

Education A830: Data in International Development

Emmerich Davies

Class: T 3:30-5:30 PM

Office: Gutman 412 or <https://harvard.zoom.us/my/eeemda>

Email: emmerich_davies_escobar@gse.harvard.edu

Class Room: Gutman 437

Office Hours: Th 10:00AM-12:00PM [Calendly](#)

Teaching Fellow: Nicolás Riveros

Office Hours: Monday 2:30-4:00 PM [Calendly](#)

Email: nriveros@g.harvard.edu

Lab Time: Th 3-4:15 PM

Lab Location: Larsen G01

Faculty Coordinator: Wendy Angus

Email: wendy_angus@gse.harvard.edu

Canvas Site: <https://canvas.harvard.edu/courses/118573>

The production and interpretation of data has become increasingly important for policymakers, politicians, and researchers in international development. Decisions are made based on what the data says, and a lot of effort is spent on collecting data. In this class, we will take a hands on approach to understanding data. We will use several datasets common in international development – including global performance assessments such as PISA, a national performance assessment dataset, a national census, and a public opinion dataset – to better understand how the data is collected and used, what we can and cannot say with the data, and what the data tells us about education in low- and middle-income contexts. Class assignments will include problem sets and a final project applying tools and concepts learned in class to a topic of interest to students. We will also expand on the core conceptual building blocks taught in S-040 with a focus on application, interpretation, and visualization. Classes will be primarily taught using the R programming language, although users of other programming languages are welcome to take the course and use their preferred language. By the end of the course, students will understand the production and application of major datasets used in international education, as well as use the skills in this class to use and critique a broad range of data.

How We Will Learn

Learning Goals

This course has three learning goals:

1. Provide a broad introduction to some of the most common sources and types of data used in international education research today.
2. Provide a hands-on introduction to data analysis and visualization, primarily through the [tidyverse](#) and the grammar of graphics.
3. Create informed consumers of data through critique and reflection of goals 1 and 2.

Course Progression

Every week we will engage with three readings: a reading that uses the type of data we are engaging with in that week's class, a policy or technical document that uses the data we are reading about for that week's class, and a

critique of the use of data in international development. The first reading is to provide a theoretical foundation about how we should think about using that dataset, the second to provide a technical foundation for the coming problem set, and the final reading so we are aware of the limitations of that and other data.

COVID

The world is not okay. You may be dealing with illness, or its aftermath. You may know people who have lost their jobs, have become sick, have been hospitalized, or perhaps have even lost their lives. You may have increased work and/or increased family care responsibilities. You may be facing uncertain job prospects, or have lost a job you depended on before coming to Harvard. The pandemic, the U.S. and other government's inept responses to the pandemic, and further cruel migration policies prevented many of you from easily coming to the United States and have made focusing on your academic and professional lives more difficult than normal. Despite our return to campus and the promise of a better tomorrow vaccines have brought, with masks, regular testing, contact tracing, and potential periods in isolation and quarantine to protect our most vulnerable, this looks different to what many of us have been accustomed to. I want you to learn lots of things from this class and be engaged with your peers, but, more importantly, I want you to stay healthy, balanced, and grounded during this crisis. To that end, let's all aim to be flexible and supportive of each other. Please don't hesitate to ask for help if you need it and as a teaching team, we will try support you as best we can. You should also know that HGSE is committed to these same goals and has a number of resources to support students that you can find at <https://osa.gse.harvard.edu/student-support-tips-sheets>.

As part of this effort to help you learn in whatever way helps you be the best student you can be, we will record classroom sessions by default. This is being done to encourage classroom safety and manage excused absences due to the potential for quarantine and isolation due to illness. No one outside of the members of our class and teaching team is or will be permitted to access the videos. No one is permitted to download or make any derivative use of any recording. If you do not wish to be visible in the video, please contact the teaching team.

Diversity, Equity and Inclusion

HGSE is committed to a diverse, equitable, and inclusive environment. If there are aspects of the design, instruction, and/or experiences within this course that result in barriers to your inclusion or to an accurate assessment of your achievement in the course, please notify the instructor as soon as possible. If you are a student with a disability and wish to request accommodations, please contact KellyAnn Robinson, Ph.D., Associate Director of Student Support Services for an appointment. Because many accommodations require early planning, requests for accommodations should be made as soon as possible.

Communication

The primary means of communication will be over slack – especially for courses such as this, we believe the open and conversational format of slack will enhance your learning experience in the course and allow for productive conversations around class material and final projects.¹ Emmerich and Nicolás will be readily available over Slack,

¹We fully recognize that there is tremendous Slack fatigue across HGSE and Harvard, so we have spent some time thinking about the benefits and drawbacks of using this as our primary means of communication. We hope that you agree and find it useful, but please let us

and you hope you feel comfortable participating there too in a democratic spirit for the class. We also encourage you to support your peers on slack if they have questions you know the answer to. While the teaching team are the “experts”, our expertise comes at the cost of distance from the difficulties of the learning process – we once struggled to learn the material and application in the same ways that you will struggle. But we are now years (and decades in Emmerich’s case!) from remembering how we overcame those difficulties. You and your peers will have a much better sense of how you learned, and we want to encourage you to be generous with your learning.

Assignments

There will be four assignments in this course, three problems sets that build on material we cover in class, and a final project to be designed in consultation between the teaching team and groups.

Problem Sets: 25% of your grade each

There will be three problem sets over the course of the semester. We encourage you to start working on the problem sets early, especially as they involve working with large datasets that will take time to analyze on your computer. These problem sets are posted on the course canvas page and ready for you to explore and begin working on as soon as you are ready. At the beginning of each problem set, you will notice a section that reads **Do Now**. We encourage you to do that section early in the semester to troubleshoot any issues you may have. We also encourage you to work in teams, but if you do, to list your peers you worked with at the top of your problem set along with your name.

1. Problem Set 1, due on **Friday, February 10 at 5PM**
2. Problem Set 2, due on **Friday, April 14 at 5PM**
3. Problem Set 3, due on **Friday, April 28 at 5PM**

Final Project: 25% of your grade

For the final project, you will pick a dataset (or datasets) at the intersection of education and international development of your choosing. It can be a dataset we have worked with in the class, something you are working on for a project outside of class, or something different altogether. Please find some time to meet with one of the members of the teaching team early in the semester to discuss the dataset or idea you have in mind so we can give you feedback on the suitability of the dataset.

Then we want you to ask an answerable and falsifiable question that can be answered using that dataset. To do this, you’ll need to know your data – the questionnaire, sample, and variables – well to be able to ask, and answer, the question from the data at hand.

Finally, interpret the results of your inquiry for us. What do we know now that we did not know before? How should we think about this question now that you have done this work?

Some questions that students have asked in previous iterations of this class are,

- “What are the differences in educational attainment of the same ethnic group across national borders in sub-Saharan Africa?” This student used the Demographic and Health Survey (DHS) of Liberia and Sierra Leone to explore whether an ethnic group that lived in both countries had different levels of educational attainment and explore why that might be so.

know if you have any concerns.

- “What relationship does educational attainment have on beliefs about corruption in Ukraine?” Using public opinion data from Ukraine and questions about perceptions of corruption, this student explored how an individual’s perceptions of corruption changed by levels of education in Ukraine.
- “Are better resourced schools more likely to have school management committees (SMCs) in India?” Using the District Information System for Education (DISE), this student explore the relationship between School Management Committees (SMCs) and school infrastructure and resources in India.

For the final project, we encourage you to work in a group, although it is not mandatory.

To scaffold the final project, we provide these rough suggestions for *when* you should have completed steps to the final project:

1. Tuesday, January 31: Identify others in the class who have similar substantive or geographic interests
2. Friday, February 17: Meet with a member of the teaching team to discuss your question, and identify or discuss potential datasets that could be used to answer your question.
3. Tuesday, March 28: Have conducted at least a preliminary analysis of your data, including a table of summary statistics, plots of distributions or percentages of key outcomes, scatter plots of changes over time or space of your key variables of interest.
4. Week of Tuesday, April 18 and Tuesday, April 25: Present early questions and findings from your project to the class in the main class time or section.
5. Final project due **Friday, May 5 at 5PM**

These are loose guidelines and we recognize that the semester moves at different speeds for everyone. The only hard and fast deadlines are for the final presentation and submission of the final project.

The final project will be graded as follows:

1. 5% of final grade: Strength of question and ability to answer it with the dataset(s) chosen.
2. 10% of final grade: Class presentation on the week beginning Tuesday, April 18. Please note that the length of these presentations and what we can expect and fit into them will depend on how many groups are formed in the class.
3. 10% of final grade: Quality of presentation and ease of interpretation of data analysis, tables, and figures.
4. 25% of final grade: Individual write-up of the final project, including discussing what you found, how you found it, why it matters and what we learn about education and development policy and practice, the limitations of the data, what you would need to better answer the question (we recognize your ability is limited in the time of half a semester!).

We expect the final projects to be around 10 pages double-spaced. This is a loose guideline and we will by no means penalize final projects that are longer or shorter than this as long as they do a good job answering the question it sets out to answer.

Overview of the Schedule

1. **Tuesday, January 24: Why Use Data?**
2. **Tuesday, January 31: Growth Data**
3. **Tuesday, February 7: International Large Scale Assessments**

4. Tuesday, February 14: Domestic Assessments
5. Tuesday, February 21: No Class: Comparative International Education Society Annual Conference
6. Tuesday, February 28: Education Over Time Through National Censuses
7. Tuesday, March 7: Health, Education, Consumption, and Time-Use Through Household Surveys
8. Tuesday, March 21: Issues of Measurement: Identity
9. Tuesday, March 28: Public Opinion: A Loud But Noisy Signal?
10. Tuesday, April 4: COVID
11. Tuesday, April 11: Collecting Your Own Data
12. Tuesday, April 18: Bringing it All Together Part I: Student Presentations
13. Tuesday, April 25: Bringing it All Together Part II: Student Presentations

Tuesday, January 24: Why Use Data?

- **Arc of the Week:**

- We will begin the course reflecting on how data has been used in our personal and professional lives, and during our time at HGSE. We will engage with three readings that are grappling with the purpose of data in professional and personal lives.

- **Guiding Questions:**

- What are examples of how data has been used or you have used data in your personal life? Your professional life? At HGSE?
- What were the guiding principles for this use? What were the underlying philosophies and end goals that drove this use?
- Are there alternatives to this use of data?
- What does it mean to “see like a state”?

- **Readings for Class:**

- **Academic:** Mesquita, Ethan Bueno de. 2013. “The Aims of Public Policy Address: The Perils of Quantification.” Address to the Incoming Class at the Harris School of Public Policy. <https://home.uchicago.edu/bdm/PDF/aims.pdf>.
- **Policy:** The World Bank. 2021 *World Development Report 2021: Data For Better Lives*. Washington, D.C.: The World Bank. Chapter 1 “Harnessing the value of data for the poor”, pgs. 23-52. <http://hdl.handle.net/10986/35218>.
- **Critical:** Scott, James C. *Seeing like a State: How Certain Schemes to Improve the Human Condition Have Failed*. New Haven: Yale University Press, 1998. Introduction and Chapter 1, pgs. 1-52

- **Assignments:**

- Introduce yourself in the course Slack. Share your name, your program at Harvard, what you were doing before coming to Harvard, where you are from, why you are interested in taking this course, and some questions you are potentially interested in working on for the final project.

- **Lab:**

- In lab this week, we will get set-up with the basics of R, the tidyverse, and ggplot. We will ensure everyone has R installed and the main suite of packages we will be using throughout the semester. We will also begin to load and explore the growth data we will be discussing in class on Tuesday, January 31.

Tuesday, January 31: Growth Data

- **Arc of the Week:**

- The correlation between economic growth and human well-being is well established. This week, we will explore the various ways we measure economic growth, whether they tell us anything about human well-being causally, and the relationship between economic growth and educational outcomes.

- **Guiding Questions:**

- What are the primary ways we measure economic growth over time?
- What is the relationship between economic growth and human well-being?

- **Readings for Class:**

- **Academic:** Angrist, Noam, Pinelopi Koujianou Goldberg, and Dean Jolliffe. 2021. "Why Is Growth in Developing Countries So Hard to Measure?" *Journal of Economic Perspectives* 35:3: 215-42.
- **Critical:** Jerven, Morten. 2013. *Poor Numbers: How We Are Misled by African Development Statistics and What to Do about It*. 1 edition. Ithaca: Cornell University Press. Pgs. 1-32.

- **Lab:**

- In lab this week, we will begin to think about various ways to display differences between the various sources of growth data through figures and tables. This will provide an introduction to the first part of the first problem set.

Tuesday, February 7: International Large Scale Assessments

- **Arc of the Week:**

- In this week, we will explore the most common form of international large-scale assessments, PISA, and discuss other assessments. We will discuss their purpose and uses and whether they fulfill those purposes. We will also introduce a framework for thinking about the downsides of data and tracking and why this might be something we should be concerned about.

- **Guiding Questions:**

- What can PISA and other international large-scale assessments tell us about a country's educational system over time?
- How consistent are the comparisons over time?
- How can we use PISA and other international large-scale assessments to make policy?

- **Readings for Class:**

- **Academic:** Le Nestour, Alexis, Laura Moscoviz, and Justin Sandefur. 2022. “The Long-Run Decline of Education Quality in the Developing World.” Working Paper. CGD Working Paper Series. Washington, D.C.: Center for Global Development <https://www.cgdev.org/sites/default/files/long-run-decline-education-quality-developing-world.pdf>.
- **Policy:** Chen, Li-Kai, Felipe Child, Emma Dorn, and Raimundo Morales. An interview with former Peruvian Minister of Education Jaime Saavedra, September 2019. [link](#).
- **Critical:** Fontdevila, Clara. “The Politics of Good Enough Data. Developments, Dilemmas and Deadlocks in the Production of Global Learning Metrics.” *International Journal of Educational Development* 96 (January 1, 2023).

- **Lab**

- In lab this week, we will begin playing with the primary international large-scale assessment data source: PISA. We will load and begin summarizing it in our lab session.

Tuesday, February 14: Domestic Assessments

- **Arc of the Week:**

- In this week, we will explore domestic assessments that look to track and evaluate student learning *within* countries, and compare their utility to international large scale assessments (ILSAs). We will use the case of *Prova Brasil* and the *Índice de Desenvolvimento da Educação Básica* (IDEB) from Brazil – one of the most comprehensive domestic assessments tools – and the Annual Status of Education Report (ASER) from India, one of the most widely adopted and emulated ILSAs.

- **Guiding Questions:**

- What questions about the quality of a country’s education do domestic assessments allow us to answer that international large-scale assessments do not?
- How do IDEB and ASER overcome some of the challenges of assessing hard to reach populations?
- How should we think of domestic assessments in light of Samarth Bansal’s investigative journalism work?

- **Readings for Class:**

- **Academic:** Djaker, Sharnic, Alejandro J. Ganimian, and Shwetlena Sabarwal. 2022. “Primary- and middle-school teachers in South Asia overestimate the performance of their students.” [link](#).
- **Policy:** ASER. January 18, 2023. “Annual Status of Education Report (Rural) 2022.” New Delhi, India. pgs. 15-33 & 38-51. [link](#).
- **Critical:** Bansal, Samarth. December 18, 2021. “#1: Inside a Fictional Education Revolution.” <https://www.samarthbansal.com/rajasthan-education-nas/>.

- **Lab:**

- In lab this week, we will begin to explore how to represent change over *time* and *space* through the IDEB and ASER surveys.

- **Assignments:**

- Meet with the teaching team by **Friday, February 17** to discuss potential ideas and datasets that you wish to analyze for your final project. For this meeting, you should come prepared with two things:

1. The dataset(s) you wish to use
 2. Any questions you want to ask about the dataset(s)
- In the meeting, we will discuss the feasibility of your questions over the course of the semester, suggestions for other data you might want to think about, and how to refine your questions to a manageable project over the scope of the course.

Tuesday, February 21: No Class: Comparative International Education Society Annual Conference

- **Assignments:**

- Problem Set 1 is due on **Friday, February 24 at 5PM**

Tuesday, February 28: Education Over Time Through National Censuses

- **Arc of the Week:**

- In this week, we will look at the broadest sources of data national governments have to understand their populations: national censuses. Given how much policy and how much is derived from national censuses, we will explore how they empower and limit policymakers. We will also explore change over *time* – we will practice making comparisons of the same units (geographic units in this case) and time.

- **Guiding Questions:**

- What are the benefits and tradeoffs censuses have? Are there questions you wish the census asked they do not? Keep in mind in answer to this question there are tradeoffs in terms of time and money (it is expensive to field a census!) and politics (national governments cannot be seen, or do not want to) ask certain questions about their populations.
- How has educational attainment changed over time in various countries? Are things getting better? Why or why not?

- **Readings for Class:**

- **Academic:** Lieberman, Evan S., and Prerna Singh. 2017. “Census Enumeration and Group Conflict.” *World Politics*. 69(01): 1-53.
- **Policy:** Angrist, Noam, Harry Patrinos, Simeon Djankov, Pinelopi Goldberg. February 8, 2022. “Mapping the Global Learning Crisis”. *Education Next*. [link](#)
- **Critical:** Ruggles, Steven, and Diana L Magnuson. 2020. “Census Technology, Politics, and Institutional Change, 1790-2020.” *Journal of American History*. 19-51.

- **Lab:**

- In lab this week, we will begin discussing how to merge datasets over time using the example of the Indian national censuses. We will also start thinking about calculating change between units over time.

Tuesday, March 7: Health, Education, Consumption, and Time-Use Through Household Surveys

- **Arc of the Week:**

- In this week, we will explore the relationship between health and education. We will use the Demographic and Health Survey (DHS) collected by the United States Agency for International Development to explore this relationship. We will also discuss other data sources, such as the IPUMS Performance Monitoring for Action (PMA) fielded by the Bill & Melinda Gates Foundation and Johns Hopkins University.

- **Guiding Questions:**

- What is the relationship between health, education, and schooling?

- **Readings for Class:**

- **Academic:** Frye, Margaret. 2017. “Cultural Meanings and the Aggregation of Actions: The Case of Sex and Schooling in Malawi.” *American Sociological Review*.
- **Policy:** Corsi, D. J., M. Neuman, J. E. Finlay, and S.V. Subramanian. 2012. “Demographic and Health Surveys: A Profile.” *International Journal of Epidemiology* 41(6): 1602-13.
- **Critical:** Dréze, Jean, and Anmol Somanchi. June 21, 2021. “View: New barometer of India’s economy fails to reflect deprivations of poor households.” *The Economic Times*. [link](#).

- **Lab:**

- In lab this week, will begin to explore the richness of household surveys through exploring the Demographic and Health Surveys from various countries, and make comparisons *across* and *within* countries using these surveys and attempting to standardize variables across the surveys. We will also introduce presenting information spatially through the use of maps, and rely on the [The Database of Global Administrative Areas](#) for our map data.

Tuesday, March 21: Issues of Measurement: Identity

- **Arc of the Week:**

- The use of data and data analysis relies on some underlying assumptions, such as the stability of measures over time and between contexts. What happens when those assumptions are violated?

- **Guiding Questions:**

- What issues arise in the measurement of identity?
- How do these translate to the measure of other concepts we are interested in measuring such as class, income, social ties?
- What implications does this have for our inference and use of data?

- **Readings for Class:**

- **Academic:** Brubaker, Rogers. 2004. *Ethnicity without Groups*. Cambridge, MA: Harvard University Press. Chapter 2 pgs. 28-63. [link](#)

- **Policy:** Davenport, Lauren. “The Fluidity of Racial Classifications.” *Annual Review of Political Science* 23(1): 221-40.
- **Critical:** Zuberi, Tukufu. 2003. *Thicker Than Blood: How Racial Statistics Lie*. 1 edition. Minneapolis: Univ Of Minnesota Press. Ch. 6. [link](#)

- **Lab:**

- In lab this week, we will begin to explore how group identification can change over time and by who is measuring identity using

- **Assignments:**

- Problem Set 2 is due at **Friday, March 24 at 5PM**

Tuesday, March 28: Public Opinion: A Loud But Noisy Signal?

- **Arc of the Week:**

- In this week, we will consider what public opinion data can tell us about citizens demands for education policy. We will begin by considering theory from Western Europe about what citizens are telling their politicians and pollsters about education, before turning to a problem set on education polling in sub-Saharan Africa. We will use the Afrobarometer survey from multiple countries across sub-Saharan African to understand what citizens think about education

- **Guiding Questions:**

- What does it mean for public opinion on education to be loud and noisy?
- Does public opinion on education follow a similar pattern in sub-Saharan Africa as it does in Western Europe?
- Given the sampling frame of the Afrobarometer data, how representative is the data of the population as a whole?
- What are Zuberi’s concerns with statistics that attempt to classify identity

- **Readings for Class:**

- **Academic:** Busemeyer, Marius R., Julian L. Garritzmann, and Erik Neimanns. 2020 *A Loud but Noisy Signal? Public Opinion and Education Reform in Western Europe*. Cambridge University Press. Chapter 3. pgs. 57-99
- **Policy:** Sahgal, Neha, Jonathan Evans, Ariana Monique Salazar, Kelsey Jo Starr, and Manolo Corichi. 2021. “Religion in India: Tolerance and Segregation.” Pew Research Center. [link](#).
- **Critical:** The Economist. September 29, 2022. “A study of lights at night suggests dictators lie about economic growth.” *The Economist*. [link](#)

- **Lab:**

- In lab this week, we will spend time understanding public opinion surveys and how to best interpret public opinion data like Likert scales.

Tuesday, April 4: COVID

- **Arc of the Week:**
- **Guiding Questions:**
 - Given what we have learned about data collection so far, what tools are available to policy-makers, researchers, and governments to understand how COVID has impacted citizens in low-income countries?
 - What additional challenges would a pandemic like COVID pose for data collection?
- **Readings for Class:**
 - **Academic:** Angrist, Noam, Peter Bergman, and Moitshepi Matsheng. 2020. “School’s Out: Experimental Evidence On Limiting Learning Loss Using ‘Low-Tech’ In A Pandemic.” *National Bureau of Economic Research*. Working Paper 28205. <https://www-nber-org.ezp-prod1.hul.harvard.edu/papers/w28205>.
 - **Policy:** IDInsight. February 2, 2021. “Economic Effects of COVID-19: Rapid Surveys of Rural Households in India.”

Tuesday, April 11: Collecting Your Own Data

- **Arc of the Week:**
 - For this week, we will turn our attention to ourselves and attempt to collect our own data. We will engage with readings that lay bare the process, from a classic in the field of sociology and emancipatory data collection, to more technocratic pieces of data collection. For lab and in the problem set, we will work in groups to design, deploy, and collect our own data using Qualtrics.
- **Guiding Questions:**
 - What are the challenges in collecting your own data?
 - What are the tradeoffs that you need to contend with and how do you resolve these tradeoffs?
- **Readings for Class:**
 - **Academic:** DuBois. W.E.B. 1967. *The Philadelphia Negro*. Schocken Books. New York. Chapters 1-2, 8, & Appendix A (pgs. 1-9, 83-96, & 400-410). [link](#).
 - **Policy:** Zwane, Alix Peterson, Jonathan Zinman, Eric Van Dusen, William Pariente, Clair Null, Edward Miguel, Michael Kremer, et al. 2011. “Being Surveyed Can Change Later Behavior and Related Parameter Estimates.” *Proceedings of the National Academy of Sciences of the United States of America*. 108(5): 1821-26.
 - **Critical:** Phillips, Trisha. 2021. “Ethics of Field Experiments.” *Annual Review of Political Science* 24(1): 277-300.

Tuesday, April 18: Bringing it All Together Part I: Student Presentations

- **Arc of the Week:**

- In this week, we will spend class and lab time on student presentations of their independent projects. We expect you to be active and engaged participants while your peers present the projects they have worked on throughout the semester. To facilitate your own work on your final project and engagement with your peer's presentations, there will be no readings for the class.

- **Lab:**

- Lab will serve as open office hours in which you can ask questions about your final project or the problem set.

- **Assignments:**

- Problem Set 3 due **Friday, April 21 at 5PM**

Tuesday, April 25: Bringing it All Together Part II: Student Presentations

- **Arc of the Week:**

- In this week, we will spend class and lab time on student presentations of their independent projects. We expect you to be active and engaged participants while your peers present the projects they have worked on throughout the semester. To facilitate your own work on your final project and engagement with your peer's presentations, there will be no readings for the class.

- **Lab:**

- Lab will serve as open office hours in which you can ask questions about your final project.

- **Assignments:**

- Final project due **Friday, May 5 at 5PM**