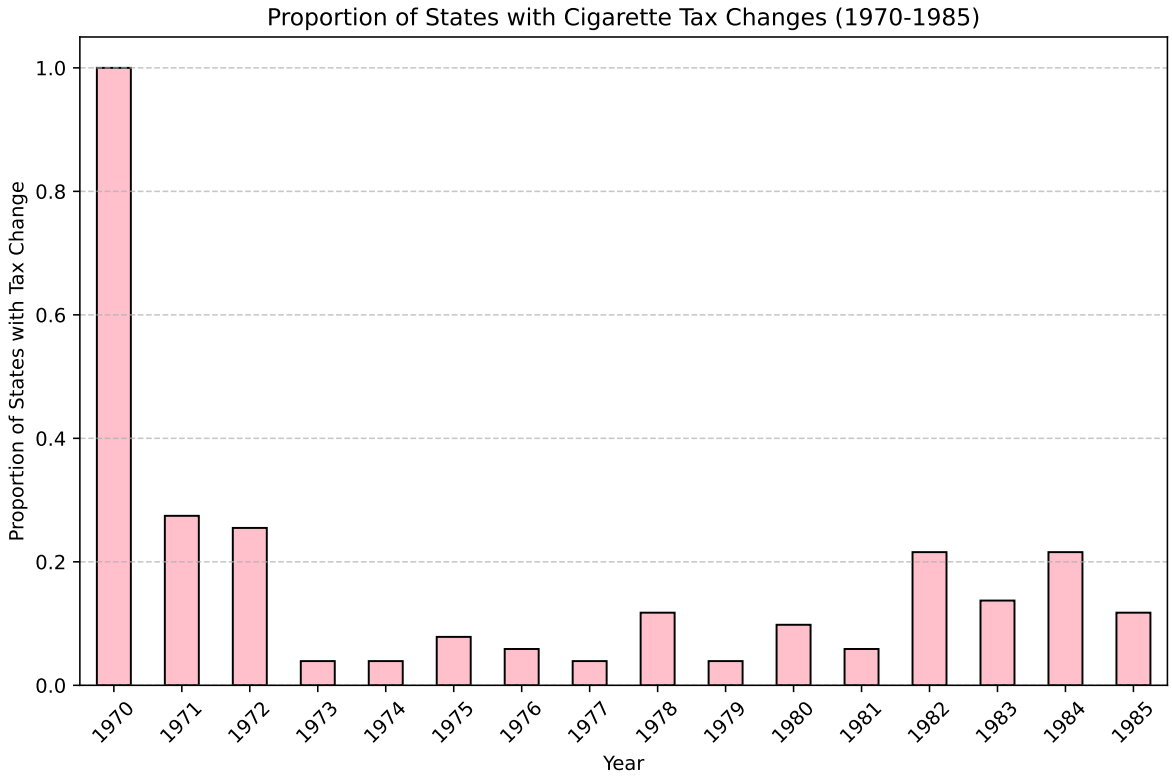
Homework 3-2

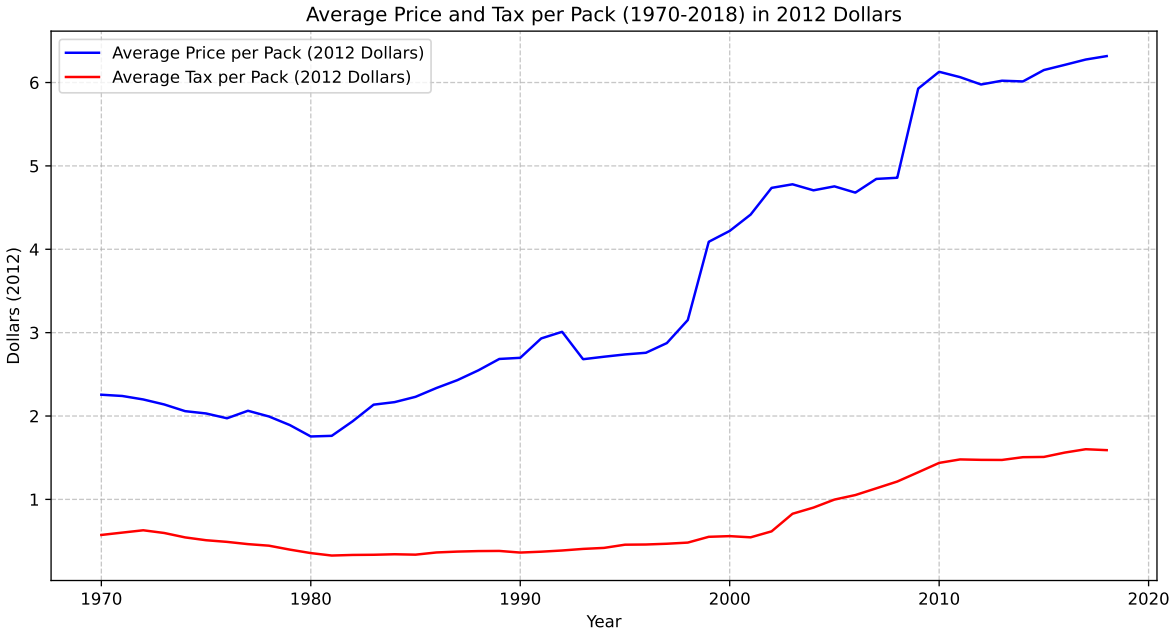
Sarina Tan

The link to my repository:
https://github.com/sarina-tan/HLTH470hw3/tree/main

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

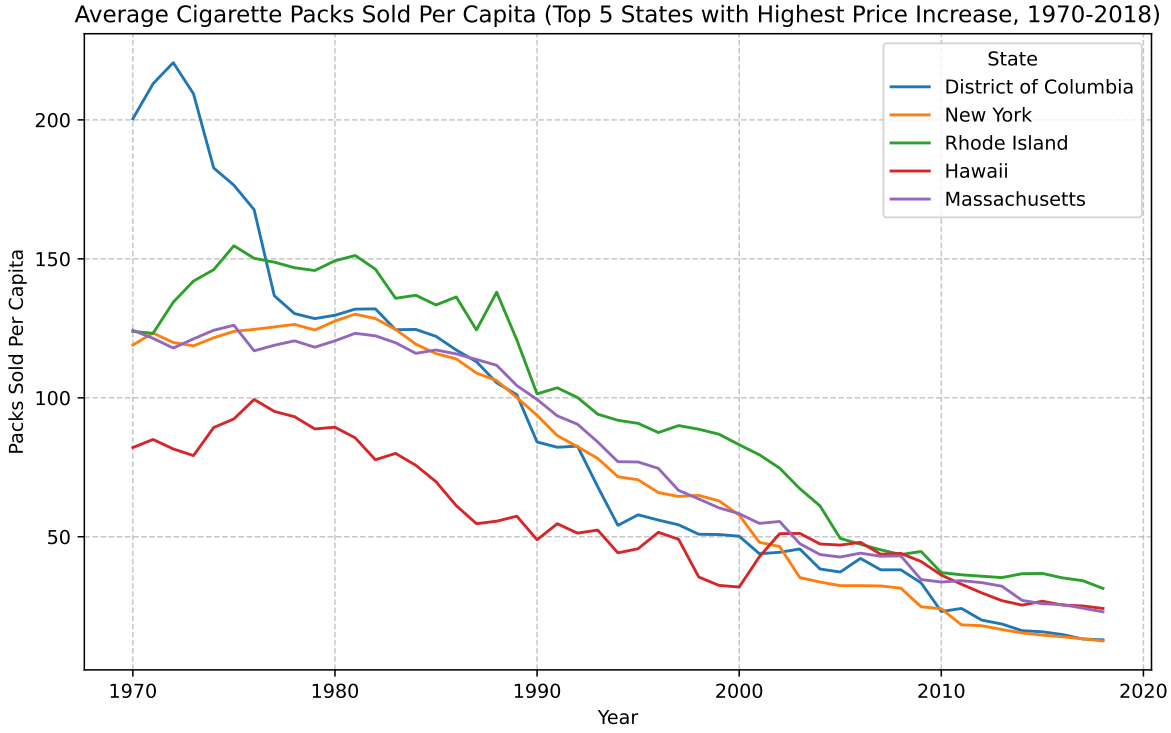


2. Plot on a single graph the average tax (in 2012 dollars) on cigarettes and the average price of a pack of cigarettes from 1970 to 2018.



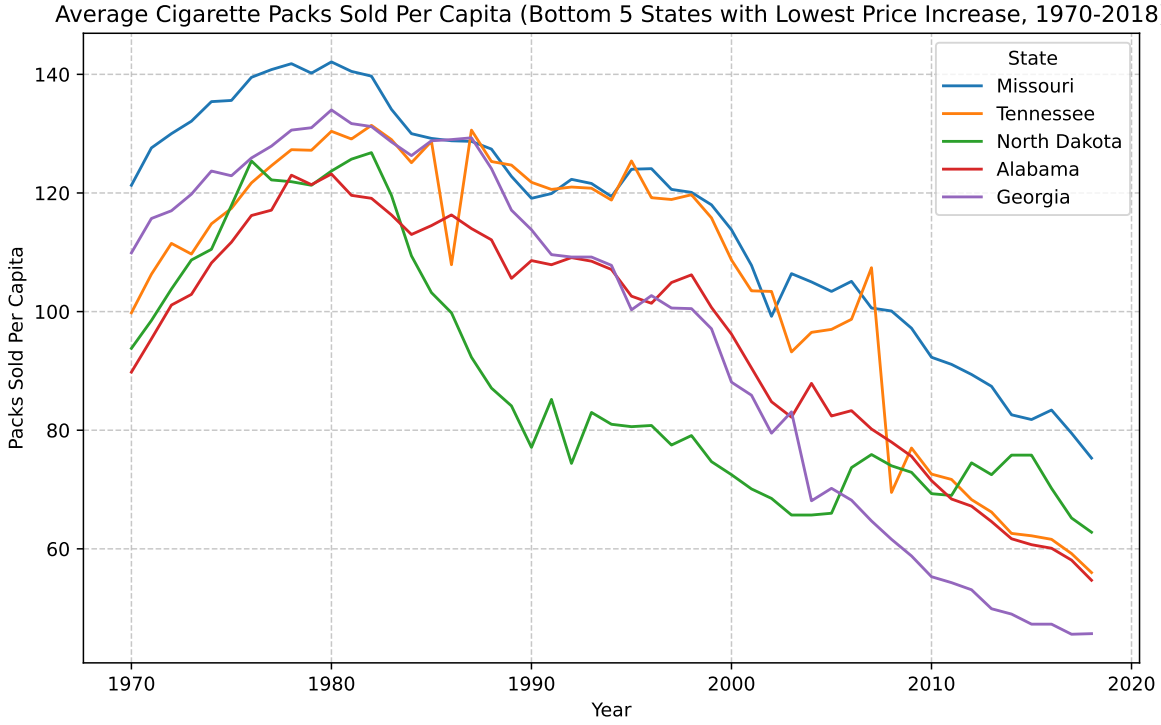
3. Identify the 5 states with the highest increases in cigarette prices (in dollars) over the time period. Plot the average number of packs sold per capita for those states from 1970 to 2018.

Top 5 states with highest price increase: [‘District of Columbia’, ‘New York’, ‘Rhode Island’, ‘Hawaii’, ‘Massachusetts’]



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.

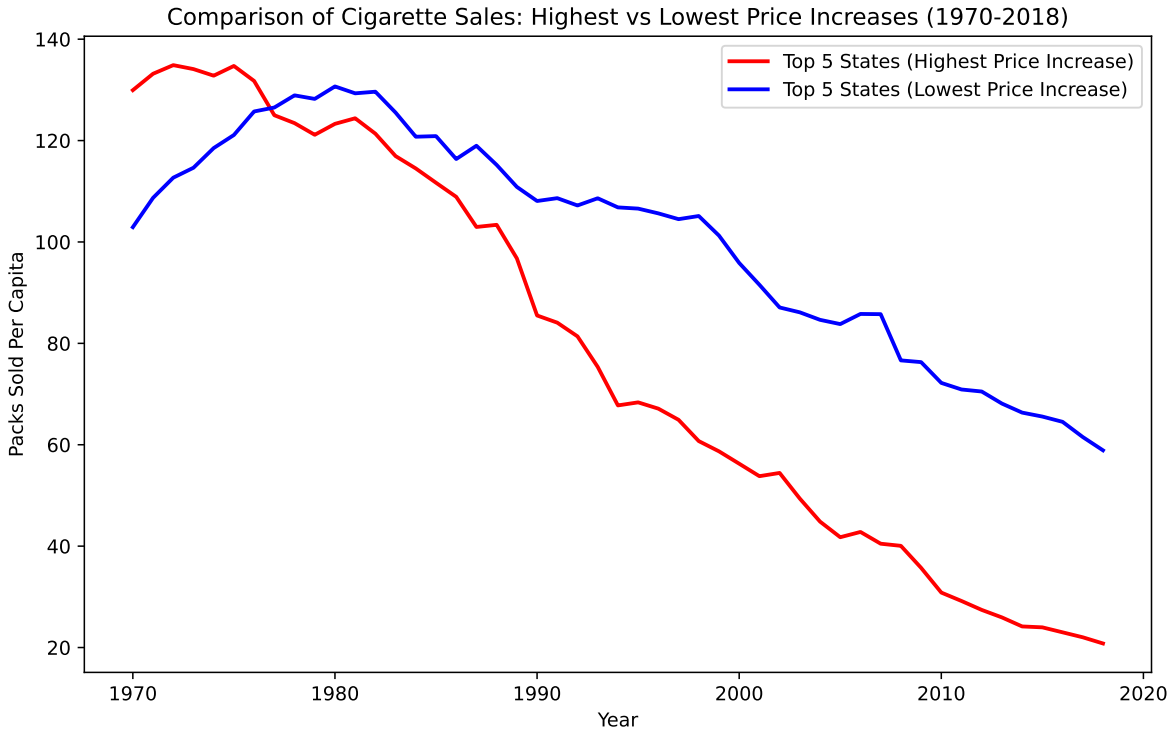
Bottom 5 states with lowest price increase: ['Missouri', 'Tennessee', 'North Dakota', 'Alabama', 'Georgia']



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

Top 5 states with highest price increase: [‘District of Columbia’, ‘New York’, ‘Rhode Island’, ‘Hawaii’, ‘Massachusetts’]

Bottom 5 states with lowest price increase: [‘Missouri’, ‘Tennessee’, ‘North Dakota’, ‘Alabama’, ‘Georgia’]



The five states with the largest increases in cigarette prices saw a sharp decline in packs sold per capita from around 1970 to 2018, indicating the influence of higher prices and likely stricter tobacco control policies. In contrast, the five states with the smallest price increases experienced more stable cigarette consumption over time, with only a gradual decline. This suggests that states with smaller price hikes, many of which have historical ties to tobacco production, may have been less proactive in using price-based policies to discourage smoking.

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 Regression Results

Dep. Variable:	log_sales	R-squared:	0.294
Model:	OLS	Adj. R-squared:	0.293
Method:	Least Squares	F-statistic:	445.1
Date:	Tue, 18 Mar 2025	Prob (F-statistic):	6.98e-83
Time:	20:10:33	Log-Likelihood:	263.40
No. Observations:	1071	AIC:	-522.8
Df Residuals:	1069	BIC:	-512.8
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	5.3854	0.028	193.692	0.000	5.331	5.440
log_price	-0.8094	0.038	-21.098	0.000	-0.885	-0.734

Omnibus:	89.160	Durbin-Watson:	0.183
Prob(Omnibus):	0.000	Jarque-Bera (JB):	466.536
Skew:	0.128	Prob(JB):	4.93e-102
Kurtosis:	6.223	Cond. No.	10.0

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
Estimated Price Elasticity of Demand: -0.809

The estimated price elasticity of demand is -0.809, meaning that a 1% increase in price is associated with a 0.8% decrease in cigarette consumption.

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?

7.1 First-Stage Regression

First-Stage Regression Results:

OLS Regression Results

Dep. Variable:	log_price	R-squared:	0.617
Model:	OLS	Adj. R-squared:	0.617
Method:	Least Squares	F-statistic:	1725.
Date:	Tue, 18 Mar 2025	Prob (F-statistic):	2.80e-225
Time:	20:10:33	Log-Likelihood:	1020.7
No. Observations:	1071	AIC:	-2037.
Df Residuals:	1069	BIC:	-2027.
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	1.1819	0.012	100.663	0.000	1.159	1.205
log_tax	0.3328	0.008	41.537	0.000	0.317	0.349

Omnibus:	6.850	Durbin-Watson:	0.303
Prob(Omnibus):	0.033	Jarque-Bera (JB):	5.505
Skew:	0.081	Prob(JB):	0.0638
Kurtosis:	2.689	Cond. No.	8.72

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

7.2 Second-Stage Regression

Second-Stage (IV) Regression Results:

OLS Regression Results

```
==============================================================================
Dep. Variable:              log_sales   R-squared:                       0.236
Model:                            OLS   Adj. R-squared:                  0.235
Method:                 Least Squares   F-statistic:                     330.3
Date:                Tue, 18 Mar 2025   Prob (F-statistic):           1.56e-64
Time:                        20:10:33   Log-Likelihood:                 221.17
No. Observations:                1071   AIC:                            -438.3
Df Residuals:                    1069   BIC:                            -428.4
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          5.4660      0.037    149.749      0.000       5.394       5.538
log_price_hat -0.9231      0.051    -18.175      0.000      -1.023      -0.823
==============================================================================
Omnibus:                       83.338   Durbin-Watson:                   0.157
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              430.014
Skew:                           0.023   Prob(JB):                     4.20e-94
Kurtosis:                       6.104   Cond. No.                         12.7
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

IV-Estimated Price Elasticity of Demand: -0.923

OLS-Estimated Price Elasticity: -0.809
Difference between OLS and IV Estimates: 0.114
```

Using cigarette taxes as an instrument for prices, the estimated price elasticity of demand for cigarettes between 1970 and 1990 is -0.923. Compared to the OLS estimate of -0.809, this IV estimate suggests a greater sensitivity to price changes. The difference indicates that the OLS regression likely underestimated the true elasticity, potentially due to endogeneity bias—where factors like state-level anti-smoking campaigns or cultural attitudes toward smoking may simultaneously impact both cigarette consumption and pricing. By leveraging taxes as an instrument, the IV approach isolates the effect of external price shifts driven by policy, offering a more accurate measure of consumer responsiveness to price changes.

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

```
=== First-Stage Regression: Log(Price) ~ Log(Tax) ===
                            OLS Regression Results
==============================================================================
Dep. Variable:              log_price   R-squared:                       0.617
Model:                            OLS   Adj. R-squared:                  0.617
Method:                 Least Squares   F-statistic:                     1725.
Date:                Tue, 18 Mar 2025   Prob (F-statistic):          2.80e-225
Time:                        20:10:33   Log-Likelihood:                 1020.7
No. Observations:                1071   AIC:                            -2037.
Df Residuals:                    1069   BIC:                            -2027.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          1.1819      0.012    100.663      0.000       1.159       1.205
log_tax        0.3328      0.008     41.537      0.000       0.317       0.349
==============================================================================
Omnibus:                        6.850   Durbin-Watson:                   0.303
Prob(Omnibus):                  0.033   Jarque-Bera (JB):                5.505
Skew:                           0.081   Prob(JB):                       0.0638
Kurtosis:                       2.689   Cond. No.                         8.72
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

The first stage regression shows that cigarette taxes are strongly associated with a positive relationship with cigarette prices, with a statistically significant coefficient, confirming that taxes are a strong and relevant instrument for price.

```
=== Reduced-Form Regression: Log(Sales) ~ Log(Tax) ===
                            OLS Regression Results
==============================================================================
Dep. Variable:              log_sales   R-squared:                       0.236
Model:                            OLS   Adj. R-squared:                  0.235
```

Method:	Least Squares	F-statistic:	330.3
Date:	Tue, 18 Mar 2025	Prob (F-statistic):	1.56e-64
Time:	20:10:33	Log-Likelihood:	221.17
No. Observations:	1071	AIC:	-438.3
Df Residuals:	1069	BIC:	-428.4
Df Model:	1		
Covariance Type:	nonrobust		

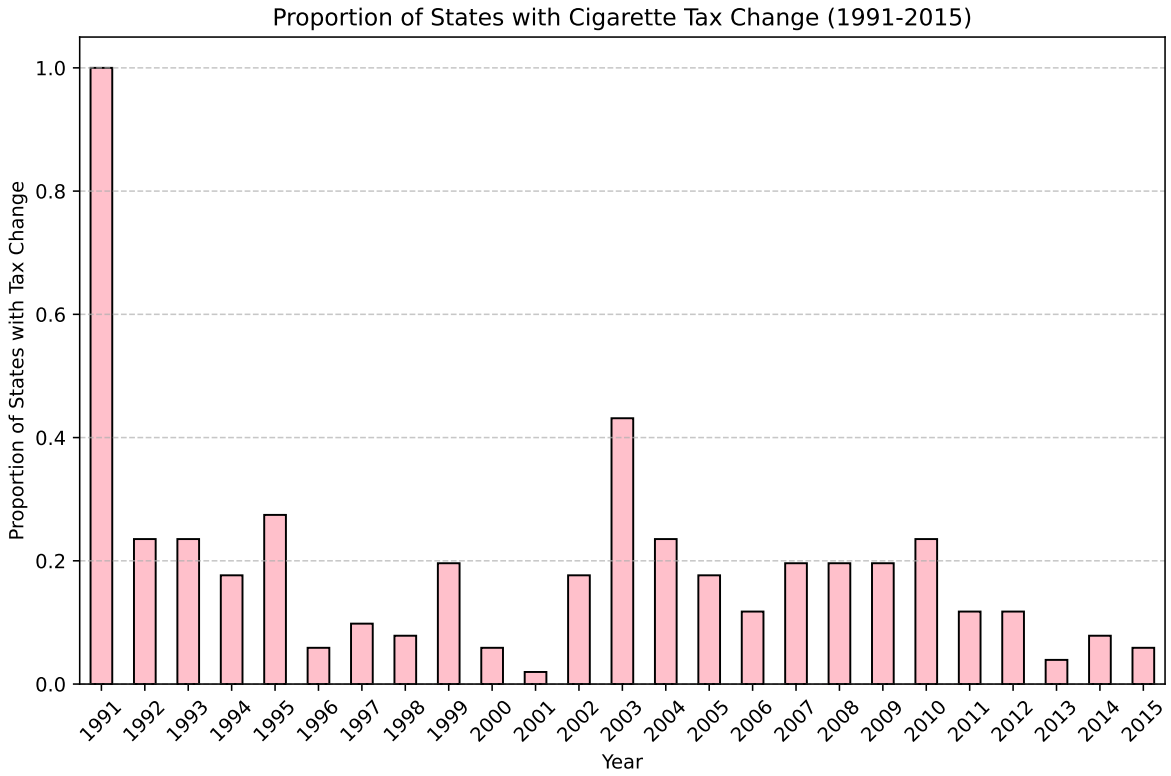
	coef	std err	t	P>|t|	[0.025	0.975]
const	4.3750	0.025	176.627	0.000	4.326	4.424
log_tax	-0.3072	0.017	-18.175	0.000	-0.340	-0.274

Omnibus:	83.338	Durbin-Watson:	0.157
Prob(Omnibus):	0.000	Jarque-Bera (JB):	430.014
Skew:	0.023	Prob(JB):	4.20e-94
Kurtosis:	6.104	Cond. No.	8.72

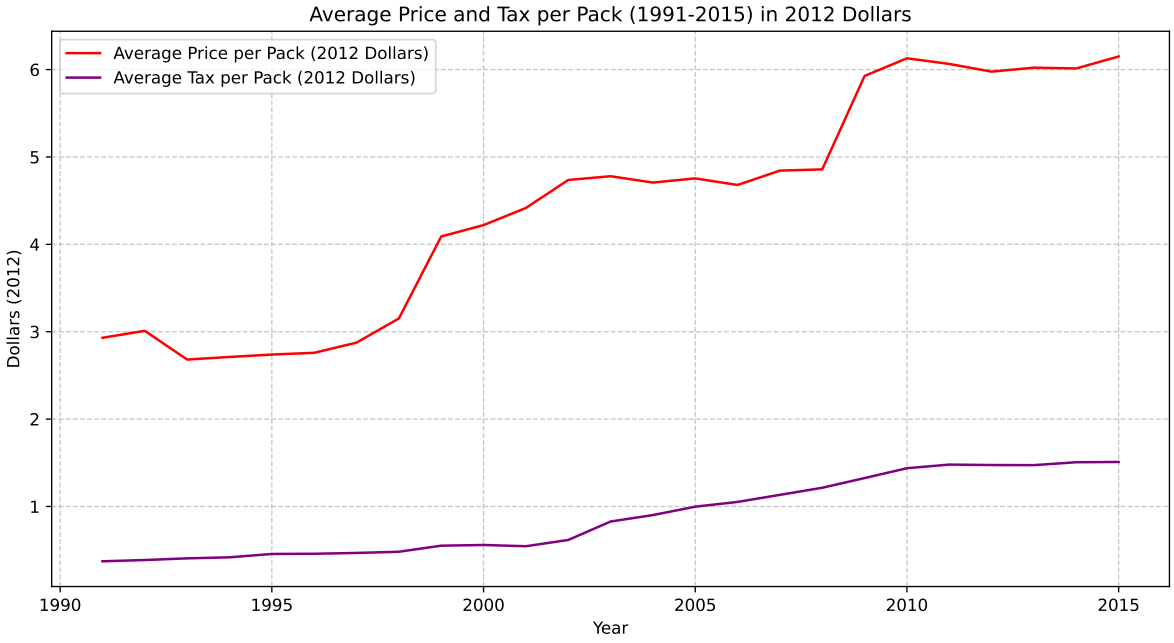
Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

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

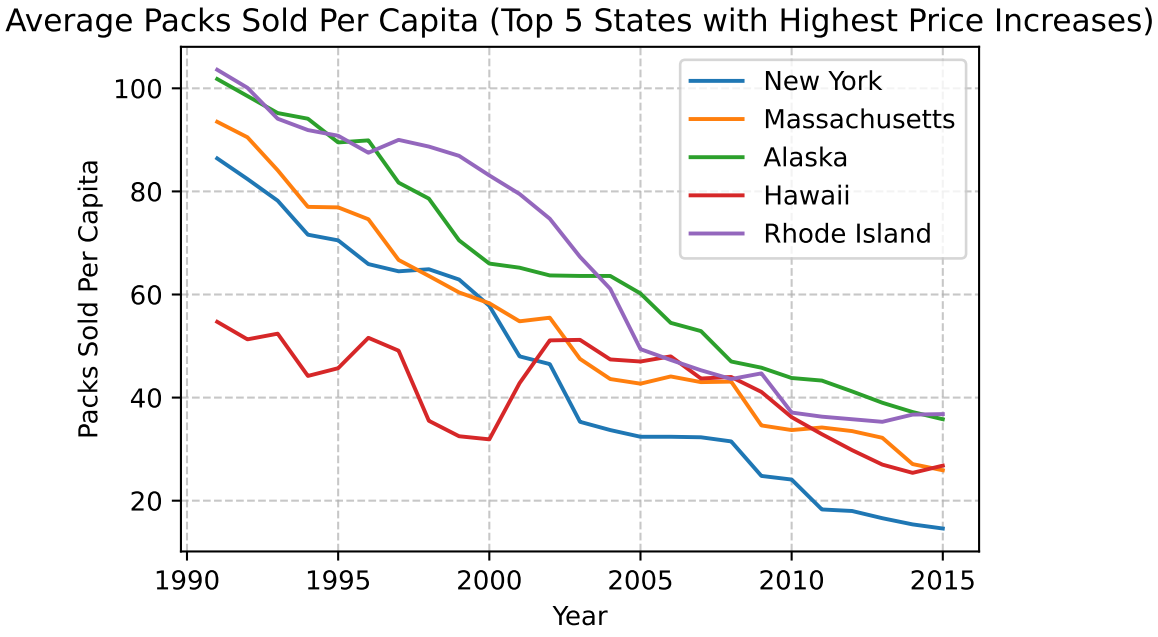
Question 1



Question 2


Average Price and Tax per Pack (1991-2015) in 2012 Dollars
Average Price per Pack (2012 Dollars)
Average Tax per Pack (2012 Dollars)
Dollars (2012)
Year
1990
1995
2000
2005
2010
2015
1
2
3
4
5
6

Question 3


Average Packs Sold Per Capita (Top 5 States with Highest Price Increases)
New York
Massachusetts
Alaska
Hawaii
Rhode Island
Packs Sold Per Capita
Year
1990
1995
2000
2005
2010
2015
20
40
60
80
100

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

```
=== Price Elasticity Estimate for 1970-1990 ===
                            OLS Regression Results
==============================================================================
Dep. Variable:              log_sales   R-squared:                       0.294
Model:                            OLS   Adj. R-squared:                  0.293
Method:                 Least Squares   F-statistic:                     445.1
Date:                Tue, 18 Mar 2025   Prob (F-statistic):           6.98e-83
Time:                        20:10:33   Log-Likelihood:                 263.40
No. Observations:                1071   AIC:                            -522.8
Df Residuals:                    1069   BIC:                            -512.8
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          5.3854      0.028    193.692      0.000       5.331       5.440
log_price     -0.8094      0.038    -21.098      0.000      -0.885      -0.734
==============================================================================
Omnibus:                       89.160   Durbin-Watson:                   0.183
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              466.536
Skew:                           0.128   Prob(JB):                    4.93e-102
Kurtosis:                       6.223   Cond. No.                         10.0
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
Estimated Price Elasticity: -0.809

=== Price Elasticity Estimate for 1991-2015 ===
                            OLS Regression Results
==============================================================================
Dep. Variable:              log_sales   R-squared:                       0.561
Model:                            OLS   Adj. R-squared:                  0.561
Method:                 Least Squares   F-statistic:                     1630.
Date:                Tue, 18 Mar 2025   Prob (F-statistic):          4.20e-230
Time:                        20:10:33   Log-Likelihood:                -256.00
No. Observations:                1275   AIC:                             516.0
Df Residuals:                    1273   BIC:                             526.3
```

```
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          5.6083      0.035    159.600      0.000       5.539       5.677
log_price     -0.9968      0.025    -40.370      0.000      -1.045      -0.948
==============================================================================
Omnibus:                       23.003   Durbin-Watson:                   0.208
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               43.688
Skew:                           0.011   Prob(JB):                     3.26e-10
Kurtosis:                       3.907   Cond. No.                         8.90
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
Estimated Price Elasticity: -0.997

=== Elasticity Comparison ===
Elasticity (1970-1990): -0.809
Elasticity (1991-2015): -0.997
Difference: 0.187
```

Compared to my elasticity estimated for 1970-1990, my estimated price elasticity of demand was -0.997 for 1991-2015. The demand became more elastic in 1991-2015. This suggests consumers are more price-sensitive, though demand is still inelastic (absolute value is less than 1). This could mean that while price matters more in purchasing decisions, other factors (like addiction or brand loyalty) still play a role.