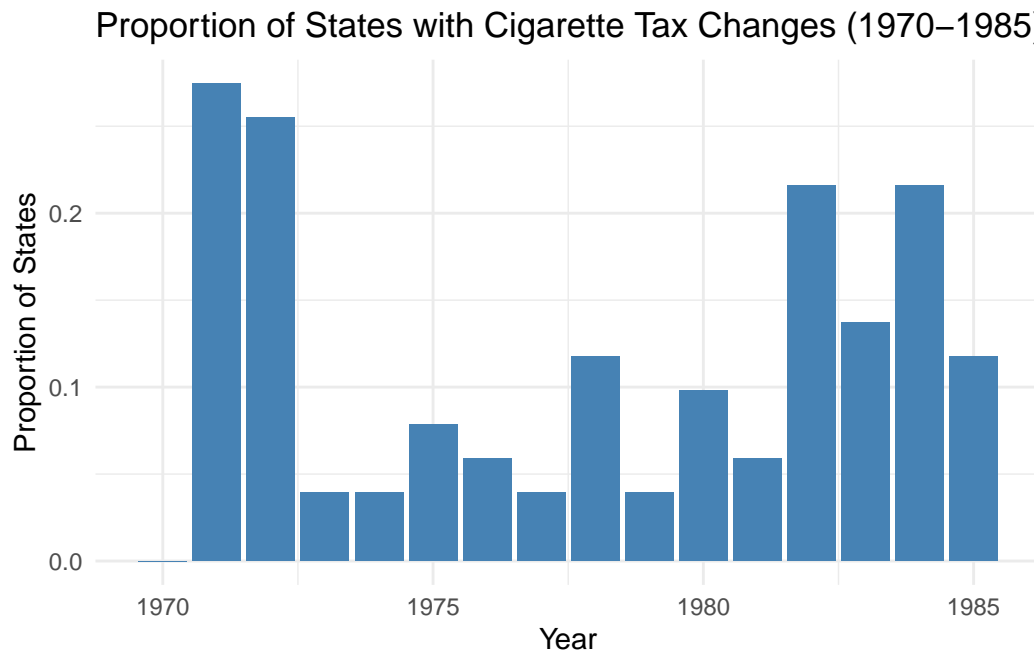# Homework 3
## Research Methods, Spring 2025

Genevieve DeBell

My answers to the homework questions are described below. The GitHub repository for this work is available here. Enjoy!

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 Cigarette Tax Changes (1970–1985)

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.

```
Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
i Please use `linewidth` instead.
```

Average Tax and Price of Cigarettes (1970–2018, Adjusted to 20
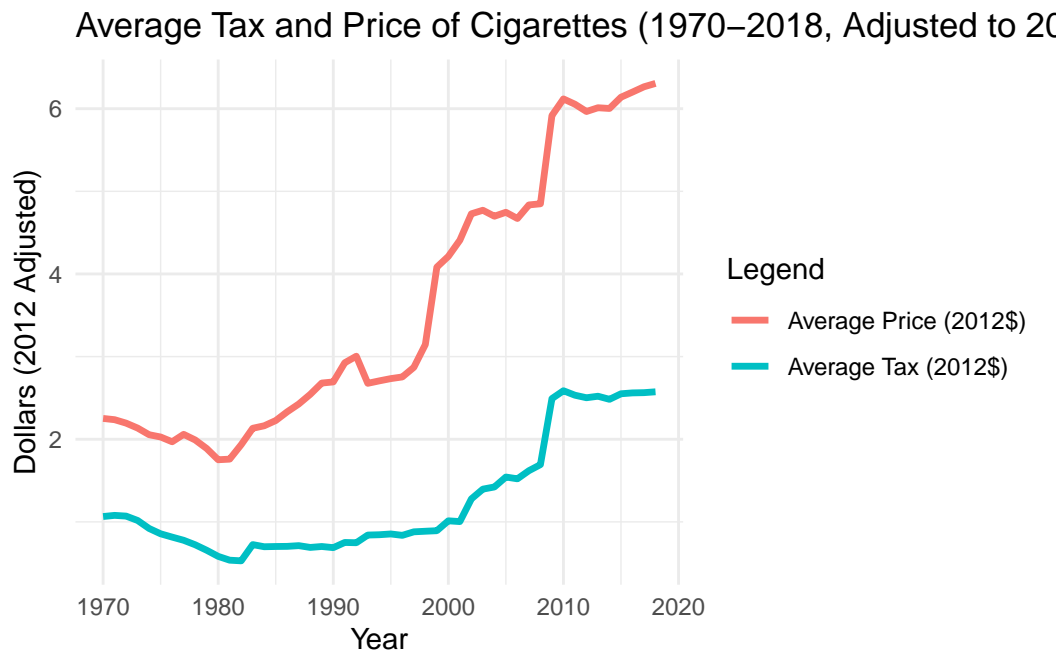
3. Identify the 5 states with the highest increases in cigarette prices (in dollars) over the time period. Plot the average number of packs sold per capita for those states from 1970 to 2018.

Average Packs Sold Per Capita (1970–2018)

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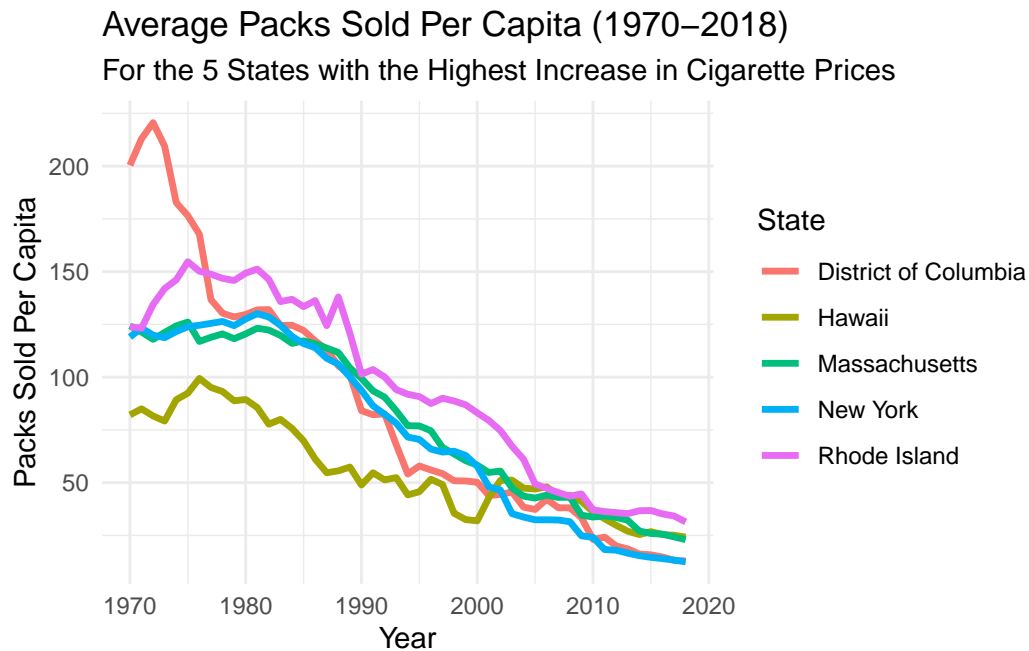
4. Identify the 5 states with the lowest increases in cigarette prices over the time period. Plot the average number of packs sold per capita for those states from 1970 to 2018.

Average Packs Sold Per Capita (1970–2018)

For the 5 States with the Lowest Increase in Cigarette Prices
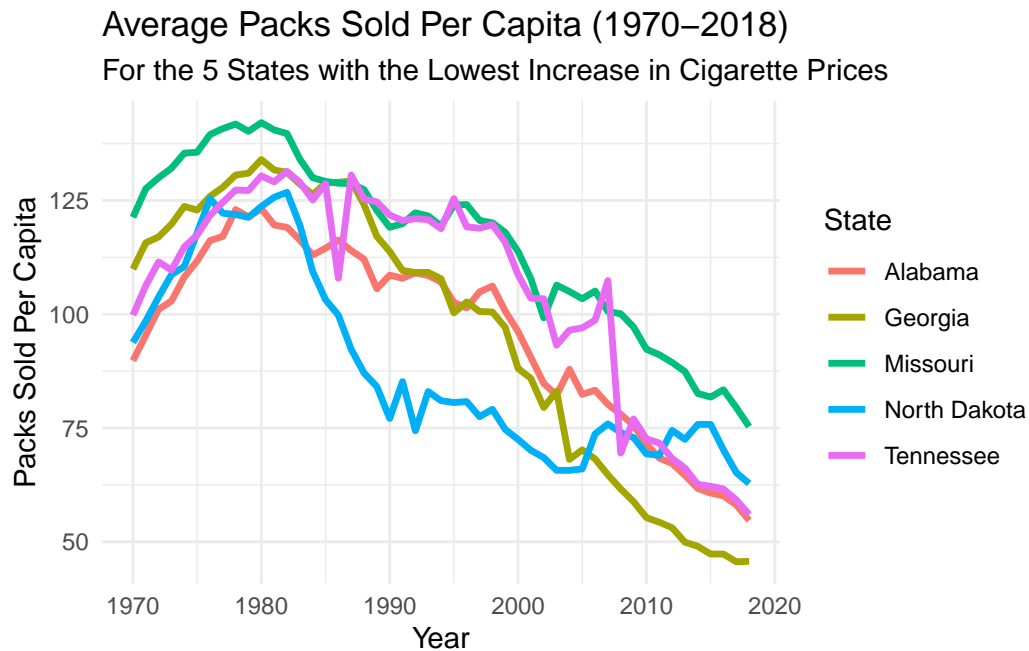
5. Compare the trends in sales from the 5 states with the highest price increases to those with the lowest price increases.

Comparison of Cigarette Sales in States with High vs. Low Pri
Average Packs Sold Per Capita (1970–2018)
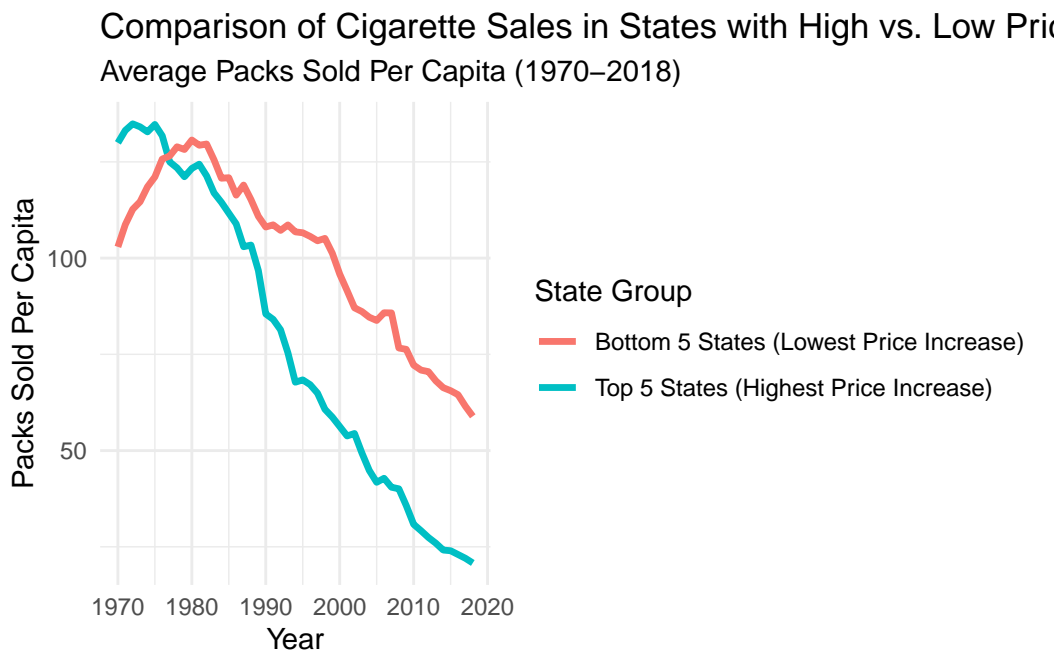
6. Focusing only on the time period from 1970 to 1990, regress log sales on log prices to estimate the price elasticity of demand over that period. Interpret your results.

```
Call:
lm(formula = log_sales ~ log_price, data = demand_data)

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)  5.42738    0.02975   182.4   <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. 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
                 Instr.  : ln_total_tax
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 -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: 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.
```

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

```
OLS estimation, Dep. Var.: ln_price
Observations: 1,071
Standard-errors: IID
             Estimate Std. Error  t value  Pr(>|t|)
(Intercept)  0.787829   0.005421 145.3208 < 2.2e-16 ***
ln_total_tax 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)   4.748839   0.009202  516.09178 < 2.2e-16 ***
ln_total_tax -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 6-8 focusing on the period from 1991 to 2015.

```
Call:
lm(formula = log_sales ~ log_price, data = demand_data)

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)  5.42738    0.02975   182.4   <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


TSLS estimation - Dep. Var.: ln_sales
                  Endo.    : ln_price
                  Instr.   : ln_total_tax
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 -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: 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.


OLS estimation, Dep. Var.: ln_price
Observations: 1,071
Standard-errors: IID
             Estimate Std. Error  t value  Pr(>|t|)
(Intercept)  0.787829   0.005421 145.3208 < 2.2e-16 ***
ln_total_tax 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)   4.748839   0.009202 516.09178 < 2.2e-16 ***
ln_total_tax -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
```