Lecture 4: Environmental Econometrics

Prof. Theising
Environmental Economics
Econ 4075

Moving into Module 2: Measuring Environmental Benefits

- Recall from previous lectures: economists' goal is to maximize social net benefits from environment
  - Policy can correct market failures, but to do so, must measure and compare costs & benefits

- Professional economists spend relatively more time and effort quantifying and monetizing environmental benefits than costs
  - **Any ideas why?**

Why Do Economists use Econometrics?

- Economic framework + Tricks from statistics = **econometrics**
  - Statistical toolkit applied to questions related to problems and phenomena within economics.

Why Do Economists use Econometrics?

- Economic framework + Tricks from statistics = **econometrics**
  - Statistical toolkit applied to questions related to problems and phenomena within economics.

- Often, econometrics has a specific focus on *causality*
  - Why such a focus on causality? Correlations are important, too…

Why Do Economists use Econometrics?

- Economic framework + Tricks from statistics = **econometrics**
  - Statistical toolkit applied to questions related to problems and phenomena within economics.

- Often, econometrics has a specific focus on *causality*
  - Why such a focus on causality? Correlations are important, too…
  - We want to ***know*** with some degree of confidence that an intervention/policy etc. affects an outcome. Resources are constrained and interventions cost money!

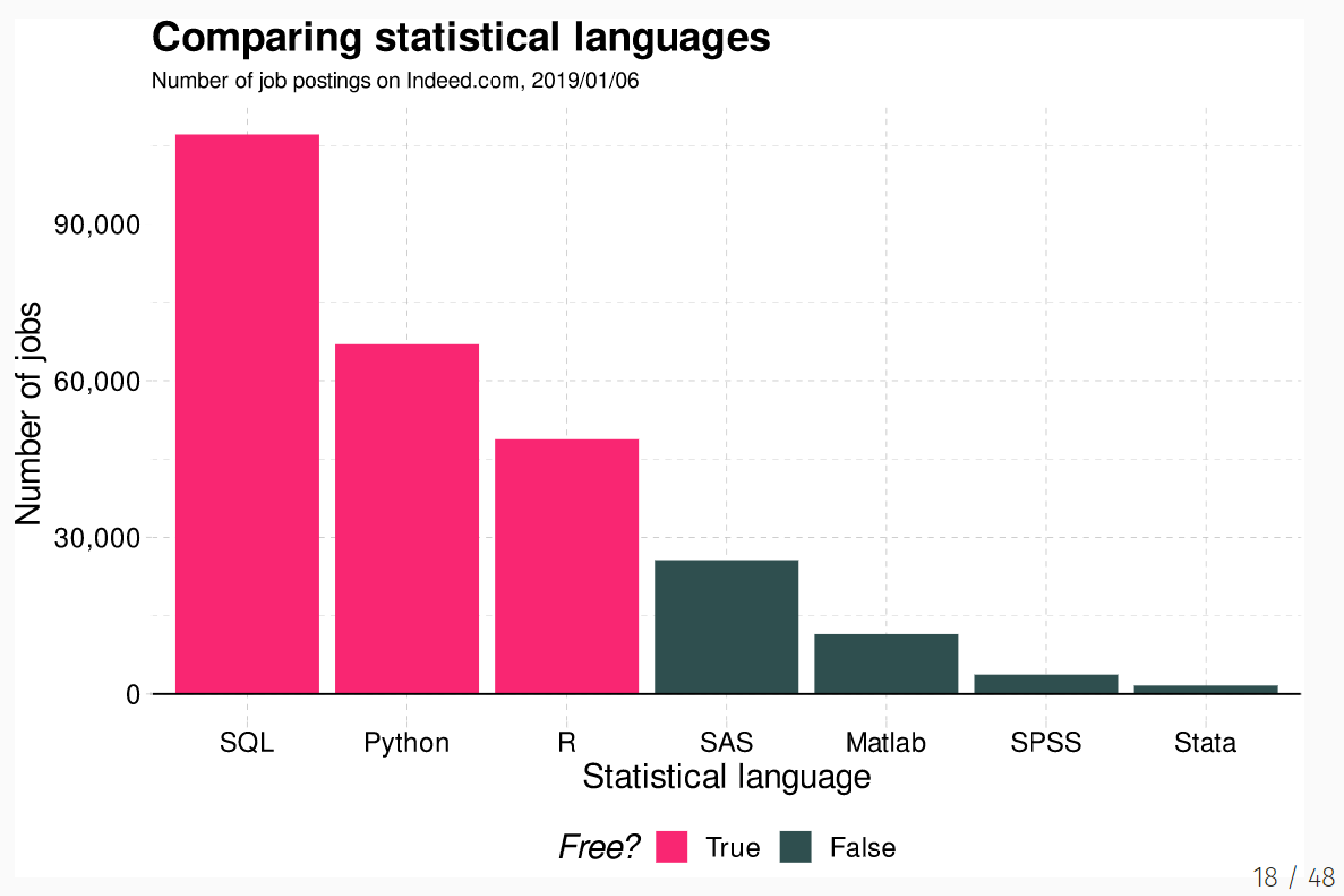
- Other times, we may formally model consumers' utility functions (to measure willingness-to-pay) or firms' profit functions (to measure marginal costs)
  - Added structure of the model: counterfactual analysis possible

Econometrics, in General

- Experience with econometrics can get you a job! A really good job (where you could improve business decisions), a good pre-doc or research assistantship, etc.

- There are amazing resources out there.
  - Dr. Ed Rubin's undergraduate econometrics course, machine learning for economists course, and graduate level econometrics course are open source and taught in R.
  - Dr. Scott Cunningham's Substack, Journal of Human Resources threads, Podcast, and Causal Inference: The Mixtape book
  - Dr. Nick Huntington-Klein's website has a TON of resources on econometrics, methods, causality, and examples with replication
  - Dr. Patty Champ et al. A Primer in Nonmarket Valuation (PDF version available here for free, and Amazon link here)

An aside:



[Source]

When estimating the benefits of a policy intervention, we want to know the causal effect of the policy on the outcome of interest. In this simple representation, δ would be the effect of cleanup on Y.

$$Y_i = \boldsymbol{X}\beta + \delta Cleanup_i + \phi_i + \varepsilon_i$$

Imagine a policy being proposed to clean up the Cuyahoga River, what are possible outcomes of interest?

Source: [history.com]

$Y =$

When estimating the benefits of a policy intervention, we want to know the causal effect of the policy on the outcome of interest. In this simple representation, δ would be the effect of cleanup on Y.

$$Y_i = \boldsymbol{X}\beta + \delta Cleanup_i + \phi_i + \varepsilon_i$$

Imagine a policy being proposed to clean up the Cuyahoga River, what are possible outcomes of interest?

Source: history.com

$Y_i \equiv$

1. Water Quality Index
2. Ecological Health
3. Human Health
4. Pr(lighting the river on fire)

Whatever the outcome might be, we want to know that δ is the best unbiased estimate of the effect of $cleanup$ on Y.

A stylized example (…*and review of regression*)

- Let’s build a simple model of students’ performance in college

- Cumulative college GPA is a function of students’ ability, hours studied (H), and let’s add **lifetime lead pollution exposure** (P)

$$GPA = f(I, SAT, H, P)$$

- We do not directly observe ability, but can proxy for it using students’ SAT scores and family income (I)

- We know GPA will increase as a function of some variables, decrease with others. **What is the expected marginal impact of lead exposure?**

Example (continued…)

- As a first pass, let's write down a linear regression model of GPA production:

$$GPA_i = \beta_0 + \beta_1 I_i + \beta_2 SAT_i + \beta_3 H_i + \beta_4 P_i + \varepsilon_i$$

- Left hand side: dependent variable (GPA)

- Right hand side: independent variables (I, SAT, H, P) and error term (ε_i)

- **How does lead pollution affect GPA here? What parameter do we care about?**

Example (continued…)

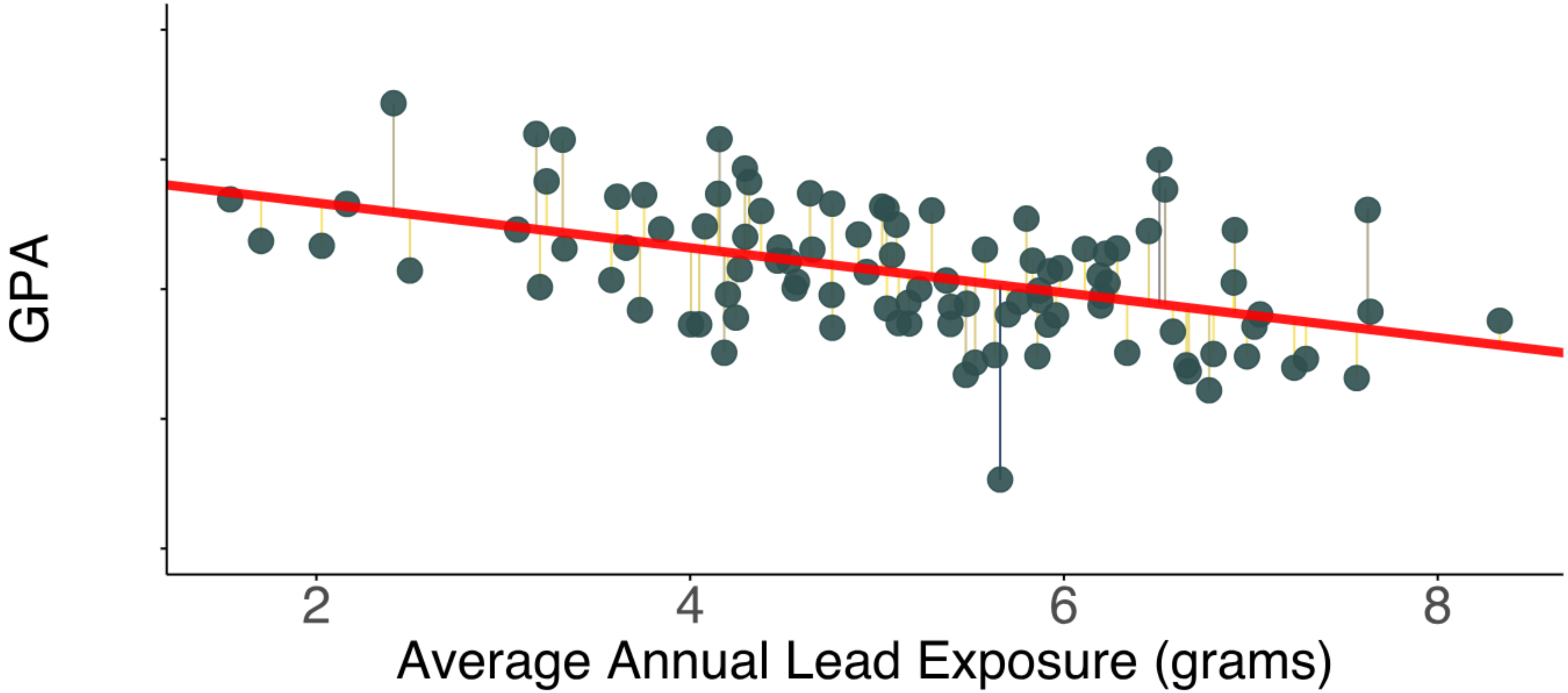
$$GPA_i = \beta_0 + \beta_1 I_i + \beta_2 SAT_i + \beta_3 H_i + \beta_4 P_i + \varepsilon_i$$

- We want to know how pollution affects GPA
- In a linear regression model, $\beta_4 = \frac{\partial GPA_i}{\partial P_i}$
- In words: "on average, for a 1 unit increase in lead pollution, students' GPA will change by β_4."
- How do we estimate $\widehat{\beta_4}$?

Example (continued…)

Answer: **Ordinary Least Squares (OLS)**

Recall: the OLS estimate is the combination of $\hat{\beta}$s that minimizes the SSE



Example (continued...)

- So what's missing?...

$$GPA_i = \beta_0 + \beta_1 I_i + \beta_2 SAT_i + \beta_3 H_i + \beta_4 P_i + \varepsilon_i$$

Example (continued…)

- So what's missing?...

$$GPA_i = \beta_0 + \beta_1 I_i + \beta_2 SAT_i + \beta_3 H_i + \beta_4 P_i + \varepsilon_i$$

- We've buried a lot of important stuff in the error term (ε_i):
  - What is student's major?
  - How is the student's home studying environment?
  - Does the student regularly attend office hours?

- While we know that OLS is the **b**est **l**inear **u**nbiased **e**stimator, unbiasedness only holds under certain conditions.

Example (continued…)

- For an OLS estimate to be unbiased and give, on average, the causal effect of some *x* on some *y*, we need *x* to be uncorrelated with the error term.

$$E[x\varepsilon] = 0 \qquad \leftrightarrow \qquad corr(x, \varepsilon) = 0$$

- In words: the error term contains all variables that determine *y*, but we omitted from the regression model.
- If *x* is correlated with any omitted variables that end up in the error term, then our estimate suffers from **omitted variable bias.**

Example (continued…)

- Say we estimate the **naïve** regression of the following form:

$$GPA_i = \beta_0 + \beta_1 P_i + \varepsilon_i$$

- What would happen if we took a sample of real world data and used OLS to estimate $\widehat{\beta_1}$?

Example (continued…)

- Say we estimate the **naïve** regression of the following form:

$$GPA_i = \beta_0 + \beta_1 P_i + \varepsilon_i$$

- What would happen if we took a sample of real world data and used OLS to estimate $\widehat{\beta_1}$?

- Who is more likely to be exposed to lead pollution?
  - Answer: children in poorer households. (Likely to live in older housing; less ability to remediation paint/pipes; less ability to move away)

- Is the correlation between lead exposure and household income a problem for our estimate of $\widehat{\beta_1}$?

Example (continued...)

- Is the correlation between lead exposure and household income a problem for our estimate of $\widehat{\beta_1}$? YES!!!

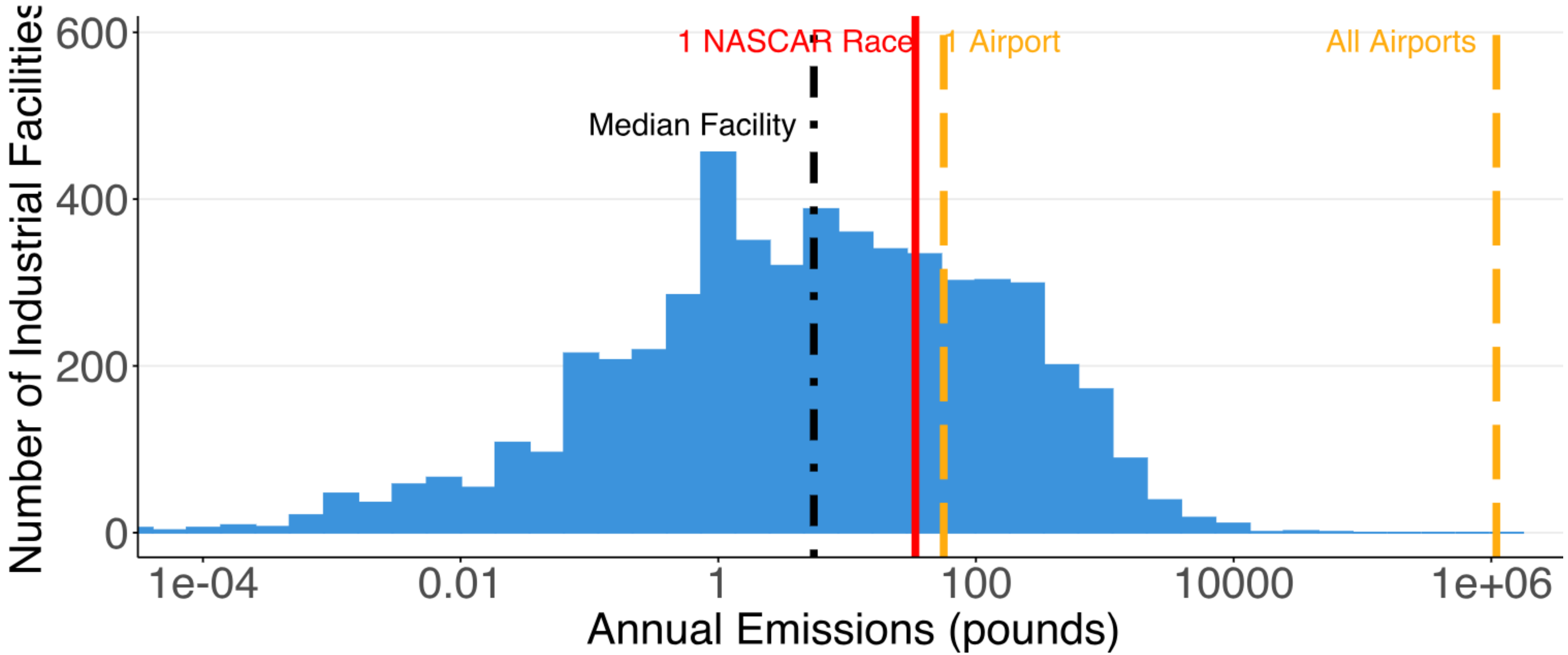
- Household income also matters for GPA, and in the naïve regression, is included in ε_i. Thus $corr(x, \varepsilon) \neq 0$.

- Why might household income be correlated with GPA?
  - Pick your story: after-school tutoring, better high schools, parental pressure, etc., etc.

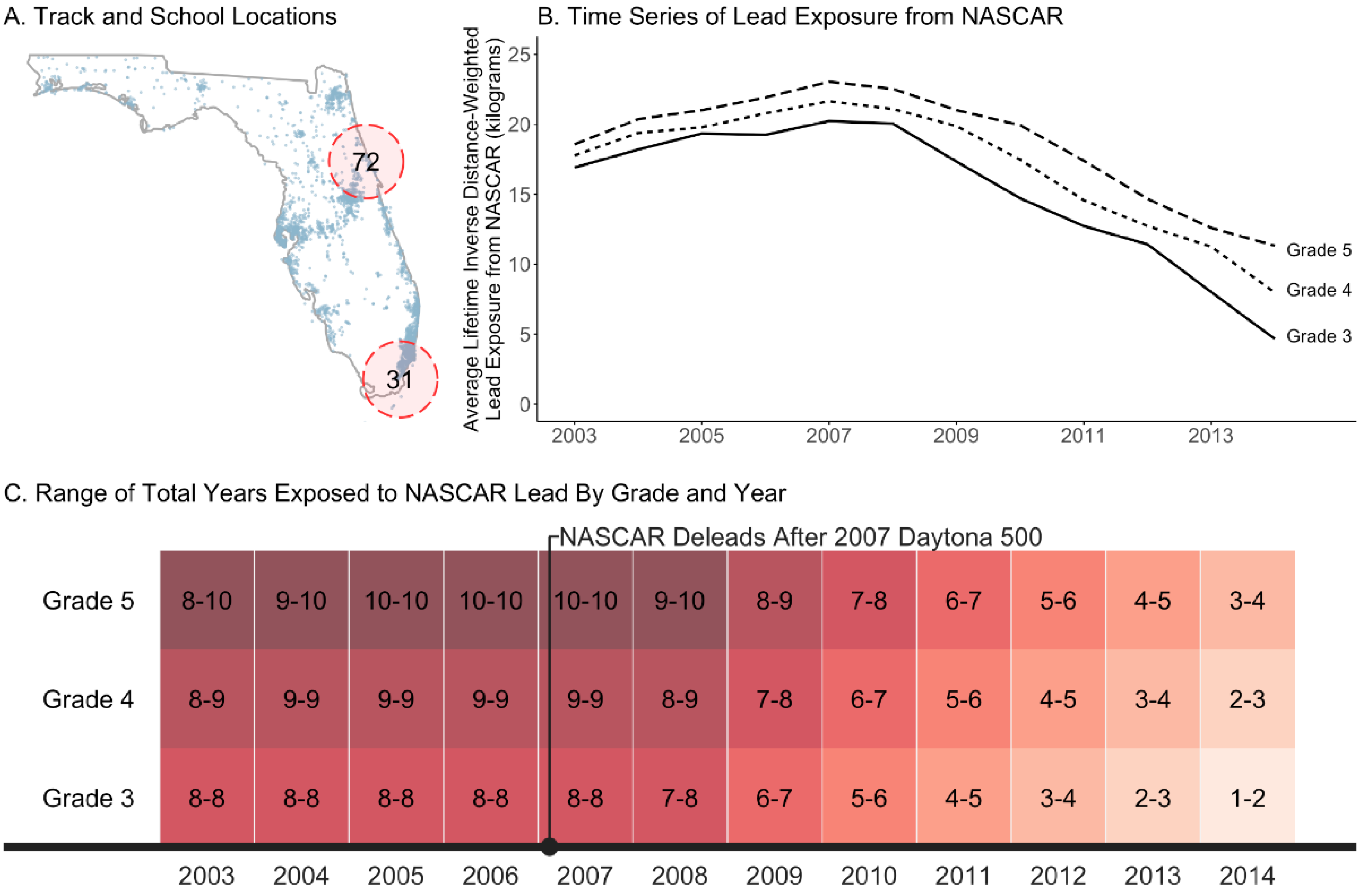
- The econometric result is a biased estimate of $\widehat{\beta_1}$.
  - It confounds the effect of income with the effect of lead exposure.
  - Because students with higher lifetime exposure are from poorer households, the naïve $\widehat{\beta_1}$ overstates the negative effects of lead

A real world example: lead and education

Prior to 2007, one NASCAR race would result in the ambient emissions of more lead than many industrial facilities’ annual output:



Florida has publicly-available, school-level data



Data: observations are school-by-year-by-grade

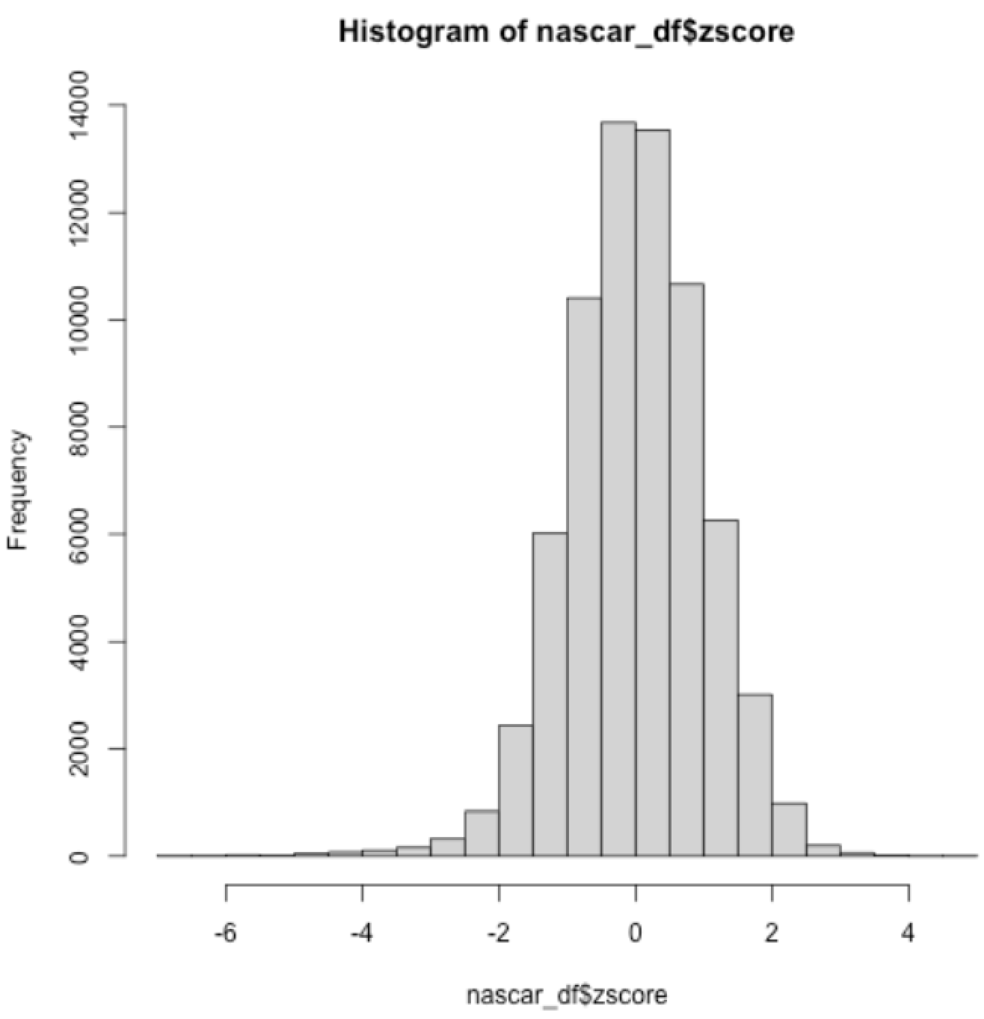
```
nascar_df
```

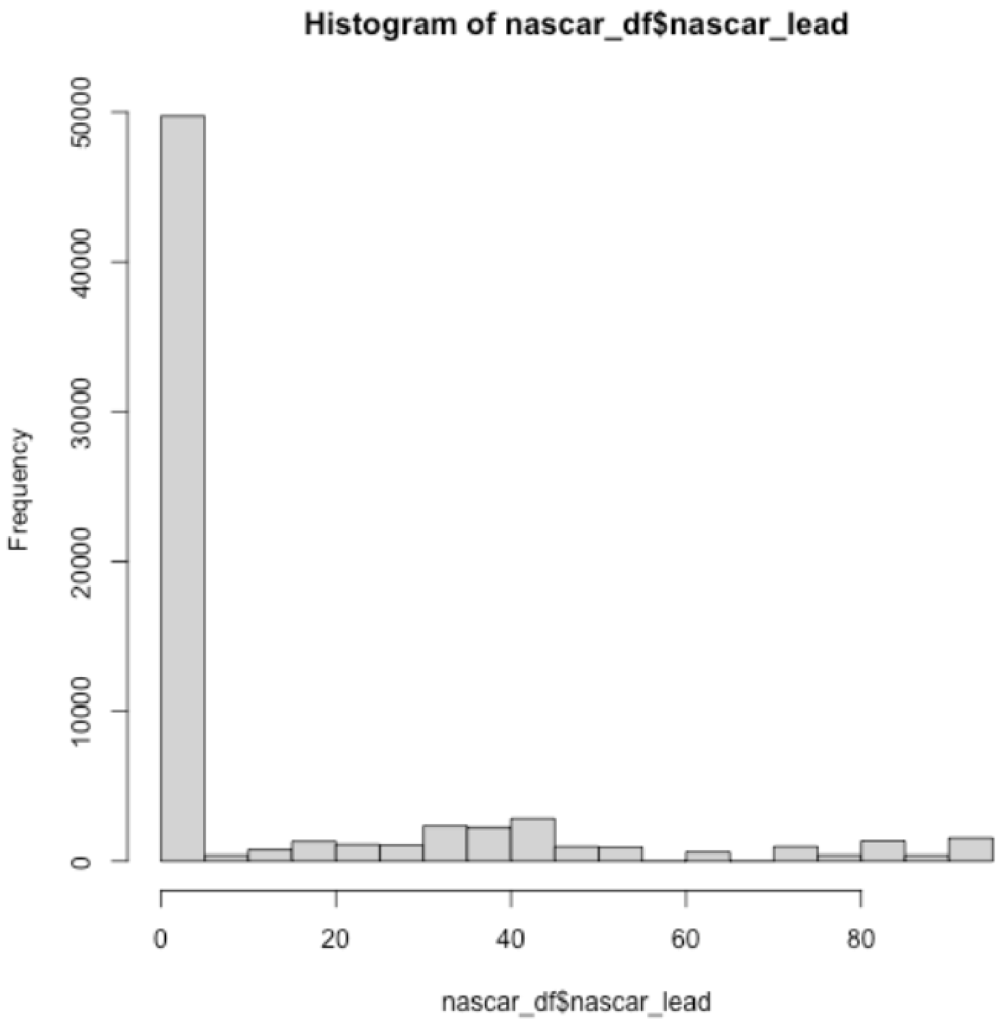
```
## # A tibble: 68,858 × 12
##    school_id school_name   grade  year zscore nascar_lead nascar_lead_weighted years_leaded indust…¹ media…² unemp…³ num_s…⁴
##        <dbl> <chr>         <dbl> <dbl>  <dbl>       <dbl>                <dbl>        <dbl>    <dbl>   <dbl>   <dbl>   <dbl>
##  1        56 HAMILTON ELEM     3  2003 -0.186        72.2                 2.53            8  822328.   49267    0.05     112
##  2        56 HAMILTON ELEM     4  2003  0.101        80.4                 2.81            8  822639.   49267    0.05     117
##  3        56 HAMILTON ELEM     5  2003 -0.206        88.0                 3.08            8  822909.   49267    0.05     120
##  4        56 HAMILTON ELEM     3  2004 -0.686        74.0                 2.59            8  967077.   50842    0.04     131
##  5        56 HAMILTON ELEM     4  2004 -0.633        82.4                 2.88            8  967352.   50842    0.04     105
##  6        56 HAMILTON ELEM     5  2004  0.352        90.5                 3.17            8  967663.   50842    0.04     109
##  7        56 HAMILTON ELEM     3  2005 -1.14         77.0                 2.69            8 1061570.   52390    0.03     110
##  8        56 HAMILTON ELEM     4  2005 -0.649        84.7                 2.97            8 1062071.   52390    0.03     137
##  9        56 HAMILTON ELEM     5  2005 -0.336        92.0                 3.26            8 1062346.   52390    0.03      97
## 10        56 HAMILTON ELEM     3  2006 -0.333        79.9                 2.80            8 1164072.   56655    0.02     133
## # … with 68,848 more rows, and abbreviated variable names ¹industrial_lead, ²median_income, ³unemp_rate, ⁴num_students
```

Variables definitions

- **zscore**: the school's score for the average student in terms of standard deviations above or below the state-wide average
- **nascar lead**: lifetime exposure to lead emissions from NASCAR tracks within 50 miles
- **industrial lead**: lead emissions from industrial sources (e.g. factories) within 50 miles
- **median income**: the school district's median incoe
- **num students**: the number of students at the school
- **school id, school name, grade, and year**: self-explanatory

Summarizing the data variation





Let's run the naïve regression…

$$zscore_{sgy} = \beta_0 + \beta_1 nascar\_lead\_weighted_{sgy} + \varepsilon_{sgy}$$

- *sgy:* (school, grade, year)

Let's run the naïve regression…

$$zscore_{sgy} = \beta_0 + \beta_1 nascar\_lead\_weighted_{sgy} + \varepsilon_{sgy}$$

- *sgy:* (school, grade, year)

```
## Estimation Results
##   parameter                  estimate
## 1 beta_0 (Intercept)            0.002
## 2 beta_1 nascar_lead_weighted  -0.004
```

- In words: "10 additional kg of lead exposure from nearby NASCAR races is associated with a school having an average test score 0.004 SDs lower."

We know we can do better…

- Schools near raceways are probably different than schools further away

- We need to control for school attributes that are correlated with both test scores and being spatially close to a NASCAR race

- Two variables in our data meet that standard:
  - Lead emissions from non-race sources (*other_lead)*
  - Median income in the school district (*income)*

Let's run the naïve regression…

$$zscore_{sgy} = \beta_0 + \beta_1 nascar\_lead\_weighted_{sgy} + \beta_2 other\_lead_{sgy} + \beta_3 income_{sgy} + \varepsilon_{sgy}$$

```
## Estimation results
##   parameter                    estimate
## 1 beta_0 (Intercept)           -0.846
## 2 beta_1 nascar_lead_weighted  -0.0008  (versus -0.004 above)
## 3 beta_2 other_lead            -0.00000006 (other lead = bad!)
## 4 beta_3 income                0.00002 (rich family = good!)
```

Omitted variable bias: new estimate of $\widehat{\beta_1}$ is only <u>**20%**</u> of the naïve estimate's magnitude

Do you find this estimate credible?

Bringing this back to the realm of quantifying monetary benefits…

Modelling required to quantify key benefit channel of reducing lead exposure:

Less lead -> Higher IQ -> Better schooling outcomes -> College attendance -> Higher lifetime wages

Bringing this back to the realm of quantifying monetary benefits…

Modelling required to quantify key benefit channel of reducing lead exposure:

Less lead -> Higher IQ -> Better schooling outcomes -> College attendance -> Higher lifetime wages

Economic Analysis of the Proposed Reconsideration of the Dust-Lead Hazard Standards and Post-Abatement Clearance Levels

June 2023

Table A-1: Summary of IQ Point Dollar Values (2021$)		
Estimate Parameter	**Discount Rate**	
	3%	**7%**
IQ Value	$24,688	$5,726
Additional Education Costs and Lost Earnings	$1,588	$728
IQ Value without Additional Education Costs and Lost Earnings	**$23,100**	**$4,997**

Environmental Econometrics

- Econometrics used within the field of environmental economics is not inherently different than approaches used in other fields.

Environmental Econometrics

- Econometrics used within the field of environmental economics is not inherently different than approaches used in other fields.

- However, in many cases the research questions are different and, therefore, require different methods.

Environmental Econometrics

- Econometrics used within the field of environmental economics is not inherently different than approaches used in other fields.
- However, in many cases the research questions are different and, therefore, require different methods.
- Because the research questions are different, environmental economists have advanced specific areas of econometrics to address hurdles that other fields may not face. In particular: econometric tools to estimate the value of **nonmarket goods and services**, specifically nonuse or nonlocal goods.

What is a nonmarket good?

"The problem of *non-market valuation* is to compute a monetary value for environmental resources or services that, in the general set up of market economies, are not subject to exchange, and therefore do not have an observable market price.

For instance, we might be interested in knowing the monetary value of preserving a wooded area in an urban landscape, relative to area's value in a residential use."

- Phaneuf & Requate (2017)

Revealed preference approaches

"Revealed preference methods are based on the premise that, though there are no direct markets for some environmental goods, people interact with the environment in a variety of observable ways.

Through their participation in private good markets that interact with the environment, people often leave behind behavioral footprints that reveal information about their environmental preferences."

Revealed preference methods are based on actual behavior reflecting utility maximization subject to constraints… we can infer monetary value "from observations of people acting in real-world settings where they must live with the consequences of their choices."

Travel Cost models

- Often focuses on nonmarket value of land/water for recreational uses
- **Variation in travel distance (the “cost”) reveals the economic value of “a trip”.**
- A “trip” contains multiple “amenities” and is a “bundled” good.
  - Activities, hiking, biking, fishing
  - Trail miles, park benches, access
  - Quality and quantity of these amenities

$$U_{ij} = -\lambda price_{ij} + X\beta + \phi + \varepsilon_{ij}$$

$$price_{ij} = \delta distance_{ij} + \gamma Z + \epsilon$$

Hedonic Valuation

- Instead of using variation in travel costs to reveal value of nonmarket environmental attributes, can we use variation in price?
- When you purchase or rent a place of residence, you are buying a bundle of characteristics
  - Square footage, building age, # of parks nearby, distance to Trader Joe's, etc.
- We can use econometric tools to carefully **measure the price variation across residential locations that is attributable to gradients of environmental quality** (say: the marginal value of distance to public green space)
- Added value compared to travel cost models: can measure use values of environmental attributes from a market in which (almost) everyone participates: housing

$$price_i = X\beta + \phi + \varepsilon_i$$

Averting Behavior / Defensive Expenditures

- People are exposed to various levels of pollution. This exposure presents health risks, perhaps in the form of lost productivity, lost workdays, lost leisure time, etc.
- To help protect themselves from this potential risks to their health, people can **change their behavior or make defensive expenditures that will reduce the severity or length of the illness.**
  - E.g. skip exercising outside during bad air pollution days, drink bottled water to avoid tap water, use air conditioning to protect against high temperatures
- These expenses or actions tell us something about individuals' willingness to pay to avoid pollution exposure!

$$Costs_{ij} = X\beta + \phi + \varepsilon_{ij}$$

Stated preference approaches

- Because markets for nonmarket goods don’t exist, sometimes it is useful to create them!
- Hypothetical markets can address:
  - Changes in nonmarket goods outside of their historically observed levels
  - Creation of entirely new goods and services (new park, etc.)
  - **Nonuse values**. If people don’t travel to them, live near by them, or even use them, but just care that they exist.
- On the other hand, **hypothetical markets --> hypothetical valuation**

$$U_{ij} = -\lambda price_{ij} + X\beta + \phi + \varepsilon_{ij}$$

In practice: econometrics in the environmental realm

- Linking research questions to econometric approaches requires careful thought about what the question is, what you believe the answer might look like, and the available data.

- To best address a research question, start by thinking about what the "ideal" data might look like. Learning how "real-world" data compares to the "ideal" data often guides the approach that should be taken

- Lastly, think about unique events that happen in our everyday lives that act as natural experiments. Did something occur suddenly and without expectation? Do states have very different policies while the communities on either side of the border remain the same?

Reminders

- Reflection Post #1 due Sunday @ 11:59pm

- **Tomorrow: Travel Cost model.** Be ready to discuss Hanauer & Reid ("Valuing urban open space using the travel-cost method and the implications of measurement error")

- *Optional reading*: Gellman, Walls, & Wibbenmeyer ("Welfare Losses from Wildfire Smoke: Evidence from Daily Outdoor Recreation Data")

On reading academic econ articles…

- As we discuss nonmarket valuation (and going forward, env. policy), will rely on research as illustrative case studies
  - Each class we will be reviewing articles, their methods, and their conclusions

- For those with less previous exposure: academic economics articles may require a bit of a learning curve

- But there is a lot of great advice out there on how to best get started.
  - How to Read Journal Articles like a Professor