Toward contrastive (and other types of) explanations in GeoAI

Ben Adams

University of Canterbury

New Zealand

WELLBEING BUDGET 2020

REBUILDING TOGETHER

BUDGET 2020

14 May 2020

StatsNZ 133 wellbeing indicators

- **Environment** (air quality, cities and settlements, climate, water and sanitation, cultural ecosystem services, ...)
- **Cultural** (sense of belonging, language development, ...)
- **Economic** (child poverty, income, …)
- **Social** (corruption, health equity, justice equity, …)
- **Contextual** (migration, population distributions, ...)

StatsNZ 133 wellbeing indicators

- **Environment** (air quality, cities and settlements, climate, water and sanitation, cultural ecosystem services, …)
- **Cultural** (cultural sense of belonging, engagement, …)
- **Economic** (child poverty, income, …)
- **Social** (corruption, health equity, justice equity, …)
- **Contextual** (immigration, population distribution, …)

High-frequency sensors (social, environmental, …)

Heterogeneous spatiotemporal data

Toward a wellbeing sensor network?

- Prime candidate for **GeoAI technologies**
- Indicators will be used for **decision making** around government funding and policy
- Models **must be explainable** to non-technical people

Algorithm Charter for Aotearoa New Zealand

- Transparency
- Partnership
- People
- Data
- Privacy, Ethics and Human Rights
- Human oversight

New Zealand Government

Stats NZ
Tatauranga Aotearoa

https://data.govt.nz/use-data/data-ethics/government-algorithm-transparency-and-accountability/algorithm-charter/

Outline

- Historical context of explanatory AI for geography
- Explainable AI (XAI) and Explanation in AI
- Important types of explanations
- Bringing more explanatory AI into GeoAI

Some context

- “There is a growing and increasingly urgent need for a major new revolution in the provision of smart tools able to make good and optimal use of the geographic information that now exists.”

Some context

- “There is a growing and increasingly urgent need for a major new revolution in the provision of smart tools able to make good and optimal use of the geographic information that now exists.”
- Openshaw & Openshaw, *Artificial Intelligence in Geography*, 1997 (**23 years ago!**)

Some more context

- “AI techniques, if properly applied, should also allow researchers to spend a greater proportion of their time on creative thinking and less on technical drudgery. As with any set of tools, the techniques of AI cannot replace a hard-earned understanding of some phenomenon and will almost certainly be overvalued and misused by some practitioners.”

Some more context

- “AI techniques, if properly applied, should also allow researchers to spend a greater proportion of their time on creative thinking and less on technical drudgery. As with any set of tools, the techniques of AI cannot replace a hard-earned understanding of some phenomenon and will almost certainly be overvalued and misused by some practitioners.”
- Terry Smith, “Artificial intelligence and its applicability to geographical problem solving” 1984 **(36 years ago!)**

3 applications of AI to geographic problem solving (Smith, 1984)

- Explanation

- Engineering

- Teaching

What about GeoAI?

- GeoAI currently driven by deep learning research
- Most work to date falls in the category of **Engineering**
- Is prediction is enough?
- Discovery of explanatory models (Gahegan 2020)

Gahegan, M. (2020). Fourth paradigm GIScience? Prospects for automated discovery and explanation from data. *International Journal of Geographical Information Science*, *34*(1), 1-21.

Different types of explanatory AI

- Explanation a la Smith (1984) and Gahegan (2020)
- Explainable AI (XAI) (Biran & Cotton 2017)
- Explanation in AI (Miller 2019)

Explanation- 4 Key Findings (Miller 2019)

- **Explanations are contrastive**
- **Explanations are selected**
- **Probabilities probably don’t matter**
- **Explanations are social**

- *Explanations are contextual.*

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, *267*, 1-38.

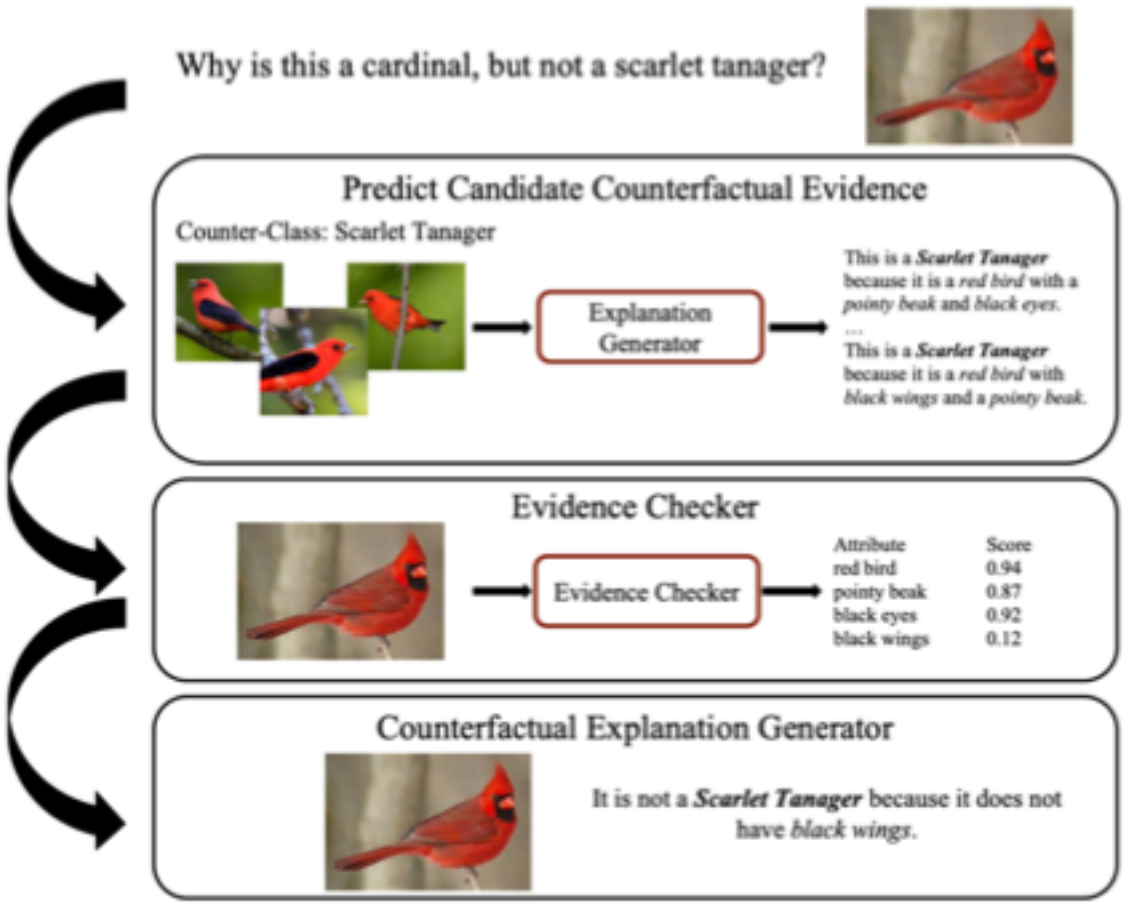
Contrastive explanation

- Explaining the cause of an event relative to some other event.

- Why event P (**fact**) happened instead of some event Q (**foil**).

- **These are counterfactual *outcomes*, not *causes*.**

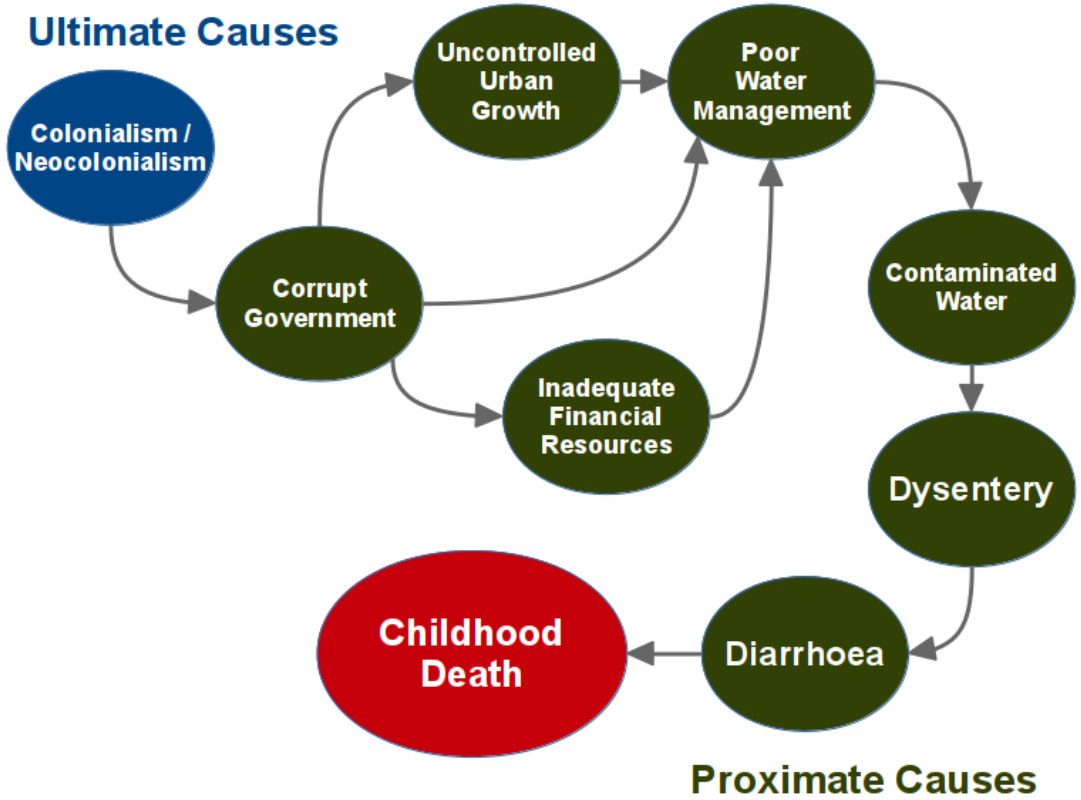
Counterfactual (contrastive) explanations



Hendricks, L.A., et al. 2018 “Generating Counterfactual Explanations with Natural Language“

4 types of explanatory questions
(Van Bouwel and Weber, 2002)

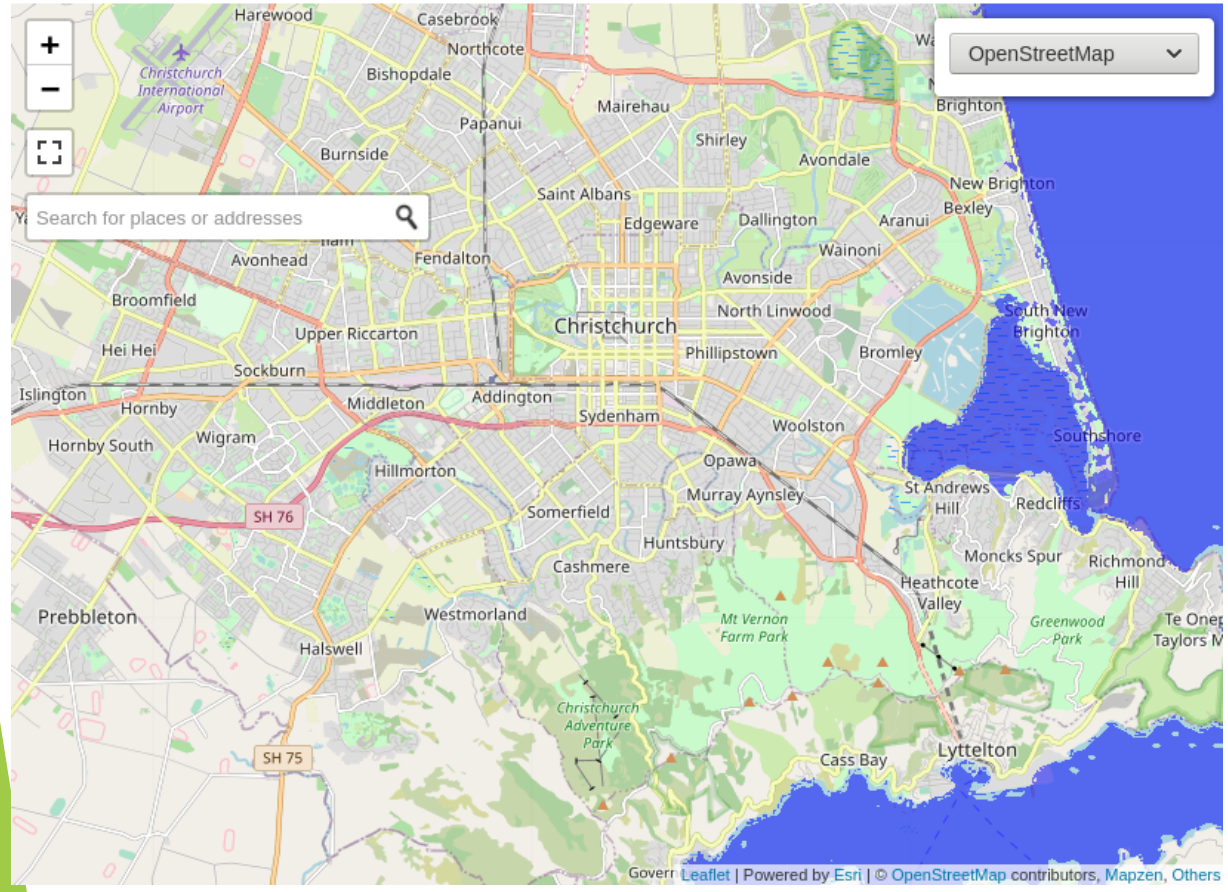
- Type 1, **Plain fact**: Why does object *a* have property *P*?

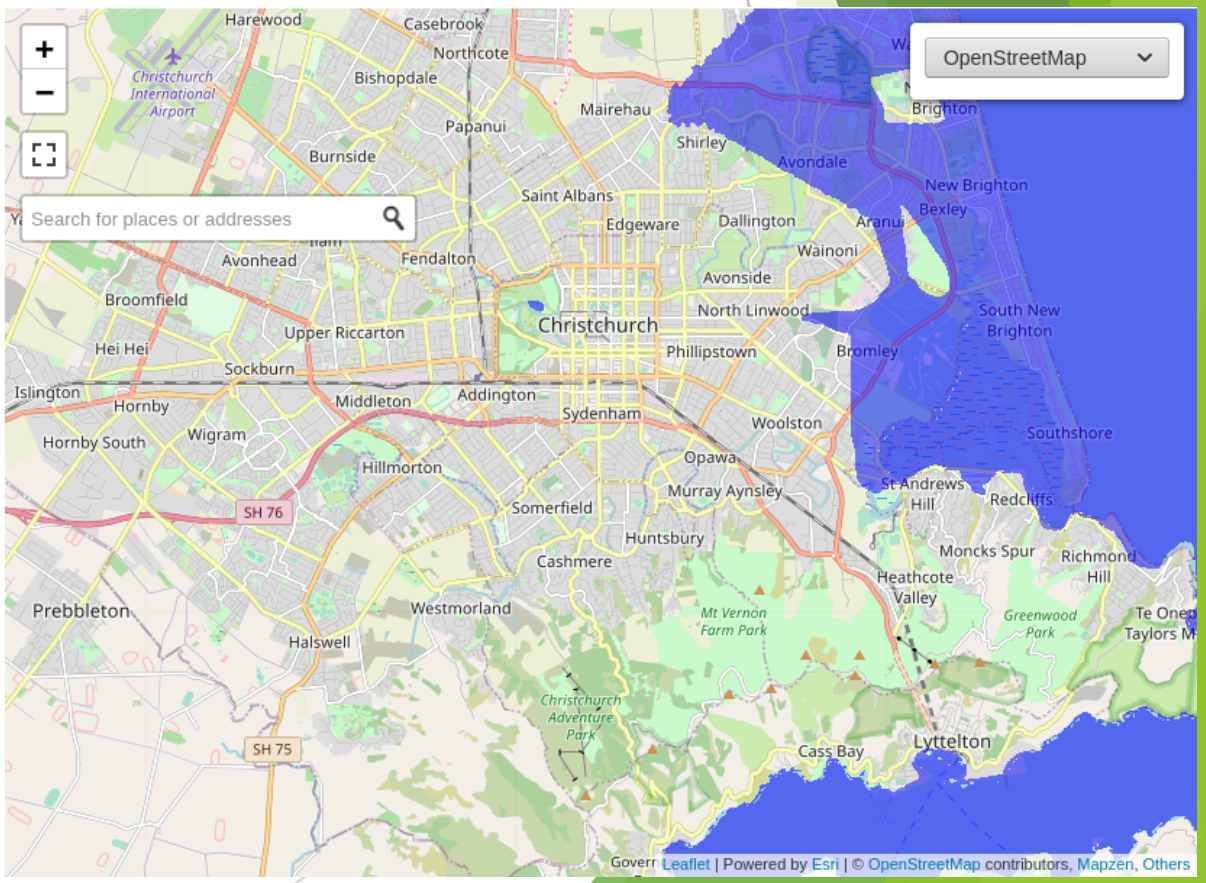


Van Bouwel, J., & Weber, E. (2002). Remote causes, bad explanations?. *J. for the Th. of Soc. Beh.*, *32*(4), 437-449.
Graph from http://michaelminn.net/tutorials/correlation/

4 types of explanatory questions
(Van Bouwel and Weber, 2002)

- Type 2, **P-contrast: Why does object *a* have property *P*, rather than property *Q*?**





4 types of explanatory questions
(Van Bouwel and Weber, 2002)

- Type 3, **O-contrast: Why does object *a* have property *P*, while object *b* has property *Q*?**

4 types of explanatory questions
(Van Bouwel and Weber, 2002)

- Type 4, **T-contrast: Why does object *a* have property *P* at time *t*, but property *Q* at time *t'*?**

Benefits of contrastive explanation

- Easier to generate than complete explanations
- Lay people find them more intuitive
- Pitched at the appropriate ‘level of explanation’

Different types of explanations

- **Functional explanations**
  - Phenomena that have dependence relations
  - Derived from *functions* or *goals*
- **Mechanistic explanations**
  - Physical phenomena
  - Derived from *parts* or *processes*
- **Geographic phenomena can have both!**

T. Lombrozo (2010) Causal-explanatory pluralism: how intentions, functions, and mechanisms influence causal ascriptions. *Cogn. Psychol.* 61(4), 303-332.

Workshop submissions

- *Spatially-explicit population projections*
- *Forecasting criminogenic environments*
- *Greenspace and academic performance*
- *Travel behavior*

- **All extremely relevant to measuring wellbeing!**

Workshop submissions

- *Spatially-explicit population projections*
- *Forecasting criminogenic environments*
- *Greenspace and academic performance*
- *Travel behavior*

- Emphasize predictive capability.
- Discuss explanation by looking at counterfactual causes, not effects.
- *But why do the models make the predictions that they do?*

What could contrastive explanations look like for **Population projection**?

- The paper describes some *contrasts in the cause*:

"Do housing choices differ between migrants and native-born?"

- Answered by looking at the output of the model
- Contrasts in the effect help to understand *how the model works*.

"Why does the model show migrants settling in one neighborhood but not another?"

What could contrastive explanations look like for **Forecasting criminogenic environments**?

- The paper says the machine learning may “generate more accurate forecasts than more traditional statistical models”
- Also, focuses on contrasts in the cause:

“… generate a counterfactual for what would have occurred in the absence of COVID-19 and the associated stringencies.”

- Contrasts in the effect, e.g.:

“Why does the model show an increase in crime in one environment but not another in the absence of COVID-19?”

Explanation **selection**

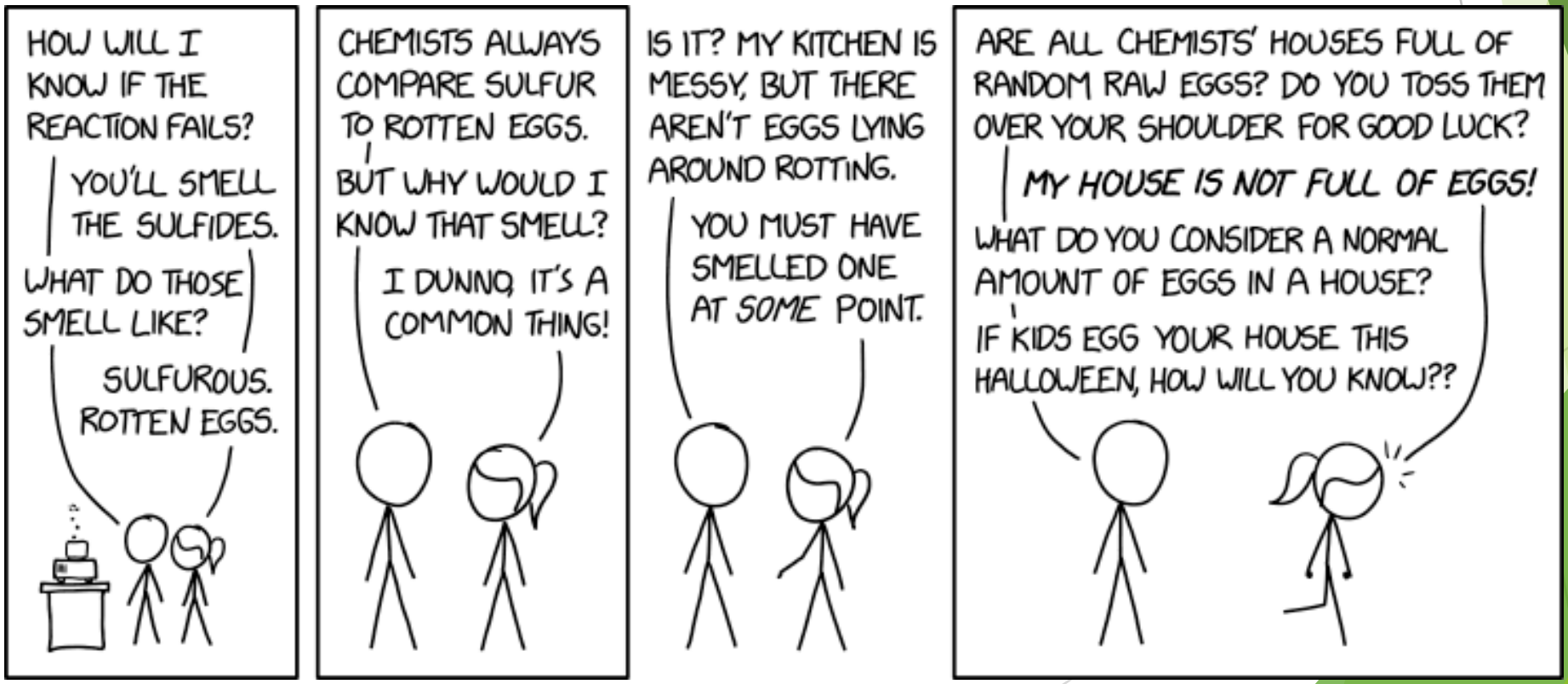
- From the **many possible causes** of an event how is one selected as ***the* explanation**?
  - Simulating counterfactuals useful to derive an explanation.
  - What events to mutate when simulating?
  - Need biases (biases can be good!)

- How do we **evaluate the explanation** that is given?

B.F. Malle (2004) *How the Mind Explains Behavior: Folk Explanations, Meaning, and Social Interactions*, MIT Press.

Social explanation

- Conversational GeoAI agents
- Interactive explanation



Questions to ask ourselves about GeoAI

- Are we **evaluating contributions** in GeoAI in the best way?
  - Is machine learning **model evaluation sufficient**?
  - **Other standards for success** based on model usability?

- Are we starting with **geographic problems** needing to be solved and thinking about the **everyday use of the model**?

- **Who needs to understand** the models we are building and why?

More questions

- **What kinds of explanations** do people need from the models we build? (functional, mechanistic, etc.)
- **What are the biases** that we should use to select counterfactuals?
- **What kinds of interaction** is most useful in explanatory AI systems for Geography?

Conclusions

- GeoAI is not a new field
- Wellbeing sensor network as a grand challenge
- Start with who will be making decisions based on the model
- Creating tools that *explain* GeoAI models to those users
  - Use contrastive explanations
  - Understand selective bias to find relevant explanations
  - Utilize interaction and conversational modes

Closing quote

- “As a geographer, your interest in AI should be purely to serve your geographical concerns. … If you become a general expert in AI and forget all about your geography, then you will almost certainly fail to do anything useful in a geographical context with your AI skills.”
  - Openshaw & Openshaw 1997

Thank you!

Questions? / thoughts

(or explanations?)