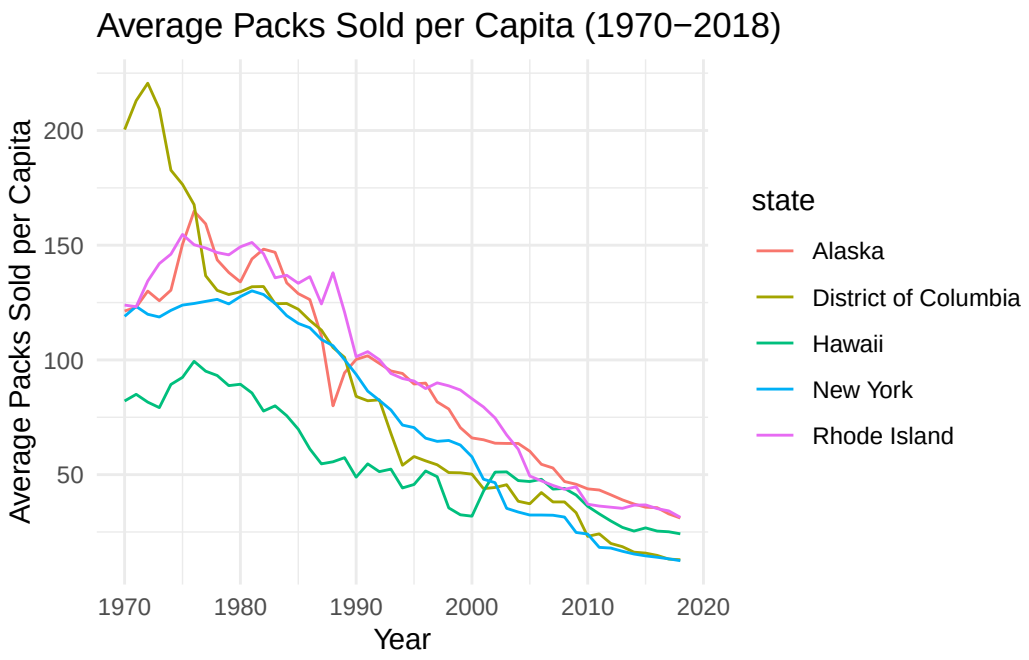


2. Plot on a single graph the average tax (in 2012 dollars) on cigarettes and the average price of a pack of cigarettes from 1970 to 2018.

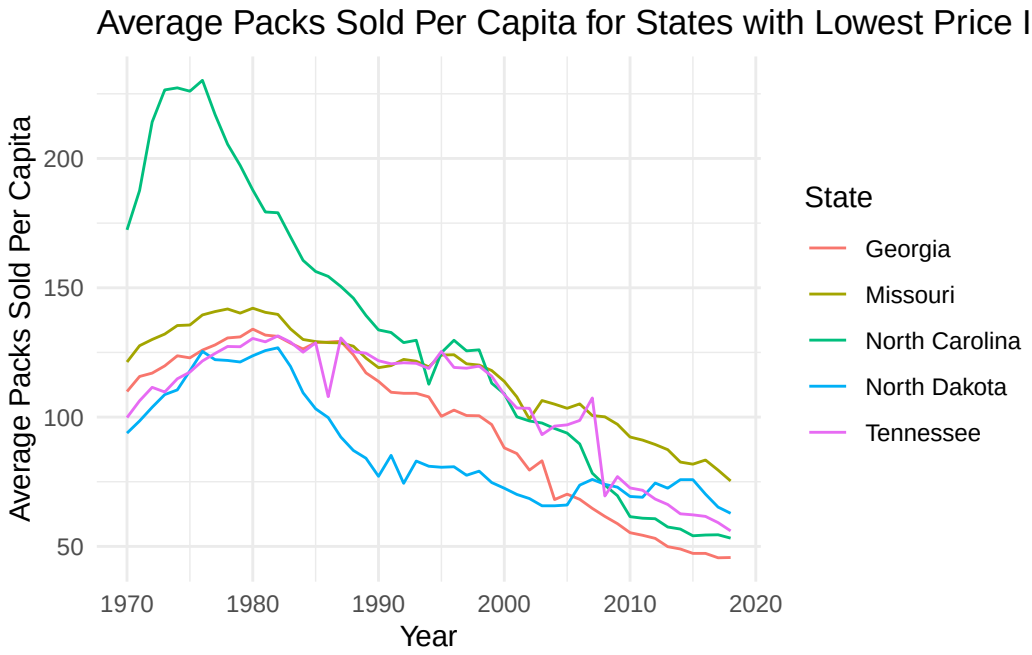


3.Identify the 5 states with the highest increases in cigarette prices (in dollars) over the time period. Plot the average number of packs sold per capita for those states from 1970 to 2018.

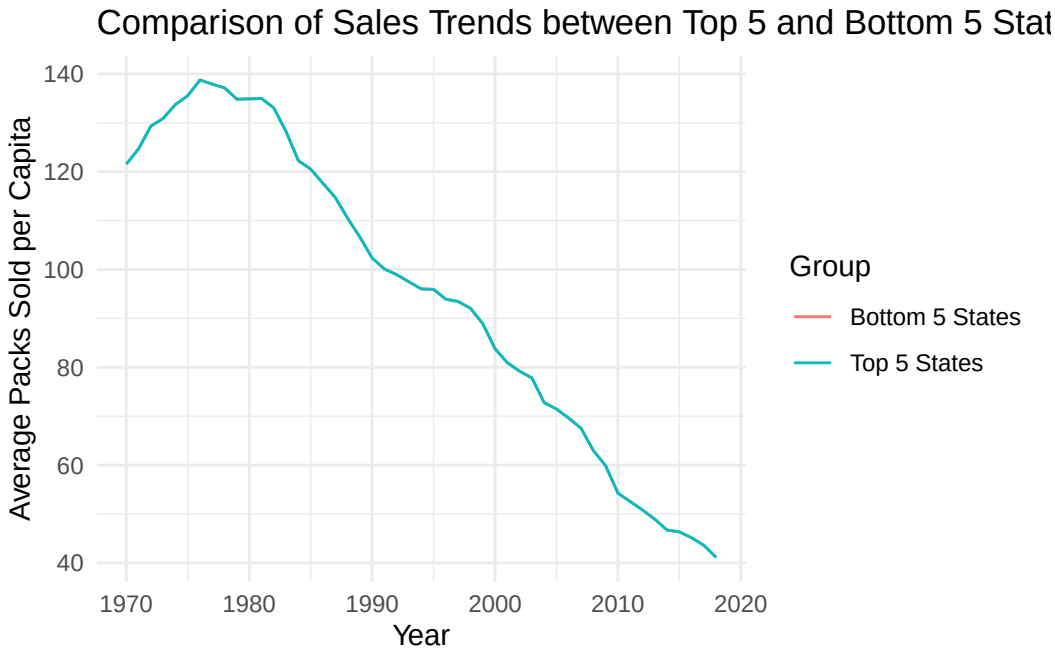
```
Selecting by price_increase
`summarise()` has grouped output by 'Year'. You can override using the
`.groups` argument.
```



4.Identify the 5 states with the lowest increases in cigarette prices over the time period. Plot the average number of packs sold per capita for those states from 1970 to 2018.



5.Compare the trends in sales from the 5 states with the highest price increases to those with the lowest price increases.



Estimate ATEs

Now let's work on estimating a demand curve for cigarettes. Specifically, we're going to estimate the price elasticity of demand for cigarettes. When explaining your findings, try to limit your discussion just to a couple of sentences. 6.Focusing only on the time period from

1970 to 1990, regress log sales on log prices to estimate the price elasticity of demand over that period. Interpret your results.

```
OLS estimation, Dep. Var.: ln_sales
Observations: 1,071
Standard-errors: IID
               Estimate Std. Error  t value  Pr(>|t|)
(Intercept)    5.385438   0.027804 193.6917 < 2.2e-16 ***
ln_price_2012 -0.809438   0.038366 -21.0980 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RMSE: 0.189214   Adj. R2: 0.293322
```

The coefficient of log(price_cpi_2012) is 5.385438 -0.8094384 with a p-value of 0 6.981617e-83 indicating that the relationship between log price and log sales per capita is significant significant.

7.Again limiting to 1970 to 1990, regress log sales on log prices using the total (federal and state) cigarette tax (in dollars) as an instrument for log prices. Interpret your results and compare your estimates to those without an instrument. Are they different? If so, why?

```
TSLS estimation, Dep. Var.: ln_sales, Endo.: ln_price_2012, Instr.: ln_tax_2012
Second stage: Dep. Var.: ln_sales
Observations: 1,071
Standard-errors: IID
                  Estimate Std. Error  t value  Pr(>|t|)
(Intercept)       5.375575   0.050825 105.7659 < 2.2e-16 ***
fit_ln_price_2012 -0.795524   0.071235 -11.1676 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RMSE: 0.189226   Adj. R2: 0.293235
F-test (1st stage), ln_price_2012: stat = 436.8     , p < 2.2e-16 , on 1 and 1,069 DoF.
                      Wu-Hausman: stat =   0.053709, p = 0.816775, on 1 and 1,068 DoF.
```

IV regression results are different from regression without IV. It shows an inverse relationship between cigarette prices and cigarette sales per capita. This suggests that as cigarette prices increase, sales per capita decrease. Using IV deals with the endogenity issue that exist in the regression, resulting in a more reliable result.

8.Show the first stage and reduced-form results from the instrument.

```
OLS estimation, Dep. Var.: ln_price_2012
Observations: 1,071
Standard-errors: IID
            Estimate Std. Error  t value  Pr(>|t|)
(Intercept) 0.801304   0.005889 136.0703 < 2.2e-16 ***
ln_tax_2012 0.260060   0.012443  20.9009 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RMSE: 0.127093   Adj. R2: 0.289437

OLS estimation, Dep. Var.: ln_sales
Observations: 1,071
Standard-errors: IID
             Estimate Std. Error   t value  Pr(>|t|)
(Intercept)  4.738119   0.009995 474.04052 < 2.2e-16 ***
ln_tax_2012 -0.206884   0.021119  -9.79629 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RMSE: 0.215714   Adj. R2: 0.081519
```

9.Repeat questions 1-3 focusing on the period from 1991 to 2015.

```
OLS estimation, Dep. Var.: ln_sales
Observations: 1,275
Standard-errors: IID
               Estimate Std. Error  t value  Pr(>|t|)
(Intercept)    5.608303   0.035140 159.6000 < 2.2e-16 ***
ln_price_2012 -0.996814   0.024692 -40.3697 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RMSE: 0.295775   Adj. R2: 0.561101

TSLS estimation, Dep. Var.: ln_sales, Endo.: ln_price_2012, Instr.: ln_tax_2012
Second stage: Dep. Var.: ln_sales
Observations: 1,275
Standard-errors: IID
                  Estimate Std. Error  t value  Pr(>|t|)
(Intercept)        5.82027   0.039371 147.8322 < 2.2e-16 ***
fit_ln_price_2012 -1.15008   0.027811 -41.3536 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RMSE: 0.300218   Adj. R2: 0.547816
F-test (1st stage), ln_price_2012: stat = 5,503.6, p < 2.2e-16, on 1 and 1,273 DoF.
                        Wu-Hausman: stat =   191.5, p < 2.2e-16, on 1 and 1,272 DoF.

OLS estimation, Dep. Var.: ln_price_2012
Observations: 1,275
Standard-errors: IID
            Estimate Std. Error  t value  Pr(>|t|)
(Intercept) 1.289867   0.004267 302.2891 < 2.2e-16 ***
ln_tax_2012 0.513550   0.006922  74.1860 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RMSE: 0.145512   Adj. R2: 0.811999

OLS estimation, Dep. Var.: ln_sales
Observations: 1,275
Standard-errors: IID
             Estimate Std. Error  t value  Pr(>|t|)
(Intercept)  4.336813   0.008211 528.1884 < 2.2e-16 ***
ln_tax_2012 -0.590626   0.013320 -44.3396 < 2.2e-16 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RMSE: 0.28   Adj. R2: 0.606669
```

10.Compare your elasticity estimates from 1970-1990 versus those from 1991-2015. Are they different? If so, why?

```
log(price_cpi_2012)
         -0.8094384
```

```
log(price_cpi_2012)
         -0.9968136
```

The difference in elasticity estimates exist. Possible reasons for the difference is probably policy interventions like taxing cigarettes.

```
[1] "cig.data"         "cig.data.change"  "top.bottom.price"
```