Get to the Bottom
Causal Analysis for User Modeling

Shi Zong

The Ohio State University

Joint work with Branislav Kveton, Shlomo Berkovsky, Azin Ashkan, Zheng Wen

UMAP 2017, Bratislava, Slovakia

Collaborators

Branislav Kveton
Adobe Research

Shlomo Berkovsky
CSIRO

Azin Ashkan
Google Inc.

Zheng Wen
Adobe Research

Motivation

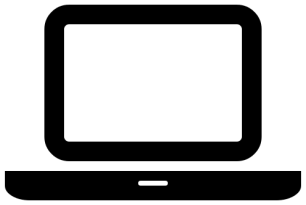
- Weather affects our mood and behavior

Sunny

- Higher temperatures bring a depressed person up
- People tend to go outside when it is sunny

Rainy

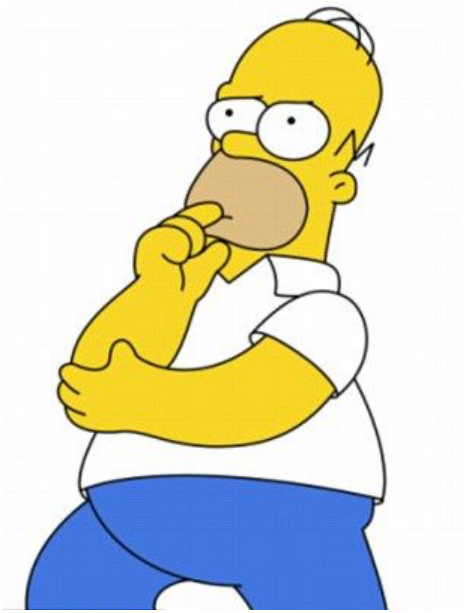
- A lack of sunlight could make you sad (Seasonal Affective Disorder)
- Rain can cause pain



Causal Analysis – An Example

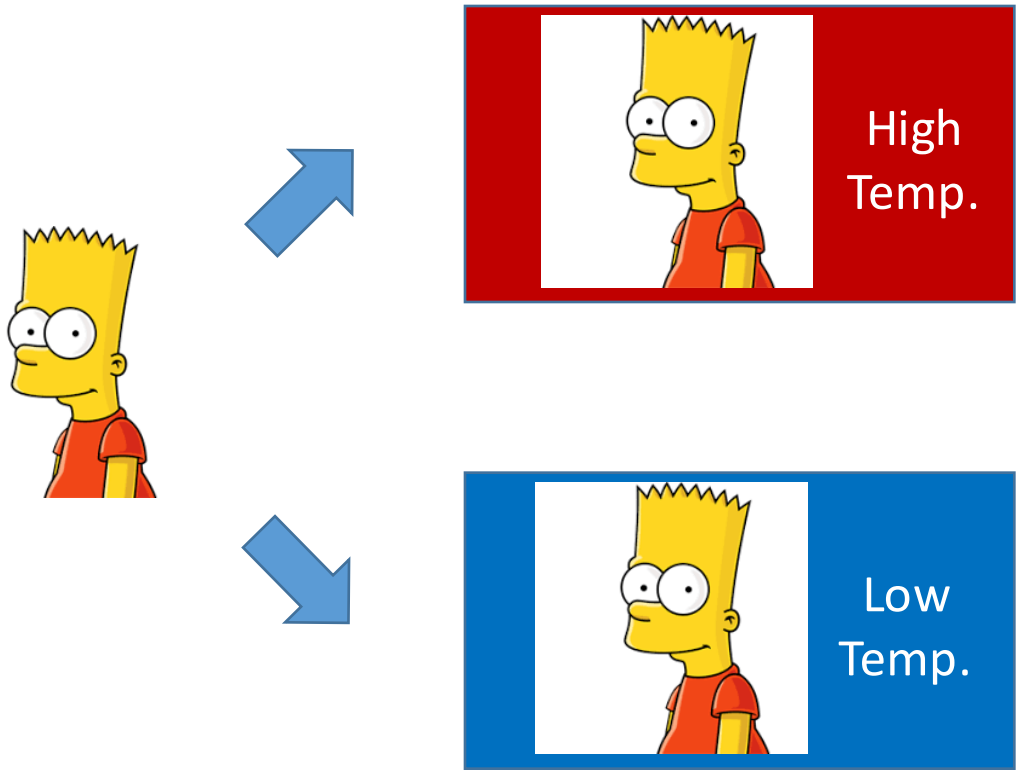
Goal: Study relations between weather and user TV watching patterns

Question: Does high temperature cause watching more Drama?



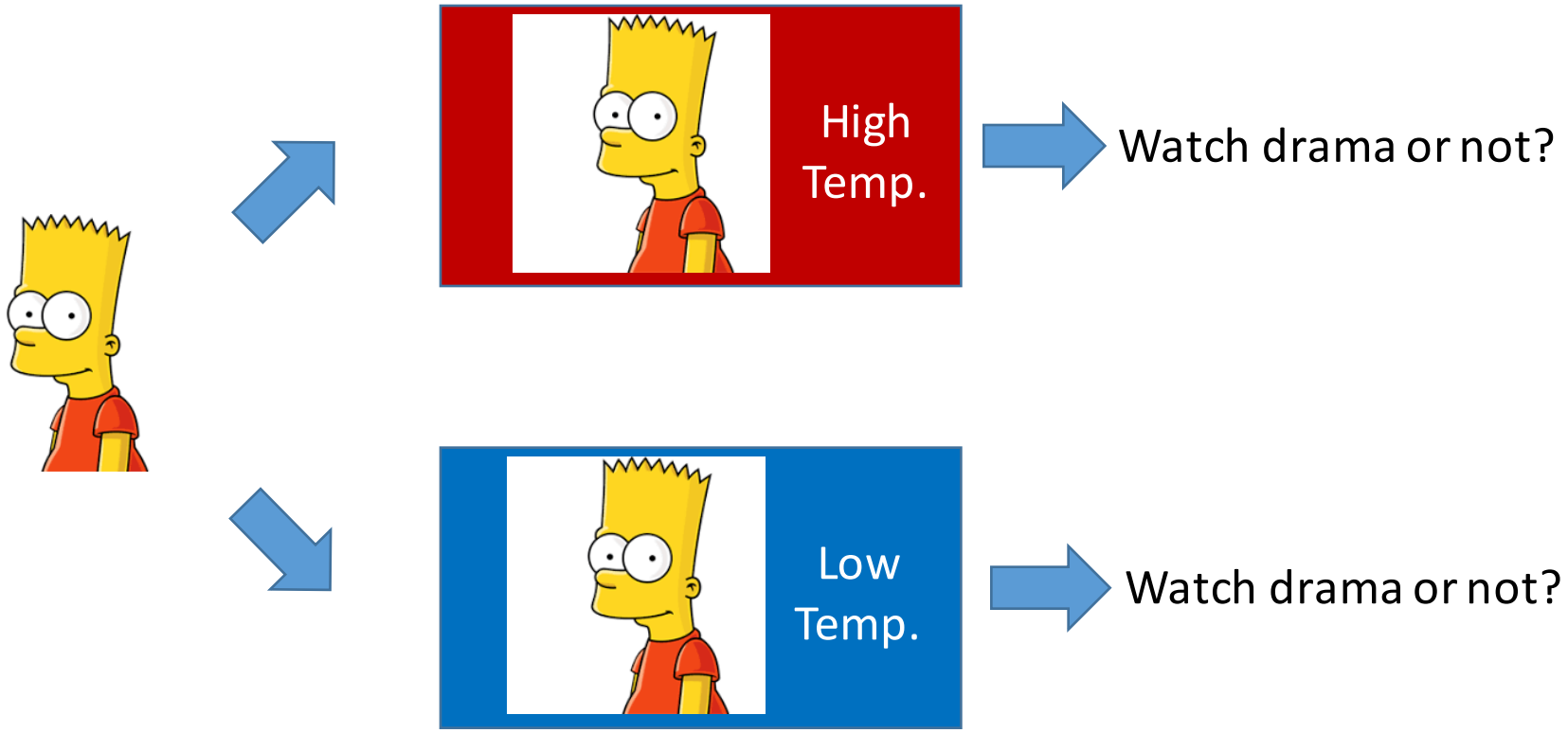
Causal Analysis – An Example

- Does high temperature cause watching more Drama?



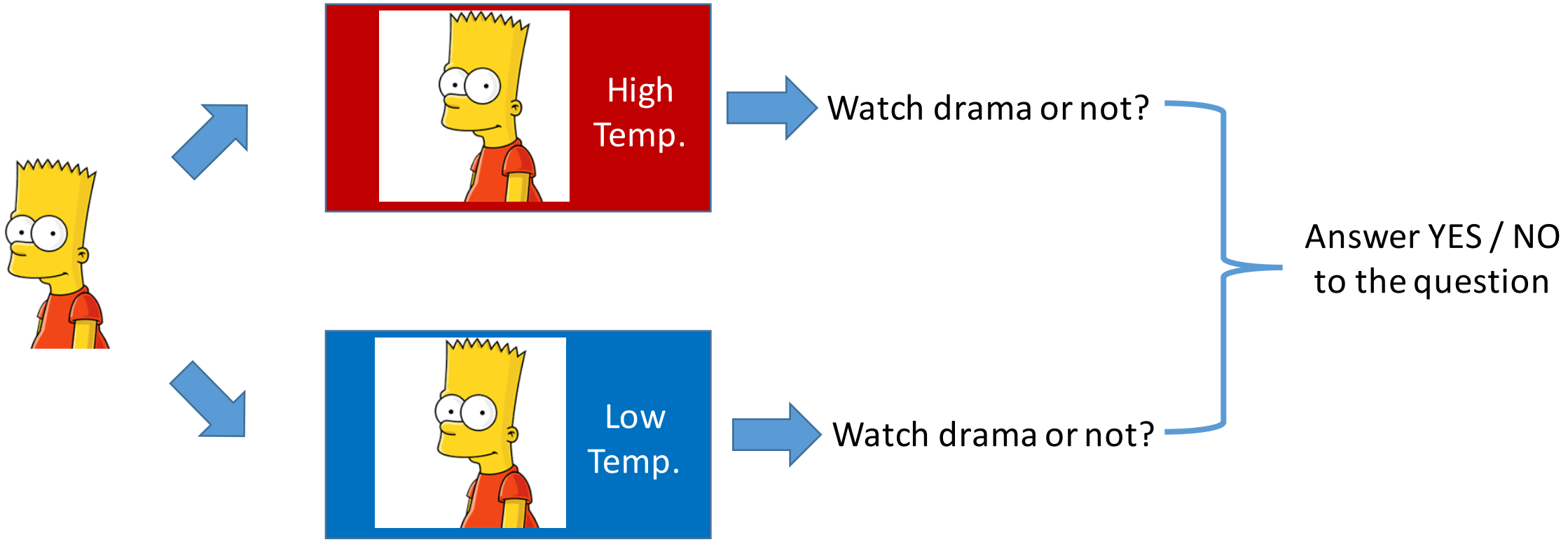
Causal Analysis – An Example

- Does high temperature cause watching more Drama?



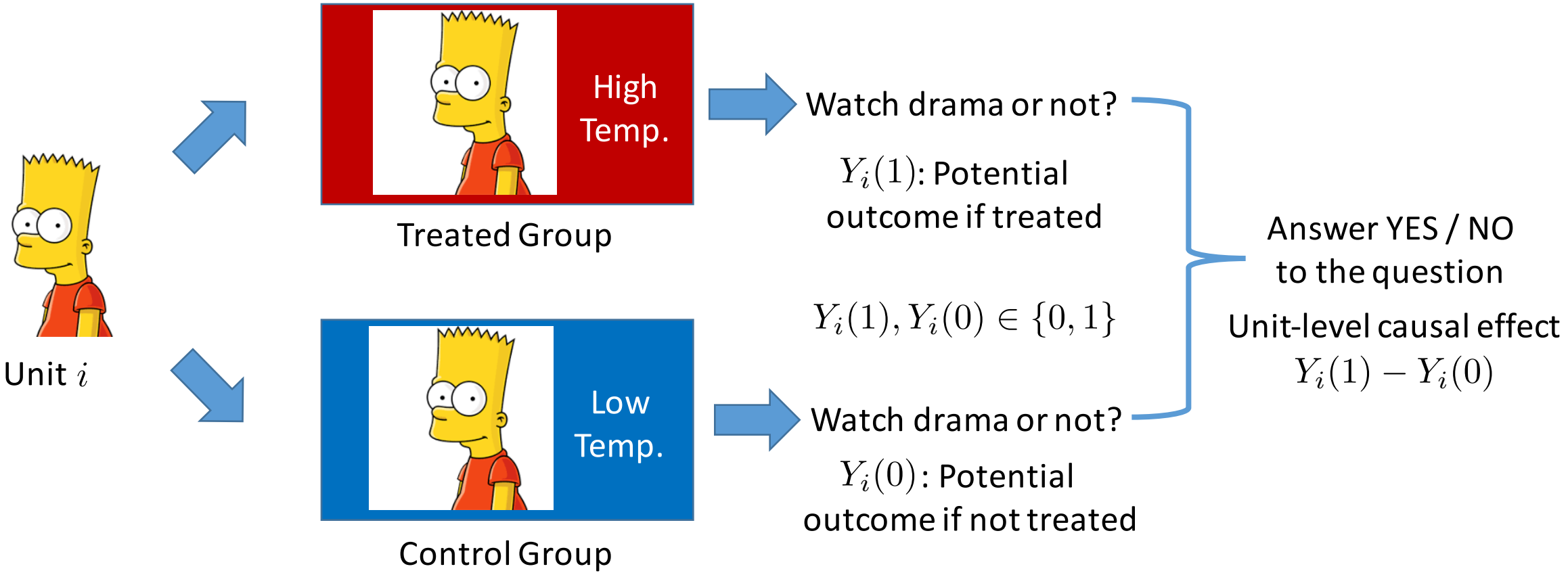
Causal Analysis – An Example

- Does high temperature cause watching more Drama?



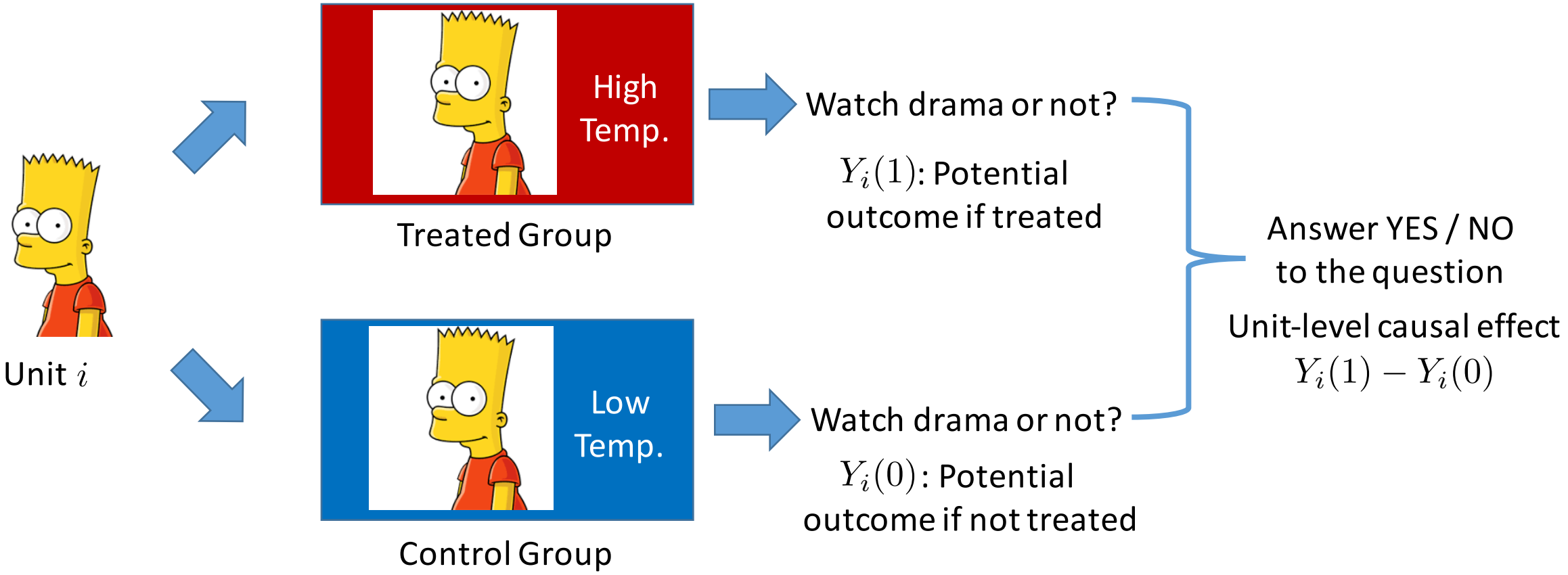
Causal Analysis – Some Terminologies

- Does high temperature cause watching more Drama?



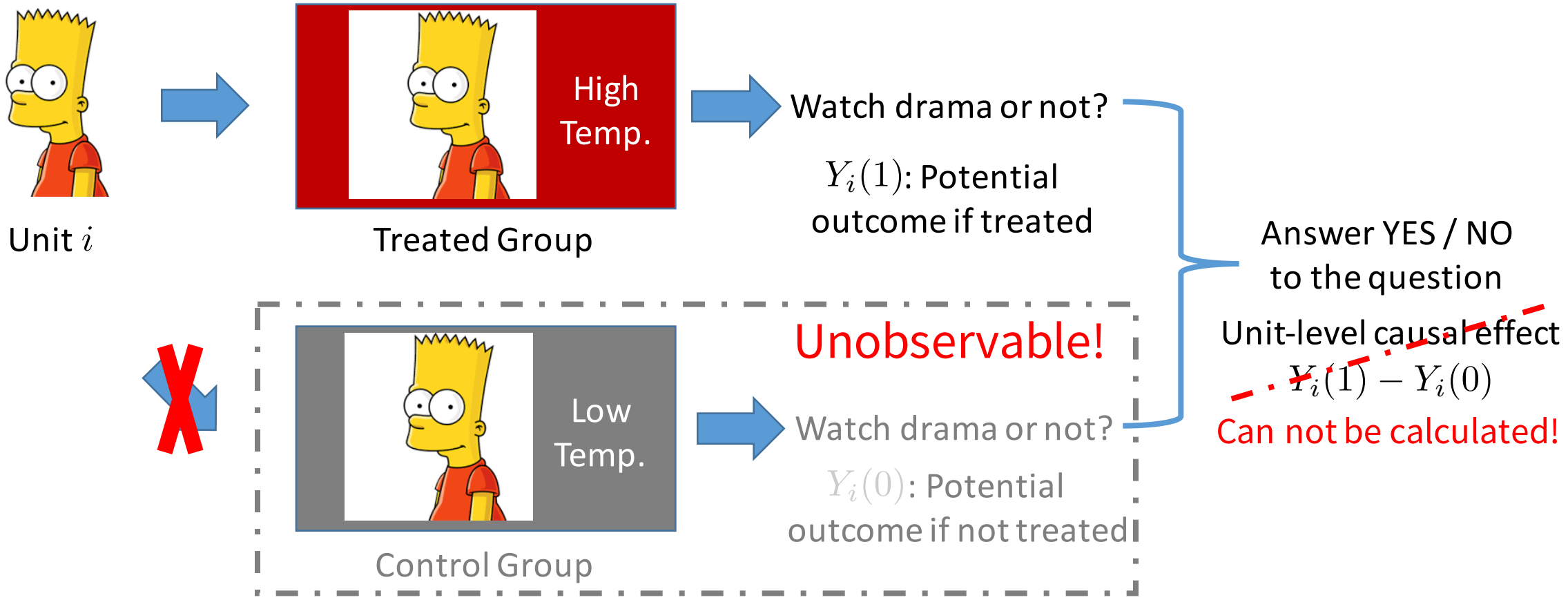
Causal Analysis – Main Problem

- Missing Counterfactuals
  - Individual can't be observed with and without treatment at the same time



Causal Analysis – Main Problem

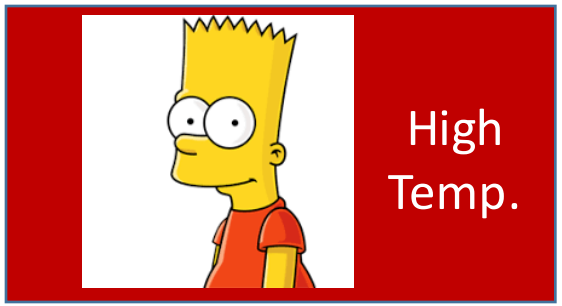
- Missing Counterfactuals
  - Individual can't be observed with and without treatment at the same time



Solution: Find a Similar "Control" Person!

- Matching treated units with units that are similar to treated units (matching on covariates)

Unit i

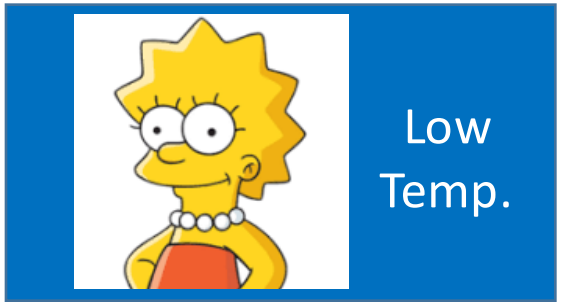


Treated Group

Watch drama or not?

$Y_i(1)$: Potential outcome if treated

Matched Unit $\pi(i)$



Control Group

Watch drama or not?

$Y_{\pi(i)}(0)$: Potential outcome if not treated

Answer YES / NO to the question

Unit-level causal effect

$Y_i(1) - Y_{\pi(i)}(0)$

Measuring Group Effect of Treatments

- Average treatment effect on treated (ATT)

Unit i

Unit-level causal effect

$$Y_i(1) - Y_{\pi(i)}(0)$$

Potential outcome if treated

All treated units

$$i : T_i = 1$$

ATT

$$\frac{1}{n_T} \sum_{i:T_i=1} \left(Y_i(1) - Y_{\pi(i)}(0)\right)$$

Potential outcome if not treated

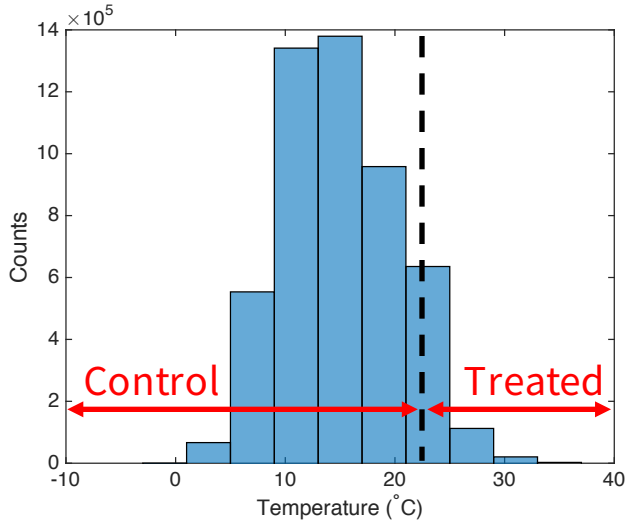
Number of treated units

- Interpretation
  - Expected increase/decrease in the frequency of watching due to treatment

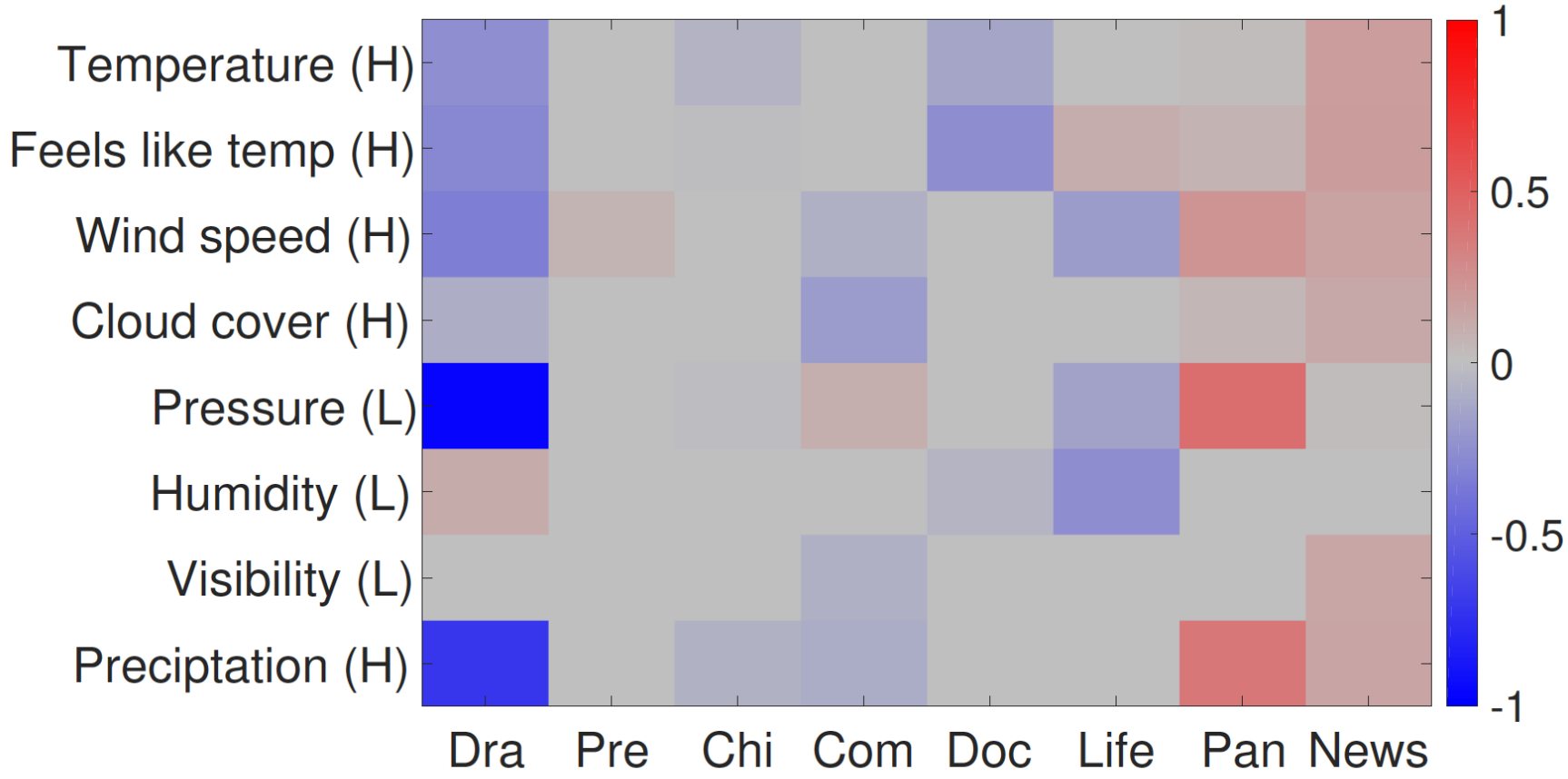
Value	Meaning
Significantly larger/smaller than 0	Significant positive/negative causal effects observed
Near 0	No causal effects observed

Our Modeling Assumptions

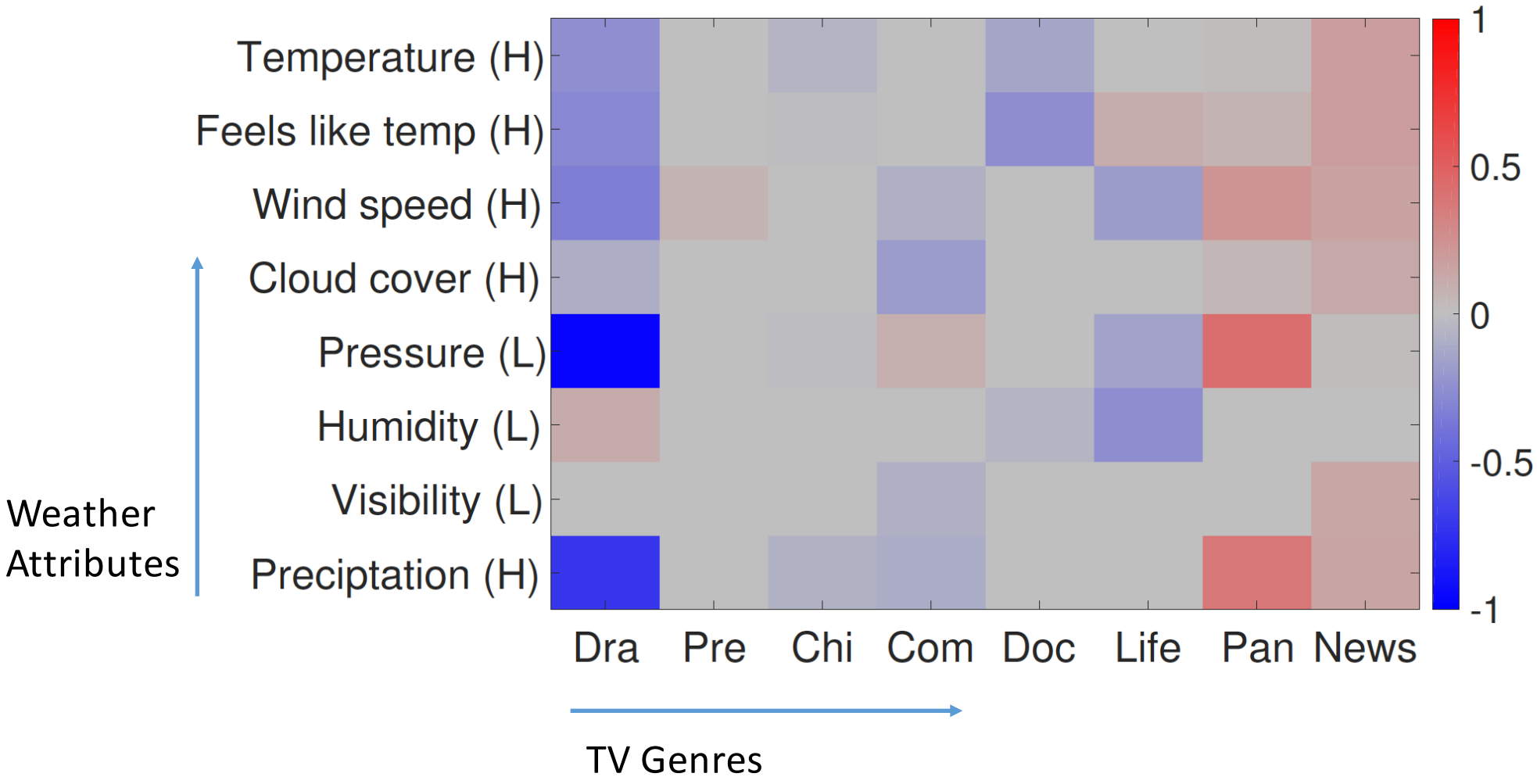
- Matching on covariates (measuring similarity)
  - A popular method is *Nearest-Neighbor Matching* (NNM)
  - In our work, covariates are chosen as *location*, *time* and *user preferences*
- Building treated group and control group
  - Treated group: top 20% of values for a specific attribute
  - Control group: others



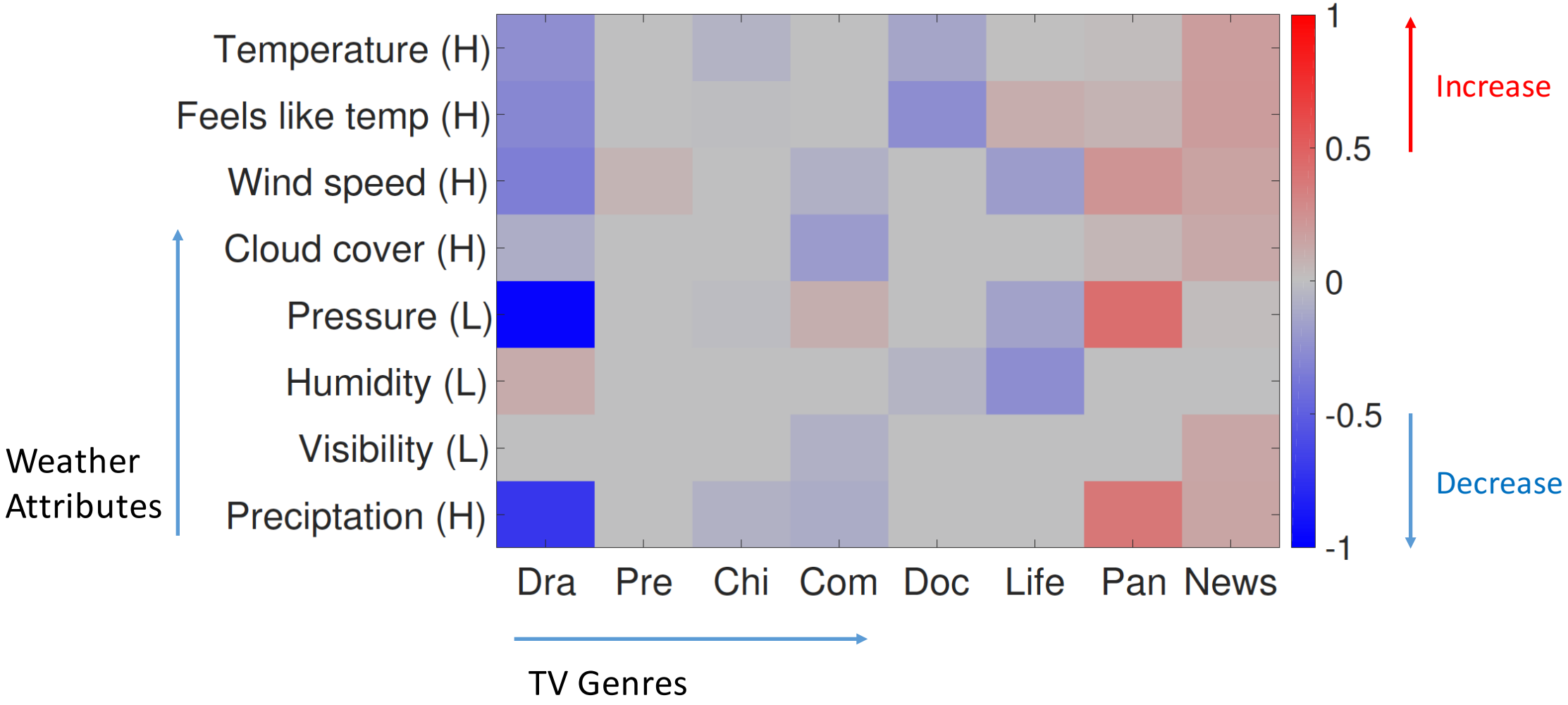
Experimental Results – Whole Population (~2M treated units)



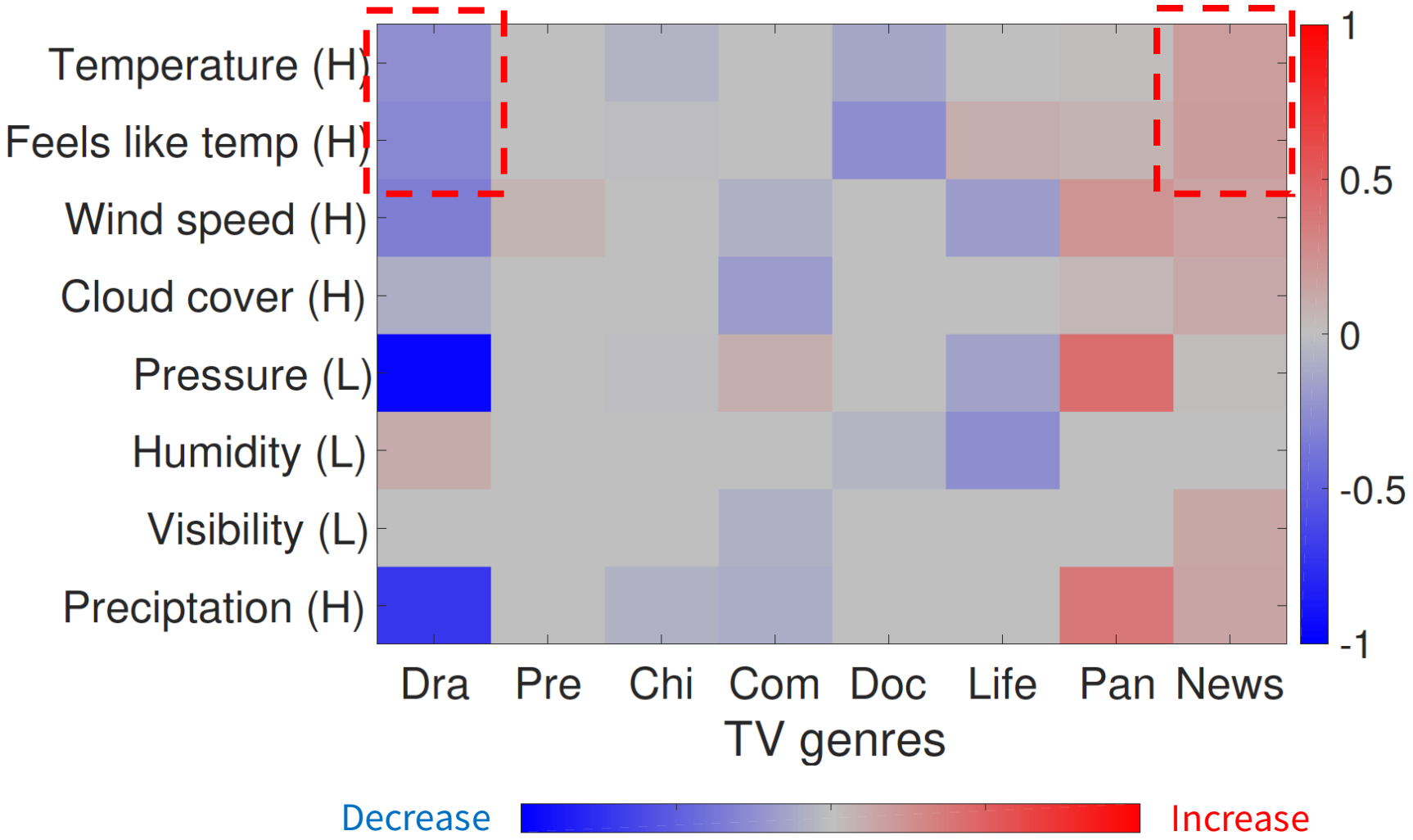
Experimental Results – Whole Population (~2M treated units)



Experimental Results – Whole Population (~2M treated units)

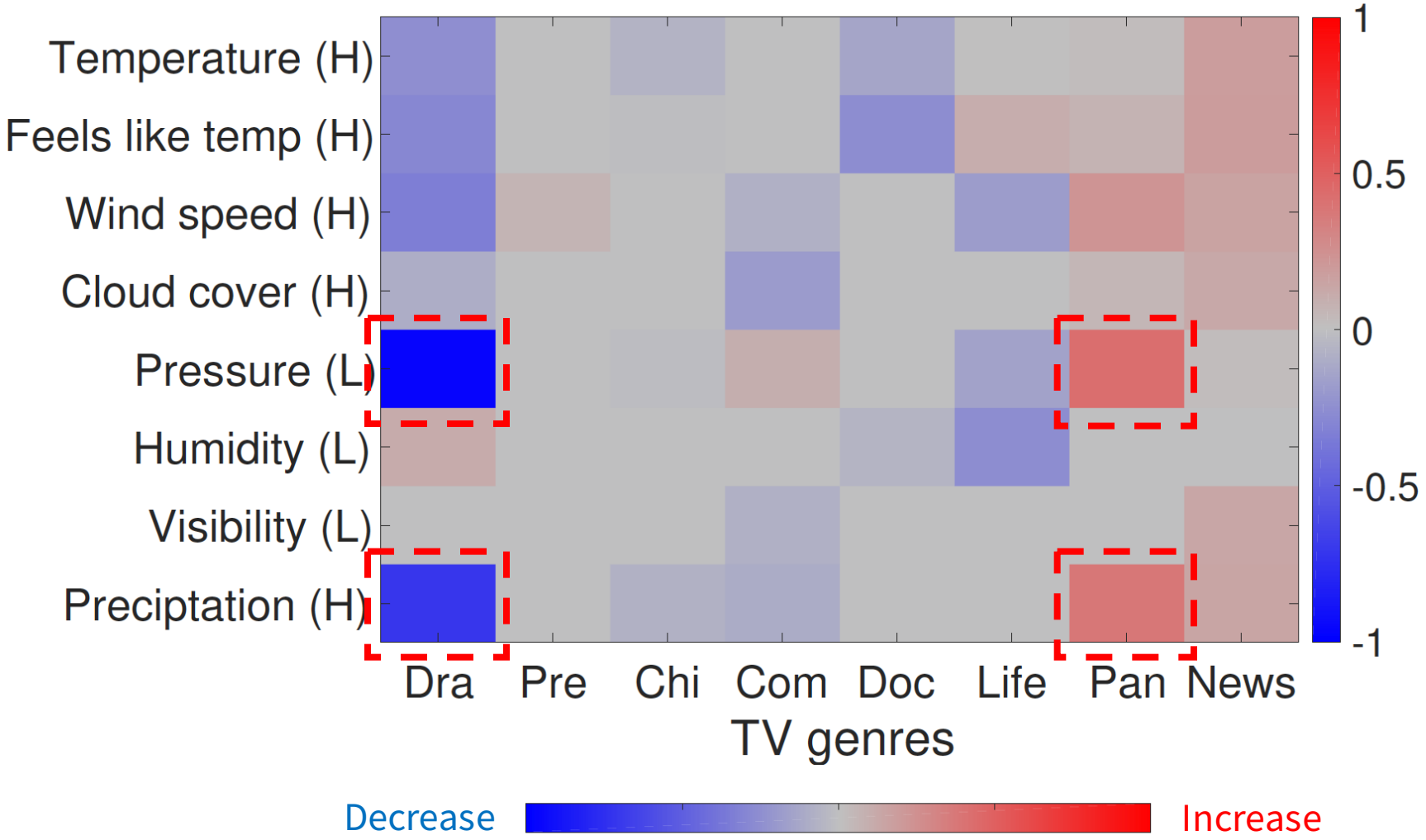


Experimental Results – Whole Population (~2M treated units)



- When it is hot …
  - People prefer watching *News*, rather than *Drama*

Experimental Results – Whole Population (~2M treated units)



- When it is hot …
  - People prefer watching *News*, rather than *Drama*
- When it is rainy …
  - People tend to watch *Panels*, instead of *Drama*

Conclusion

- Take-home message
  - Observe causal relations between weather and user TV watching behavior
  - Next-generation recommender system design may take weather into consideration
- Contribution
  - First large-scale formal causal analysis
    -- Some weather attributes cause significant changes in TV watching patterns
  - User modeling based on causal analysis
    -- We compare between different levels of user granularity and different types of users

- Thanks!

Dataset

- Newly-built large-scale dataset of Australia watching events in 2012
  - 1,296,392 household level participants
  - 21,406,768 records of people watching events
  - 8 attributes of weather
- Rich information of user behavior & weather
  - User behaviors: date, anonym IP address, watching video ID
  - Weather attributes: temperature, feels like temperature, wind speed, cloud cover, pressure, humidity, visibility, precipitation