Homework 3

ECON 470, Spring 2025

Camila Castaneda

Here is a link to my repository: {https://github.com/ctcasta/homework3}

1. Present a bar graph showing the proportion of states with a change in their cigarette tax in each year from 1970 to 1985.

Proportion of States with a Change in Cigarette Tax (1970–19
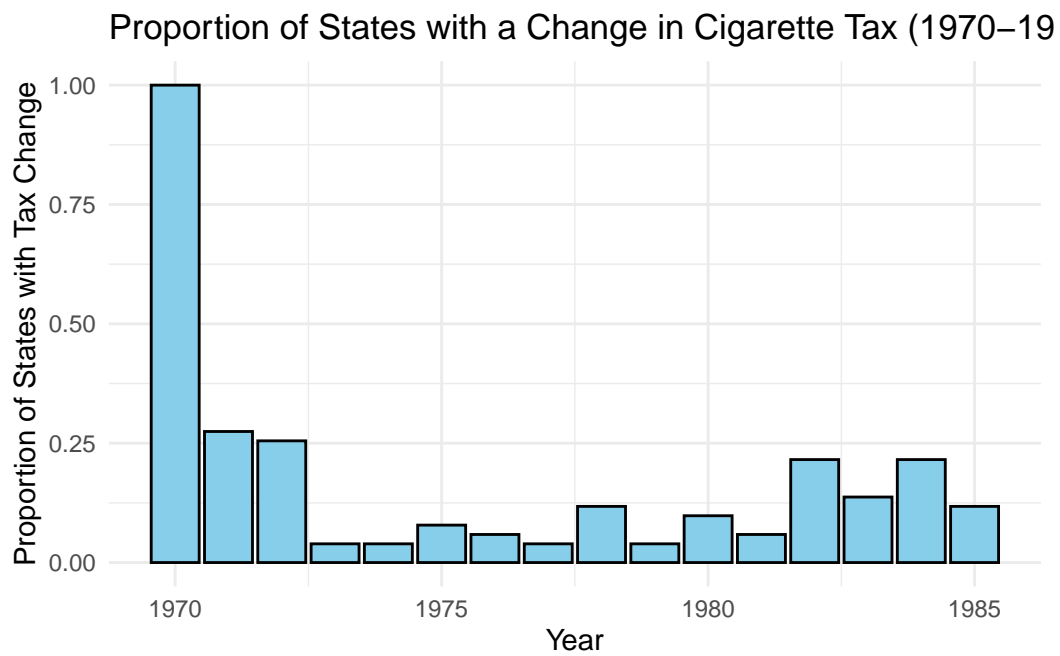
2. Plot on a single graph the average tax (in 2012 dollars) on cigarettes and the average price of a pack of cigarettes from 1970 to 2018.

Average Cigarette Tax and Price Adjusted to 2012 Dollars

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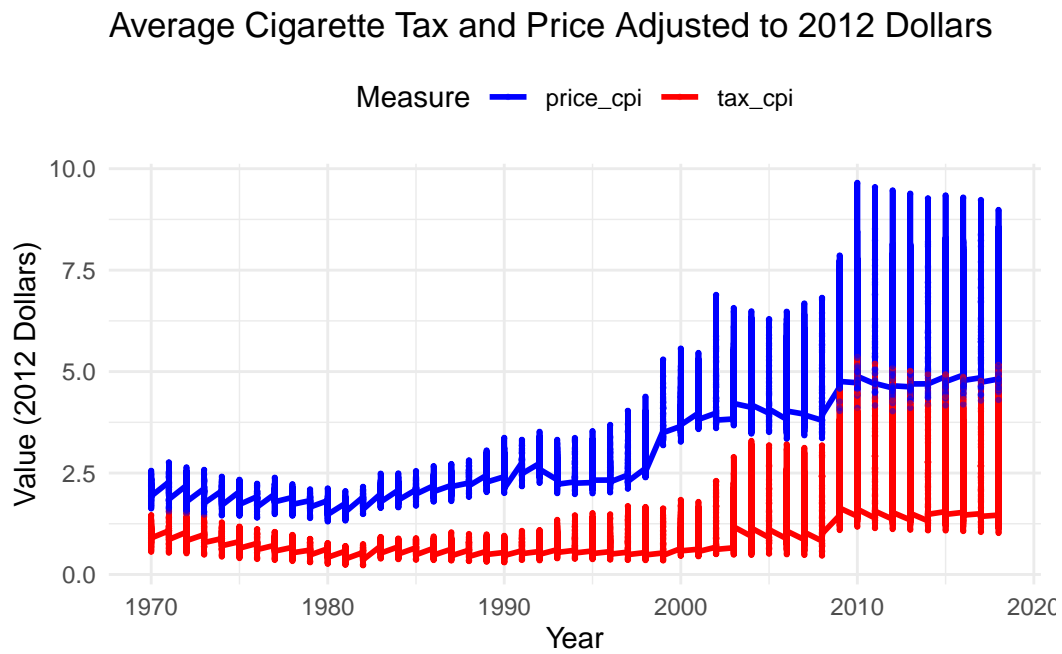
3. Identify the 5 states with the highest increases in cigarette prices (in dollars) over the time period. Plot the average number of packs sold per capita for those states from 1970 to 2018.

Average Packs Sold Per Capita (Top 5 States with Highest Pric

State — Alaska — Hawaii — Massachusetts — New York — Rhode Islan
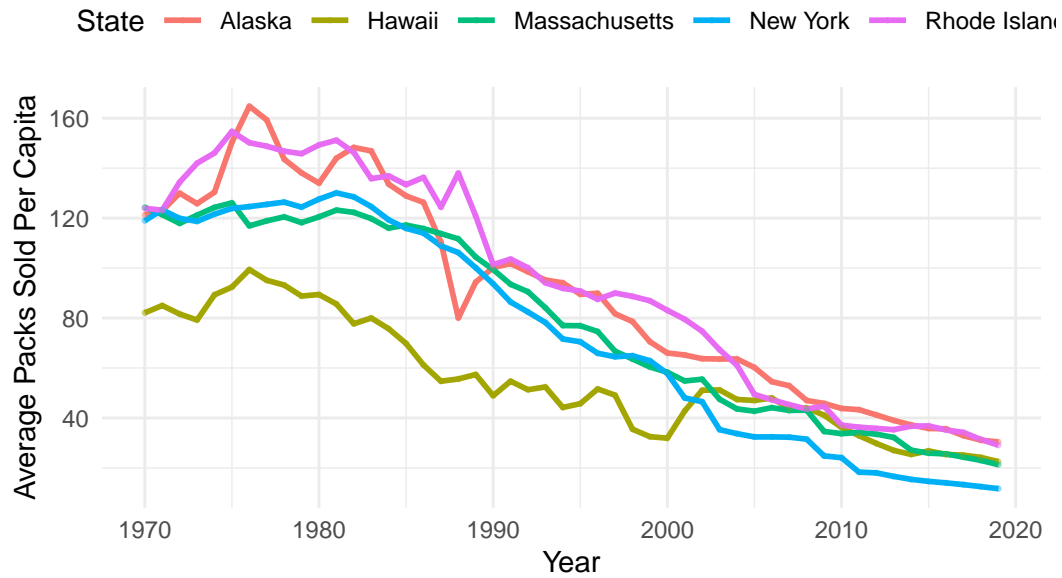
4. Identify the 5 states with the lowest increases in cigarette prices over the time period. Plot the average number of packs sold per capita for those states from 1970 to 2018.

Average Packs Sold Per Capita (Top 5 States with Lowest Pric
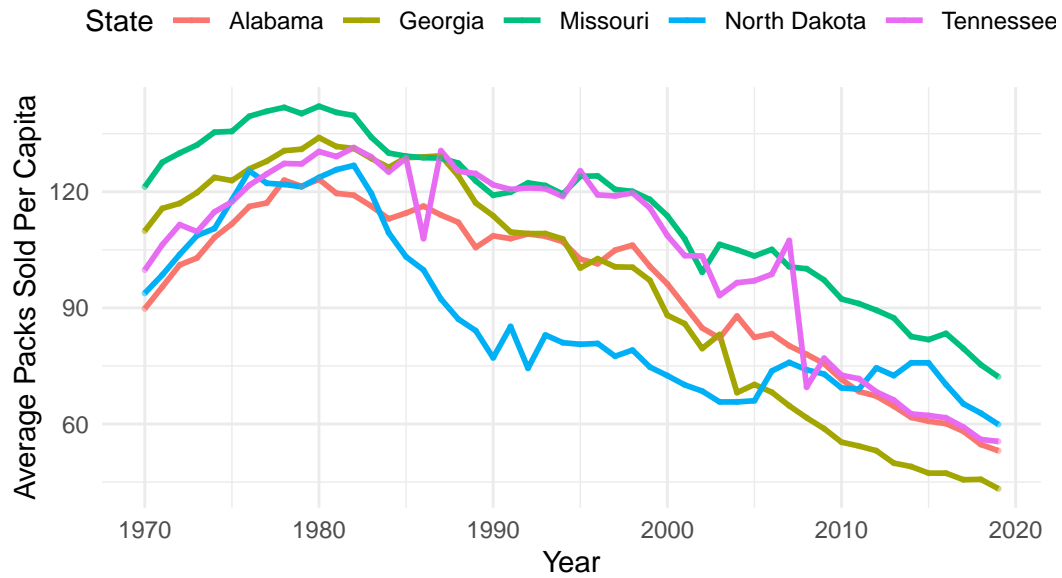
5. Compare the trends in sales from the 5 states with the highest price increases to those with the lowest price increases.

Comparison of Cigarette Sales in States with High vs. Low Pri...

| Group | Bottom 5 Price Increase | Top 5 Price Increase |
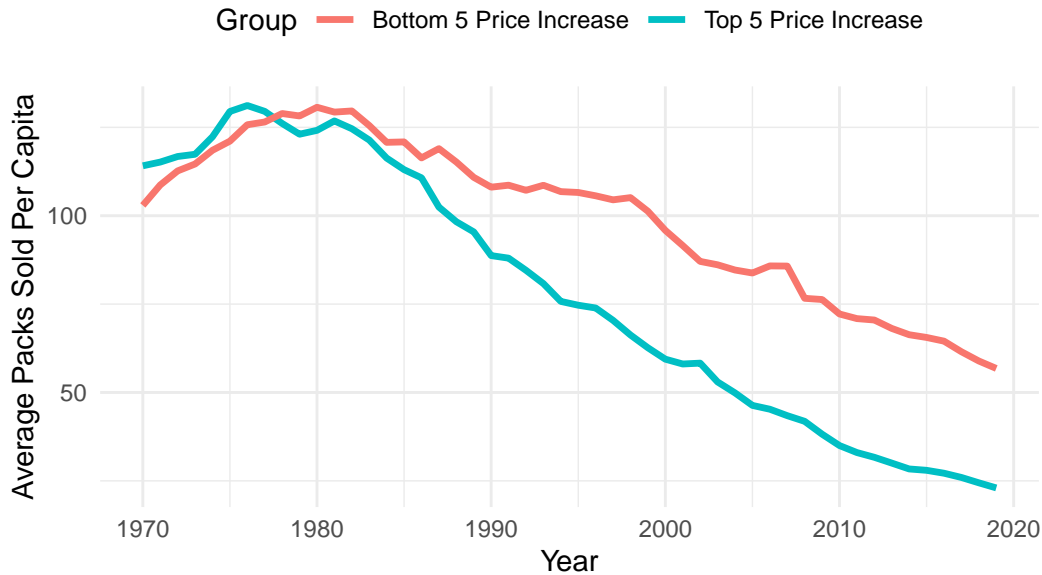
The following is focused on data from the years 1970-1990

6. Regress log sales on log prices to estimate the price elasticity of demand over that period. Interpret your results.

```
Call:
lm(formula = log_sales ~ log_price, data = final.data.70.90)

Residuals:
     Min       1Q   Median       3Q      Max
-0.68335 -0.08598 -0.00284  0.08778  0.83516

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.38544    0.02780   193.7   <2e-16 ***
log_price   -0.80944    0.03837   -21.1   <2e-16 ***
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1894 on 1069 degrees of freedom
Multiple R-squared:  0.294, Adjusted R-squared:  0.2933
F-statistic: 445.1 on 1 and 1069 DF,  p-value: < 2.2e-16
```

7. 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.: log_sales
                   Endo.    : log_price
                   Instr.   : log_total_tax
Second stage: Dep. Var.: log_sales
Observations: 1,071
Standard-errors: IID
               Estimate Std. Error  t value  Pr(>|t|)
(Intercept)    5.465991   0.035234 155.1349 < 2.2e-16 ***
fit_log_price -0.923078   0.049025 -18.8286 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RMSE: 0.189989   Adj. R2: 0.287523
F-test (1st stage), log_price: stat = 1,725.3, p < 2.2e-16 , on 1 and 1,069 DoF.
                  Wu-Hausman: stat =    14.3, p = 1.614e-4, on 1 and 1,068 DoF.
```

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

```
Call:
lm(formula = log_price ~ log_total_tax, data = final.data.70.90)

Residuals:
     Min        1Q    Median        3Q       Max
-0.224556 -0.063007  0.001792  0.064653  0.297333

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.181922   0.011741  100.66   <2e-16 ***
log_total_tax 0.332789   0.008012   41.54   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.09339 on 1069 degrees of freedom
Multiple R-squared:  0.6174,    Adjusted R-squared:  0.6171
F-statistic:  1725 on 1 and 1069 DF,  p-value: < 2.2e-16
```

```
Call:
lm(formula = log_sales ~ log_total_tax, data = final.data.70.90)

Residuals:
     Min       1Q   Median       3Q      Max
-0.75589 -0.08447  0.00043  0.09596  0.80589

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)    4.37498    0.02477  176.63   <2e-16 ***
log_total_tax -0.30719    0.01690  -18.18   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.197 on 1069 degrees of freedom
Multiple R-squared:  0.2361,    Adjusted R-squared:  0.2353
F-statistic: 330.3 on 1 and 1069 DF,  p-value: < 2.2e-16
```

9. The following repeats questions 6-8 but is focused on data from the years 1991-2015

6.2

```
Call:
lm(formula = log_sales ~ log_price, data = final.data.91.15)

Residuals:
     Min       1Q   Median       3Q      Max
-0.92230 -0.17004  0.00664  0.17869  1.10282

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.60830    0.03514  159.60   <2e-16 ***
log_price   -0.99681    0.02469  -40.37   <2e-16 ***
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.296 on 1273 degrees of freedom
Multiple R-squared:  0.5614,    Adjusted R-squared:  0.5611
F-statistic:  1630 on 1 and 1273 DF,  p-value: < 2.2e-16
```

7.2

```
TSLS estimation - Dep. Var.: log_sales
                  Endo.   : log_price
                  Instr.  : log_total_tax
Second stage: Dep. Var.: log_sales
Observations: 1,275
Standard-errors: IID
              Estimate Std. Error  t value  Pr(>|t|)
(Intercept)    5.76890   0.037898 152.2221 < 2.2e-16 ***
fit_log_price -1.11294   0.026728 -41.6390 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RMSE: 0.298334   Adj. R2: 0.553474
F-test (1st stage), log_price: stat = 8,390.2, p < 2.2e-16, on 1 and 1,273 DoF.
                  Wu-Hausman: stat =   164.5, p < 2.2e-16, on 1 and 1,272 DoF.
```

8.2

```
Call:
lm(formula = log_price ~ log_total_tax, data = final.data.91.15)

Residuals:
    Min       1Q   Median       3Q      Max
-0.27511 -0.07478 -0.01372  0.07629  0.39324

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.376565   0.003416   403.0   <2e-16 ***
log_total_tax 0.431717   0.004713    91.6   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.122 on 1273 degrees of freedom
```

```
Multiple R-squared:  0.8683,    Adjusted R-squared:  0.8682
F-statistic:  8390 on 1 and 1273 DF,  p-value: < 2.2e-16




Call:
lm(formula = log_sales ~ log_total_tax, data = final.data.91.15)

Residuals:
     Min       1Q   Median       3Q      Max
-0.82897 -0.14423  0.00604  0.14668  1.19203

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)    4.236866   0.007842   540.3   <2e-16 ***
log_total_tax -0.480477   0.010820   -44.4   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.28 on 1273 degrees of freedom
Multiple R-squared:  0.6077,    Adjusted R-squared:  0.6074
F-statistic:  1972 on 1 and 1273 DF,  p-value: < 2.2e-16
```

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

Table 1: Comparison of Slope Estimates for 1970-1990 and 1991-2015

	1970-1990	1991-2015
Slope Estimate	-0.9230776	-1.112943