Image source: https://selfimagemedia.com/2017-digital-marketing-trends-everyone-talking/

Analyzing Google Trends Data

Final Research Project

—

Chirag Vithalani
CUNY (City University of New York)
New York, NY 10017
Advisor / Guide: Honey Berk

Contents

Trends come and go,
Not many of them ever last for long,
It's just the majority thing at the time,
Where you find more just joining the line..

Source : http://hellopoetry.com/poem/485153/trends/

Introduction to Google Trends

Have you ever wondered what internet users search for? If you manage an e-commerce website, this question has likely crossed your mind as you work to attract visits and make sales. Trillions of searches are done every year – and data about them can offer valuable insights. A simple tool that provides this wealth of data is none other than Google Trends. This handy data explorer offers both real time and non-real time information on what users around the world are searching for through Google's search engine – for free! Whether it's a news story or an upcoming sports event, Google Trends allows us to see trending topics as well as data on topics and keywords you're interested in. Human thoughts are complex propositions, that way this is a useful tool which tells users how frequently any word or phrase has been searched in different parts of the world at different time durations.

Google Trends data is an unbiased sample of our Google search data. It's anonymized (no one is personally identified), categorized (determining the topic for a search query) and aggregated (grouped together). This allows us to measure interest in a particular topic across search, from around the globe, right down to city-level geography.

The use of Google Trends is fairly simple and very intuitive. We start by entering a search term or topic in the query box then we can select from several filtering options:

- Region. Allows us to define your search by worldwide, country, state or even city/metro area.
- Time Frame. Lets us select a variety of predefined time frames (last seven days, one year, one month, etc.) or set a custom period. The data goes back to 2004, which makes this an amazingly interesting source as it contains the rise of the Internet, the evolution and decline of Myspace, and other fun historical changes.
- Categories. We can limit the terms and search volume to a certain category. Having this dimension allows us to look at specific trends and discover new themes/searches.
- Engines. This option lets us choose between Web Search, News Search, YouTube Search, and Shopping search, which offer great flexibility depending on the brand and vertical. This also allows us to focus on the right intent — learning through video search or buying through shopping search, for example.

The results are broken out into two separate graphs: historic trending (interest over time) and localized (Interest by subregion) behavior. Google trends can reveal many things about collectively us as a society. I say collectively because Google Trends will provide data only when lots of people make the same search otherwise it shows the message “Hmm, your search doesn't have enough data to show here.”

A steadily increasing number of Internet users visit websites of search engines every day. Each query request can be seen as an individual vote: using search engines, users leave information about their interests codified as search terms. Thus, search engines can collect our interests on the smallest possible scale—the scale of individual requests. On larger time scales, our interest forms trends. Aggregated search volume data can be used for uncovering such trends that affect our life on large scales.

Limitations

Some limitations of Google Trends are it captures data only where of course, Internet available and more importantly Google is available. For instance, Google is unavailable in China despite being the country with highest population and internet availability.

Secondly, Google Trends is not greatly useful where English is not first preference language. For instance, India as a country is so diverse and multilingual. There is no language called “Indian”. India speaks hundreds of languages and dialects. Some are extinct, while some are still in use with considerable speakers. Officially India has 23 constitutionally recognized official languages. So in this case, Google Trends will capture data correctly only for English speaking users.

Objective

Below is a report on “Analysis of Google Trends data”. I gathered Google Trends data different ways and then tried to get meaning out it. Initially, Google search data didn't seem to be a proper source of information for "serious" research. But as we will see, trails we leave as we try to query on the Internet are extremely revealing.

The idea of this project is to analyze multiple topics and some places merge it with other external data. The analysis starts with a simple subject and gradually more complex subjects are covered. Actual development is published at
http://rpubs.com/chirag/GoogleTrendsDataAnalysis

Data gathering

Although data from Google Trends is nicely formatted and easily available. There is no official API available to query this data. An API to accompany the Google Trends service was announced by Marissa Mayer, then vice president of search-products and user experience at Google. This was announced in 2007, and so far has not been released. Philippe Massicotte has written unofficial gTrendsR API to get this data and package is maintained by him. Fortunately, a package is updated but the documentation is not, so I reached out to him and then I was able to make it work using trial and error approach.

Search quotas

Google has incorporated quota limits for Trends searches. This limits the number of search attempts available per user/IP/device. Details of quota limits have not yet been provided, but it may depend on geographical location or browser privacy settings. It has been reported in some cases that this quota is reached really quickly if you are not logged in into your Google account before trying to access the trends service.

Challenges

Sometimes while doing high volume query search with gTrendsR API stops responding and doesn’t give any result at all (even for smaller and simple queries). It returns something like

$status_code == 200 is not TRUE

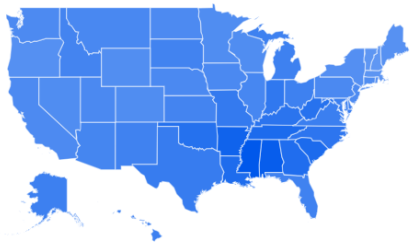
In that case, we need to wait for many hours and then it would start responding properly. Unfortunately, it is not documented to what is that magic - upper limit - a number which is allowed.

Visualization

Visualization is done with the mix of ggplot2, plotly and pure HTML in R Markdown. Plotly is leading open source tools for data visualization and it has good community support hence it is preferred for plotting. The map is created through mapchart.net

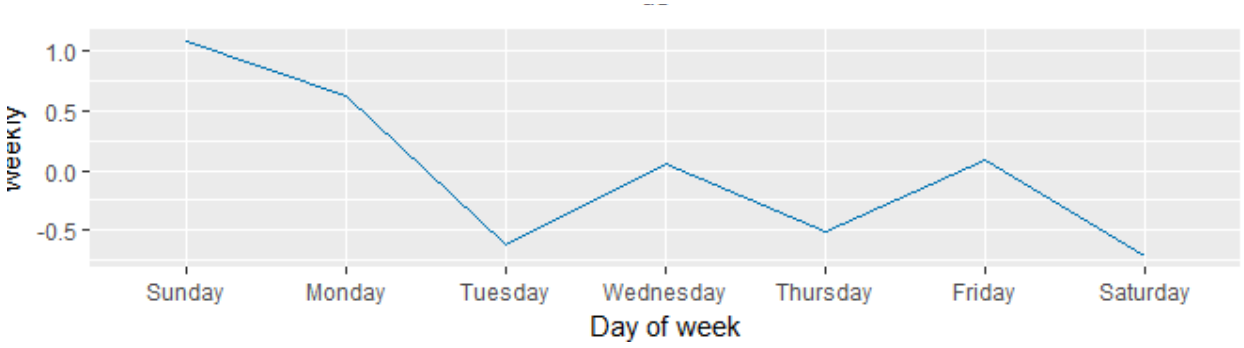
In search of God

Just like all programming books starts with “Hello World !” program. I start with a simple query, searching ‘God’ in google trends tool and we can see the states that make the most Google searches mentioning “God” are Mississippi, Alabama, and Arkansas which are part of the Bible Belt. The *Bible Belt* is an informal region in the southeastern and south-central United States in where Christian church attendance across the denominations is generally higher than the nation's average



1	Mississippi	100
2	Alabama	96
3	Arkansas	91
4	South Carolina	84
5	Tennessee	84

None of this is surprising, confirms or proves the obvious, but it is fascinating that search data could reveal such a clear pattern. The power of Google searches is not that they can tell us that *God* is popular in southern states, any survey would tell us that. As we will see, the power in Google trends is that users tell the Google search engine things they might not tell anyone else.



We can see in this data that search historically picks up on Sundays. For some reason, people look for God more on Sundays. There is a nice detailed analysis done on subject with the title “Philosophical and religious questions people ask Google”.

Movie search trends v/s Lifetime Gross

Next, I choose to analyze academy awards nominated movies. I wanted to compare if there is some pattern in how people search for movies and the business that movie does. So I choose below four movies to do the analysis. I had to keep "*Dunkirk"* out as Google search data wasn't clear about the search of "*Dunkirk"* movie, as there is place with same name.

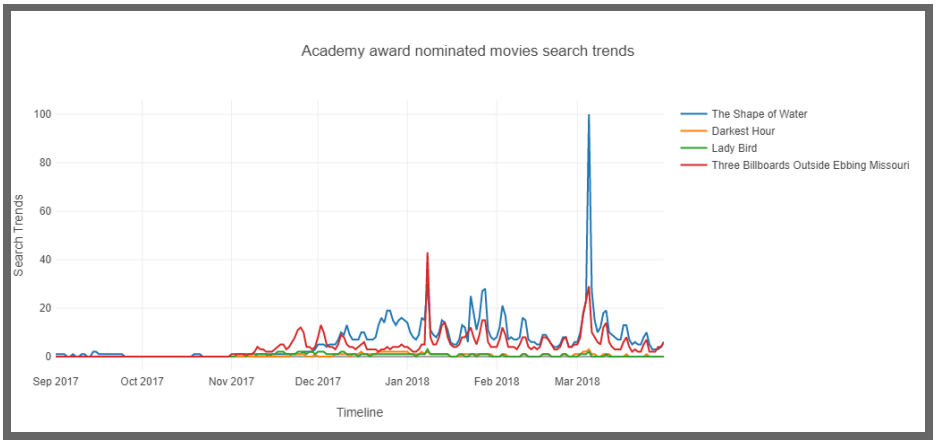

The Shape of Water


Three Billboards


Lady Bird


Darkest Hour

I got Google search data from trends tool and the data of business that movie did from BoxOfficeMojo.com



As we can see "*Shape of water*" is highest searched movie among all four and that is also movie which had highest gross business. Second highest searched movie is "*Three Billboards Outside Ebbing, Missouri*" whereas second most earner was movie "*Darkest Hour*".

Least searched movie among four is "*Lady Bird"* and that is the one which did least business. so looks like Google search does reasonably good job with choosing first and last winner. In addition of Google search of movie names, we would have checked how frequently YouTube search of movie (to see trailer) was done and how that comparison results.

Quit smoking search and predict with prophet package

Each new year we all make resolutions and we keep making them each year. We all search lose weight and quit smoking each year. These search queries peak around January each year.

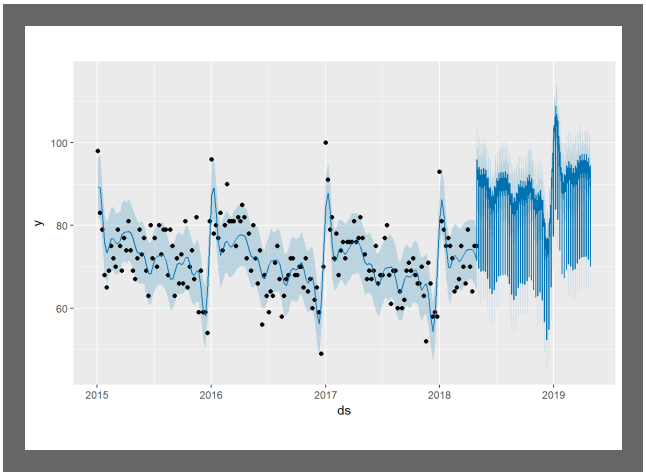


Image of left shows the forecasting (using Facebook's time series forecasting prophet package) of "*quit smoking*" query, which shows that this trend continuing in January 2019. This data consistent throughout the world. Using this graph we spot the yearly trend and seasonality much clearer and easy to analyze.

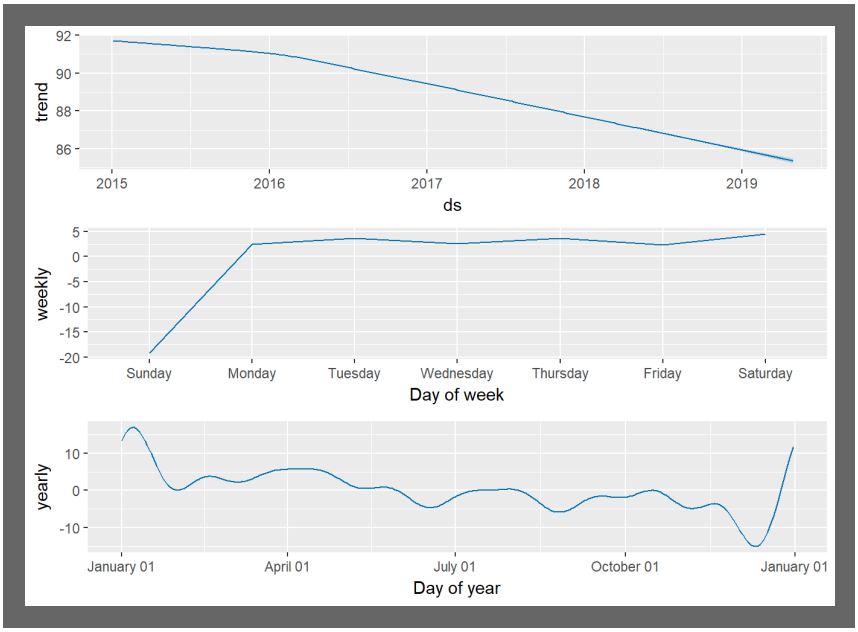
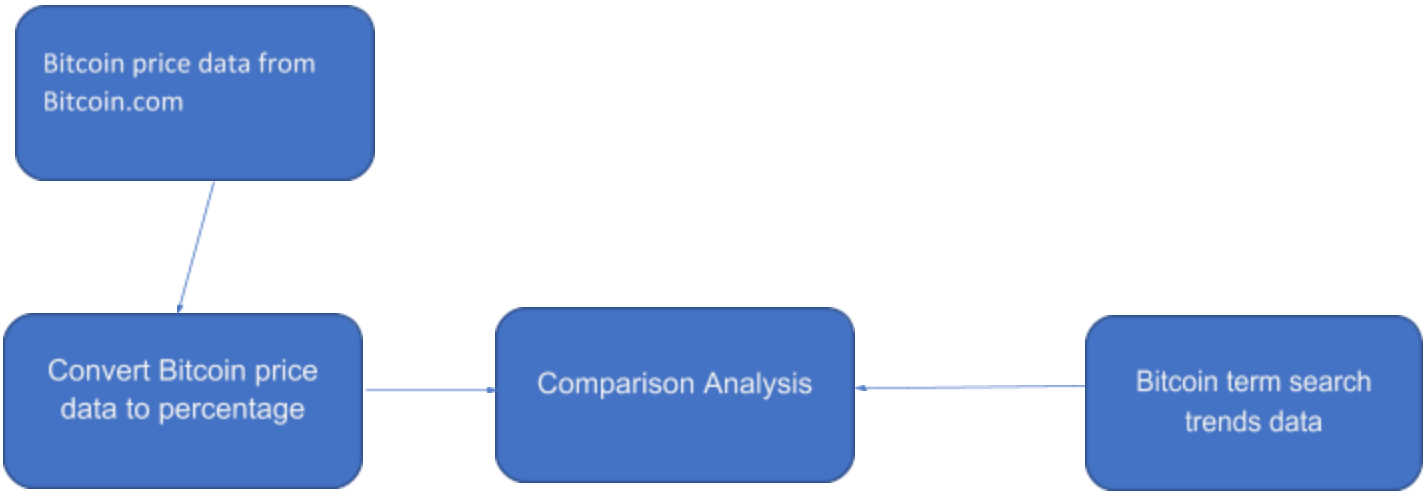


Image on left side shows same data in different time spans. As we can see overall trend is over the years is going down. Somehow we search about "quit smoking" more on except on Sunday.

Interesting trend query on same subject is "quit smoking magnet" which is on rise.

Bitcoin search trends v/s Bitcoin price

Bitcoin price hit it's all-time high in mid-December 2017, touching around $20,000 USD. Thus, it will come as no surprise that the number of people searching for 'Bitcoin' on Google also peaked at the exact same time. The term "Bitcoin" was the second most popular search for global news in 2017. It would be interesting to see how Google search Trends about term 'Bitcoin' data compares with actual Bitcoin price. Following are steps I followed.



Filtering

Here although Bitcoin price data is available from the year 2010, there was no price movement. Price started moving since last 2-3 years otherwise, it was almost flat. So, I decided to show price from 2016 onwards.

When Bitcoin price data was available daily, Google trends for Bitcoin term search data was available on weekly basis. So, using merge function I choose only weekly price for which trends data available.

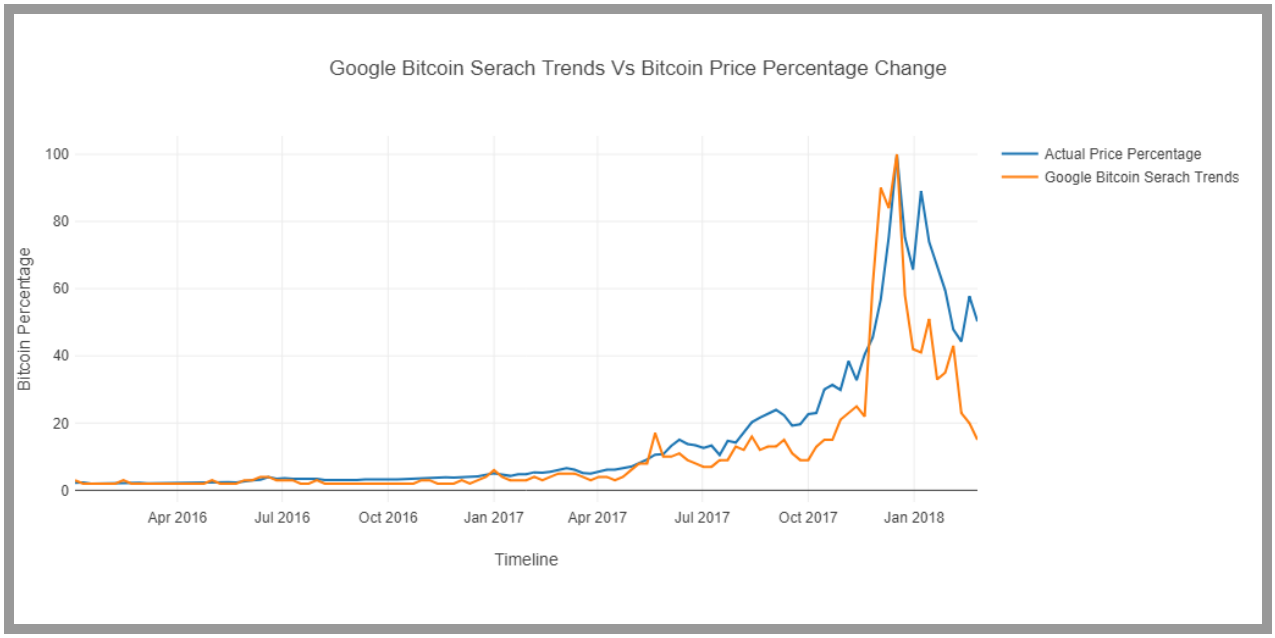
Conversion

Bitcoin price data was available as the absolute value; whereas Google trends data is available as a relative value from 1 to 100. i.e. 1 being least search, 100 being term searched most. So, I decided to convert Bitcoin price value to percentage value. So that way I have a dependent variable in the same format for comparison.

Also, date format was different so I converted to common date format so data can be joined with the date column.

Plotting for comparison

Below is Bitcoin price and Bitcoin search trends comparison using line chart. It seems there is a correlation between how many people search about the Bitcoin and the price of Bitcoin. Also, it shows around December-January the price and searches for Bitcoin was highest. It doesn't mean more people search for Bitcoin; more its value will be. But reverse might be true.

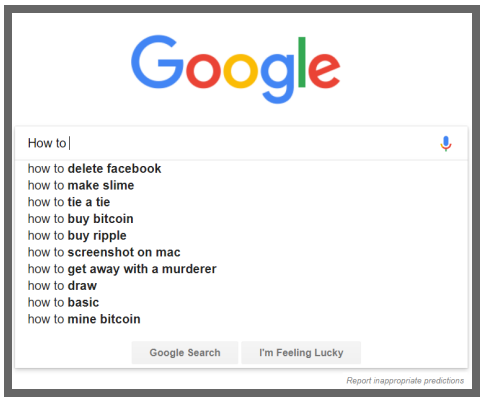


As we can see search for Bitcoin and price both picked around last week of December 2017 and starting of January 2018. Although It's not a perfect indicator Google Trends sometimes lags and sometimes leads bitcoin' price.

We can say there's a strong correlation between bitcoin's price and the performance of the search term "Bitcoin" on Google. Maybe price follows interest, and therefore, more buyers and greater search volume. If that is the case trends chart shows a reversal to the uptrend for interest in Bitcoin.

How to…search analysis

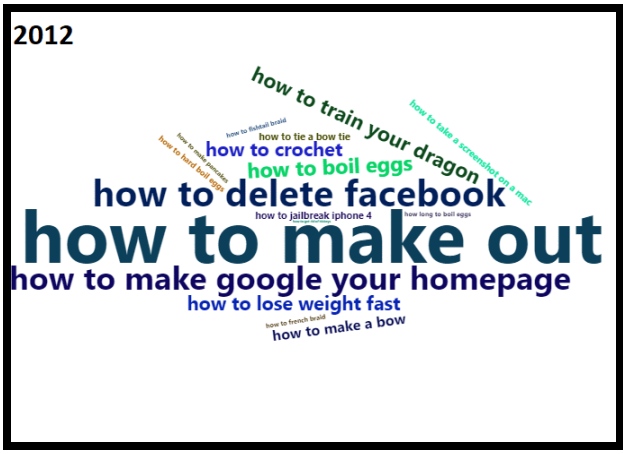
“How to...” is one the most important queries we do on Google. It can reveal what people are looking for on the Internet in different parts of the world.



This analysis is divided into two parts.

- First is to find in general what people are searching every year and to see how it changed over years. This is analysis is performed only for United States region.
- Second is find unique and unusual “How to” queries which people ask only in particular region. This analysis was done for the whole world which zeroed down some interesting facts.

Visualization



There was a question about how to present “how to..” search query analysis in a meaningful way. So, I decided to use word cloud as it gives greater prominence to words that appear more frequently. I wrote a reusable function which would get google trends data using gtrendsR
Even after generating word cloud for each year; it is overwhelming to see six (2012 to 2017) word cloud images. So, I decided to make GIF image animation, displaying each year at the top. I used ezgif.com for that. Additionally, this information displayed in tabular format. Unfortunately PDF format report will not show animation.

Queries like “how to boil eggs”,”take a screenshot”,”tie a tie”,”write a check” we have been asking since many years. In 2016, being election year, one of the most asked query was “how to vote” or “how to register to vote”.

Challenges

Some of the “How to” queries are the title of the movie or name of game. These are not actual queries and doesn’t give any meaningful information. For now, those queries remove manually.

Unique or unusual how to…queries

What is Unique and unusual is an abstract concept. So, I went through all “How to” queries data for all the countries of the world. I choose queries which are unique in the region to compare to rest of the world.

Topic	Country
how to make paper flowers	Bhutan
how to take pictures of northern lights	Iceland
how to become good teacher	India
how to get twins	Kenya
how to hack facebook	Myanmar
how to make carrot oil	Nigeria
how to handle wife	Pakistan
how to identify AIDS	Sri Lanka
how to delete telegram account	Uzbekistan
how to measure infiltration rate	Zimbabwe

- The most surprising part of this project was to see Sri Lanka being very a tiny country of the world searches “how to identify AIDS” more than any country in the world. I am really curious to know why this is happening. I reached out and tweeted about this to health minister of Sri Lanka.
- Before working on this project, I didn’t know what is aurora/Northern lights. I found out that it is a natural light display in the Earth’s sky, predominantly seen in the high-latitude regions like Iceland. That is the reason people from Iceland search “how to take pictures of northern lights”.

- We see “how to handle wife” query most searched in “Pakistan”. I feel this is funny and serious query at same time which points gender inequality.
- Right side shows other countries (which are mostly south asian countries) where people are searching for same.

location	Percentage of Hits
Pakistan	100
Sri Lanka	69
United Arab Emirates	54
India	42
Bangladesh	34

How to spell & pronounce

Google Trends is a unique and useful tool we can use to keep track of what people want to know about. When people do Google search; they misspell words and collectively Google can tell which words are misspelled more frequently. Similarly, people also use google when they want to know the spelling of some complex word with query "how to spell..".

Misspelled words

There is an article with the title "*Google reveals top "how to spell' searches by Canadian province*" which gave an interesting analysis of which words are misspelled by Canadians. Similar data analysis for the United States was given by Google trends twitter handle. Below shows how one small incident (in this case tweet) can make people search about partifulcar query.

Story behind "how to spell sound of sniff"

If we query "how to spell" right now then top Google trends result in the United States is "how to spell the sound of a sniff" which is a strange result but there is a story behind it. Macaroni Tony with Twitter handle @BigBeard_Ali, who is followed by 14K users tweeted below on 11 Feb 2018.

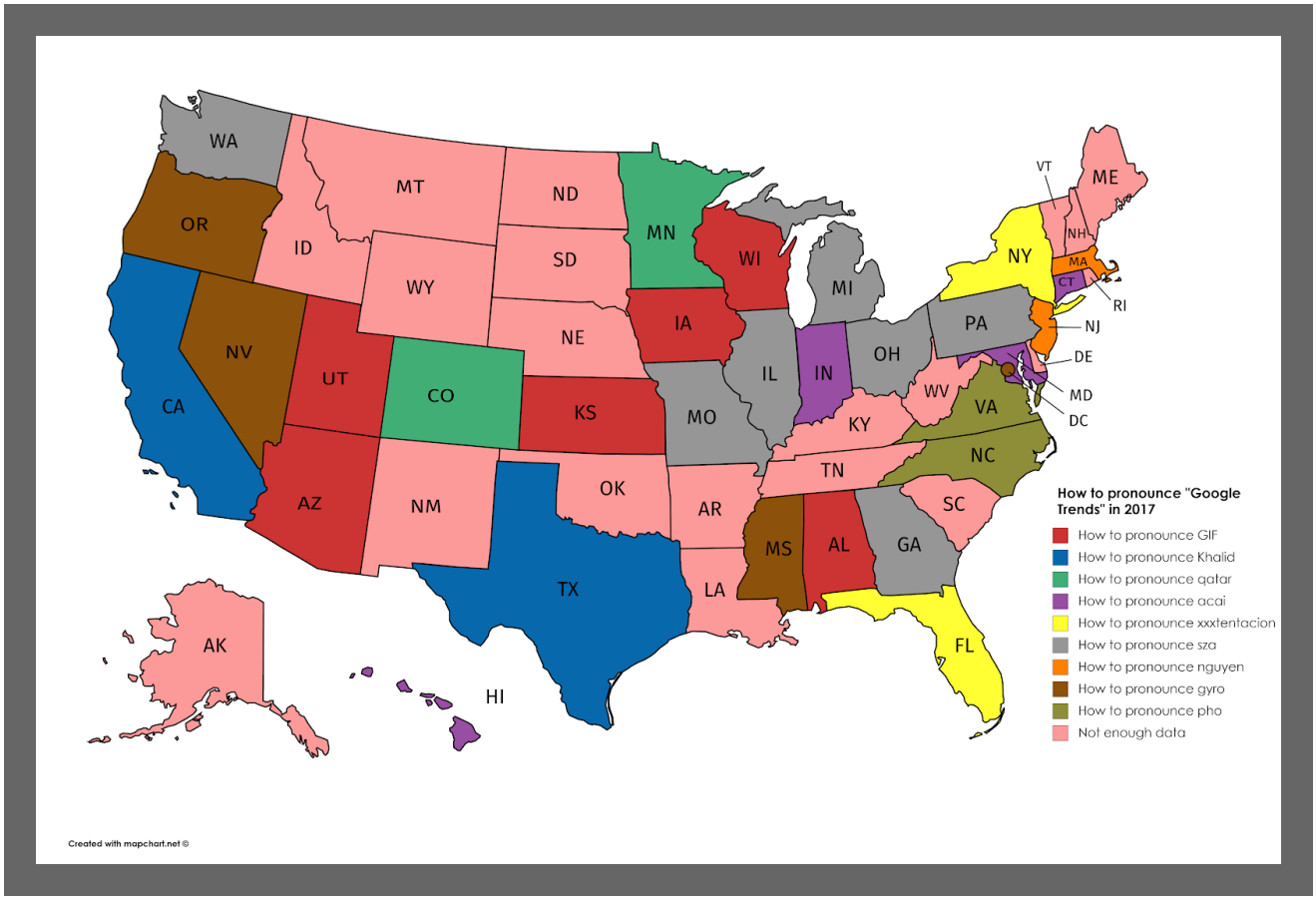
This tweet had around 6K re-tweets and 10K likes and this led people to search for "how to spell sound of sniff". I was able to see that this query picked only after this tweet. It is amazing that only one tweet with a small amount of prize money can have such ripple effect.

How to pronounce searches in the United States

Inspired by above analysis, I decided to do how to pronounce analysis. It would be fun to understand which are the words people find difficult to pronounce. In last twelve months, most difficult words people find difficult to pronounce are people's names.

- First, is singer-songwriter named "Sza" 
- Followed by Wonder Woman fame "Gal Gadot" 
- Followed by American rapper "XXXTentacion" 

Below is state wise breakdown showing each state find which word difficult to pronounce.



This also contains food-related items like 'pho', 'gyro' and acaí. "how to pronounce GIF" shown on the map is also all time high worldwide. The map is created through mapchart.net. At world stage top query is also "how to pronounce Pyeongchang" which hosted the 2018 Winter Olympics and the 2018 Winter Paralympics.

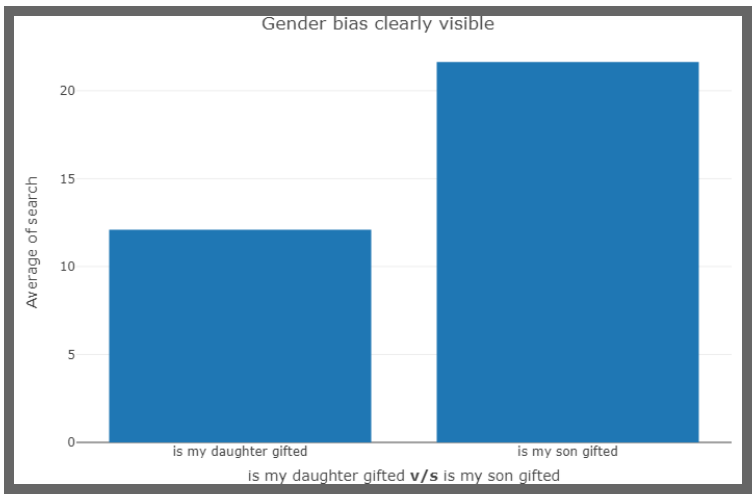
Sons vs. Daughters: Gender Bias

In the 21st century we would like to think that we treat boys and girls same. At least in United States - which is arguably one most the most liberal country in the world. We would like to think that American parents have similar standards and similar dreams for their sons and daughters and there will not be any gender bias. But a study from Seth Stephens-Davidowitz suggests that is not the case.

Google searches suggest parents have different concerns for male and female children. As he points out and the same is shown below with Google trends data that our parents expect (maybe unknowingly) boys to be smarter and girls to be thinner. They are more excited by the intellectual potential of their sons and they are more concerned about the weight and appearance of their daughters.

Giftedness

Mostly all parents like to believe that their kid has special talent. But as below graph shows people in US queried more “is my son gifted” than “is my daughter gifted”. This search difference is almost 50%.



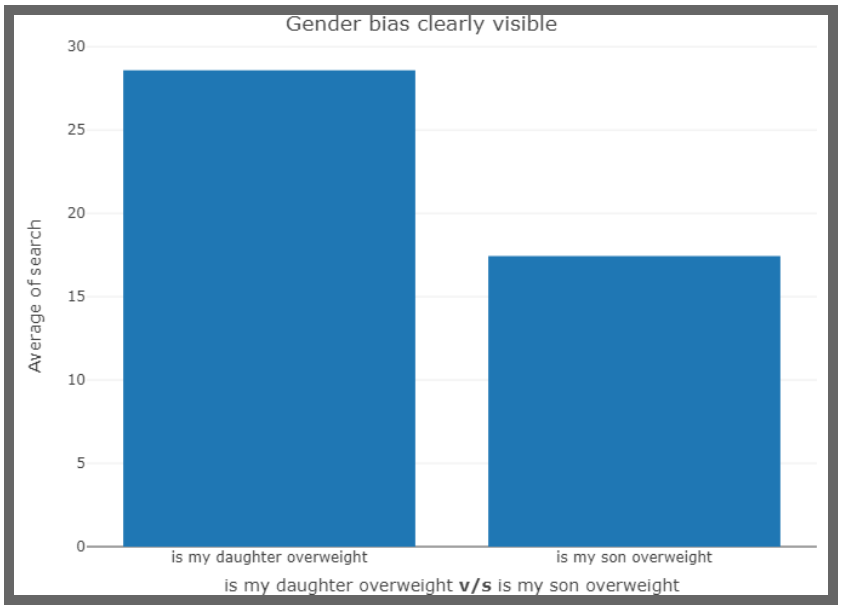
Unfortunately, this is despite the fact that that in real life this is exactly opposite. As David Walsh, an American psychologist, who specializes in parenting, points out as below

“*Girls talk earlier than boys, have larger pre-school vocabularies, and use more complex sentence structures. Once in school, girls are one to one-and-a-half years ahead of boys in reading and writing. Boys are twice as likely to have a language or reading problem and three to four times more likely to stutter. Girls do better on tests of verbal memory, spelling and verbal*

fluency. On average, girls utter two to three times more words per day than boys and even speak faster---twice as many words per minute."

Overweight

Childhood obesity has undoubtedly become one of the most complex public health problems facing future generations. In last 12 months parents search "Is my daughter overweight?" almost twice as frequently as they search "Is my son overweight?". This gap is narrowed if we examine data from last 5 years. Which means we are becoming more biased. At least one thing is sure that parents worry more about overweight girls than overweight boys. Just as with giftedness, this gender bias is not grounded in reality. That is despite the fact that trends show - in reality, boys are more overweight than girls.



The volume of these two queries are only in the United States; so it was not possible to do the relative comparison. Parents were more likely to ask about sons rather than daughters on every matter. There are more searches for "is my son behind" or "stupid" than comparable searches for daughters. Also, parents are more likely to ask whether a son is "happy" and slightly more likely to ask whether a daughter is "depressed".

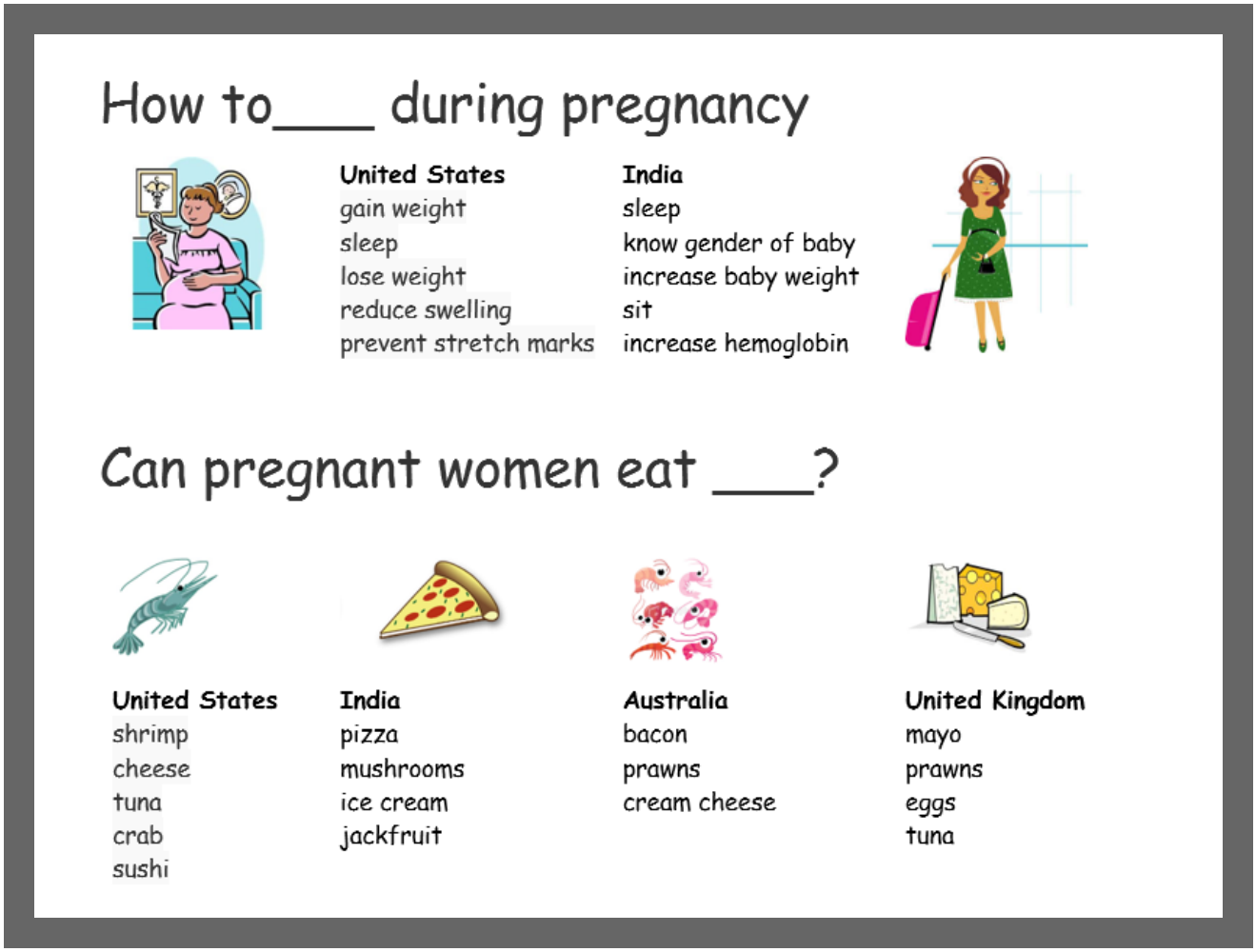
Google search data also tell us that mothers and fathers are more likely to wonder whether their daughter is "beautiful" or "ugly." How Google is expected to know whether a child is beautiful or ugly is hard to say :)

What pregnant Women Want

Unlike “what women want” which no one can claim to know about, we think we would have a good idea “What pregnant Women Want”. It would be the cravings for pickles and chocolate, the avoidance of wine, nausea and stretch marks, and good supporting husband.

It looks like Google trends can put more light on this question and surely we can expect different answers from different parts of the world.

We can start with questions about “*what pregnant women can do”* safely. The top questions women ask in the United States are: Can pregnant women “eat shrimp, tuna, fish or crab”, “drink wine, coffee or tea” or “take Tylenol”? There are (comparatively) less frequently asked questions too like “fly”, “take bath” or “paint”.



But if we inquire the same question in other countries, they don't look much like the United States or one another. Whether pregnant women can "drink wine" is not among the top ten questions in Canada, Australia or the United Kingdom. Australia's unique concern is mostly related to eating cream cheese.The differences in questions have less to do with what is safe to do and more to do with information coming from different sources in each country including old stories, local custom, and neighborhood trivial talk.

We can see another clear difference when we look at the top searches for "how to ___ during pregnancy?" In the United States, the top search is (gain or lose) weight-related queries whereas in India women are more concerned about how to sleep or weight of the baby. In India, it is illegal to find the gender of a baby before birth and that is why one of the top queries is to know any trick about knowing the gender. Also, a baby with more weight is considered as "healthy" that is the reason mother is more concerned about baby's weight.

While the cultural manifestations of pregnancy may be different, the physical experience tends to be similar everywhere. I tested how often various symptoms were searched in combination with the word "pregnant." For example, how often is "pregnant" searched in conjunction with "nausea," "back pain" or "constipation"?

Canada's symptoms were very close to those in the United States. Symptoms in countries like Britain, Australia and India were all roughly similar, too. Preliminary evidence suggests that no part of the world has stumbled upon a diet or environment that drastically reduces a pregnancy symptom.

We can extend this analysis and check what expectant *fathers are searching for*. In the United States, right now the top search on the subject is "be nice to me my wife is pregnant shirt".

Conclusion

Philosophers speculated about a tool called “cerebroscope,” a mythical device that would display a person’s thoughts on a screen, people have been looking for tools to expose the workings of human nature. Google Trends data is one of such tool which is an anonymous, categorized, and unbiased sample of Google search data. It tracks trillions of searches per year , making it one of the most useful, real-time data indicators of human interest by region and category. Google Trends is most often used to understand brand health and monitor changes in consumer interests along competitive metrics and factors such as seasonality.

Search engine query data offer insights into our life on the smallest possible scale of individual actions. In order to investigate whether Internet search volume is correlated with another aspect of our life, I used search volume data provided by the search engine Google. I tried to focus on the incredible amount of information about the localized behavior we can get from Google Trends. We got answers to questions we never asked from people we never considered. On top of that, we got information on historic behavior — we can’t ask panels how they felt many years ago!

One of the reasons why I preferred to use Google Trends as my source for information instead of the standard surveys or focus groups is the fact that we are leveraging the largest panel in the world (the internet). It’s honest, trusted and not influenced/skewed. Google Trends is arguably the best publicly available data we have. I say more trusted because it is default search engine we all use. If people were using the different search engine then this kind of analysis would have been difficult. I admire Google as a company for keeping this kind of data publicly available for free. Taking the meaning of open to next level. Due to this openness, anyone can do analysis with one’s own interest. Ideas can come from anyone. Data analysis is no longer restricted elite group of researchers and academias.

Opportunities are endless. For many people, Google is more than just a simple search engine — it’s one of their closest confidants. The evidence is provided by Jeremy Ginsberg that Google Trends data can be used to track influenza-like illness in a population. Because the relative frequency of certain queries is highly correlated with the percentage of physician visits in which a patient presents with influenza-like symptoms, an estimate of weekly influenza activity can be reported.

This kind of data analysis can answer taboo subjects like what percent of American men are gay? or issues related to child sexual exploitation or child abuse can be analyzed more reliably because the Internet is the first thing we reach out. The advantage of this data source, of course, is that most people are making these searches in private.

Google Trends offer an unprecedented peek into people’s psyches. People can unburden themselves of some wish or fear without a real person reacting in dismay. Most important thing is, we should ask the right questions !

Appendix

Code https://github.com/chirag-vithlani/Capstone

References

1. Google Trends Help Center https://support.google.com/trends/?hl=en#topic=6248052
2. What is Google Trends data — and what does it mean? https://medium.com/google-news-lab/what-is-google-trends-data-and-what-does-it-mean-b48f07342ee8
3. Everybody Lies, book by Seth Stephens-Davidowitz (ISBN-10: 0062390856)
4. Google reveals top 'how to spell' searches by Canadian province https://www.ctvnews.ca/canada/google-reveals-top-how-to-spell-searches-by-canadian-province-1.3447494
5. Timing the “Bitcoin Bubble”—How Google Trends offers insight into the $100 billion question https://techburst.io/timing-the-bitcoin-bubble-how-google-trends-offers-insight-into-the-100-billion-question-4e50f4e68308
6. Boys' and Girls' Brains: What's the Difference? by David Walsh, Ph.D. http://www.pbs.org/parents/experts/archive/2012/09/boy-and-girl-brains-whats-the.html
7. Philosophical and religious questions people ask Google https://my.visme.co/projects/6xor0j07-philosophical-and-religious-questions-people-ask-google#s1
8. The term “Bitcoin” was the second most popular search for global news in 2017 https://trends.google.com/trends/topcharts#vm=trendingchart&cid=045735de-6767-48be-8780-c2c50ffee8ae&geo=&date=2017&cat=
9. most misspelled words in America https://twitter.com/GoogleTrends/status/869618902942101504
10. Childhood Obesity (PDF) https://neohospitals.org/TheCenterForHealthAffairs/MediaCenter/NewsReleases/2013/June/~/media/75B1305DF6964F2CB6F1AFEA4735C42C.ashx
11. Meaning of open https://googleblog.blogspot.com/2009/12/meaning-of-open.html

More reading

My Attempt at Predicting the Opening Gross for the Avengers: Infinity War Using Linear Regression
https://medium.com/@pmin91/how-i-tried-to-predicted-the-opening-gross-for-the-avengers-infinity-war-using-data-science-3fd2beb9512d