Small N

POLI SCI 210

Introduction to Empirical Methods in Political Science

AI Prompts

- Necessary vs. sufficient

- INUS, SUIN causes

- Causal mechanism

- Case selection for [study]

- When to use one method/approach over another

- Mixed methods?

So far

- Course mostly about **statistical** and **causal inference**

- **This week:** Small N or qualitative methods

- **Tuesday:** Reservations with large N + logic of causation

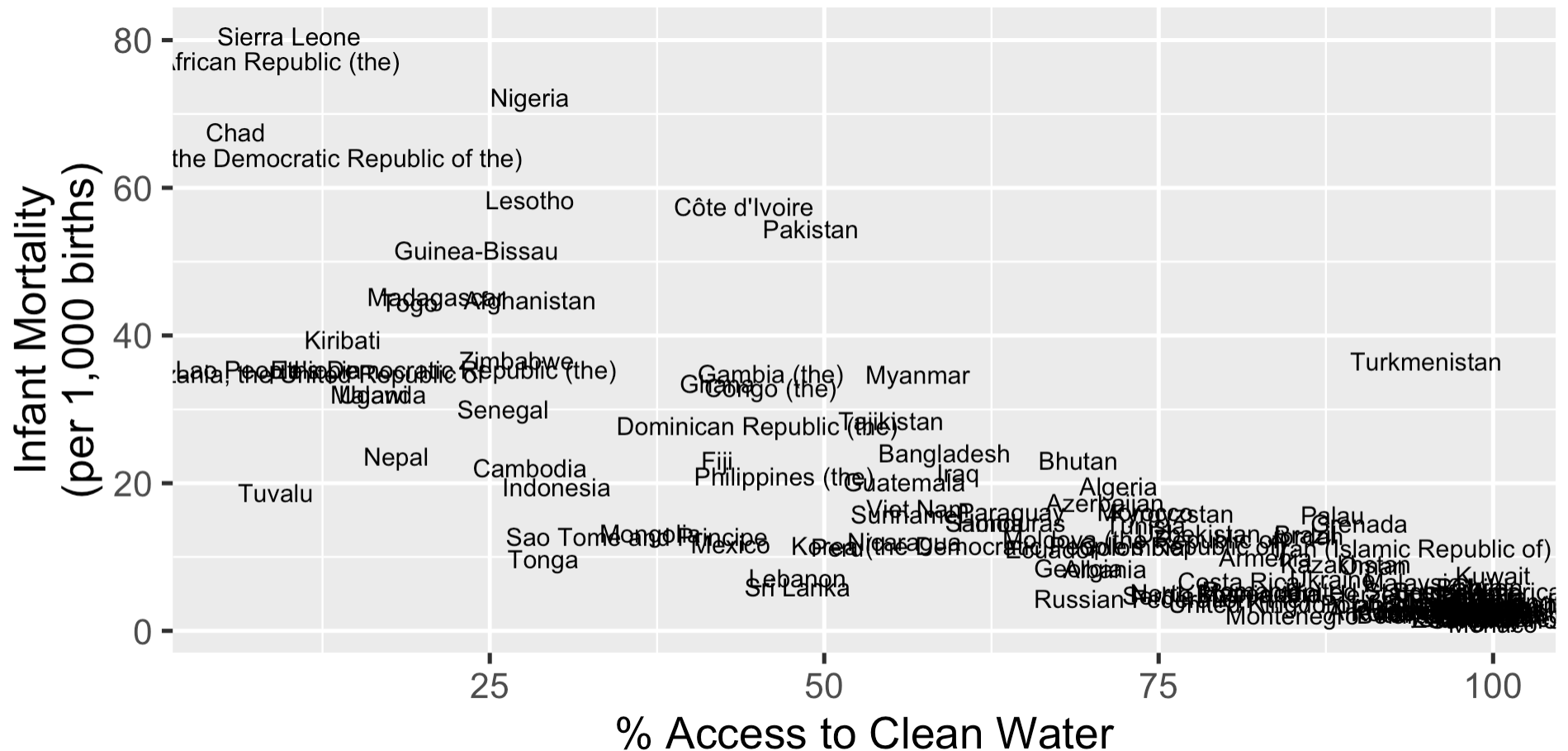- **Thursday:** Case selection + techniques
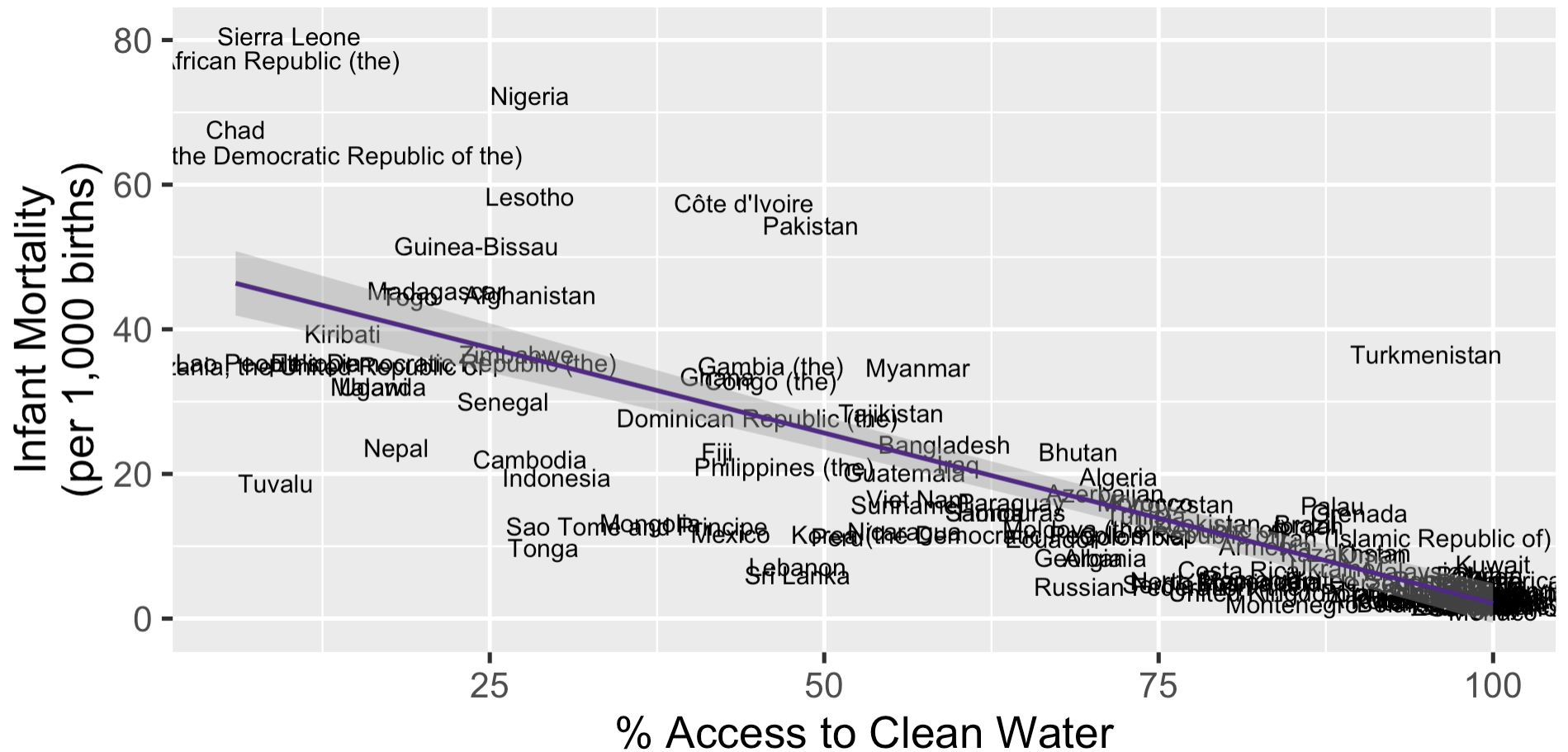
Questions

What have you learned so far?

Any reservations with the methods we have discussed?

Any other methods you would like to learn about?
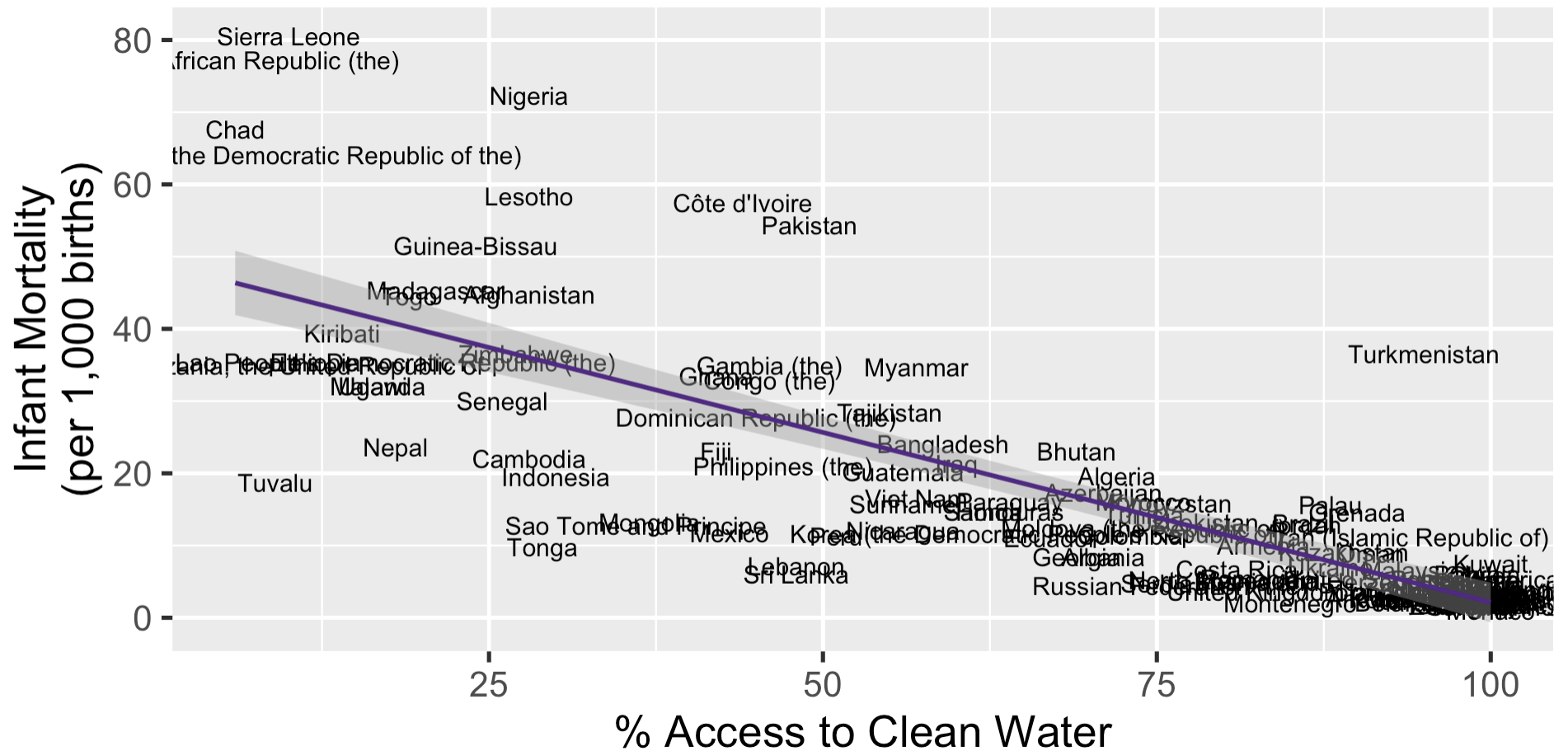
Back to regression: QoG 2024

Scatter plot. X-axis: % Access to Clean Water (25, 50, 75, 100). Y-axis: Infant Mortality (per 1,000 births) (0, 20, 40, 60, 80).

Country labels include: Sierra Leone, African Republic (the), Nigeria, Chad, (the Democratic Republic of the), Lesotho, Côte d'Ivoire, Pakistan, Guinea-Bissau, Madagascar, Togo, Afghanistan, Kiribati, Zimbabwe, Lao People's Democratic Republic (the), Ethiopia, Malawi, Uganda, Senegal, Gambia (the), Congo (the), Myanmar, Ghana, Dominican Republic (the), Tajikistan, Nepal, Cambodia, Fiji, Bangladesh, Bhutan, Indonesia, Philippines (the), Guatemala, Iraq, Algeria, Tuvalu, Viet Nam, Azerbaijan, Kyrgyzstan, Palau, Grenada, Sao Tome and Principe, Mongolia, Honduras, Paraguay, Moldova (the), Uzbekistan, Brazil, Tonga, Mexico, Korea (the Democratic People's Republic of), Nicaragua, Ecuador, Oman, Iran (Islamic Republic of), Lebanon, Georgia, Albania, Armenia, Kazakhstan, Kuwait, Sri Lanka, Russian Federation, Costa Rica, Ukraine, Malaysia, North..., United Kingdom, Montenegro, Turkmenistan, Monaco

Fit regression

Scatter plot with regression line showing Infant Mortality (per 1,000 births) on the y-axis (0 to 80) versus % Access to Clean Water on the x-axis (25 to 100). Countries plotted include Sierra Leone, Central African Republic (the), Nigeria, Chad, Congo (the Democratic Republic of the), Lesotho, Côte d'Ivoire, Pakistan, Guinea-Bissau, Madagascar, Afghanistan, Togo, Kiribati, Zimbabwe, Lao People's Democratic Republic (the), Ethiopia, Tanzania, Gambia (the), Myanmar, Turkmenistan, Malawi, Uganda, Ghana, Congo (the), Senegal, Tajikistan, Dominican Republic (the), Nepal, Cambodia, Fiji, Bangladesh, Bhutan, Indonesia, Philippines (the), Guatemala, Iraq, Algeria, Tuvalu, Viet Nam, Azerbaijan, Suriname, Paraguay, Samoa, Honduras, Kyrgyzstan, Palau, Grenada, Sao Tome and Principe, Mongolia, Mexico, Nicaragua, Korea (the Democratic...), Moldova (the), Ecuador, Colombia, Armenia, Kazakhstan, Brazil, Iran (Islamic Republic of), Oman, Kuwait, Tonga, Lebanon, Georgia, Albania, Costa Rica, Malaysia, Russian Federation, United Kingdom, Montenegro, Monaco.

More questions

- Can we claim *access to clean water* **reduces** *infant mortality*?

- What do we need to believe?

- Random assignment? Conditional independence? Quasi-experiment?

Critique 1: *Correlation does not imply causation*
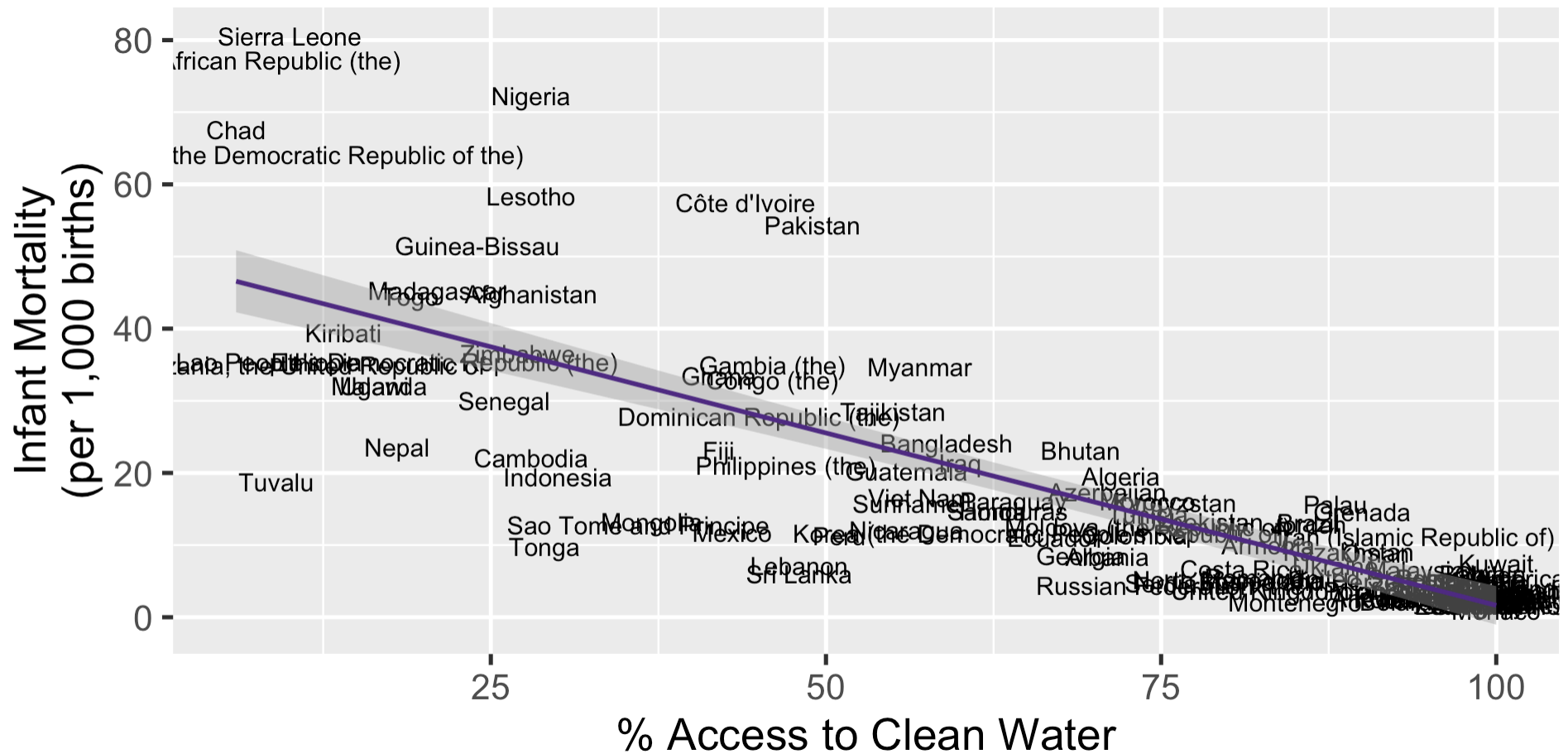
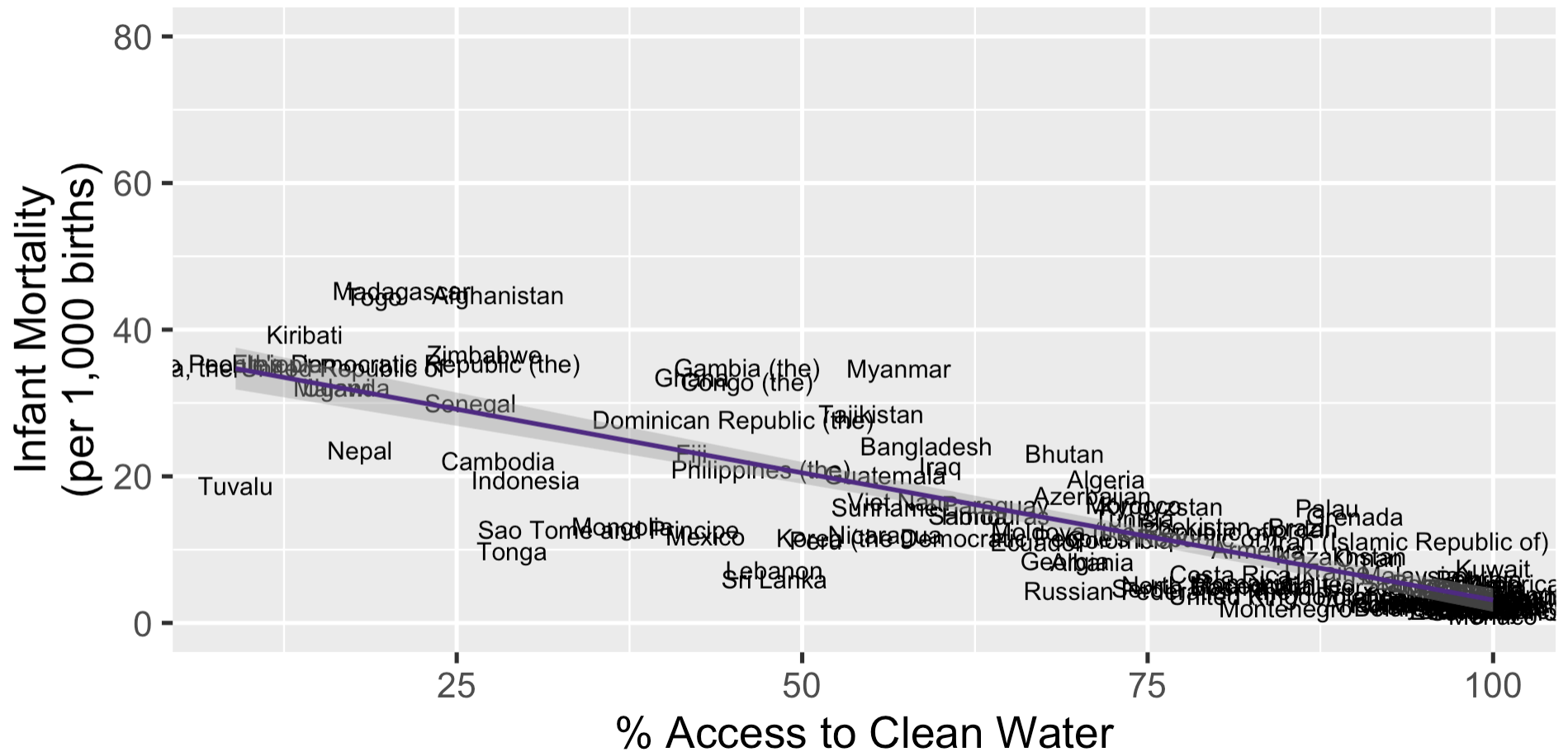We need more knowledge to make causal claims

Remove Turkmenistan

Remove Turkmenistan

Remove high infant mortality
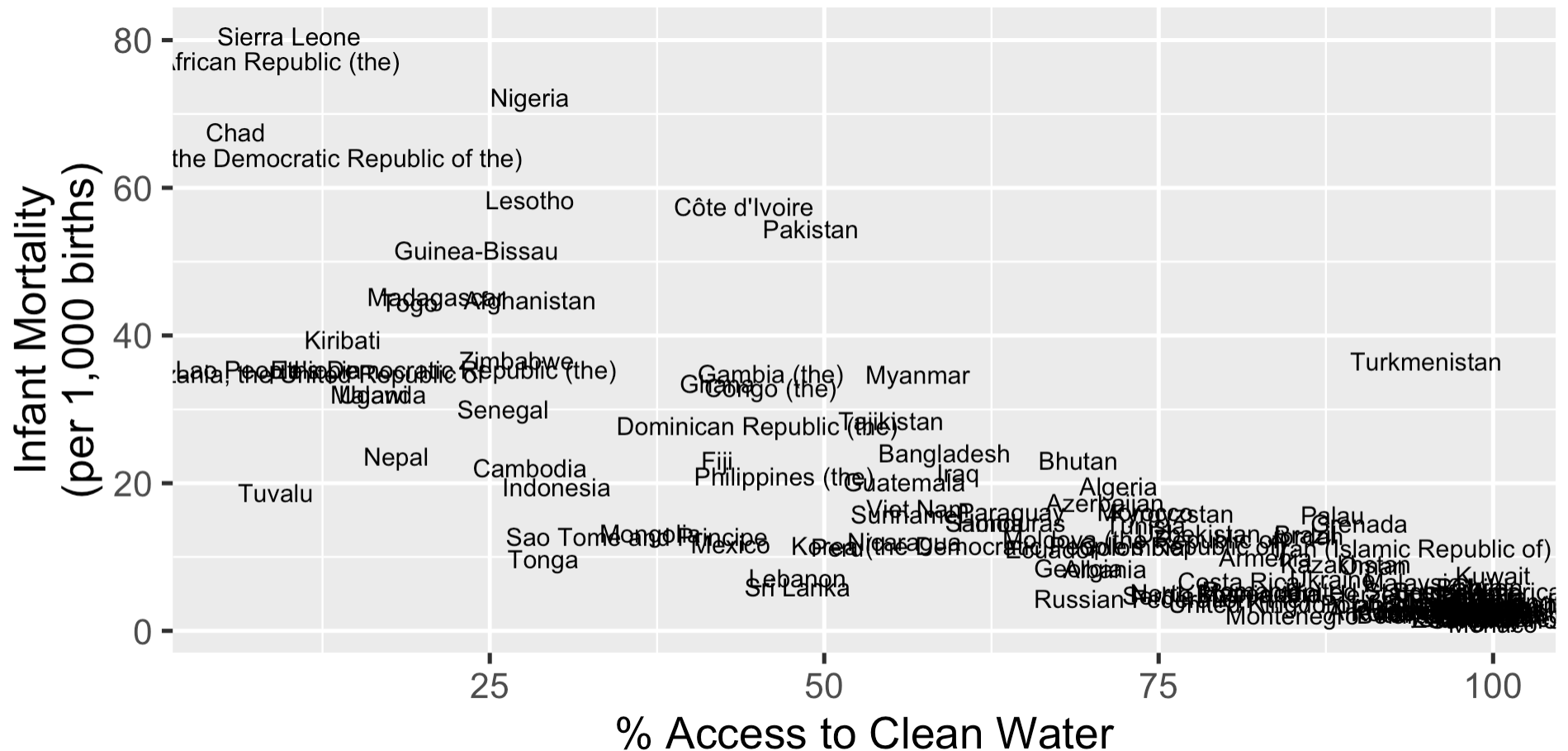
Remove high infant mortality

More more questions

Let's say you believe this relationship is causal

- How is it that one thing causes the other?

- Direction of causality?

- Is this true everywhere?

- Are we comparing apples and apples or apples and oranges and pick-up trucks?
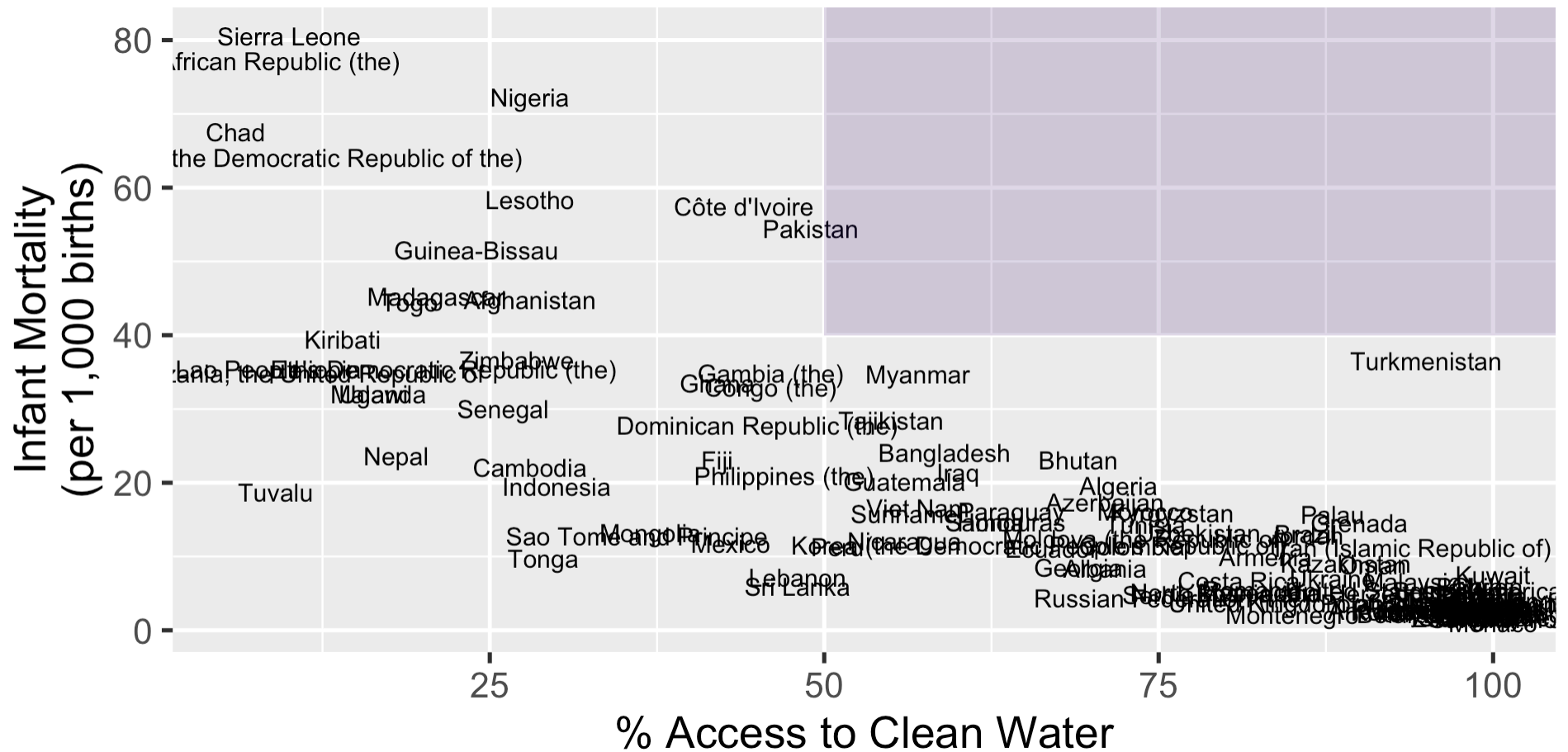
Critique 2: *Large N findings are what happens **on average***
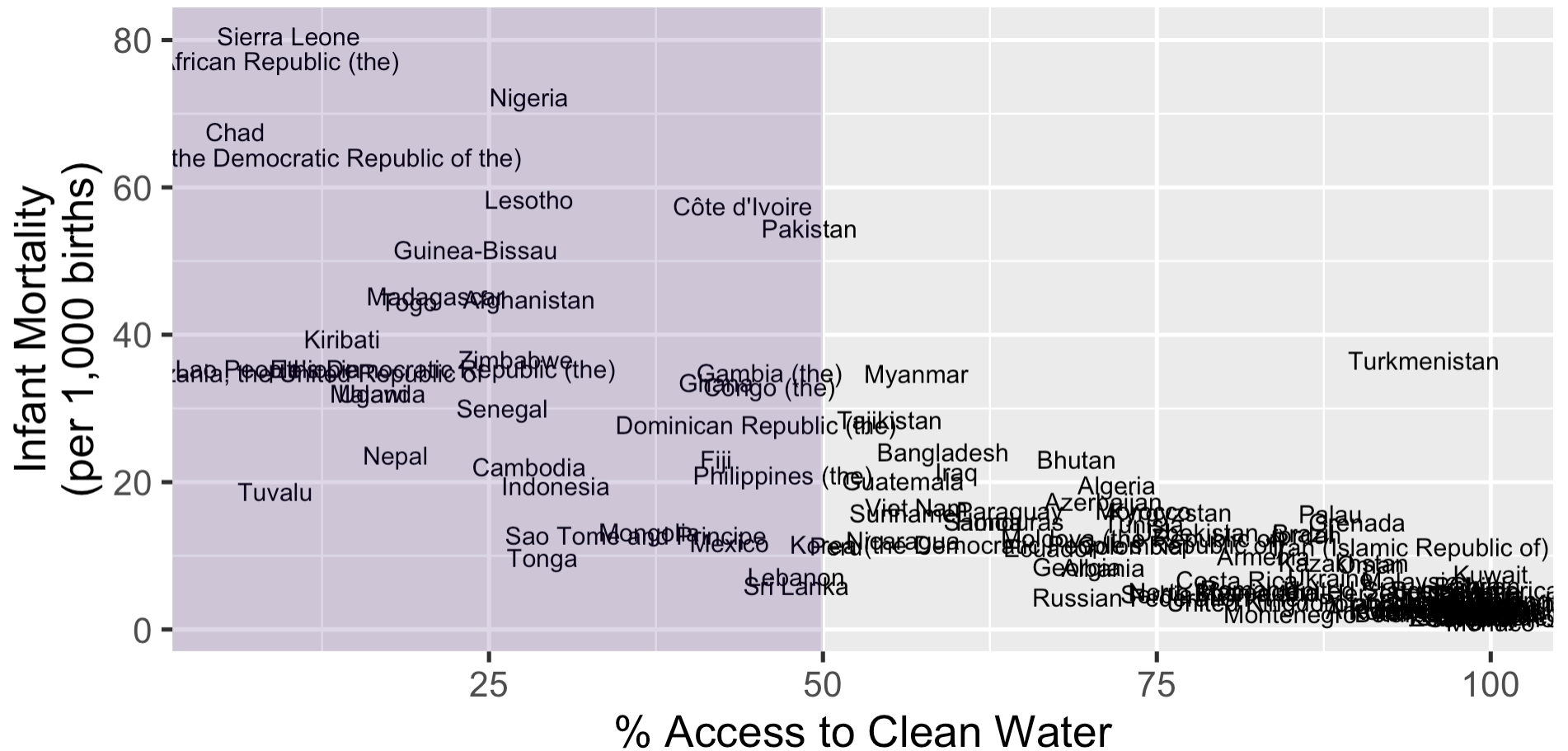
Averages are sensitive to **sample composition**

Another way to think about it

# Another way to think about it

Another way to think about it

Y-axis: Infant Mortality (per 1,000 births), 0, 20, 40, 60, 80

X-axis: % Access to Clean Water, 25, 50, 75, 100

Sierra Leone
African Republic (the)
Nigeria
Chad
the Democratic Republic of the)
Lesotho
Côte d'Ivoire
Pakistan
Guinea-Bissau
Madagascar
Togo
Afghanistan
Kiribati
Lao Peo Ethio Democratic Republic of
zania, the United Republic of (the)
Zimbabwe
Malawi
Uganda
Senegal
Gambia (the)
Myanmar
Ghana
Congo (the)
Tajikistan
Dominican Republic (the)
Nepal
Cambodia
Fiji
Philippines (the)
Bangladesh
Bhutan
Tuvalu
Indonesia
Guatemala
Iraq
Algeria
Viet Nam
Azerbaijan
Suriname
Paraguay
Kyrgyzstan
Palau
Sao Tome and Principe
Honduras
Moldova (the)
Uzbekistan
Brazil
Grenada
Mongolia
Nicaragua
Ecuador
Oman
Iran (Islamic Republic of)
Tonga
Mexico
Korea (the Democ
Armenia
Kazakhstan
Oman
Kuwait
Lebanon
Georgia
Albania
Costa Rica
Ukraine
Malaysia
Sri Lanka
Russian Fed
North
United Kingdom
South Africa
Montenegro
Turkmenistan
Monaco

Causal statements

Before:

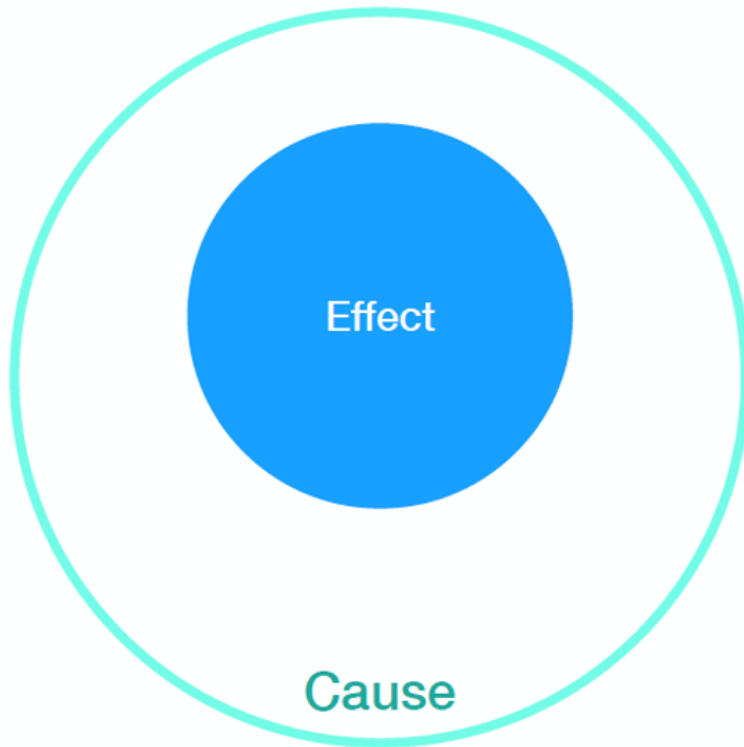- Access to clean water and infant mortality are negatively correlated

Now:

- High access to clean water is **sufficient** for low infant mortality
- Low access to water is **necessary** for high infant mortality
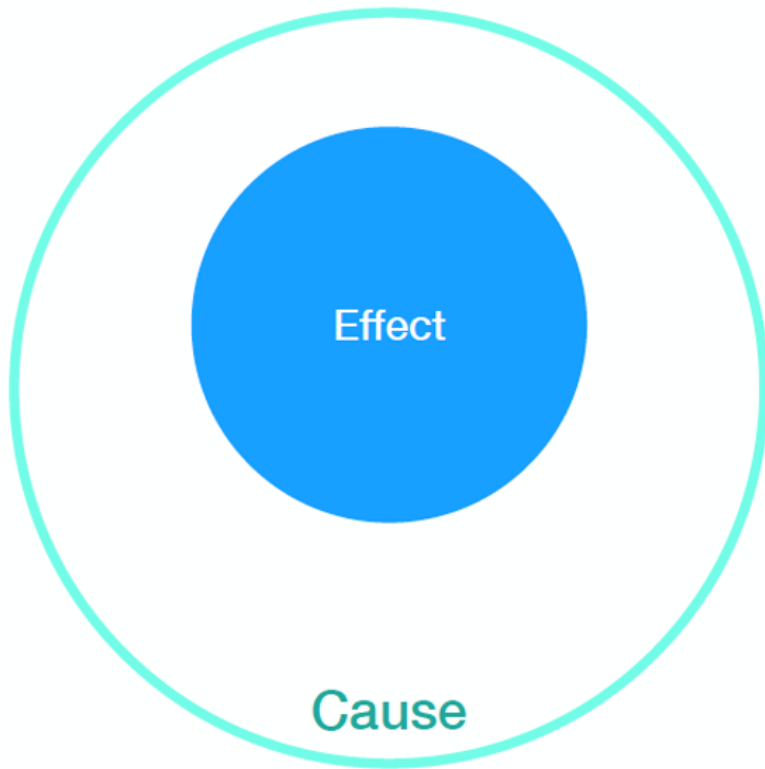
Necessary: Effect happens
when we see the cause,
but not always
(never see effect without cause)

Sufficient: Cause can explain
the effect, but not always
(never see cause without effect)

Necessary and Sufficient: Cause and effect always observed together (never observe one without the other)

Effect

Cause

Effect

Cause

Effect

Cause

Necessary: Effect happens when we see the cause, but not always (never see effect without cause)

Sufficient: Cause can explain the effect, but not always (never see cause without effect)

Logic of causation

Large N: Effect of causes

- "What is the relationship between X and Y?"

- Focus on averages, associations

- Quantify uncertainty

- Impose assumptions for causal interpretation

See Mahoney (2008) for more

Logic of causation

Small N: Causes of effect

- "Why does Y happen?"

- **Equifinality:** Multiple paths to the same outcome

- Focus on **causal mechanisms**

See Mahoney (2008) for more

Equifinality

Example: Paths to high economic growth

1. Democracy AND industrialization AND rule of law

<div align="center">OR</div>

2. Democracy OR Autocracy AND natural resources

Multiple paths also imply **combinations of causes**

Multiple causation

INUS: **I**nsufficient but **N**ecessary part of an **U**nnecessary but **S**ufficient set of conditions for the outcome

Example: *Unprotected intercourse* is **insufficient** for *HIV transmission*, but it can be a **necessary** part in one of the ways in which HIV is transmitted

aka **support factors**

Multiple causation

SUIN: Sufficient but **U**nnecessary part of a condition set that is **I**nsufficient but **N**ecessary for the outcome

Example: *Electoral fraud* is a **sufficient** condition for *democratic erosion*, which in turn is a **neccessary** condition for war

aka **precipitating causes**

Causal mechanism

> A process that explains **how** an outcome occurs.

Explaining **how** or **why** cause and effect are related entails identifying a **causal mechanism**

Large N relies **data set observations** (DSOs) that do not say much about the **mechanism**

Small N focuses on **causal process observations** (CPOs) that help us establish causation more directly
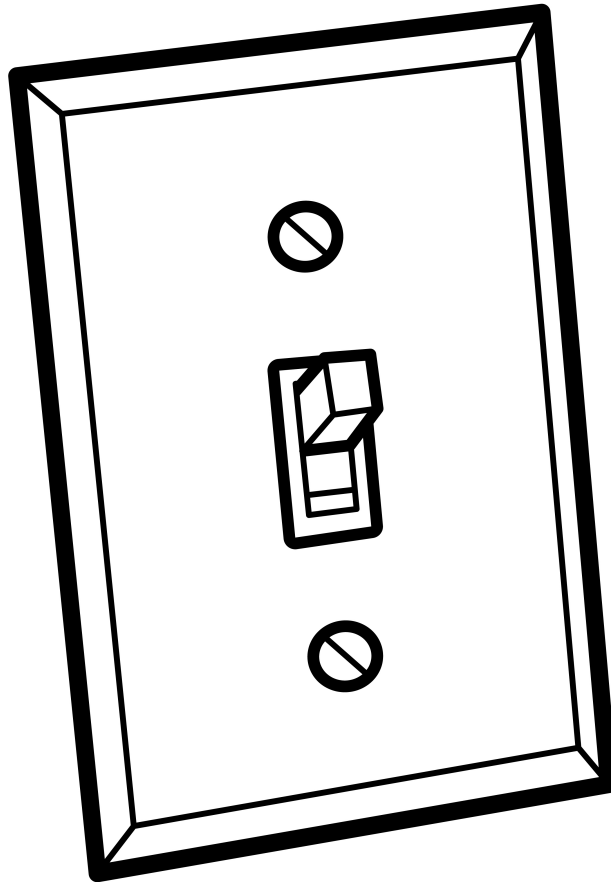
See Brady, Collier, and Seawright (2006) and Beck (2010) for more

Causal Process Observation

> An insight or piece of data that provides information about context, process, or mechanism, and that contributes distinctive leverage in causal inference.

Brady and Collier (2004, p. 227)

Wait a minute

Causal Process Observation

> An insight or piece of data that provides information about context, process, or mechanism, and that contributes distinctive leverage in causal inference.

CPOs are not a real thing!

Neither are the assumptions we make in Large N!

We make assumptions not because they are true, but because they are useful

Brady and Collier (2004, p. 227)

Case selection

We only need to look at a few cases to identify key CPOs

But how do we choose?

How not to do it

sketchplanations.com/looking-under-the-lamppost

How to do it?

Next time

Small N

POLI SCI 210

Introduction to Empirical Methods in Political Science

Last time

sketchplanations.com/looking-under-the-lamppost

Today: Case selection

Large N: Random sampling

Does it work here?

Asymptotic properties will not kick in with a few cases

Randomly chosen cases will not be *representative*

Need to sample intentionally!

Case selection methods

1. Most similar

2. Most different

3. Extreme cases

4. Deviant case

5. Diverse cases

6. Influential cases

See Seawright and Gerring (2008) for more

Most similar

Case	X_1	X_2	Y
1	+	+	+
2	-	+	-

- Needs at least two cases
- Similar in every variable except for key explanatory and outcome variable
- **Example:** Economic development in South Korea and North Korea since the war
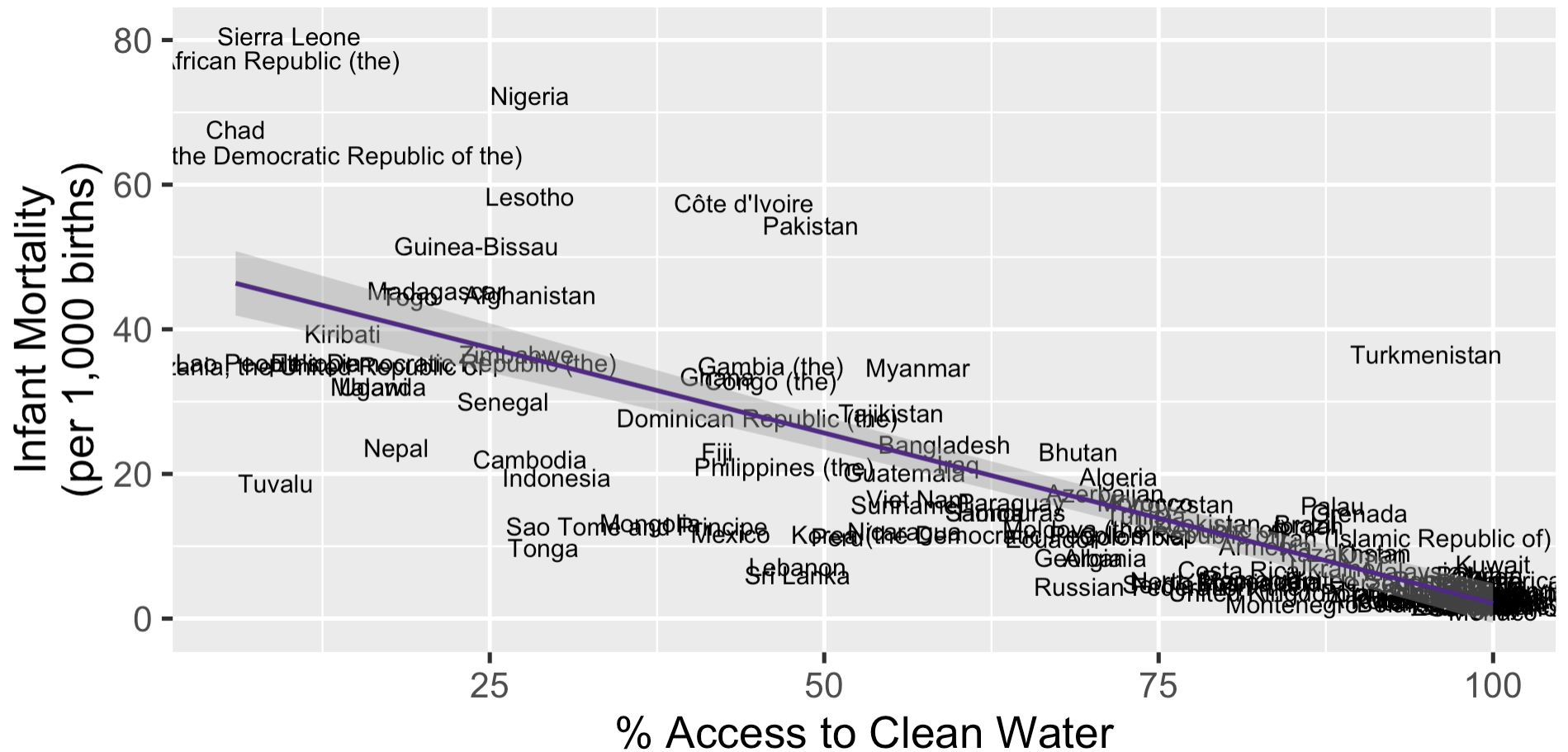
Most different

Case	X_1	X_2	Y
1	+	+	+
2	+	-	+

- Cases differ on everything except for key explanatory and outcome variable

- **Example:** Japan and Australia driving on the left

Typical case

- Common or representative case that a theory explains

- **Representative** by definition

- **Example:** France, Russia, China as *typical* cases of social revolutions

- Used to confirm causal mechanism proposed by the theory

- **Trap:** Usually, you don't want to look at a typical case **only**

Typical cases are "on the line"

Extreme case

- Extreme value on explanatory OR outcome variable

- Far away from the "mean"

- Valuable because it is *rare* in one dimension

- **Example:** Qatar and Saudi Arabia as cases of extreme autocracy (Polity score of -10)

Deviant case

- Surprising value on some variable

- Poorly explained by the theory

- Used to explore potential alternative explanations

- Once explained, not deviant anymore

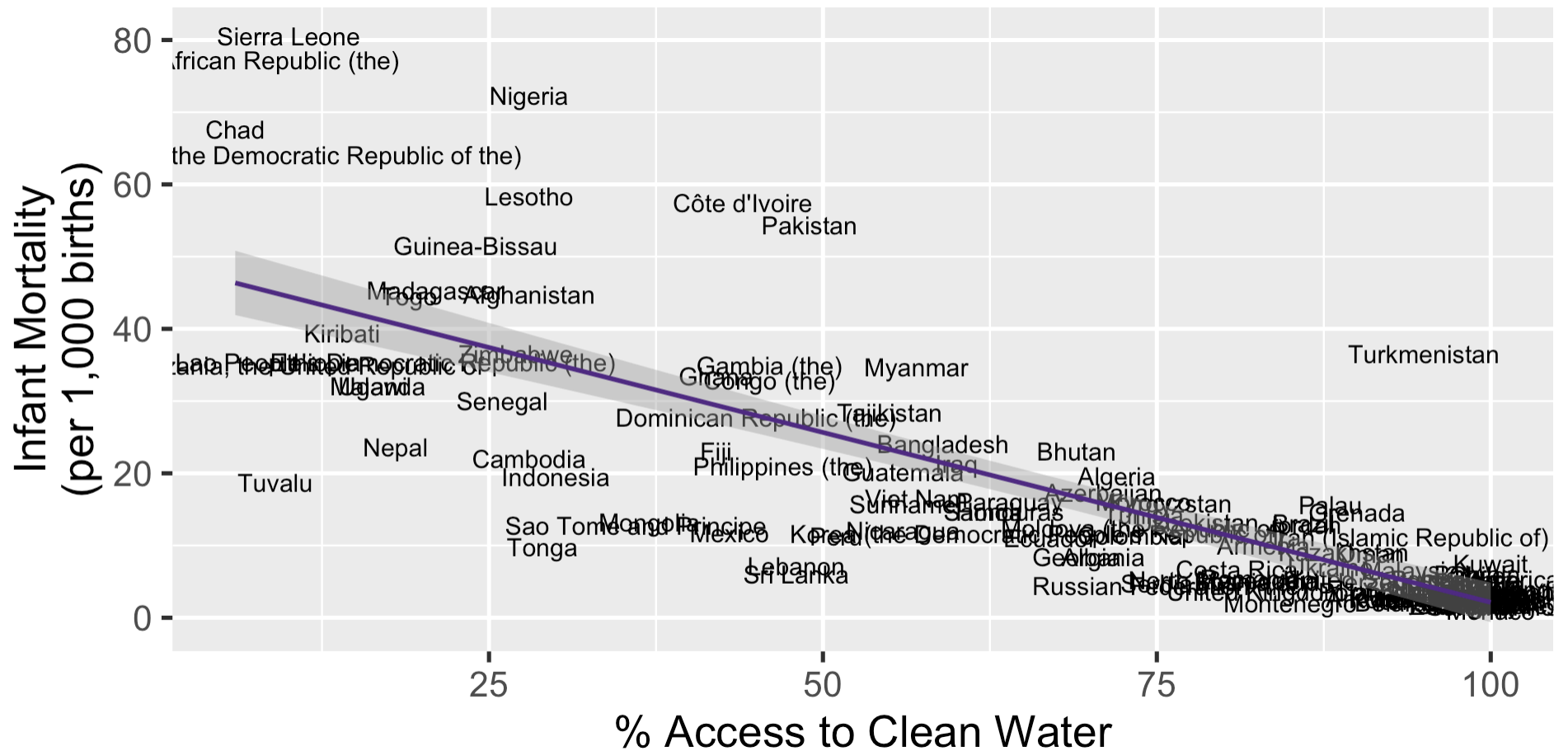- **Example:** Norway as an oil-state that is not an autocracy

Diverse cases

- Choose cases to maximize variance along relevant dimensions

- At least two cases

- Represent full range of values characterizing X and Y

- **Example:** Colonial heritage influences economic development. Look at French, Spanish, British colonies

Influential case

- **Concern:** results may be driven by one or a few cases

- Choose a case to check the assumptions behind proposed causal mechanism

- **Example:** Relationship between mass education and democratization *influenced* by countries who expanded education before becoming democracies

-

Practice

Scatter plot titled "Practice" showing the relationship between % Access to Clean Water (x-axis, 25 to 100) and Infant Mortality (per 1,000 births) (y-axis, 0 to 80) for various countries, with a downward-sloping purple regression line.

Countries labeled include: Sierra Leone, African Republic (the), Nigeria, Chad, the Democratic Republic of the), Lesotho, Côte d'Ivoire, Pakistan, Guinea-Bissau, Madagascar, Afghanistan, Togo, Kiribati, Zimbabwe, Gambia (the), Myanmar, Congo (the), Turkmenistan, Tanzania, Ethiopia, Lao People's Democratic Republic (the), Ghana, Tajikistan, Malawi, Uganda, Senegal, Dominican Republic (the), Nepal, Cambodia, Fiji, Bangladesh, Bhutan, Indonesia, Philippines (the), Guatemala, Iraq, Algeria, Tuvalu, Viet Nam, Azerbaijan, Suriname, Paraguay, Kyrgyzstan, Palau, Sao Tome and Principe, Mongolia, Samoa, Honduras, Kazakhstan, Brazil, Grenada, Mexico, Korea (the Democratic Republic), Nicaragua, Moldova (the), Ecuador, Colombia, Iran (Islamic Republic of), Tonga, Georgia, Oman, Kuwait, Lebanon, Russian Federation, Costa Rica, Malaysia, Sri Lanka, United Kingdom, Montenegro, Monaco

Moving on

Now we know how to select cases for small N analysis

But what do we do with these cases?

Or, what if we can only conduct **within-case analysis**?

Qualitative methods

Ordered by increasing *level of immersion*

1. Process tracing

2. Interviews

3. Focus groups

4. Participant observation

5. Ethnography

Becoming a detective

The Mole Agent Official Trailer Releasing September 1

https://youtu.be/5Z1hPH3b5RA?si=9sntmHBU5n_Ek1qG

Process-tracing

- Usually based on **archival research**

- Extensive analysis of **primary** or **secondary** sources

- **Primary:** Original document or first-hand account

- **Secondary:** Summaries, critiques, opinions, analyses of those who did not participate or witness directly

- Find *pieces of evidence* that assist on *establishing causality*

Process tracing tests

Process Tracing Tests for Causal Inference

| | | **SUFFICIENT FOR AFFIRMING CAUSAL INFERENCE** | |
		No	**Yes**
NECESSARY FOR AFFIRMING CAUSAL INFERENCE	**No**	**1. Straw-in-the-Wind**	**3. Smoking-Gun**
		a. Passing: Affirms relevance of hypothesis, but does not confirm it.	**a. Passing:** Confirms hypothesis.
		b. Failing: Hypothesis is not eliminated, but is slightly weakened.	**b. Failing:** Hypothesis is not eliminated, but is somewhat weakened.
		c. Implications for rival hypotheses: Passing *slightly* weakens them. Failing *slightly* strengthens them.	**c. Implications for rival hypotheses:** Passing *substantially* weakens them. Failing *somewhat* strengthens them.
	Yes	**2. Hoop**	**4. Doubly Decisive**
		a. Passing: Affirms relevance of hypothesis, but does not confirm it.	**a. Passing:** Confirms hypothesis and eliminates others.
		b. Failing: Eliminates hypothesis.	**b. Failing:** Eliminates hypothesis.
		c. Implications for rival hypotheses: Passing *somewhat* weakens them. Failing *somewhat* strengthens them.	**c. Implications for rival hypotheses:** Passing *eliminates* them. Failing *substantially* strengthens.

See Collier (2011) for more

Interviews

- **Structured:** Stick to script. Easy to compare across subjects

- **Semi-structured:** Start with guide, allow for deviations. Somewhat comparable

- **Unstructured:** Let people say whatever they want. Harder to compare and more time consuming to analyze

Key informants more important than a **representative sample**

Random sampling?

Non-random sampling?

Convenience   Quota   Respondent-driven

Stratum (e.g., locality)
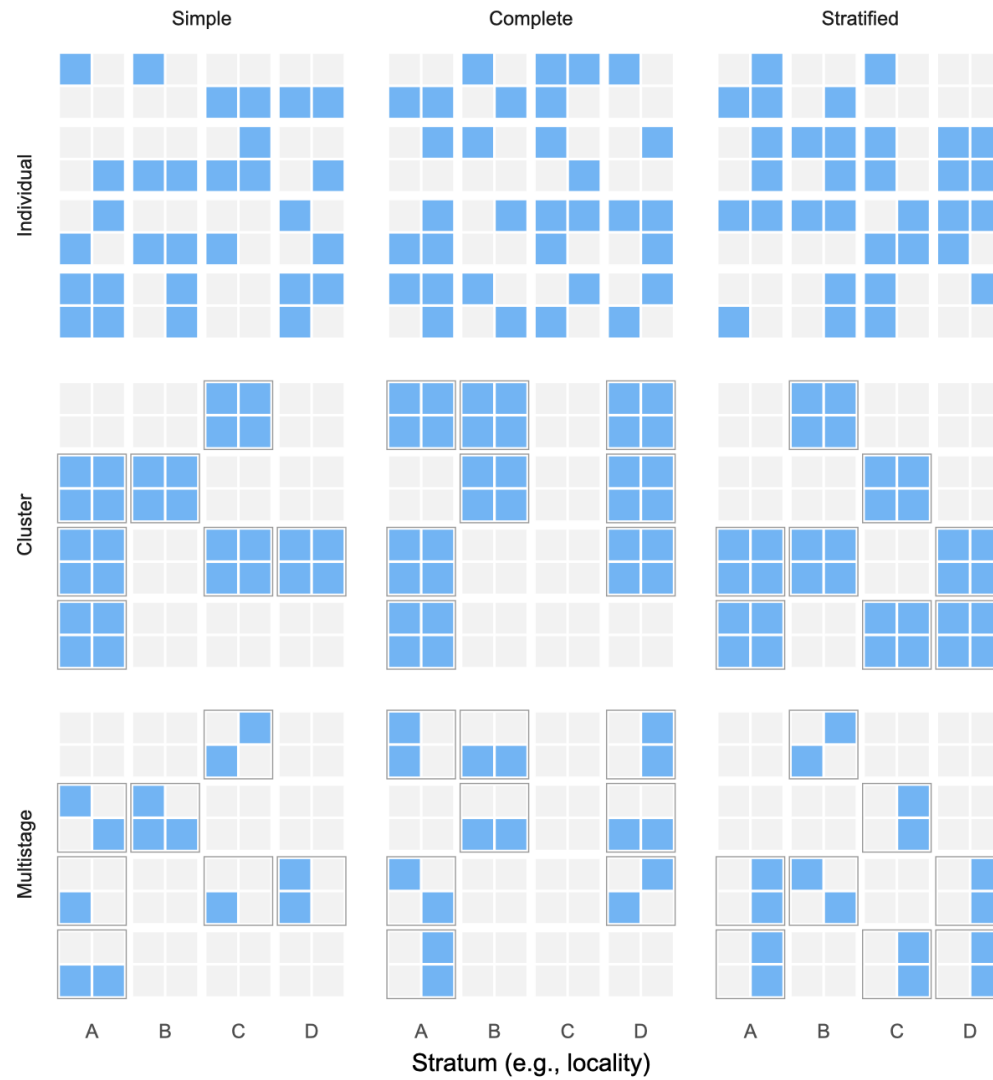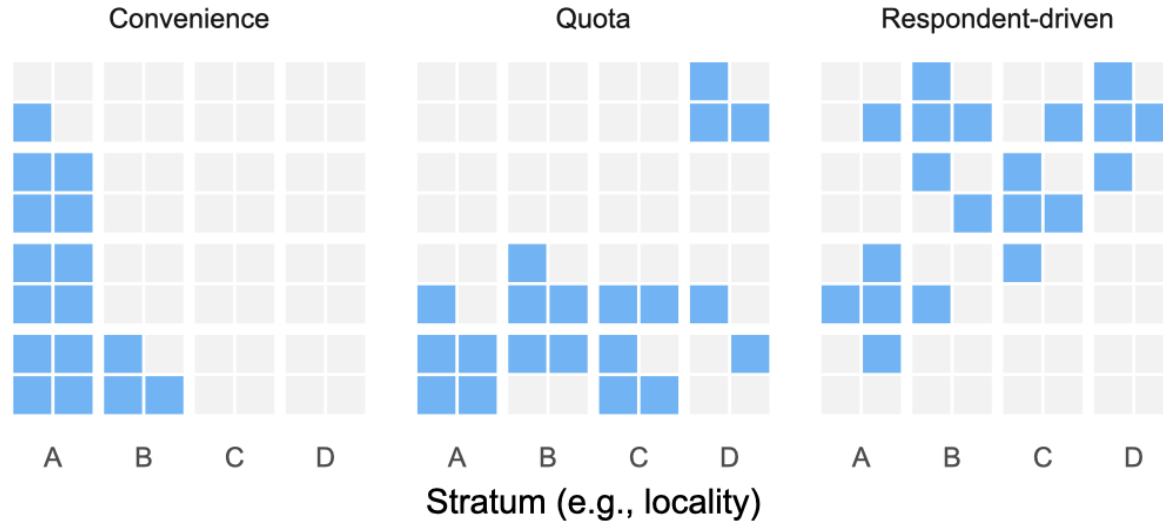
Places to start:

- Key informants
- Send out call on relevant platform
- Snowball sampling

Focus groups

- Essentially an interview with multiple people in a controlled space

- Interactions in group discussion provide data

- Researcher acts as facilitator

Why focus groups?

Asch Conformity Experiment

https://youtu.be/TYIh4MkcfJA?si=pucARllaLPgLA3Bo

Participant observation

- Researcher acts as an observer for an *extended period* of time

- Unlike focus groups, observe subjects *in their natural habitat*

- Must determine **level of involvement** and **whether to reveal identity**

Types of participant observation

Involvement	Overt	Covert
		Identity
Passive	City hall meeting	Public space, hidden camera
Active	Visiting rural areas	Community center

Ethnography

- **Deep immersion** in a culture or community

- Combination of interviews, focus groups, participant observation

- **Thick description:** Produce vast amount of data (notes) to capture the complexity of behavior and its context

And I mean THICC

A CLUTTERED LIFE: Middle-Class Abundance

https://youtu.be/3AhSNsBs2Y0?si=ehUIf9COf9x9KAtz

Positionality

> A person's social location and how it shapes their perspective

- Affects researcher's relationship with and access to participants
- **Large N:** Same protocol, same results
- **Small N:** Same protocol, *different* results
- Can be an advantage and a disadvantage

Summary

- Only one week on qualitative methods 🙁
- Cause of effects vs. effects of causes
- **Inference:** Necessary and sufficient vs. correlational
- **Case-selection:** Avoid convenience traps!
- **Techniques:** Many available, often combined, but *positionality* matters