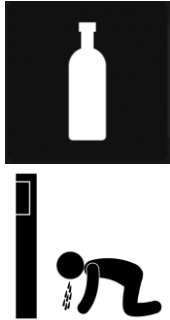# Mining Google Trends for Entertainment Analytics

Qing Wu,
Principal Economist, Google
Los Angeles, May 23, 2019

Google Trends: Vodka and Hangover
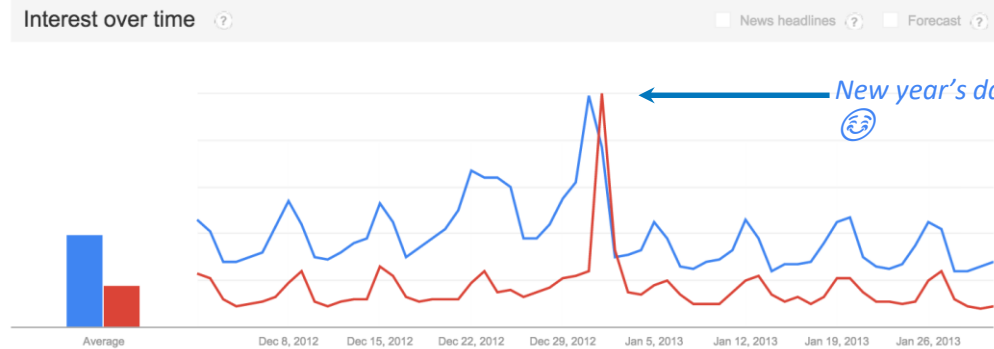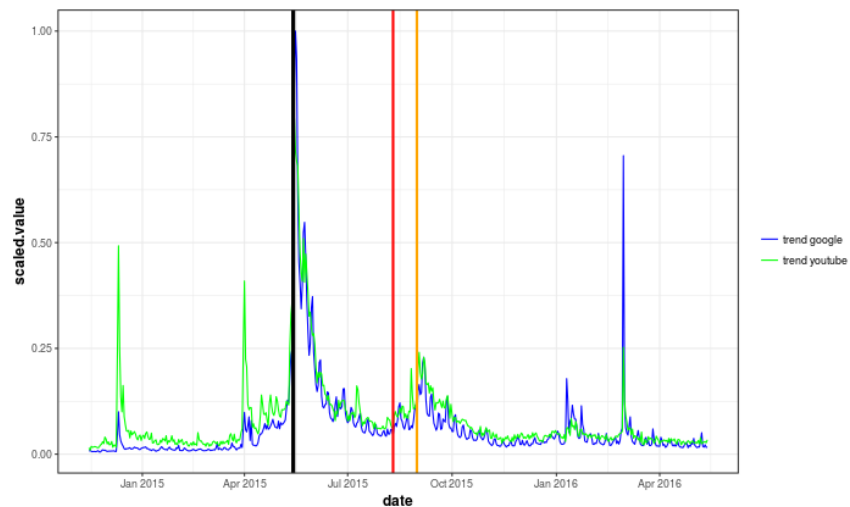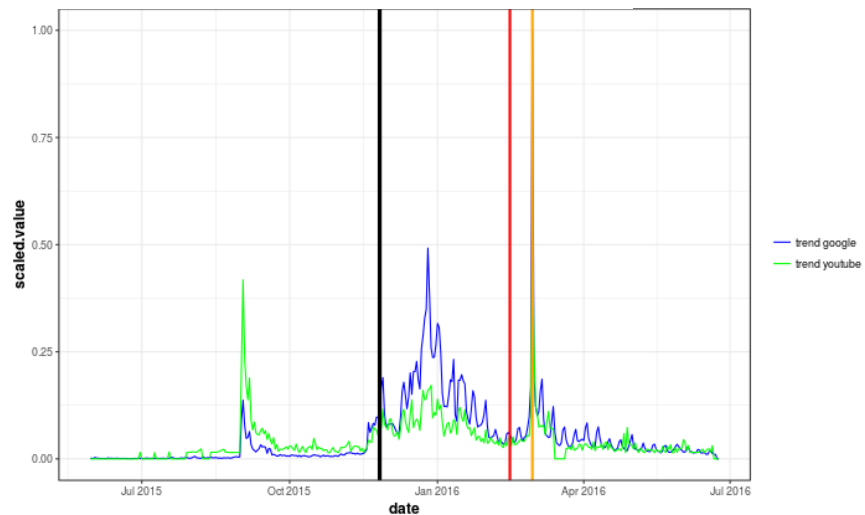

Google

1. Consumer Insights

2. Predictions

3. Causal inferences

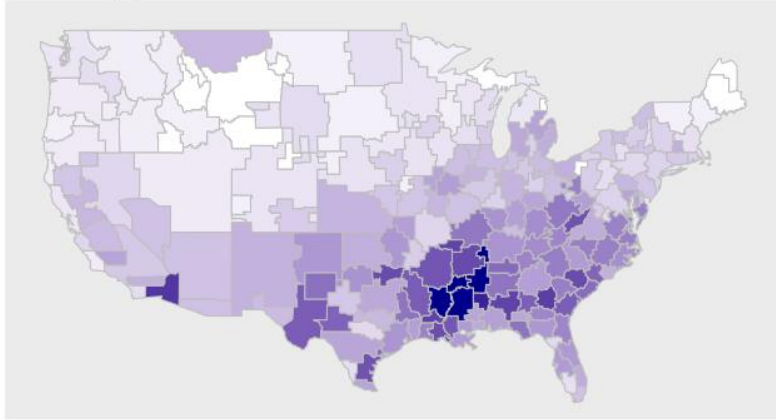# Google and Youtube Search Trends for Movies

Mad Max: Fury Road

Danish Girl

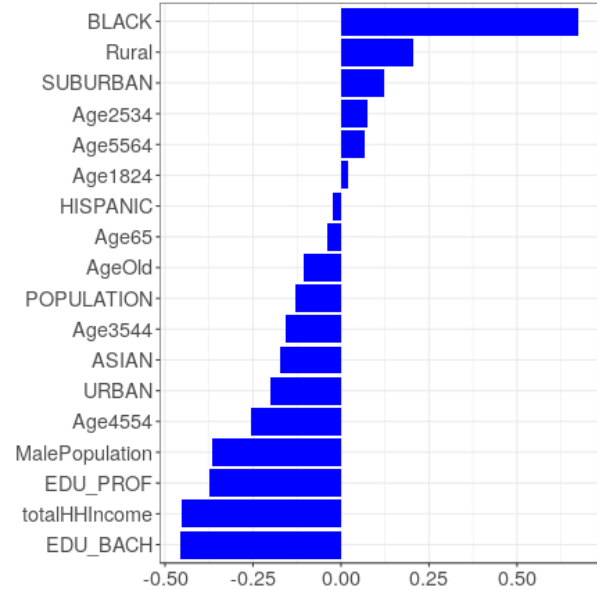# Geo Variances: Madea Family Funeral

search intensity by dma
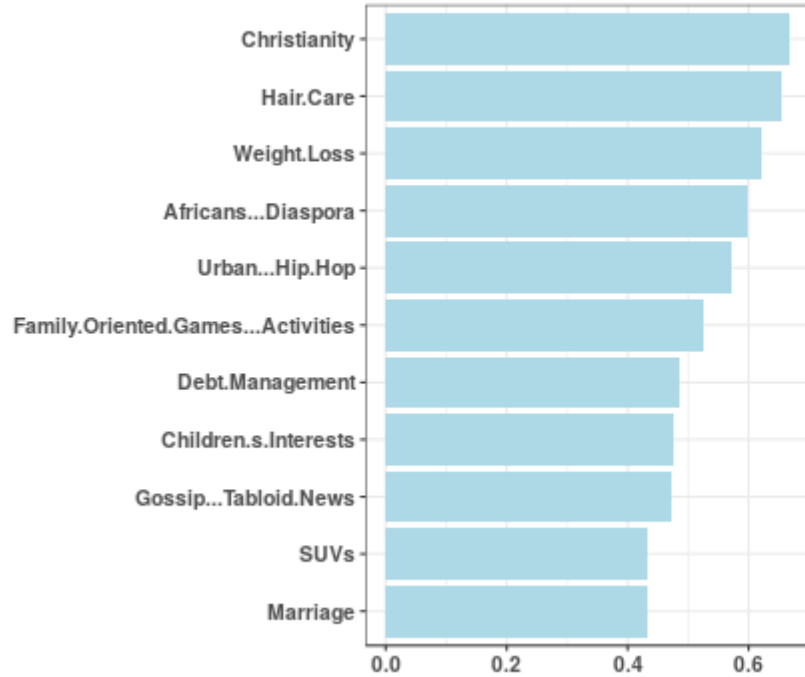

Correlation with demographics

Google

Positively Correlated

Christianity	
Hair.Care	
Weight.Loss	
Africans...Diaspora	
Urban...Hip.Hop	
Family.Oriented.Games...Activities	
Debt.Management	
Children.s.Interests	
Gossip...Tabloid.News	
SUVs	
Marriage	

0.0   0.2   0.4   0.6

Negatively Correlated

Bicycles...Accessories	
Yoga...Pilates	
Investing	
Buddhism	
Beer	
Green.Living...Environmental.Issues	
Wine	
World.News	
Hiking...Camping	
Hockey	

-0.4   -0.2   0.0
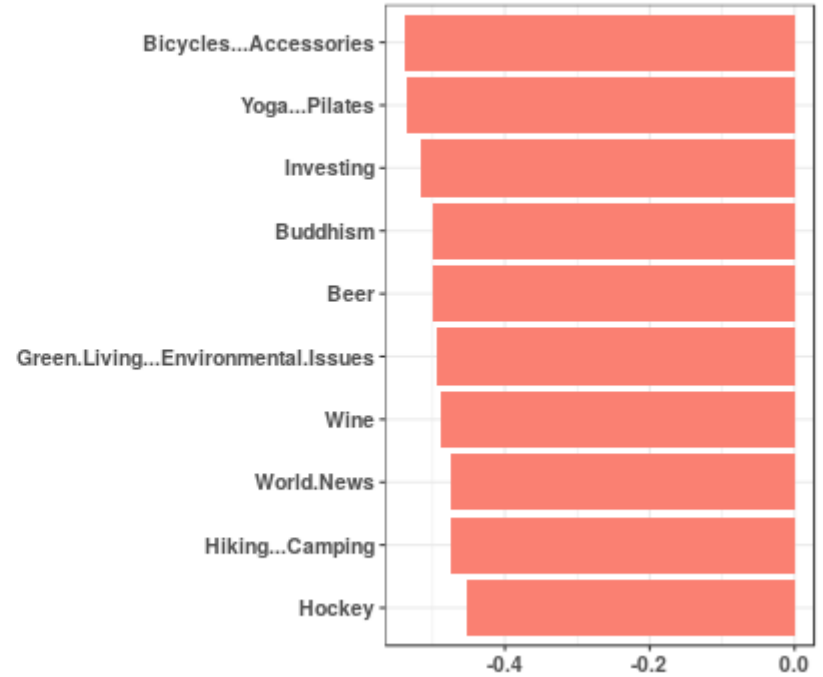
Google
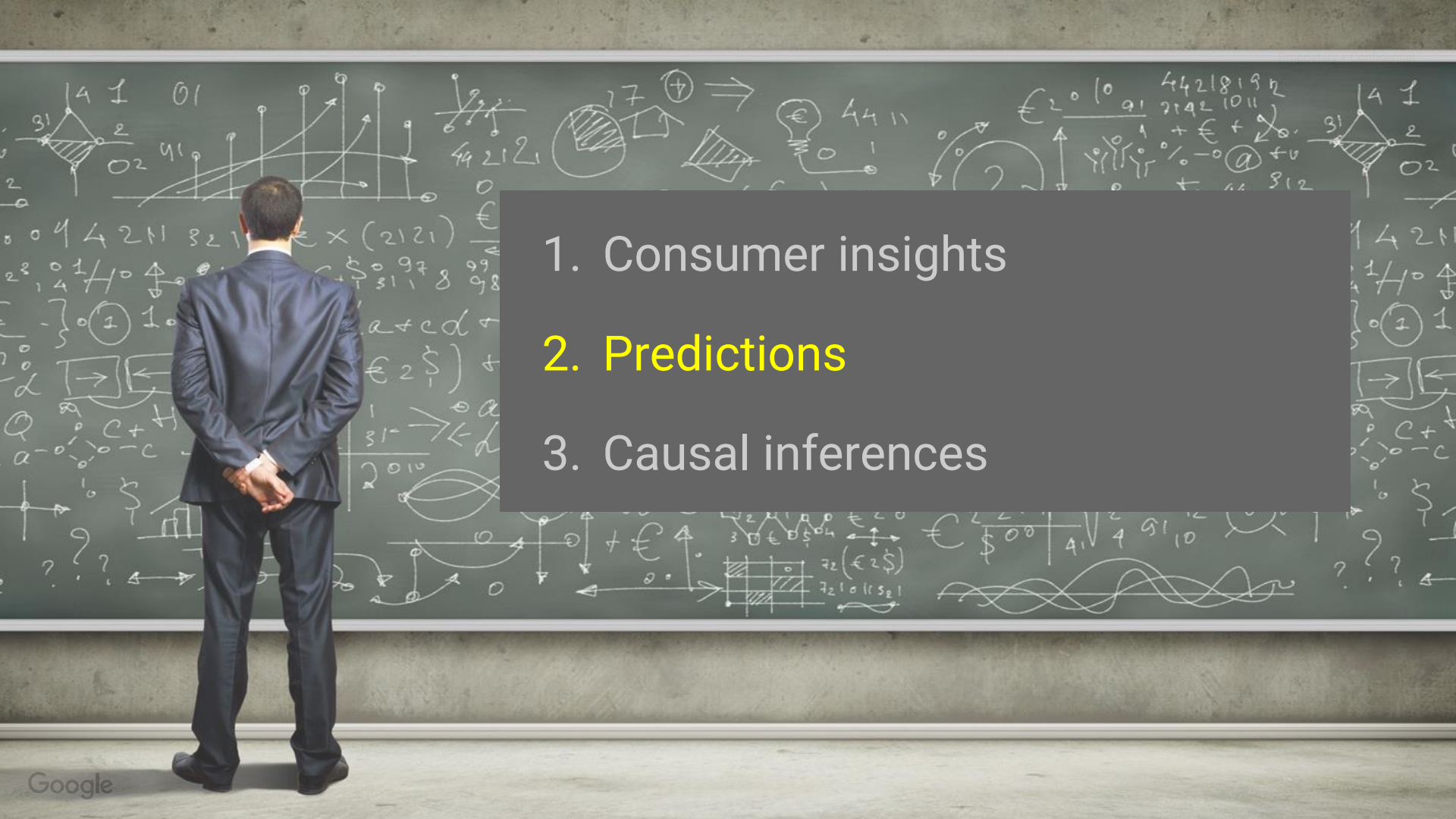
Spread of Trend: K-Pop Girl Band Blackpink
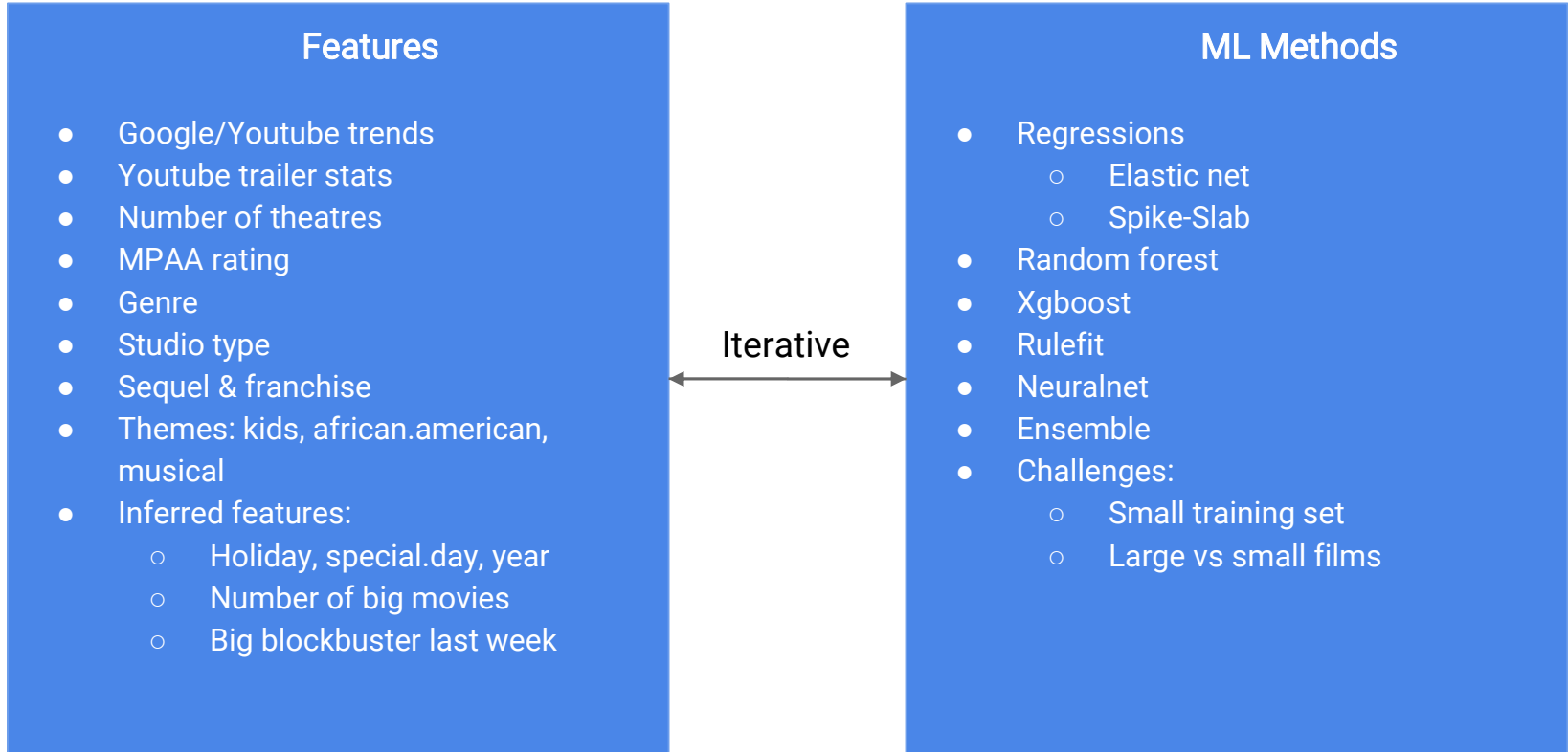

blackpink 2016

1. Consumer insights

2. Predictions

3. Causal inferences

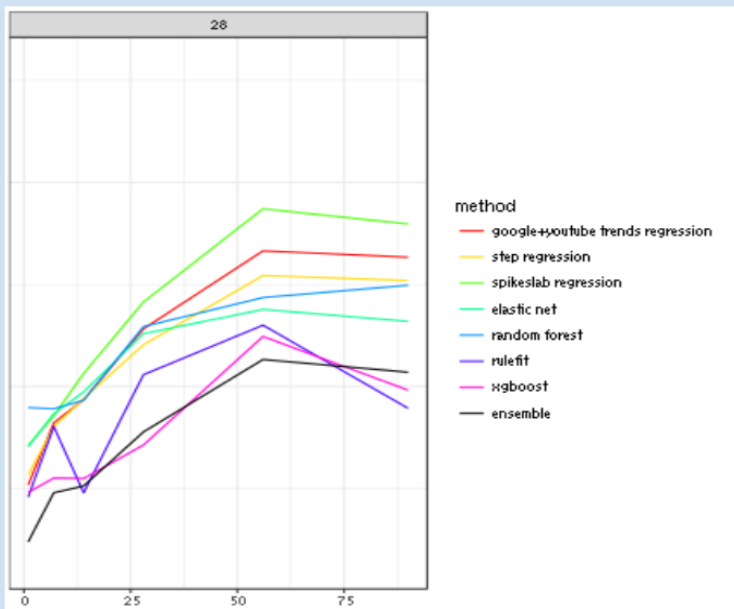Can we use Google trends to help with predictions in the entertainment industry?

- Opening weekend boxoffice
- TV viewership
- Video game sales
- Concert ticket sales

Machine Learning: Opening Weekend Boxoffice Prediction

Features

- Google/Youtube trends
- Youtube trailer stats
- Number of theatres
- MPAA rating
- Genre
- Studio type
- Sequel & franchise
- Themes: kids, african.american, musical
- Inferred features:
  - Holiday, special.day, year
  - Number of big movies
  - Big blockbuster last week

Iterative

ML Methods

- Regressions
  - Elastic net
  - Spike-Slab
- Random forest
- Xgboost
- Rulefit
- Neuralnet
- Ensemble
- Challenges:
  - Small training set
  - Large vs small films
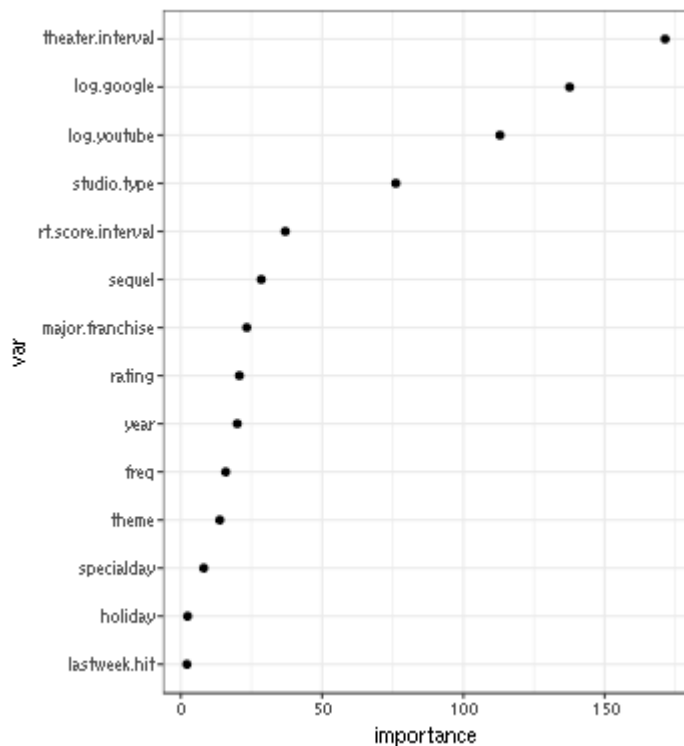
Google

Ensemble Did the Best Again!

Accuracy increases significantly in the final week before release

Accuracy flat between 1.5 months and 3 months.
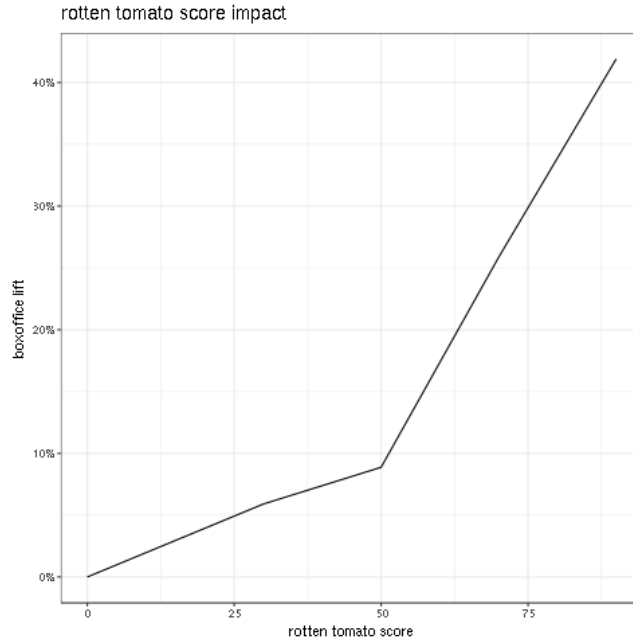
Insights: Variable Importance from Random Forest
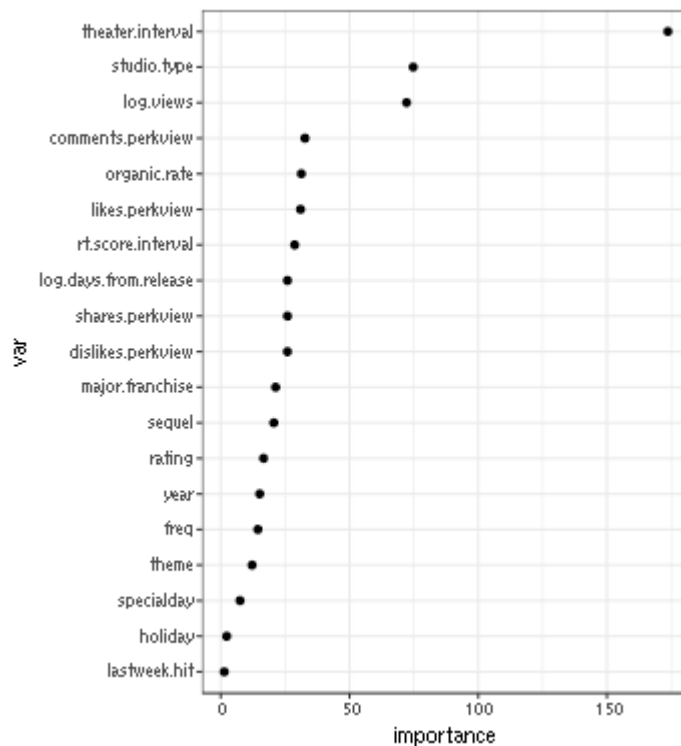
Impact of Rotten Tomato Score
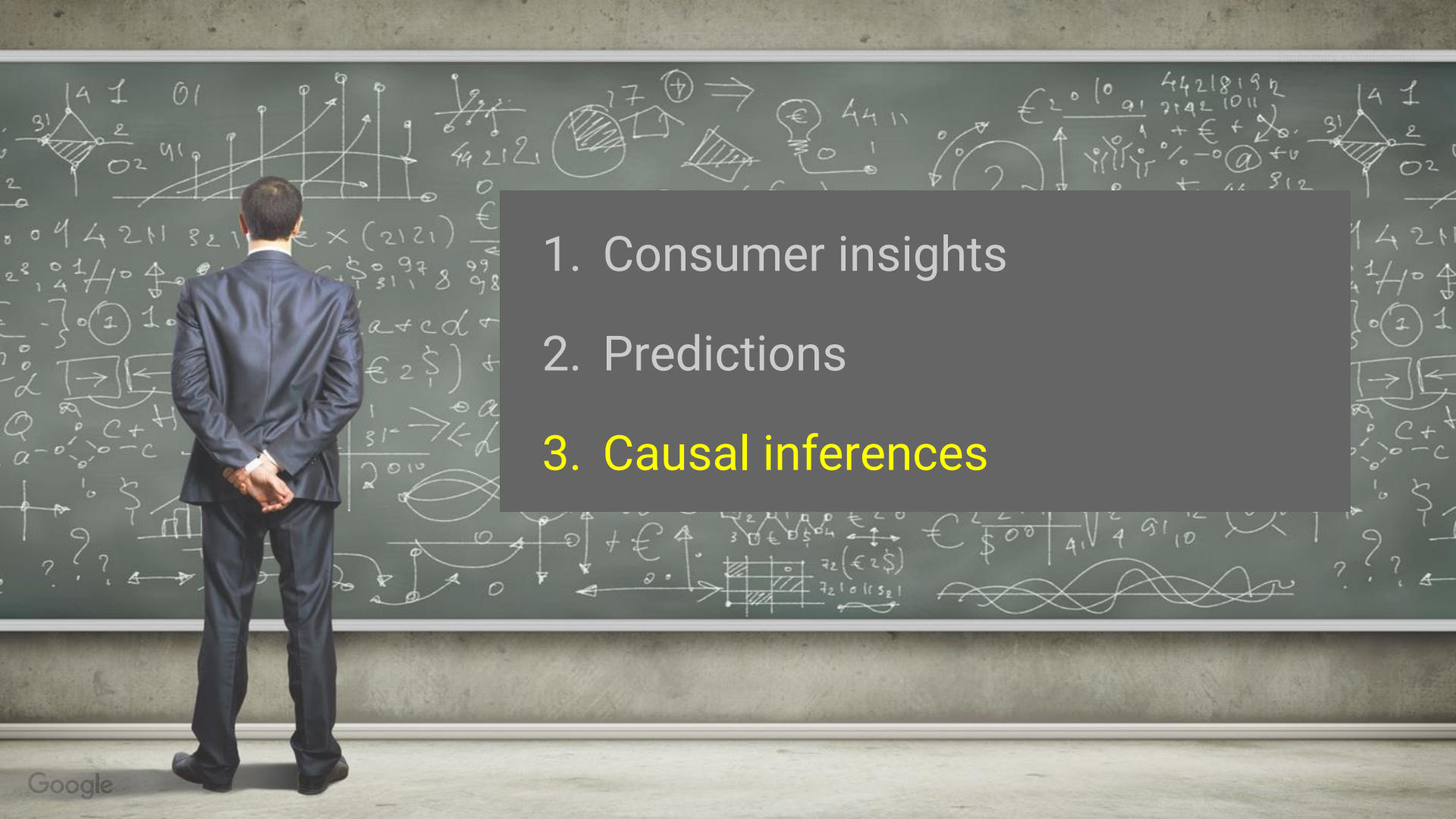
Google

1. Consumer insights

2. Predictions

3. Causal inferences
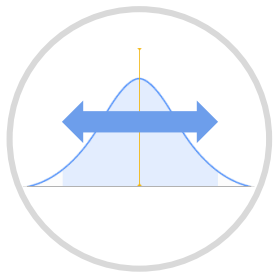
Google
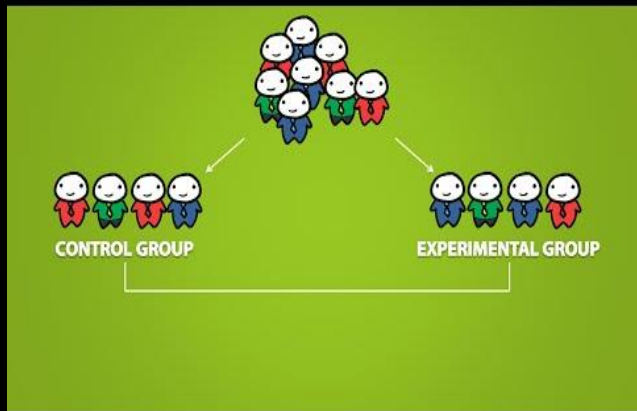
Randomized control and experiment is the ultimate approach to draw causal inference and estimate the advertising effectiveness.

GEO EXPERIMENTS

NATURAL EXPERIMENT

Google

Geo Experiment Process

Test Design

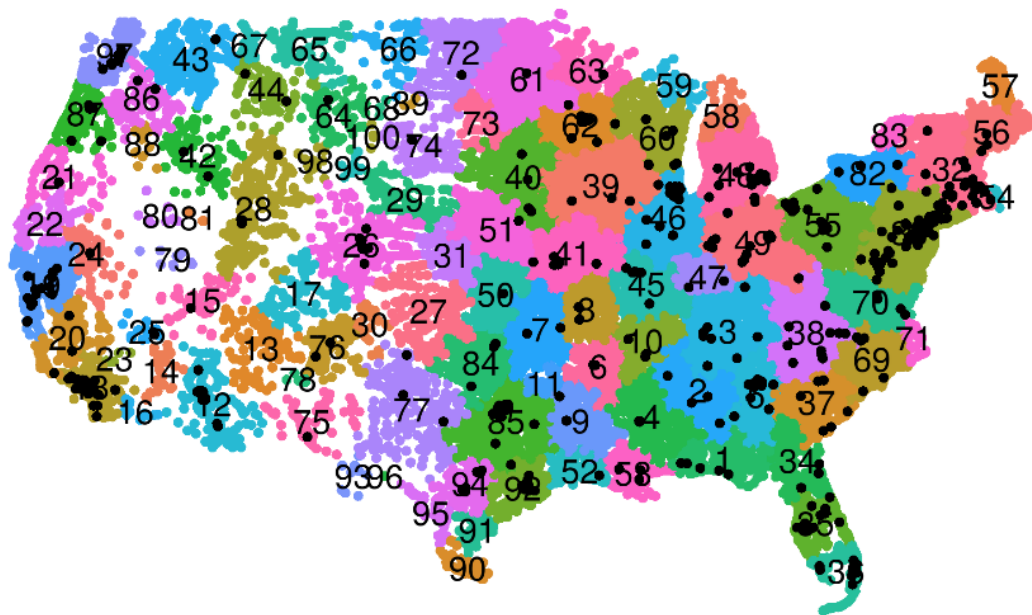Optimize parameters (duration, accuracy, geos) to maximise statistical power.

Experiment

Geo-target ad-spend to randomized sub-regions and expose changes.
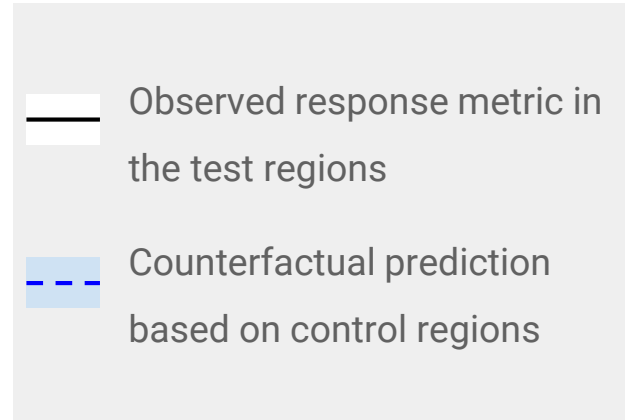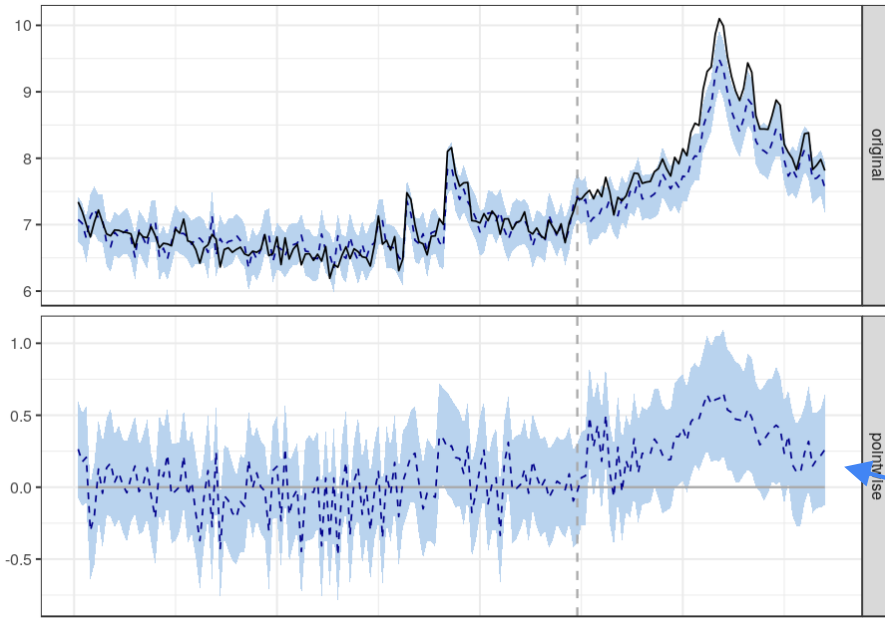
Analysis

Measure incremental revenue and visits and draw effectiveness conclusion.

Google

Designing an Experiment

- Non-overlapping sub-regions

- Consistent with ad serving

- Similarly sized in terms of population / queries

- Minimize contamination between GMAs

- Matching pre-treatment trends

Google

Time Based Methodology (CausalImpact)

Observed response metric in the test regions

Counterfactual prediction based on control regions

Impact X%

Pre test

Test

But experiments are expensive!
Looking for an "natural experiment"
An event where ad exposures are randomly boosted in some cities but not others!

Super Bowl!

Using Super Bowl Fans as the Instrument

- 10-20% more viewers in home cities of competing teams
- Advertisers purchase ad slots long before it is known who will compete
- In 2 cities *"randomly chosen"* out of 32 cities with NFL teams we see a 10-20% elevation in viewership

2012	2013
Boston Final TV Rating: 56.7	Baltimore Final TV Rating: 59.6
Baltimore TV Rating: 47.9	Baltimore TV Rating: 48

Procedure from "Super Returns"

- Outcome: Google trends, opening weekend boxoffice
- 10 Super Bowl games from 2004-2013.
- 72 movies that advertised during games.
- 72 placebo movies that did not advertise during game.
- Compare outcomes in home cities of competing teams to outcome in other cities.

Google

Bump in Home City Queries by Week

Google
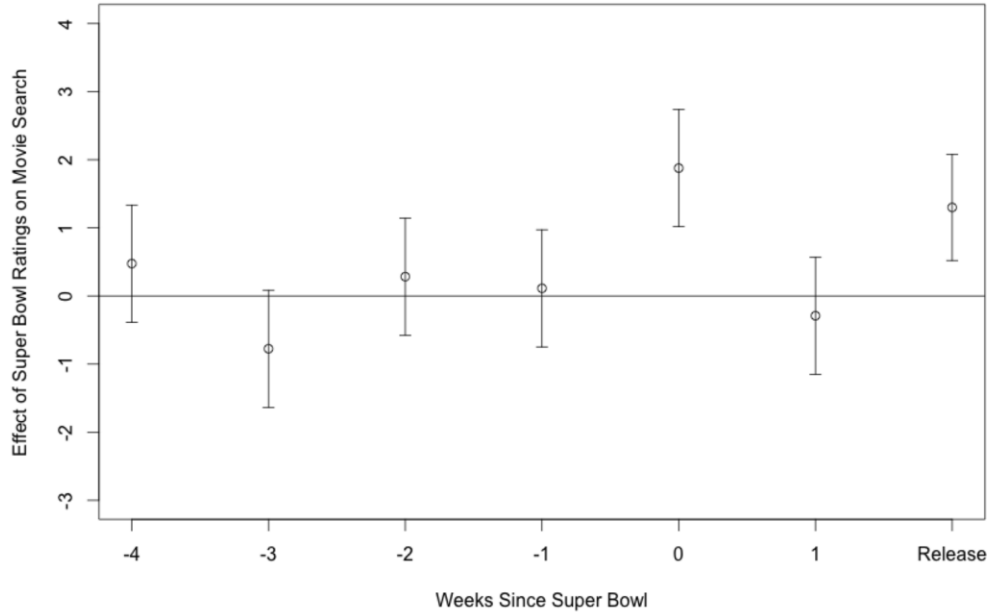
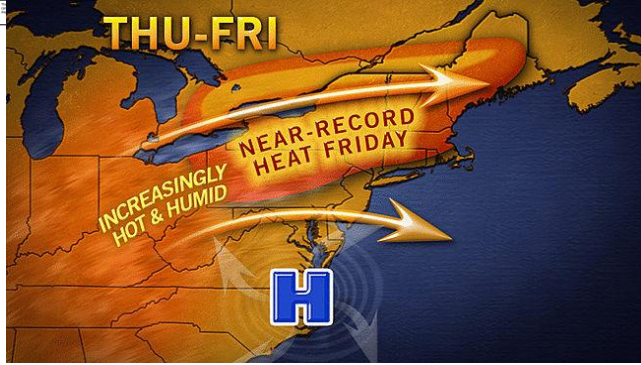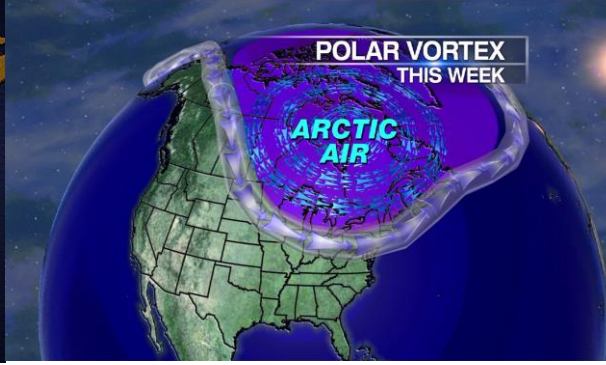## Back-of-Envelope Impact Estimate

- 21% ticket sales increase
- Average opening weekend revenue: $40M
- Super Bowl slot: $3M
- 40-50% goes to studio
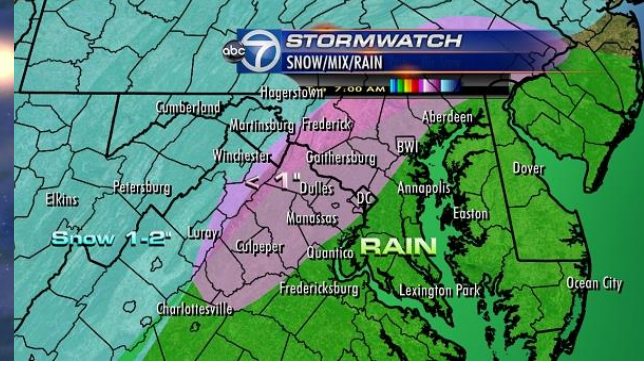- ROI: 1.4 to 1, or 40%

Robustness Test

- Placebo movies: similar movies that weren't advertised in the Super Bowl
- Runners up: teams that just missed qualifying for Super Bowl.
- In both cases: no effect.

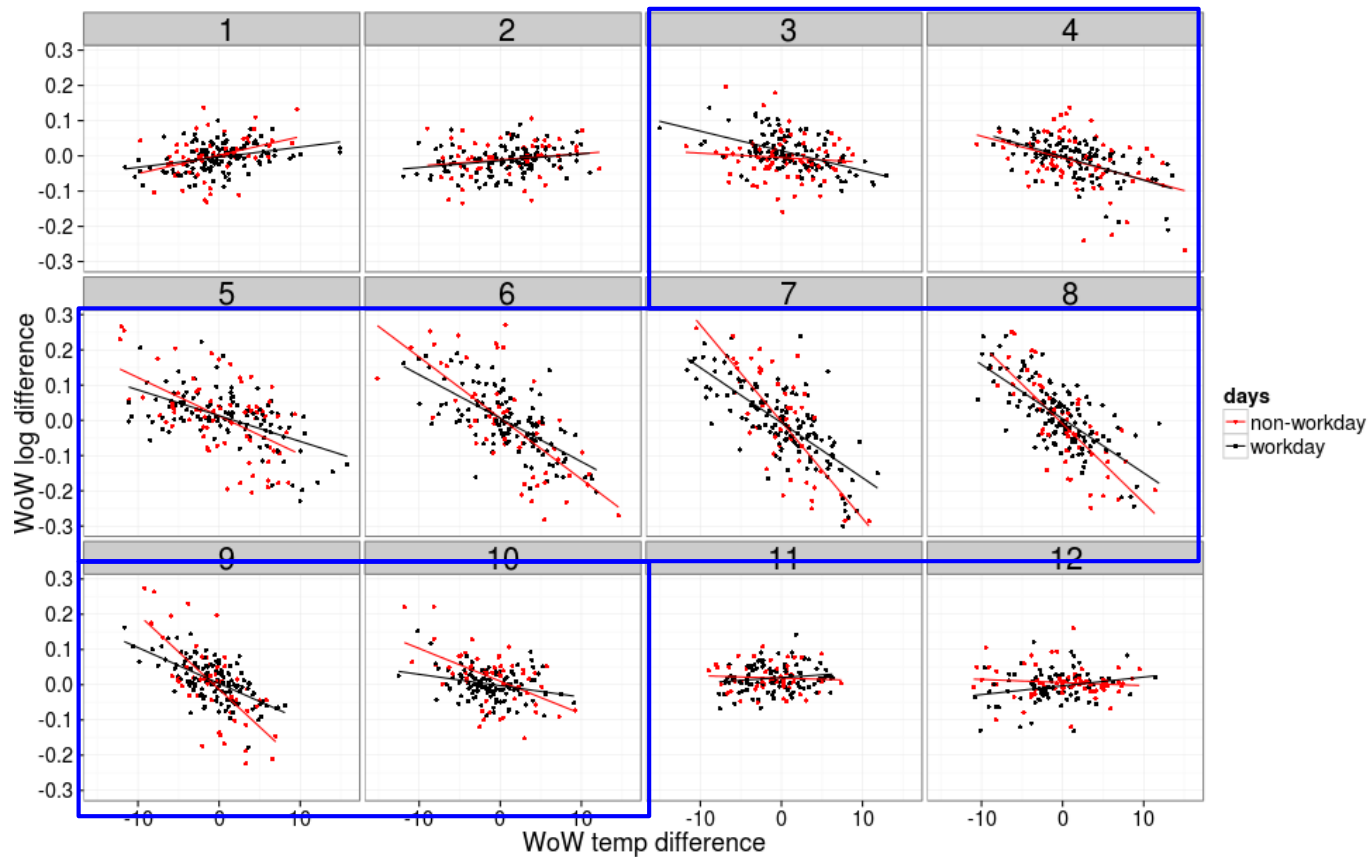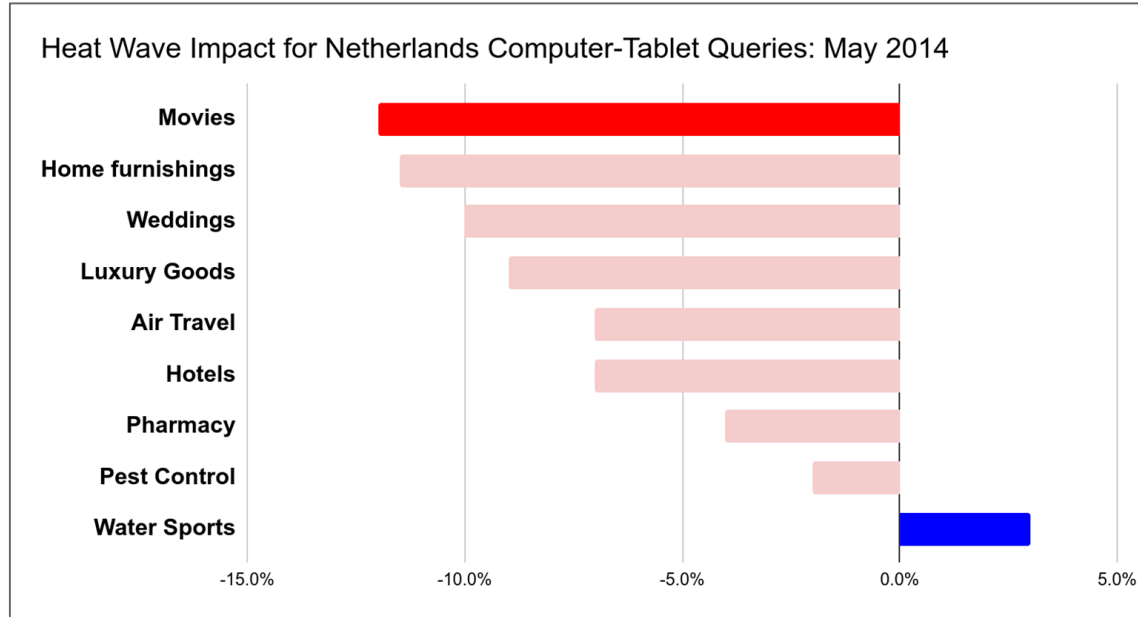**heat**   **cold**   **rain / snow**

Weather Determines Whether You Stay Indoors or Not

# Difference-in-Difference Method: Traffic vs. Temperature

Google

Weather Impact Example: Heatwave in Netherlands

Heat Wave Impact for Netherlands Computer-Tablet Queries: May 2014

Google

But Is that the Case in US?

From February to May warm weather in Washington DC has negative impact on movie searches. But from July to August, it is much more muted. The impact is slightly positive over the weekends as people seek ACs in the theaters.

Google
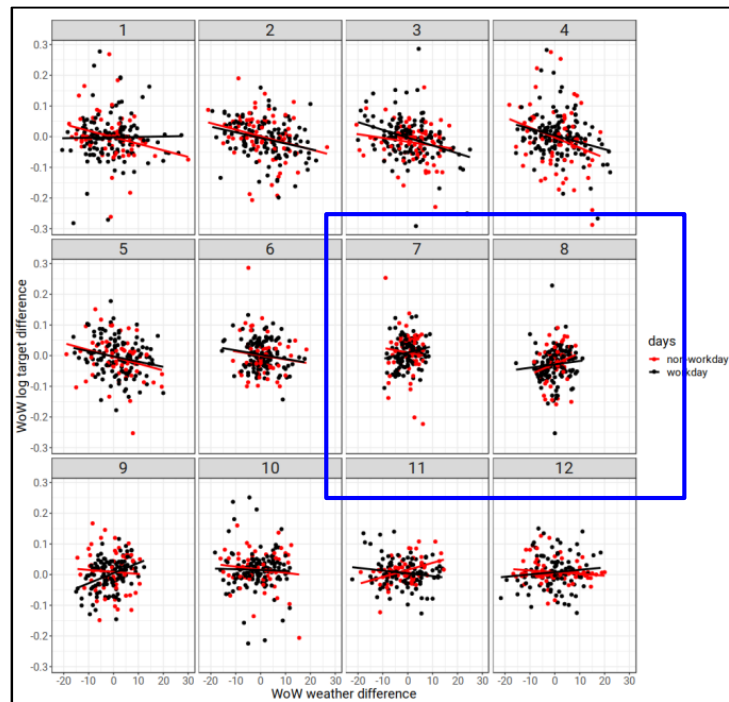
Using data to get insights and make decisions!

THANKS!