

# **Theory and Data**

**POLI SCI 210**

Introduction to Empirical Methods in Political Science

# Teaching Assistants

 Chloe Bernadaux

 [chloebernadaux2028@u.northwestern.edu](mailto:chloebernadaux2028@u.northwestern.edu)

 Scott Hall 215

 Lucas Camara

 [lucascamara2028@u.northwestern.edu](mailto:lucascamara2028@u.northwestern.edu)

 Scott Hall 219

Check the syllabus for discussion sections and TA office hours

# Sections

	<b>Section 60</b>	<b>Section 61</b>	<b>Section 63</b>	<b>Section 65</b>
<b>Time</b>	Th 1:00– 1:50pm	Th 1:00– 1:50pm	Fr 2:00– 2:50pm	Fr 3:00– 3:50pm
<b>Room</b>	University Hall 218	<b>University</b> <b>Hall 318</b>	University Hall 418	University Hall 418
<b>TA</b>	Bernadaux	Camara	Camara	Bernadaux

# Questions?

# AI Prompts

- What is a theory?
- What makes for a good theory (in social science, political science)?
- What are some common theories in political science?
- What is data? What is a variable?
- Different kinds of data/types of variables
- How do political scientists collect data?

**Reminder:** These are just suggestions, feel free to be creative with your prompts!

# General roadmap

- **Weeks 1-2:** Build common language to talk about different methods
- **Week 3 onward:** Talk about different methods and when/why/how they are useful or important

# **Last week: The scientific method**

**Step 0:** Puzzle/Question

**Step 1:** Theory/Model

**Step 2:** Hypothesis/Implications

**Step 3:** Observation/Testing

**Step 4:** Conclusion/Evaluation

# **This week**

**Step 0: Puzzle/Question**

**Step 1: Theory/Model**

**Step 2: Hypothesis/Implications**

**Step 3: Observation/Testing**

**Step 4: Conclusion/Evaluation**



# This week

Step 0: Puzzle/Question

**Step 1: Theory/Model TUESDAY**

Step 2: Hypothesis/Implications

Step 3: Observation/Testing

Step 4: Conclusion/Evaluation

**THURSDAY:** Start moving toward the next steps with a conversation about *data*

# Part I: Theory

**What is a theory?**

**What is a theory?**

-\\(ツ)/-

# What is a theory?

-(ツ)-

I have no idea

# What is a theory?

-\\(ツ)/-

I have no idea

My *theory* is that no one really knows

# What is a theory?

-(ツ)-

I have no idea

My *theory* is that no one really knows

so they make it complicated

# But it sure is important!

Most common reasons why a peer-reviewed political science journal rejects an article:

1. Having no theory
2. Weak theory
3. Insufficient theoretical innovation



# theory

[Add to list](#)[Share](#)

/ˈθɪəri/



/ˈθɪəri/

[IPA guide](#)

Other forms: **theories**

When you have a *theory*, you have a set of beliefs or principles that might not be proven yet. Does anyone have a good *theory* for where missing socks go when you do laundry?

A *theory* is a set of accepted beliefs or organized principles that explain and guide analysis and one of the ways that *theory* is defined is that it is different from practice, when certain principles are tested. For example, you could be a musician who plays well but who doesn't have a lot of experience with the theory of music. This word is a noun and comes from the Greek *theoria*, which means "contemplation or speculation."

[vocabulary.com/dictionary/theory](https://www.vocabulary.com/dictionary/theory)

# Dictionary definitions

1. A **belief** that can **guide behavior**
2. A **well-substantiated explanation** of some aspect of the natural world; an **organized system of accepted knowledge** that **applies in a variety of circumstances** to explain a **specific set of phenomena**
3. A **tentative insight** into the natural world; a concept that is **not yet verified** but that if true **would explain certain facts or phenomena**

# Textbook definition

*A scientific theory*

# Textbook definition

*A scientific theory* is a set of **logically consistent statements**

# Textbook definition

*A scientific theory* is a set of **logically consistent statements** that tell us **why** the empirical social and political phenomena we observe, or the relationships between them, **occur in the way they occur**

A theory is a **set of interrelated propositions** about empirical reality

Components:

1. **Concepts** that define basic terms
2. **Assumptions** that relate concepts to each other
3. **Generalizations** that connect statements to observations

# Toy example: Democratic peace

**Proposition:** Democracies are hesitant to fight each other

**Concepts:**

- *Democracy*: Leaders elected in free-fair elections
- *Fighting*: Militarized Interstate Disputes (MID)

**Assumptions:**

- Liberal democratic culture of negotiation and compromise
- Democratic publics dislike wars

**Generalization:** No war between democracies

**Good?**

**Good?**

Doesn't seem that complicated



# Good?

Doesn't seem that complicated

But many things that look like theory are not

ASQ Forum

What Theory is *Not*

**Robert I. Sutton**

*Stanford University*

**Barry M. Staw**

*University of California at  
Berkeley*

This essay describes differences between papers that contain some theory rather than no theory. There is little agreement about what constitutes strong versus weak theory in the social sciences, but there is more consensus that references, data, variables, diagrams, and hypotheses are not theory. Despite this consensus, however, authors routinely use these five elements in lieu of theory. We explain how each of these five elements can be confused with theory and how to avoid such confusion. By making this consensus explicit, we hope to help authors avoid some of the most common and easily averted problems that lead readers to view papers as having inadequate theory. We then discuss how journals might facilitate the publication of stronger theory. We suggest that if the field is serious about producing stronger theory, journals need to reconsider their empirical requirements. We argue that journals ought to be more receptive to papers that test part rather than all of a theory and use illustrative rather than definitive data.

Sutton, Robert I. and Barry M. Staw. 1995. “What Theory is Not.” *Administrative Science Quarterly* 40 (3): 371-384

# References are not theory

Literature A suggests X.

# References are not theory

Literature A suggests X. Literature B says Y.

# References are not theory

Literature A suggests X. Literature B says Y. Borrowing from both, we should expect XYZ.

- Referencing theories in previous work helps to contextualize new explanations
- But often used as a smokescreen for no theory

This pattern is consistent with findings that aggression provokes the ‘fight’ response (Frijda, 1986) and that anger is a contagious emotion (Schacter and Singer, 1962; Baron, 1977).

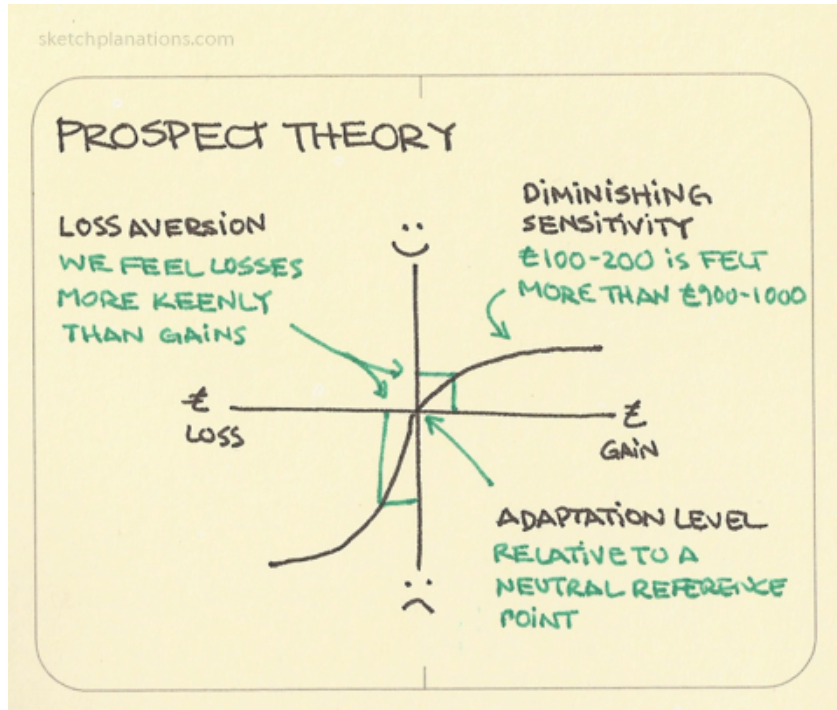
# Data is not theory

Recent empirical evidence suggests that the collective bargaining process, the union-management contract, and union-management relations in general all have important consequences for the quality of worklife of unionized workers. Moreover, previous work has investigated the relationship between union strength and construction workers' reactions to their work. She found that union strength (operationalized in terms of workers' relative wages) was positively related to both pay satisfaction and perceived job security. Finally, the union's ability to formally increase members' participation in job-related decisions has been frequently cited as contributing to the unionization of teachers and other professionals.

- Helps us set expectations, but doesn't really explain why things happen

# Diagrams are not theory

- Shows connections
- Does not explain them



# Also not theories

- **Lists of variables or concepts:** Do not explain why
- **Models or paradigms:** Simplify, but do not explain (e.g. rational choice)
- **Hypotheses:** Translate theory to expectations



# Honestly

Theory is either:

1. explanation
2. speculation

We need to sound more sophisticated!

# Honestly

Theory is either:

1. explanation
2. speculation

# Honestly

Theory is either:

1. **Logical** explanation
2. speculation

# Honestly

Theory is either:

1. **Logical** explanation
2. **Informed** speculation

**Real question:** What makes a good theory?

# Honestly

Theory is either:

1. **Logical** explanation
2. **Informed** speculation

**Real question:** What makes a good theory?

Content warning: Video with curse words in the next slide

The REAL F Around and Find Out Graph



<https://youtu.be/mH9NzJwldok?si=XviVZMs3ZL5Qh4tH>

# Good theories

- Give a **new** explanation
- Give a **different** explanation
- Not too different to sound like **non-sense**
- But different enough to try to change your mind

**Trap:** Pressure to say something new leads to overly complicated theories

---

*Symposium: “What is Good Theorizing?”*

## **Fuck Nuance**

**Kieran Healy**

### **Abstract**

Nuance is not a virtue of good sociological theory. Although often demanded and superficially attractive, nuance inhibits the abstraction on which good theory depends. I describe three “nuance traps” common in sociology and show why they should be avoided on grounds of principle, aesthetics, and strategy. The argument is made without prejudice to the substantive heterogeneity of the discipline.

### **Keywords**

theory, nuance, models, fuck

---

Sociology Department, Duke University, Durham, NC, USA

Healy, Kieran. 2017. “**Fuck Nuance.**” *Sociological Theory* 35 (2): 118-127



# Nuance traps

## 1. Nuance of the *fine-grain*

Theory as *extremely detailed* description of the world

## 2. Nuance of the *conceptual framework*

Expansion of an *irrefutable theoretical system*

## 3. Nuance of the *connoisseur*

Nuanced theory as a sign of *sophistication*

# Rule of thumb

Good theories explain more than what they assume

See [Clark \(2020\)](#) for a more eloquent presentation

# Takeaways

1. Theory is logical explanation or informed speculation
2. Theories explain why we observe what we observe
3. It feels complicated because
  - Good theorizing is hard
  - Pressure to always say something new

# **Theory and Data**

**POLI SCI 210**

Introduction to Empirical Methods in Political Science

# Part II: Data

# data

 Add to list

 Share

  /ˈdeɪrə/   /ˈdætə/ [IPA guide](#)

*Data* is information such as facts and numbers used to analyze something or make decisions. Computer *data* is information in a form that can be processed by a computer.

*Data* is the plural of Latin *datum*, "something given." In scientific use, the word *data* is often considered to be a plural noun meaning "pieces of information": The data are reliable. But most people think of *data* as a mass noun meaning "information" and use the word with a singular verb and pronoun: The data is reliable because it was tested by experts.

## Definitions of *data*

1. **noun** a collection of facts from which conclusions may be drawn

*"statistical data"*

synonyms: [information](#)

[see more](#) ▾

# In empirical research

The term *data* is used in two ways:

## 1. Data as *information*

“The data suggests that inflation keeps growing”

## 2. Data as the *raw material* from which we draw inferences

“The data contains 1,000 responses to an online survey”

# In empirical research

The term *data* is used in two ways:

## 1. Data as *information*

“The data suggests that inflation keeps growing”

## 2. Data as the *raw material* from which we draw inferences

“The data contains 1,000 responses to an online survey”

The second is more correct since “data doesn’t speak.”



# Data usually comes as a rectangle

country	year	cases	population
Afghanistan	1999	745	19987071
Afghanistan	2000	2666	20595360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	213766	1280428583

We call this a **dataset**, **dataframe**, or **database**

# Elements of a dataset

A single *observation* is a scalar

$$a = 12$$

Several scalars together make a vector

$$\vec{b} = [12 \quad 14 \quad 15]$$

# Elements of a dataset

Vectors have *direction*

This is a **row** vector

$$\vec{b} = [12 \quad 14 \quad 15]$$

This is a **column** vector

$$\vec{c} = \begin{bmatrix} 12 \\ 14 \\ 15 \end{bmatrix}$$

# Elements of a dataset

Several vectors of equal length make a **matrix**

$$A = \begin{bmatrix} 12 & 14 & 15 \\ 115 & 22 & 127 \\ 193 & 29 & 219 \end{bmatrix}$$

The number of *rows* and *columns* are the **dimensions**

This is a  $3 \times 3$  matrix ( $r \times c$ )

# A dataset is just a matrix with labels

country	year	cases	population
Afghanistan	1999	745	19987071
Afghanistan	2000	2666	20595360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	213766	1280428583

# A dataset is just a matrix with labels

country	year	cases	population
Afghanistan	1999	745	19987071
Afghanistan	2000	2666	20595360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	213766	1280428583

A row is an **observation** or **unit**

# A dataset is just a matrix with labels

country	year	cases	population
Afghanistan	1999	745	19987071
<b>Afghanistan</b>	<b>2000</b>	<b>2666</b>	<b>20595360</b>
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	213766	1280428583

A row is an **observation** or **unit**

# A dataset is just a matrix with labels

country	year	cases	population
Afghanistan	1999	745	19987071
Afghanistan	2000	2666	20595360
<b>Brazil</b>	<b>1999</b>	<b>37737</b>	<b>172006362</b>
Brazil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	213766	1280428583

A row is an **observation** or **unit**



# A dataset is just a matrix with labels

country	year	cases	population
Afghanistan	1999	745	19987071
Afghanistan	2000	2666	20595360
Brazil	1999	37737	172006362
<b>Brazil</b>	<b>2000</b>	<b>80488</b>	<b>174504898</b>
China	1999	212258	1272915272
China	2000	213766	1280428583

A row is an **observation** or **unit**

# A dataset is just a matrix with labels

country	year	cases	population
Afghanistan	1999	745	19987071
Afghanistan	2000	2666	20595360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
<b>China</b>	<b>1999</b>	<b>212258</b>	<b>1272915272</b>
China	2000	213766	1280428583

A row is an **observation** or **unit**

# A dataset is just a matrix with labels

country	year	cases	population
Afghanistan	1999	745	19987071
Afghanistan	2000	2666	20595360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	212258	1272915272
<b>China</b>	<b>2000</b>	<b>213766</b>	<b>1280428583</b>

A row is an **observation** or **unit**

# A dataset is just a matrix with labels

country	year	cases	population
Afghanistan	1999	745	19987071
Afghanistan	2000	2666	20595360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	213766	1280428583

A column is a **variable**

# A dataset is just a matrix with labels

country	year	cases	population
Afghanistan	1999	745	19987071
Afghanistan	2000	2666	20595360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	213766	1280428583

A column is a **variable**

# A dataset is just a matrix with labels

country	year	cases	population
Afghanistan	1999	<b>745</b>	19987071
Afghanistan	2000	<b>2666</b>	20595360
Brazil	1999	<b>37737</b>	172006362
Brazil	2000	<b>80488</b>	174504898
China	1999	<b>212258</b>	1272915272
China	2000	<b>213766</b>	1280428583

A column is a **variable**

# A dataset is just a matrix with labels

country	year	cases	population
Afghanistan	1999	745	19987071
Afghanistan	2000	2666	20595360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	213766	1280428583

A column is a **variable**

# A dataset is just a matrix with labels

country	year	cases	population
Afghanistan	1999	745	19987071
Afghanistan	2000	2666	20595360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	213766	1280428583

What does the third row, fourth column represent?



# A dataset is just a matrix with labels

country	year	cases	population
Afghanistan	1999	745	19987071
Afghanistan	2000	2666	20595360
Brazil	1999	37737	<b>172006362</b>
Brazil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	213766	1280428583

What does the third row, fourth column represent?

# Types of variables

1. **Nominal/categorical:** Cannot be ordered in any way
2. **Ordinal:** Ordered
3. **Numerical:** Ordered + equidistant

EMPS further distinguishes numerical variables:

1. **Discrete:** Countable integers
2. **Continuous:** Infinitesimally uncountable, real numbers

But technically everything numerical is discrete!

# Practice

**Varieties of Democracy's** *regime type* measure:

0: Closed autocracy – No multiparty elections

1: Electoral autocracy – De-jure elections but not free and fair

2: Electoral democracy – Free and fair elections with some flaws

3: Liberal democracy – Free and fair elections guaranteed

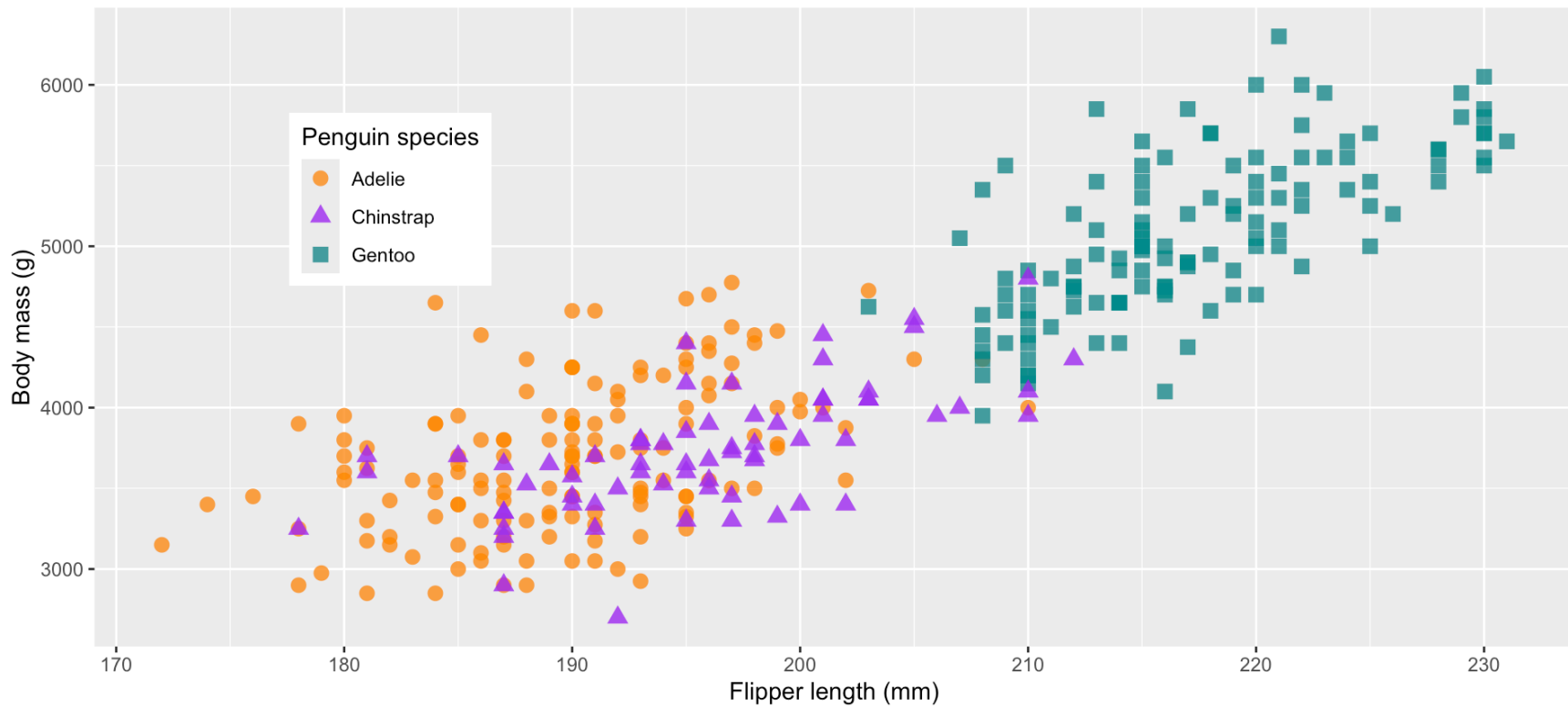
More at [v-dem.net](http://v-dem.net)

<https://ourworldindata.org/regimes-of-the-world-data>

# Practice 2

## Penguin size, Palmer Station LTER

Flipper length and body mass for Adelie, Chinstrap and Gentoo Penguins



How many different variables do you see? What type are they?

# Reminder

Variable types depend on what we are **willing to assume**

# Variables vs. concepts

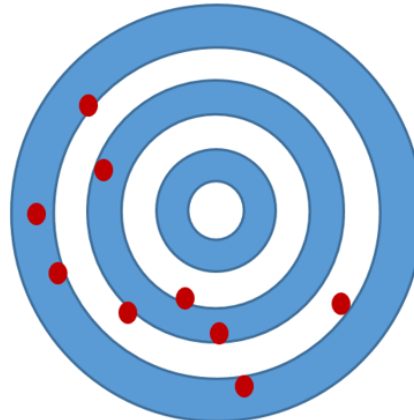
- A **concept** is an element defined by a theory
- A **variable** is the measurement of a concept
- The translation from concept to measurement is called **operationalization**
- Because concepts are abstract, there is room for **measurement error**

# Measurement error

Reliable, not valid



Not reliable, not valid



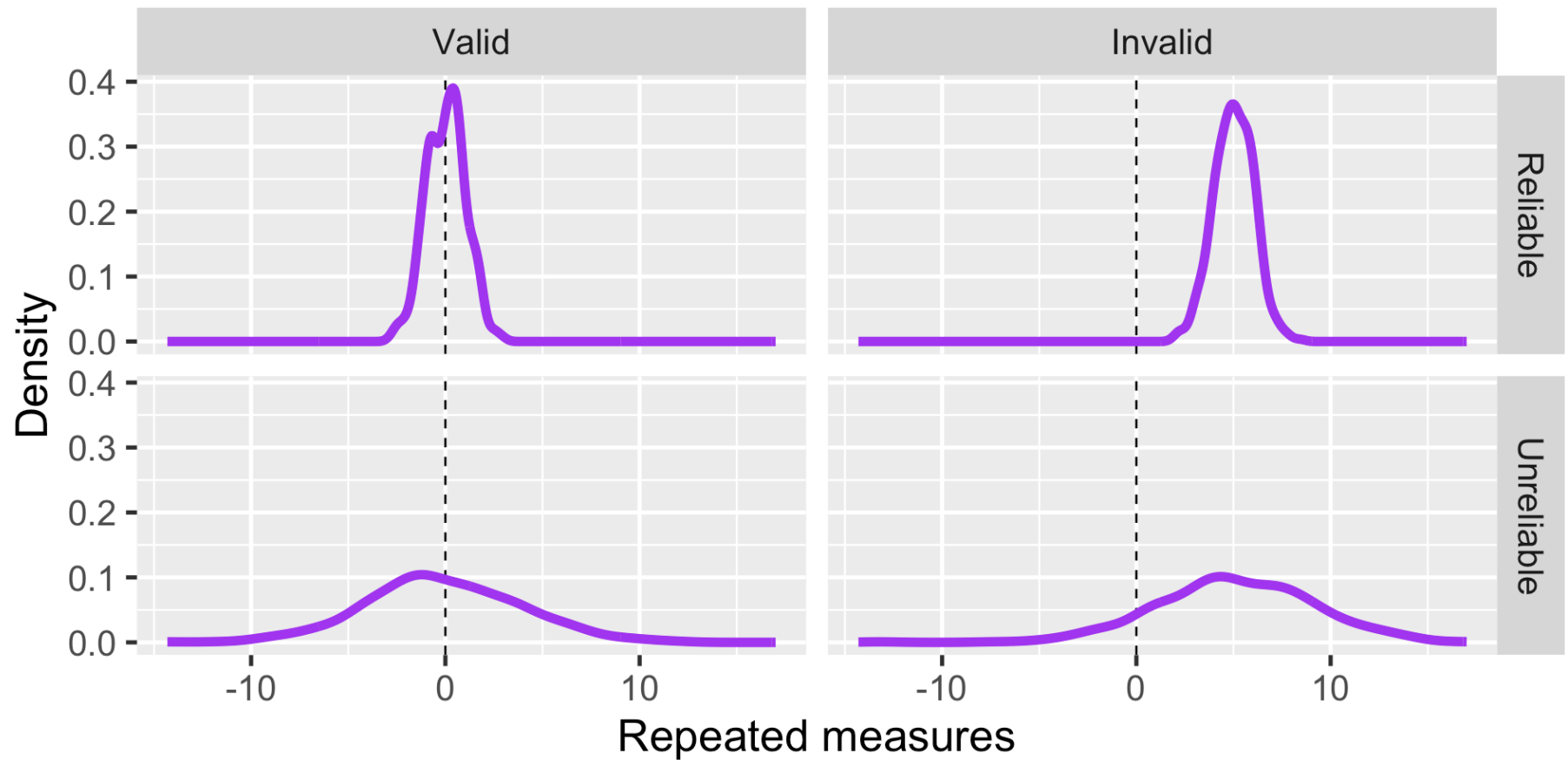
Reliable and valid



- **Reliable:** Hypothetical repeated measurements are close to each other
- **Valid:** Distance from center averages to zero



# Another way to look at it



# Takeaways

- Variables are how we measure concepts
- A variable can be categorical, ordinal, numerical
- Depends mostly on what we are willing to assume
- Gap between concept and variable produces measurement error

# **Next week**

## **Inference**

How do we summarize data?

How do we use summaries to evaluate hypotheses?