

EC970: The Evolution of Segregation and Education

Class Introduction

1/23/2024

Plan for today

- Introductions
- Syllabus
- Reading and writing economics
- Upcoming due dates

(if there's time) metrics!



The basics

- We meet 9-10:15 am in Sever 304
- Office hours are in Clover
 - 10:30-11:30 Tuesday
 - 9-11 Friday
 - I prefer if you make an appointment, but technically walk-in
- No laptops allowed



Resources you should bookmark

- Course repo: <https://github.com/daniellecgw/segregation-ed/>
 - Contains assignment descriptions, handouts, and syllabus
- Syllabus: https://github.com/daniellecgw/segregation-ed/blob/main/syllabus_segregation-ed.pdf
- Canvas page: <https://canvas.harvard.edu/courses/126957>
 - Turn in assignments here
- Stata / R page: <https://canvas.harvard.edu/courses/19323>
- Schedule office hours:
<https://daniellecgw.github.io/teaching/segregation-ed>
- Assorted other resources listed on syllabus

Course structure

- Discussion and writing based class
 - Class participation **15% of your grade!**
 - Come to class prepared
- Most classes: 30-45 min discussing papers
 - Remaining time on metrics or other review
 - Exception: this week (metrics time)

What are we going to do?

- Learn about school segregation and integration over time
- Learn how to use common applied microeconomics tools
- Learn how to read economic papers critically
- Learn how to write research papers

What are we going to do?

- Learn about school segregation and integration over time
- Learn how to use common applied microeconomics tools
- Learn how to read economic papers critically
- Learn how to write research papers

How are we going to do it?

- Read a lot of papers
- Talk about a lot of papers
- Write a lot of papers
- Do Stata / R PSETS

What are we going to do?

- Learn about school segregation and integration over time
- Learn how to use common applied microeconomics tools
- Learn how to read economic papers critically
- Learn how to write research papers

How are we going to do it?

- Read a lot of papers
- Talk about a lot of papers
- Write a lot of papers
- Do Stata / R PSETS
- This class is a lot of work.
Plan ahead

What I expect from you

- Come to class on time and be present
- Do the readings
- Turn in your work on time
- Don't cheat
- Treat each other with respect



Attendance and Extensions: Attendance is mandatory. Extensions are discouraged, but allowed under exceptional circumstances.

- **You get one freebie absence with no questions asked.**
- If you are sick, please don't come to class. If you must come to class sick (even if it's just the sniffles!), I ask that you wear a mask. Please e-mail me to let me know if you are sick. If you don't e-mail me **before class**, I will assume you are using your one freebie absence.
- If absent for any reason (with the exception of your one freebie absence), you will be expected to write an additional response paper on the day's reading and schedule office hours with me to make up the class.
- Any other absences and **any** extensions must be cleared in advance with me.
 - **Acceptable reasons to ask for an absence / extension include**, but are not limited to, medical emergencies, scheduled family gatherings (such as a wedding), or the death of a loved one.
 - I ask that, except in the case of emergencies, you discuss planned absences / extensions with me at least **two weeks in advance** by coming up to me after class.
 - There will be a **maximum of one extension not exceeding three days** granted for each assignment.
 - **No extensions will be granted for the final paper** – this deadline is set by the department. Plan ahead accordingly!

Late Work Policy: Each student has a “late bank” of 48 hours they can use throughout the semester. If you turn in a response paper 2 hours late, you have 46 hours in your late bank remaining. If you then turn in a short paper 20 hours late, you now have 26 hours remaining. **It is up to you to keep track of how many hours you have in your “late bank”** throughout the semester. You may ask me at any point to clarify how many hours you have remaining.

- You may not use your late bank on the final paper or the Stata / R PSETs. These deadlines are set by the department.
- You may not use your late bank to turn in a response paper after 9am on the day that the paper is discussed. Response papers will not be accepted after 9 am on the day that the paper is discussed.

Note that using your late bank on certain assignments may prevent me from giving you timely feedback.

If your late bank is empty, I will grade your assignment as follows:

- Turned in .5-24 hours after deadline: A- max grade
- 24-48 hours: B max grade
- 48-72: C max grade
- After 72 hours: assignment not accepted

Plagiarism: "Students who, for whatever reason, submit work either not their own or without clear attribution to its sources will be subject to disciplinary action, up to and including requirement to withdraw from the College." – *Harvard College Handbook for Students*. If you cheat, you will fail this class.

Collaboration: Discussion of papers and problem sets with others is permitted and encouraged. Copying others is not. Every student is expected to turn in their own unique work.

Generative AI: ChatGPT and other generative AI are tools you should learn to use **responsibly**. Examples of acceptable use include:

- “Polishing” a draft
- Asking for simple explanations of concepts (e.g. “can you explain fixed effects for me?” or “how do I interpret β in OLS?”)

Examples of unacceptable use include:

- Writing code for Stata / R problem sets: as researchers, we must be responsible stewards of data. Writing code manually will help you understand the nuts and bolts of the assumptions you make when you manipulate and analyze data, which you must acknowledge when conducting research.
- Writing first drafts: ChatGPT will often make up citations or describe identification strategies you didn’t do. Your research is the result of *you* analyzing a data set and running (likely) a regression-based analysis. ChatGPT cannot and should not do this for you.
- Writing response papers: We will cover this in more depth in class.

A note of caution: citation fabrication, or citation faking, happens when AI generates fictional citations. Accurate citations are vital for trust. Responsible AI use requires verification by researchers and writers to ensure legitimacy.

What you can expect from me

Emails: You can expect me to answer questions within 24 hours, with the exceptions of weekends: questions sent on Friday, Saturday or Sunday can expect a response by the end of day Monday.

- In return, I expect you to budget your time appropriately: a good rule of thumb is to ask any urgent questions at least **36 hours in advance** of a deadline so that I have time to respond and you have time to digest.
- **Please include EC970 in the subject line** to help me find your e-mail faster.

Come to Office Hours!

Research is hard. Please do not hesitate to come to office hours— I want to make conducting economics research feel approachable and accessible for all of you. In the second week of class, everyone is required to schedule an OH slot with me. [Schedule OH here.](#)

Accessibility: Please contact me if you need special arrangements to use a laptop or any other accommodation.

Assignments

- Class participation – 15%
- Stata / R PSETS – 10%
- 3 short papers – 15%
- Response papers – 8%
- Leading class discussions – 4%
- Final paper prep assignments – 10%
- Final presentation – 5%
- Final paper – 33%



Leading Class Discussions and Response Papers 12%

The purpose of these exercises is to get you practicing two important skills in research: thinking critically about papers and generating new ideas. Don't put too much pressure on yourself – I do not expect all of these ideas to be good!

Response papers: Throughout the semester, you will write **four** 2 page (size 12, Times New Roman, 1 inch margins, double-spaced) response papers. These are due **5 pm before the day we cover the paper in class.** Most weeks, you will summarize the paper, then discuss pros, cons, and directions for future research. Some weeks, there will be specific prompts (for example, defining key terms). If there is a specific prompt for the next weeks' response paper, it will be posted by 5 pm on Friday.

You pick which papers you respond to, subject to the following conditions:

- Respond to a starred ★ (not *) reading
- Respond to a maximum of one paper per week
- Do not turn in a response paper the week of January 25 (first week of class) or April 23 (last week of class)
- Do not turn in a response paper the weeks you are leading a class discussion

Leading Class Discussions: Each student will present twice for 5 minutes on a paper from the class syllabus. The student will then lead (with some help) the class discussion (this will last around 30-40 additional minutes). Students will submit their preferences at the beginning of the term. Slides are due **5 pm before the day we cover the paper in class.**

Stata / R Minicourse Problem Sets (4) 10%

As part of EC970, all students participate in the Stata / R minicourse. You will complete four problem sets. This minicourse is an excellent opportunity to learn how to code and handle data *responsibly*. The [canvas webpage](#) for the course is an excellent resource.

Coding Resources

- The [canvas webpage](#) for the Stata / R mini course has excellent resources for coding.
- There are dedicated TFs for the Stata / R minicourse. Make use of them! Go to their office hours. They will give you better coding help (especially in Stata) than I.
- For R, I recommend [R for Data Science](#), a free, online, well-written and frequently updated textbook.
- [Econometrics with R](#) is another free, online, frequently updated textbook that focuses specifically on econometrics.

You have the option of writing your problem sets in either R or Stata. I use R. I rarely use Stata. I like R because it is free, you can use it to read in Stata files, and you can make beautiful graphs with ggplot. Downsides of R are that some econometrics package run smoother in Stata. However, at the end of the day, the most important thing is that you pick a language and learn it well. Most graduate student's language of choice is entirely dependent on the language that their first boss had them code in. My first job was in data science, and Stata is almost never used in non academic settings, so now I use R.

Short Papers 15%

Why Study School Integration? 5%

This assignment encourages students to think critically about what outcomes to prioritize when studying school integration. Students will write a short paper (3-4 pages) responding to two works that summarize the consequences of school integration.

Measuring Segregation 5%

This assignment introduces students to measures of segregation. Students will write a short paper (4-6) that discusses the pros and cons of different measures.

Explain a Model 5%

Students will write a short paper (4-6 pages) that outlines a framework (“model”) of behavior based on the abstract of a paper from the reading list. The goal of this assignment is to give students practice in conceptualizing mechanisms that drive empirical results.

Final Research Paper 43%

Throughout this course, you will build a final research paper. The final product should be **15-18 pages** (size 12, Times New Roman, 1 inch margins, double-spaced). This paper will be on a topic of your choosing that falls in the realm (broadly defined) of this class.

- ***Three Ideas*** 2%
- ***Data Description and Summary Statistics*** 3%
- ***Intro, Empirical Strategy and Rough Draft*** 2%
- ***Rough Draft and Peer Review*** 3%
- ***Final Paper*** 33%

Final Presentation 5%

Each student will present on their final research paper for 15 minutes total. This time includes questions from the audience – when you practice your talk (and you should absolutely practice your talk), it should take around 10-12 minutes. Both asking and responding to questions in research seminars are important skills, and all students are required to ask at least one question (and more questions are encouraged!).

Writing (and Reading) Economics

Adapted from Michael Holcomb

Writing (and Reading) Economics

- The best papers are clearly organized, with each piece contributing to a unified whole
- Introduction – What is your research question? Why should we care?
 - Answering the second question takes experience; sometimes motivated by previous literature
- Literature review – What have people previously said about this topic? Why is this incomplete, and where does your paper fit in?
- Setting – What is the relevant background about the time, place, institutions, etc. that you're studying?
 - For example, court cases in the 1960s and 1970s
 - Usually narrative and some descriptive statistics

Adapted from Michael Holcomb

Writing (and Reading) Economics, con't

- Data – What are the sources of data you're using? What are their strengths/limitations?
 - Is the data only one year of data from 1979? How many observations?
- Methods (identification strategy) – What is the selection/OVB problem? What will you do to get around it? What are the strengths/weaknesses of your approach?
- Results – What did you find? How do we interpret your figures and tables?
- Conclusion – Bring it all together and remind us what we learned
- Discussion/policy implications – What can we generalize from your paper? How might this inform policy?

Adapted from Michael Holcomb

Writing (and Reading) Economics, con't

- Note that these sections can often be combined – for example, intro + lit review; setting + data; conclusion + discussion/policy implications
- Not every paper has policy implications, but it's generally nice to mention outside extensions
- Lately, economics papers have gotten very long, with long appendices

Adapted from Michael Holcomb

Reading checklist

- Key question
- What's new (compared to previous research)
- Key finding(s)
- Setting/Details
- Conceptual framework/model
- Data
- Methods
- Tables/coefficients of interest
- Problems?
- Thoughts? Relevance? Implications?

Adapted from Michael Holcomb

Reading questions

1. What is the main question or hypothesis?
2. How does the paper fit into the themes of the course? What did it teach us about school integration or segregation?
3. What is the main identification problem in the paper? Did the author(s) convincingly overcome it, and if so, how?
4. If the paper uses data, what are the strengths and the limitations of the data source(s)?
5. How would you revise/re-do/extend this paper? Was anything confusing or unclear?

Adapted from Michael Holcomb

Designing Research

The Effect, Chapter 1-2

What is empirical economic research?

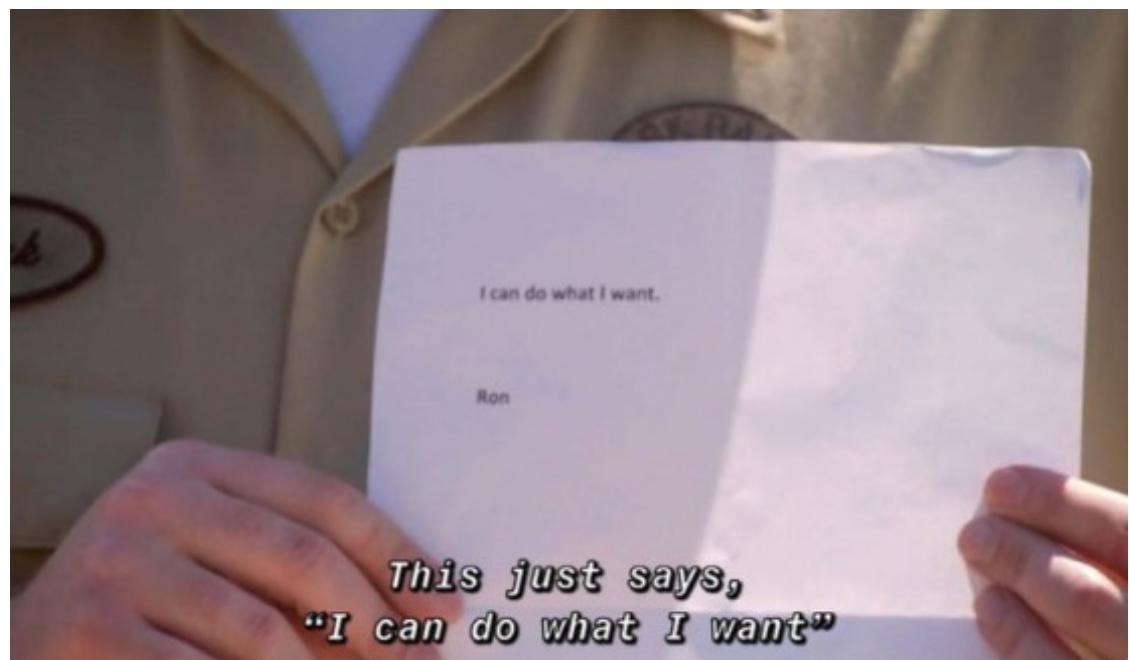
- Empirical = backed by data
 - “Uses structured observations to attempt to answer questions”
- Economic = ??
 - Outcome definition: research question relates to economic outcomes
 - E.g. labor market (job) (you can relate most things to jobs if you think hard enough)
 - Methods definition: uses math that focuses on causal inference*
 - Causal inference = you can infer that what you’re studying caused the other thing you’re studying (and not some mystery third thing)
 - * We will also read some descriptive papers
- You’ll end this class by writing a research paper
 - That paper will focus on an answerable research question

Ok, how do I do research?

Easy, three step process:

1. Craft research project
2. Figure out data you need
3. Figure what calculations you need to perform on your data

So easy! So simple!



Beauty of empirical micro economic research is that empirical micro tools can be applied to most topics

Economics is a cool way of framing the world around us!

But you can't do **anything** you want...

Good vs. bad research questions

Good:

1. Can be answered
2. Improves understanding of how world works

How to do it:

- Come up with theory of how world works (“data generating process”)
- Come up with testable hypothesis
 - “If this is my explanation of how the world works, then what should I observe in the world? Do I observe it?”

Test: if you find an unexpected result, does it change your understanding of the world?

- If yes and you still hold onto original theory, not good research question

Example: Graves 2024

- Bad: “Is my dog the cutest dog in the world”
 - Theory: My dog is the cutest dog in the world
 - Hypothesis: If I ask people if they think my dog is the cutest dog in the world, they will say yes
 - Data: People’s responses
 - Unexpected result: most people say no
 - This doesn’t change my understanding of the world
 - Those people obviously don’t get it...

Example: Reber 2011

"From separate and unequal to integrated and equal? School desegregation and school finance in Louisiana" (REStat 93:2 2011)

- Good: "Did court ordered desegregation improve the quality of schools Black children attended in Louisiana?" (Reber 2011)
 - Theory: "One motivation for desegregating public schools was a concern that—despite the legal requirement that black schools be “equal” to white schools—black schoolchildren would never receive adequate resources in their schools so long as they were separate from whites, and it was hoped that educational opportunities for minority children could be improved in part by tying their fate to that of the white students in their districts." pg. 1
 - Hypothesis: If this is true, we would expect to see an increase in per pupil funding and teacher-student ratios at schools Black students attended
 - Data: Funding, teacher-student ratios from state department of education reports
 - Result: funding, teacher-student ratios increase
 - Test: unexpected result: no change or decrease
 - This would change our understanding of the world!
- "Why are schools segregated?" is a **bad** research questions because it is too broad
 - Testing too many hypotheses!

Ok, how do I do research? (revised)

Easy, three step process:

1. Craft research project

1. Come up with research question

1. Think about theory of how world works (data generating process)
2. Come up with hypothesis to test such “that the result of the test tells us something about the theory”

OR

- Learn something unusual -> what research questions would this allow me to answer?

Check:

- Feasibility, scale, design
- Consider potential results

2. Figure out data you need

3. Figure what calculations you need to perform on your data

So easy! So simple!

Describing Variables and Relationships

The Effect, Chapter 3-4

Brief recap on basics

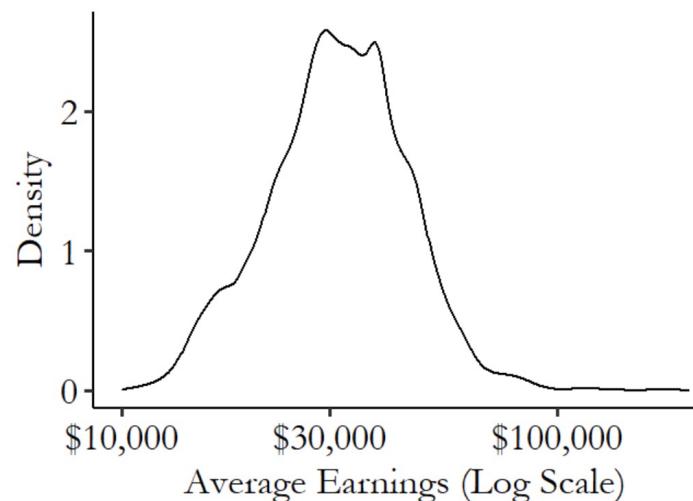
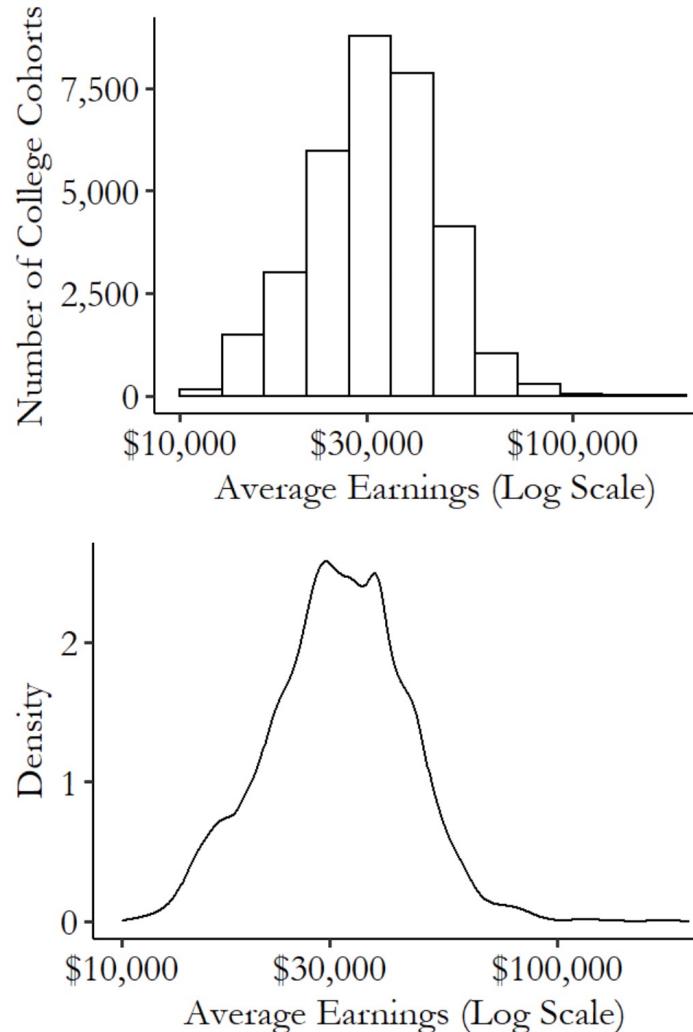
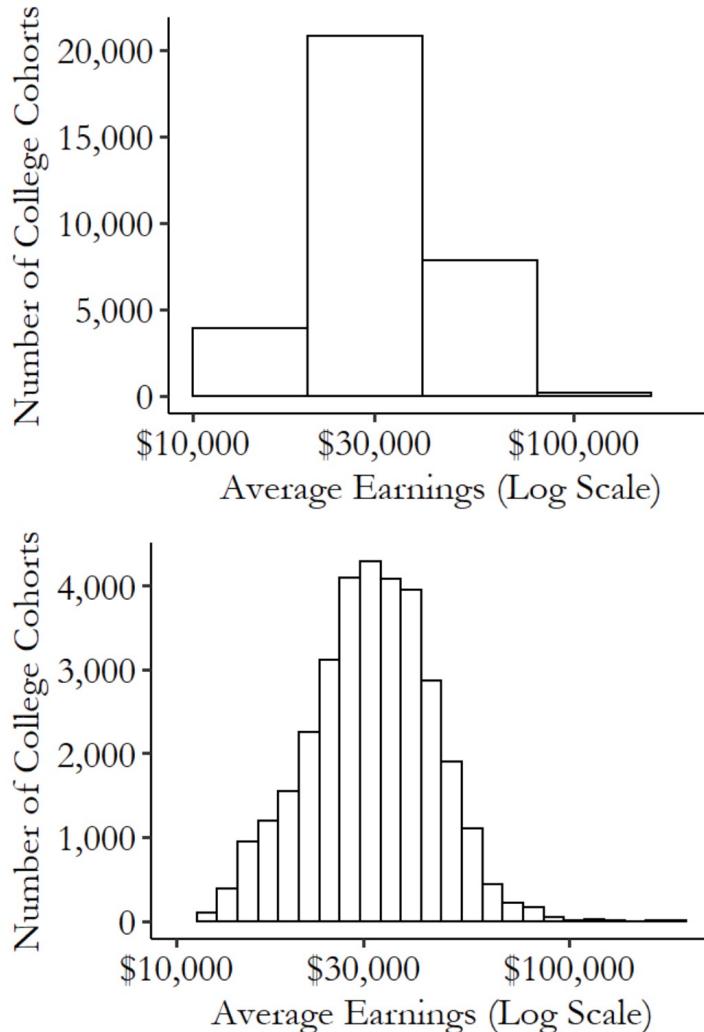
Type of variable affects type of analysis you do

1. Continuous
2. Count
3. Ordinal (first place, second place, third place)
4. Categorical (e.g. binary)

Ways to summarize data (“moments”)

- Median, mean, percentile, range, distribution

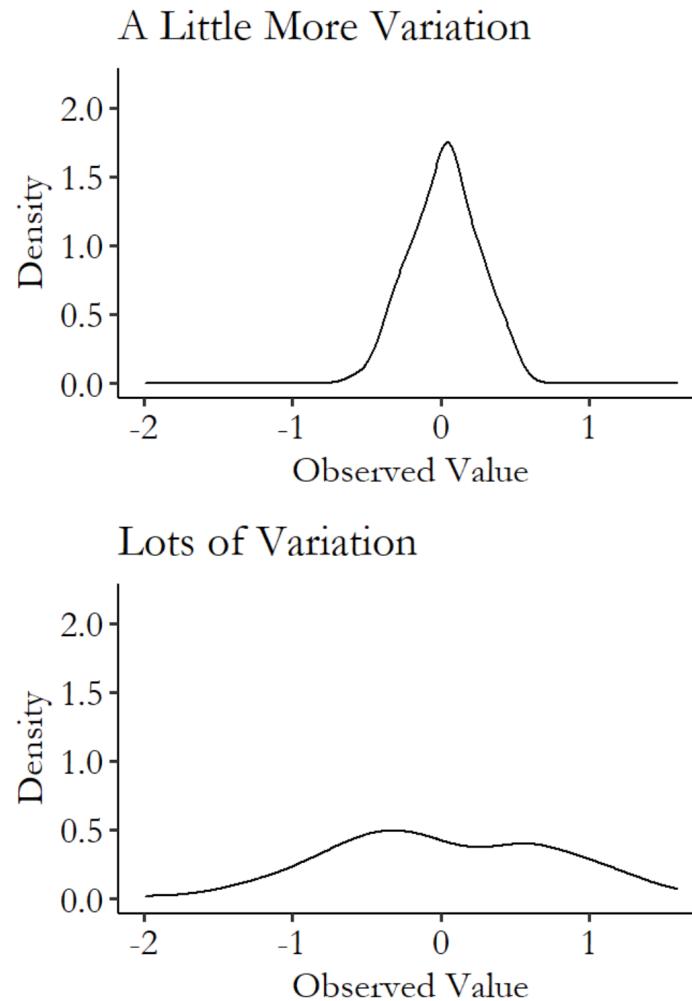
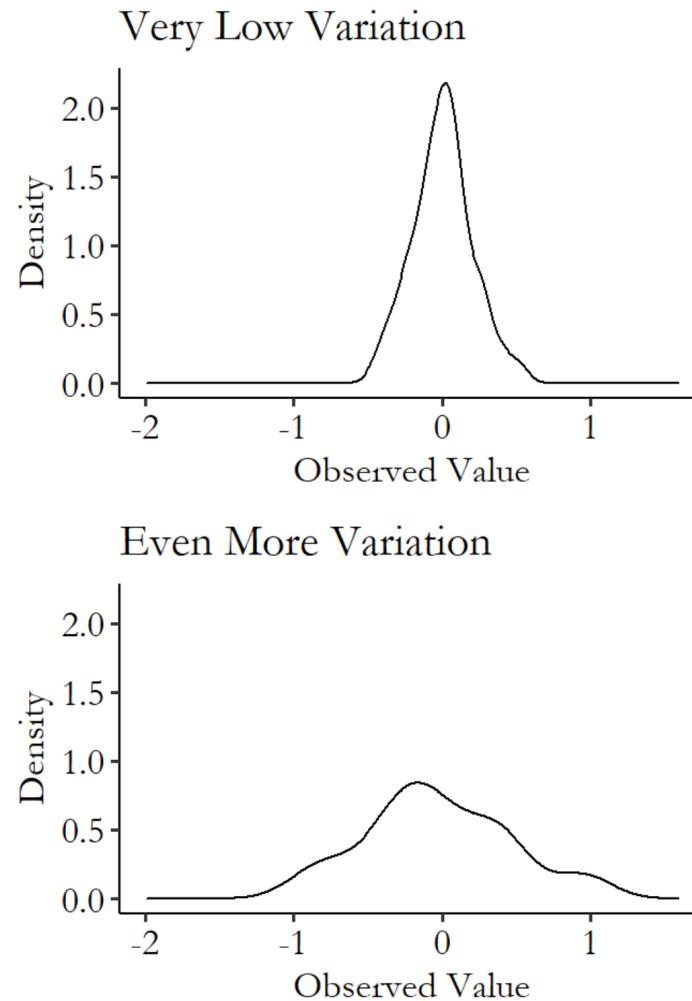
Distribution



- True distribution not known
- Observations give us an idea of what our variables' distributions are

Figure 3.3: Distribution of Average Earnings across US College Cohorts

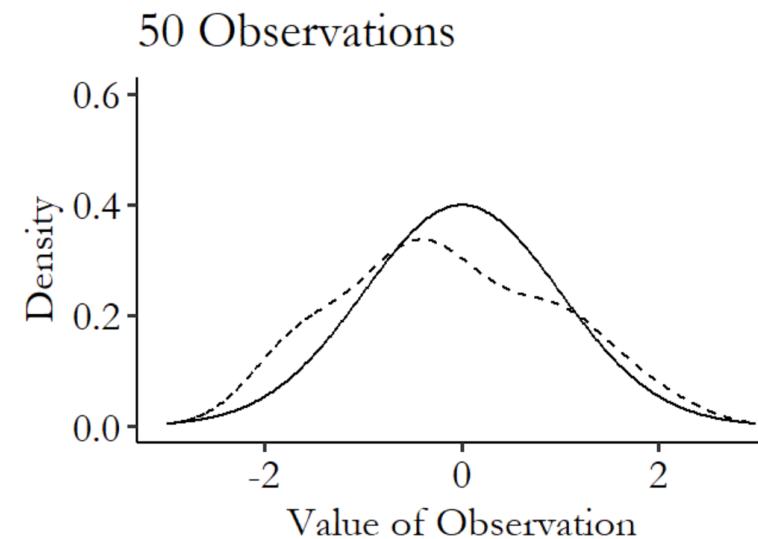
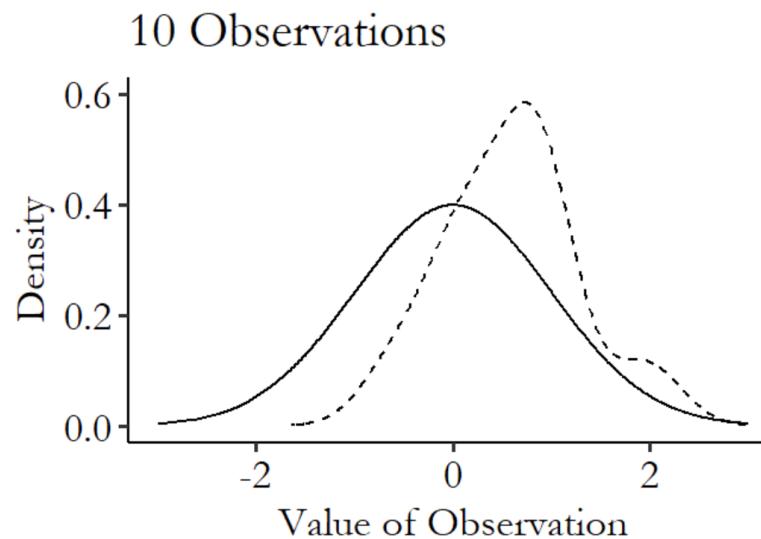
Variance



- How confident we are in observations vs. reality dependent on variance and number of observations

Figure 3.6: Four Variables with Different Levels of Variation

Theoretical Distribution



- Dotted: observed
- Line: actual

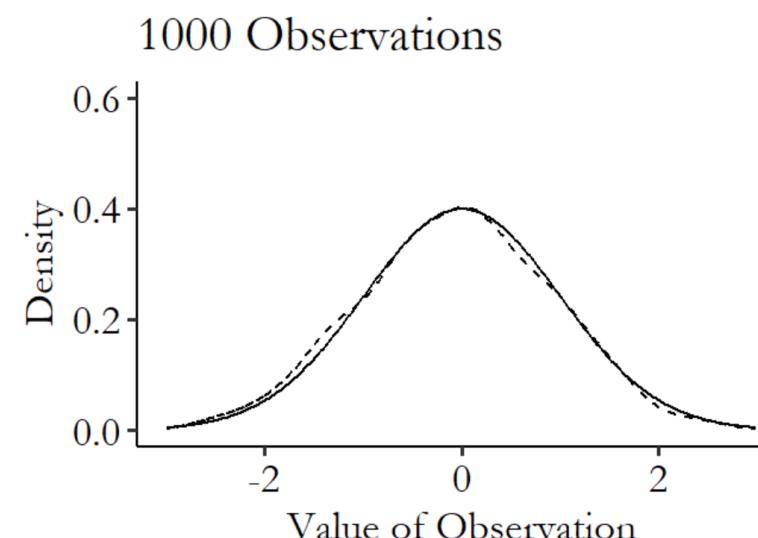
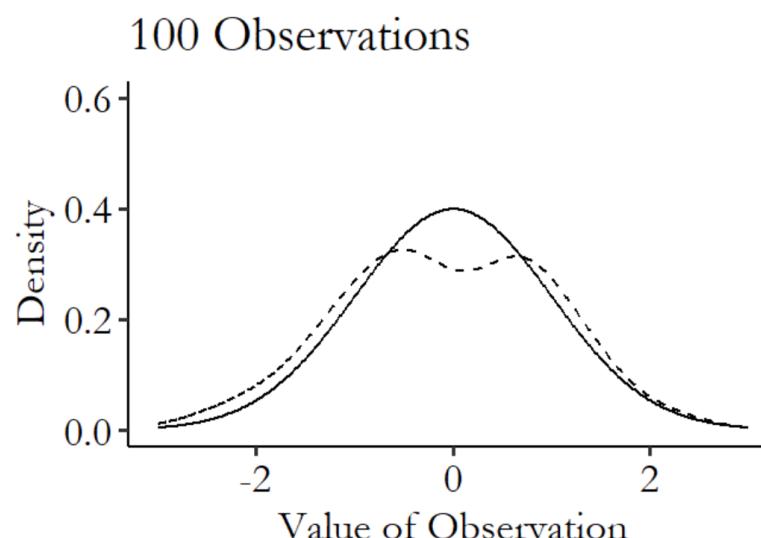
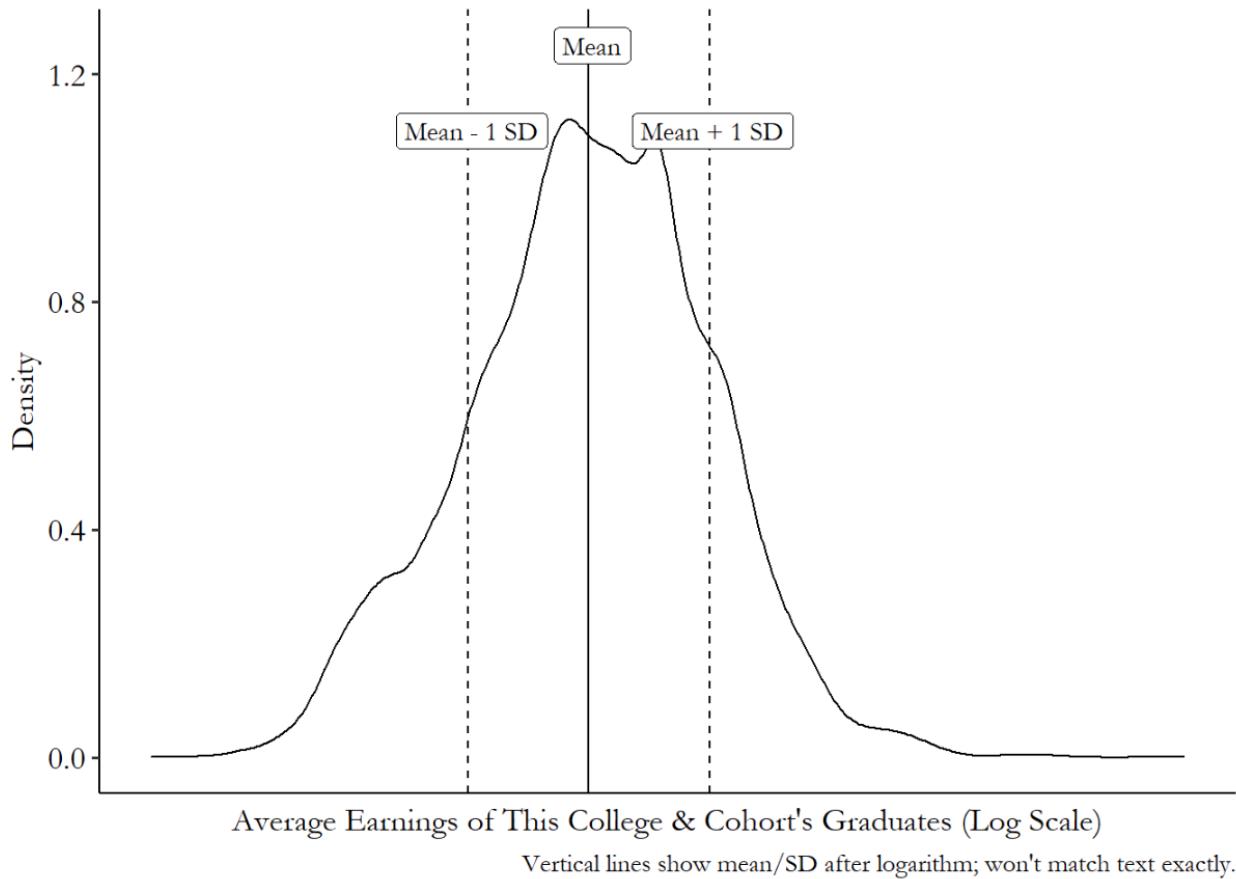


Figure 3.11: Trying to Match the Theoretical Distribution

Standard deviations

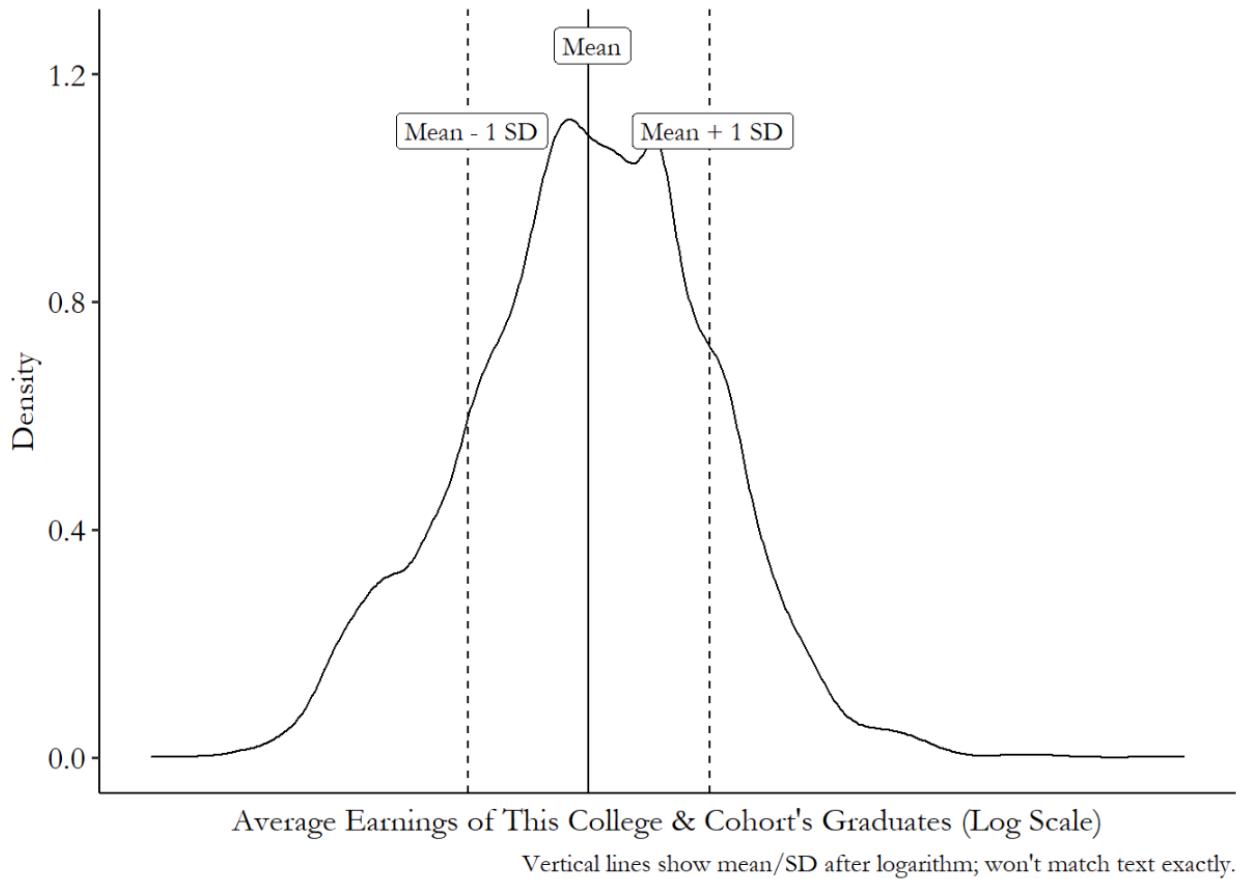
So how weird is being one standard deviation away from the mean? Well, roughly a third of people are between you and the average. Make of that what you will.



- Gives us an idea of variance of distribution without graph
- How your would interpretation of result change if there's a large standard deviation?

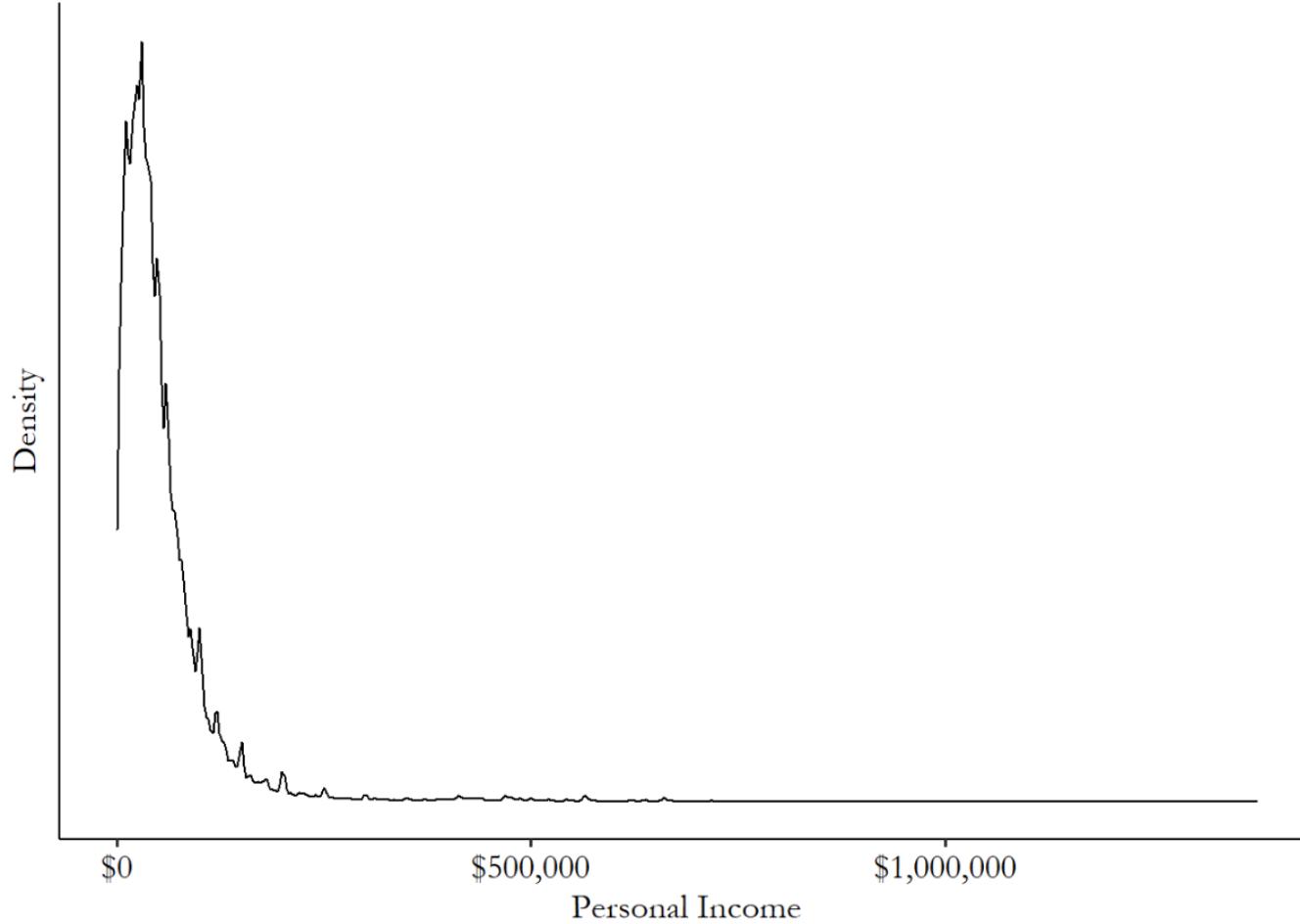
Standard deviations

So how weird is being one standard deviation away from the mean? Well, roughly a third of people are between you and the average. Make of that what you will.



- Gives us an idea of variance of distribution without graph
- How your would interpretation of result change if there's a large standard deviation?

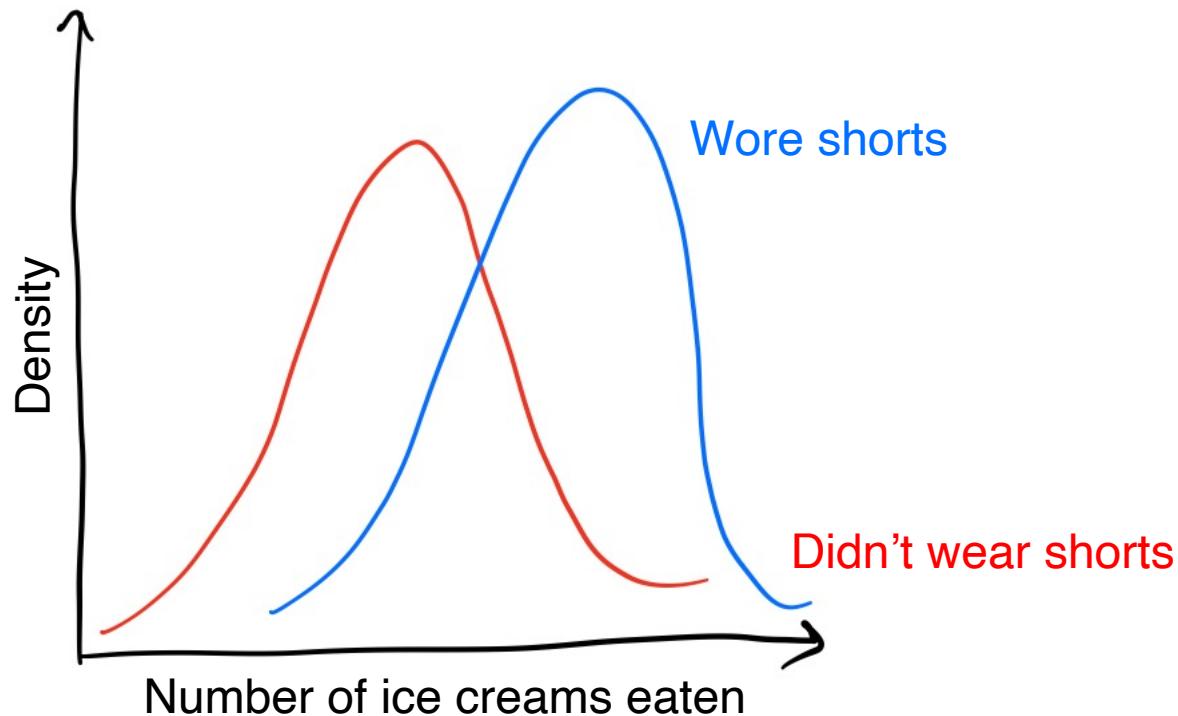
Sometimes distributions are ugly...



- More on this later
- Always check your data!

Relationships between variables

- Most of the time, what we're doing is comparing or summarizing relationship between the distribution of two or more variables



Distribution of ice creams eaten conditional on getting heatstroke

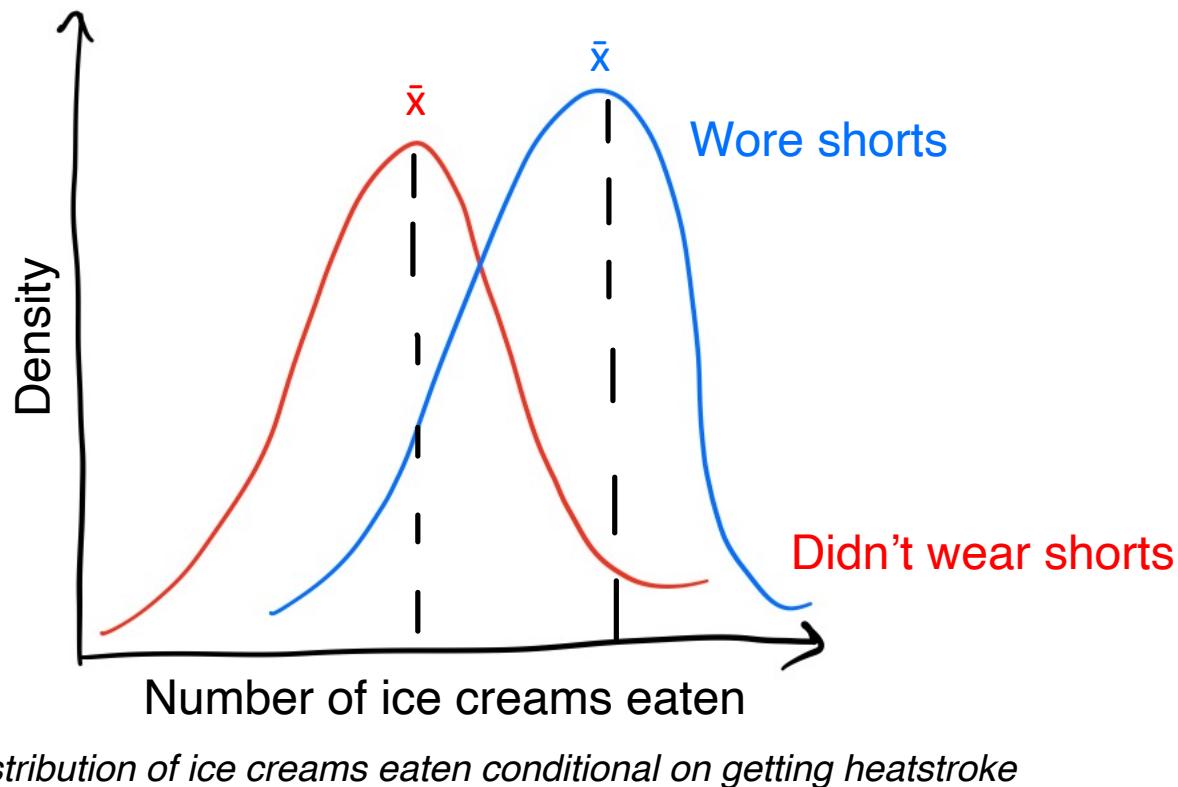
Does eating ice cream cause you to wear shorts?

Null hypothesis: **distribution of wore shorts = distribution of didn't wear shorts**

If **blue** different than **red** (according to some test statistic), we **reject our null hypothesis** and conclude wearing shorts makes you eat more ice cream

Relationships between variables

- Most of the time, what we're doing is comparing or summarizing relationship between the distribution of two or more variables



Does eating ice cream cause heatstroke?

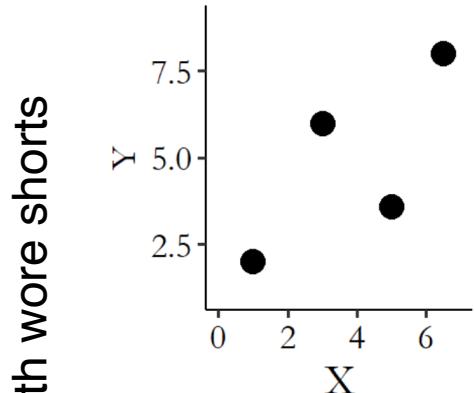
Null hypothesis: distribution of wore shorts = distribution of didn't wear shorts

If blue different than red (according to some test statistic), we **reject our null hypothesis** and conclude wearing shorts makes you eat more ice cream

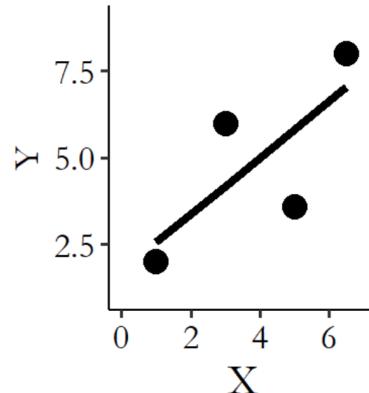
Often simplify by comparing conditional means (can't observe theoretical distribution)

Line time

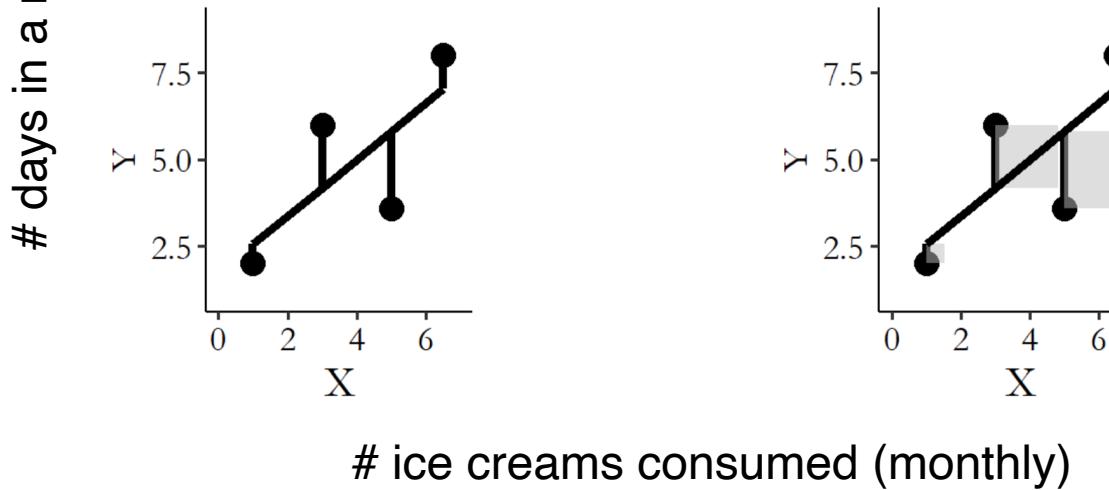
Let's fit a line to four points



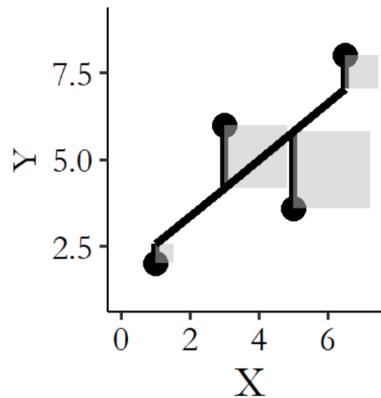
Add the OLS line



Residuals are from point to line



Goal: minimize squared residual



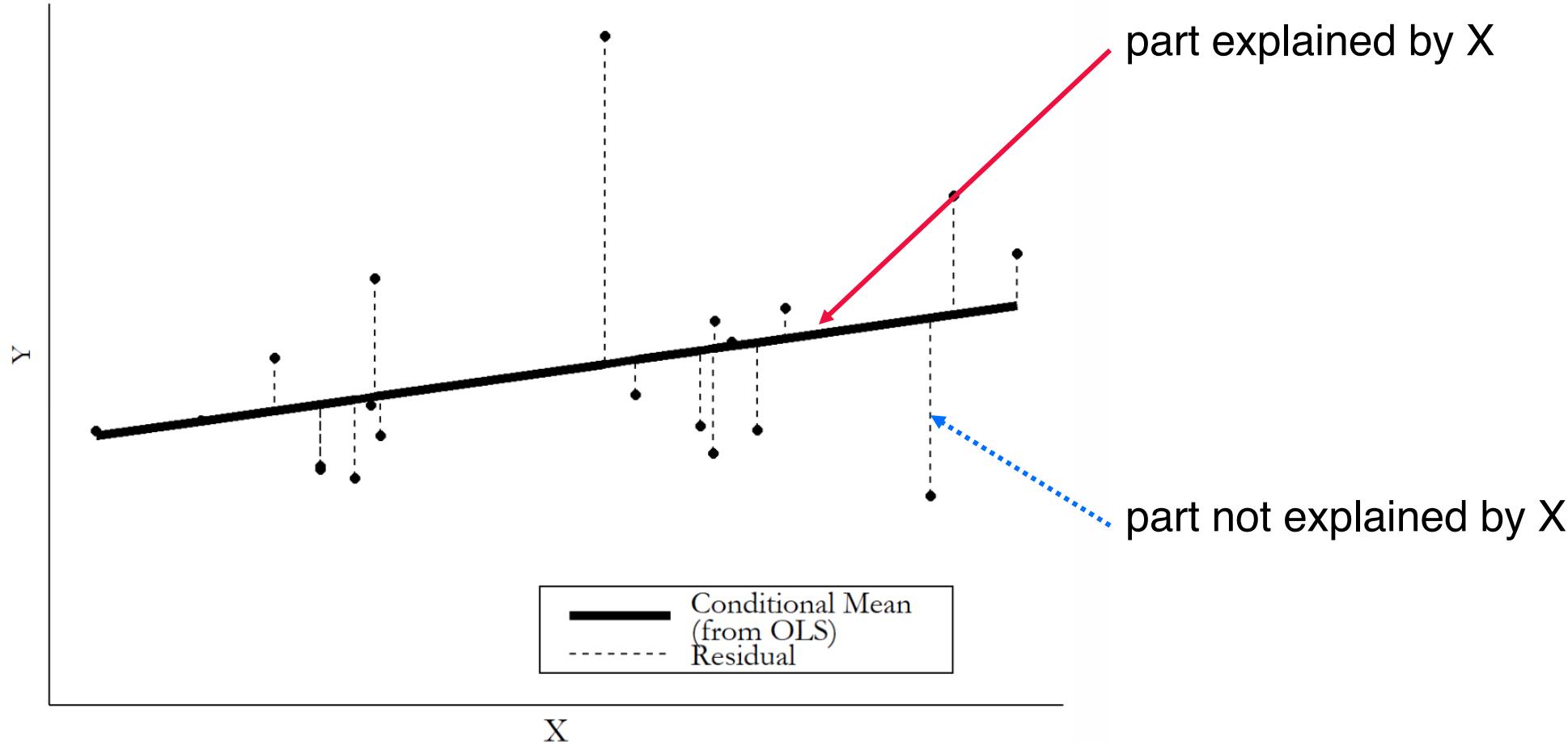
- We don't see theoretical distribution – instead, fit a line to observations

$$(7-2.5) / (6-1) = .9$$

$$y = 2.5 + .9x$$

For every ice cream you eat in a month, you wear shorts .9 extra days

OLS revisited



But eating ice cream doesn't make you wear shorts

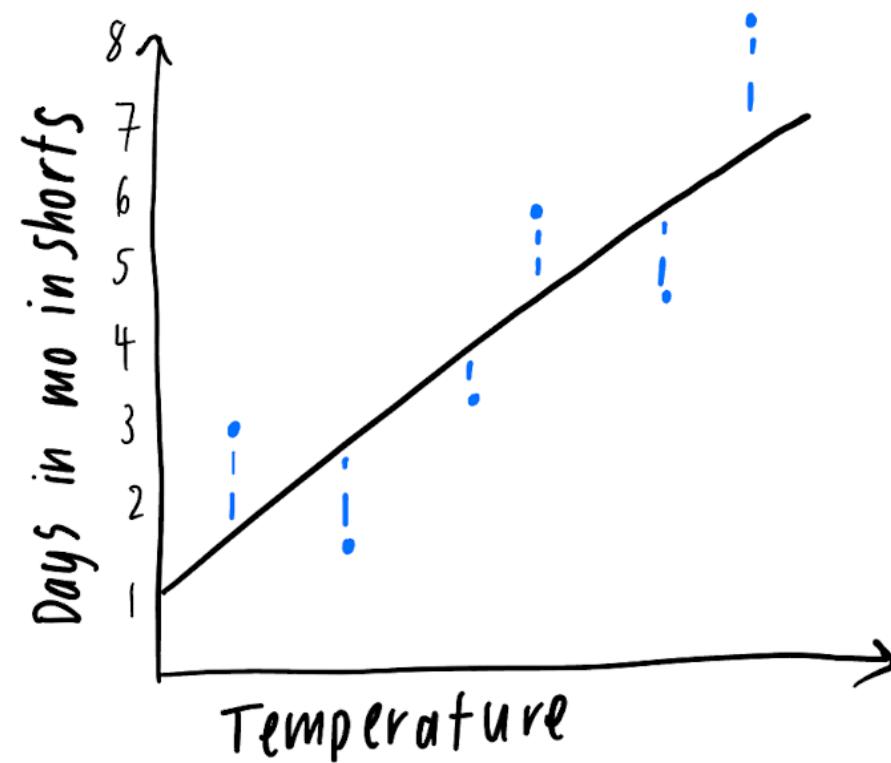
- And wearing shorts doesn't cause you to eat ice cream...
(reverse causality)
- What do you think is going on?

But eating ice cream doesn't make you wear shorts

- And wearing shorts doesn't cause you to eat ice cream...
(reverse causality)
- What do you think is going on?
- Need to control for weather
 - We will then compare **conditional conditional means**

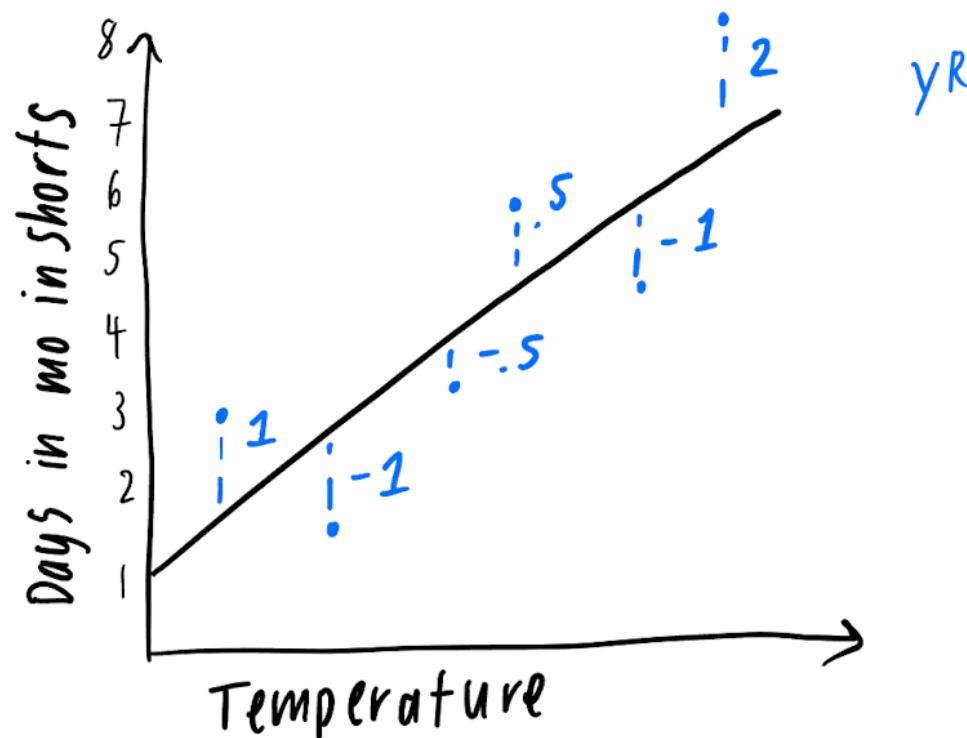
Controlling for variable(s)

1. Get the mean of Y conditional on Z



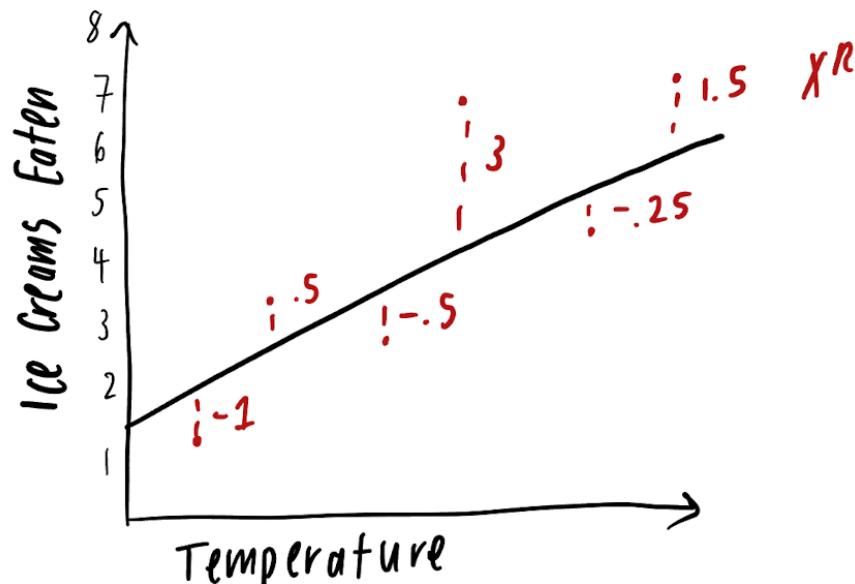
Controlling for variable(s)

1. Get the mean of Y conditional on Z
2. Subtract out that mean to get the residual of $Y = YR$



Controlling for variable(s)

1. Get the mean of Y conditional on Z
2. Subtract out that mean to get the residual of Y = Y_R
3. Get the mean of X conditional on Z
4. Subtract out that conditional mean to get the residual of X = X_R



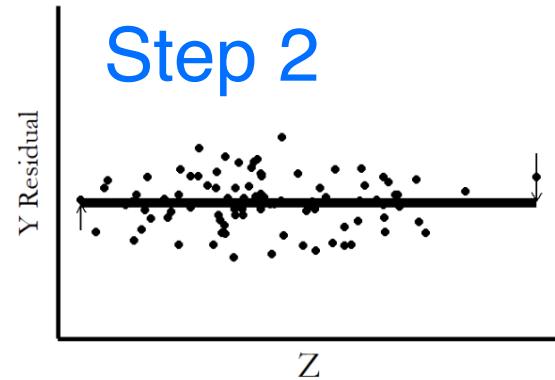
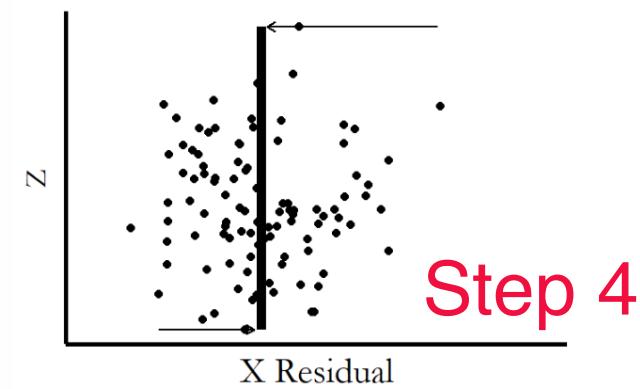
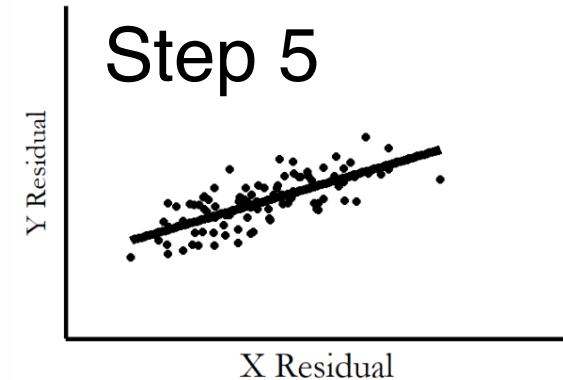
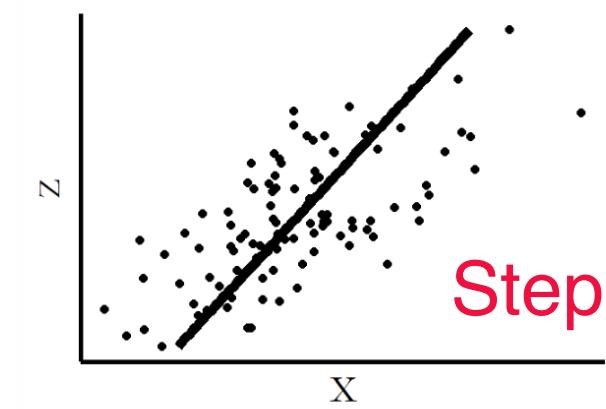
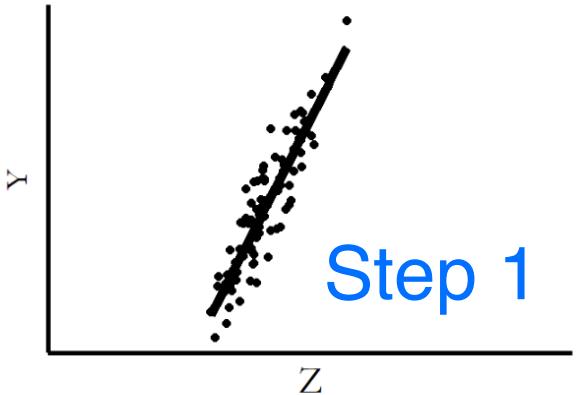
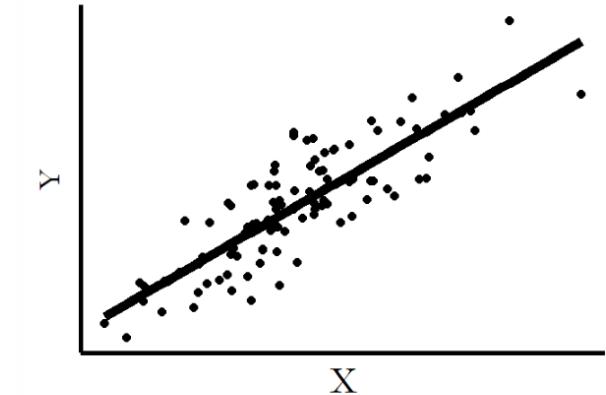
Controlling for variable(s)

1. Get the mean of Y conditional on Z
2. Subtract out that mean to get the residual of Y = YR
3. Get the mean of X conditional on Z
4. Subtract out that conditional mean to get the residual of X, XR
5. Describe the relationship between XR and YR

This is the part of the relationship between X and Y not explained by Z

= *Mean of Y conditional on X conditional on Z*

How much of a relationship is there between ice cream and shorts that isn't explained by weather?



Upcoming deadlines and wrap-up

Upcoming deadlines

- Send me your preferences for paper to present by 5pm on Thursday (Jan 25)
- “Why Study School Integration?” due 5pm Monday (Jan 29)
- Come to OH this week or next (I will keep track)

Why Study School Integration? The Evolution of Segregation and Education

Due Date: Monday, January 29 at 5pm

5% of final grade

EC 970: Sophomore Tutorial
Harvard University Department of
Economics

Danielle Graves

In this assignment, you will respond to the arguments made by both sets of authors. Are their viewpoints necessarily in conflict with each other? What are the strengths and weaknesses of each argument? How do the authors cite data and economic methods to back up their findings?

What do the authors list as the benefits and weaknesses of school integration? When does it fail? When does it succeed? What does it mean for school integration to succeed or fail? What do *you* think matters when studying school integration?

This should be 3-4 pages in length (double-spaced, Times New Roman 12 pt font, 1-inch margins).

Please submit your papers to canvas as a word or PDF document by **5pm on Monday, January 29**. Please name your file lastname_a1.

Assignment structure

You will write a 3-4 page paper responding to the following:

- Diette, Timothy M.; Hamilton, Darrick; Goldsmith, Arthur H.; Darity, William A. 2021. "Does the Negro Need Separate Schools? A Retrospective Analysis of the Racial Composition of Schools and Black Adult Academic and Economic Success". *RSF : Russell Sage Foundation journal of the social sciences*, 2021, 7(1): 166-186 [link](#)
- "Rucker Johnson: Why School Integration Works." [link](#)
This is a Youtube video summarizing his 265 page book. It runs 1:15

My goal is to familiarize and engage you with these arguments before we dive into the extensive literature on school integration. I'm not looking for a rigid paragraph-by-paragraph structure.

Think of this as a warm up for the class—a long-form response paper. In future papers, I will ask you to incorporate citations to back up your points, but there's no requirement to include citations to other sources for this paper.

What we did today

- Syllabus, course policies, assignments
- Writing Economics
 - Structure of papers – generally a clear roadmap: intro, lit review, setting, data, methods, results, conclusion, discussion and policy implications
- Assignment for next Monday