23) Google Trends, Google Correlate, and Google Surveys

Vitor Kamada

August 2019

Reference

Davidowitz & Varian (2014). **A Hands-on Guide to Google Data**

https://trends.google.com/trends/?geo=US

https://www.google.com/trends/correlate

https://surveys.google.com/

Google Trends

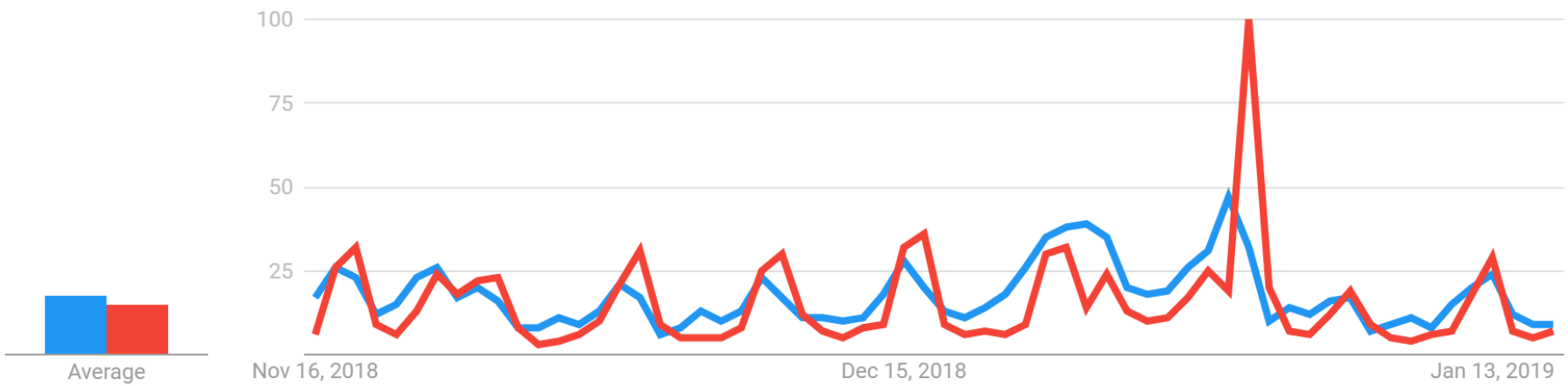
martini recipe
Search term

hangover cure
Search term

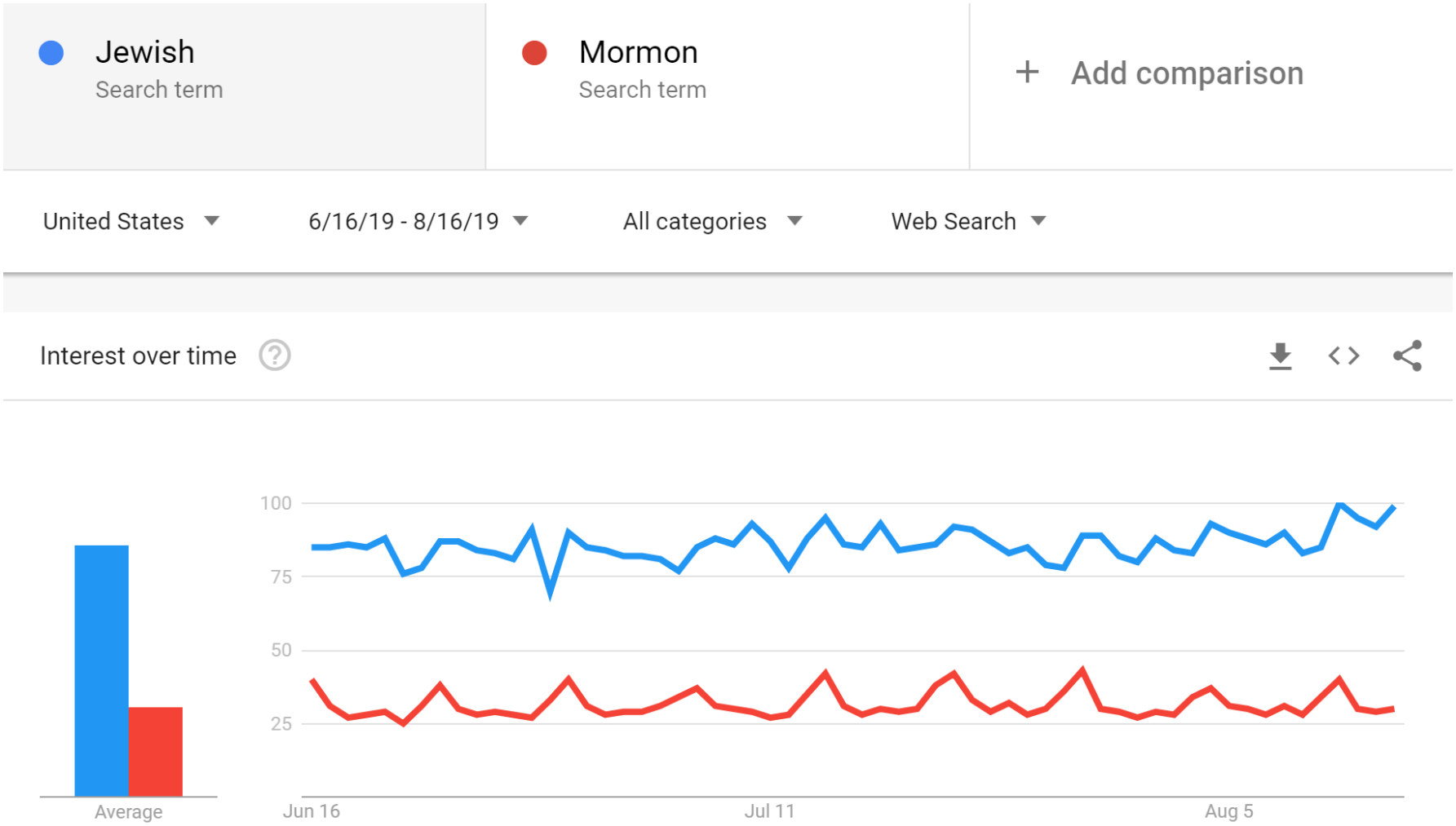
+ Add comparison

United States | 11/16/18 - 1/16/19 | All categories | Web Search

Interest over time



Jewish has 3.2 times more searches than Mormon

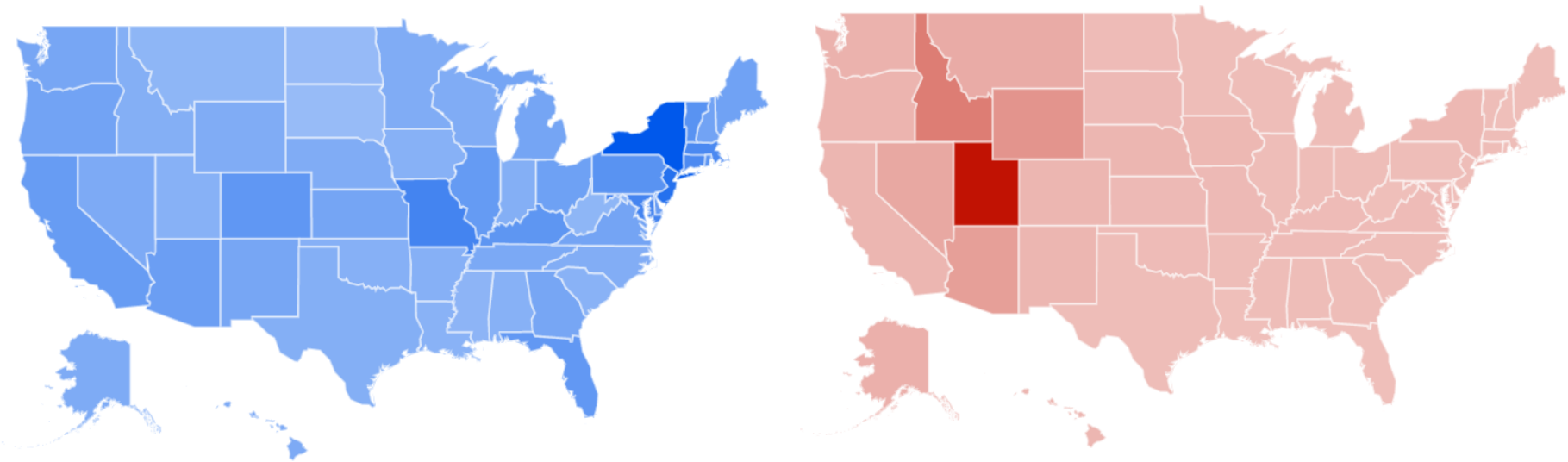


Jewish population is not 3.2 times larger than the Mormon

Comparison Jewish and Mormon

Data are normalized

The highest-scoring state has 100



Map correlates ($R^2 = 0.88$) with the proportion of a state's population that is Jewish

Daily Data vs State Data

$$\text{Range} = [0, 100]$$

100 vs 50 means twice more searches

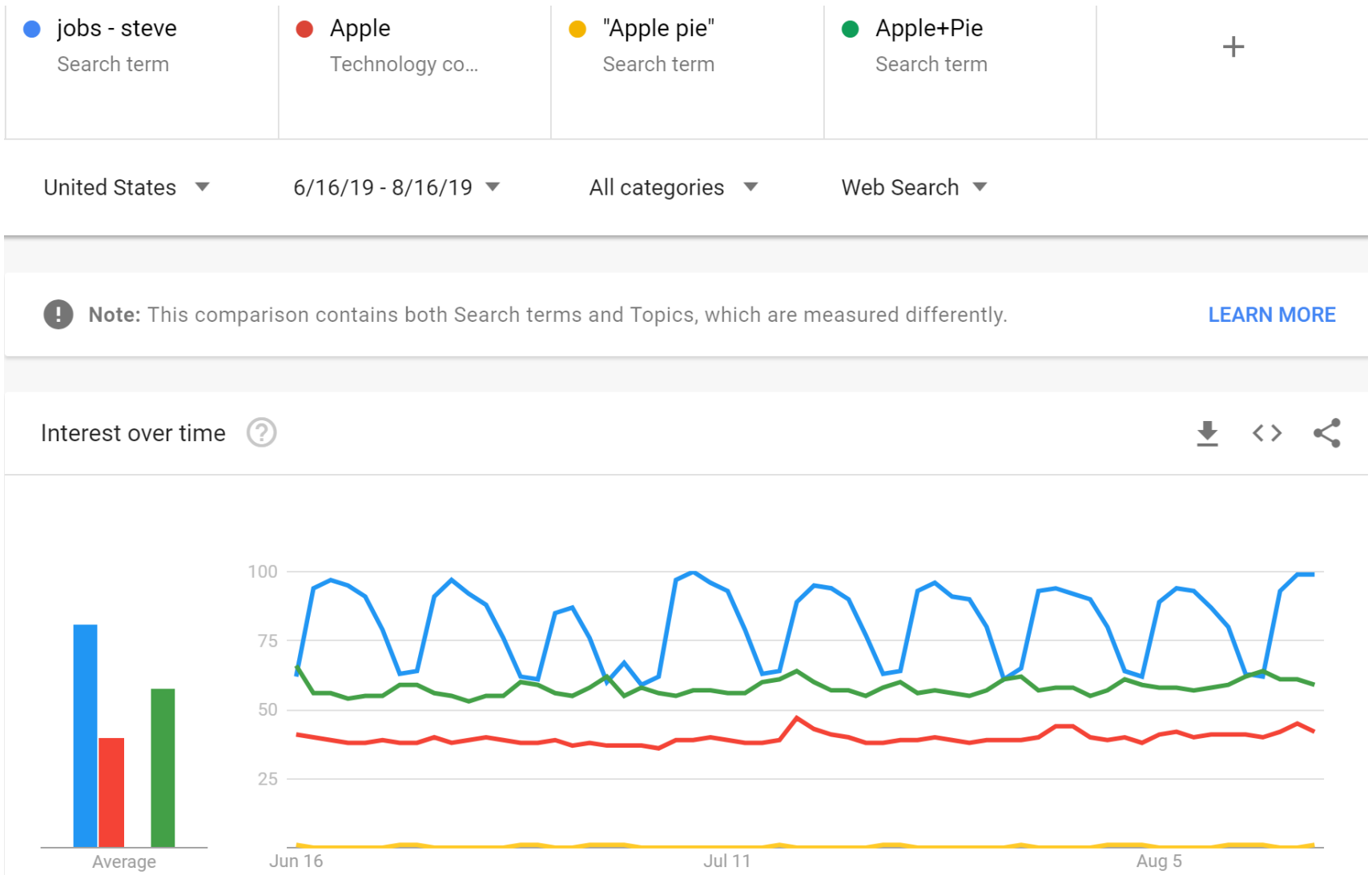
3	Day	martini recipe	hangover cure: (United States)
4	11/16/2018	17	6
5	11/17/2018	26	26
6	11/18/2018	23	32
7	11/19/2018	12	9
8	11/20/2018	15	6
9	11/21/2018	23	13
10	11/22/2018	26	24

3	Region	Jewish: (6/16/19 - 8/16/19)
4	New York	100
5	District of Colum	82
6	New Jersey	79
7	Maryland	62
8	Missouri	62
9	Massachusetts	58
10	Connecticut	57
11	Pennsylvania	50
12	Vermont	50
13	Rhode Island	47
14	Kentucky	47
15	Colorado	46
16	Florida	45

Data is cached each day

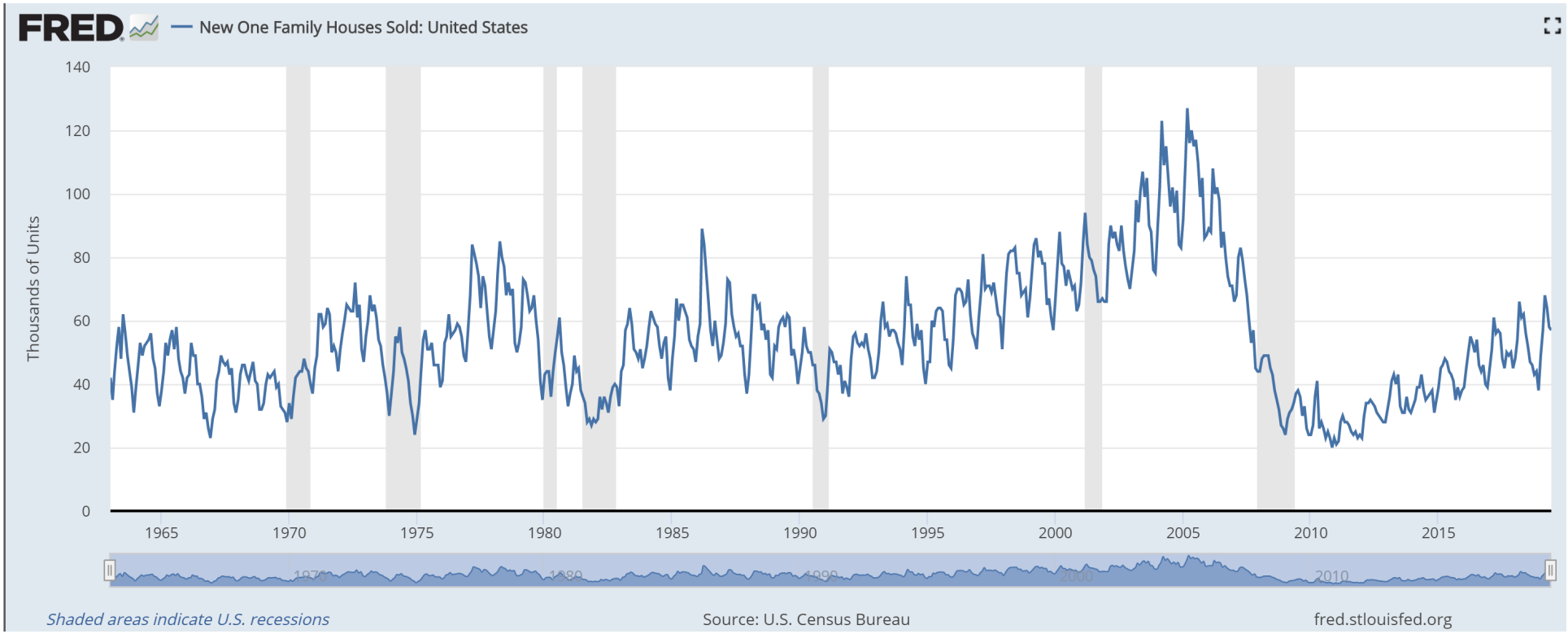
Average multiple samples: get from different days

Match Types



A space means "and"

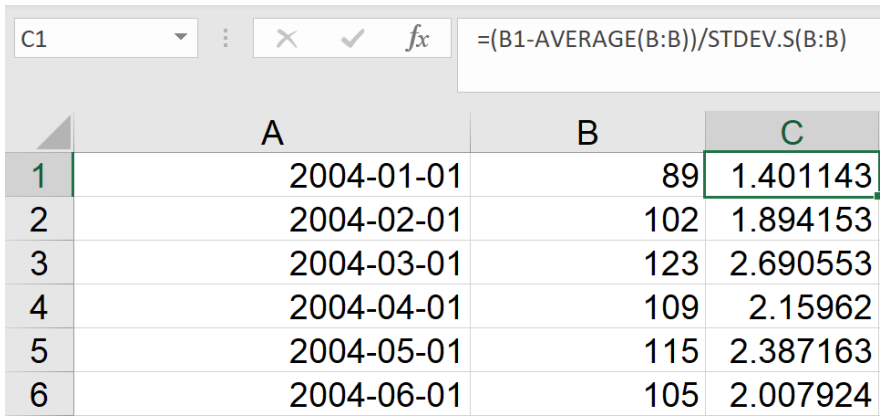
http://research.stlouisfed.org/fred2/series/HSN1FNS



Normalized Data

Right click -> Format Cells -> Custom -> Enter yyyy-dd-mm

1	FRED Graph Observations	
2	Federal Reserve Economic Data	
3	Link: https://fred.stlouisfed.org	
4	Help: https://fred.stlouisfed.org/help-faq	
5	Economic Research Division	
6	Federal Reserve Bank of St. Louis	
7		
8	HSN1FNSA	New One Family Houses S
9		
10	Frequency: Monthly	
11	observation_date	HSN1FNSA
12	1963-01-01	42
13	1963-02-01	35
14	1963-03-01	44
15	1963-04-01	52
16	1963-05-01	58
17	1963-06-01	48

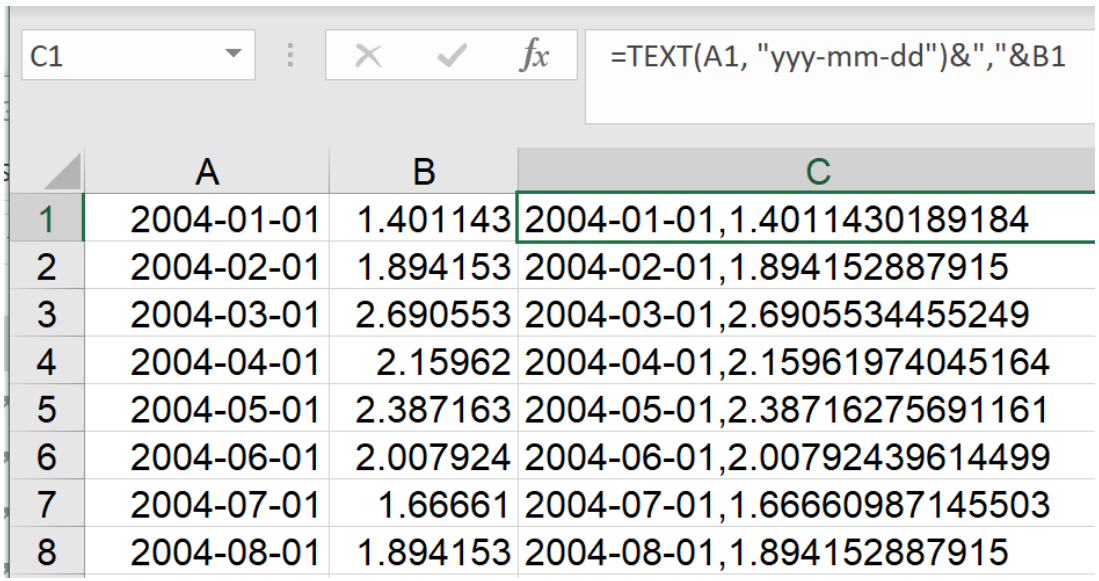


	A	B	C
1	2004-01-01	89	1.401143
2	2004-02-01	102	1.894153
3	2004-03-01	123	2.690553
4	2004-04-01	109	2.15962
5	2004-05-01	115	2.387163
6	2004-06-01	105	2.007924

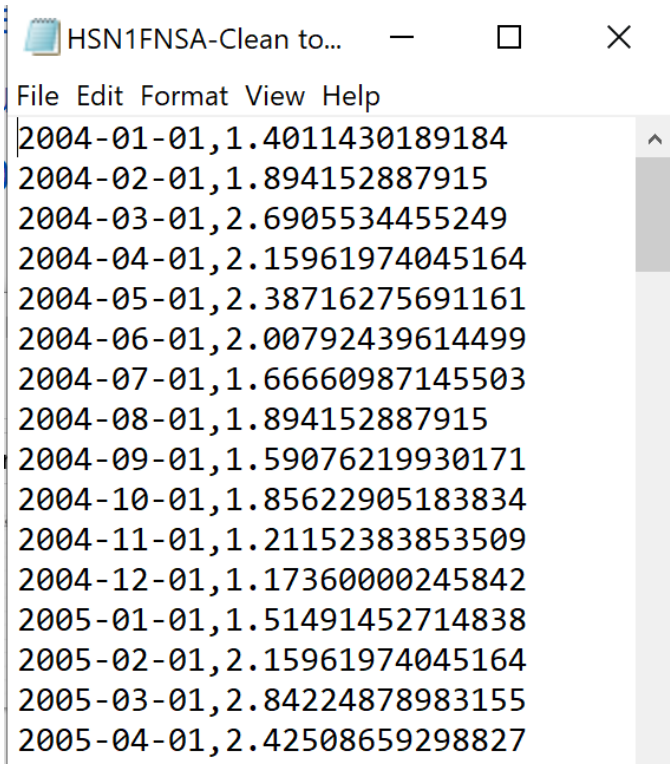
=(B1-AVERAGE(B:B))/STDEV.S(B:B)

Copy Column "C" to Text file

=TEXT(A1, "yyy-mm-dd")&","&B1



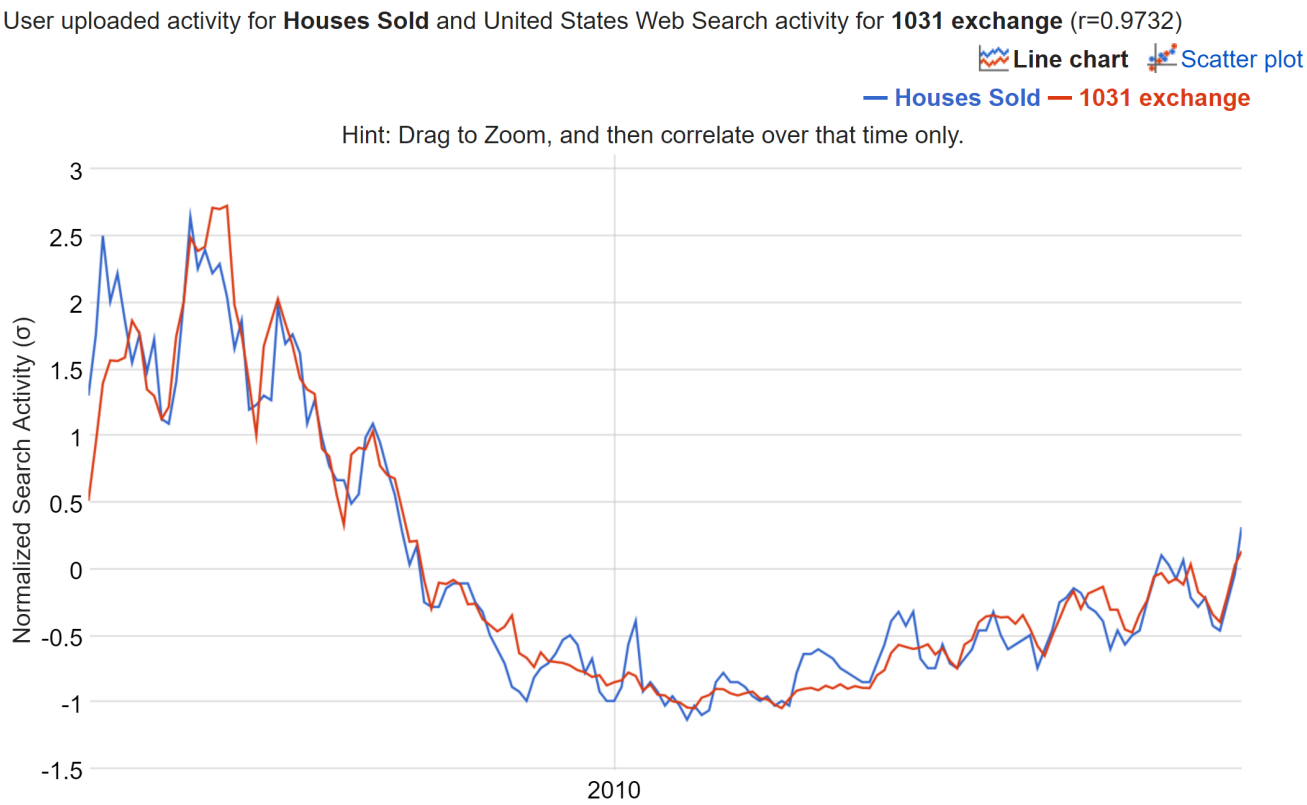
	A	B	C
1	2004-01-01	1.401143	2004-01-01,1.4011430189184
2	2004-02-01	1.894153	2004-02-01,1.894152887915
3	2004-03-01	2.690553	2004-03-01,2.6905534455249
4	2004-04-01	2.15962	2004-04-01,2.15961974045164
5	2004-05-01	2.387163	2004-05-01,2.38716275691161
6	2004-06-01	2.007924	2004-06-01,2.00792439614499
7	2004-07-01	1.66661	2004-07-01,1.66660987145503
8	2004-08-01	1.894153	2004-08-01,1.894152887915



Google Correlate

Correlated with **Houses Sold**

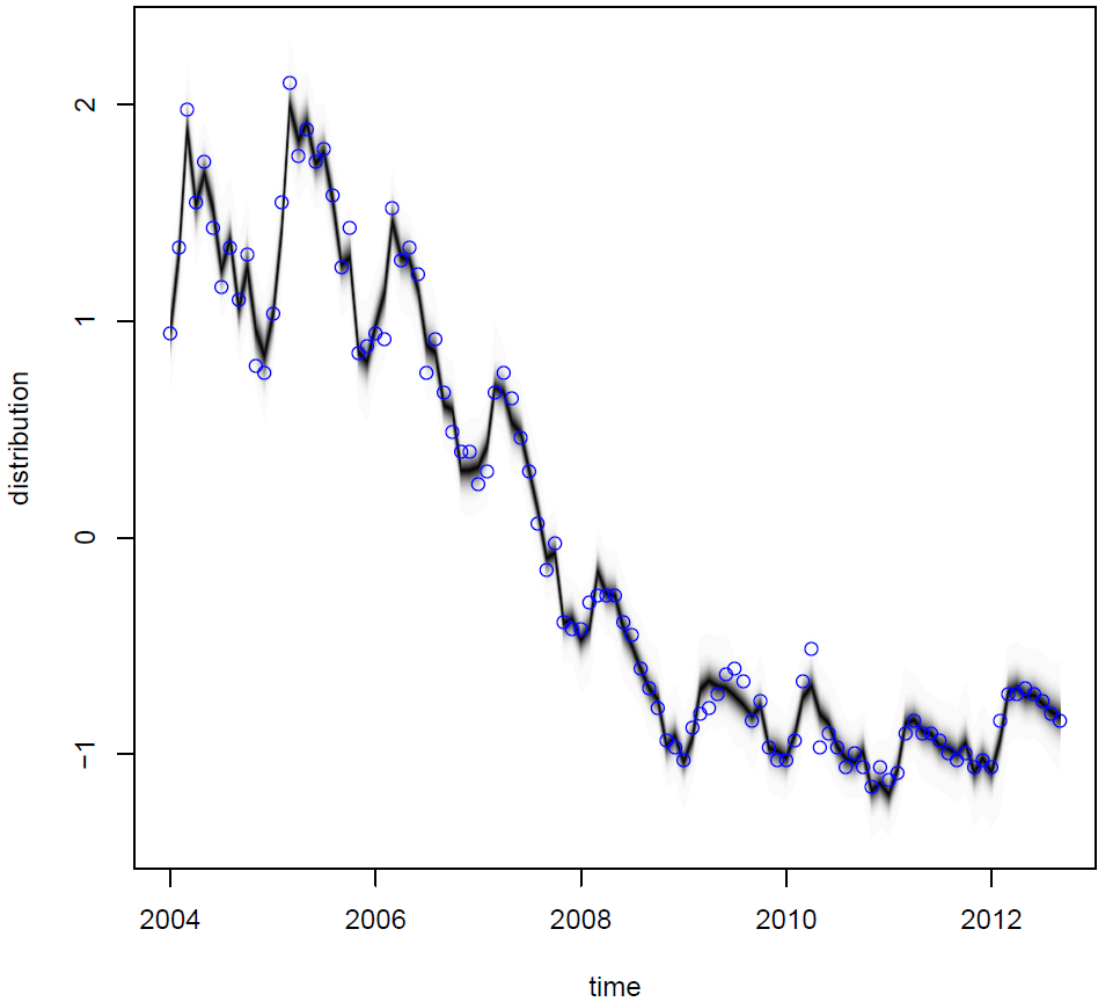
- 0.9732 1031 exchange
- 0.9696 1031
- 0.9688 real estate test
- 0.9686 exhaust sound
- 0.9650 new home construction
- 0.9642 new housing development
- 0.9640 home communities
- 0.9638 appreciation rate
- 0.9611 manufactured
- 0.9611 irs 1031



Bayesian Structural Time Series

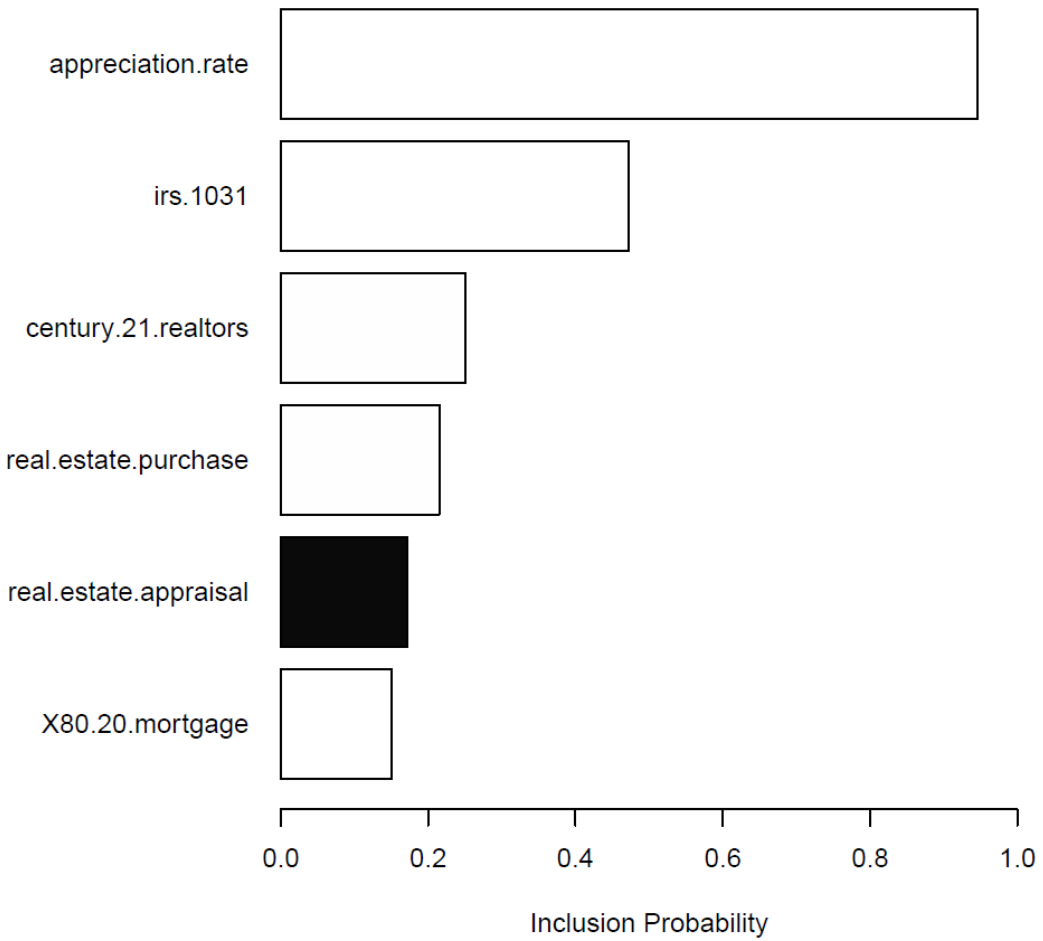
```
# read data from correlate and make it a zoo time series
dat <- read.csv("Data/econ-HSN1FNSA.csv")
y <- zoo(dat[,2],as.Date(dat[,1]))
# use correlates as possible predictors
x <- dat[,3:ncol(dat)]
# set a few parameters
numiter <- 4000
npred <- 5
# describe state space model consisting of
# trend and seasonal components
ss <- AddLocalLinearTrend(list(),y)
ss <- AddSeasonal(ss,y,nseasons=12)
# estimate the model
model <- bsts(y~.,state.specification=ss,data=x,
niter=numiter,expected.model.size=npred,ping=0,seed=123)
```

Posterior Distribution and Actual Outcome

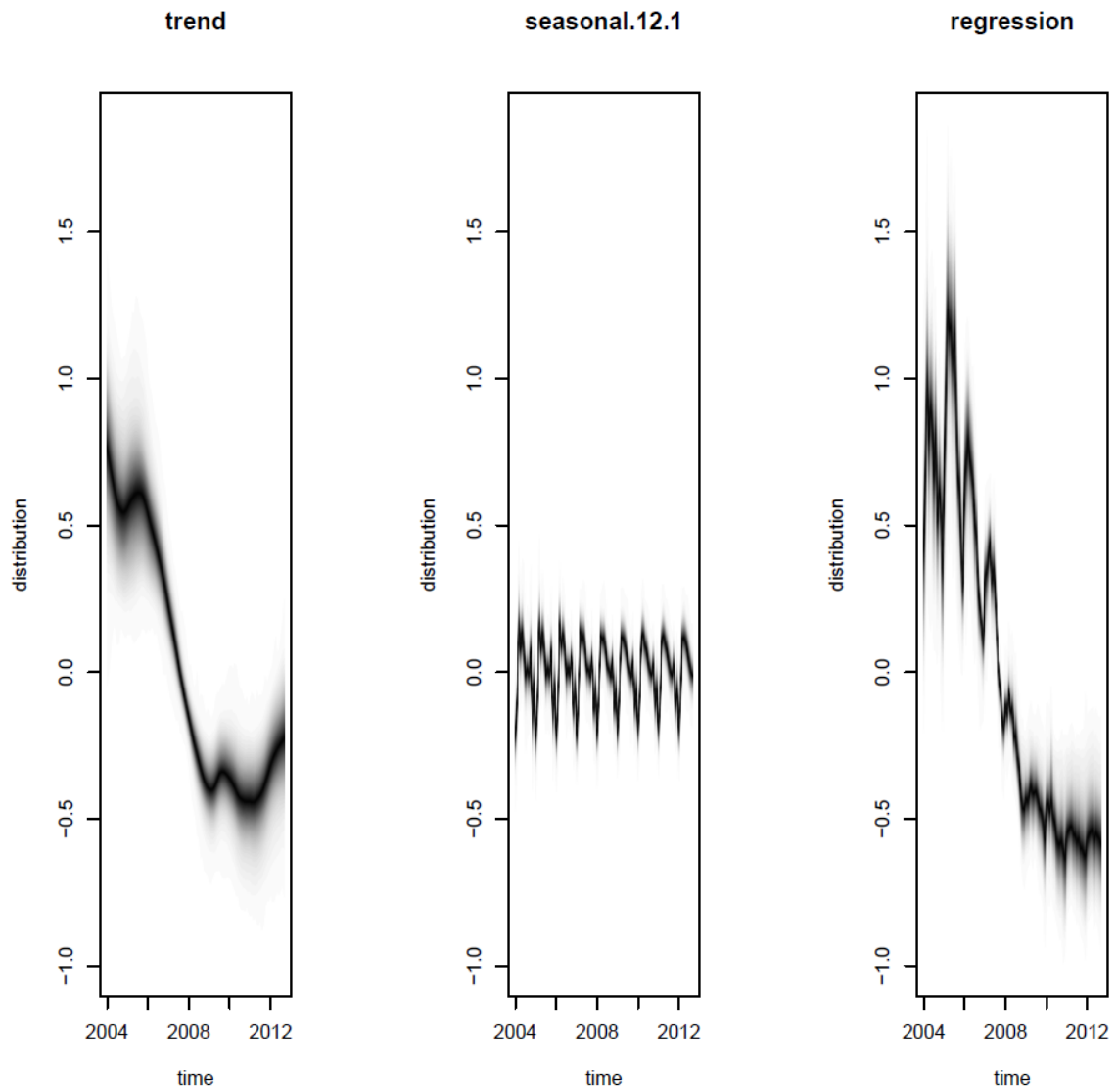


```
plot(model,"coef",inc=.15)
```

Top Predictors



```
plot(model,"comp")
```



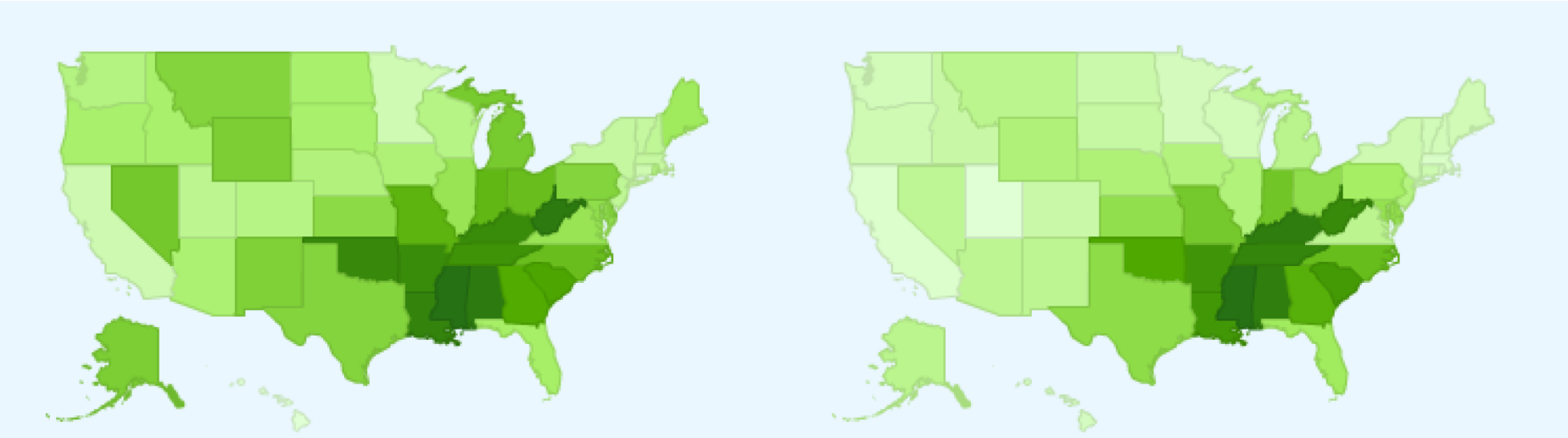
Life Expectancy by State

urlk.org/other/state-indicator/life-expectancy/

Abnormally Short Lifespans: put a minus sign in front the entries in the CSV file

User uploaded activity for **negative life expectancy** and United States Web Search activity for **blood pressure medicine** (r=0.9092)

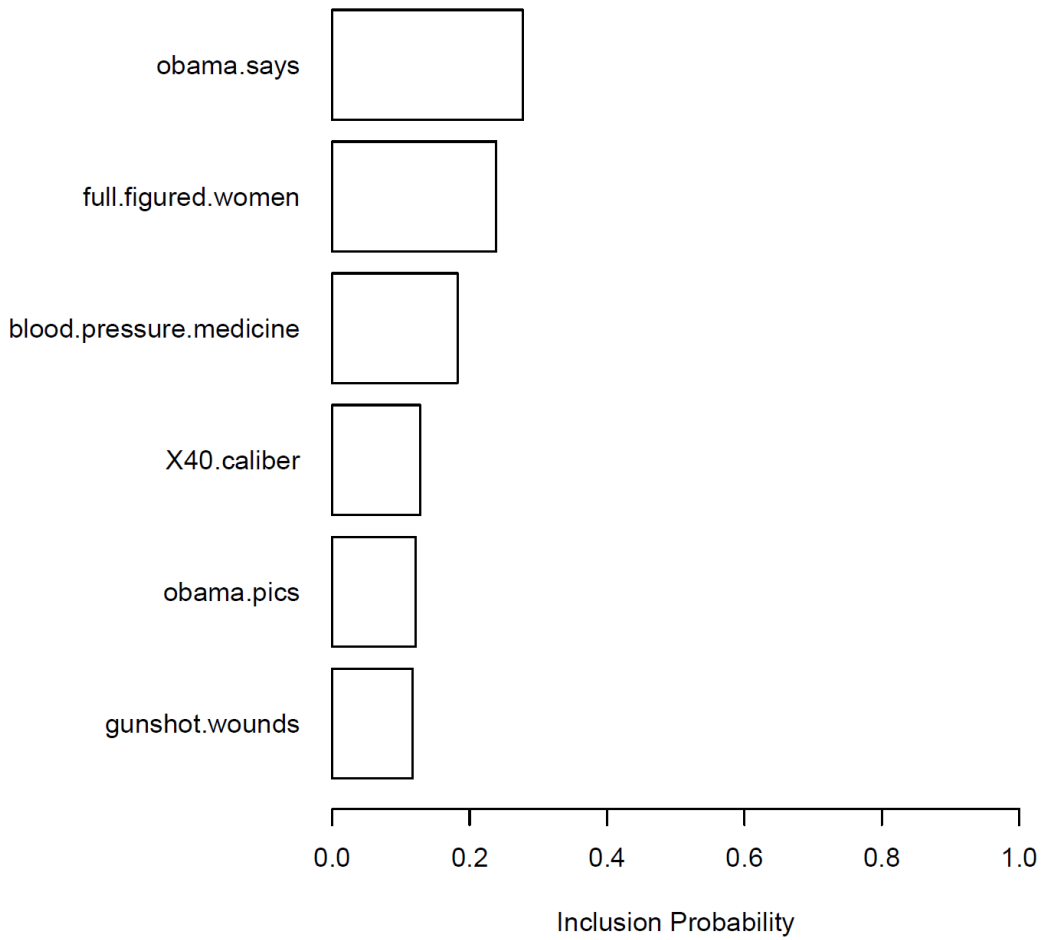
State maps Scatter plot



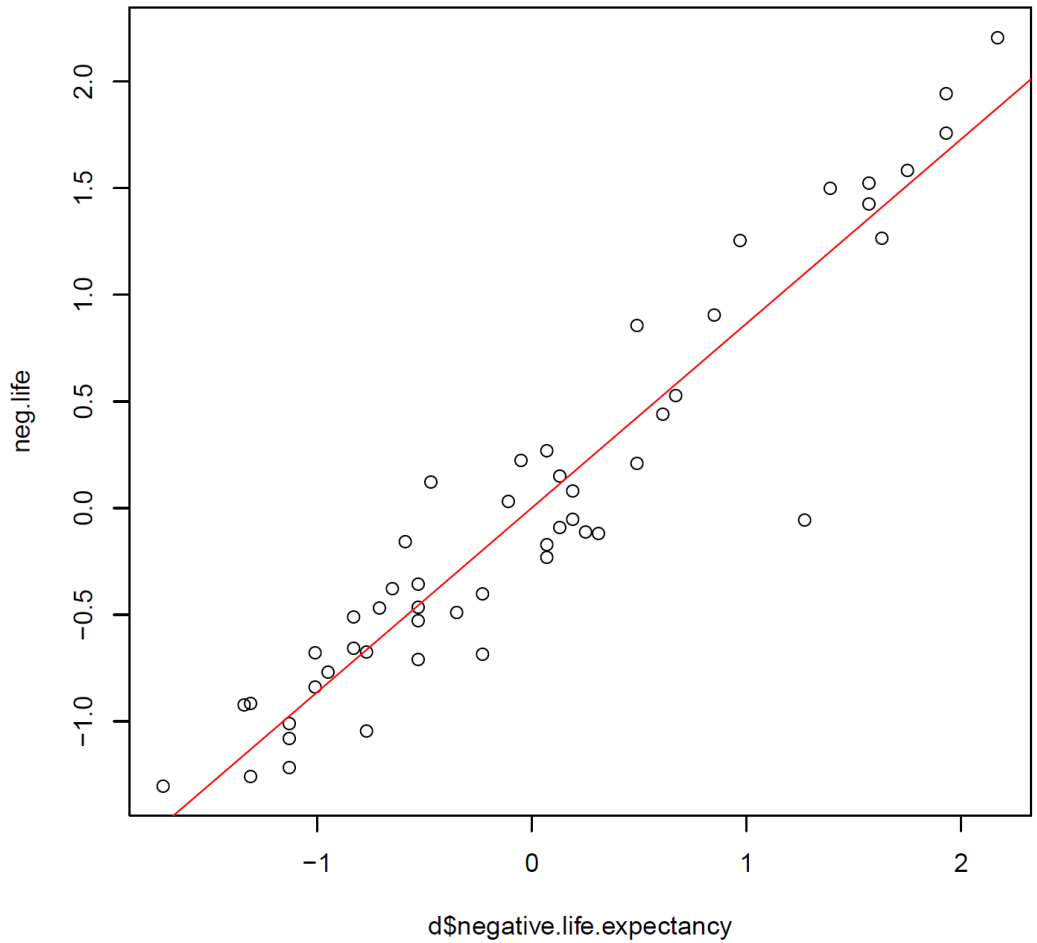
```r
# library(BoomSpikeSlab)
dat <- read.csv("Data/correlate-negative_life_expectancy.csv")
d <- dat[,-1]
reg <- lm.spike(negative.life.expectancy ~ .,niter=4000,data=d)
plot(reg,inc=.10)
```

Correlated with **negative life expectancy**

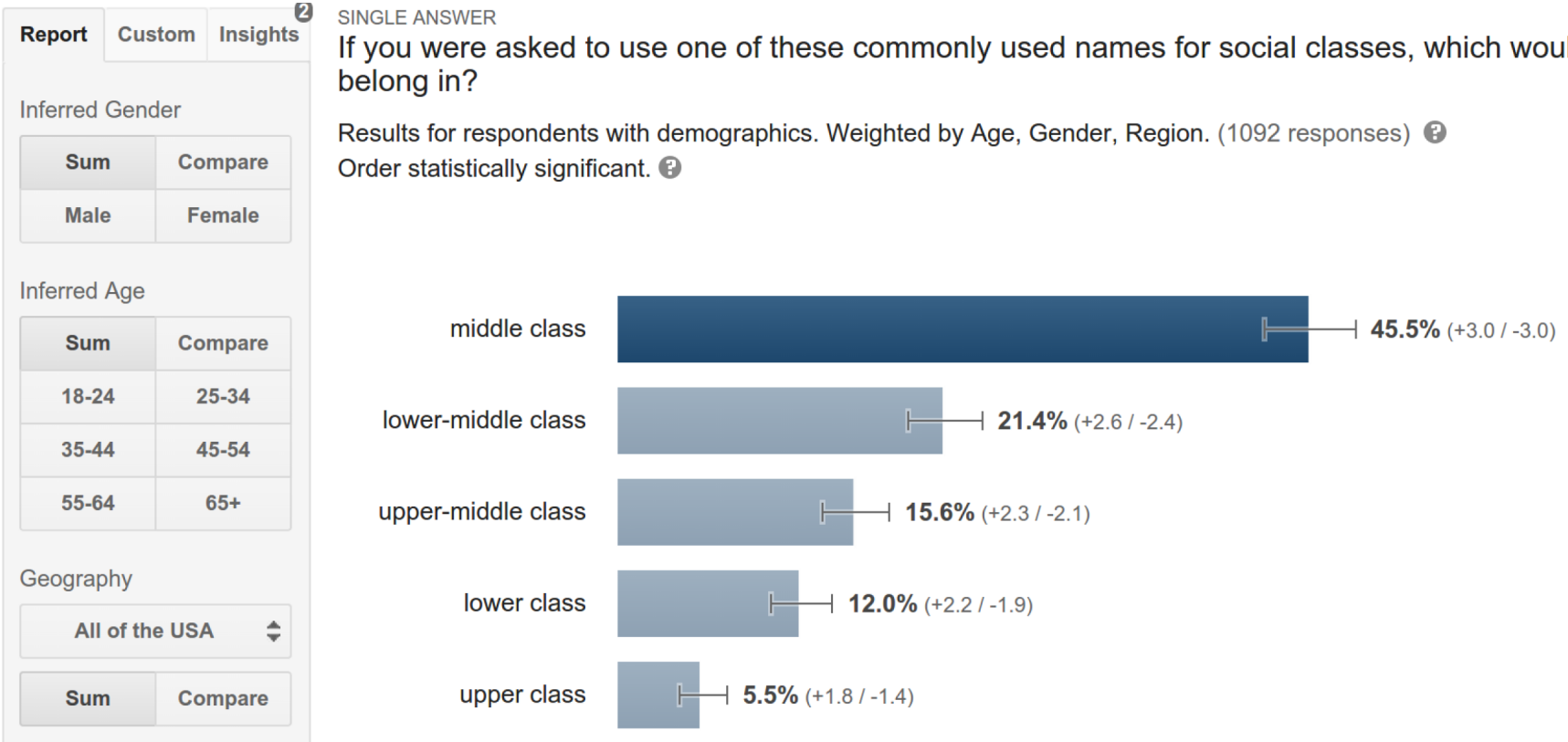
0.9092 blood pressure medicine
0.8985 obama a
0.8978 major payne
0.8975 against obama
0.8936 king james bible online
0.8935 about obama
0.8928 prescription medicine
0.8920 40 caliber
0.8919 .38 revolver
0.8916 reprobate
0.8911 performance track
0.8910 lost books of the bible
0.8905 glock 40 cal
0.8898 lost books
0.8896 the mark of the beast
0.8892 obama says
0.8891 obama said
0.8882 sodom and
0.8882 the antichrist



```r
temp <- predict(reg,newdata=d)
neg.life <- rowMeans(temp)
plot(neg.life~d$negative.life.expectancy)
reg1 <- lm(neg.life~d$negative.life.expectancy)
abline(reg1,col=2)
```



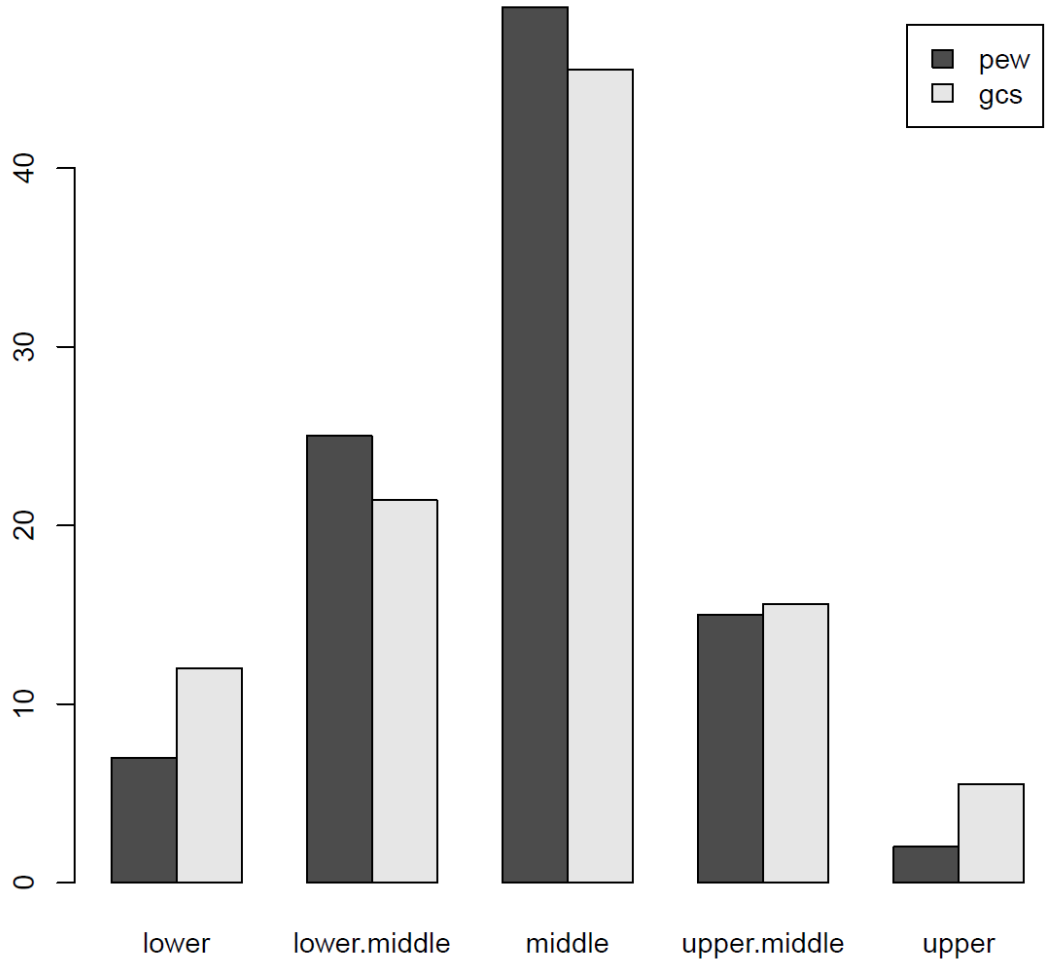
Google Surveys (GS)



City: IP address **Age and gender**: web site visits

Income: location and Census data

Pew and GCS(GS) Answers



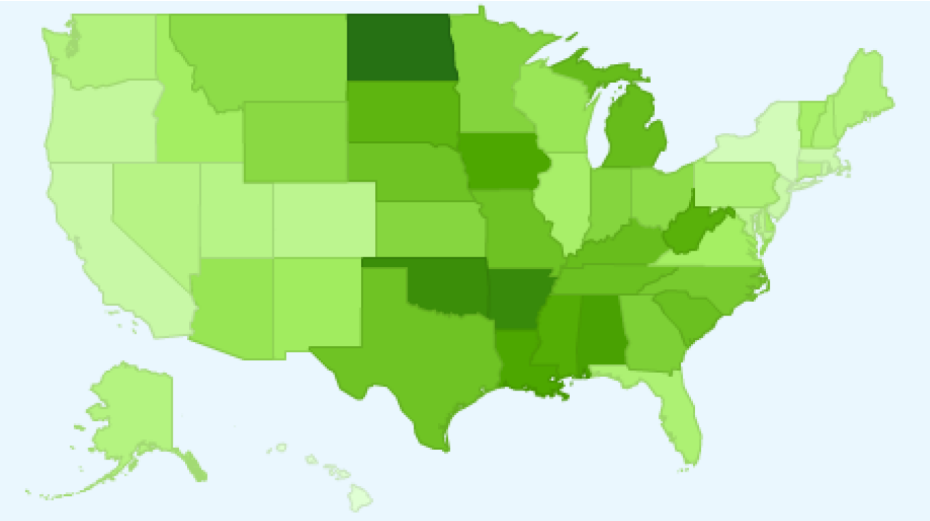
GCS surveys similar to surveys published by reputable organizations

Pricing of Google Surveys

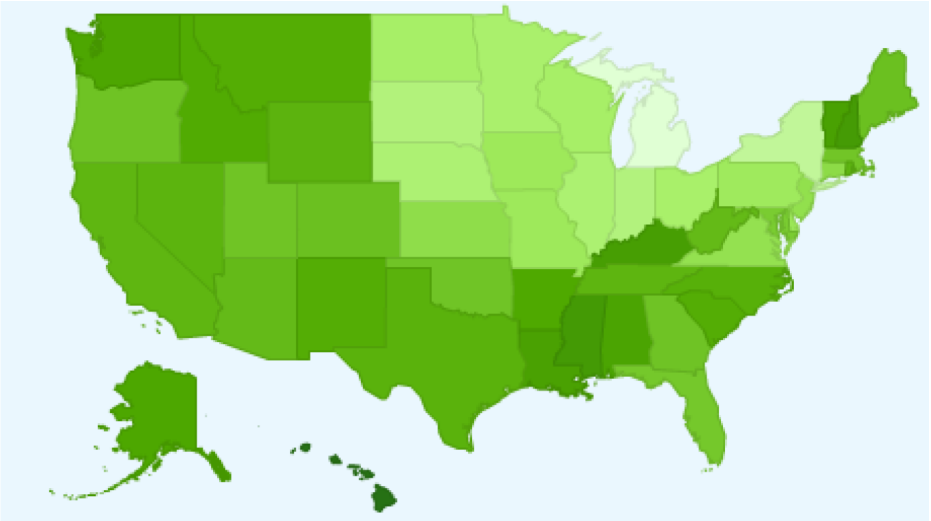
Currency	General population or Android-smartphone users		Age, gender, or location targeted		Postal-code targeted*		User-list targeted*	
	1	2-10	1	2-10	1	2-10	1	2-10
USD	0.10	1.00	0.15	1.50	0.60	6.00	1.00	10.00
AUD	0.13	1.30	0.20	2.00	0.75	7.50	1.25	12.50
CAD	0.12	1.20	0.20	2.00	0.75	7.50	1.25	12.50
CHF	0.10	1.00	0.14	1.40	0.58	5.80	0.96	9.60
EUR	0.09	0.90	0.13	1.30	0.51	5.10	0.86	8.60
GBP	0.08	0.80	0.11	1.10	0.46	4.60	0.76	7.60
ILS	0.32	3.20	0.47	4.70	1.89	18.90	3.15	31.50
JPY	11.20	112.00	16.70	167.00	67.00	670.00	111.60	1116.00

I prefer to buy products that are assembled in America [Agree or disagree]

chevrolet



toyota



Most Responsive: Kernshaw, SC; Summersville, WV; Grundy, VA; Chesnee, SC

Least Responsive: Calipatria, CA;Fremont, CA; Mountain View, CA; San Jose, CA

Predictors for “assembled in America" question