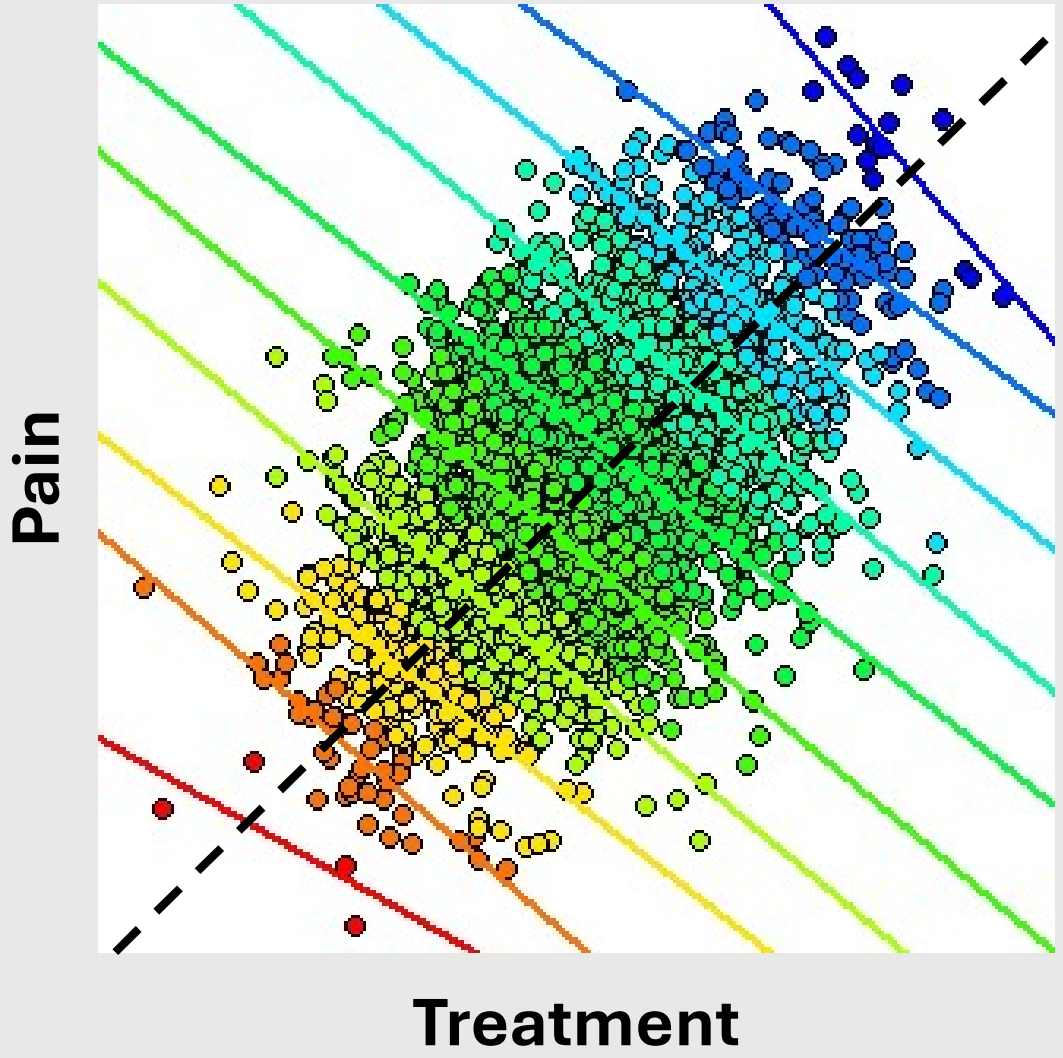
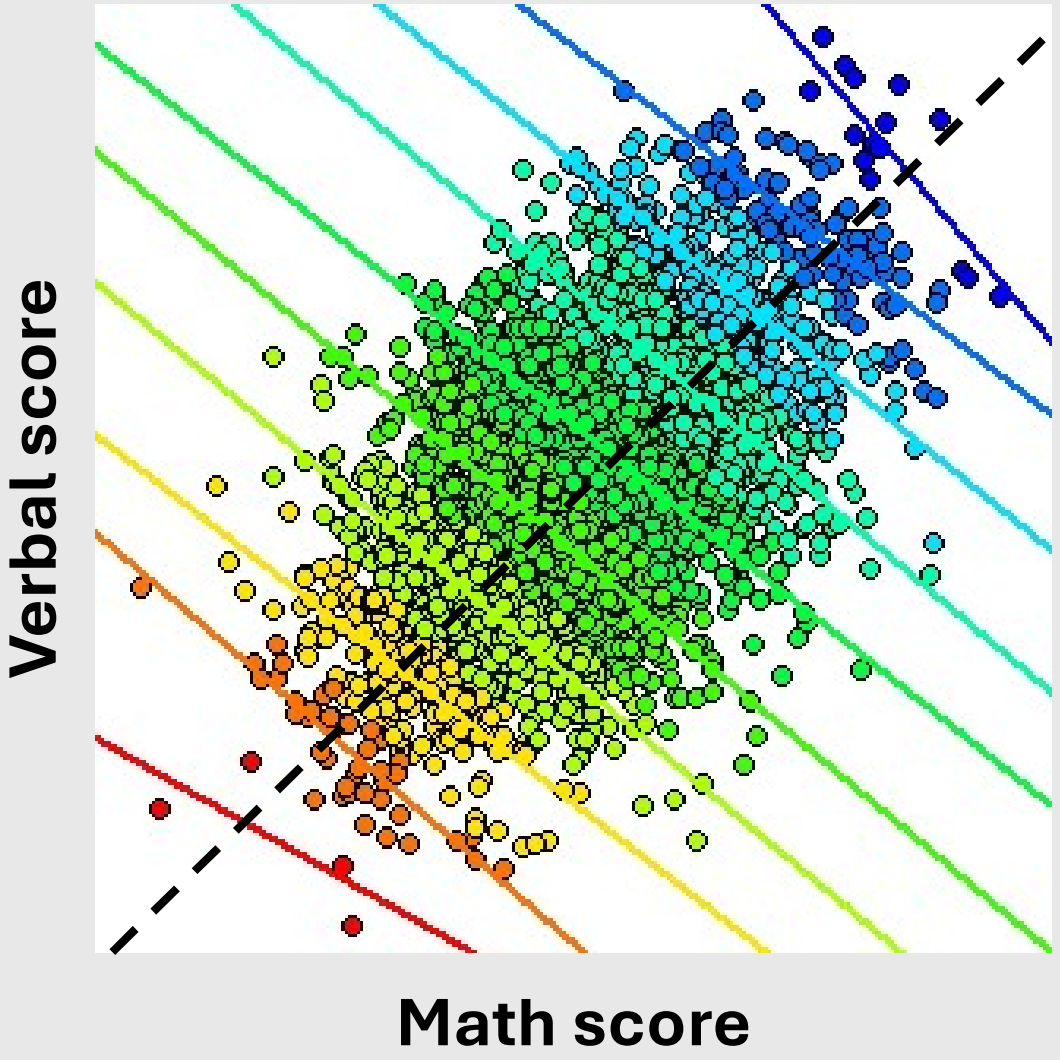
Multi-Stage Simpson's Paradox Machine

Bob Horton
cybertory@gmail.com
https://github.com/rmhorton/multistage_simpson



Color by disease severity



Color by school

The American Statistician, 68(1):8-13, February 2014.

TECHNICAL REPORT
R-414

Comment: Understanding Simpson's Paradox

Judea PEARL

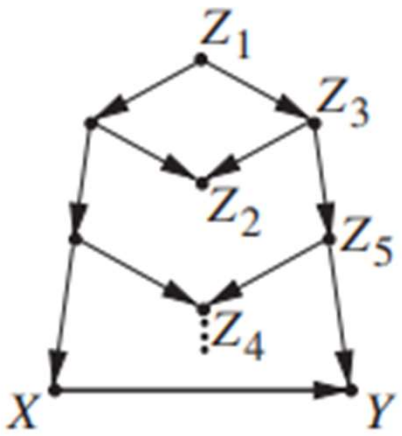
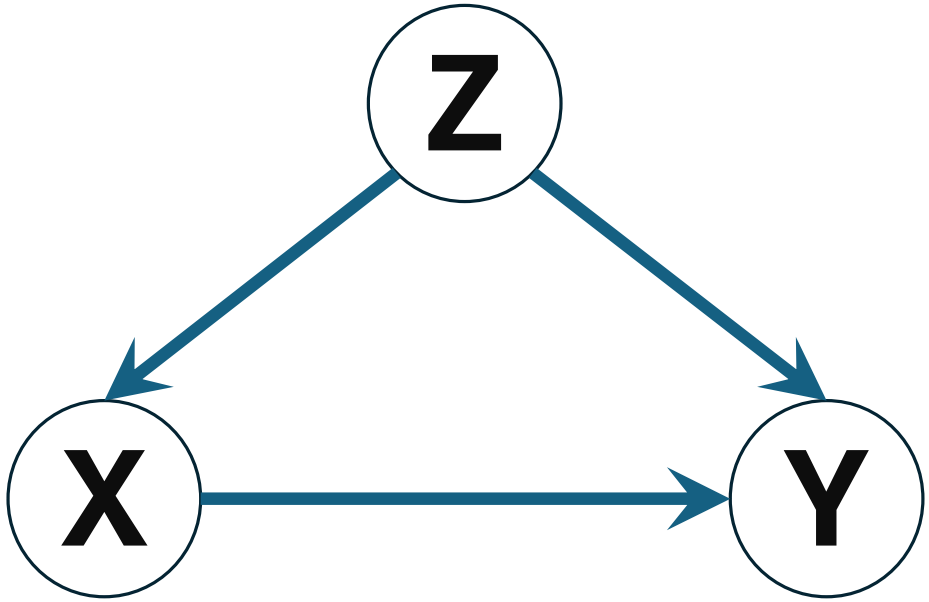
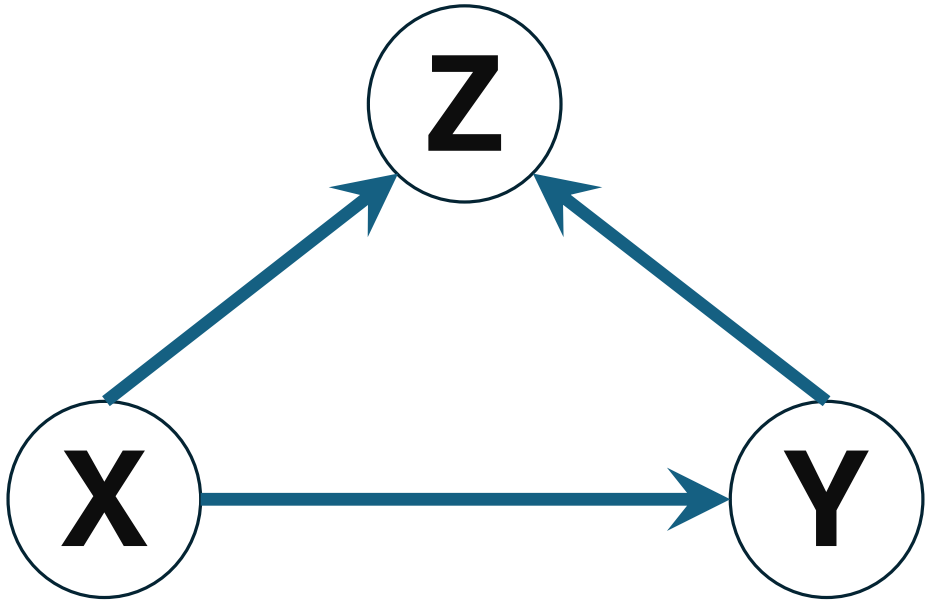


Figure 3. A multistage Simpson's paradox machine. Commulative conditioning in the order (Z_1, Z_2, Z_3, Z_4, Z_5) creates reversal at each stage, with the correct answers alternating between disaggregated and aggregated data.

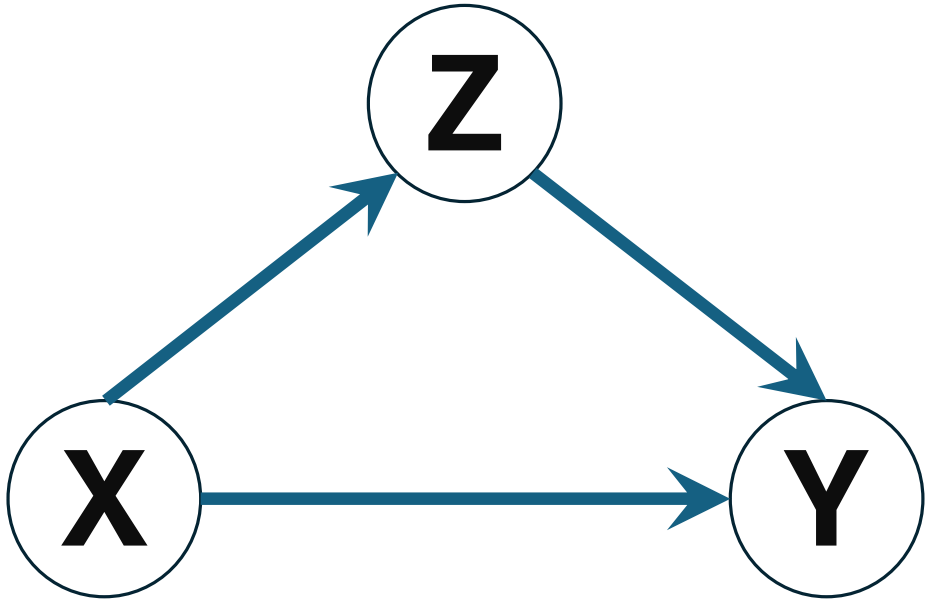
Confounder



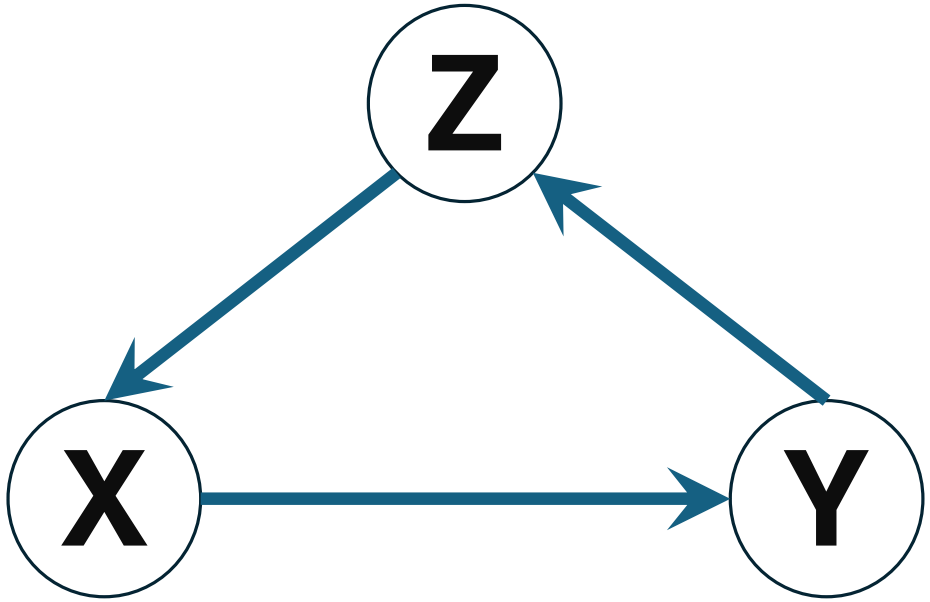
Collider



Mediator

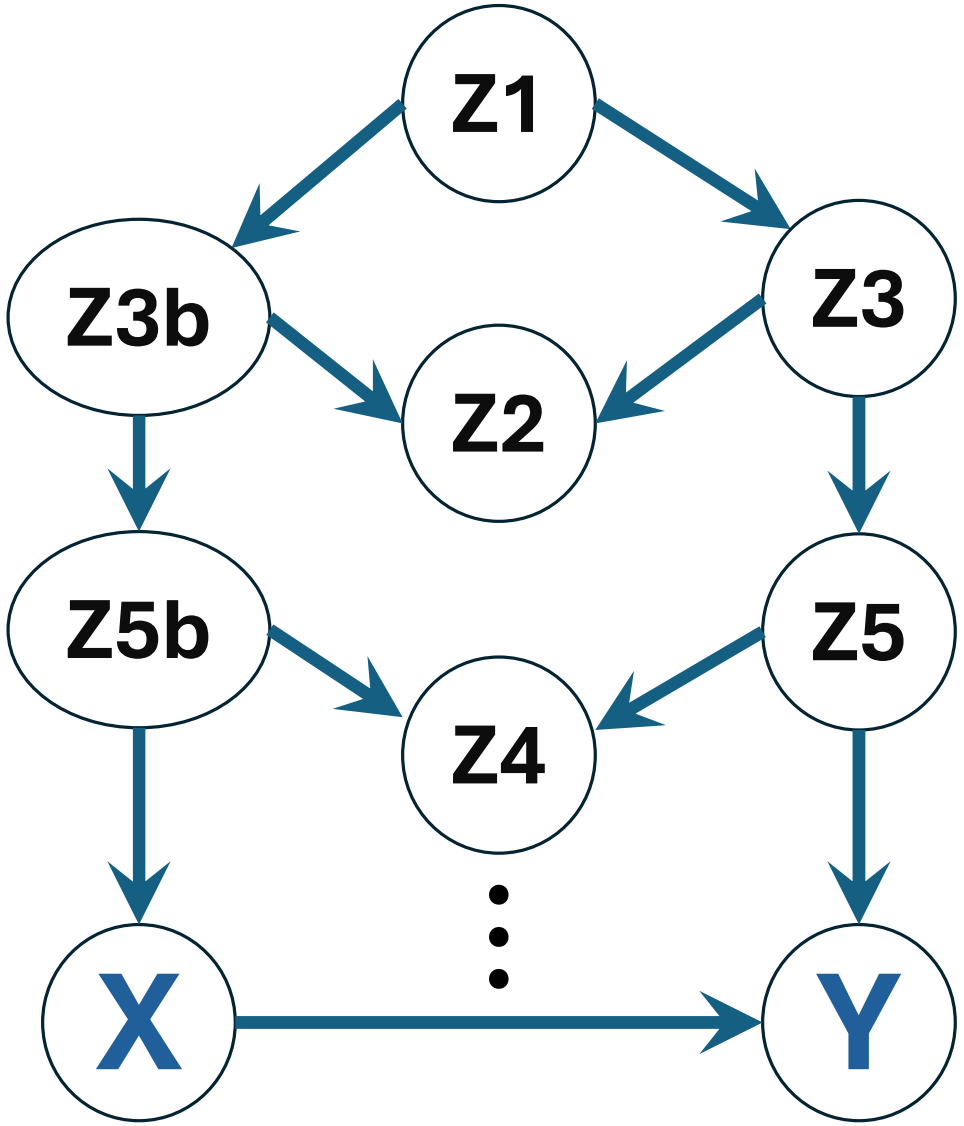


Loop



We can't handle loops (think temporal paradoxes in Star Trek). Sometimes we can remove them by unrolling in time, but that is a different topic.

Multi-Stage Simpson's Paradox Machine



Demo

Feedback

Did the firehose approach work?

- **Yes**: I'm going home with lots to think about
- **Maybe**: I don't know yet
- **No**: it just confused everything

What next?

EBMs

- Analyze Python objects in R
  - extract parameters
  - smooth the curves
  - predict in SQL

Causal-model based simulations

- Temporal unrolling
- Incorporating domain knowledge
  - causal graph
  - Extrapolation