

Nowcasting with Google Trends

Hal Varian
March 16, 2022
UC Berkeley
Hobart, Australia
July 12, 2022

Opinions expressed in these slides are solely those of the author.

12 years ago...

Predicting the Present with Google Trends

HYUNYOUNG CHOI and HAL VARIAN

Google, Inc., California, USA

In this paper we show how to use search engine data to forecast near-term values of economic indicators. Examples include automobile sales, unemployment claims, travel destination planning and consumer confidence.

3000 citations in Google Scholar!

Google Trends

Searches for [hangover]

Which day of the week are there the most searches for [hangover]?

1: Sunday

2: Monday

3: Tuesday

4: Wednesday

5: Thursday

6: Friday

7: Saturday

Hangover-vodka time series from Trends

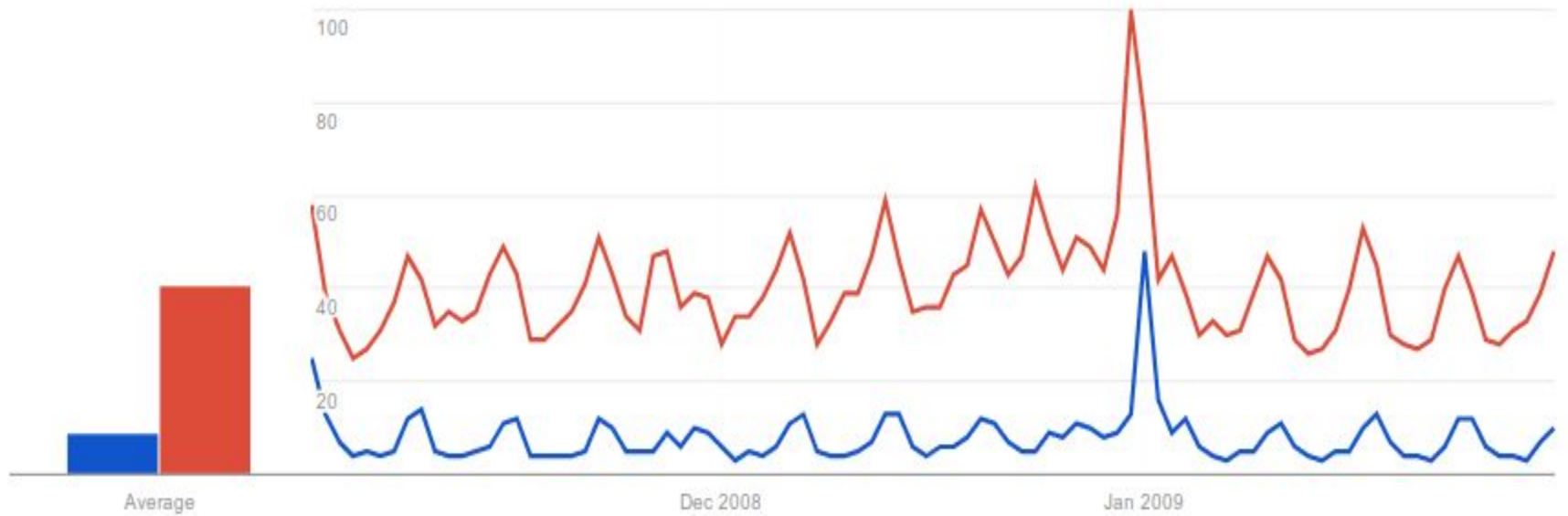
Web Search Interest: **hangover, vodka**. United States, Nov 2008 - Jan 2009.



Interest over time ?

The number 100 represents the peak search volume

News headlines Forecast ?

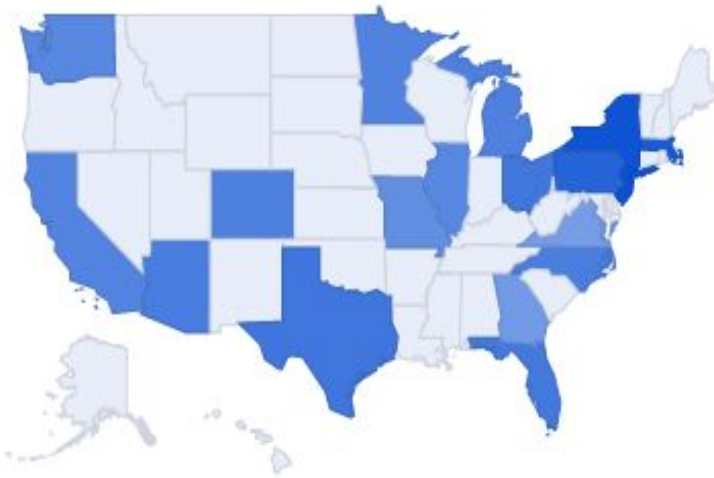


Hangover geo

Regional interest ?



Worldwide > United States



0 100

Subregion | Metro | City

▶ View change over time ?

Embed

Related terms ?



Top

Rising

cure hangover	100	
hangover cures	65	
the hangover	50	
cure a hangover	40	
hangover remedies	35	
hangover food	15	
hangover symptoms	15	
love hangover	15	
cure for hangover	15	
best hangover cure	15	

Embed

Limitations of Google Trends

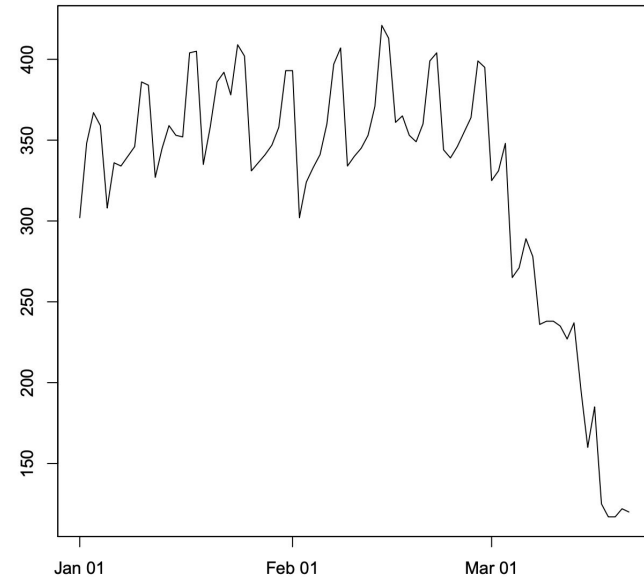
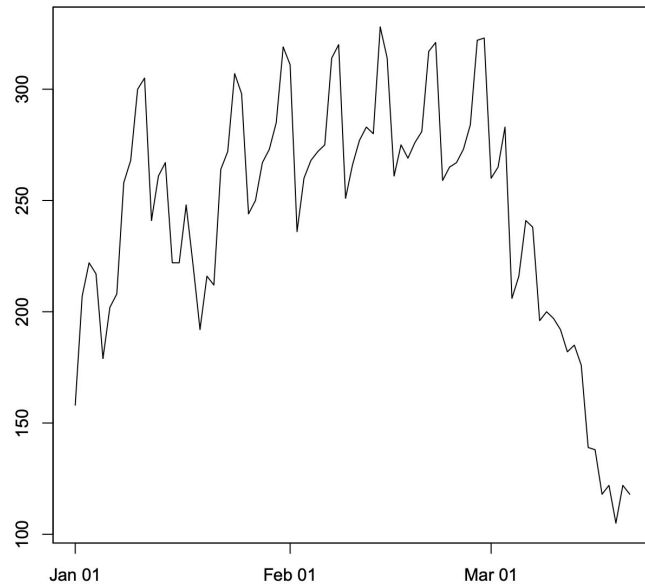
- Only popular queries (from distinct IP addresses)
- Only aggregate data (country, state, metro)
- Largest value of index normalized to 100
- Other values are relative to largest index
- Data is a sample, cached and refreshed daily
- API is available to qualified researchers
- Cap on number of daily downloads
- Topics and categories can be searched as well
- Can search for queries from web, images, news, shopping, youtube

Nowcasting and forecasting

- John Battelle : “Google is a database of intentions”
 - Intentions usually precede action
 - Perhaps Google queries can predict actions?
- Most searches lead quickly to actions, e.g., an impulse purchase [**nowcasting**]
- But some intentions are far in advance of actions
 - “Buy a new house”
 - “Buy a new car”
 - “Vacation in Hawaii”
 - Tourism in general
- May be able to prediction actions months in advance [**forecasting**]

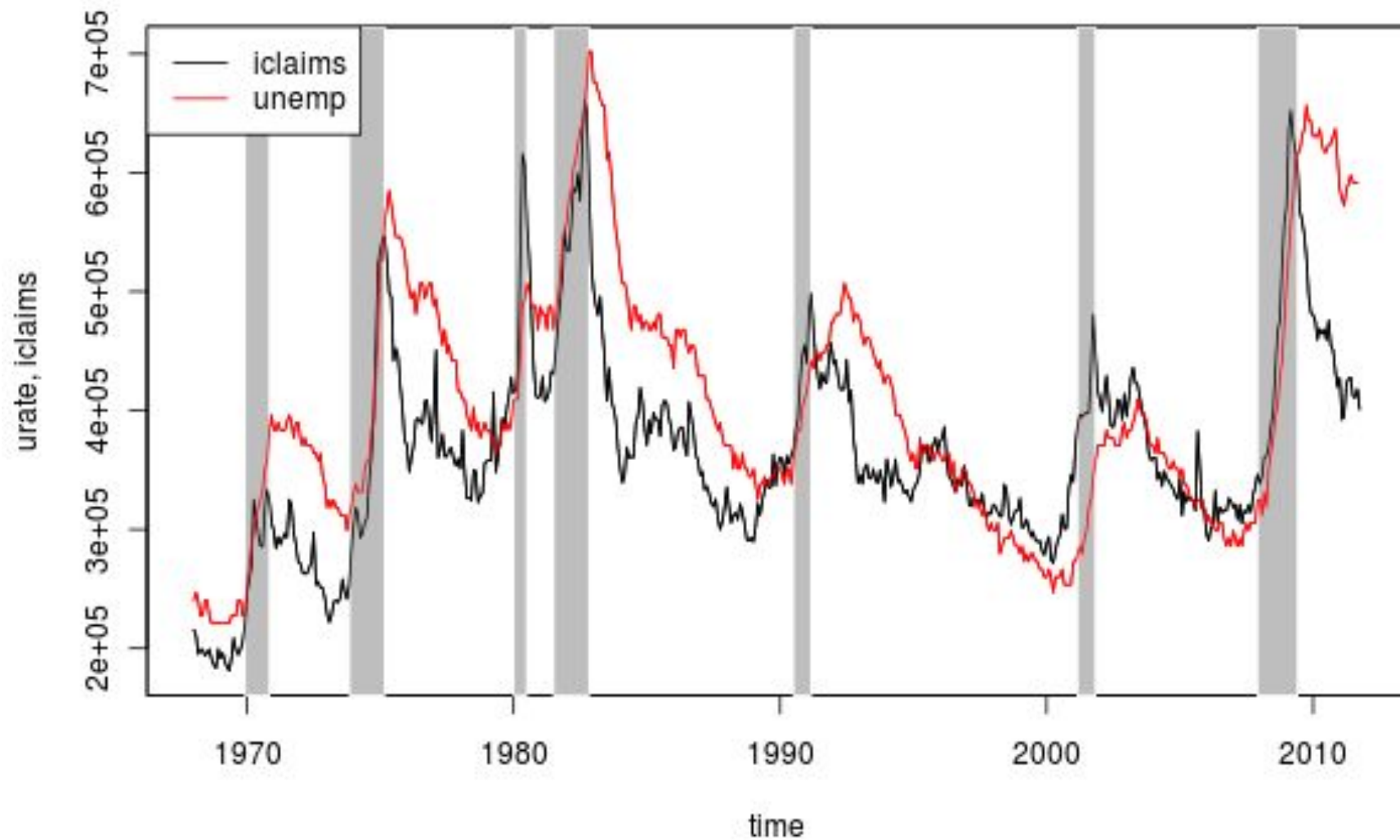
Real time economic metrics

- Economic data are released periodically and infrequently (monthly, quarterly).
- Data are revised as new information becomes available.
- A real-time signal would enable rapid response
- Example: Fed and DAU traffic data April 2020
- Two cities Daily Average Users



Example: initial claims
for unemployment
benefits

Unemployment and Initial Claims



What would you do if you became unemployed?

Go to Google
and look for
information
about
unemployment
benefits!



michigan unemployment,
idaho unemployment,
pennsylvania unemployment,
illinois unemployment,
new jersey unemployment,
unemployment filing,
rhode island unemployment,
unemployment office,
filing unemployment,
state unemployment,mi,
unemployment,connecticut,
department of unemployment,

Challenge: build a model to nowcast indicators

- Model combines Kalman filter + regression
- **Kalman filter:** estimate trend and seasonal components
 - prediction this period = prediction last period + error adjustment
 - one step ahead prediction
- **Spike and slab regression:** choose best predictors
 - spike: probability a predictor is in the regression
 - slab: probability distribution of coefficient conditional on being in regression
- Fit models using Markov chain Monte Carlo (**MCMC**)
- Put everything together to get **Basic Structural Model**
 - $y_t = \text{level}_t + \text{trend}_t + \text{seasonal}_t + \text{regression}_t + \text{error}_t$

Basic structural model + regression

$$y_t = \underbrace{\mu_t}_{\text{level}} + \underbrace{\gamma_t}_{\text{seasonal}} + \underbrace{\beta^T \mathbf{x}_t}_{\text{regression}} + \epsilon_t \quad \text{observation eqn}$$

$$\mu_t = \mu_{t-1} + \delta_{t-1} + u_t \quad \text{level evolution}$$

$$\delta_t = \delta_{t-1} + v_t \quad \text{trend evolution}$$

$$\gamma_t = - \sum_{s=1}^{S-1} \gamma_{t-s} + w_t \quad \text{seasonal evolution}$$

See Harvey, chapter 7

Bayesian structural time series

- BSTS is available in R or Python
 - CRAN, Pystan, Github
- Technical description in NBER volume in [*Economic Analysis of the Digital Economy*](#)



Created by Steve Scott

```
# Build a state space model
```

```
ss <- AddLocalLinearTrend(list(), y)
```

```
ss <- AddSeasonal(ss, y, nseasons=52)
```

```
# estimate the model
```

```
model <-
```

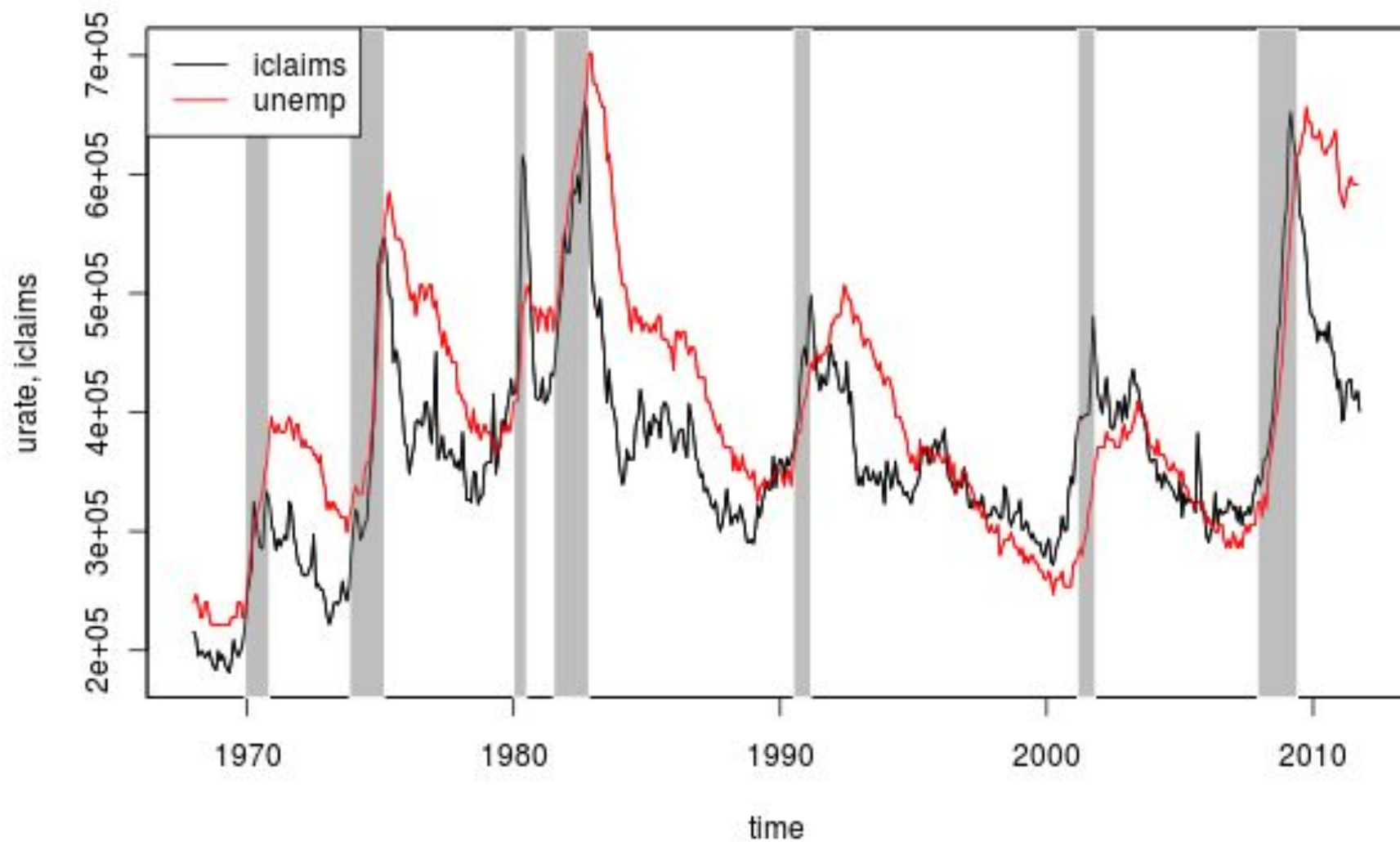
```
bsts(y~., state.specification=ss, data=x, niter=3000,
```

```
ping=200, expected.model.size=5, seed=123)
```

Examples

Initial claims for
unemployment
benefits

Unemployment and Initial Claims



Estimating model using BSTS

```
y <- my.data$ResponseVariable
```

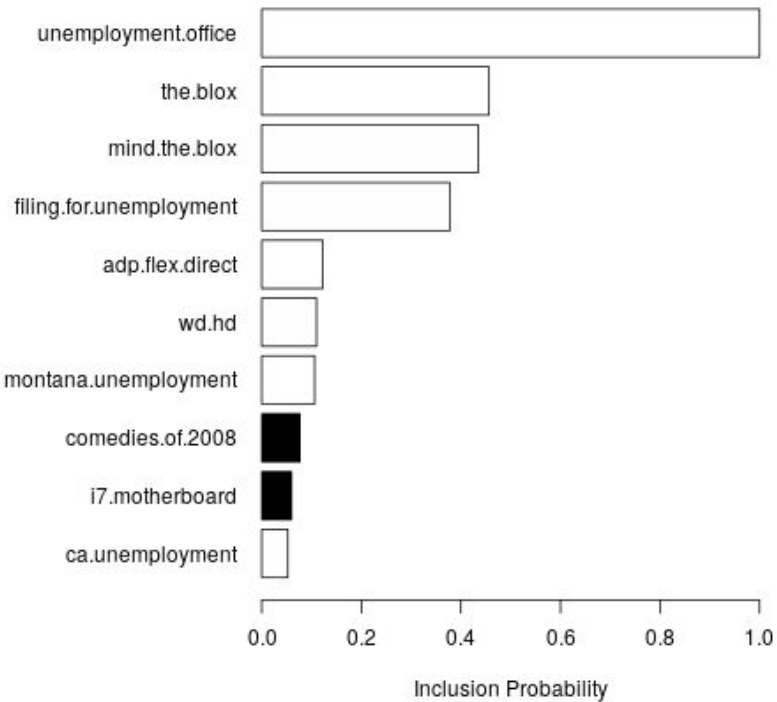
```
ss <- AddLocalLinearTrend(  
  list(),    ## No previous state specification.  
  y)        ## Peek at the data for scaling.
```

```
ss <- AddSeasonal(  
  ss,        ## Adding state to ss.  
  y,        ## Peek at the data for scaling.  
  nseasons = 7) ## 7 "seasons" for day of week effect
```

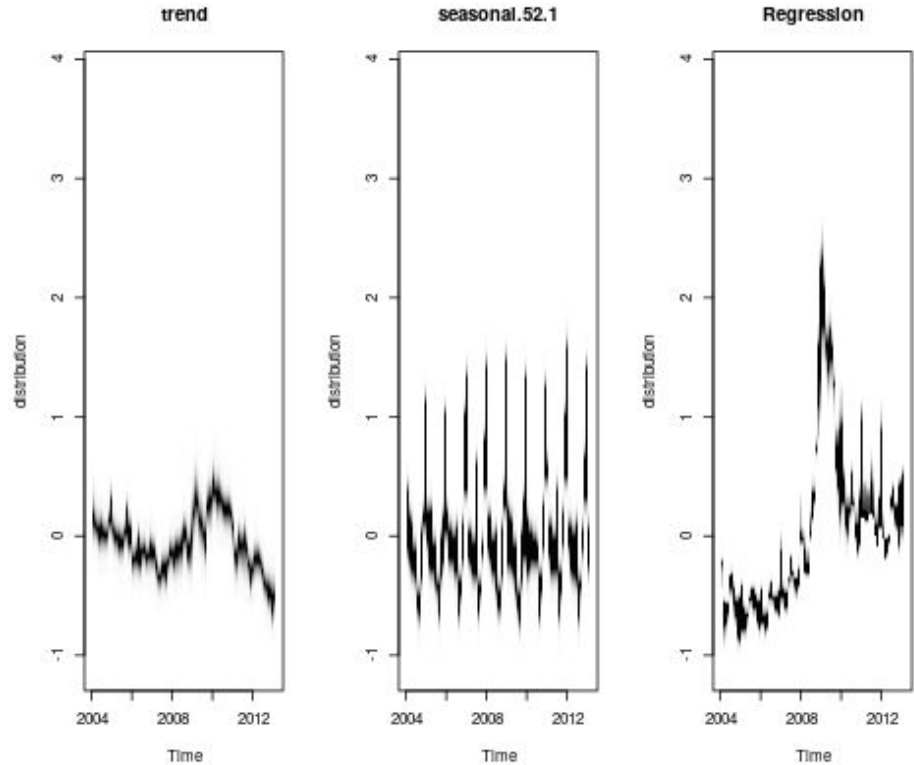
```
model <- bst(y ~ .,          ## regression formula like 'lm'  
             state.specification = ss, ## time series spec  
             niter = 1000,      ## MCMC iterations  
             data = my.data,  
             expected.model.size = 1) ## spike-slab
```

Model output

Probability of inclusion

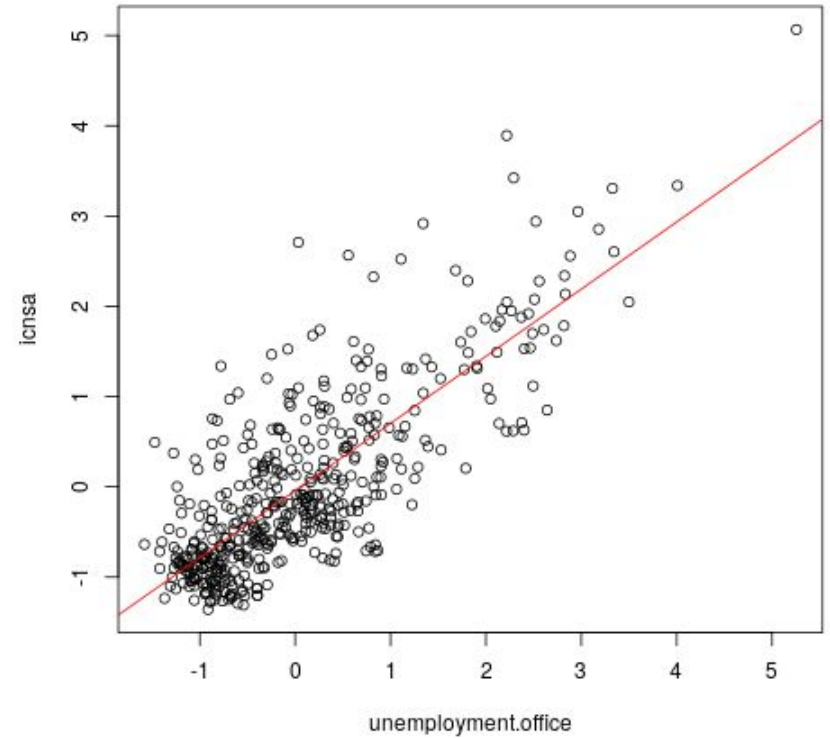
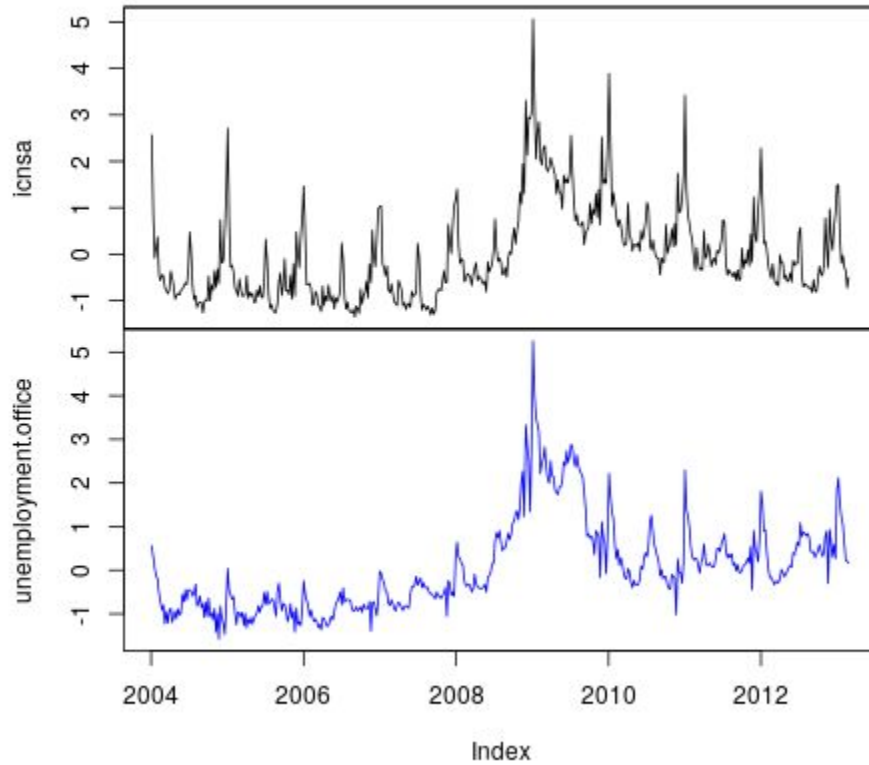


Component decomposition



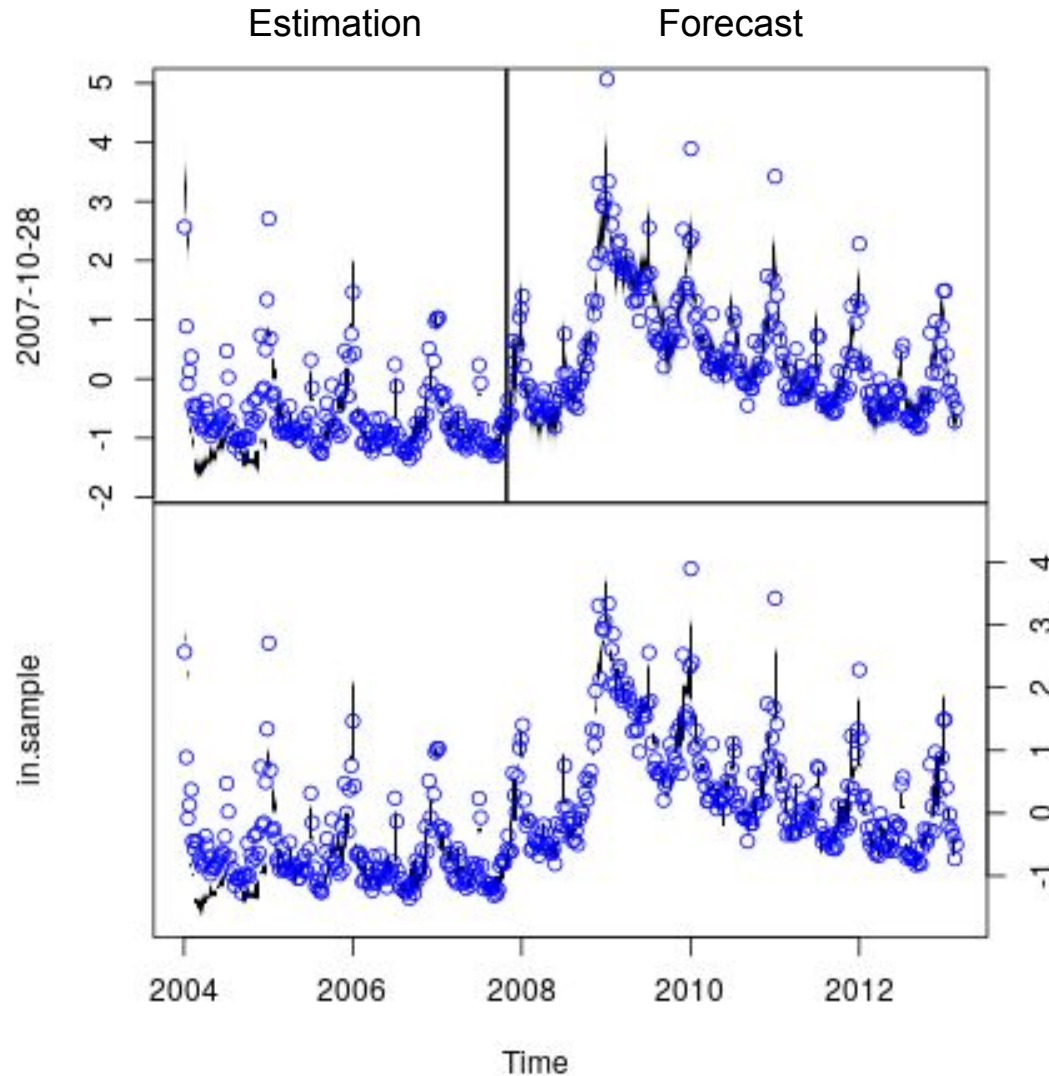
Actual and forecast

Actual and forecast



Out of sample forecast

Can specify “cutpoints” so model parameters can be estimated over a region, then they are “frozen” and can capture out-of-sample forecasts.

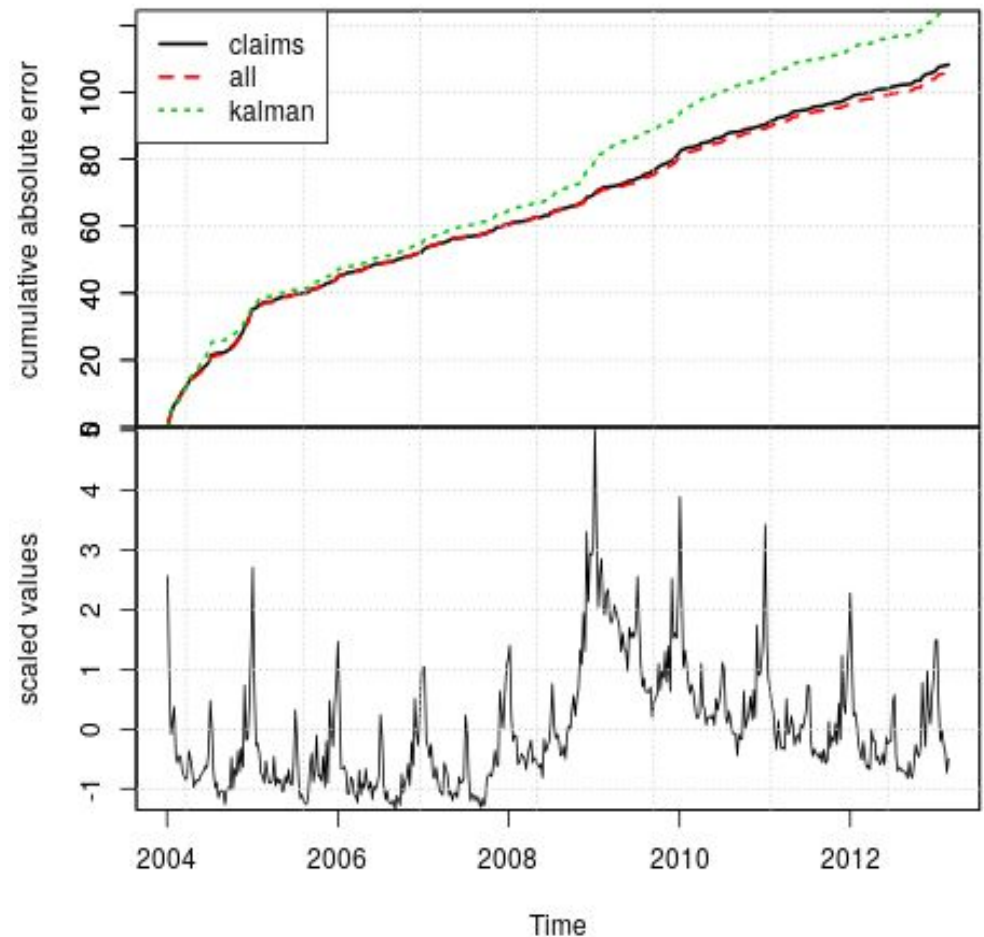


Compare 3 models

Look at 1-step ahead absolute prediction errors

1. Kalman only
2. All predictors
3. “unemployment related” only
 - michigan unemployment
 - idaho unemployment
 - pennsylvania unemployment
 - illinois unemployment
 - new jersey unemployment
 - unemployment filing
 - rhode island unemployment
 - unemployment office
 - filing unemployment
 - state unemployment
 - unemployment connecticut
 - department of unemployment

Up until recession hit, all models worked equally well. After that, regression performed better than Kalman, even if we restrict to just a few predictors



Housing

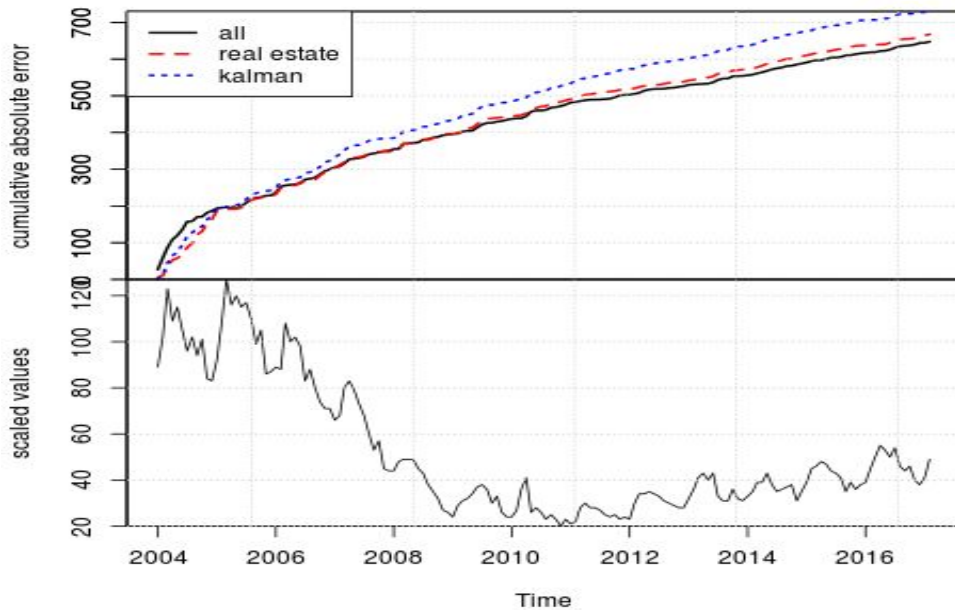
Use query categories

Real_Estate__Apartments_and_Residential_Rentals
Real_Estate__Commercial_and_Investment_Real_Estate
Real_Estate
Real_Estate__Property_Development
Real_Estate__Property_Inspections_and_Appraisals
Real_Estate__Property_Management
Real_Estate__Real_Estate_Agencies
Real_Estate__Real_Estate_Listings
Real_Estate__Timeshares_and_Vacation_Properties

Advantage: exact queries may not persist in future but categories will (likely) exist

Three models:

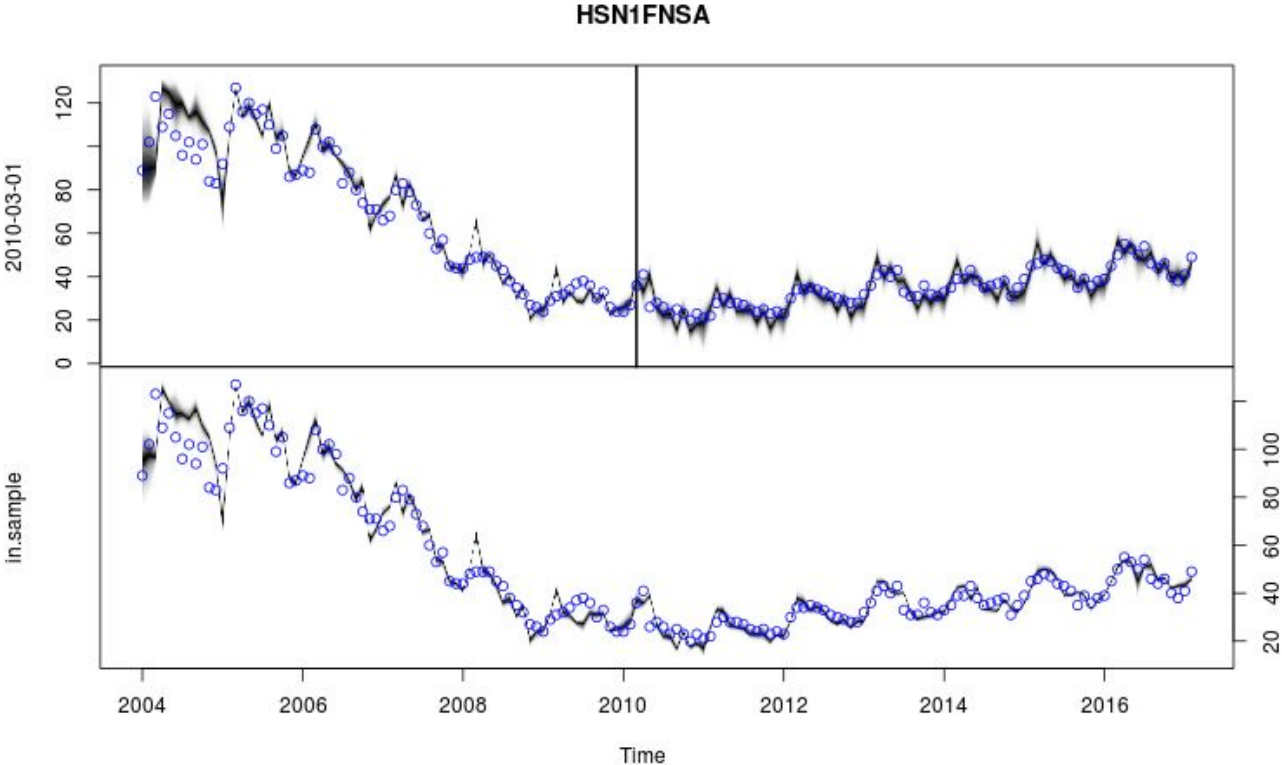
- 1) 148 commercial categories,
- 2) 9 real estate categories
- 3) Kalman only



Restricting to real estate had little impact on prediction error

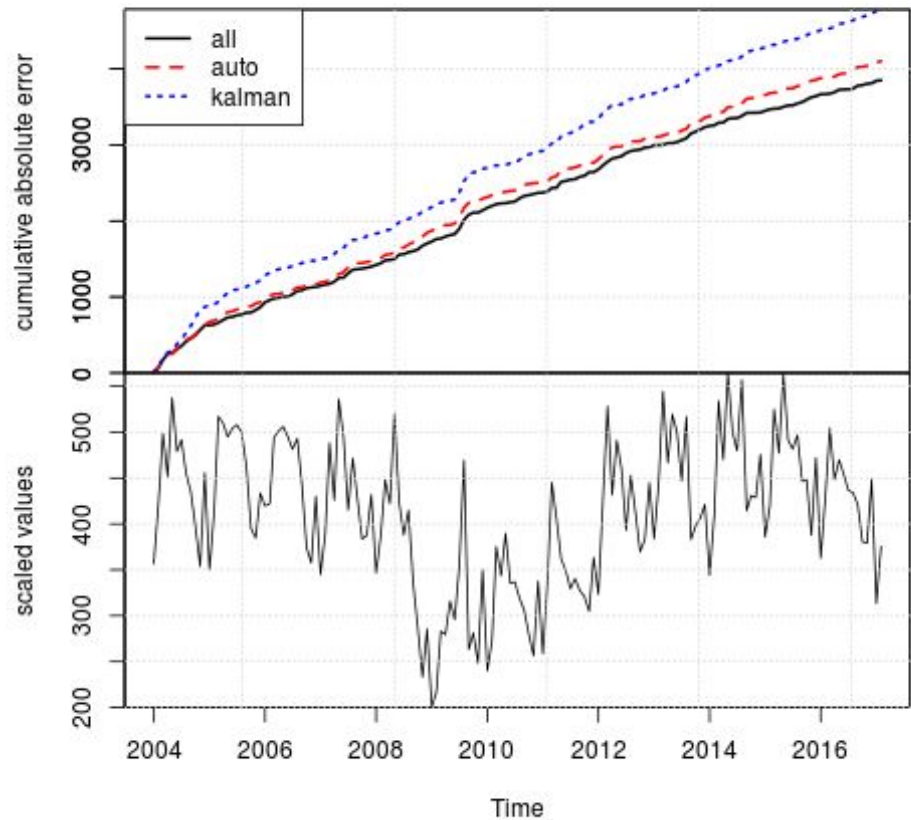
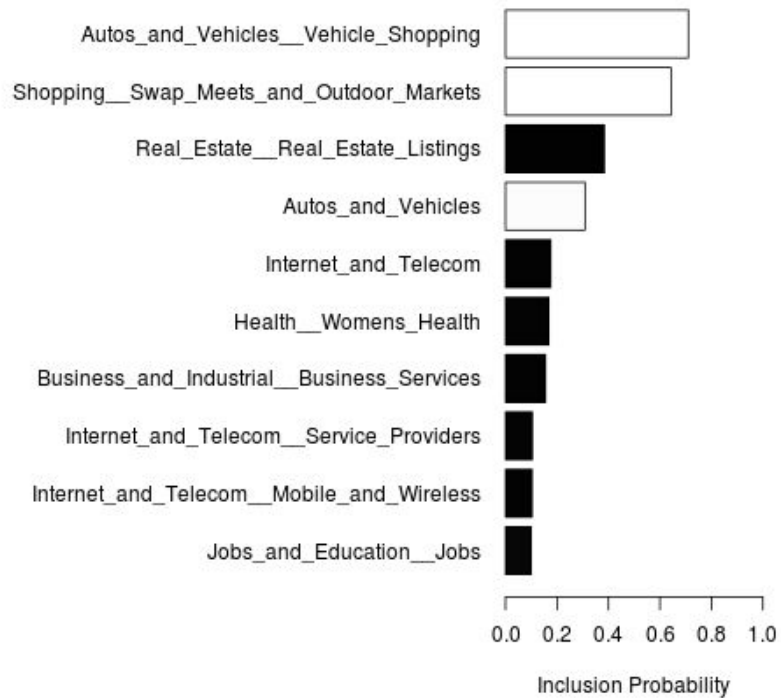
Out of sample prediction for housing

Estimate up until time t and then freeze all the posterior distributions. This freezes regression and the variance posteriors, but allows for Kalman updating.

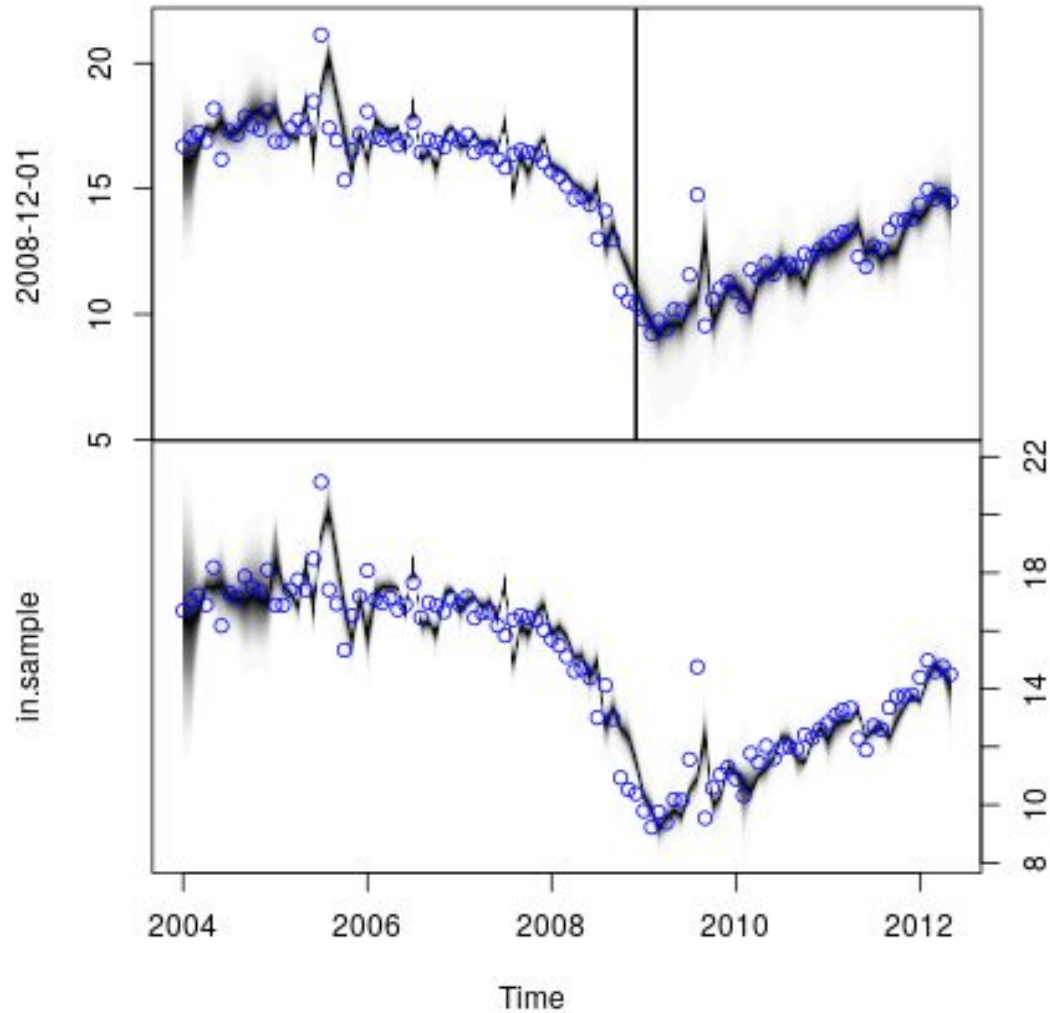


Automobiles

Example for motor vehicle sales



Out of sample prediction of automobile sales



OECD Tracking GDP

Tracking Activity in Real Time with Google Trends

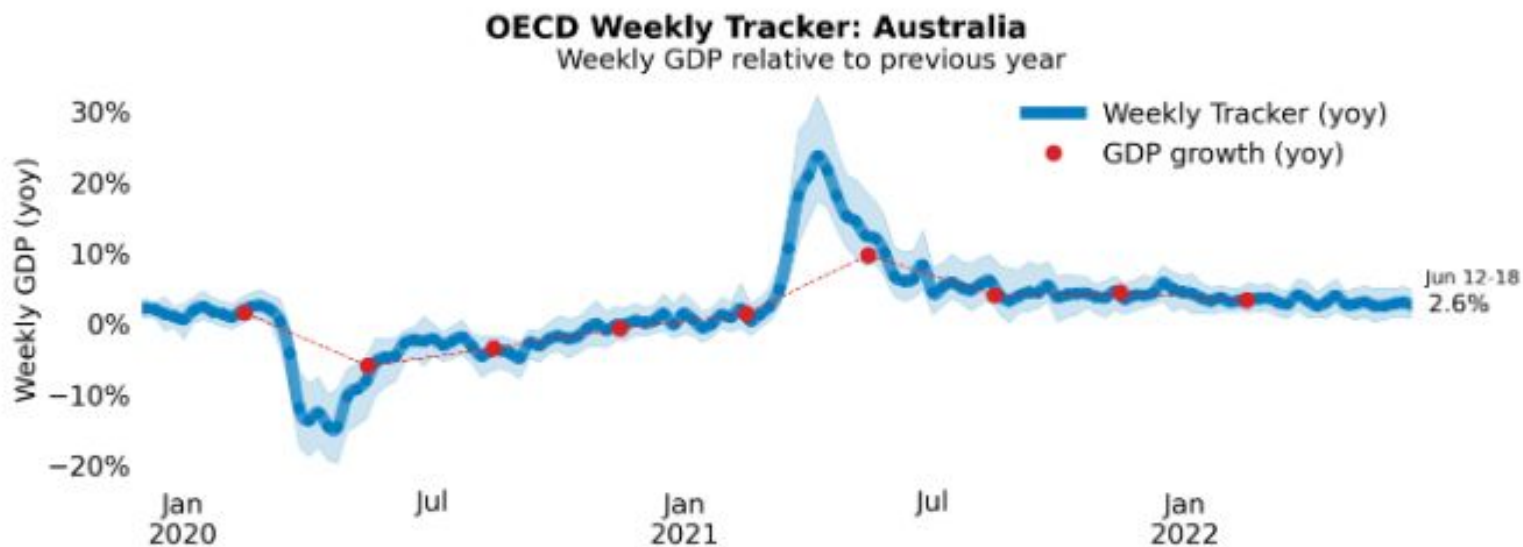
Nicolas Woloszko, OECD



Weekly Tracker by country

Select a country:

Australia



Note: The Weekly Tracker provides an estimate of weekly GDP based on Google Trends search data and machine learning.
Source: OECD Weekly Tracker (Woloszko, 2020), <https://www.oecd.org/economy/weekly-tracker-of-gdp-growth/>;
OECD Quarterly National Accounts.

Some predictors from the OECD GDP Tracker

See Nicolas Woloszko, [Tracking Activity in Real Time with Google Trends](#)

'Crisis / Recession': "Economic crisis", "Crisis", "Recession", "Financial crisis", 'Krach'
'Unemployment / unemployment benefits':

"Unemployment", "Unemployment benefits", 'Welfare & Unemployment'

'Student loan', 'Credit & Lending', 'Loan', 'Interest', 'Mortgage', 'Auto Financing'

'Consumption items' : 'Food & Drink', 'GPS & Navigation', 'Performing Arts', 'Luggage topic',
'Vehicle Brands', 'Birthday', 'Travel', 'Energy & Utilities', 'Vehicle Shopping', 'Tobacco Products',
'Health', 'Pharmacy', 'Carpooling & Ridesharing', 'Sports', 'Animal Products & Services',
'Fitness', 'Weddings', 'Car Rental & Taxi Services', 'Autos & Vehicles', 'Tourist Destinations',
'Home & Garden', 'Events & Listings', 'Grocery & Food Retailers', 'Vehicle Licensing &
Registration', 'Timeshares & Vacation Properties', 'Home Appliances', 'Mass Merchants &
Department Stores', 'Car Electronics', 'Fashion & Style', 'Trucks & SUVs', 'Home Furnishings',
'Footwear', 'Cruises & Charters', 'Hotels & Accommodations', 'Luggage & Travel Accessories',
'Fast Food', 'Book Retailers', 'Veterinarians', 'Spas & Beauty Services', 'Acting & Theater',
'Travel Agencies & Services'

"Jobs" : 'Waiter', 'Job Listings', 'Resumes & Portfolios', 'Jobs topic', 'Temporary jobs', etc

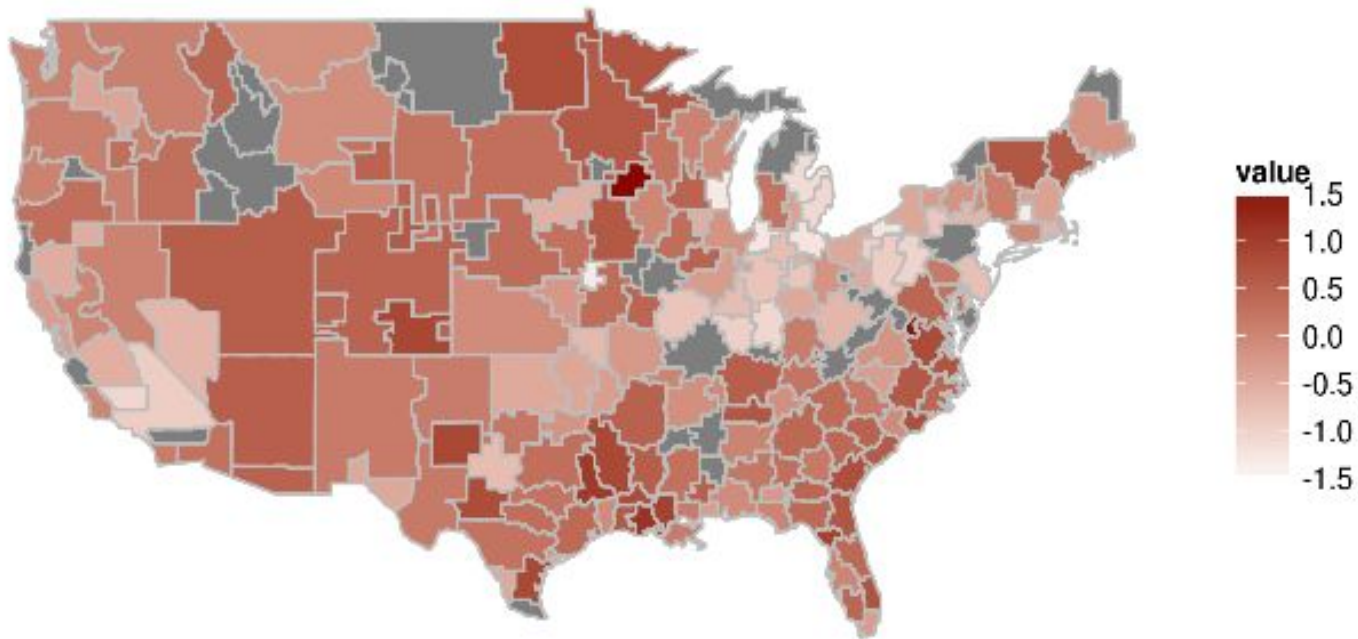
Cross section

Where are the happiest cities in the US?

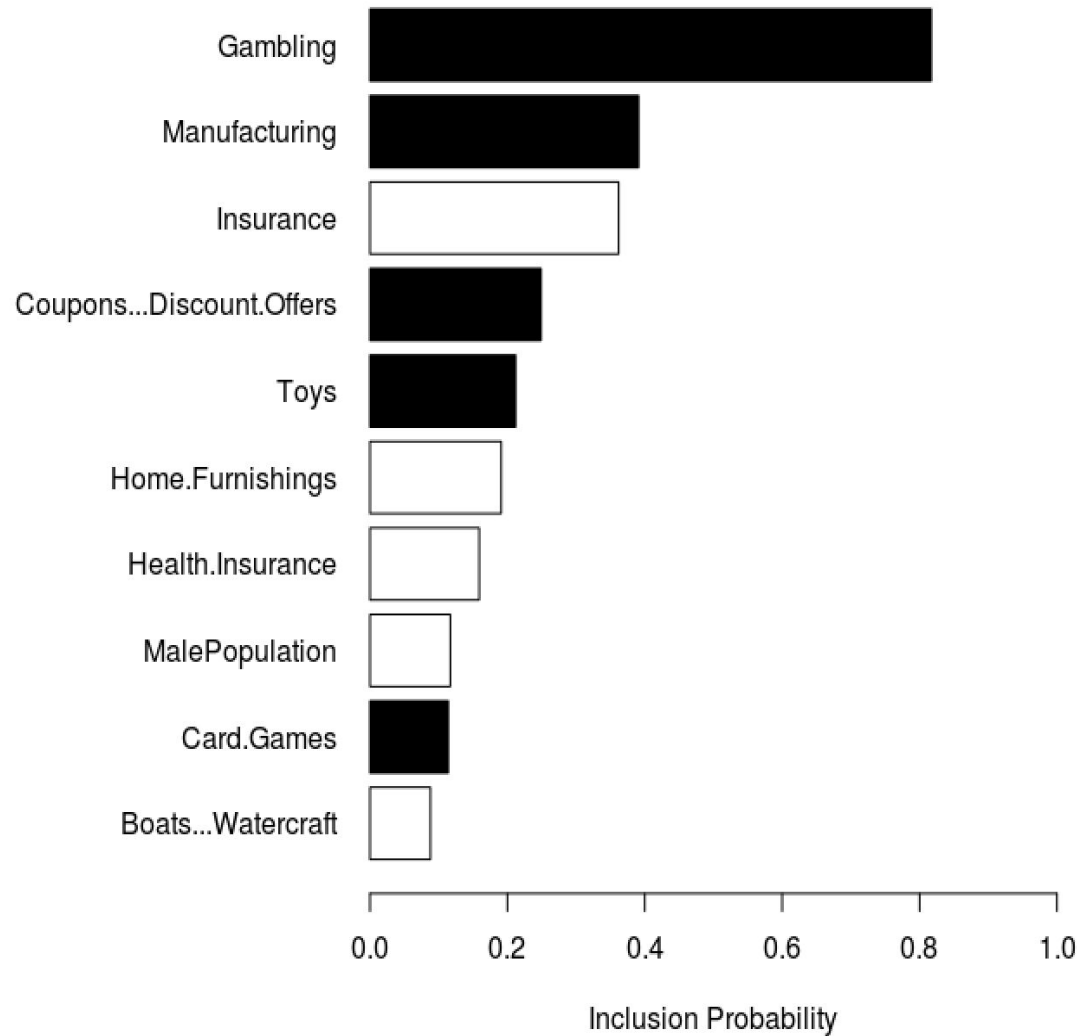
Ed Glaeser, Joshua Gottlieb, and Oren Ziv, “Unhappy Cities”, NBER 20291

CDC Behavioral Risk Factor Surveillance System survey: “In general, how satisfied are you with your life?” Answers for 174 cities.

What kinds of queries are predictive of happy and unhappy cities?



Probability of inclusion in regression for happy cities



Google Surveys

Google Surveys

How it works:



Ask your questions online



People across the internet respond in exchange for access to high-quality content or Google Play credit



Get aggregated and analyzed results in one simple online interface

Fair Use Digital Circulation Strategy Information Overload

The Work of Art in the Age of Mechanical Reproduction

Matthew Dodd from the January 16, 2013 issue

Jurgen Habermas R&D Android cops beat The Weekender mathewi Tim Carmody attracting young readers tweets, collaboration tags the medium is the message blog plagiarism horse-race coverage advertising the other longer Book Review....



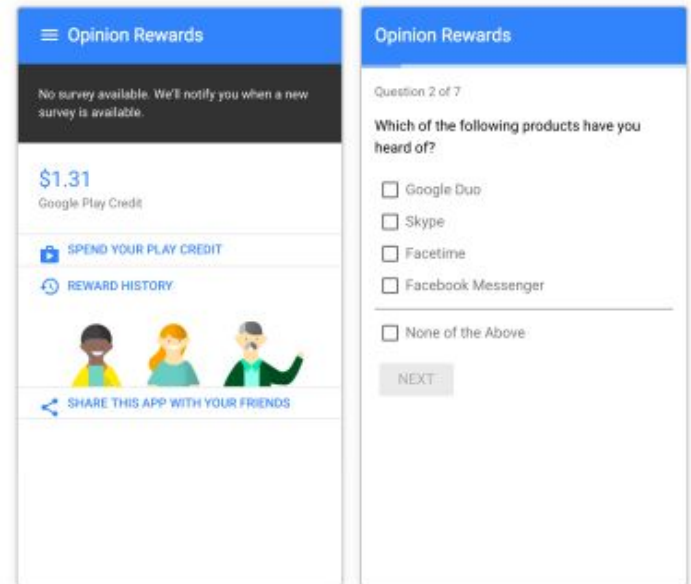
Privacy put the paper to bed Fuego news.me photo source: proimos/flickr

Please complete the following survey to access this premium content.

Question 5 of 5:

If you could replace one piece of furniture in your home today, what would you change?

Enter your answer



The mobile app Google Opinion Rewards.

Comparing Google Survey to Pew Survey

Report Custom Insights ²

Inferred Gender

Sum	Compare
Male	Female

Inferred Age

Sum	Compare
18-24	25-34
35-44	45-54
55-64	65+

Geography

All of the USA

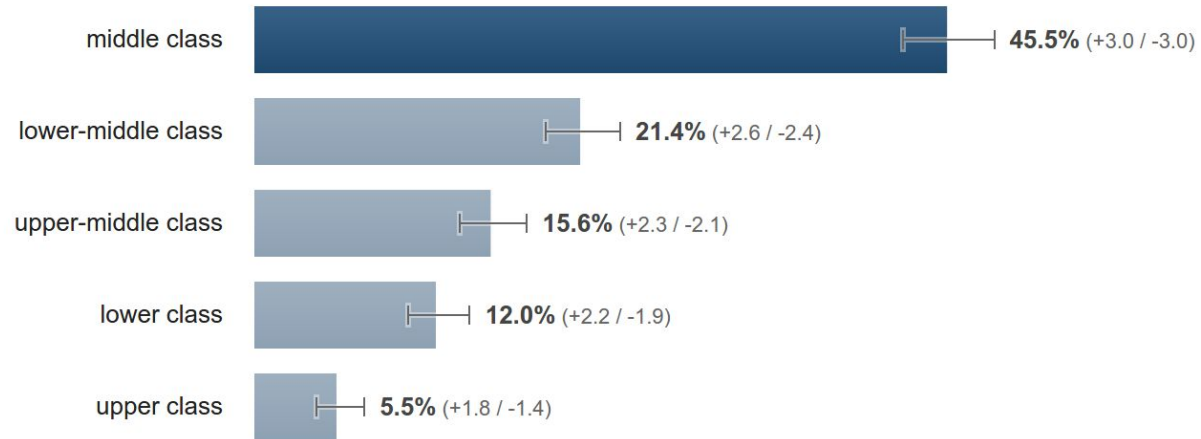
Sum	Compare
-----	---------

SINGLE ANSWER

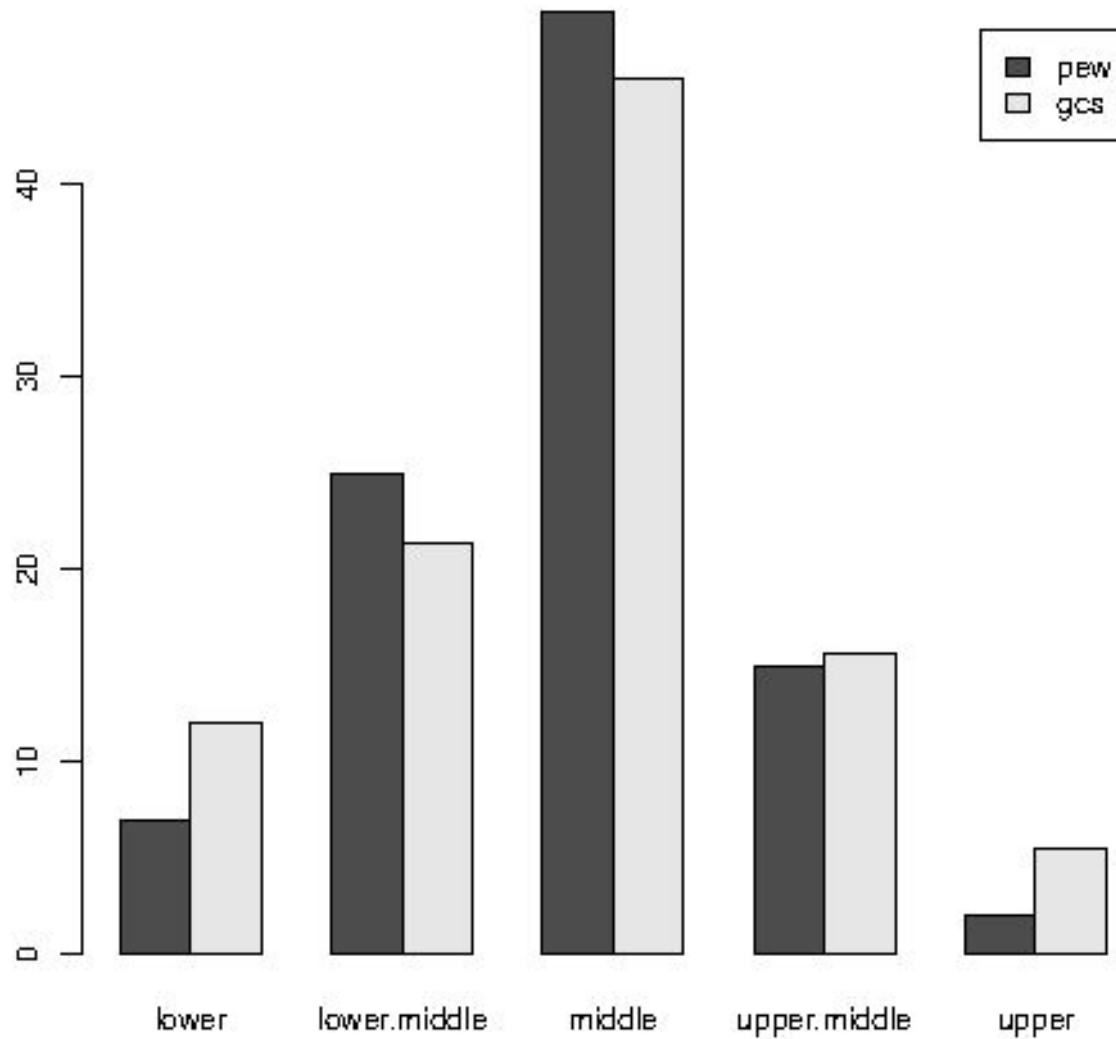
If you were asked to use one of these commonly used names for social classes, which would you say you belong in?

Results for respondents with demographics. Weighted by Age, Gender, Region. (1092 responses) ?

Order statistically significant. ?



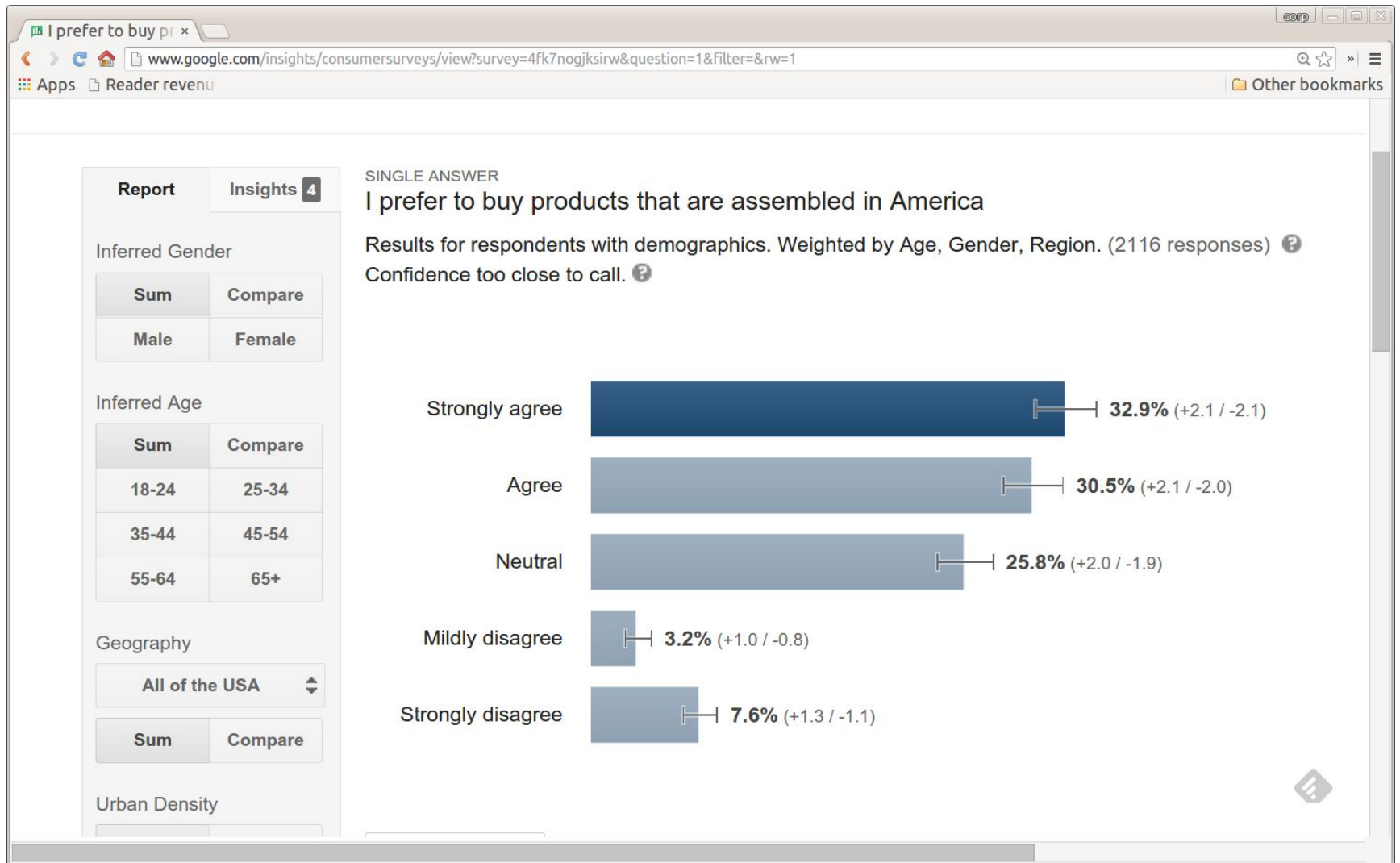
Comparisons of Pew and GCS



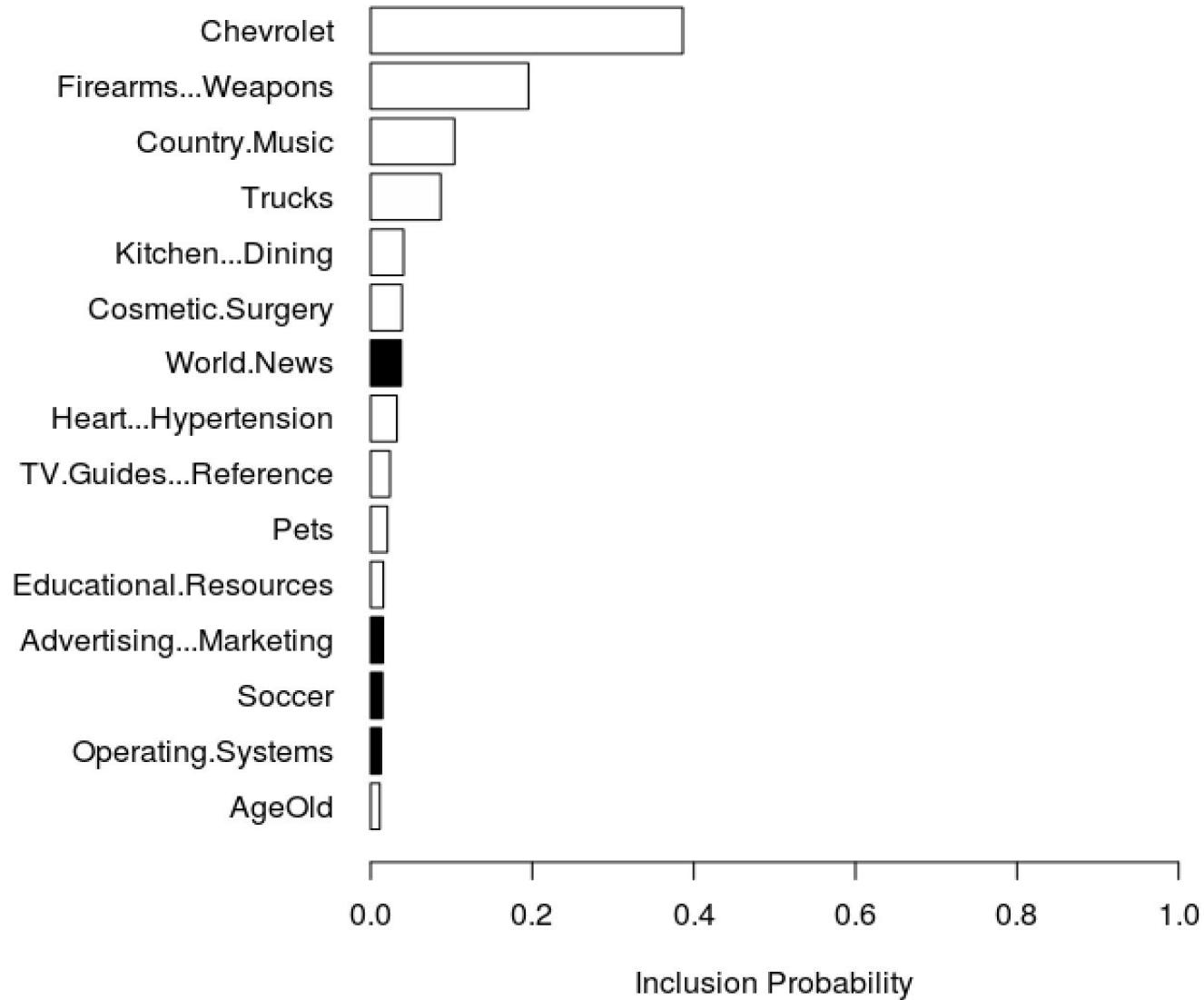
How this changes surveys

- Good
 - Anyone can do them
 - The cost is dramatically lower
 - Results come back in a few hours
 - Surveys can be replicated ... or not
 - You can measure sensitivity to wording or framing
- Bad
 - Not a random sample
 - But often works well

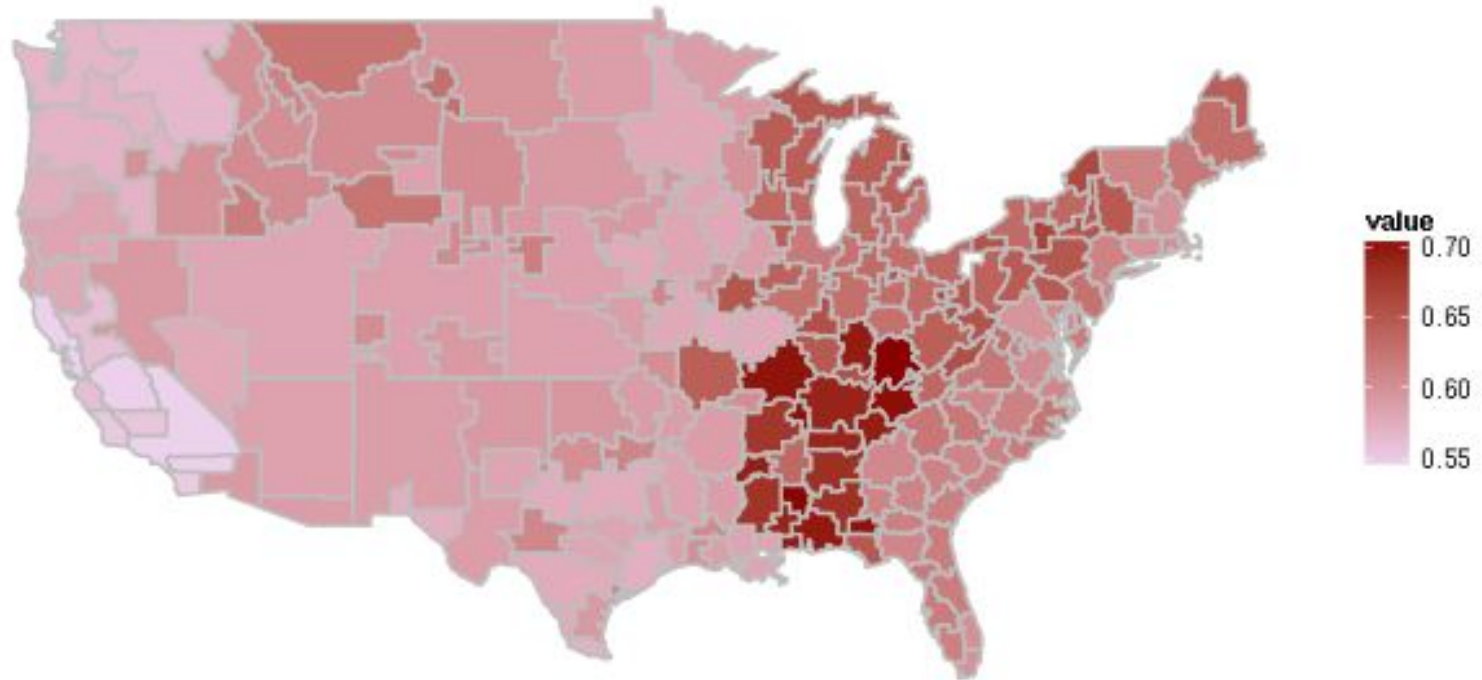
Attitude surveys: Assembled in America



Predictors of survey response



Assembled in America by DMA



Top and bottom cities' predicted score

Top

Kershaw, SC: 83.2 %

Summersville, WV: 82.8 %

Grundy, VA: 82.8 %

Chesnee, SC: 82.7 %

Duffield, VA: 82.5 %

Norton, VA: 82.3 %

Jonesville, VA: 82.2 %

Walnut Cove, NC: 82.2 %

Weston, WV: 82.2 %

Ennice, NC: 82.1 %

Bottom

Calipatria, CA: 40.2 %

Fremont, CA: 40.2 %

Mountain View, CA: 40.8 %

San Jose, CA: 41.4 %

Berkeley, CA: 41.4 %

Redmond, WA: 41.5 %

Glendale, CA: 41.5 %

Cupertino, CA: 41.6 %

Palo Alto, CA: 41.7 %

Daggett, CA: 41.9 %

Nowcasting surveys

How will you spend your stimulus check?

Because of the coronavirus crisis, most families will receive a rebate of \$1,200 for individuals, \$2,400 for couples, and \$500 for each child.

Thinking about your (family's) financial situation this year, will the rebate lead you mostly to increase spending, mostly to increase saving, or mostly to pay off debt?

1. Increase spending
2. Increase saving
3. Pay off debt
7. Don't expect to receive the rebate
9. Don't know

Rebates - April 3-6

	2020		2008	
	All	Response	All	Response
Pay off debt	25	52	53	54
Increase savings	16	34	27	27
Increase spending	6	13	19	19
Don't know	36		2	
Don't expect rebate	17		9	

Source: Google Survey; Sahm, Shapiro, and Slemrod: <http://www-personal.umich.edu/~shapiro/papers/aejep2012.pdf>. Note: data for 2008 rebates is from Michigan surveys in May and June.

Rebates - April 10-14

	2020		2008	
	All	Response	All	Response
Pay off debt	25	52	53	54
Increase savings	13	27	27	27
Increase spending	6	13	19	19
Don't know	40		2	
Don't expect rebate	16		9	

Source: Google Survey; Sahm, Shapiro, and Slemrod: <http://www-personal.umich.edu/~shapiro/papers/aejep2012.pdf>. Note: data for 2008 rebates is from Michigan surveys in May and June.

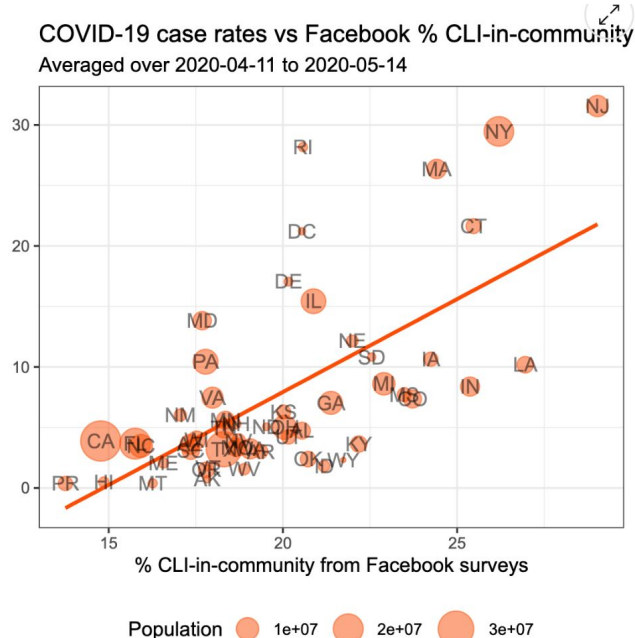
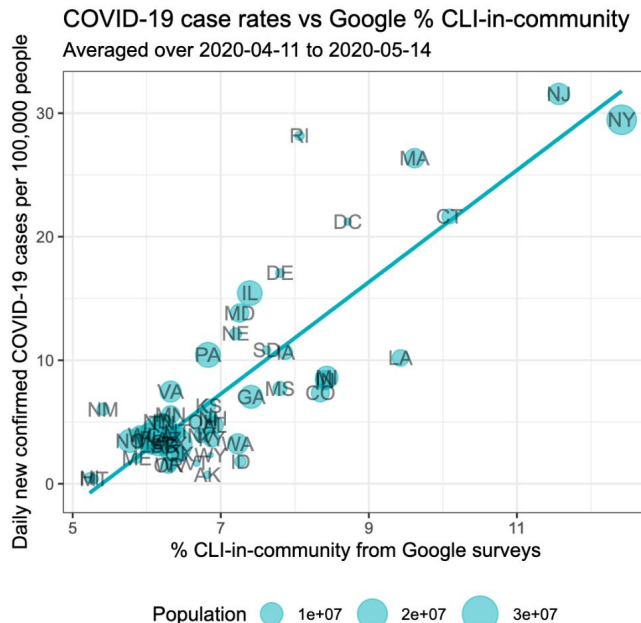
Google and Facebook Symptoms Survey

“From April 11 through May 14, we ran Google surveys in over 600 counties per day, with a target of at least 1,000 responses per county. The average number of responses per day was over 600,000, and at its peak, over 1.2 million! Survey question used was a “proxy questions” [Ryan Tibshirani \(CMU\)](#)

Original question: Do you or anyone in your household have a fever of at least 100 °F, along with cough, shortness of breath, or difficulty breathing?

Proxy question: Do you know of **someone in your community** who is sick with a fever, along with cough, shortness of breath, or difficulty breathing right now?

$r = .84$

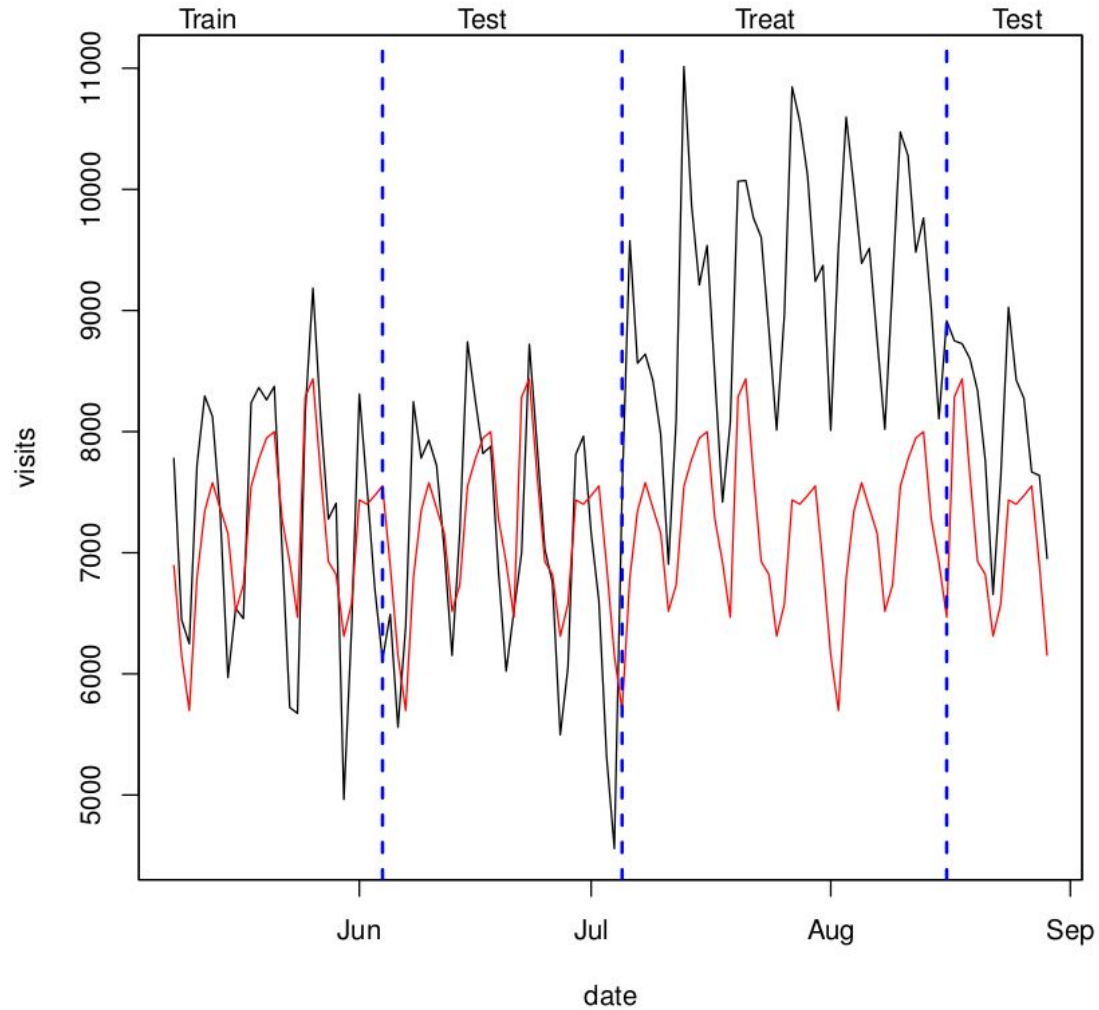


$r = .64$

Causal Inference

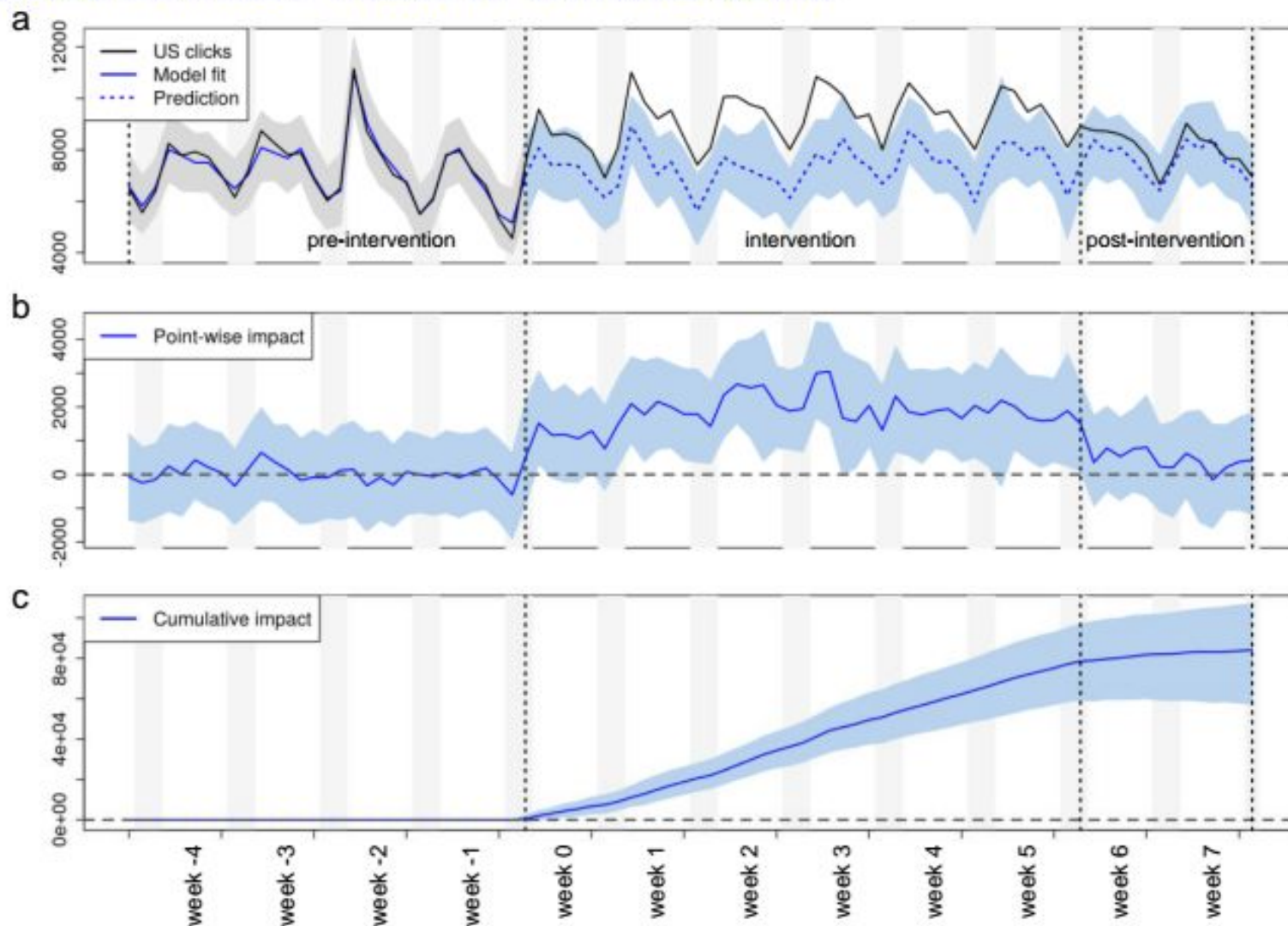
Causal inference

Can build model of counterfactual: train, test, treat, compare. Related to “synthetic control” in Abadie (2003, 2010).



Causal Impact

Google advertiser. Treated vs. Untreated regions



The End

Survey question about stimulus check?

Because of the coronavirus crisis, most families will receive a rebate of \$1,200 for individuals, \$2,400 for couples, and \$500 for each child. Thinking about your (family's) financial situation this year, will the rebate lead you mostly to increase spending, mostly to increase saving, or mostly to pay off debt?

1. Increase spending
2. Increase saving
3. Pay off debt
7. Don't expect to receive the rebate
9. Don't know

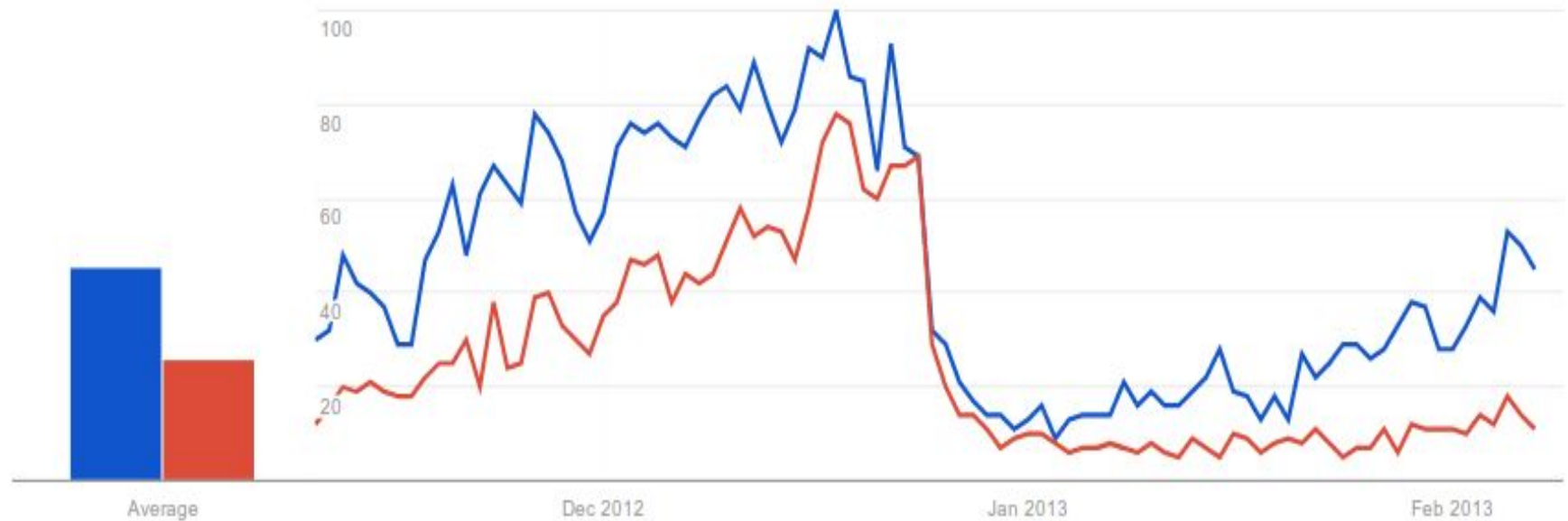


Gift for boyfriend v Gift for girlfriend

Interest over time ?

The number 100 represents the peak search volume

News headlines ? Forecast ?



For boyfriend For girlfriend

Gift for husband v Gift for wife

Web Search Interest: gift for husband, gift for wife. United States, Past 90 days.

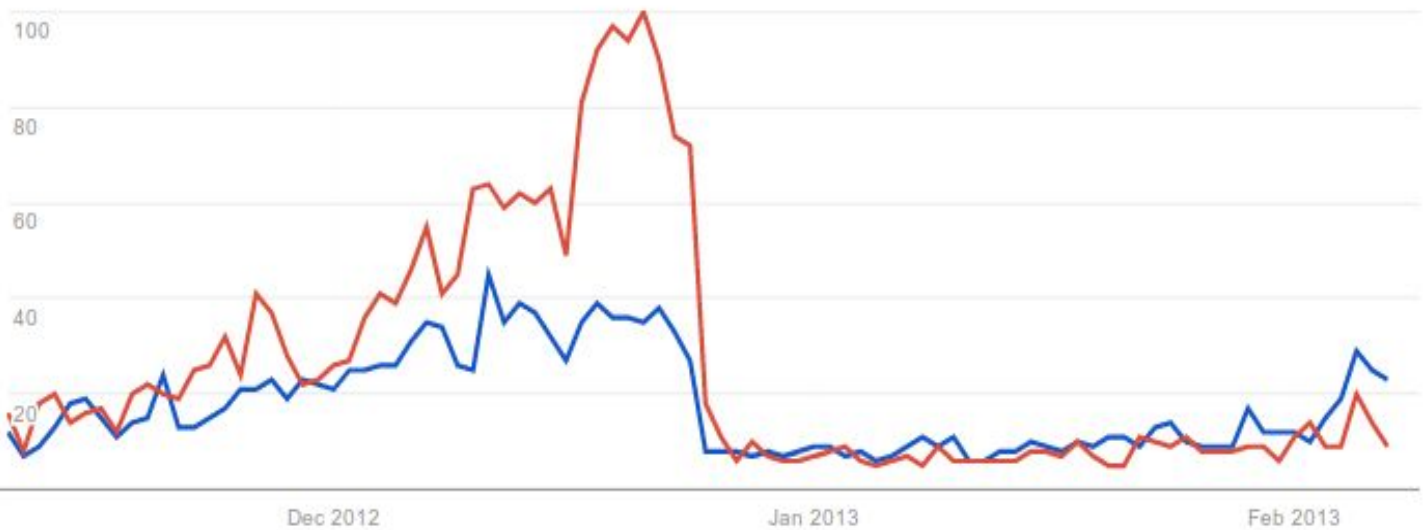


Interest over time ?

The number 100 represents the peak search volume

News headlines ?

Forecast ?



Average

Dec 2012

Jan 2013

Feb 2013

For husband
For wife

Searches for [civil war]

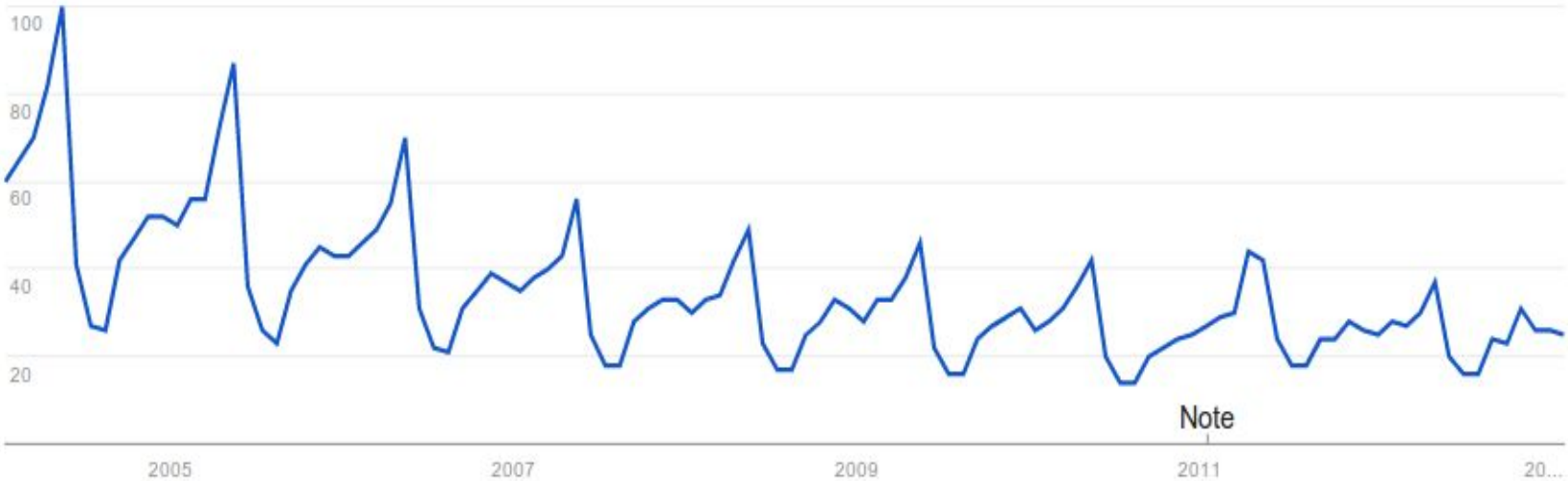
Web Search Interest: **civil war**. United States, 2004 - present.



Interest over time ?

The number 100 represents the peak search volume

News headlines Forecast ?



Note

Searches for [term paper]

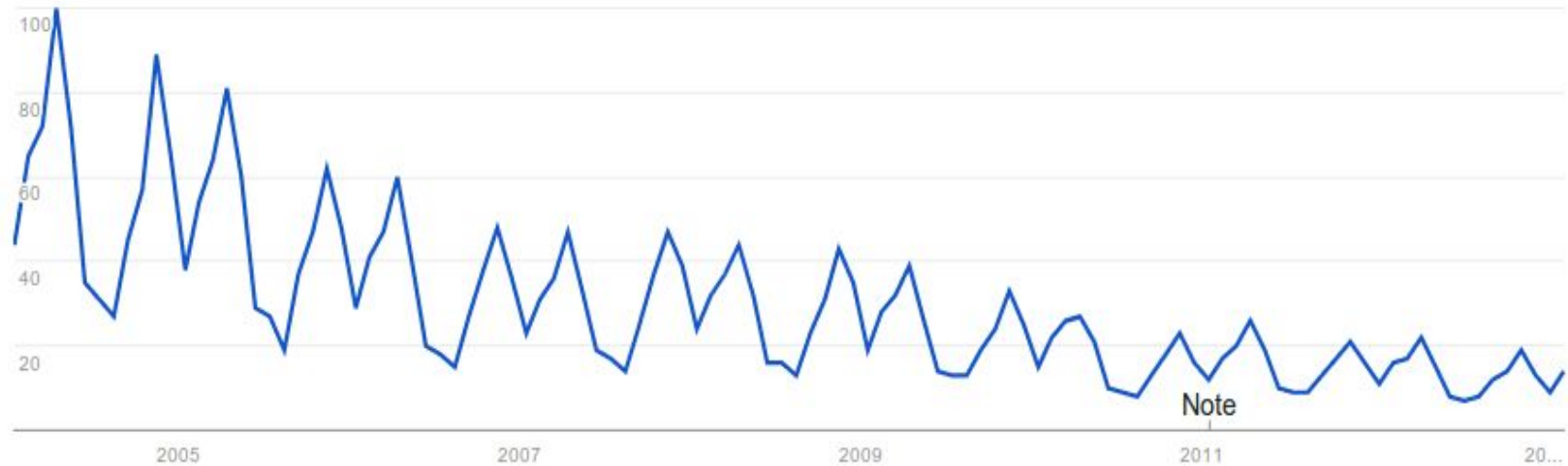
Web Search Interest: **term paper**. United States, 2004 - present.



Interest over time [?](#)

The number 100 represents the peak search volume

News headlines Forecast [?](#)



Note

Embed