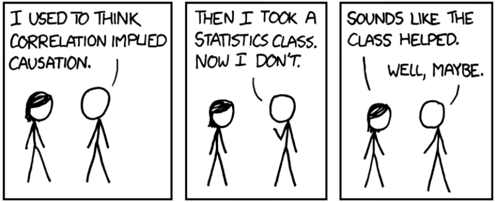
What Makes a(n Economics) Paper a Paper?

Types of Economics Research

- Empirical/applied microeconomics
  - New data, broadly defined
  - Causal estimates of policy impacts, broadly defined
- Microeconomic theory
  - A new mathematical model of human behavior (almost no one can do this)
- Macroeconomics
  - I literally have no idea what makes a macro paper a paper, maybe growth regressions

How Not to Write an Economics Paper



source: xkcd

- Data Assignment 1 asked you to look at the relationship between two variables
  - Economists often call this type of analysis is "descriptive" or "observational"
  - They don't mean it in a nice way

Economics Is a Method

- Gary Becker talks about "the economic approach to human behavior"
  - He's talking about theory, not empirics
  - Equally true of applied micro today
- It is natural to have research questions you are interested in (a research agenda)
  - To write an economics paper, you need more than a compelling research question
  - Need to be able to convince people that you are expanding what we know on a topic
  - Can't usually do this by examining correlations in widely available data sets
  - Correlations can reflect **treatment effects** or **selection bias**

Types of "Impact of X on Y" Papers

- **Randomized experiments**
  - A **treatment** of interest is randomly assigned, as in a medical trial
  - When this is true, regressing Y on X does (more or less) tell us about causal effect
- **Quasi-experimental research designs**
  - Difference-in-differences: compare changes in (non-random) treatment, comparison groups
  - Instrumental variables: find a source of **exogenous variation** in treatment (an **instrument**)
  - Regression discontinuity design: find a setting where access to program/policy depends on a continuous eligibility criterion with a strict cutoff, compare just above to just below
- **Natural experiments:** find a setting where policy is "as good as random" (similar to IV)

Randomized Experiments (RCTs): Examples

DO LABOR MARKET OPPORTUNITIES AFFECT YOUNG WOMEN'S WORK AND FAMILY DECISIONS? EXPERIMENTAL EVIDENCE FROM INDIA*

ROBERT JENSEN

Do labor market opportunities for women affect marriage and fertility decisions? We provided three years of recruiting services to help young women in randomly selected rural Indian villages get jobs in the business process outsourcing industry. Because the industry was so new at the time of the study, there was almost no awareness of these jobs, allowing us in effect to exogenously increase women's labor force opportunities from the perspective of rural households. We find that young women in treatment villages were significantly less likely to get married or have children during this period, choosing instead to enter the labor market or obtain more schooling or postschool training. Women also report wanting to have fewer children and to work more steadily throughout their lifetime, consistent with increased aspirations for a career. *JEL* Codes: I21, J12, J13, J16, J22.

Reshaping Adolescents' Gender Attitudes: Evidence from a School-Based Experiment in India†

By DIVA DHAR, TARUN JAIN, AND SEEMA JAYACHANDRAN*

This paper evaluates an intervention in India that engaged adolescent girls and boys in classroom discussions about gender equality for two years, aiming to reduce their support for societal norms that restrict women's and girls' opportunities. Using a randomized controlled trial, we find that the program made attitudes more supportive of gender equality by 0.18 standard deviations, or, equivalently, converted 16 percent of regressive attitudes. When we resurveyed study participants two years after the intervention had ended, the effects had persisted. The program also led to more gender-equal self-reported behavior, and we find weak evidence that it affected two revealed-preference measures. (*JEL* D63, D91, I21, J13, J16, 012)

Should You Run an RCT for Your Empirical Project?

- Randomized experiments are difficult to implement, and expensive, and time consuming
  - Almost every economist I know tells PhD students and assistant professors not to run RCTs
- Ways you might base your empirical project on an existing RCT (not an exhaustive list):
  - Find information on communities in sample, link to another source of data (this is hard)
    - Example: Professor Ozier's paper on spillover effects of medication to treat intestinal parasites
  - Discover an RCT that was conducted but not (fully) analyzed (also hard)
    - Example: Baranov, Bhalotra, Biroli, and Maselko (2020) estimate the medium-term impacts of treating maternal depression by following up participants from an earlier RCT in Pakistan
  - Find and then address a glaring omission in an RCT (with replication data available)
    - This should probably be a last resort, as far as choosing a project topic goes

Difference-in-Differences

- In difference-in-differences (or diff-in-diff or DiD), we compare **changes** in outcomes between treatment and comparison groups that are (usually) not randomly selected
  - Pre-treatment vs. post-treatment outcomes for a(n eventually) treated group of individuals, states, whatever and an untreated (as in not ever treated in your data) comparison group
- Identification assumptions (required for causal interpretation):
  - Selection bias reflects unchanging characteristics
  - Time trends and shocks are common to treated, untreated units
  - Typically described as a **common trends** assumption:
    - In the absence of treatment, treated and untreated units were changing at the same rate
- Diff-in-diff is often possible when other identification strategies (RCT, RD) are not

Difference-in-Differences: Empirical Specifications

- Simplest possible 2×2 diff-in-diff specification (with two periods of data):

$$Y_{i,t} = \alpha + \beta T_i + \gamma Post_t + \delta T_i \times Post_t + \varepsilon_{it}$$

  where:

  - $Y_{i,t}$ is outcome for unit i in period t
  - T_i is a dummy equal to one for units that are **ever** treated
  - $Post_t$ is a dummy for time periods after treatment starts
  - $T_i \times Post_t$ indicates the observations in your data that are actually treated (ever treated units in the post-treatment period)
  - ε_{it} is a conditionally mean-zero error term

Difference-in-Differences: Empirical Specifications

- Often use individual (unit) fixed effects instead of ever-treated-group dummy:

$$Y_{i,t} = \alpha + \nu_i + \gamma Post_t + \delta T_i \times Post_t + \varepsilon_{it}$$

  where:

  - $Y_{i,t}$ is outcome for unit i in period t
  - ν_i is a vector of individual fixed effects for states, people, whatever
  - $Post_t$ is a dummy for time periods after treatment starts
  - $T_i \times Post_t$ indicates the observations in your data that are actually treated (ever treated units in the post-treatment period)
  - ε_{it} is a conditionally mean-zero error term

Difference-in-Differences: Empirical Specifications

- With only two periods, we can also use the change in outcomes as the dependent variable:

$$\Delta Y_{i,t} = \alpha + \delta T_i + \varepsilon_{it}$$

  where:

  - $\Delta Y_{i,t}$ is the change in outcomes for unit i, pre vs. post
  - T_i is a dummy equal to one for units that are **ever** treated
  - ε_{it} is a conditionally mean-zero error term

- Don't need individual fixed effects (with only two periods) because they are differenced out

Difference-in-Differences: Empirical Specifications

- If you have many periods of data (which is good!), you can also use time fixed effects:

$$Y_{i,t} = \alpha + \nu_i + \phi_t + \delta T_i \times Post_t + \varepsilon_{it}$$

  where:

  - $Y_{i,t}$ is outcome for unit i in period t
  - ν_i is a vector of individual fixed effects for states, people, whatever
  - ϕ_i is a vector of time period fixed effects
  - $T_i \times Post_t$ indicates the observations in your data that are actually treated (ever treated units in the post-treatment periods)
  - ε_{it} is a conditionally mean-zero error term

- If treatment starts at different times in different locations/units, life is more complicated

Difference-in-Differences: Examples

"Momma's Got the Pill": How Anthony Comstock and *Griswold v. Connecticut* Shaped US Childbearing

By Martha J. Bailey*

The 1960s ushered in a new era in US demographic history characterized by significantly lower fertility rates and smaller family sizes. What catalyzed these changes remains a matter of considerable debate. This paper exploits idiosyncratic variation in the language of "Comstock" statutes, enacted in the late 1800s, to quantify the role of the birth control pill in this transition. Almost 50 years after the contraceptive pill appeared on the US market, this analysis provides new evidence that it accelerated the post-1960 decline in marital fertility. (*JEL* J12, J13, K10, N31, N32)

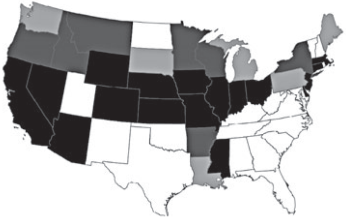


Figure 4. Geographic Distribution of Comstock Sales Laws by Type Circa 1960

Notes: No shading: states with no laws mentioning the "prevention of conception." Light gray: states banning only advertising or the distribution of information (Table 1, columns 2 and 3). Dark gray: states banning advertising and the sale of contraceptives but with physician exceptions (Table 1, columns 3 and 4). Black: states with sales bans and advertising bans with no exceptions for physicians (Table 1, columns 3 and 4)

Difference-in-Differences: Examples

Women's inheritance rights reform and the preference for sons in India☆

Sonia Bhalotra [a], Rachel Brulé [b], Sanchari Roy [c,*]

[a] *University of Essex, UK*
[b] *NYU Abu Dhabi, United Arab Emirates*
[c] *King's College London, UK*

A R T I C L E I N F O

JEL classification:
O12
K11
I21

Keywords:
Inheritance rights
Ultrasound
Female foeticide
Sex-selection
Son preference
Gender
India

A B S T R A C T

We investigate whether legislation of equal inheritance rights for women modifies the historic preference for sons in India, and find that it exacerbates it. Children born after the reform in families with a firstborn daughter are 3.8–4.3 percentage points *less* likely to be girls, indicating that the reform encouraged female foeticide. We also find that the reform increased excess female infant mortality and son-biased fertility stopping. This suggests that the inheritance reform raised the costs of having daughters, consistent with which we document an increase in stated son preference in fertility post reform. We conclude that this is a case where legal reform was frustrated by persistence of cultural norms. We provide some suggestive evidence of slowly changing patrilocality norms.

Fig. 1. The Five Reforming States (in green) and rest of India (in blue). Notes: The southern state of Andhra Pradesh split into two states - Andhra Pradesh and Telengana - in 2014. For the purposes of our analysis, these two are considered together as the undivided state of Andhra Pradesh that reformed in 1986. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Should You Use Diff-in-Diff for Your Empirical Project?

- Diff-in-diff is a promising approach when a policy is implemented in different places at different times, or implemented in some places but not in other reasonably similar places
  - Both the U.S. and India have sufficiently decentralized governance to allow for this
- Sometimes a policy change takes effect everyone, but will only impact some areas
  - Example: Professor Godlonton's work on informal birth attendants in Malawi
- Diff-in-diff is useful when:
  - You know of a policy change you want to study
  - You know of a data set with observations before and after the policy change, and with some variation in who was impacted by the policy (e.g. based on geography, age, gender etc.)

Instrumental Variables

- You'd like to run the regression

$$Y_i = \alpha + \delta T_i + \varepsilon_{it}$$

  to estimate the impact of treatment T_i on Y_i, but you "can't" because of selection bias

- An **instrument** is a variable that:
  - Predicts take-up of treatment T_i (first-stage)
  - Is as-good-as-random (exogeneity)
  - Doesn't have a direct effect on Y_i (exclusion restriction)

- If you have a good instrument, then (and only then) you can use **instrumental variables**

Instrumental Variables: Specifications

- IV estimated via two-stage least squares (2SLS) given instrument Z_i:

$$T_i = \alpha + \beta Z_i + \nu_i \quad \text{[first stage]}$$

$$Y_i = \gamma + \delta_{IV} \hat{T}_i + \varepsilon_i \quad \text{[second stage]}$$

- We can also look at the impact of our instrument on the outcome (through treatment):

$$Y_i = \alpha + \lambda T_i + \nu_i \quad \text{[``reduced form'']}$$

- IV estimate (from 2SLS second stage) is ratio of reduced form λ to first stage β

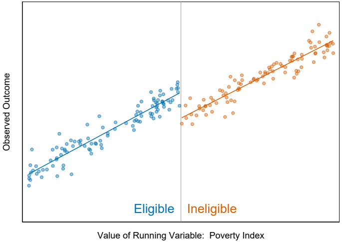
Instrumental Variables: Examples

- Weather as an instrument for the size of political rallies:
  - A famous paper by Madestam, Shoag, Veuger, and Yanagizawa-Drott (2013) shows that the Tea Party movement was stronger, and Republicans subsequently increased their vote share, in places where there as good weather on April 15, 2009 (the day of the first rally)
  - Magdalena Larreboure and Felipe González have a working paper showing similar effects of weather on the day of a protest from the 2017 Women's March on the 2018 House elections
- Sex composition of children as an instrument for future births:
  - Angrist and Evans (1998) use the sex composition of a couples' first two children (same sex?) to instrument for subsequent fertility, estimating impacts on labor supply

Should You Use IV for Your Empirical Project?

- Do you have an instrument?!?
  - Is it exogenous?
  - Does it satisfy the exclusion restriction? Really?
    - Exclusion restrictions cannot be tested, which is both good and bad
    - If you want to know if you idea for an instrument is convincing and satisfies the exclusion restriction, explain it to a professor who teaches ECON 255 and see if their face lights up
- You also need a **strong** first-stage relationship
  - This can be tested empirically
  - A rule-of-thumb is that your first-stage F-statistic should be at least 10 if not much higher

Regression Discontinuity



Assignment to treatment depends on a continuous variable (like an index, score, or vote share) with a sharp, known cutoff; those just above and just below the cutoff are otherwise similar

- Idea is to control flexibly for the **running variable**, estimate the jump at the cutoff

Regression Discontinuity: Empirical Specifications

- If you have many periods of data (which is good!), you can also use time fixed effects:

$$Y_i = \alpha + \delta D_i + \beta X_i + \gamma X_i \times D_i + \varepsilon_{it}$$

where:

  - Y_i is outcome for unit i
  - D_i is a dummy for being above (or below) the cutoff, and hence treated
  - X_i is the value of the running variable
  - ε_{it} is a conditionally mean-zero error term
- The key issue is choice of **bandwidth** around the discontinuity (to include in analysis)
- Also important to demonstrate absence of manipulation around the discontinuity

Regression Discontinuity: Examples

Journal of Public Economics 194 (2021) 104315

Contents lists available at ScienceDirect

Journal of Public Economics

journal homepage: www.elsevier.com/locate/jpube

Can gender quotas in candidate lists empower women? Evidence from a regression discontinuity design ☆

Manuel Bagues [a], Pamela Campa [b,*]

[a] The University of Warwick, Economics Department, United Kingdom
[b] SITE, Stockholm School of Economics, Sweden

A R T I C L E I N F O

Article history:
Received 20 September 2018
Revised 23 August 2020
Accepted 29 September 2020
Available online 23 December 2020

Keywords:
Gender quotas
Regression discontinuity design
Women in politics

A B S T R A C T

We provide a comprehensive analysis of the short- and medium-term effects of gender quotas in candidate lists using evidence from local elections in Spain. In the context of a closed list system with proportional representation, quotas were introduced in 2007 in municipalities with more than 5000 inhabitants, and were extended in 2011 to municipalities with more than 3000 inhabitants. Using a Regression Discontinuity Design, we find that quotas increased the share of women in candidate lists by around 8 p.p. and among council members by 4 p.p. However, within three rounds of elections, we do not observe any significant variation in several proxies of politicians' quality, the probability that women reach powerful positions such as party leader or mayor, or the size and composition of public finances. Overall, our analysis suggests that quotas in candidate lists fail to remove the barriers that prevent women from playing an influential role in politics.

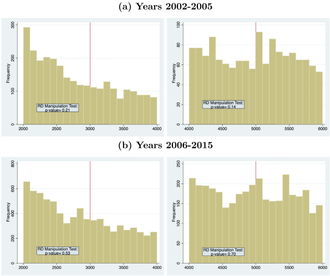


Fig. 5. Histograms of population. Note: Histograms of population in bins of 100 individuals for municipalities with a population close to the 3000 threshold (left-hand side) and municipalities with a population close to the 5000 threshold (right-hand side). Each figure also reports the result from the density test proposed by Cattaneo et al. (2019) performed at the corresponding cutoff.

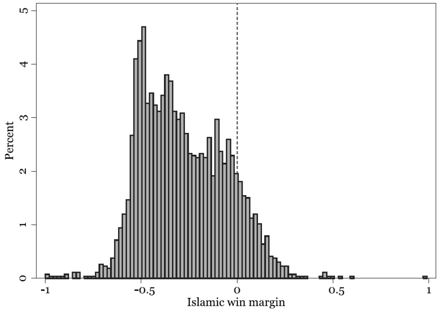
Regression Discontinuity: Examples

Econometrica, Vol. 82, No. 1 (January, 2014), 229–269

ISLAMIC RULE AND THE EMPOWERMENT
OF THE POOR AND PIOUS

BY ERIK MEYERSSON[1]

Does Islamic political control affect women's empowerment? Several countries have recently experienced Islamic parties coming to power through democratic elections. Due to strong support among religious conservatives, constituencies with Islamic rule often tend to exhibit poor women's rights. Whether this reflects a causal relationship or a spurious one has so far gone unexplored. I provide the first piece of evidence using a new and unique data set of Turkish municipalities. In 1994, an Islamic party won multiple municipal mayor seats across the country. Using a regression discontinuity (RD) design, I compare municipalities where this Islamic party barely won or lost elections. Despite negative raw correlations, the RD results reveal that, over a period of six years, Islamic rule increased female secular high school education. Corresponding effects for men are systematically smaller and less precise. In the longer run, the effect on female education remained persistent up to 17 years after, and also reduced adolescent marriages. An analysis of long-run political effects of Islamic rule shows increased female political participation and an overall decrease in Islamic political preferences. The results are consistent with an explanation that emphasizes the Islamic party's effectiveness in overcoming barriers to female entry for the poor and pious.



Should You Use RD for Your Empirical Project?

- RD requires a policy or treatment where eligibility depends on a cutoff
  - More like an RCT than DD or IV: no debate about whether something is an RD
- If you think you know of a potential RD, you need to check for:
  - Manipulation around the cutoff (endogenous selection invalidates an RD design)
  - Sufficient data around the cutoff (sample restricted to narrow bandwidth)
- Key to many quasi-experimental designs is knowing about policies, policy changes, and other sources of variation in institutional arrangements that allows for impact evaluation
  - Look for information on policy changes in existing papers

When to Write a Descriptive Paper

- Economists write descriptive papers when there is something inherently interesting about the data set, or the summary statistics (e.g. gender gaps) one can derive from the data
    - This is subjective (it does help if you are famous)
- This typically involves collecting your own data or identifying a new source of data
    - Running a lab experiment (or similar)
    - Building a data set from archival or secondary sources
    - Gaining access to a new(ish) source of "big data"
    - Scraping (or otherwise collecting) data from the web
- These types of papers are often motivated by a theoretical model ("test of theory") or a hypothesis from another discipline ("bringing new data to an old question")

Descriptive Papers: Examples

MANAGEMENT SCIENCE
Vol. 60, No. 2, February 2014, pp. 434–448
ISSN 0025-1909 (print) | ISSN 1526-5501 (online)

informs
http://dx.doi.org/10.1287/mnsc.2013.1776

Gender Differences in Willingness to Guess

Katherine Baldiga
Ohio State University, Columbus, Ohio 43210, kbaldiga@gmail.com

We present the results of an experiment that explores whether women are less willing than men to guess on multiple-choice tests. Our test consists of practice questions from SAT II history tests; we vary whether a penalty is imposed for a wrong answer and the salience of the evaluative nature of the task. We find that when no penalty is assessed for a wrong answer, all test takers answer every question. But, when there is a penalty for wrong answers, women answer significantly fewer questions than men. We see no differences in knowledge of the material or confidence in the test takers, and differences in risk preferences explain less than half of the observed gap. Making the evaluative aspect of the test more salient does not impact the gender gap. We show that, conditional on their knowledge of the material, test takers who skip questions do significantly worse on our test.

Data, as supplemental material, are available at http://dx.doi.org/10.1287/mnsc.2013.1776.

Keywords: economics; behavior; behavioral decision making; microeconomic behavior; education systems
History: Received December 17, 2012; accepted May 8, 2013, by Uri Gneezy, behavioral economics. Published online in *Articles in Advance* October 7, 2013.

Table 4 Mean Number of Questions Skipped by Treatment and Gender

	Male means	Female means	*p*-value[a] men vs. women
Unframed Low penalty	2.000 (3.259)	3.679 (4.452)	0.008
SAT framed Low penalty	1.063 (1.702)	2.035 (3.336)	0.033
p-value[a] Unframed vs. SAT	0.042	0.008	

[a]From Fisher–Pitman permutation tests for two independent samples, testing the null of equality.

Descriptive Papers: Examples

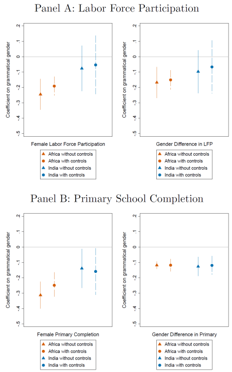
Gendered Language

Pamela Jakiela and Owen Ozier*

January 26, 2022

Abstract

Languages use different systems for classifying nouns. *Gender languages* assign nouns to distinct sex-based categories, masculine and feminine. We construct a new data set, documenting the presence or absence of grammatical gender in more than 4,000 languages which together account for more than 99% of the world's population. We find a robust negative relationship between prevalence of gender languages and women's labor force participation and educational attainment both across and within countries. We also demonstrate that grammatical gender is associated with both weaker legal support for women's equality and reduced female bargaining power within the household.



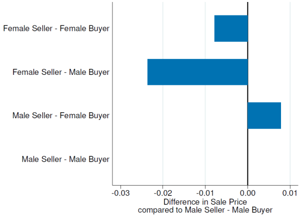
Descriptive Papers: Examples

The Gender Gap in Housing Returns
Paul Goldsmith-Pinkham and Kelly Shue
NBER Working Paper No. 26914
March 2020
JEL No. D14,D31,G4,G51,J16,R2

ABSTRACT

Housing wealth represents the dominant form of savings for American households. Using detailed data on housing transactions across the United States since 1991, we find that single men earn 1.5 percentage points higher unlevered returns per year on housing relative to single women. The gender gap grows significantly larger after accounting for mortgage borrowing: men earn 7.9 percentage points higher levered returns per year relative to women. Approximately 45% of the gap in housing returns can be explained by gender differences in the location and timing of transactions. The remaining gap arises primarily from gender differences in execution prices: data on repeat sales reveal that women buy the same property for approximately 2% more and sell for 2% less. Women experience worse execution prices because of differences in the choice of initial list price and negotiated discount relative to the list price. Gender differences in upgrade and maintenance rates, and preferences for housing characteristics and listing agents appear to be less important factors. Overall, the gender gap in housing returns is economically large and can explain 30% of the gender gap in wealth accumulation at retirement.

Figure 6: Transaction price by seller-buyer gender pairing



Note: This figure plots the average difference in log transaction prices for each possible seller-buyer gender pair, relative to transactions involving single male sellers and single male buyers. These estimates come from a regression of the form in Table 4 column 4, but allowing for the buyer and seller gender group indicators to interact. We plot only the coefficients representing single male or female buyers and sellers, with male seller-male buyer as the omitted base coefficient.

Descriptive Papers: Examples

AEA Papers and Proceedings 2018, 108: 175–179
https://doi.org/10.1257/pandp.20181101

GENDER ISSUES IN ECONOMICS

Gendered Language on the Economics Job Market Rumors Forum†

By Alice H. Wu*

Women are underrepresented in math-intensive fields (Ceci et al. 2014; Kahn and Ginther 2017), and analysts have noted that the representation gap is as large or larger in economics than in STEM (science, technol- issues. Anonymity presumably eliminates social pressures that constrain participants' speech in other public settings, leading to a record of postings that reveal what participants believe but would not otherwise openly express.

Table 1—Top 10 Words Most Predictive of Female/Male

Most *female*		Most *male*	
Word	ME	Word	ME
Hotter	0.422	Homo	−0.303
Pregnant	0.323	Testosterone	−0.195
Plow	0.277	Chapters	−0.189
Marry	0.275	Satisfaction	−0.187
Hot	0.271	Fieckers	−0.181
Marrying	0.260	Macroeconomics	−0.180
Pregnancy	0.254	Cuny	−0.180
Attractive	0.245	Thrust	−0.169
Beautiful	0.240	Nk	−0.165
Breast	0.227	Macro	−0.163

Notes: The model was trained on a 75 percent sample of gendered posts that contain only female or only male classifiers from the comprehensive list. ME—the marginal effect of word *w* is the change in probability that a post is discussing a female, when it contains an additional word *w*. The words that predict *Female* (*Male*) are sorted in descending (ascending) order of the ME.

Should You Write a Descriptive Paper for Your Empirical Project?

- Can you find a new source of data that speaks to a relevant topic?
  - Part of the contribution is (often) in putting together a data set, or finding a data set that has not yet been used in economic analysis (e.g. the *Ethnologue* in the plow paper)
  - Where to look for data:
    - On the web (e.g. Alice Wu on EJMR)
    - In the library (particularly for historical sources)
    - Look at the data sources you use in your own life (hobbies, interests, etc.)
  - Can you collect your own data by running a (simple) online experiment or survey?
- Can you combine multiple existing data sets in a new way?
  - Data Assignment 2 is an example of a replicate and extend research design
  - World Bank and many NGOs regularly release new country-level data sets

Widely Used Data Sources

- World Development Indicators
- American Community Survey (ACS)
- Demographic and Health Surveys (DHS)
- IPUMS
- Panel Study of Income Dynamics (PSID)
- National Longitudinal Survey of Youth (NLSY)
- Afrobarometer
- India Human Development Survey
- World Banks Living Standard Measurement Surveys (LSMS)
- Replication files from recently published papers

The End!