

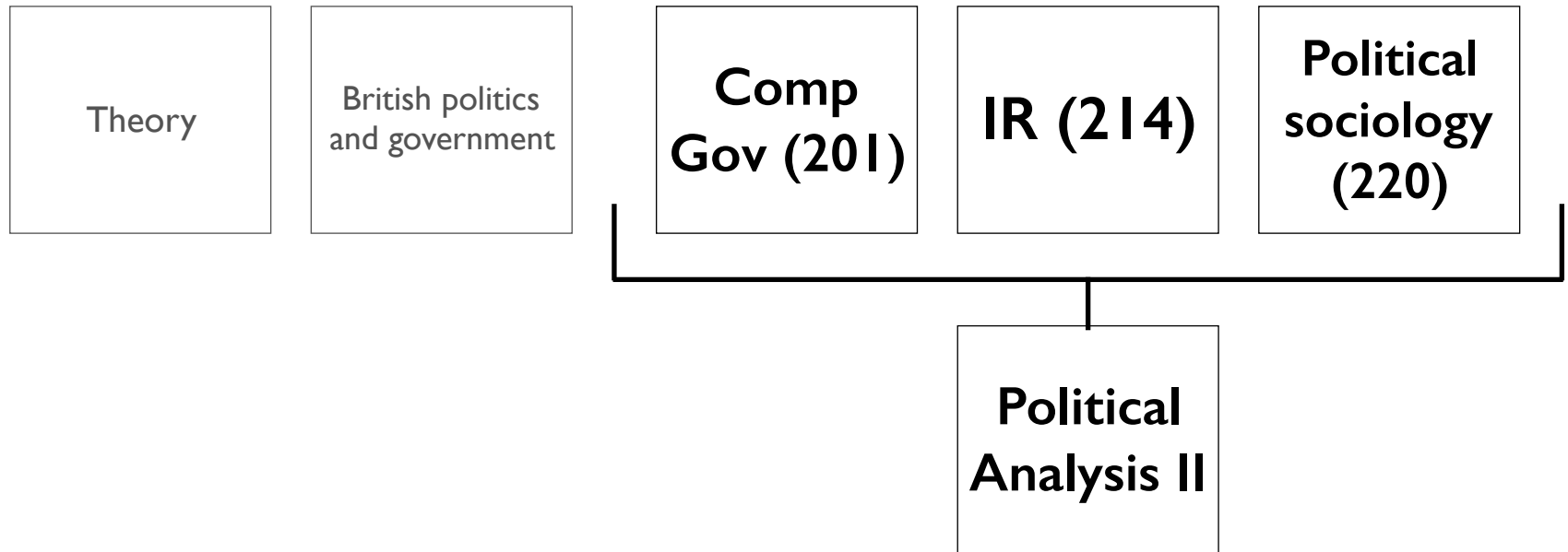
# Review and Preview

Political Analysis 2: Week 1

9 October, 2018

Andy Eggers

## Political Analysis 2 is a required component of 3 core papers in Politics



**Goal: give you the skills to**

1. **critically assess** empirical claims in academic articles (also policy research, the media, etc.)
2. do your **own analysis** for a dissertation (or for research reports, business plan, etc.)

# Our goals for you, in more detail

**At Oxford:** In tutorial essays and exams, you should be able to

1. Explain briefly the basis for empirical claims: *What is the comparison being made? What type of design is it (e.g. cross-country regression, field experiment)?*
2. Highlight potential shortcomings and weigh evidence: *Under what assumptions does a given study tell us what we want to know? Which claims are more credible and why?*

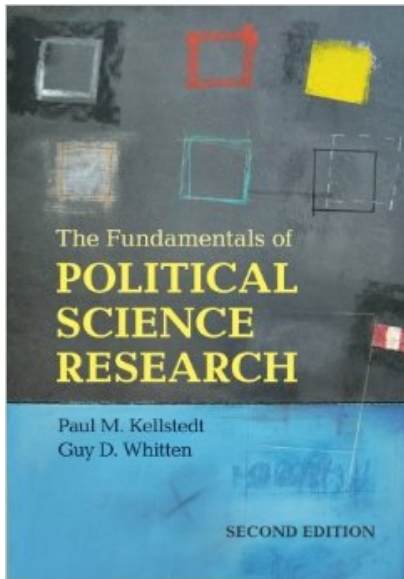
You should also be able to design and execute original data analysis for a **dissertation project**.

Not just  
reporting who  
said what.

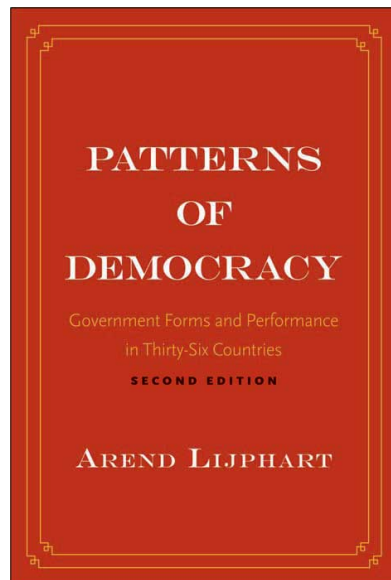
**After Oxford:** You should be able to

1. Examine quantitative evidence and make a decision
2. Produce your own analysis of a new problem (or learn what you need to know to do it)

# Review of Political Analysis I



Main text for content



Main context for application

A table from Lijphart:

270 EFFECTIVE GOVERNMENT AND POLICY-MAKING

TABLE 15.2

Multivariate regression analyses of the effect of consensus democracy (executives-parties dimension) on five indicators of violence, with controls for the effects of the level of economic development, logged population size, and degree of societal division, and with extreme outliers removed

Performance variables	Estimated regression coefficient	Absolute t-value	Countries (N)
Political stability and absence of violence (1996–2009)	0.189***	3.360	34
Internal conflict risk (1990–2004)	0.346**	2.097	32
Weighted domestic conflict index (1981–2009)	-105.0*	1.611	30
Weighted domestic conflict index (1990–2009)	-119.7**	2.177	33
Deaths from domestic terrorism (1985–2010)	-2.357**	1.728	33

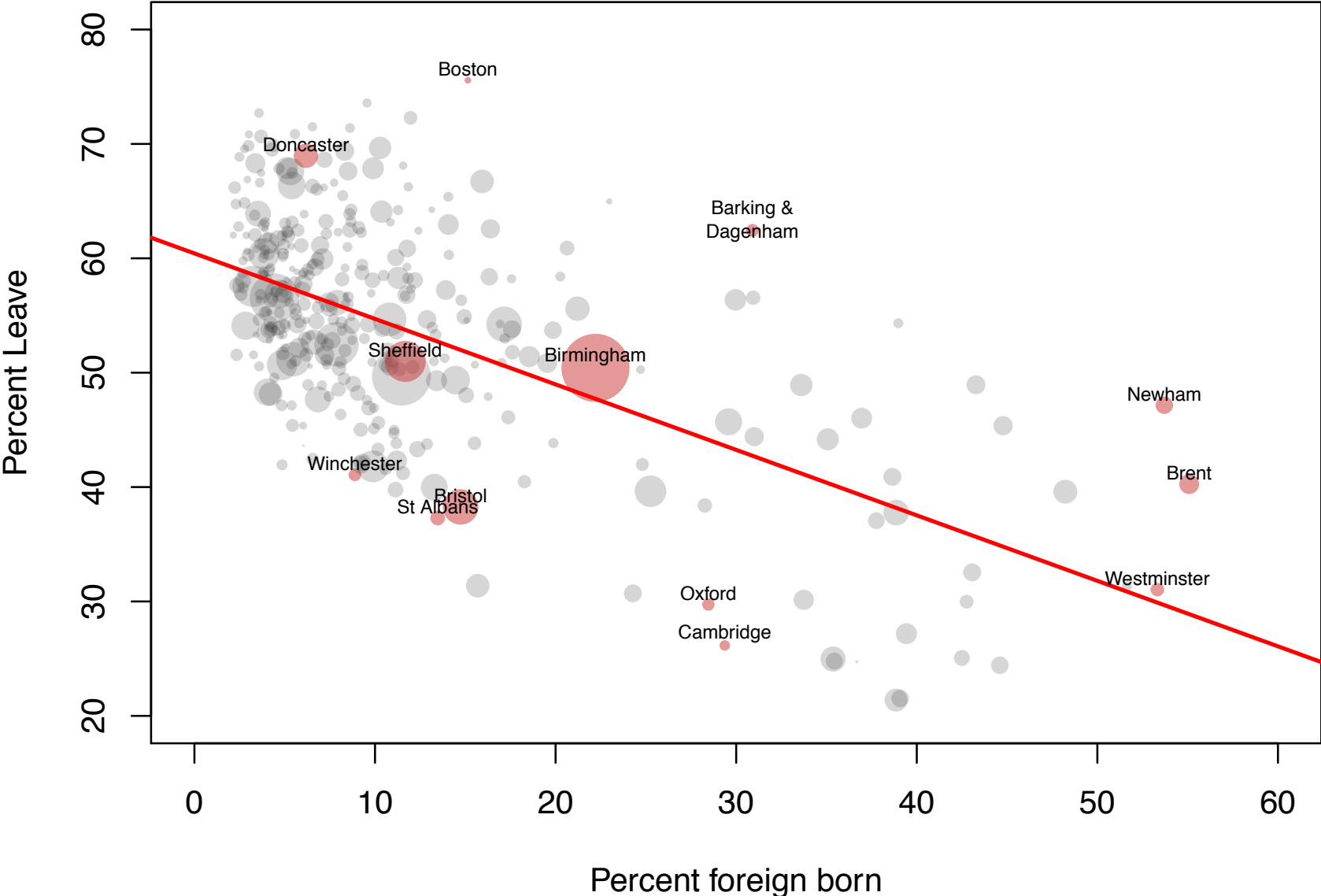
\* Statistically significant at the 10 percent level (one-tailed test)

\*\* Statistically significant at the 5 percent level (one-tailed test)

\*\*\* Statistically significant at the 1 percent level (one-tailed test)

Source: Based on data in Kaufmann, Kraay, and Mastruzzi 2010; PRS Group 2004; Banks, 2010; and GTD Team 2010

# Review of Political Analysis I



# Review of Political Analysis I

*Dependent variable:*

---

Percent supporting Leave

	(1)	(2)	(3)
Percent foreign born	-0.582*** (0.041)	-0.174*** (0.027)	0.013 (0.038)
Percent w. bachelors degree		-0.988*** (0.035)	-1.066*** (0.035)
Mean age			0.755*** (0.115)
Constant	60.937*** (0.618)	83.139*** (0.857)	52.808*** (4.701)
Observations	344	344	344
R <sup>2</sup>	0.375	0.813	0.834

*Note:*

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

# What is regression?



Hannah Fry  @FryRsquared · Oct 4

FFS 

To briefly explain how Linear Regression helped us reverse engineer the BSR equation, let's break it down. Linear Regression is an AI equation that finds the proper coefficients for an equation by sorting through massive amounts of data. The equation looks something like  $BSR = X(a) + Y(b) + Z(c) \dots$  and so and so forth.

 125

 849

 3.2K



## Review: regression coefficients

Performance variables	Estimated regression coefficient	Absolute t-value	Countries (N)
Political stability and absence of violence (1996–2009)	0.189***	3.360	34

We assume a linear relationship between the **dependent variable** (outcome) and one or more **independent variables**, and use **ordinary least squares regression (OLS)** to find what the best-fitting **coefficients (slopes)** are.

$$\text{PolStab}_i = \beta_0 + \beta_1 \text{ExecParties}_i$$

*Bivariate regression: week 5*

$$\text{PolStab}_i = \beta_0 + \beta_1 \text{ExecParties}_i + \beta_2 \text{EconDev}_i + \beta_3 \text{LogPop}_i + \beta_4 \text{SocietalDiv}_i$$

*Multivariate regression: week 6*



# Review: bivariate regression

Bivariate regression equation  
to be estimated:

$$\text{PolStab}_i = \beta_0 + \beta_1 \text{ExecParties}_i$$

Command in R:

```
lm(polstab ~ execparties, data = lijphart)
```

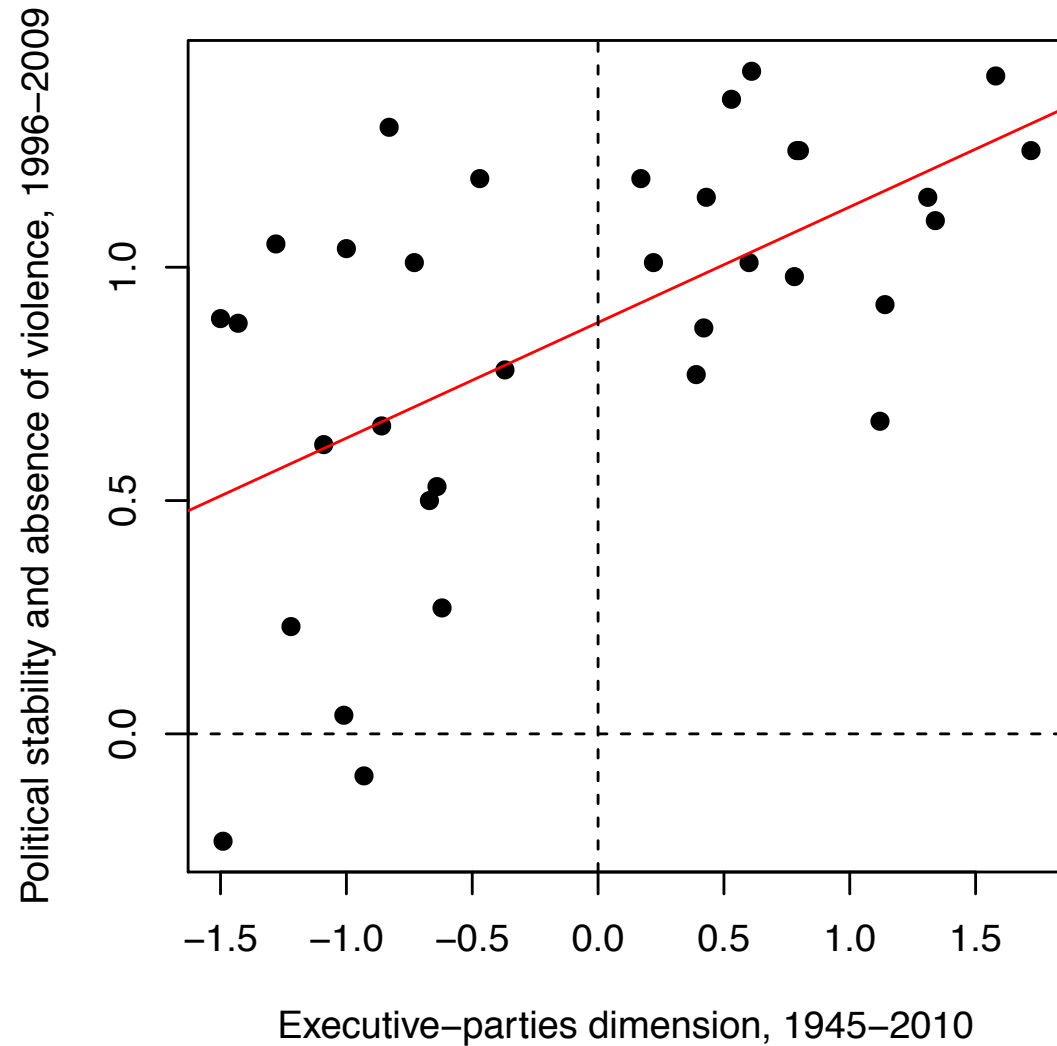
Resulting equation:

$$\text{PolStab}_i = 0.88 + 0.25 \text{ExecParties}_i$$

In figure at right:

- black dots are countries
- red line is the regression line

Which coefficient is the  
intercept, which is slope?



# Review: multivariate regression

Multivariate regression equation to be estimated:

$$\text{PolStab}_i = \beta_0 + \beta_1 \text{ExecParties}_i + \beta_2 \text{EconDev}_i + \beta_3 \text{LogPop}_i + \beta_4 \text{SocietalDiv}_i$$

Command in R:

```
lm(polstab ~ execparties + econdev + logpop + societaldiv, data = lijphart)
```

Resulting equation:

$$\text{PolStab}_i = -0.28 + 0.20\text{ExecParties}_i + 2.80\text{EconDev}_i - 0.13\text{LogPop}_i - 0.08\text{SocietalDiv}_i$$

These are “partial regression coefficients”. They indicate how much the predicted outcome changes for a one-unit change in the independent variable, holding other variables constant.

# Review: inference

The **regression coefficient** tells us our best estimate of the linear relationship between two variables (possibly controlling for other variables).

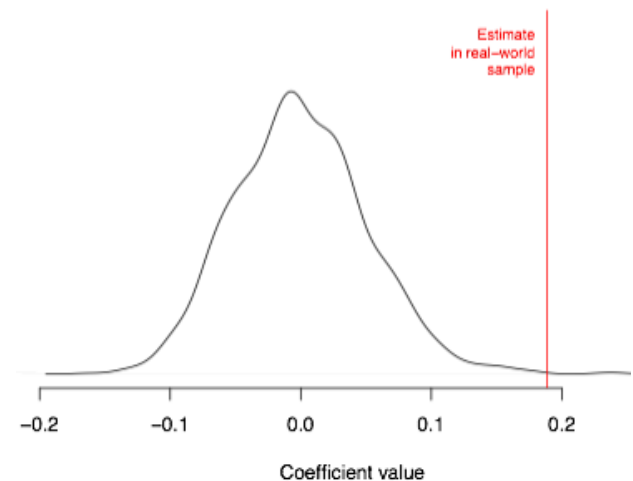
We think of our data as a **sample** (from the **population**, or “data generating process”).

The **p-value** tells us how likely it is that we would get a coefficient at least this large in a **sample** if the coefficient in the **population/DGP** was 0 (no relationship).

In regression tables, stars mean low p-values, which mean *statistically* strong relationships.

From week 7 of last year’s lectures

Assumed sampling distribution of coefficient on consensus democracy under null hypothesis



Performance variables	Estimated regression coefficient	Absolute t-value	Countries (N)
Political stability and absence of violence (1996–2009)	0.185 ***	3.360	34

# Review: summarizing information

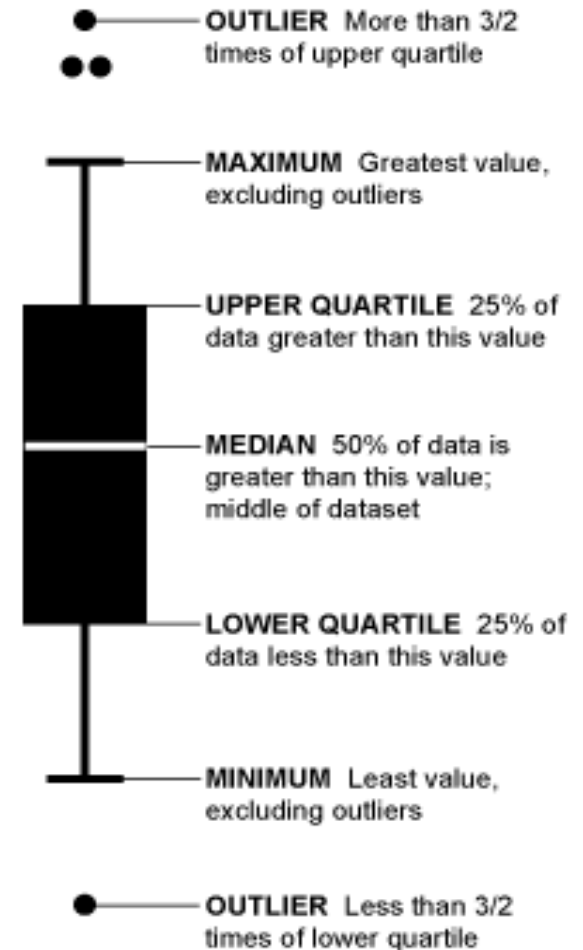
Regressions, correlations, etc, focus on relationships among variables.

Important to also understand different types of variables:

- **interval variables** (aka continuous): values are placed on real numerical scale. *Examples:* GDP, Polity score, number of casualties
- **ordinal variables** (aka categorical): values can be ordered but are not put on a scale. *Example:* high, medium, and low levels of religiosity
- **nominal variables:** values are not ordered. *Example:* religion is Muslim, Hindu, Buddhist, etc

And ways we summarize information about them:

Frequency distributions, histograms, measures of central tendency (mean, median, mode), percentiles, variance/standard deviation, box plot (see figure at right)



# Review: measurement

Before we analyze data, we must collect it.

Before we collect data, we must

- define a concept, i.e. the thing we want to measure (**conceptualization**)
- determine a procedure for measuring the concept (**operationalization**)

Behind every measure is

- lots of hard work, difficult choices
- (usually) a theory or research question.

How do we measure democracy? Bureaucracy?

Two kinds of critiques of measures:

- A measure is not **valid** if it is fundamentally aiming at the wrong thing: as designed, it does not correspond to the concept.
- A measure is not **reliable** if it is error-prone: it may be right for some units, but not for others.

What is a critique of Lijphart's measure of consensus democracy from the standpoint of

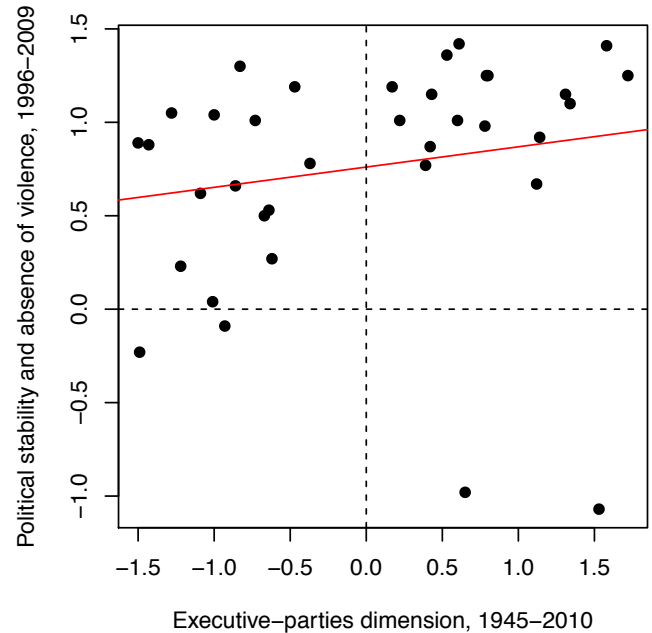
- validity?
- reliability?



## All empirical claims

- have a **scope** of application: set of units (e.g. countries) within which claim is supposed to hold (*population*)
- are based on a **sample**: set of units (e.g. countries) used to establish/test the claim.

For Lijphart, the scope is *advanced democracies*; the sample is “36 countries continuously democratic since 1989 or earlier” (excluding outliers in particular cases).



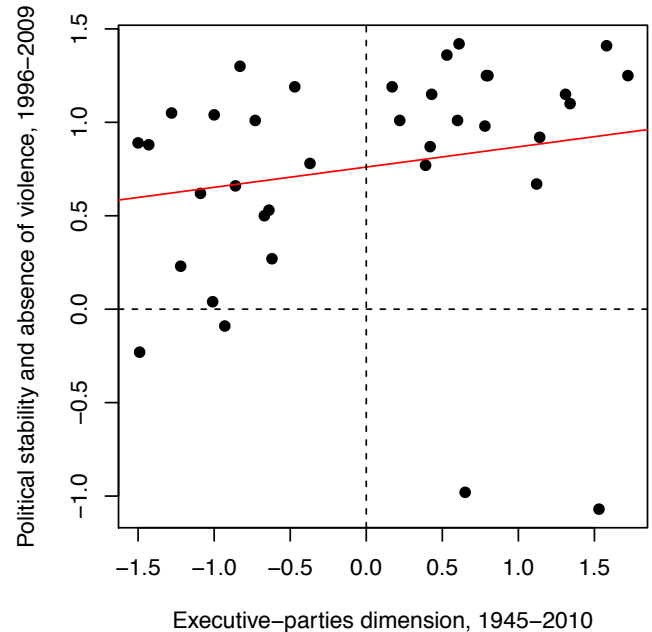
Lijphart excludes India and Israel  
(regression line shown includes them)

## Review: case selection (2)

Useful critiques of case selection:

- Sample is chosen based on the outcome (*selecting on the dependent variable*)
- Sample is chosen to support a point of view, not test a conjecture (*cherry picking*)
- Claim may hold true in the sample examined, but not more broadly (*over-generalization*)

All variations on one point: “The sample analyzed is not representative of the population of interest.”



Lijphart excludes India and Israel  
(regression line shown includes them)

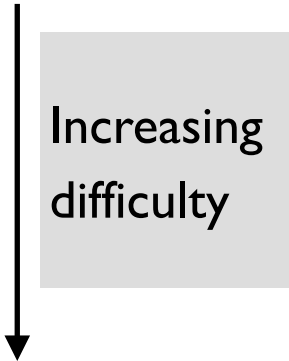
# Review: research design

Some research ask *descriptive* questions: “what is going on?”

Some asks *forward causal* questions: “what is the effect of X?”

Some asks *explanatory* questions: “why Y?”

Often a combination.



Increasing  
difficulty

In your reading (and writing), take note of:

- the research question
- the *type* of question
- the *research design*, i.e. the comparisons and/or measurements being made to answer the question. e.g., *cross-country regression, case study survey experiment*

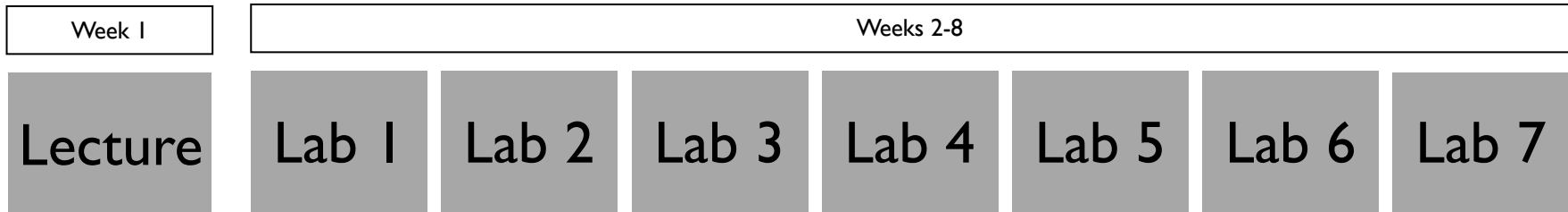
Some key distinctions among designs:

- Is it an *experiment* (i.e. researcher manipulates something) or an *observational study*?
- What are the units? (Countries, individuals, cities, country-years)



# Preview: Political Analysis 2

## Overall Structure



### Themes we will emphasize:

- How to interpret statistical results beyond a simple multivariate regression
  - Interactions
  - Non-linearities
  - Logistic regression
- How to assess what control variables should be included
- Identifying and comparing research designs
- Better understanding of inference (p-values, standard errors)

# Political Analysis 2: assessment

Two options:

- (1) Take-home exam:** You will be (a) given some data to analyze & asked to do some analysis; (b) shown some results from someone else's analysis; (c) asked to interpret the results. Exam distributed Week 8.
- (2) Essay based on original data analysis:** Find data related to something you find interesting; write an essay of 2000 words. The first step in a dissertation project? (See course outline for more details, including about **consultation email**.)

In both cases, assignment due *12pm Friday, week 2 of HT.*



## Review/preview: data analysis in RStudio

Before first lab:

1. Refresh your memory about data analysis in R.

Suggestions:

A. Review the lab worksheets from HT last year, especially Lab 4

B. Do a `swirl` tutorial (1. R Programming; 2. Regression models; 3. Statistical inference)

C. On Datacamp: try “data analysis and statistical inference” course

2. Skim the article we will be analyzing: Brambor, Clark and Golder (2007) “Are African party systems different?”  
*Electoral Studies*