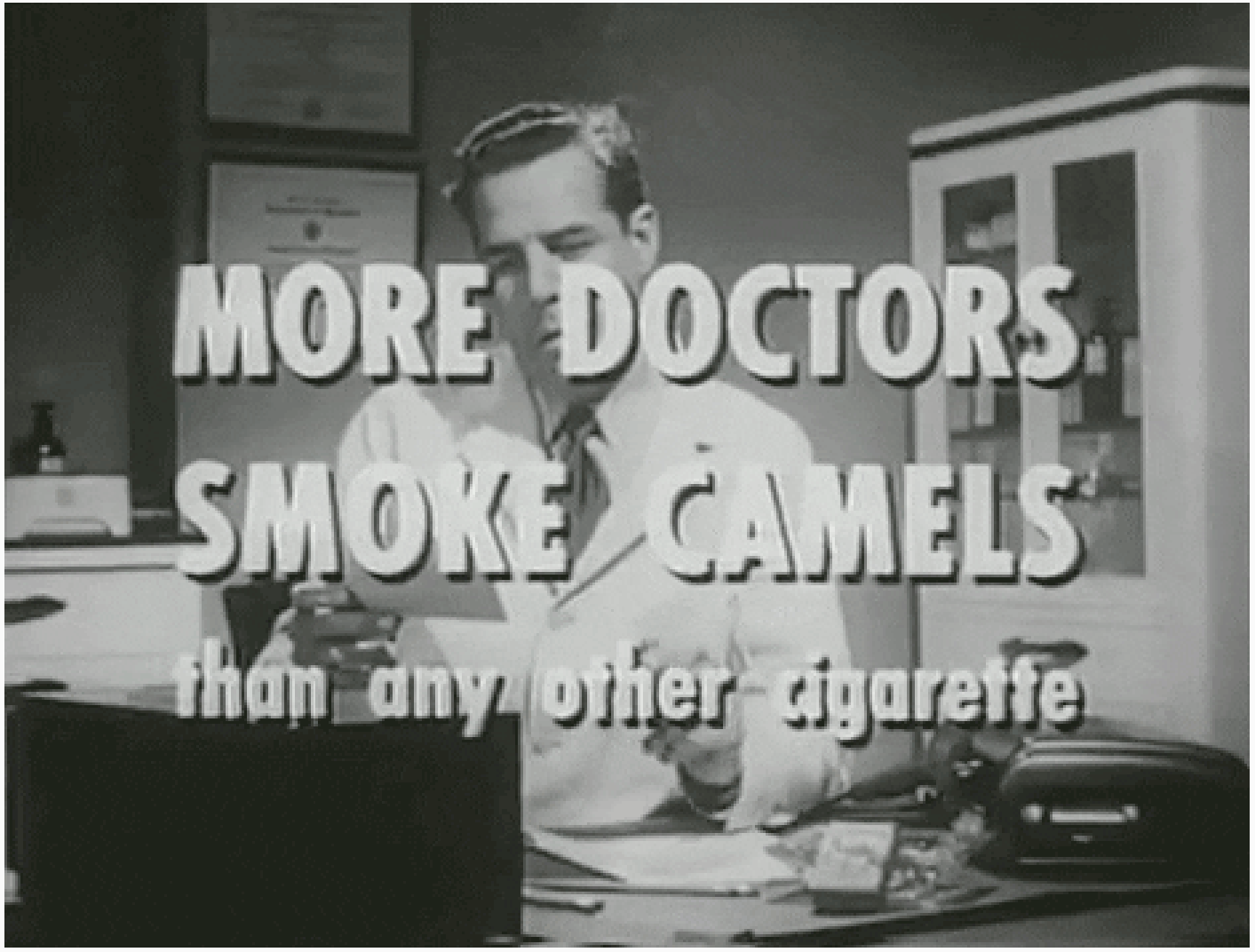
Module 2: Demand for Cigarettes and Instrumental Variables

Part 1: CDC Data on Smoking and Cigarettes

Ian McCarthy | Emory University
Econ 470 & HLTH 470

History of Smoking



History of Smoking

- Widespread smoking began in late 1800s
- Lung cancer becoming more common after 1930s
- First evidence of link in 1950s
- Surgeon general's report in 1964
- Very important in causal inference! (Section 5.1.1 of Causal Inference Mixtape)

Why it matters

1. Extreme public health concerns
    - Lung cancer prevalence
    - Fetal and baby health
2. Economic questions
    - Is it an information problem?
    - Externalities (second-hand smoke)
    - Moral hazard due to insurance

In our case

We want to focus on estimating demand for cigarettes. By this, I mean estimating price elasticity of demand.

We'll show that standard OLS isn't going to do this very well.

Cigarette Data

The Data

- Data from CDC Tax Burden on Tobacco
- Visit GitHub repository for other info: Tobacco GitHub repository
- Supplement with CPI data, also in GitHub repo.

Summary stats

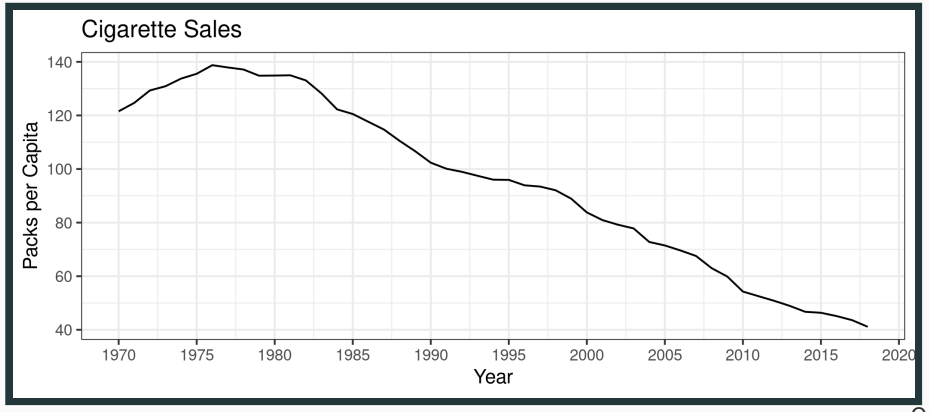
We're interested in cigarette prices and sales, so let's focus our summaries on those two variables

```
stargazer(as.data.frame(cig.data %>% select(sales_per_capita, price_cpi, cost_per_pack)), type="html")
```

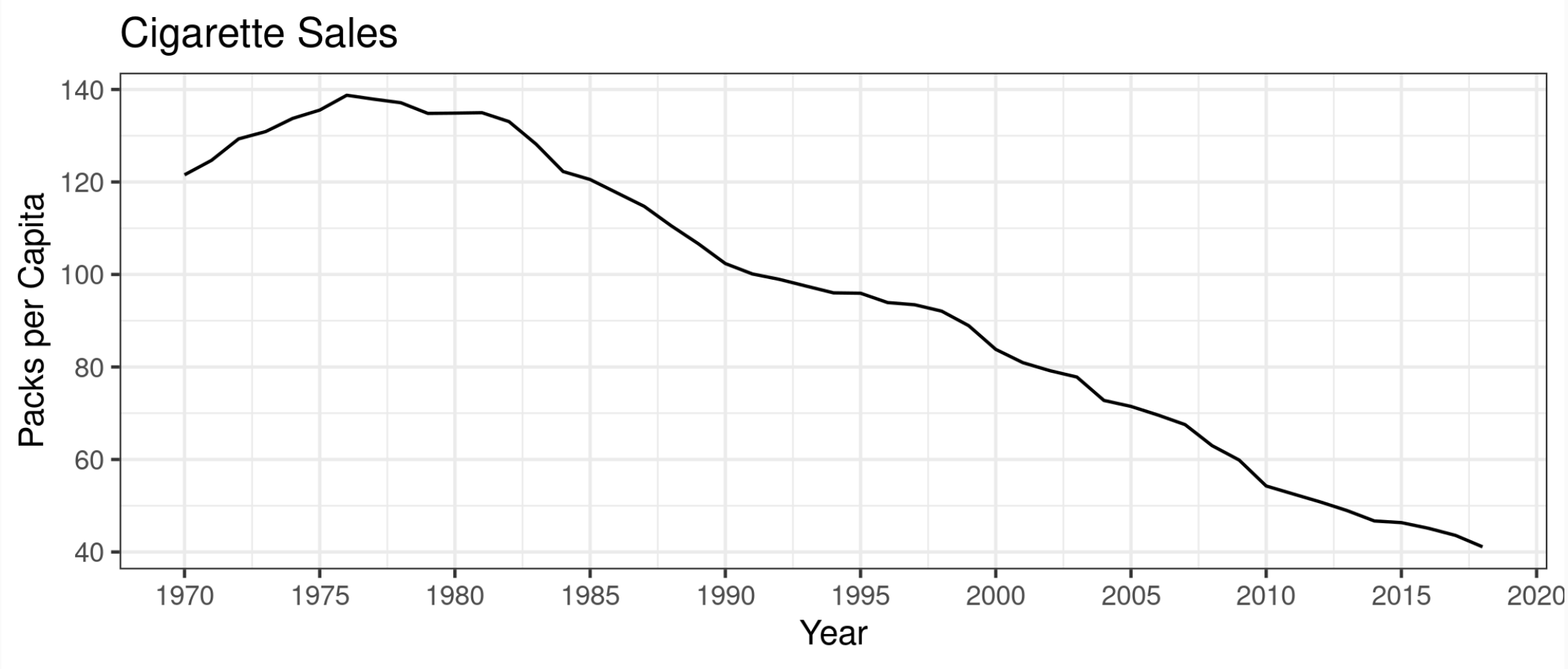
Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
sales_per_capita	2,499	95.150	41.133	12.500	63.050	122.400	296.200
price_cpi	2,499	3.396	1.641	1.307	2.088	4.520	9.651
cost_per_pack	2,499	2.678	2.238	0.287	0.780	4.237	10.376

Cigarette Sales

```
cig.data %>%
  ggplot(aes(x=Year,y=sales_per_capita)) +
  stat_summary(fun.y="mean",geom="line") +
  labs(
    x="Year",
    y="Packs per Capita",
    title="Cigarette Sales"
  ) + theme_bw() +
  scale_x_continuous(breaks=seq(1970, 2020, 5))
```

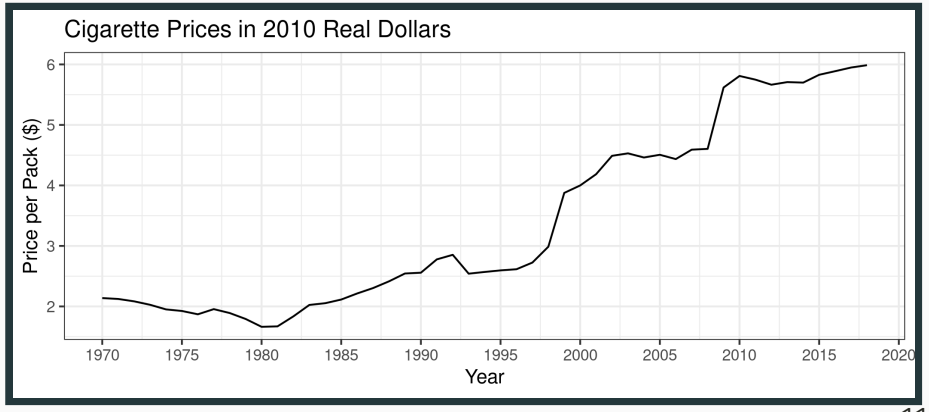
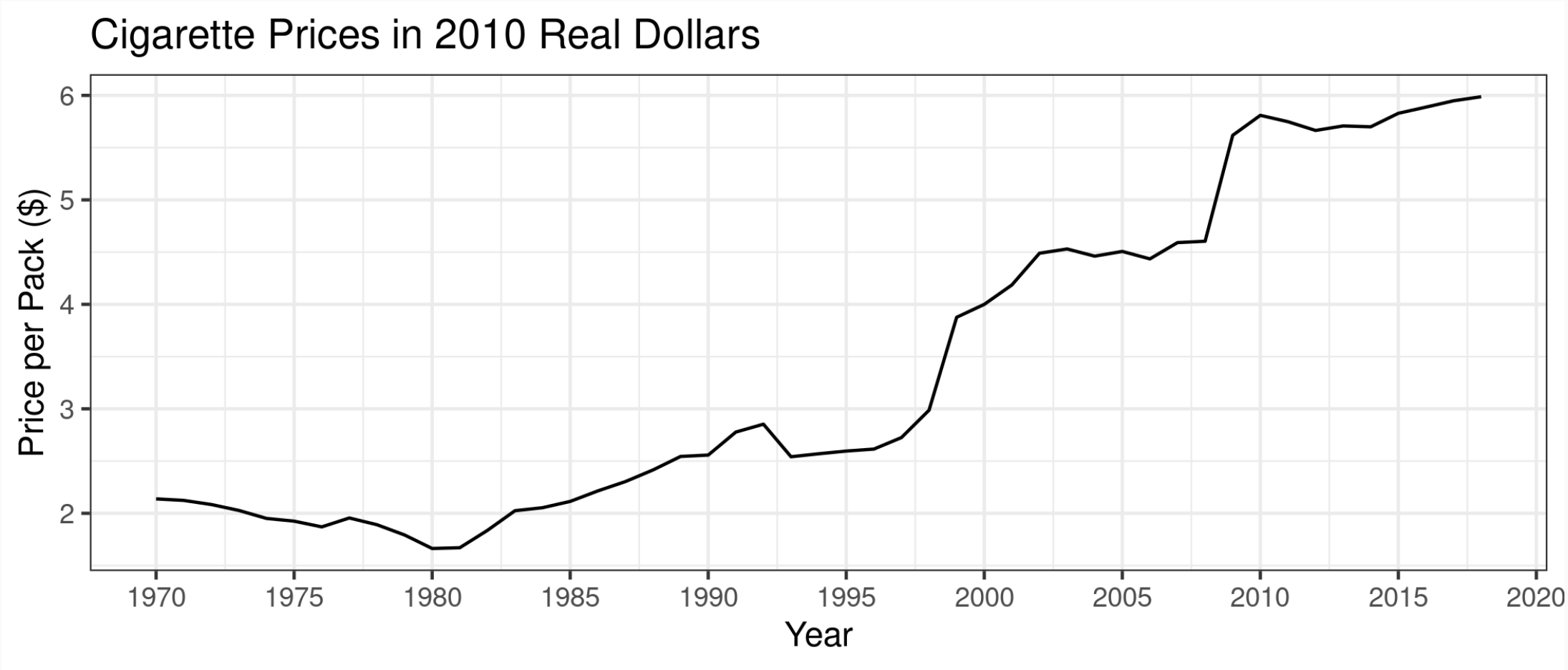


Cigarette Sales



Cigarette Prices

```
cig.data %>%
  ggplot(aes(x=Year,y=price_cpi)) +
  stat_summary(fun.y="mean",geom="line") +
  labs(
    x="Year",
    y="Price per Pack ($)",
    title="Cigarette Prices in 2010 Real Dollars"
  ) + theme_bw() +
  scale_x_continuous(breaks=seq(1970, 2020, 5))
```



Cigarette Prices