



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

Analyzing Google Trends Data

Mid-term report

Chirag Vithalani

CUNY (City University of New York)

New York, NY 10017

Advisor / Guide: Honey Berk

Contents

Introduction	3
Data gathering	3
Visualization	3
Bitcoin price versus Bitcoin term search	4
Filtering	4
Conversion	4
Plotting for comparison	4
How to...search analysis	5
Visualization	5
Challenges	5
Unique or unusual how to...queries	6
Summary	6

Introduction

Google search data don't seem to be a proper source of information for "serious" research. But we can see trails we leave, as we pursue knowledge or query on the Internet are extremely revealing. It can show many things about collectively us as a society.

Below is a mid-term report on "Analysis Google Trends" project. It discusses the milestone achieved, challenges faced, working code examples and road ahead.

The idea of this project is to analyze multiple topics and merge it with other external data. Actual development is published at <http://www.rpubs.com/chirag/GoogleTrendsAnalysis>

Data gathering

Although data from Google Trends is nicely formatted and easily available. There is no official API available to query this data. [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 was able to make it work using trial and error approach.

Visualization

Visualization is done with the mix of 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.

Bitcoin price versus Bitcoin term search



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.

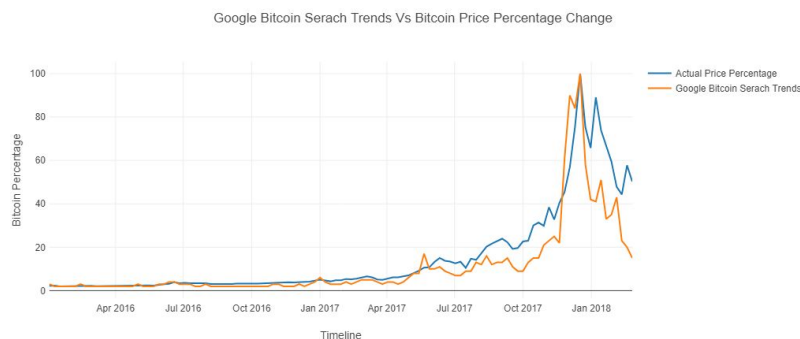
When Bitcoin price data was available daily, Google trends for Bitcoin term search data was available on weekly basis. So, using merge function I choose only weekly price for which trends data available.

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.

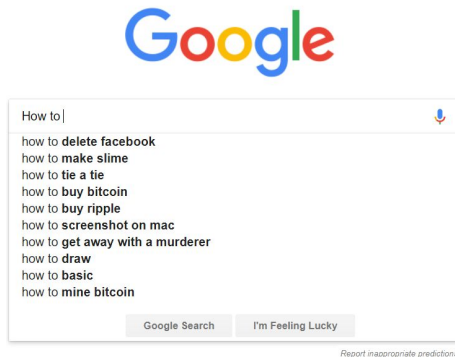
Plotting for comparison

Below is Bitcoin price and Bitcoin search trends comparison using line chart. It seems there is a correlation between how much people search about 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.



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.



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 done only for US region.
- Second is find unique and unusual “How to” queries which people ask only in particular region. This analysis done for the whole world 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.

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. Challenge is to remove them without doing any manual work.

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 chose queries which are unique in region to compare to rest of the world. I am displaying this data on a world map and when we mouse hover it will show actual query.

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

- We can see Sri Lanka being very tiny country of the world searches “how to identify AIDS” more than any country in the world.
- We see “how to handle wife” query being searched more in south Asian countries which points gender inequality.

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

- Northern light 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”.

Summary

We can say original intent of project – to analyze google trends data – is heading in right direction. After initial hiccups, now we know to work with gTrendsR package in ‘R’ and R Markdown. This sets the stage for more detailed and in-depth analysis of other topics.