

# Developing a social science research question

Bernie Hogan

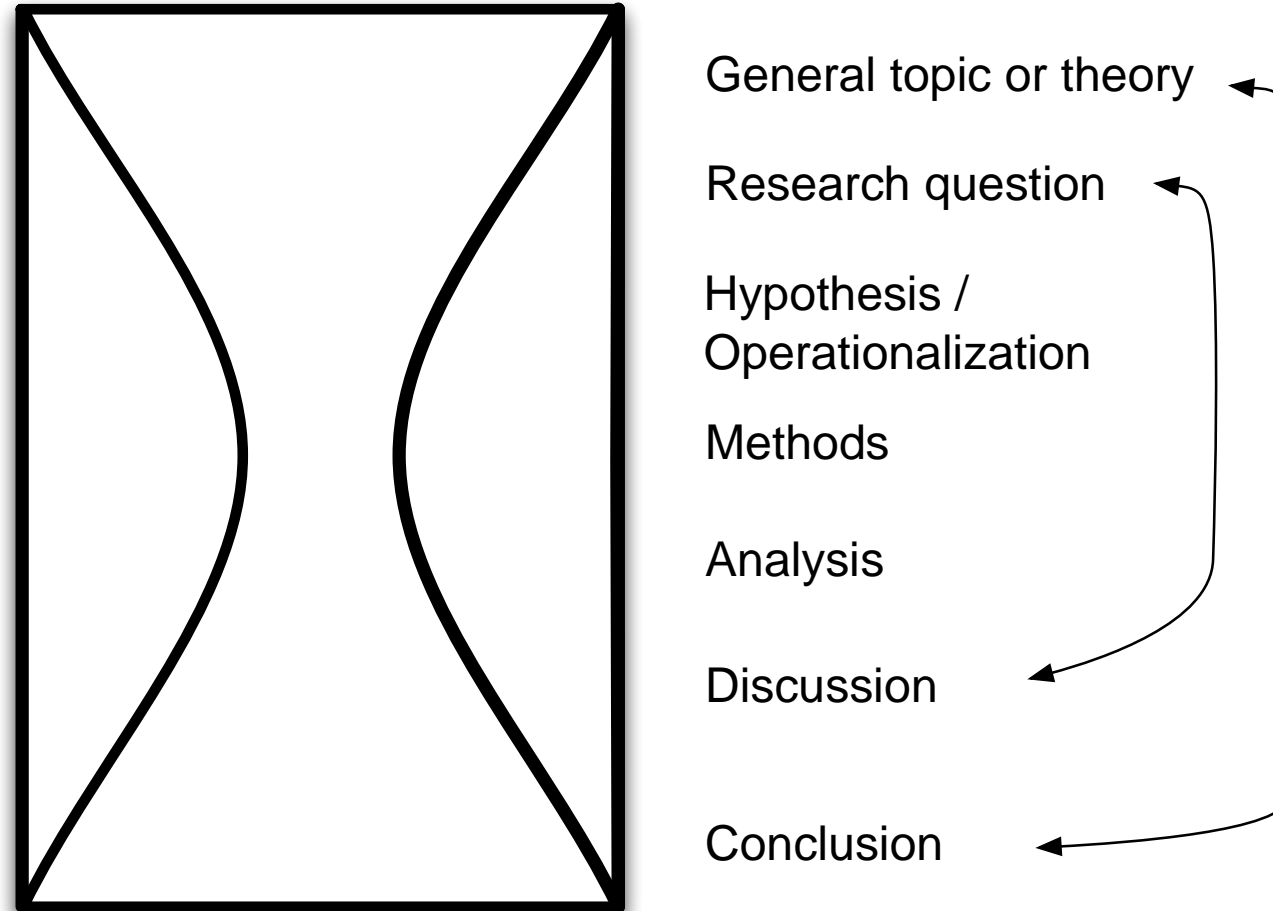
Python for Social Data Science, Week 4, Lecture 3 [afternoon]

November 1, 2018

# Why bother with Research Questions? (or why big data won't solve all our problems)

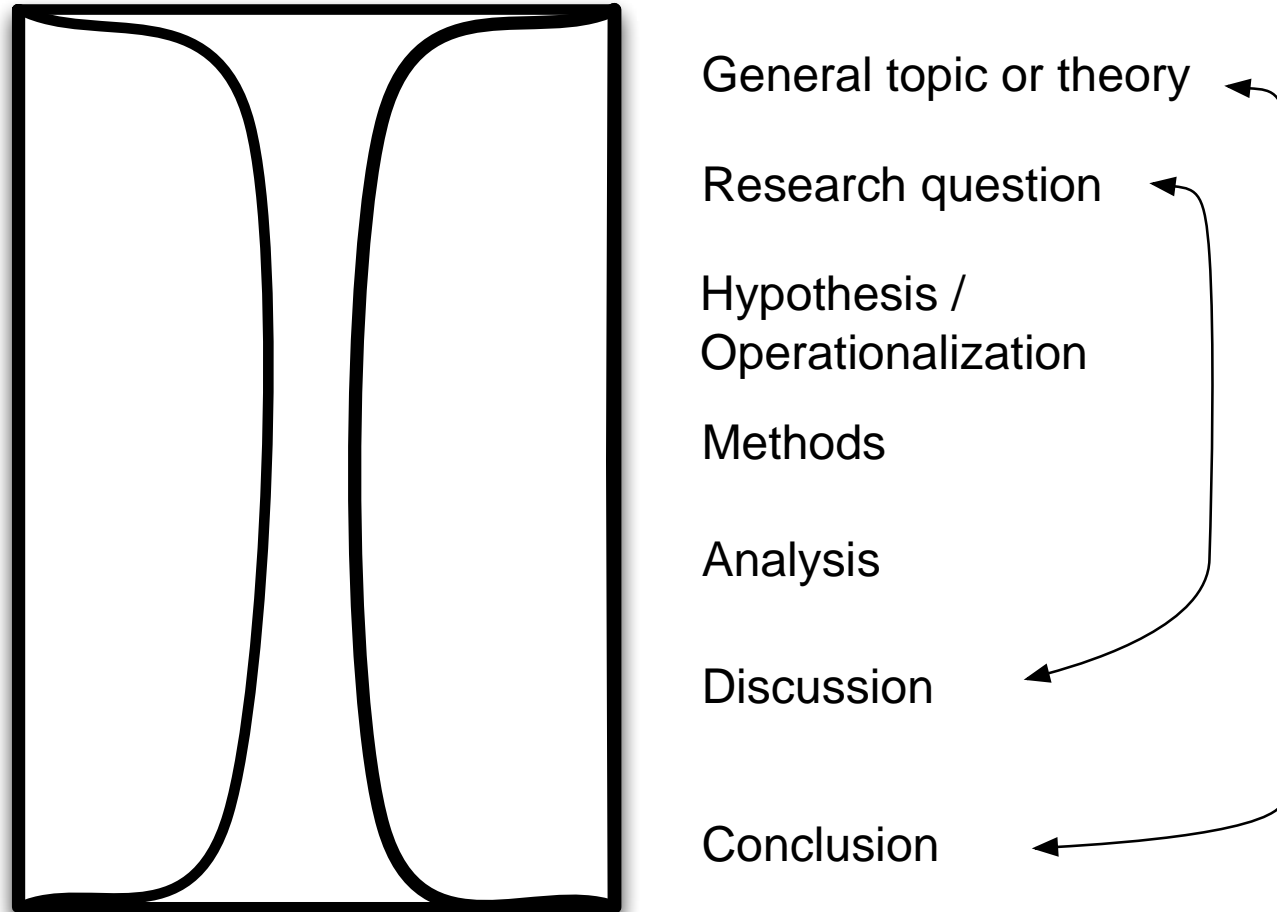
- We need a mechanism to focus our inquiry from a general topic or problem towards specific constructive claims.
- This needs to be more general than a specific falsifiable hypothesis but less general than a broad theory.
- Research questions clarify our intentions about the research site or problem.
- This enables us to be in command of our research activities, rather than let the phenomenon or data drive the research activity.

# The hourglass model of research



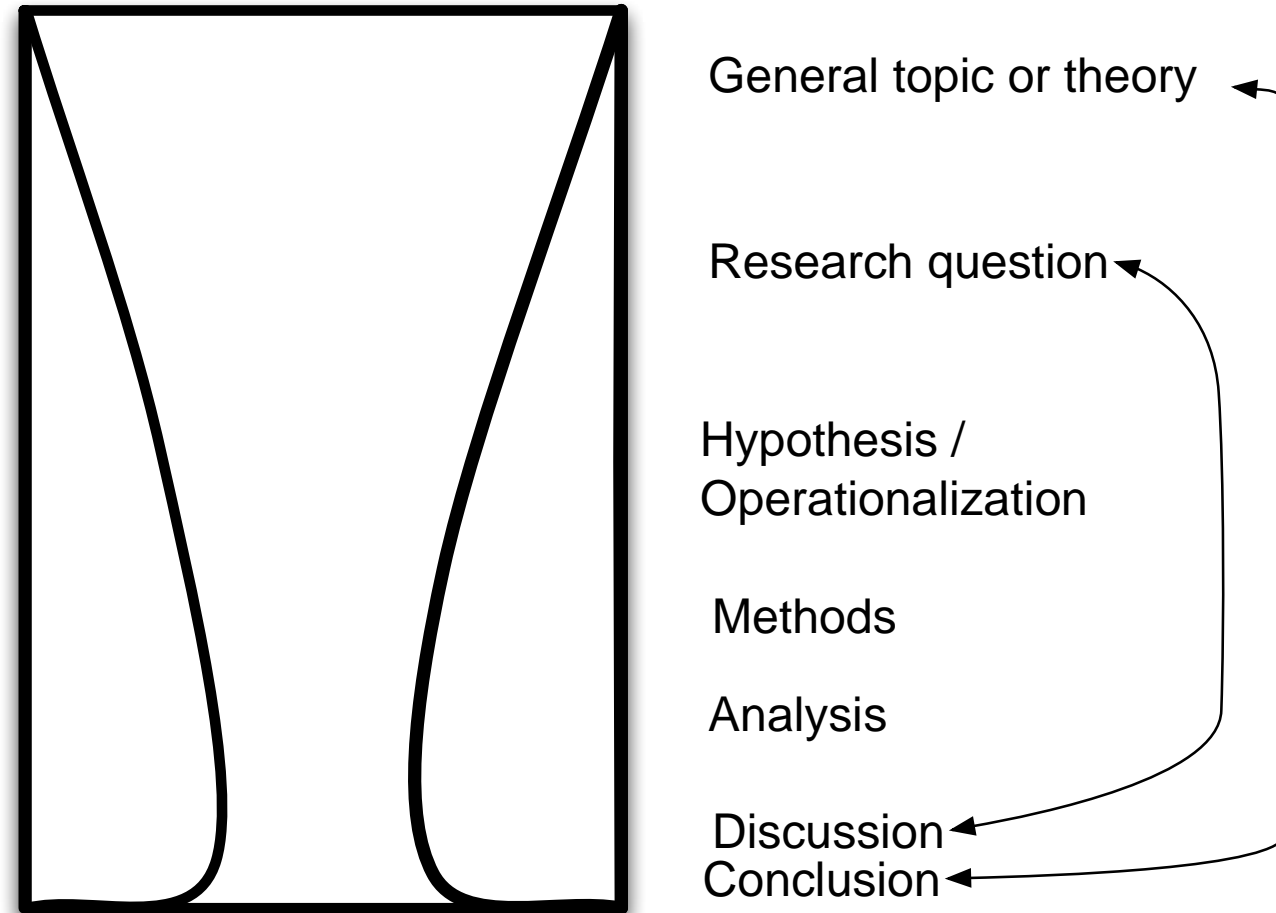
Ideal: Research question enables a smooth transition from general topic to specific issue.

# The hourglass model of research



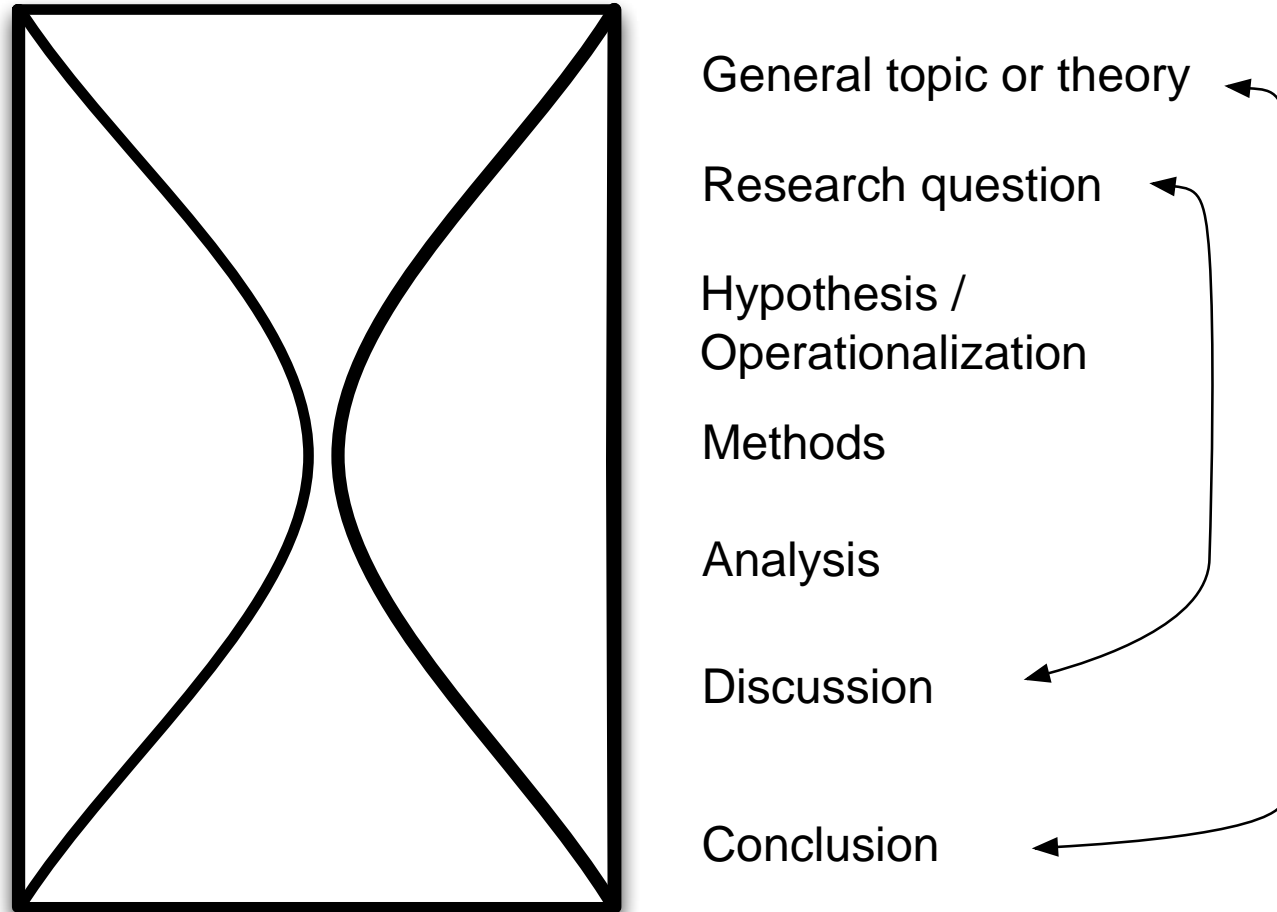
Problem: Analysis is not properly contextualised or framed.

# The hourglass model of research



Problem: Not bringing back to the topic and the RQs

# The hourglass model of research



Problem: Analysis is too specific to support the research questions.

# How to choose a research topic

- Select an area where:
  - You have or believe you can have more than a passing level of experience or expertise.
  - You can draw strength to work on this topic even when times are tough.
  - There is enough that is not known that your work can make a meaningful novel claim.

# Research question as resource

- A good research question enables, by providing most of the following:
  - An ability to resolve given **resources** (i.e. what do I need in terms of time, data, researchers, etc...)
  - Specific **language that clarifies** what aspect of phenomena are under scrutiny.
  - A puzzle, an **unknown answer** or paradox where systematic inquiry can contribute.
  - The capacity to reach back towards the **wider implications** for other research.
  - Generativity: first the question should generate **specific technical hypotheses**, and the answer should generate **additional interest in the topic**.



# Research questions and operationalisation

- Operationalisation is a key part of research in computational social science. It is the act of translating a concept which we understand from theory into something measurable.
- For example, the study this morning on emotional contagion was really a study on sentiment analysis.
- The study on darknet markets was a study of cannabis, opiates and cocaine (but not other drugs).
- Operationalisation needs to be considered with care to support the general ideas advanced at the beginning of the paper.

# A simple typology drawing on Earl Babbie

| Type        | General   | Example   |
|-------------|---|---|
| Descriptive | Describe the <i>{existence of / prevalence of / share of}</i> X                     | How much time do people spend on social media?                              |
| Exploratory | How does X <i>{compare to / break down by / is distributed over / works with}</i> Y | How does social media use vary by age group?                                |
| Explanatory | What role does Z play in the relationship between X and Y                           | Do income levels mediate the relationship between social media use and age? |

# Qualities of Research Questions

| Type                      | General                               | Example   |
|---------------------------|---------------------------------------|---|
| Causal                    | Does X cause Y                        | Do videogames increase aggression in children?  |
| Non-causal experiential   | What is X? How is X interpreted?      | How do visible minority children see themselves when playing video games with white protagonists? |
| Non-causal policy-related | What does X mean [for Y stakeholder]? | What constitutes <i>social exclusion</i> on the internet?   |
| Non-causal evaluative     | What makes X useful?                  | Does the inclusion of computers in libraries reduce digital skills divides in rural UK            |

# A Research Question is not a hypothesis!

- A research **question** is a focusing device and is an essential component of most all social science research.
- A hypothesis is a falsifiable and testable **statement**.
- Typically a hypothesis suggests both a specific operationalisable domain and direction:
  - Research question: Do design elements of video games have an effect on the child's mood after playing?
  - Hypothesis 1: Video games with flashing elements make children more hyperactive.
  - Hypothesis 2: Any effect of video games on a child's mood dissipates within an hour.

# Inductive, abductive and deductive RQs

- Inductive reasoning draws from the concrete and observed towards the abstract.
  - Example: What strategies do software pirates use to coordinate their capture, uploading and distribution of torrents?
- Deductive reasoning draws from abstract principles towards what can be observed.
  - Example: Do Facebook users deliberately enhance their social capital by posting about life events?
- Abductive reasoning is about starting from observations to draw the most likely inference. Differs from inductive in that it is not seeking general principles but plausible explanations that are continually updated
  - Example: If blood sugar is high, what is the likely cause? Diabetes would be a first plausible one. It's not definite as we are generalizing from outcomes not causes, but helps us with priors.

# Research questions and “ground truth”

- There is no ground truth in an absolute sense.
- We have systems where knowledge is contained, and we seek to benchmark against this. These benchmarks are often reinforced through intraclass correlation / reliability statistics.
- Testing a system versus ground truth is a common data science task, but can be a bit uninteresting from a social scientific perspective unless further applied.
- How well does X perform is really “against what benchmark”, wherein we ask about the hidden biases of the benchmarks.

# RQs and Scope

- Typically research questions embed notions of scope. They are the place where we transition between general ideas and a specific population where we investigate ideas.
- Examples of scope:
  - Age / Period / Cohort (also crucial for modeling)
  - Geography
  - Population features (non-users, women, personality types, etc...)
- This scoping occurs partially because of practical issues:
  - Time required
  - Cost of research
  - Accessibility

# RQs and comparison groups

- Be mindful of what you are *comparing to* or *comparing within*.
- For example, researching the careers of celebrities before and after their appearance on the Muppet Show should imply a reasonable comparison group.
- Within here we can start to think about counterfactuals, both within and between sets. Within-set counterfactuals might have to do with algorithmic outcomes (e.g., if we tune this parameter how does the classification change). Between-set counterfactuals tends to counter selection bias (e.g., what if X was in the other group).
- See: <https://arxiv.org/ftp/arxiv/papers/1711/1711.00399.pdf>



# Pitfalls

- Too many research questions!
  - You do not need an RQ for every single variable or variant in your work. It overcomplicates the analysis and makes storytelling difficult.
- Overly broad research question
  - A great deal of work starts as descriptive or exploratory. But work should not be presented in this way.
  - E.g., “How does capitalism affect self-presentation online”
- Overly qualified research question
  - “How do senior citizens in Leicester make use of their social networks when they cannot access the Internet from home”.

# Summary

- A research question is a focusing device for transitioning between general phenomena and theory to practical and attainable claims of scientific knowledge.
- Research questions often, but not necessarily, lead to hypotheses.
- They should be introduced early in a paper to show the core focus and reintroduced late in a paper to show that the work was done.

# Exercise

- Part 1. (5 mins) Take five minutes and write down two research questions for your work.
- Part 2. (5 mins) Pick a partner and explain your research question to the other person. Ask for their feedback.
- Part 3. (10 mins) Round table; random draw – you will be asked to present **the other person's** research question. Be sure you pay attention to what they said!