Homework 3 - Submission 3

ECON 470

Baran Pasa

Homework 3 Summary Statistics and ATE Analysis

Link to Github

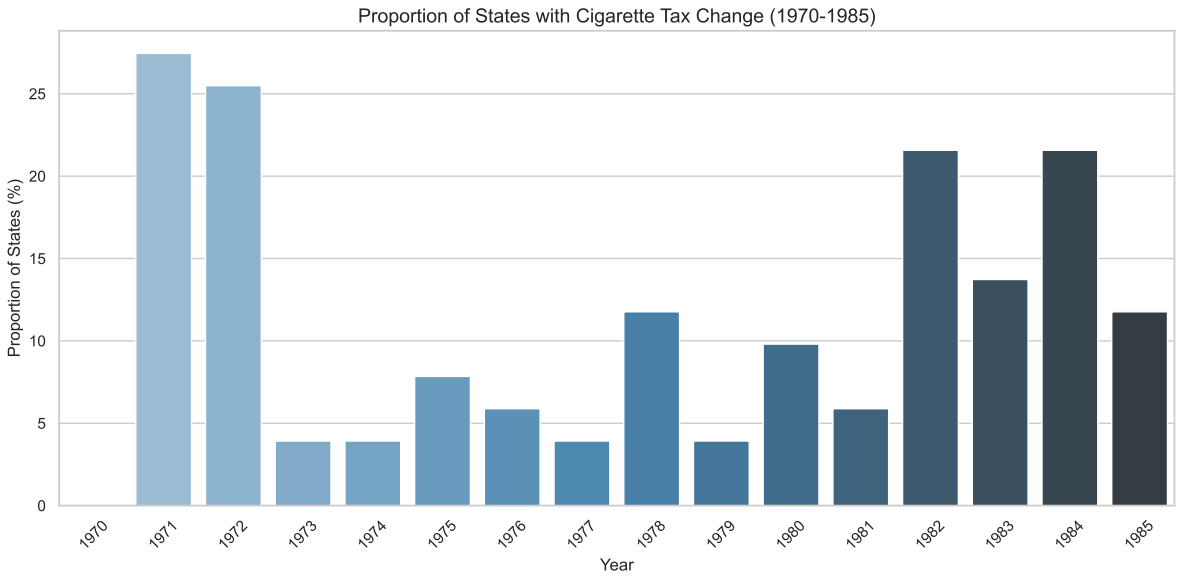
Collecting and Cleaning Data

CDC Tax Burden on Tobacco Data was collected from a provided repository and inflation data was collected from the BLI database. Raw data was downloaded and then put into real dollars using 2012 as the base year.

1. Summarizing the Data

1. Proportion of States with a change in their cigarette tax each year from 1970 to 1985

As seen in the figure below, there is a large spike in proportion of states at 1983. It is possible that this conincides with a major policy shift on cigarettes.

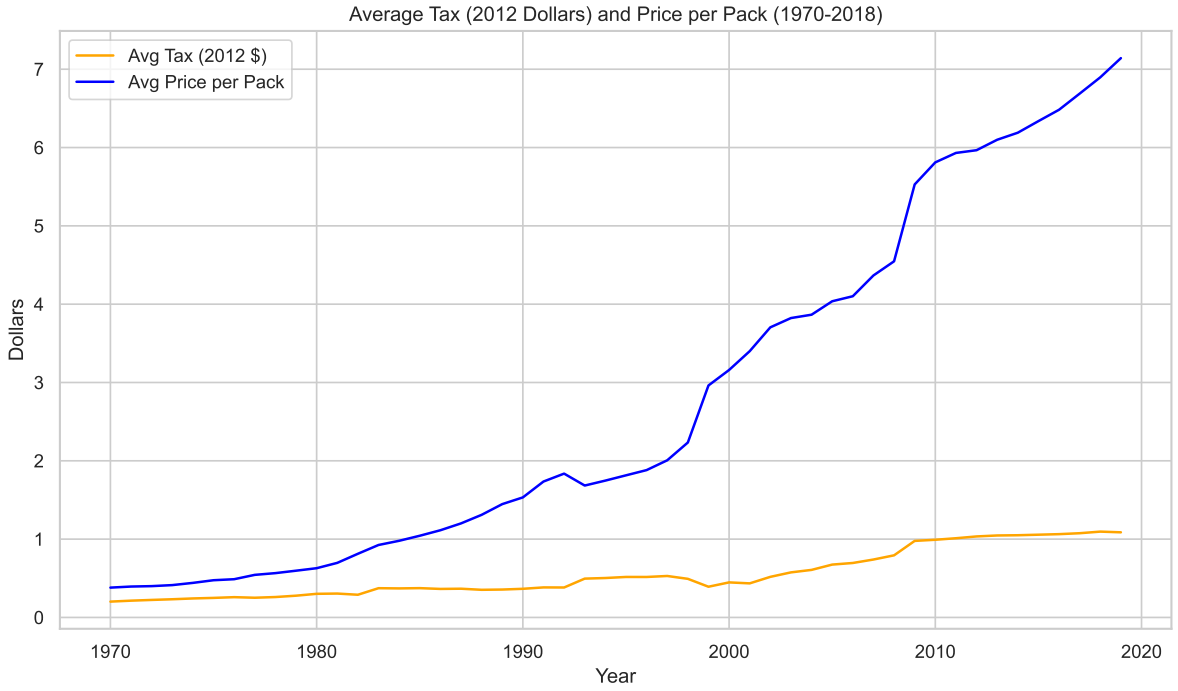


Update for submission 2:

A fix in the data.py code made for this bar graph to accurately display the data.

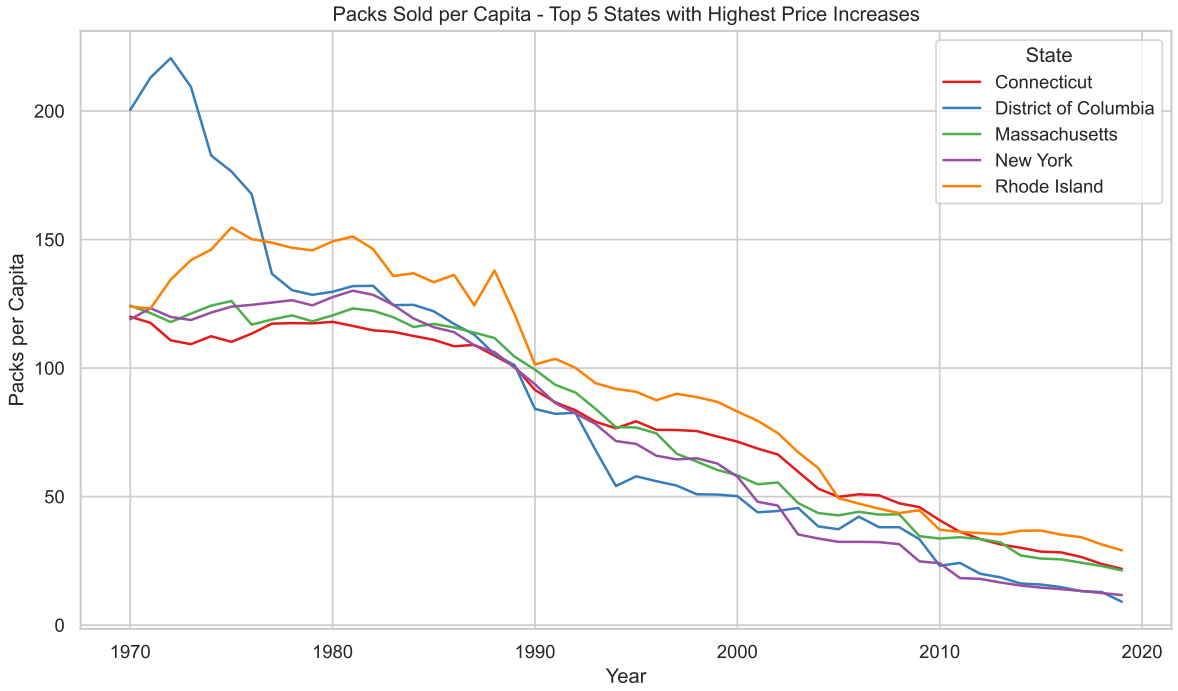
2. Average Price per Pack, 1970-2018 (in 2012 dolalrs)

While both tax and price per pack have risen in past years, price per pack has grown exponentially more. This makes sense as the government has tried to use economic disincentives to sway people away from cigarettes.



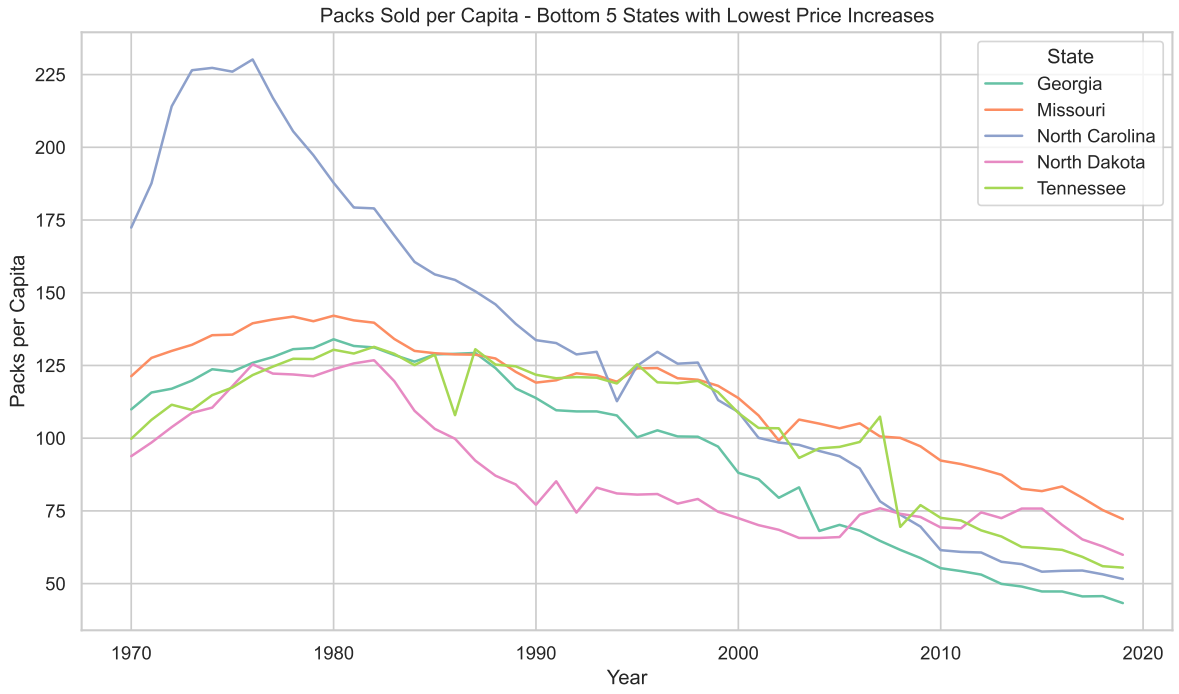
3. Top 5 states with highest increases in cigarette prices

In the top 5 states with the highest price increases, there has been a significant decrease in the number of packs per person. This might point to some movement in the right direction for policy makers.



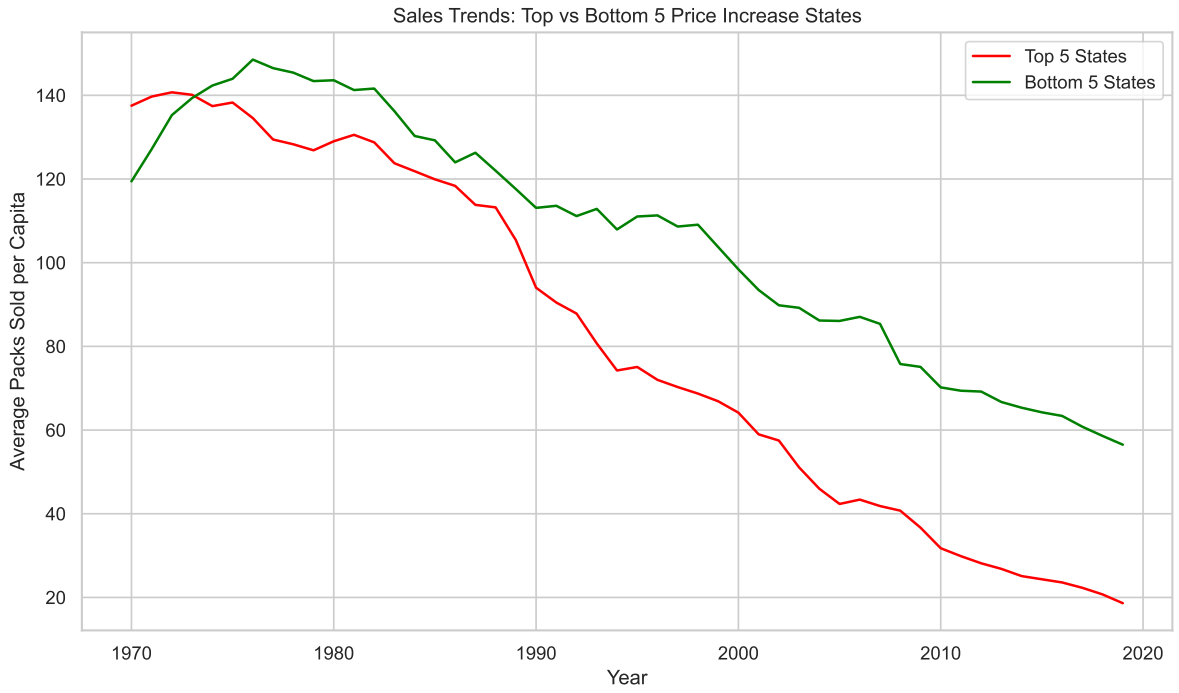
4. Top 5 states with lowest increases in cigarette prices

Decline in the top 5 states with the lowest increases mimic the previous graph, just a higher number of packs per person. While they directionally are comaprable, their scales are different.



5. Comparison of trends between these 10 states:

Contrasting the 10 states, it is easier to see the differences between the states with the highest taxes and those with the lowest. These graphs do show some indication that increasing the tax burden might decrease the number of packs per person, but there are many outside variables that have not been taken into account.



ATE

For submission 2, I switched over to the pyfixest model, making it much easier to run my regressions.

```
Unable to display output for mime type(s): text/html
```

```
Unable to display output for mime type(s): text/html
```

6.

	log_sales (1)
coef	
log_price	-0.809***
	(0.038)
Intercept	5.427***
	(0.030)
stats	
Observations	1071
S.E. type	iid
R^2	0.294

Significance levels: * p < 0.05, ** p < 0.01, *** p < 0.001. Format of coefficient cell: Coefficient (Std. Error)

With a log_price coefficient of -0.172, the regression indicates that for every 1 unit increase of price, sales decrease by -0.172 units. This does make sense, as an increase in price should decrease the demand.

7.

	log_sales (1)	(2)
coef		
log_price	-0.809***	-0.867***
	(0.038)	(0.116)
Intercept	5.427***	5.471***
	(0.030)	(0.089)

	log_sales	
	(1)	(2)
stats		
Observations	1071	1071
S.E. type	iid	iid
R^2	0.294	-

Significance levels: * p < 0.05, ** p < 0.01, *** p < 0.001. Format of coefficient cell: Coefficient (Std. Error)

8.

	log_sales
	(1)
coef	
tax_dollar	-1.127***
	(0.065)
Intercept	5.101***
	(0.018)
stats	
Observations	1071
S.E. type	iid
R^2	0.217

Significance levels: * p < 0.05, ** p < 0.01, *** p < 0.001. Format of coefficient cell: Coefficient (Std. Error)

9.

	log_sales			
	(1)	(2)	(3)	(4)
coef				
log_price	-0.809***	-0.867***	-0.997***	-1.169***
	(0.038)	(0.116)	(0.025)	(0.042)
Intercept	5.427***	5.471***	5.660***	5.908***
	(0.030)	(0.089)	(0.036)	(0.061)
stats				
Observations	1071	1071	1275	1275
S.E. type	iid	iid	iid	iid

	log_sales			
	(1)	(2)	(3)	(4)
R^2	0.294	-	0.561	-

Significance levels: * p < 0.05, ** p < 0.01, *** p < 0.001. Format of coefficient cell: Coefficient (Std. Error)

10. Further Interpretations