Homework 3
Research Methods, Spring 2024

Taeyoung Yim

```
cig.data <- readRDS('data/output/TaxBurden_Data.rds')

cig.data <- cig.data %>% group_by(state) %>% arrange(state, Year) %>%
  mutate(tax_change = tax_state - lag(tax_state),
  tax_change_d = ifelse(tax_change == 0,0,1),
  price_cpi_2012 = cost_per_pack*(218/index),
  total_tax_cpi_2012 = tax_dollar*(218/index),
  ln_tax_2012 = log(total_tax_cpi_2012),
  ln_sales = log(sales_per_capita),
  ln_price_2012 = log(price_cpi_2012)
)
```

## 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()`).
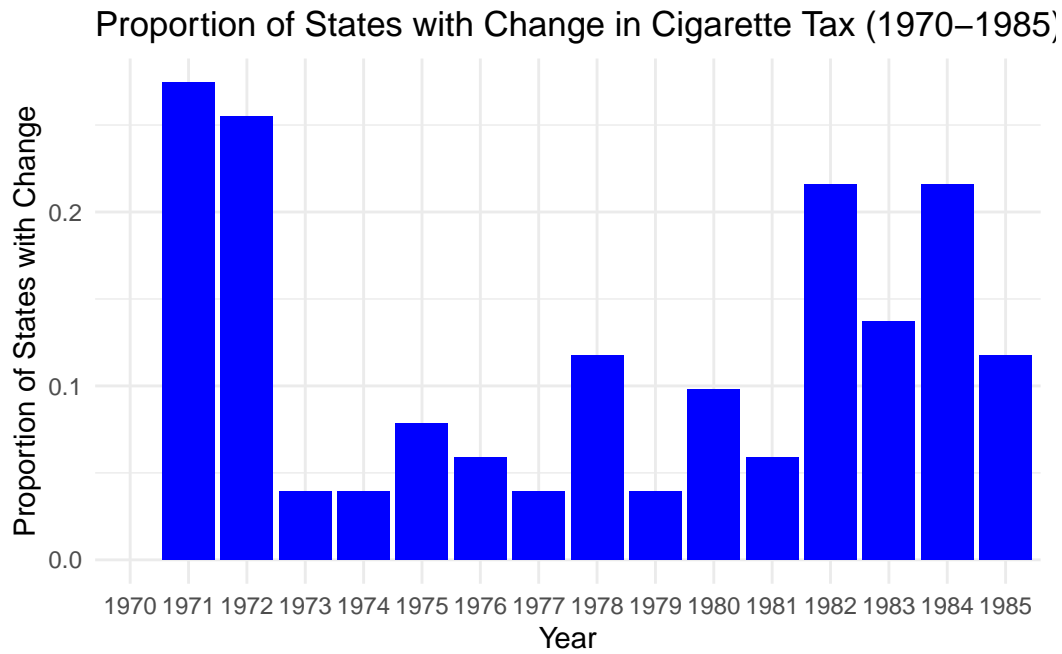
2. Plot on a single graph the average tax (in 2012 dollars) on cigarettes and the average price of a pack of cigarettes from 1970 to 2018.
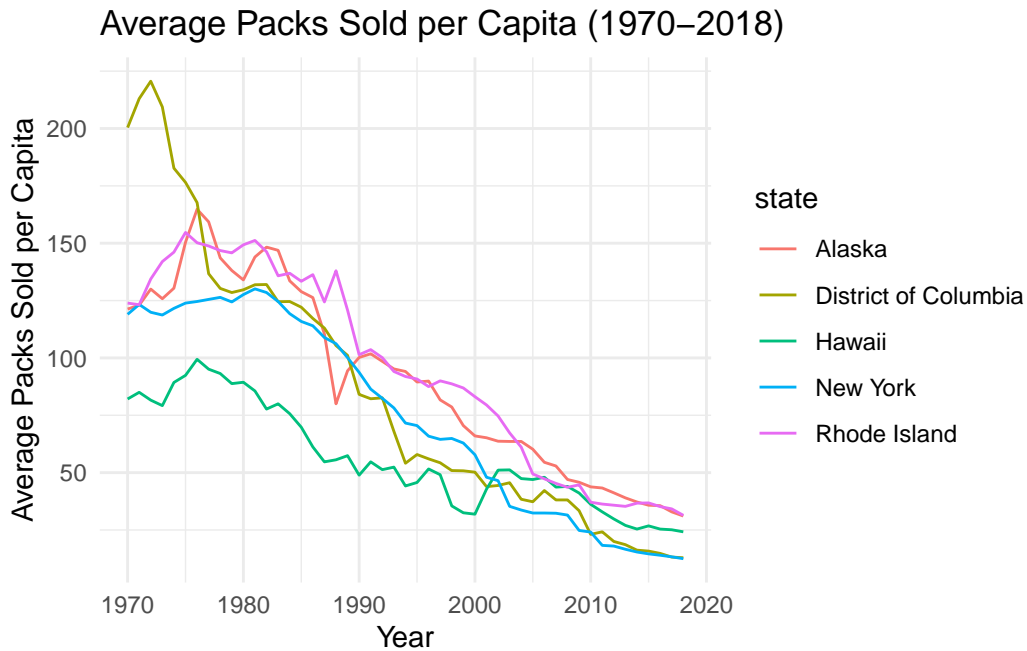
## Average Tax and Price of Cigarettes (1970–2018)
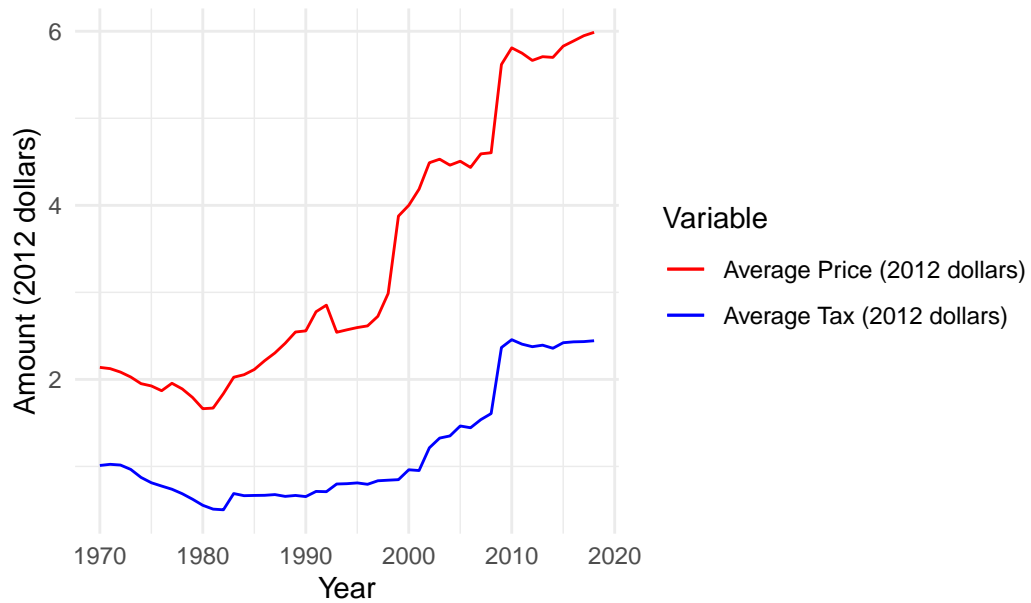
3.Identify the 5 states with the highest increases in cigarette prices (in dollars) over the time period. Plot the average number of packs sold per capita for those states from 1970 to 2018.

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



Average Packs Sold per Capita (1970–2018)

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 for States with Lowest Price In
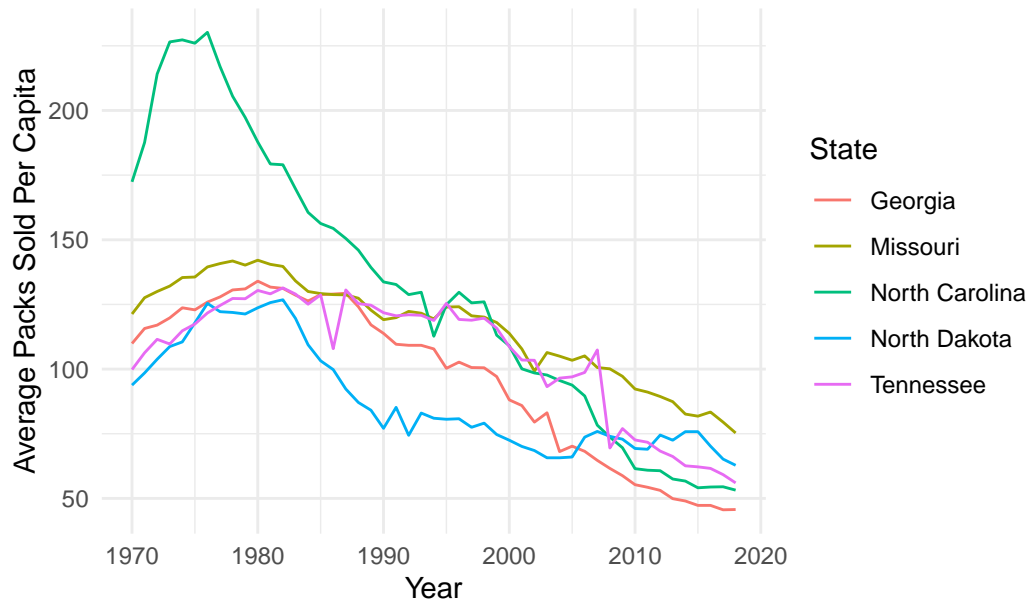
5.Compare the trends in sales from the 5 states with the highest price increases to those with the lowest price increases.

## Comparison of Sales Trends between Top 5 and Bottom 5 Stat
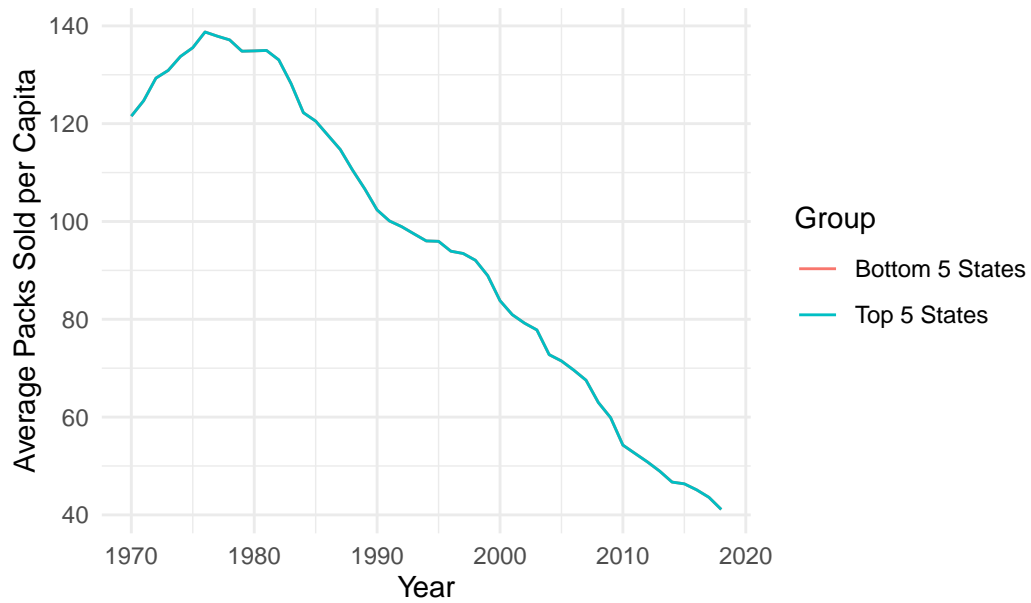
## Estimate ATEs

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.

```
Call:
lm(formula = log(sales_per_capita) ~ log(price_cpi_2012), data = filtered_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)   5.38544   0.02780   193.7   <2e-16 ***
log(price_cpi_2012) -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

`geom_smooth()` using formula = 'y ~ x'
```
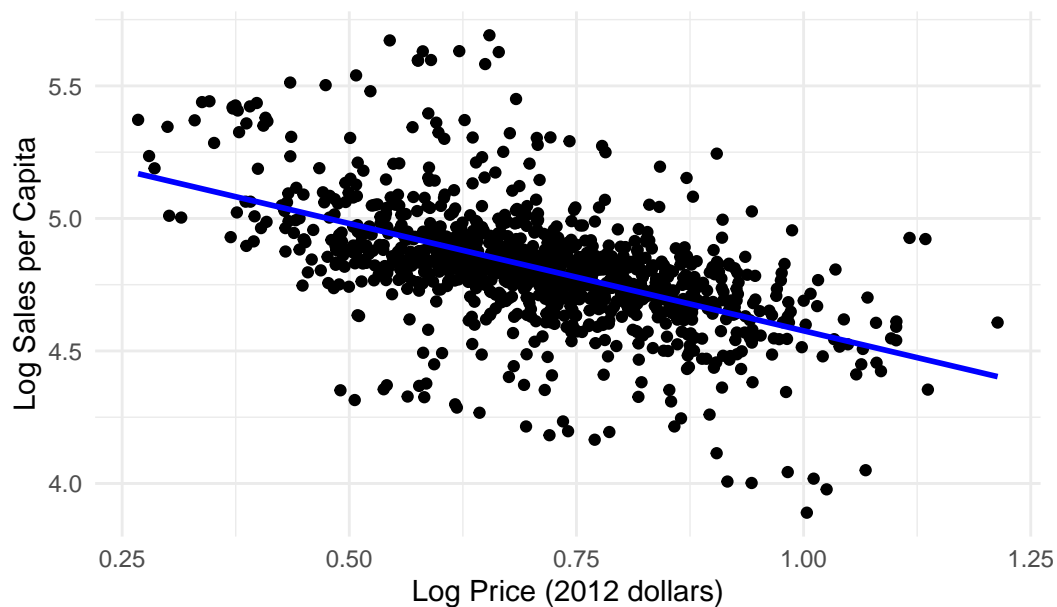
## Price Elasticity of Demand for Cigarettes (1970–1990)



```
# A tibble: 2 x 5
  term               estimate std.error statistic  p.value
  <chr>                 <dbl>     <dbl>     <dbl>    <dbl>
1 (Intercept)            5.39    0.0278     194.   0
2 log(price_cpi_2012)   -0.809   0.0384     -21.1  6.98e-83
```

The coefficient of log(price_cpi_2012) is 5.385438 -0.8094384 with a p-value of 0 6.981617e-8

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?

```
Attaching package: 'plm'

The following objects are masked from 'package:dplyr':

    between, lag, lead

Oneway (individual) effect Within Model
Instrumental variable estimation

Call:
plm(formula = log(sales_per_capita) ~ log(price_cpi_2012) | log(total_tax_cpi_2012),
    data = filtered_data, model = "within")

Balanced Panel: n = 51, T = 21, N = 1071

Residuals:
      Min.    1st Qu.     Median    3rd Qu.       Max.
-0.4595449 -0.0366605  0.0056037  0.0456071  0.4118392

Coefficients:
                    Estimate Std. Error z-value Pr(>|z|)
log(price_cpi_2012) -0.326532   0.068545 -4.7638  1.9e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    13.476
Residual Sum of Squares: 7.6596
R-Squared:      0.60197
Adj. R-Squared: 0.58204
Chisq: 22.6935 on 1 DF, p-value: 1.9001e-06
```

The instrumental variable (IV) regression results indicate a statistically significant inverse relationship between cigarette prices (adjusted for inflation) and cigarette sales per capita. This suggests that as cigarette prices increase, sales per capita decrease. The use of the total cigarette tax as an instrument for cigarette prices helps address potential endogeneity issues, leading to more reliable estimates compared to ordinary least squares (OLS) regression.

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

```
Call:
lm(formula = log(price_cpi_2012) ~ log(total_tax_cpi_2012), data = filtered_data)

Residuals:
     Min       1Q   Median       3Q      Max
-0.23046 -0.09207 -0.02919  0.08019  0.48675

Coefficients:
                         Estimate Std. Error t value Pr(>|t|)
(Intercept)              0.801304   0.005889   136.1   <2e-16 ***
log(total_tax_cpi_2012)  0.260060   0.012443    20.9   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1272 on 1069 degrees of freedom
Multiple R-squared:  0.2901,    Adjusted R-squared:  0.2894
F-statistic: 436.8 on 1 and 1069 DF,  p-value: < 2.2e-16
```

```
Call:
lm(formula = log(sales_per_capita) ~ log(total_tax_cpi_2012),
    data = filtered_data)

Residuals:
     Min       1Q   Median       3Q      Max
-0.86239 -0.09798  0.00549  0.09359  0.95094

Coefficients:
                         Estimate Std. Error t value Pr(>|t|)
(Intercept)              4.738119   0.009995 474.041   <2e-16 ***
log(total_tax_cpi_2012) -0.206884   0.021119  -9.796   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2159 on 1069 degrees of freedom
Multiple R-squared:  0.08238,   Adjusted R-squared:  0.08152
F-statistic: 95.97 on 1 and 1069 DF,  p-value: < 2.2e-16
```
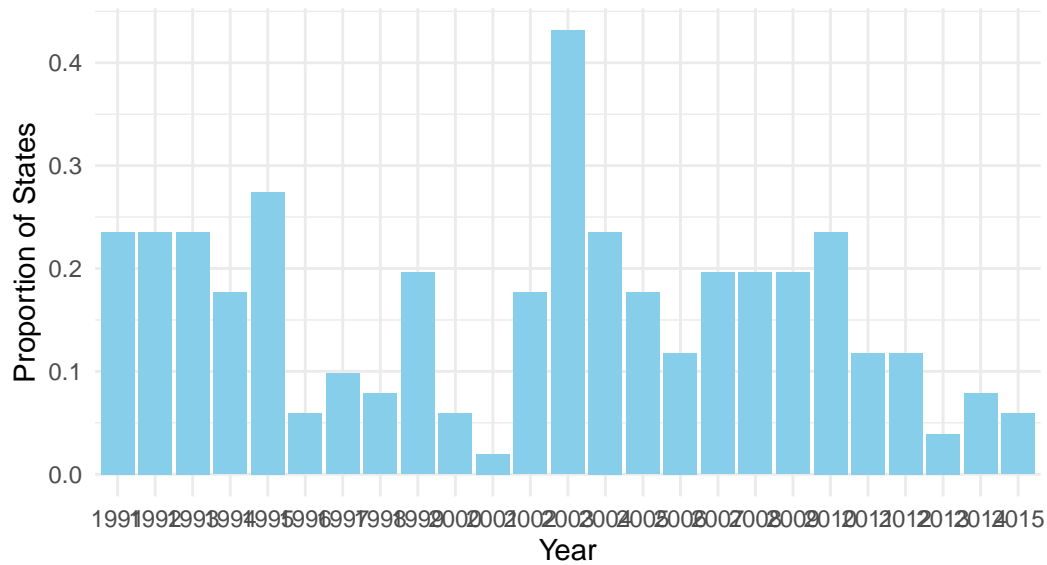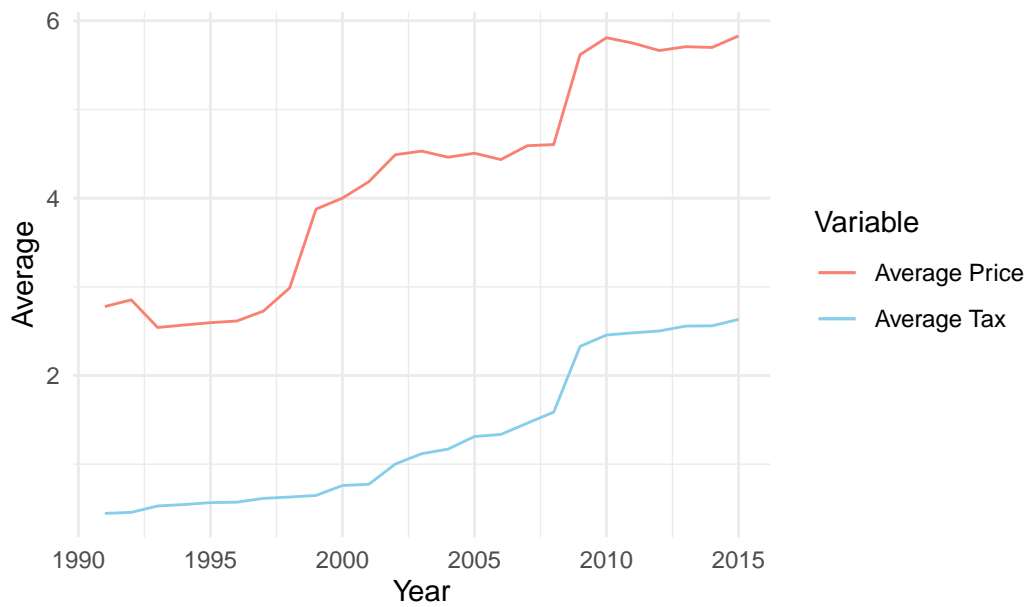
9.Repeat questions 1-3 focusing on the period from 1991 to 2015.

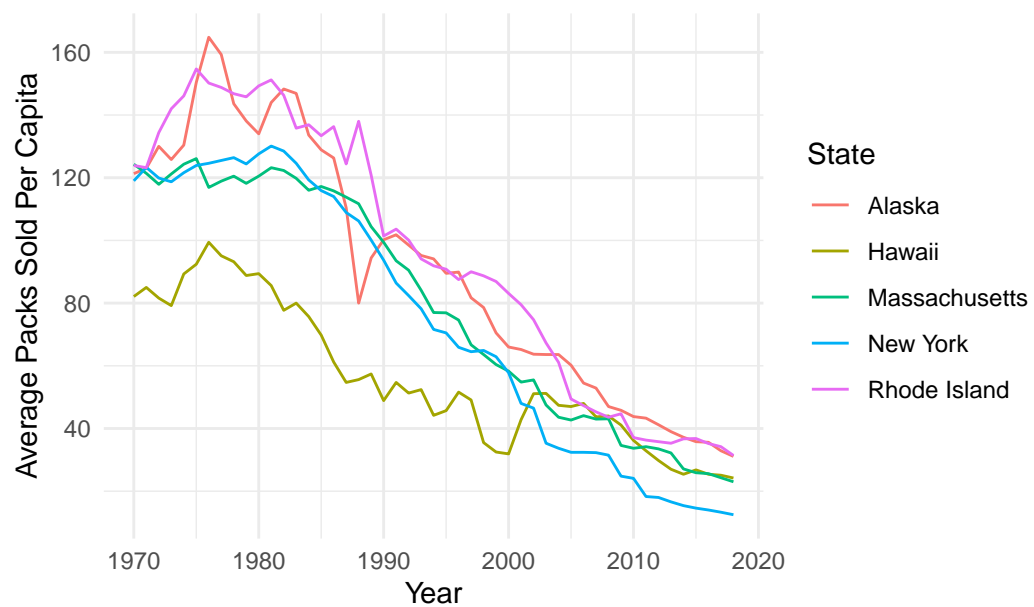## Proportion of States with Cigarette Tax Change (1991–2015)

Data Source: CDC

Average Tax and Price of Cigarettes (1991–2015)

Average Packs Sold Per Capita for States with Highest Price Ir

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


Elasticity estimates differ between the periods 1970-1990 and 1991-2015.
The difference in elasticity estimates is significant.
Possible reasons for the difference may include changes in consumer preferences, policy inter
```