Elements of quantitative research

Francisco Villamil

Research Design for Social Sciences
MA Computational Social Science, UC3M
Fall 2023

Roadmap

Theories and research questions

Concepts and operationalization

Measurement

Describing variables and relationships

Example: Wartime civilian deaths

Paper discussion

Research and RQs

- Research answering questions
- Re-cap on research types
  1. Normative vs. positive research
  2. Positive: theoretical vs. empirical
  3. Empirical: Descriptive, explanatory (and predictive?)

What is a research question?

- Any question we can answer
- Sometimes we say that we derive an RQ from a topic, and a theory from an RQ
  - $\rightarrow$ Topic $>$ RQ $>$ Theory

What is a research question?

- Any question we can answer
- Sometimes we say that we derive an RQ from a topic, and a theory from an RQ
  - $\rightarrow$ Topic $>$ RQ $>$ Theory
- In reality, an RQ can be thought of as an operationalization of an argument
  - $\rightarrow$ Previous evidence $>$ Argument $>$ RQ $>$ Hypotheses
- **Even though** this 'argument' can be something anecdotical that we later abstract into a proper theoretical argument

What is a research question?

- Any question we can answer
- Sometimes we say that we derive an RQ from a topic, and a theory from an RQ
  - → Topic > RQ > Theory
- In reality, an RQ can be thought of as an operationalization of an argument
  - → Previous evidence > Argument > RQ > Hypotheses
- **Even though** this 'argument' can be something anecdotical that we later abstract into a proper theoretical argument
  - → And it would actually look into something like this:
  - → Previous > 'Anecdotal argument' > RQ > (Proper) Theory > Hs

Example


Internacionales UMU
@internacionalUM
¿Y sabías que encuentran #trabajo más rápido que los estudiantes que no realizan #movilidad?
3 de cada 4 graduados Erasmus+ consideran que su #experiencia en el extranjero es beneficiosa para encontrar su primer trabajo, y el 80% encontraron trabajo 3 meses después de graduarse.
Translate Tweet
Erasmus+
ESTUDIOS
DIFERENCIATE DEL RESTO
Los estudiantes Erasmus+ encuentran trabajo más rápido que los estudiantes que no realizan movilidad.
ELIGE TU DESTINO
ELIGE ERASMUS
Solicitud abierta hasta el 14/12/20
erasmus.um.es
UNIVERSIDAD DE MURCIA
12:05 PM · Nov 23, 2020 · Twitter Web App

Example

- That's some descriptive evidence that could inspire an anecdote
- The **anecdotal argument** (think of a story)
- The **research question**
- The 'proper' **theory**
- The **hypotheses**?

Example

- That's some descriptive evidence that could inspire an anecdote
- The **anecdotal argument** (think of a story)
  - → My friend John who went on Erasmus has more money than my other friend who couldn't go, and also, John managed to get a job because his father is partner at a local firm
- The **research question**
  - → Is there a causal effect of Erasmus on labor market early success? Is the effect mediated by household income?
- The 'proper' **theory**
  - → Going on Erasmus does not have any causal effect on getting a first job, the relationship is explained by the confounding effect of income
  - → Or: Positive effect among high-income students because they have access to informal networks where this experience is valued
- The **hypotheses**?

Generating theories

- No recipe for this, everyone generates theories *all the time*
- Usually it refers to an analytical argument that explains something
  - → It could also be a descriptive or predictive theory, but even in those cases there's probably an explanation underneath
- Developed inductively, from descriptive data to general explanations
- This is personal, but if you can't tell a story out of the theory, you're not there yet (i.e. need to be able to travel from/to abstraction)

Generating theories

- No recipe for this, everyone generates theories *all the time*
- Usually it refers to an analytical argument that explains something
  - → It could also be a descriptive or predictive theory, but even in those cases there's probably an explanation underneath
- Developed inductively, from descriptive data to general explanations
- This is personal, but if you can't tell a story out of the theory, you're not there yet (i.e. need to be able to travel from/to abstraction)

- **Q:** How to identify a **good theory**?

Research and RQs

- Normally we think that we derive an RQ from a topic, and a theory from an RQ

Research and RQs

- Normally we think that we derive an RQ from a topic, and a theory from an RQ
- The process is more circular

Research and RQs

- Normally we think that we derive an RQ from a topic, and a theory from an RQ
- The process is more circular
- Theory ¿ RQ ¿ Hypothesis

Research and RQs

- Normally we think that we derive an RQ from a topic, and a theory from an RQ
- The process is more circular
- Theory ¿ RQ ¿ Hypothesis
- RQs as operationalization of theory

Hola

Roadmap

Theories and research questions

Concepts and operationalization

Measurement

Describing variables and relationships

Example: Wartime civilian deaths

Paper discussion

Good RQ

1. consider potential results of the analyses
   - → if you found X, does that answer the question?
   - → example: are kids who play videogames often more aggresive?
   - → does that inform a theory on the aggresiveness effect of VGs?
2. Is it feasible?
   - → do you have the data? is it possible to do it? (e.g. re-offenders)
   - → also: is there any design or strategy to answer it?
3. keeping it simple and narrow
   - → what are the causes of economic underdevelopment? vs. does exposure to natural disasters hinders economic development?

Computational methods and theory

- Limitations of *data mining*
- Focus on the what rather that on the why
  - → example of ice creams and short pants

Roadmap

Theories and research questions

Concepts and operationalization

Measurement

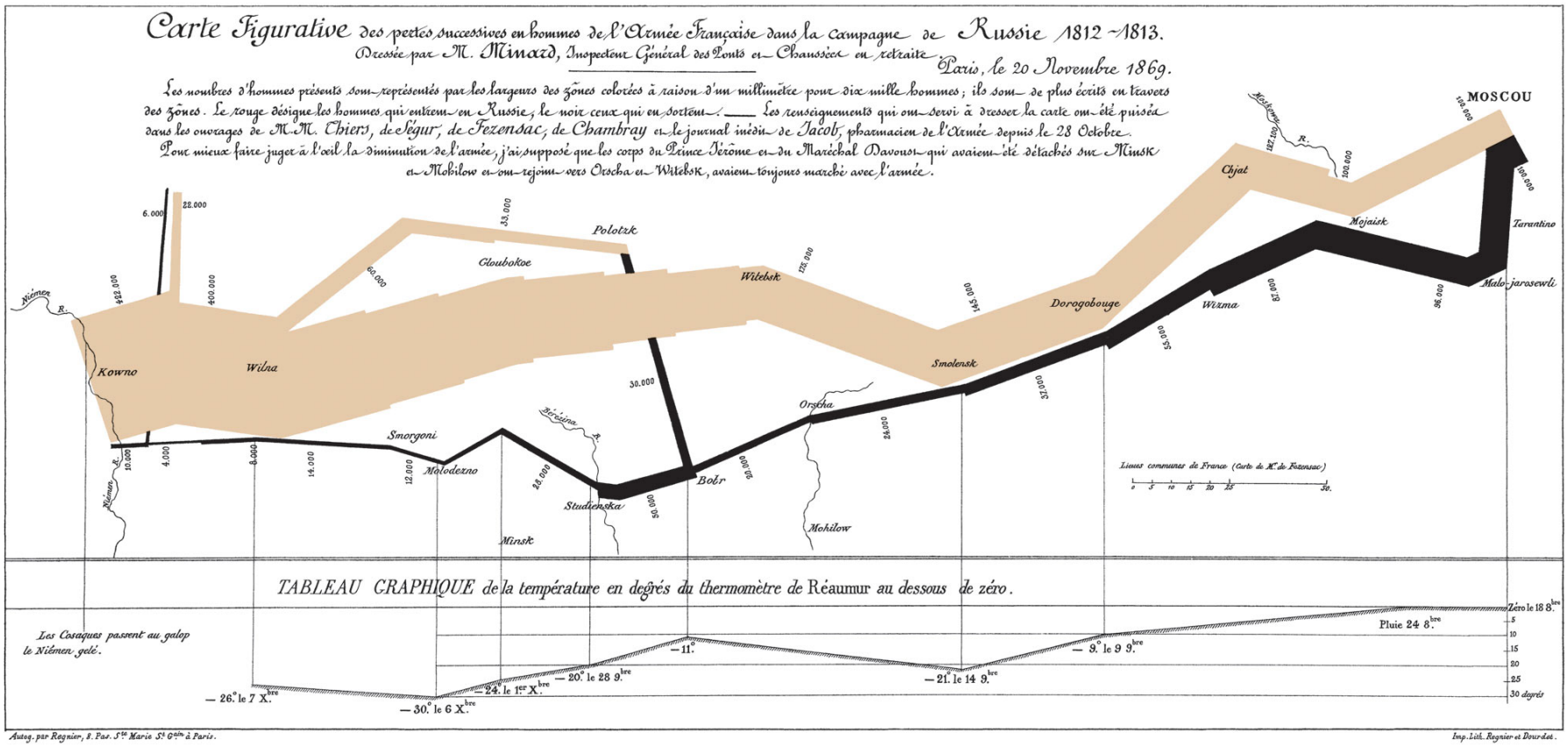
Describing variables and relationships

Example: Wartime civilian deaths

Paper discussion

Unit of analyses

A more difficult example



- How many variables and what's the unit of observation?

Extra: What's the causal argument being told here?

Could we test it with the data we have?

Introduce sampling bias (later develop with DAGs)

Missing data in DAG framework

Missing completely at random (MCAR) Missing at random (MAR)
Missing not at random (MNAR)

Roadmap

Theories and research questions

Concepts and operationalization

Measurement

Describing variables and relationships

Example: Wartime civilian deaths

Paper discussion

Describing variables

- What is a **variable**?
- Types?

Describing variables

- What is a **variable**?
- Types?
  - → Continuous

Describing variables

- What is a **variable**?
- Types?
  - → Continuous
  - → Count

Describing variables

- What is a **variable**?
- Types?
  - → Continuous
  - → Count
  - → Ordinal

Describing variables

- What is a **variable**?
- Types?
  - → Continuous
  - → Count
  - → Ordinal
  - → Categorical (binary)

Describing variables

- What is a **variable**?
- Types?
  - → Continuous
  - → Count
  - → Ordinal
  - → Categorical (binary)
  - → Qualitative (* are really a variable?)

Describing variables

- What is a **variable**?
- Types?
  - → Continuous
  - → Count
  - → Ordinal
  - → Categorical (binary)
  - → Qualitative (* are really a variable?)
- Why does it matter?
- Conceptual meaning vs statistical meaning

Describing variables

main idea: variable distribution (of values)

you probably know this because of basic statistics. it doesn't matter so much

kurtosis, normality, blah blah

but one thing though: values are real-world observations. do **plot** it, and make sure it's coherent with the **theoretical** or **expected distribution**

maybe also important: mean +/- SD, or 25/75 percentiles. for effect sizes

in a normal distribution, there not much to say. but e.g. if an independent variable is a bimodal distribution, does that say something about the causal mechanism?

e.g. think about the effect of income on X, in a super unequal society vs one in which is normally distributed

Describing relationships

- What is a **relationship**?
- Essentially that as you know about the values of one variable, you learn about the values of the other variables

  e.g. a *negative* relationship means that you know that higher values in x imply lower values in y

Describing relationships

- What is a **relationship**?
- Essentially that as you know about the values of one variable, you learn about the values of the other variables

  e.g. a *negative* relationship means that you know that higher values in x imply lower values in y

  Imagine you have a small car, and a friend of yours is coming and is bringing along his two kids. Concerned about space, you ask '*how old are they?*' And the answer is: '*They're 6 and 2.*'

  $\rightarrow$ What do you imagine about size?

Describing relationships

- What is a **relationship**?
- Essentially that as you know about the values of one variable, you learn about the values of the other variables

  e.g. a *negative* relationship means that you know that higher values in x imply lower values in y

  Imagine you have a small car, and a friend of yours is coming and is bringing along his two kids. Concerned about space, you ask '*how old are they?*' And the answer is: '*They're 6 and 2.*'

  → What do you imagine about size?

  → Now imagine you ask '*are they blonde, red-haired, or brown-haired?*'

Statistical relationship $\neq$ causal relationships

- Last example: Is there a causal relationship *age* $\rightarrow$ *size*?

Statistical relationship $\neq$ causal relationships

- Last example: Is there a causal relationship *age* $\rightarrow$ *size*?
- What if the variable you want to guess is **the time of the day**, and someone tells you that she just heard the rooster crow? Causal?

Statistical relationship $\neq$ causal relationships

- Last example: Is there a causal relationship *age* $\rightarrow$ *size*?
- What if the variable you want to guess is **the time of the day**, and someone tells you that she just heard the rooster crow? Causal?

- Why are non-causal descriptive relationships **useful**?

Describing relationships

conditional distributions conditional means conditional conditional means (controlling)

Roadmap

Theories and research questions

Concepts and operationalization

Measurement

Describing variables and relationships

Example: Wartime civilian deaths

Paper discussion

Practical example

- You want to test an argument about **wartime civilian deaths:**
  - → The intuition you have is that civilians will be more likely to be treated well (and not killed) by rebel groups during civil wars when they need their resources (e.g. labor) to survive

- Clean up the theory, decide on the main concepts
- Develop different RQ at different levels
- How can we measure the main concepts? Variables?
- What answers could we get from the data?
  - → Are we learning something about our theory?

Roadmap

Theories and research questions

Concepts and operationalization

Measurement

Describing variables and relationships

Example: Wartime civilian deaths

Paper discussion

Journal of Conflict Resolution
2021, Vol. 65(2-3) 563-590

DOI: 10.1177/0022002720963674
journals.sagepub.com/home/jcr


Roads to Rule, Roads to Rebel: Relational State Capacity and Conflict in Africa

Carl Müller-Crepon[1], Philipp Hunziker[2], and Lars-Erik Cederman[3]

Abstract

Weak state capacity is one of the most important explanations of civil conflict. Yet, current conceptualizations of state capacity typically focus only on the state while ignoring the relational nature of armed conflict. We argue that opportunities for conflict arise where relational state capacity is low, that is, where the state has less control over its subjects than its potential challengers. This occurs in ethnic groups that are poorly accessible from the state capital, but are internally highly interconnected. To test this argument, we digitize detailed African road maps and convert them into a road atlas akin to Google Maps. We measure the accessibility and internal connectedness of groups via travel times obtained from this atlas and simulate road networks for an instrumental variable design. Our findings suggest that low relational state capacity increases the risk of armed conflict in Africa.