Google Trends Data Extraction Tool

Last revision: 07 March 2023

Current version: 2.2.2

Table of Contents

TABLE OF CONTENTS ..... **II**

LIST OF FIGURES ..... **IV**

LIST OF TABLES ..... **V**

CHAPTER 1 OVERVIEW **1**

1.1 DEVELOPER 1
1.2 CITATION 1
1.3 LICENSE 1
1.4 PASSWORD 1
1.5 CODE 1
1.6 LINKS 1
1.7 DEVELOPMENT 1
1.8 CODE ATTRIBUTIONS 2
1.9 DEVELOPER ACCOUNT 2
1.10 ACCOUNT QUOTAS 2
1.11 DATA LIMITATIONS 4
1.12 DATA SCALING AND ZERO VALUES 4

CHAPTER 2 COMMON SETTINGS AND BACKGROUND INFORMATION **7**

2.1 INPUT AREAS 7
2.2 QUERY TERM SPECIFICATIONS 7
2.2.1 What is a query term? *7*
2.2.2 Query term limitations *8*
2.2.3 Query term specification modifiers *8*
2.2.4 Maximum number of query terms *9*
2.2.5 Freebase IDs *10*
2.3 COUNTRIES AND REGIONS 11
2.3.1 Background information on geographic specification *12*
2.3.2 Geographic specification *12*
2.4 DATE SELECTION 13
2.5 API KEY AND ACCOUNT NAME 14
2.6 TARGET FILE 15
2.7 USER INTERFACE ERROR MESSAGES 16
2.8 USER INTERFACE HELP MESSAGES 16
2.9 CLEAR SPECIFICATIONS 17
2.10 LOADING PREVIOUS SPECIFICATIONS 18
2.11 ABOUT 18
2.12 PROGRAM OUTPUT 19
2.12.1 Worksheets *19*
2.12.2 Charts *21*
2.12.3 Log *21*

CHAPTER 3 GOOGLE TRENDS EXTENDED HEALTH **22**

3.1 DATE SPECIFICATION 22
3.1.1 Date specification via the calendar form *23*
3.1.2 Date specification via the year/month selector *23*
3.2 SAMPLING 24
3.3 MULTI-TERM REQUESTS 25
3.4 MULTI-REGION REQUESTS 25
3.5 EXTRACTION PROCESS 26
3.6 OUTPUT 27
3.6.1 Samplings (Data) sheets *27*
3.6.2 Summary sheets *28*
3.6.3 Sampling Summary *30*
3.6.4 Master Summary sheet (Means of all … series) *32*
3.6.5 Chart *33*

CHAPTER 4 GOOGLE TRENDS WEB **35**

4.1 DATE SPECIFICATION 35

4.2 FUNCTION 35
4.3 SEARCH DOMAIN 38
4.4 CATEGORY 38
4.4.1 Category selector 38
4.4.2 Browsing categories 40
4.5 OUTPUT 40
4.5.1 Graph 40
4.5.2 Graph Averages 41
4.5.3 Rising Queries and Top Queries 42
4.5.4 Rising Topics and Top Topics 43
4.5.5 List of Region Values 44

CHAPTER 5 PUBLICATIONS AND CITATION 48

CHAPTER 6 VERSION HISTORY 49

REFERENCES 54

List of Figures

FIGURE 1 API ACCOUNT PAGE 3
FIGURE 2 ACCOUNT QUOTAS (GOOGLE TRENDS EXTENDED HEALTH INTERFACE) 4
FIGURE 3 INPUT AREAS (GOOGLE TRENDS EXTENDED HEALTH INTERFACE) 7
FIGURE 4 GOOGLE KNOWLEDGE GRAPH ENTITIES SHOWN AS TOPICS ON GOOGLE TRENDS WEBSITE 10
FIGURE 5 GOOGLE SEARCH RETURNING WHAT IT BELIEVES TO BE THE CORRECTLY SPELLED ENTITY 11
FIGURE 6 GEOGRAPHIC SPECIFICATION 12
FIGURE 7 GEOGRAPHIC LEVEL 12
FIGURE 8 COUNTRY 13
FIGURE 9 REGION 13
FIGURE 10 CALENDAR FORMS 14
FIGURE 11 API KEY FILE AUTO LOADING 15
FIGURE 12 MANUALLY SELECTING AN API KEY FILE 15
FIGURE 13 TARGET FILE SPECIFICATION 16
FIGURE 14 MESSAGES INDICATING EXTRACTION CAN BE COMMENCED 16
FIGURE 15 ERROR MESSAGES THAT STILL NEED RESOLVING 16
FIGURE 16 USER INTERFACE HELP MESSAGE 17
FIGURE 17 CLEAR SPECIFICATIONS BUTTON 17
FIGURE 18 LOAD SPECIFICATION BUTTON 18
FIGURE 19 ABOUT PAGE 19
FIGURE 20 QUERY SPECIFICATION SHEET (GOOGLE TRENDS EXTENDED HEALTH) 20
FIGURE 21 API REQUEST URL ENTERED INTO A BROWSER 20
FIGURE 22 GOOGLE TRENDS EXTENDED FOR HEALTH DATE RANGE SPECIFICATION 22
FIGURE 23 SETTING THE DATE RESOLUTION 23
FIGURE 24 USING THE CALENDAR FORM 23
FIGURE 25 GOOGLE TRENDS EXTENDED HEALTH DATE SPECIFICATION: YEAR/MONTH SELECTOR 24
FIGURE 26 SETTING THE NUMBER OF SAMPLES TO BE EXTRACTED FOR GOOGLE EXTENDED TRENDS FOR HEALTH DATA 25
FIGURE 27 SPECIFYING A MULTI-TERM REQUEST 25
FIGURE 28 SPECIFYING A MULTI-REGION REQUEST 26
FIGURE 29 EXAMPLE OF A PLOT FROM A MULTI-REGION REQUEST 26
FIGURE 30 SAMPLINGS SHEETS FROM A MULTI-REGION REQUEST 28
FIGURE 31 SUMMARY SHEET FOR A SINGLE-TERM, SINGLE REGION REQUEST 29
FIGURE 32 SUMMARY SHEET FOR ONE REGION OF A MULTI-REGION REQUEST 29
FIGURE 33 SAMPLING SUMMARY SHEET 31
FIGURE 34 MASTER SUMMARY SHEET FOR A MULTI-REGION REQUEST 33
FIGURE 35 LINE CHART FOR GOOGLE TRENDS EXTENDED HEALTH DATA 34
FIGURE 36 GOOGLE TRENDS WEB DATE RANGE SPECIFICATION 35
FIGURE 37 TRENDS API CALLS AVAILABLE VIA THE FUNCTION SELECTION 36
FIGURE 38 FUNCTIONS/API CALLS RELATED TO GOOGLE TRENDS WEBSITE CONTENT 37
FIGURE 39 SEARCH DOMAIN SETTINGS 38
FIGURE 40 CATEGORY FIELD (GOOGLE TRENDS WEB) 39
FIGURE 41 CATEGORY SELECTOR 39
FIGURE 42 BROWSING CATEGORIES 40
FIGURE 43 SAMPLE OUTPUT: GRAPH (FOUR QUERY TERMS—GOOGLE TRENDS, GOOGLE CORRELATE, GOOGLE SCHOLAR, GOOGLE ADWORDS) 41
FIGURE 44 LINE CHART FOR GOOGLE TRENDS WEB DATA 41
FIGURE 45 SAMPLE OUTPUT: GRAPH AVERAGES 42
FIGURE 46 SAMPLE OUTPUT: TOP QUERIES AND RISING QUERIES 43
FIGURE 47 SAMPLE OUTPUT: TOP TOPICS AND RISING TOPICS 44
FIGURE 48 SAMPLE OUTPUT: LIST OF REGION VALUES 45
FIGURE 49 BAR CHART FOR LIST OF REGION VALUES DATA FROM GOOGLE TRENDS WEB INTERFACE 46
FIGURE 50 MAP CHART FOR LIST OF REGION VALUES DATA FROM GOOGLE TRENDS WEB INTERFACE 47

List of Tables

TABLE 1 IMPORTANT LINKS 1
TABLE 2 CODE ATTRIBUTIONS 2
TABLE 3 QUERY TERM MODIFIERS 8
TABLE 4 CLEAR SPECIFICATIONS 18
TABLE 5 SUMMARY SHEET COLUMNS 28
TABLE 6 GOOGLE TRENDS WEB TIME RESOLUTIONS (USING DEFAULT TIME SETTINGS AVAILABLE ON GOOGLE TRENDS WEBSITE) 35
TABLE 7 GOOGLE TRENDS WEB INTERFACE FUNCTIONS PLOTTED AGAINST LIST OF API CALLS FROM HTTPS://DEVELOPERS.GOOGLE.COM/APIS-EXPLORER/#P/TRENDS/V1BETA/ 36
TABLE 8 CHANGE HISTORY FOR GOOGLE TRENDS DATA EXTRACTION TOOL 49

Chapter 1 Overview

1.1 Developer

This extraction tool for the Google Extended Trends API for Health was developed by Dr Jacques Raubenheimer (referred to below as "this developer") of the Australian Government's National Health and Medical Research Council-funded Translational Australian Clinical Toxicology programme at the University of Sydney.

The tool is an interface for connecting with, and extracting multiple samples from, the Google Extended Trends API for Health.

Current Version: 2.2.2 ©2023

1.2 Citation

Please see Chapter 5, p. 48, for details on how to cite the program.

1.3 License

The tool is released under the GPLv3 license (https://opensource.org/licenses/GPL-3.0).

1.4 Password

Some parts of the workbook (workbook structure, various worksheets worksheets) are protected with the password "GoogleTrends".

1.5 Code

The tool uses various features of Microsoft Excel. Some worksheet formulas, some data validation, a fair amount of named ranges (truth be told, almost 180 names), and a large amount of VBA (Visual Basic for Applications) code. Some code has been modified from other sources (cf. section 1.8 below), and the code is published open source, as indicated by the license above. The code is published with comments, and users are free to examine the code for their own purposes.

1.6 Links

Table 1 Important links

Site	Link
Google Trends website	https://trends.google.com
Google Trends API Explorer	https://developers.google.com/apis-explorer/#p/trends/v1beta/
Google API Terms of Service	https://developers.google.com/terms/
Google Trends API permission request	http://bit.ly/2KyqDYW

1.7 Development

The tool is an Excel workbook. Development was done on Excel 2016, 64-bit, for Windows, but it should work on any version of Excel for Windows from Excel 2010 through Excel 365. I doubt that it will work on Excel for Mac.

The tool makes extensive use of Excel Names and Named Ranges—all names and the ranges or formulas they refer to are listed in a hidden worksheet. The tool will not allow users to copy and paste onto its worksheets, as these may disrupt the Named Ranges. If you wish to copy and paste information, you will have to do it cell-by-cell, by editing the cell (F2) and then pasting your text into the content of the cell. The tool also makes use of a fair amount of Visual Basic for Applications code. Apart from the attributions below, all code was written by this developer.

1.8 Code attributions

This tool uses some code posted online for use by third parties. An indication is given Table 2 of which code snippets were modified to make them fit-for purpose in this application.

Table 2 Code attributions

Code	Modified	Author	Link
Code to export modules	Yes	Chip Pearson	http://www.cpearson.com/excel/vbe.aspx
modArraySupport	No	Chip Pearson	http://www.cpearson.com/excel/vbaarrays.htm
JSON parsing code in mJSON	No	Daniel Ferry	https://medium.com/swlh/excel-vba-parse-json-easily-c2213f4d8e7a
Code to block Cut/Copy/Paste functionality	Yes	Ken Puls	http://www.vbaexpress.com/kb/getarticle.php?kb_id=373
HTTP request code	Yes	Taeho Kang	http://tkang.blogspot.com.au/2010/09/sending-http-post-request-with-vba.html
Calendar control (CalendarForm v1.5.2)	No	Trevor Eyre	https://trevoreyre.com/portfolio/excel-datepicker/

1.9 Developer account

To use the tool, you need a valid API key associated with a developer account, as assigned by Google. To obtain an API key, complete and submit the request at http://bit.ly/2KyqDYW. Without an API key, the tool is useless.

Google will supply registered API users with an account page from where the quotas and daily usage can be monitored (https://console.developers.google.com/apis/api/trends.googleapis.com/quotas?project=[AccountName], Figure 1). This account name should be stored in the API Key file (see section 2.5, p. 14). A button on the interface will read the account name from the key file and open the account page in your browser.

1.10 Account quotas

Over and above the data limitations mentioned in section 1.11 below, Google caches data requested through the API for 24 hours, and a repeat of the request will not produce a new sampling, but will merely result in the cached data being returned again. This is a method of reducing the load on their servers, and since the data are provided by their courtesy, must be respected. This does, however, create a problem for the researcher hoping to obtain more accurate estimates from the data, and when referred to below, this will be termed the "caching limit."

In further steps to reduce the load on their servers, Google restricts users to a quota of API calls per day, referred to below as the "quota limit." This developer was restricted to 5000 calls per day. When the quota is exceeded, the API reports that the quota will be reset at midnight Pacific Time. Users can thus estimate when they can resume their queries after the quota has been exceeded.

Google[1] also restricts users to a certain number of API calls per second, below referred to as the "rate limit." The limit assigned to this developer's account was 2QPS (queries per second).

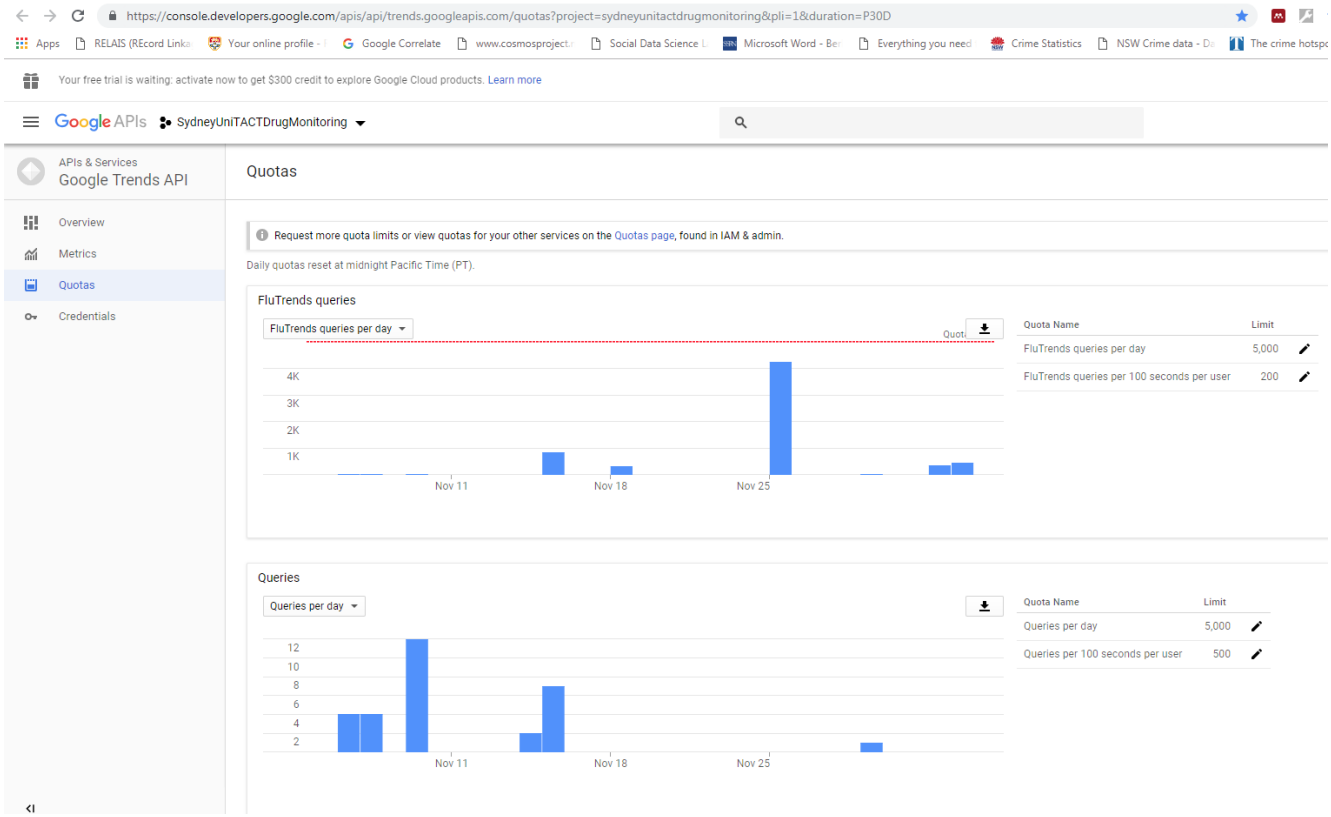


Figure 1 API Account page

Both of the account limits should be enter into the Account quotas section of the Google Trends Extended Health interface (Figure 2). The 'Reset Queries used' button allows users to manually enter a new value for the 'Queries used this session' box, e.g. when the program was left open and extractions are resumed on the next day.

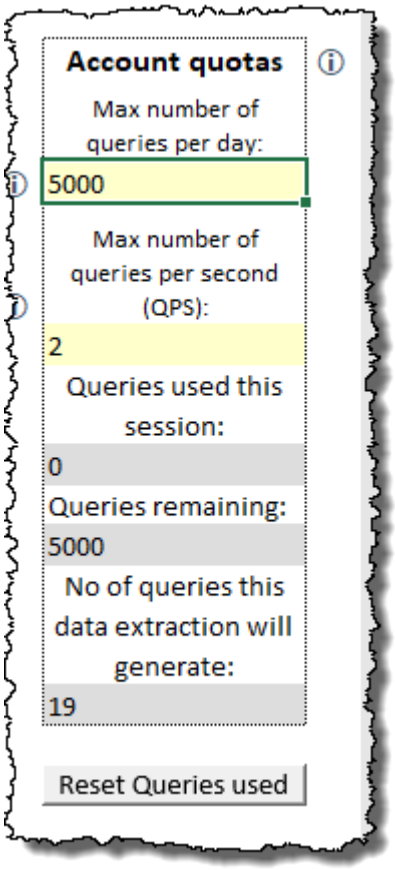


Figure 2 Account quotas (Google Trends Extended Health interface)

1.11 Data Limitations

Some important issues concerning the data should be noted. Google[1] offers the disclaimer that "Google can't guarantee the accuracy of the numbers." Caveat Emptor!

Each request to the Google Extended Trends API for Health returns a JSON string of no more than 2000 data points. This limits what can be requested (as will be seen below) and when this limit is referred to, the term "2000-point limit" will be used.

In mentioning the disclaimer above, the Google documentation[1] explains that "numbers are calculated on a uniformly distributed random sample of 10%–15% of Google web searches done since 2004 , updated once a day."

1.12 Data scaling and zero values

This section is essential reading for using Google Trends data for any analysis.

Google[1] reports that "The numbers returned are the probability of a short search-session (few consecutive searches), to satisfy the corresponding term restriction, given it was done in the restricted geography (if such exist) and during the time represented by that data point." It is important to note that the value is the probability of a short search session (Google does not define what they consider to be "short"), and not the absolute volume of searches. This is important, because probabilities can be compared, while absolute search volumes cannot. For example, if absolute volumes were compared, countries with large populations would always overshadow countries with smaller populations, but not necessarily so with probabilities. This does not mean that there are no regional differences in what is search, but rather, that the comparison of the probability of something being searched can be compared between regions using Google Trends data. Also, search volumes have changed over the years (more

people use the internet now than in 2004 when Google Trends data were first collected), but Google Trends data allow researchers to compare the probability of a term being searched now, to the probability of a term being searched in the past.

At this point, there is a separation in how the data are treated by the Google Trends API (accessed via the Google Trends Web interface) and the Google Health Trends API (accessed via the Google Trends Extended Health interface). For the Google Health Trends API, Google[1] explains that "this probability is multiplied by 10 Million in order to be more human readable. [Thus] Value[time-point, term-restriction] = P(term-restriction | time-point AND geo-restriction) * 10M." This tool reports the values accordingly.

However, for the Google Trends API (as can be seen on the https://trends.google.com website), the data are scaled so that he maximum value in the series is set to 100, and all other values as expressed relative to that (e.g., if the maximum probability was 0.00000006, it would be set to 100, and if another probability in the time series was 0.00000003, it would be set to 50, but in addition to this, these scaled values are also rounded to the nearest integer (e.g., 0. 0000000299 would also be set to 50). This has several implications, and explain why the Google Health Trends API data are to be preferred for research work above the Google Trends API data. First, the rounding means that Google Trends API data are coarser than Google Health Trends API data. Second, the additional scaling makes it difficult to compare or even aggregate data returned from the Google Trends API, while data from the Google Health Trends API can be aggregated or compared (subject to some limitations). To explain why Google Trends API data cannot be aggregated, consider two samples of monthly data, one in which the value for January 2016 is the highest, and the other in which June 2017 is the highest. For the former, the January 2016 value will be set to 100, and all other values in the time series, including June 2017, will be scaled accordingly. For the latter, the June 2017 value will be set to 100, and all other values in the time series, including January 2016, will be scaled accordingly. Presumably, for this to occur, the two original probabilities of January 2016 and June 2017 would not be very different, but the two resulting series of data, while roughly similar, would contained discernible differences. Using Google Trends data thus inherently incorporates a huge amount of chance. This will not happen with Google Health Trends data, as they are not scaled in this way. The comparison of data is also prevented by this scaling. Note that Google Health Trends data are still restricted in that they are defined by a time resolution and a geographic specification. However, as an example, if a researcher were to download, from the Google Health Trends API, daily data for 2004–2008, and then daily data for 2009–2013, these two time series can be successfully appended[*]. However, with Google Trends API data (and apart from the fact that this API will not return daily data for so long a period (weekly data can be returned for a five-year period, though—cf. Table 6, p. 35), the maximum value in the first series will be set to 100, and all other values in that series scaled relative to that maximum, and then the maximum value in the second series will be set to 100, and all other values in *that* series scaled relative to *that* maximum. What this means is that, even if the maximum probability in the first series was only half the maximum probability in the second series, they will both be expressed as 100, and the two series cannot be appended. There have been contrived processes developed to aggregate Google Trends API data[2], but it is far better to use the Google Health Trends API data, for the simple reason that multiple samples can be drawn from those data, which give users a very good estimate of the true probabilities, while the methods to aggregate Google Trends API data rely on combining individual samplings of different data, each with their own chance errors.

[*] See the discussion in section 2.2.4, p. 9) for why this might be necessary.

Similarly, if data for Country A and data for Country B are downloaded separately, these data would be comparable when downloaded from the Google Health Trends API, but not when downloaded from the Google Trends API, as the latter will have each country's maximum probability set to 100, and all other values for each country scaled accordingly, while this will not be done with the Google Health Trends data.

Furthermore, Google[1] notes that due to possible privacy concerns, "zeros may indicate no volume, or that the query didn't get enough distinct searches and is considered private" (i.e., it was very low, and was set to zero). Because of this, **the Google Trends Extended Health tool deletes all zero values from the data**. However, because of the additional scaling of the Google Trends Web data, zero values are not removed by that tool.

The default view of the workbook will show two worksheets: Google Trends Extended Health and the Google Trends Web. The Google Trends Extended Health is the main interface, and what the tool was created for. The Google Trends Web interface is an add-on to give access to some additional API calls allowed by the API, and produces some "nice to have" information.

Having said that, some settings have commonalities between the Google Trends Extended Health and the Google Trends Web interfaces. Other settings are unique, but require some explanation beyond the scope of the tools themselves. Both of these groups of settings are discussed here.

2.1 Input areas

All input areas in the user interface are specified in yellow. Some conditional formatting is used to change the appearances of inputs as they are entered. With some exceptions, which will be indicated below, all inputs must be completed before a data extraction can be run.

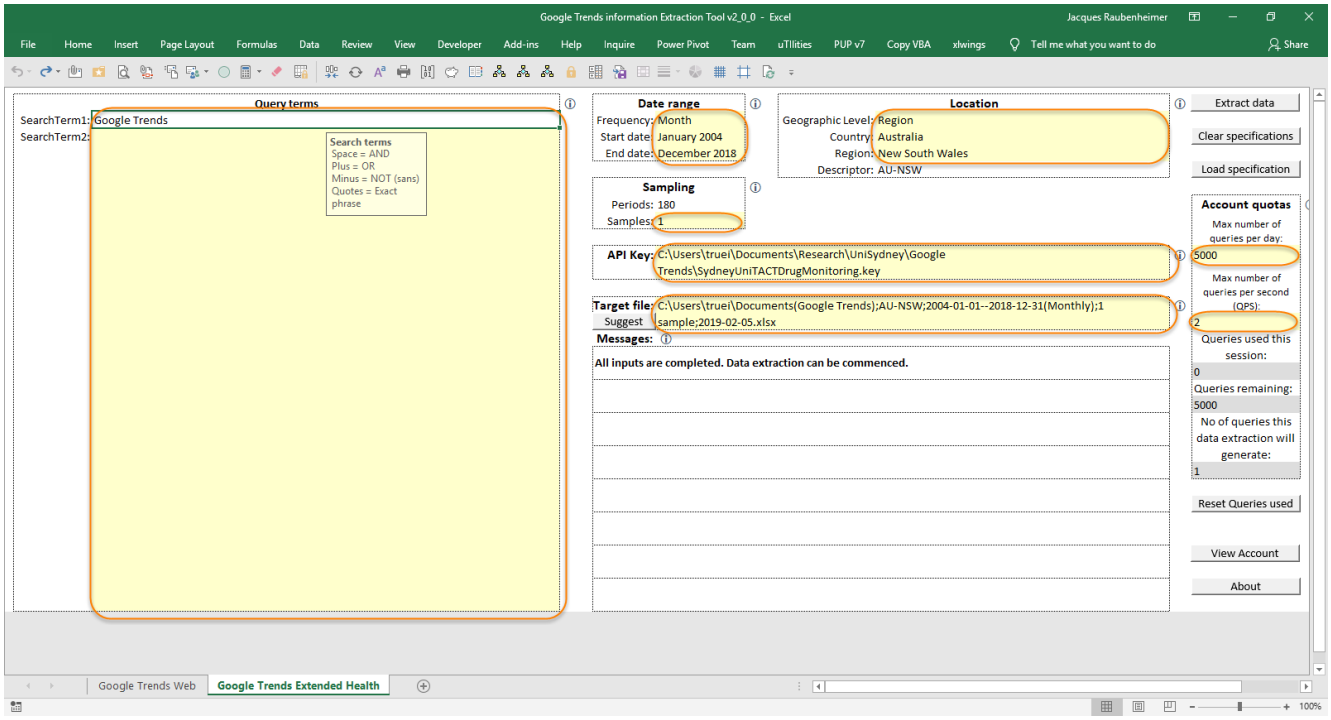


Figure 3 Input areas (Google Trends Extended Health Interface)

2.2 Query term specifications

The top left input area on both worksheets is designated for the query terms.

2.2.1 What is a query term?

Query terms are the words or phrases for which you want to extract Google search data. This is thus the most essential part of the process, and you should go to great lengths to ensure that your query term specifications are correct. This will rely on a combination of subject-matter expertise, as well as some trial and error to see what terms people are actually using on Google. For example, from the field of pharmacology where I developed the tool, people do not search for the drug MDMA by its street name "ice," because that understandably will not get them what they want (in fact, it seems as if they simply google "MDMA"). Conversely, we have discovered that with pharmaceutical agents, people sometimes

google drugs by their substance names (e.g., fentanyl), and other times by their brand names (e.g., Prozac). Note also that regional differences have to be taken into account.

It must be understood that a query term is not a single word, but a complete phrase, the structure of which will be explained below. Thus, in a previous study[3] using the tool, this developer used this as a single query term: *"cream charger" +"cream chargers" +nangs +whippits +whippit +/m/02dkhk –nang*.

2.2.2 Query term limitations

It is vital that the limitations to query terms be understood. This may be best explained by noting what query terms do *not* cover:

- Capitalisation
- Punctuation
- Synonyms
- Misspellings/typos
- Variant spellings
- Conjugations
- Parts of words
- Wildcard operators

Three of these are very important not to miss. As an example of variant spellings, if one were to request searches for "color," the results would not include searches for "colour." To obtain searches for both, the specification could have to be "color + colour" (see 2.2.3 below). Parts of words is explained by noting that the aforementioned search will not include searches for "colors, colored, coloring, colours, coloured, colouring." Wildcard operators are not permitted, so the only option is to include all variants of a word in the specification.

2.2.3 Query term specification modifiers

It is often necessary to modify the specification query in order to get the right match for what you are looking for. While the query specifications do not allow wildcard characters, they do allow the following basic Boolean modifiers:

Table 3 ***Query term modifiers***

Modifier	Space	Plus	Minus	Double Quotes
Meaning	AND	OR	NOT	Exact phrase
Specification example	Google Trends	Google +Trends	Google -Trends	"Google Trends"
Interpretation example (note that search terms are case insensitive)	"google" AND "trends"	"google" OR "trends"	"google" WITHOUT "trends"	"google trends" EXACTLY
Examples of searches:				
How does Google Home work?	×	✓	✓	×
Is there a trend in what people are googling?	×	×	×	×
Are there trends in what people have googled?	×	✓	×	×
Latest fashion trends	×	✓	×	×
What are the latest trends on Google?	✓	✓	×	×
What is Google Trends?	✓	✓	×	✓
What is trending on Google?	×	✓	✓	×

It may be instructive to notice the difference between exact phrase queries and AND queries. For example, the query specification "MDMA overdose" (sans quotes) will find "MDMA overdose," "Can an MDMA overdose kill," "overdose on MDMA," "How much MDMA will cause an overdose," etc., as all of those phrases contain both words specified in the query, while the exact phrase search '"MDMA overdose"' (i.e., including the quotes) will only find the first two, which both contain the exact two words as specified between quotes. Tying this in to what was explained before, neither will find MDMA overdoses, as the query specification uses the word "overdose," not the word "overdoses."

2.2.4 Maximum number of query terms

A distinction should be made between a single term and multiple terms. There is no documented indication what the maximum length of a single query term may be, but they can be quite long. If I were to guess based on my experience with queries in other APIs, I would say it might be in the order of 1024 characters. This means that, when stringing together phrases using the + (OR) operator, a large number of phrases could be included in a single query term.

However, the API also allows for the comparison of various query terms (cf. the two terms—Google Trends as search term and as topic—in Figure 4, p. 10). When multiple terms are specified, the API returns a series of values for each term, for the specified time range and geographic region, allowing comparison between the different query terms for the same time and place. Technically, this should allow you to compare different times or different geographic regions for the same query term as well, as can be done on the Google Trends website, but I have not set up the Extraction tool to allow for that—you would have to code your own API calls to accomplish that.

You can specify up to 30 terms for the Google Trends Extended Health API call, and 5 for the Google Trends Web API call. However, bear in mind that the 2000-point limit places a practical (or impractical!) limit on this. If you are extracting daily data, for example, then each day is a data point for each term you specify. Thus, if you specify a time frame of 1826 days (five years, unless it spans two leap years), and you enter 2 terms, that already amounts to 3652 data points—far more than the 2000 allowed. In fact, you cannot extract daily data for the full time frame for which Google Trends data are available (from 1 January 2004) in a single API call, so you would typically have to break it up into several calls, depending on the number of query terms you have specified. As an example, for a single query term, you could extract 1 January 2004 to 31 December 2008 (1827 days), then 1 January 2009 to 31 December 2013 (1826 days), and then 1 December 2014 to two days before the present, or 23 June 2019, whichever is earlier. Of course, from 26 June 2019, this approach will require a fourth extraction, and a more utilitarian extraction of 1 January 2004 to 22 June 2009 (2000 days), followed by 23 June 2009 to 13 December 2014 (2000 days) and finally 14 December 2014 through to 4 June 2020 will allow almost an additional year's worth of data to be captured for the same number of extractions. This can be done with the Google Trends Extended for Health API call, since the raw search priorities are returned, and different extractions can be appended to each other (within limits—see the chapter on the Google Trends Extended Health interface below). With the Google Trends Web interface, the 2000-point limit is a non-issue, since the resolution of the returned data is automatically adjusted based on the time frame requested (see Table 6 in the chapter on the Google Trends Web interface), and so up to five query terms can be requested under all circumstances, barring for the List of region values function (API call), which can only return values for one query term.

2.2.5 Freebase IDs

The Google Health Trends API supports Knowledge Graph entities. Knowledge Graph is a technology that was released in 2012[4] and is based on the Freebase technology developed by Metaweb Technologies, which Google purchased in 2010[4,5]. As a lay summary of what this means: Google has machine learning algorithms which classify various items to build up a base of knowledge about certain topics. Searches on Google are also linked to Knowledge Graph entities, and requesting Google Health Trends for a Knowledge Graph entity will retrieve all searches which Google's AI algorithms have linked to that entity. The Knowledge Graph entities are accessed by entering the Freebase Identifier as the query term[1]. Google does have a Knowledge Graph API[6], which could be incorporated into a tool such as this. Using the API would allow one to return the Freebase Identifier for a particular query term. Nonetheless, in the absence of that, Freebase Identifiers can easily be found by searching for the topic on Wikidata (https://www.wikidata.org/). Freebase IDs take the form "'/m/0' followed by 2 to 7 characters" (e.g., the Freebase ID for "Freebase ID" is "/m/0j9kvph")[7]. The use of these Knowledge Graph entities (via the Freebase IDs) is akin to performing an API data extraction for a topic (i.e., the related Freebase ID) vs a search term (i.e., normally specified query term) on the Google Trends website (cf. Figure 4).

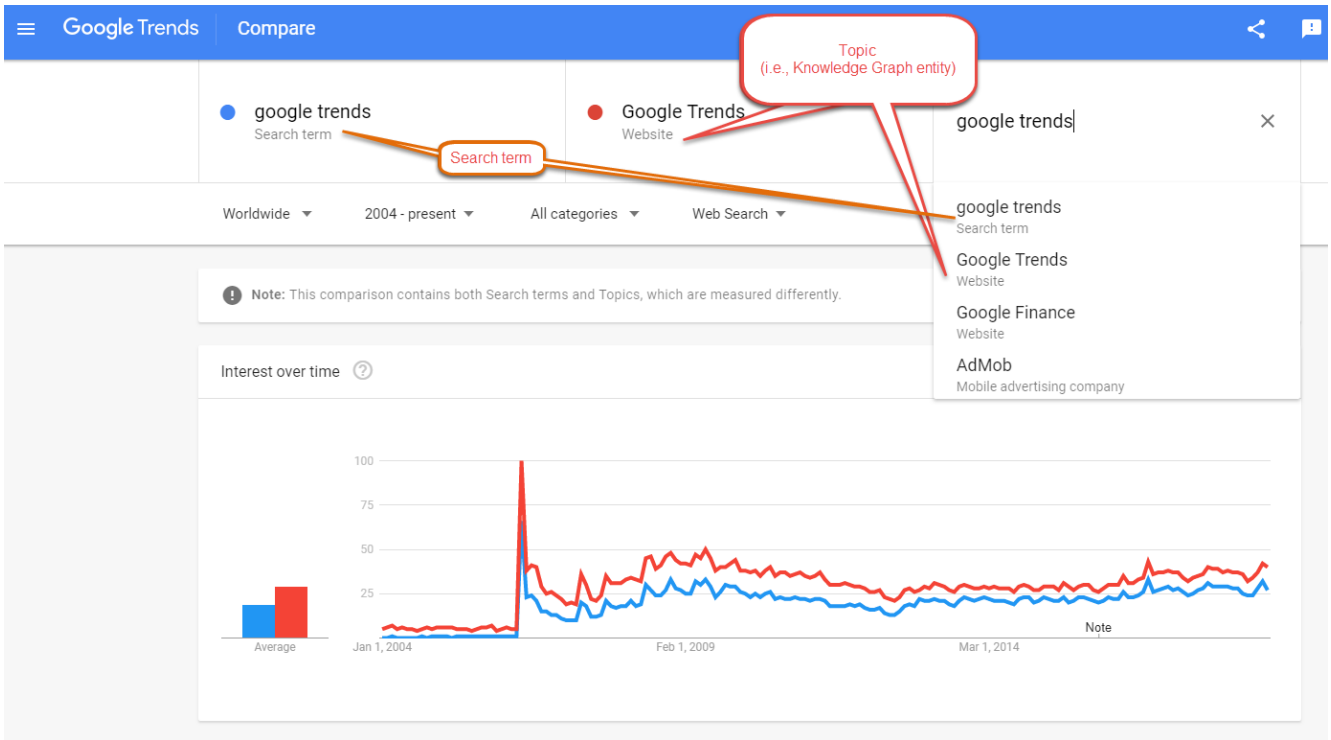


Figure 4 Google Knowledge Graph entities shown as Topics on Google Trends Website

In theory, this is useful, since the trend of the specified Freebase ID would, ideally, be the most complete specification of a particular topic. One immediately apparent advantage of using Freebase IDs is that they include misspellings and typos which, it was indicated above, a normal query term would not include (this is akin to performing a search on Google, and having it show the results for what it perceives to be the correctly spelled entity—Figure 5).

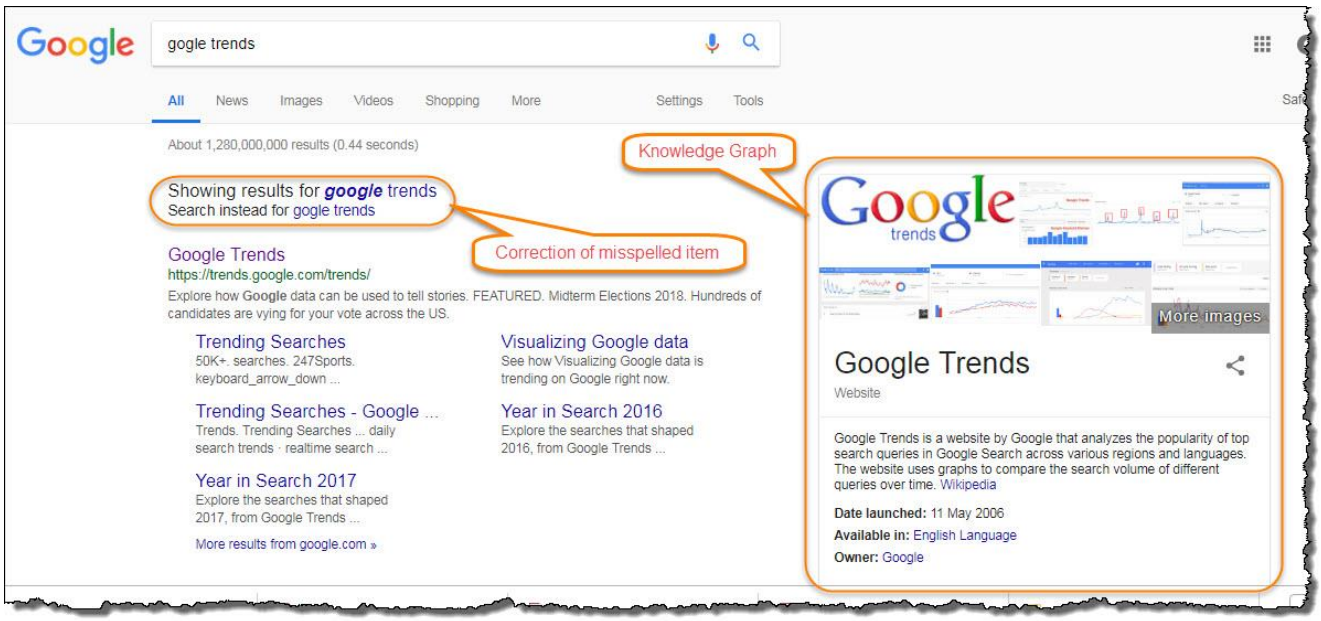


Figure 5 Google search returning what it believes to be the correctly spelled entity

However, this developer's limited experience has shown that there are several potential problems to the Freebase IDs, which always need to be borne in mind.

First, since this is proprietary information, Google would never reveal the exact algorithms that classify various search terms under a particular Freebase ID. This means that by using the Freebase ID, researchers never truly know what they are getting. Any possible errors or erroneous assumptions in that particular algorithm could affect the researcher's results.

Second, while the trends for the Freebase IDs generally do tend to track a well-designed set of query terms closely, this developer have noticed instances where the Freebase ID trend will either under- or overestimate the set of query terms (see the aforementioned point…). There have been instances where our query terms are so specific (e.g., brand names for pharmaceutical agents that have no other meaning), that we are certain that the problem lies with the algorithm missing certain information which it should be including. This is especially at the regional level, where, for example, certain pharmaceutical agents have different brand names in different regions, although these regional errors would obviously affect the worldwide totals.

Third, it seems as if there is a lot of variation for the trends of Freebase IDs in the first portion of the historical data. This developer has not investigated this in detail, but it appears as if the classifications for at least the first four years (2004 to 2008) are worrisome. This is a topic which should be investigated more closely, but in the interim, researchers should remain aware of this challenge.

Researchers should thus guard against summarily assuming that using the Freebase ID is an easy escape to doing the complex investigation required to come up with the best set of terms for the query specification.

This developer's recommendation, then, is that, where available, Freebase ID trends be included with the set of query term specifications chosen by the researcher, so that readers can make their own conclusions relevant to the topic under investigation.

2.3 Countries and regions

The Google API allows data to be returned for a variety of geographical selections.

2.3.1 Background information on geographic specification

Google[8] claims that the geographic specification "takes any of the values depicted here: http://en.wikipedia.org/wiki/ISO_3166-2#Current_codes". Rarely, though, are things so simple. This developer used the list from Wikipedia https://en.wikipedia.org/wiki/ISO_3166-2, but it appears that there are several discrepancies with actual Google data. Resolving this is a work in progress, and the current implication is that not all of the more far-flung locations provided by the tool will actually return usable data.

This developer has not included the use of Nielsen DMAs, even though Google does allow these in the API, as the DMAs are only defined for the USA. Other developers are welcome to add this (see 2.3.2.4 below).

2.3.2 Geographic specification

Geographic specification allows a breakdown of up to three levels. The choice at this level will determine whether lower levels can be set (as will be indicated by conditional formatting on the interface).

The selection the user makes will set the Descriptor value, which cannot be manually set by the user. This is what the extraction tool will use in the URL submitted to Google.

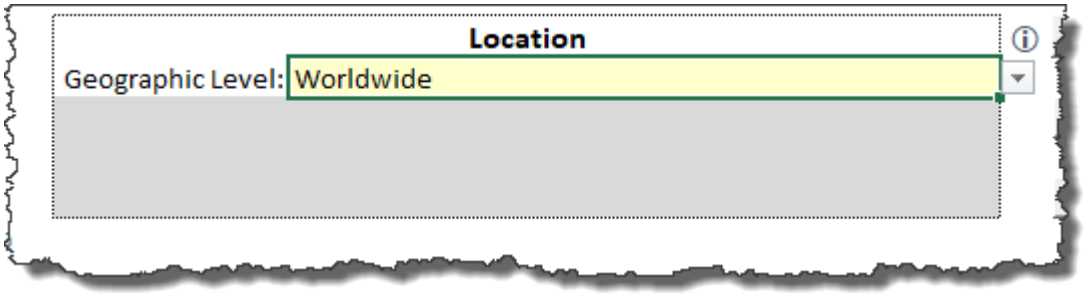


Figure 6 Geographic specification

2.3.2.1 Top geographic level

The first choice with geographic specification is the Geographic level. Choose between Worldwide, Country, or Region. Note that although the Google Trends website (https://trends.google.com) sometimes gives the option of city-level data, this option is not provided by the Google API, and my experience with working with the website is that city-level data is, at best, still unreliable. This may be better in some countries than others, though.

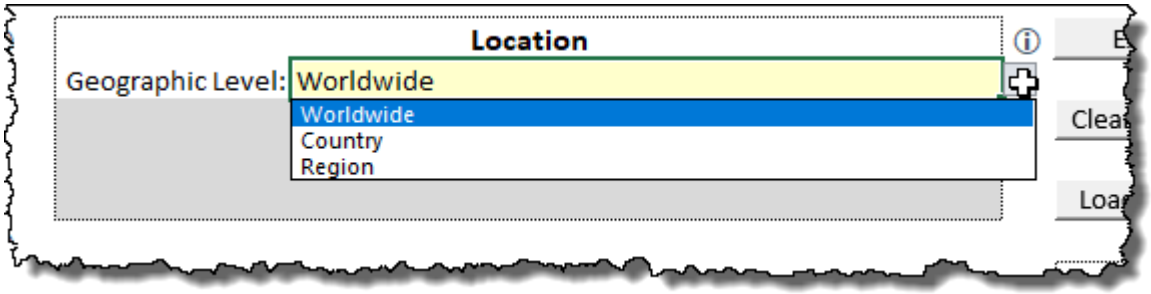


Figure 7 Geographic level

2.3.2.2 Country

If the top geographic level is set to Country or Region, then the Country must be set next. A full list of countries from ISO3166 is provided via a drop-down.

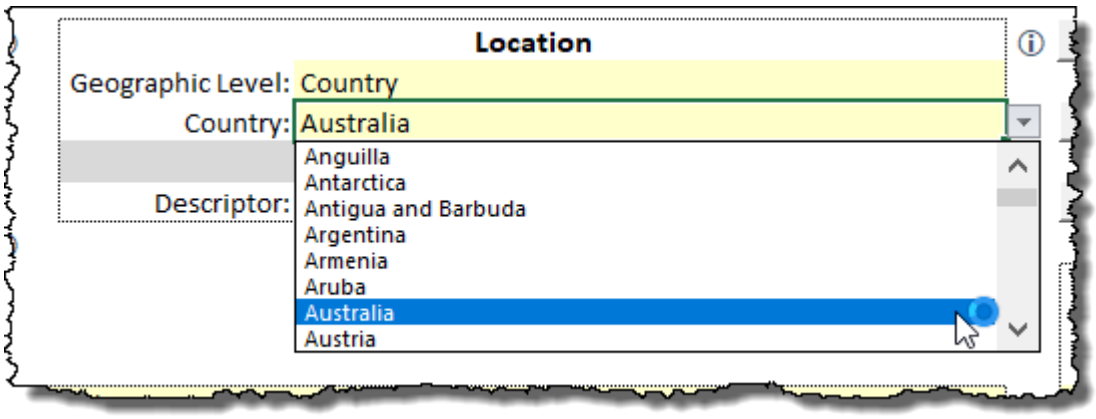


Figure 8 Country

2.3.2.3 Region

If the top geographic level is set to Region, then the Region field can be set from the drop-down list (only the regions for the selected country will be displayed), or—for the Google Trends Extended Health interface only—it can be left blank to perform a comparative extraction for all regions in that country. Because the extraction tool attempts to compare either regions or terms in the resultant output, at present, the extraction tool will not allow a simultaneous region-comparative and term-comparative extraction. Thus, if the Region field is left blank (i.e., all regions in the specified country must be compared), then only one query term can be specified. For a comparison of various query terms, a specific region must be selected.

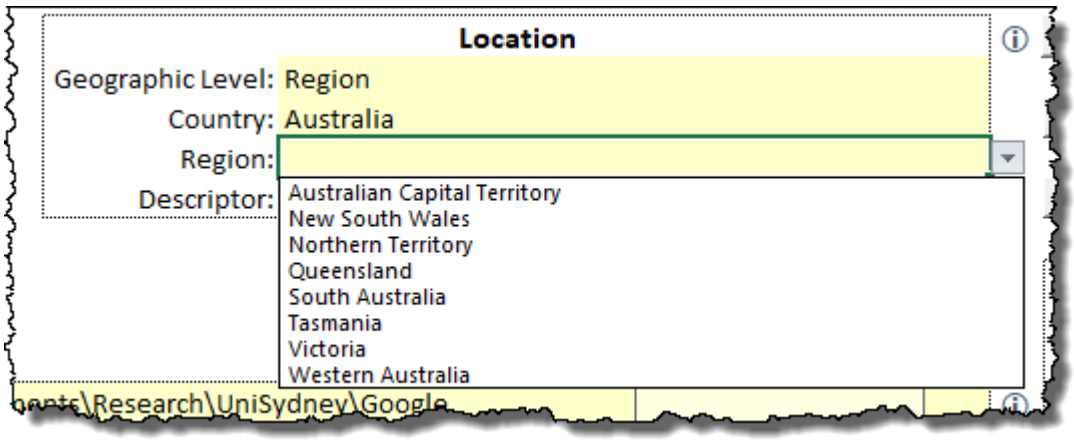


Figure 9 Region

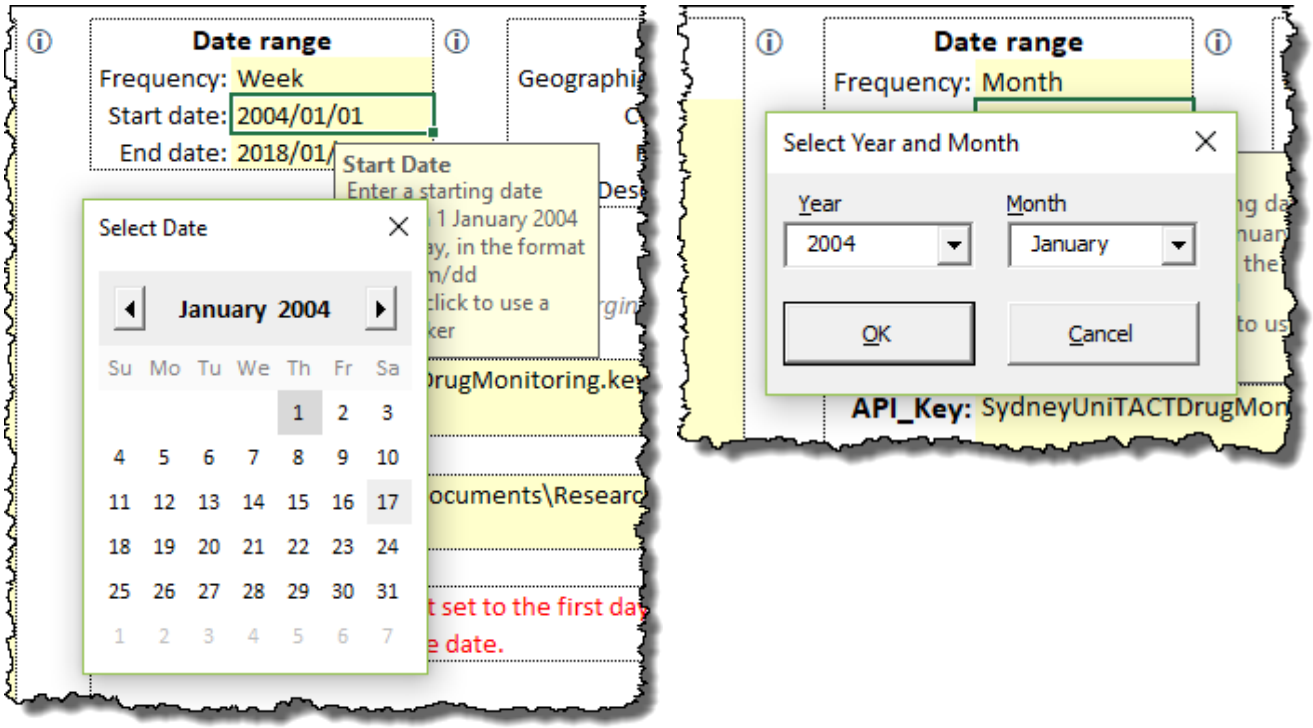
2.3.2.4 DMA

The Google Extended Trends API allows geographic restrictions to be set to the Nielsen Digital Marketing Area codes. For more information, see http://www.nielsen.com/intl-campaigns/us/dma-maps.html and https://www.thebalance.com/what-is-a-designated-market-area-dma-2315180. However, since my research indicated that this is limited to the USA, I did not incorporate it into the tool. Since the tool is open source, anyone wishing to do so, is welcome, although it would be appreciated if any such changes be shared with the community, and that this developer be notified.

2.4 Date selection

Both the Google Trends Extended Health and Google Trends Web interfaces contain areas where start and end dates should be specified. Specifics for each interface are mentioned in the relevant chapters (pp. 22, 35). You can manually enter the dates in any of the date specification cells. Furthermore, double

clicking on any of the cells will open one of two calendar tools which will allow you to easily select a date from a calendar interface. One is the calendar tool supplied by Trevor Eyre (https://trevoreyre.com/portfolio/excel-datepicker) and the other a simplified year/month selector created by this developer (Figure 10). The Google Trends Extended Health interface will offer up the calendar form if the date resolution is set to Day or Week, and the year/month selector if the resolution is set to Month or Year (offering only years in the latter case). The Google Trends Web interface will only offer the calendar form.



Calendar form **Year/Month selector**

Figure 10 Calendar forms

2.5 API key and account name

A valid API key must be supplied in a text file. For security reasons, this file is not stored in the extraction tool. While the API key file could be stored anywhere, it is generally a good idea to store it in the same folder as the extraction tool. Since Version 1.1, the file must contain only two lines of text: The first line must be only the API key supplied by Google, and the second line must be only the account name provided by Google. It can be stored as either a *.txt file, or a *.key file, although the latter is recommended. When the extraction tool is opened, it self-checks to test whether an API key file is specified. If one is specified, it tests to see if the file can be found. If the file can no longer be found, or if there is no file specified, it will test for the existence of a *.key file in the same directory, and if found, will offer to load it for you (Figure 11).

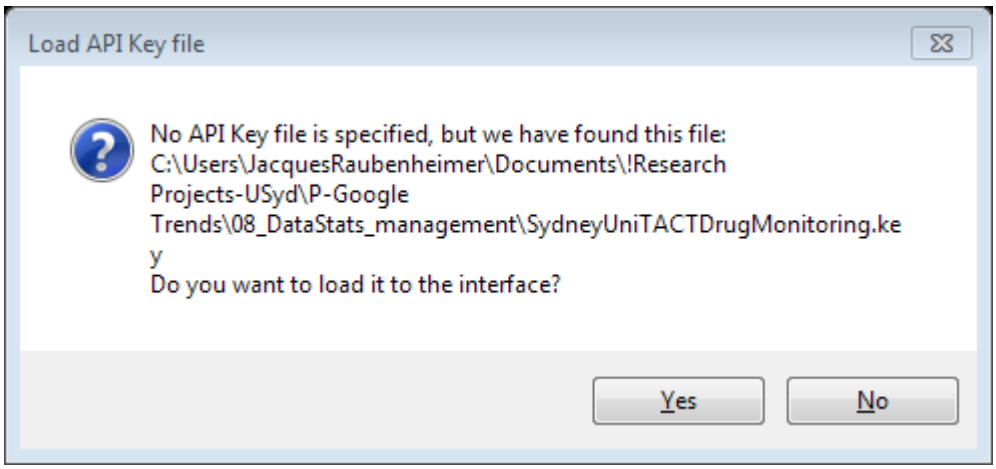


Figure 11 API key file auto loading

To manually load the API key file, simply double click on the file specification area to have a file browser window open (Figure 12), where you can navigate to the appropriate folder, and select the file.

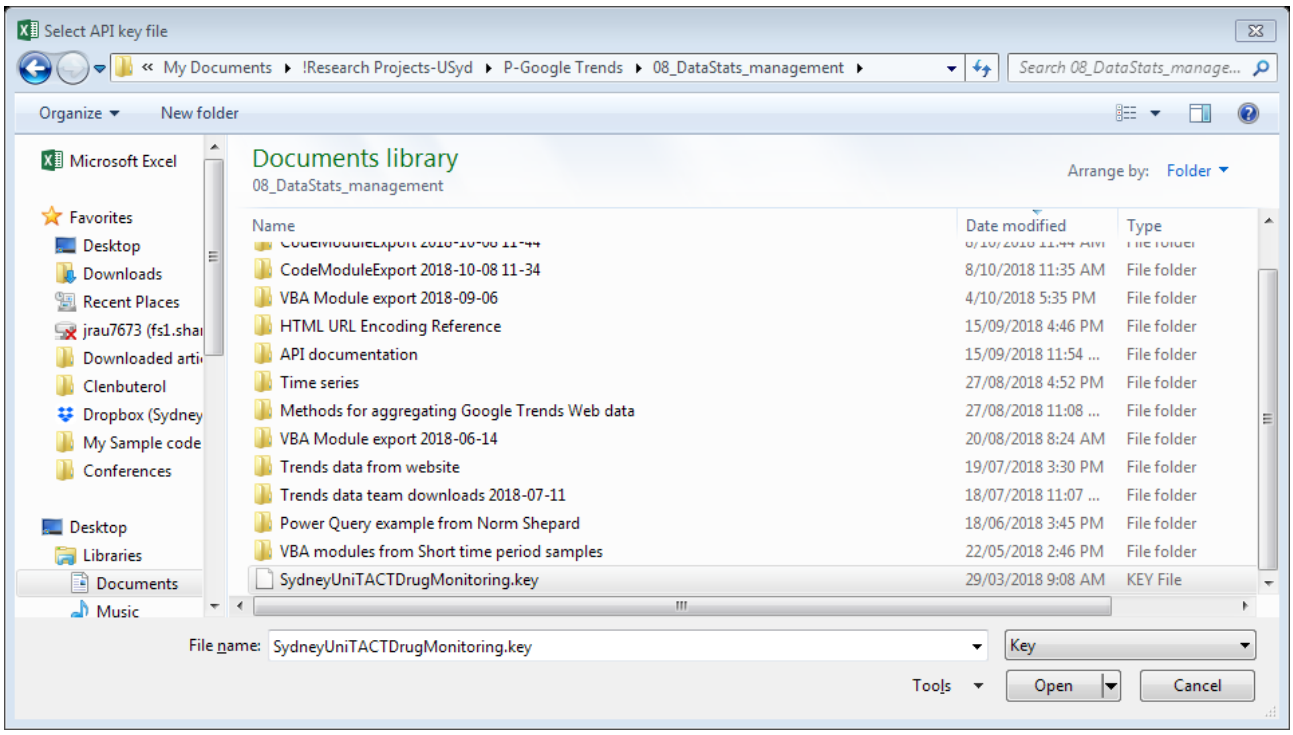


Figure 12 Manually selecting an API key file

The API key is read and used every time an extraction is done, and this extraction will then be registered against the user account. To view the account, click on the View account button (added in version 1.1), which will read the account name, and open the account page in a browser (obviously, an internet connection is required).

2.6 Target file

For both the Google Trends Extended Health and the Google Trends Web interfaces, a target file must be set in which the results of the extraction will be stored. This must be the name of an as-yet non-existent Excel workbook, in a valid directory. The tool will create and save the workbook during the extraction process. Overwriting existing files is not permitted. To select a location and enter a file name, simply double click on the file specification area to have a file browser window open, where you can navigate to the appropriate folder, and specify the file name you want. Note that the total length of the

file path with the file name and file extension may not be longer than 218 characters. This is an Excel restriction, and cannot be circumvented—the interface will warn you if the total length is over this limit. As of version 1.1.0 of the tool, a suggest button has been added (Figure 13) which creates an automatically suggested name based on the inputs set for the extraction. This button can only be used once all the inputs have been set (with the obvious exception of the target file), and may be useful when doing multiple extractions (e.g., repeated extractions on concurrent days), as the file naming is consistent. This autosuggestion may still generate file paths that exceed the 218 character limit, but will warn users when this occurs.

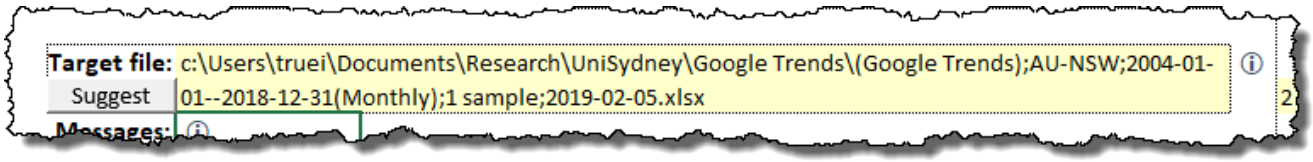


Figure 13 Target file specification

2.7 User Interface error messages

Both the Google Trends Extended Health and the Google Trends Web interfaces contain message areas showing errors, notes or warnings which specify what still needs to be set or changed in the interface before an extraction can be commenced. The aim is to get the message "All inputs are completed. Data extraction can be commenced" (Figure 14). Any error message displayed in red will mean that an extraction cannot be done yet (Figure 15).

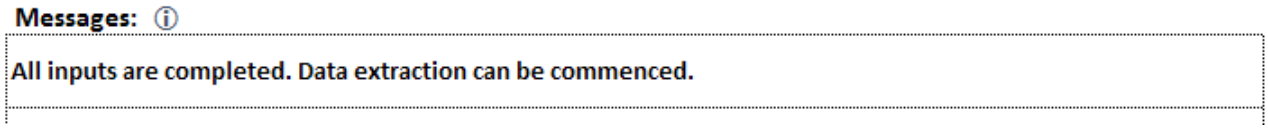


Figure 14 Messages indicating extraction can be commenced

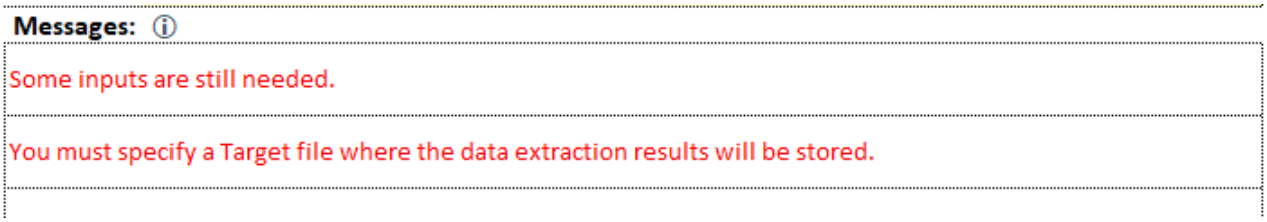


Figure 15 Error messages that still need resolving

I have detailed the process used to create this worksheet-based error checking on my blog: http://insight.trueinsight.za.com/giving-feedback-in-an-excel-worksheet-ui/.

2.8 User Interface help messages

Many elements of the user interface contain small help icons which you can click on to obtain brief help instructions. I got the idea for this from Jan-Karel Pieterse's website (https://jkp-ads.com/Articles/ExcelDataEntryWithValidationHelp.asp).

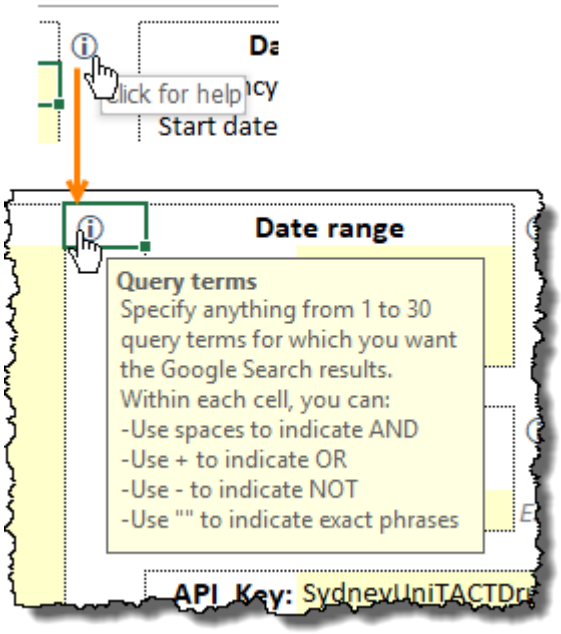


Figure 16 User interface help message

2.9 Clear Specifications

Both the Google Trends Extended Health and the Google Trends Web interfaces contain a **Clear Specifications** button which will clear all input areas (Figure 17).

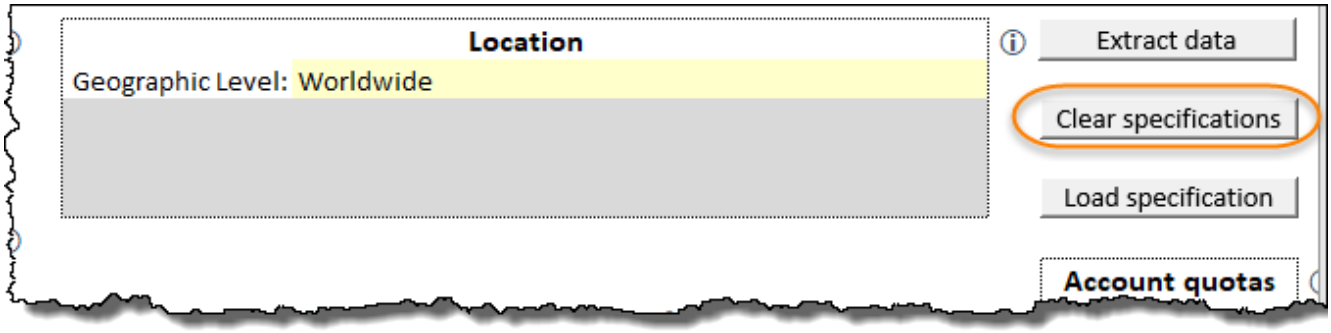


Figure 17 Clear Specifications button

The changes are listed in Table 4. Note that for some elements, the user is first asked via a message box whether the element should be cleared or not.

For the Google Trends Extended Health interface, the target file location and the query term list will be cleared, and the user will be asked whether to clear the number of samples, and whether to clear the geographic location (which will set it to Worldwide). The date resolution will be set to monthly, and the start and end dates set to the full monthly period (1 January 2004 to the last day of the preceding month).

For the Google Trends Web interface, the target file location and the query term list will be cleared, and the user will be asked whether to clear the geographic location (which will set it to Worldwide). The start and end dates will be cleared (defaulting to the entire period), the API function call will be set to Graph, and the search domain set to the default Google Search.

Table 4 Clear Specifications

Interface element	User confirmation sought?	Action
	Google Trends Extended Health	
List of Query terms	No	Cleared
Geographic location	Yes	Set to Worldwide
Target file location	No	Cleared
Date resolution	No	Set to Monthly
Start date	No	Set to 1 January 2004
End date	No	Set to last day of preceding month
No of samples	Yes	Cleared
	Google Trends Web	
List of Query terms	No	Cleared
Geographic location	Yes	Set to Worldwide
Target file location	No	Cleared
Start date	No	Set to 1 January 2004
End date	No	Cleared (i.e., default to present time)
Search domain	No	Set to Google Search
Function	No	Set to Graph

2.10 Loading previous specifications

Because it is quite conceivable that a researcher would want to repeat (or slightly modify) a previous query extraction, the tool allows users to select a previous extraction file (i.e., the file that was specified, in the Target file area), and the query settings from that file will be copied to the relevant interface. Note, though, that while both of the interfaces contain a **Load specification** button, the tool will load the specification to the relevant interface (e.g., if you click on **Load specification** from the Google Extended Health Interface, and select a file that was created by the Google Trends Web interface, the specifications will still be loaded to the Google Trends Web interface).

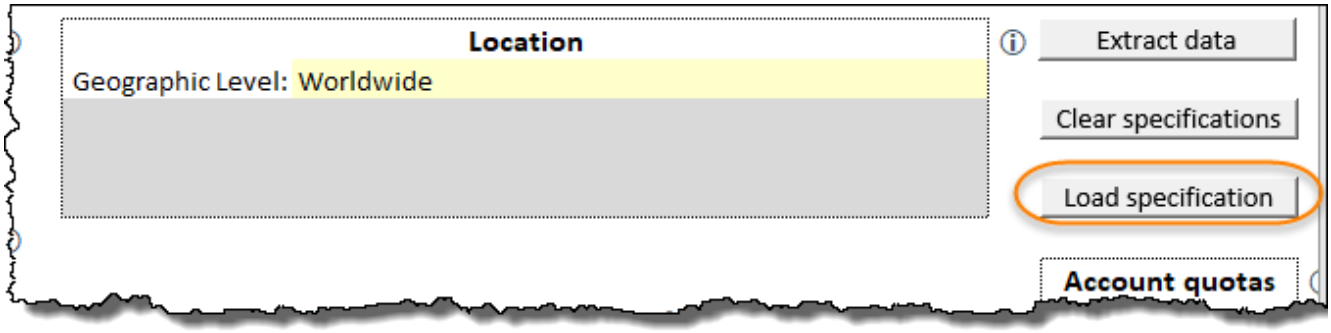


Figure 18 Load Specification button

2.11 About

Clicking on the about page takes you to a page with some information about the extraction tool. Click on Return to go back to the sheet from which you clicked About. The Help button opens this file.

This extraction tool for the Google Extended Trends API for Health was developed by Jacques Raubenheimer
of the Australian Government's National Health and Medical Research Council-funded
Translational Australian Clinical Toxicology programme at the University of Sydney
Version v2.0.0
©Jacques Raubenheimer
©2019
This work is released under the GPLv3 License: https://opensource.org/licenses/GPL-3.0

Special thanks to Google for making the Google Trends data available and providing the Google Extended Trends API for Health
To visit the Google Trends website, go to: https://trends.google.com
For the Google Trends API Explorer, go to: https://developers.google.com/apis-explorer/#p/trends/v1beta/
For the Google API Terms of Service, go to: https://developers.google.com/terms/
To request permission to use the Google Trends API, go to: http://bit.ly/2KyqDYW

The Calendar control is the public release of CalendarForm v1.5.2 from Trevor Eyre ©2015–2016: https://trevoreyre.com/portfolio/excel-datepicker/

THE UNIVERSITY OF SYDNEY

Help Return

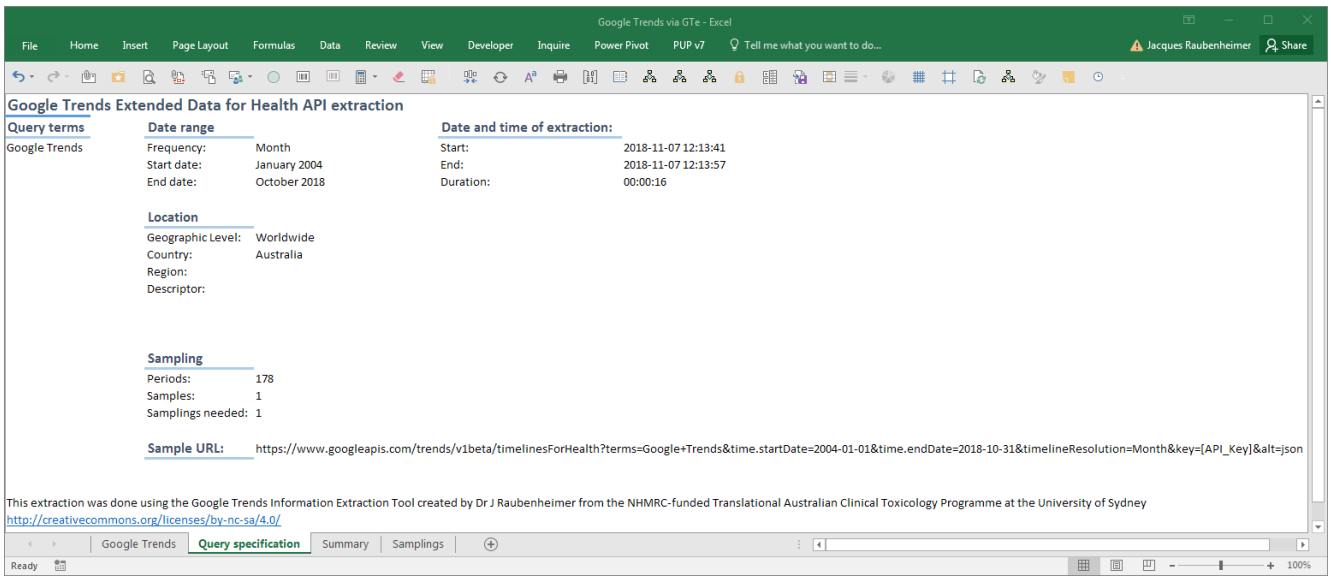
Figure 19 About page

2.12 Program output

Whenever an extraction is run, the tool creates a workbook (as specified in the Target file input area—cf. p. 15) containing the results obtained from Google's servers to the API request. The Google Trends Extended Health interface produces only one kind of output, while the Google Trends Web interface provides slightly different output based on which function (cf. p. 35) is used. More details on the output of each interface will be provided in the relevant sections, but some common elements are mentioned here.

2.12.1 Worksheets

Generally, the tool creates a workbook with worksheets for the data, possibly a chart (see below), and then two more worksheets—one is hidden, and contains the details of the request so that it might possibly be reloaded in the future (cf. section 2.10—Loading previous specifications on p. 18), the other is the **Query Specification** sheet (Figure 20) which contains a more human-readable (i.e., formatted) listing of the extraction was done. This is useful for later auditing purposes. It summarises the values from all the input areas, and records details of when the extraction was done.



Google Trends Extended Data for Health API extraction

Query terms		Date range		Date and time of extraction:	
Google Trends		Frequency:	Month	Start:	2018-11-07 12:13:41
		Start date:	January 2004	End:	2018-11-07 12:13:57
		End date:	October 2018	Duration:	00:00:16

Location

Geographic Level:	Worldwide
Country:	Australia
Region:	
Descriptor:	

Sampling

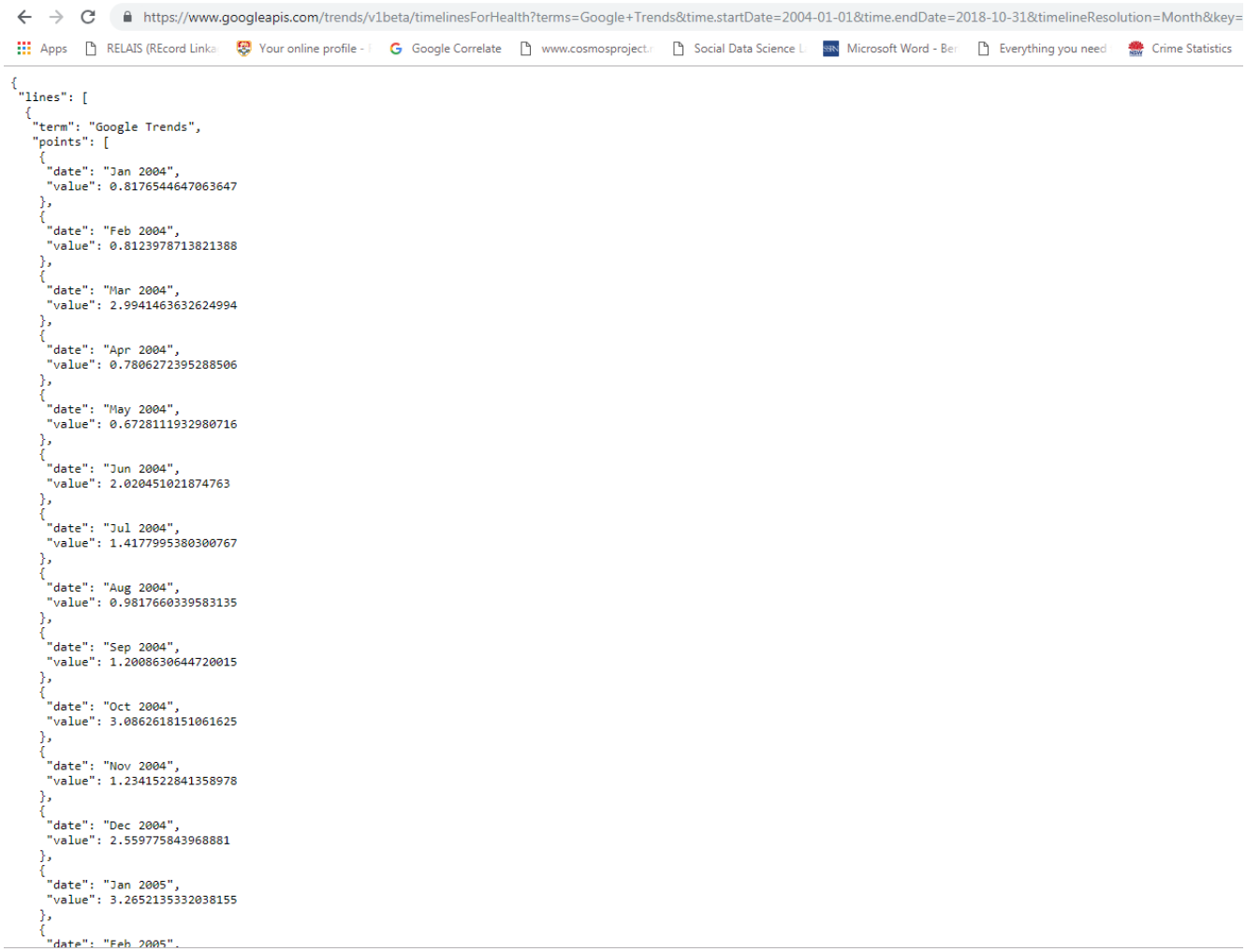
Periods:	178
Samples:	1
Samplings needed:	1

Sample URL: https://www.googleapis.com/trends/v1beta/timelinesForHealth?terms=Google+Trends&time.startDate=2004-01-01&time.endDate=2018-10-31&timelineResolution=Month&key=[API_Key]&alt=json

This extraction was done using the Google Trends Information Extraction Tool created by Dr J Raubenheimer from the NHMRC-funded Translational Australian Clinical Toxicology Programme at the University of Sydney
http://creativecommons.org/licenses/by-nc-sa/4.0/

Figure 20 Query Specification sheet (Google Trends Extended Health)

A sample of the URL containing the API request (as created from all the provided inputs) is also recorded. If the API key is added, this URL can be pasted into a browser, and should deliver a result (Figure 21)—note that this will count as one request against your quota (cf. section 1.11—Data Limitations, p. 4).



https://www.googleapis.com/trends/v1beta/timelinesForHealth?terms=Google+Trends&time.startDate=2004-01-01&time.endDate=2018-10-31&timelineResolution=Month&key=


```
{
 "lines": [
  {
   "term": "Google Trends",
   "points": [
    {
     "date": "Jan 2004",
     "value": 0.8176544647063647
    },
    {
     "date": "Feb 2004",
     "value": 0.8123978713821388
    },
    {
     "date": "Mar 2004",
     "value": 2.9941463632624994
    },
    {
     "date": "Apr 2004",
     "value": 0.7806272395288506
    },
    {
     "date": "May 2004",
     "value": 0.6728111932980716
    },
    {
     "date": "Jun 2004",
     "value": 2.020451021874763
    },
    {
     "date": "Jul 2004",
     "value": 1.4177995380300767
    },
    {
     "date": "Aug 2004",
     "value": 0.9817660339583135
    },
    {
     "date": "Sep 2004",
     "value": 1.2008630644720015
    },
    {
     "date": "Oct 2004",
     "value": 3.0862618151061625
    },
    {
     "date": "Nov 2004",
     "value": 1.2341522841358978
    },
    {
     "date": "Dec 2004",
     "value": 2.559775843968881
    },
    {
     "date": "Jan 2005",
     "value": 3.2652135332038155
    },
    {
     "date": "Feb 2005",
```


Figure 21 API request URL entered into a browser

2.12.2 Charts

Three functions produce charts. Users may have to tweak the charts somewhat to get presentation-ready output (e.g., the title might obscure data—for Figure 49, the title was adjusted to correct this problem—or there may be too many bars in the List of Region Values chart, or the user may want to create a split level horizontal date axis). Each chart is discussed in the relevant output section in the following chapters—3.6.5 (p. 33), 4.5.1 (p. 40), 4.5.5 (p. 44).

2.12.3 Log

As of version 2.0.0, the program creates and updates a log file. The file is stored in the same directory as the program file, with the name "Google Trends Information Extraction Tool.log."

Four different events are logged: Opening, closing, GTe extractions, and GTweb extractions. Opening and closing simply record the version number, and the date and time. GTe and GTweb extractions log the type of extraction, the version number, and then all the information relevant to the extraction (much the same information as is stored in the Query Specification worksheet—cf. Figure 20 above). Each piece of information consists of a tag, a colon, and then the supplied/selected information, and all the pieces of information are comma-separated.

Log entries are appended in chronological order, not in distinct sections for each type of event. This allows a scan of the file to reveal how the program was used over time, without the need of switching between sections. The trade-off is that successive entries do not necessarily share the same format, making extraction and analysis of the logged information slightly more complex. If you really want to analyse this information, copy and paste the log file into Excel (making certain that the Text-to-columns settings are off). All the data should be in Column A. Add a filter to that, filter for the type of log entry you want, and then copy and paste the information to a new worksheet, where Text-to-columns, separating by commas and colons, will parcel the information out into analysable columns.

Chapter 3 Google Trends Extended Health

The Google Trends Extended Health API data offer several advantages over the Google Trends Web data, and are to be preferred for serious research work. First, the data are not scaled to 100. This means that, for example, data for one country can be extracted, and can be compared to data for another country, or data for different time frames. This allows one to overcome the 2000-point limit, which cannot be done in any other way than the limiting direct comparison of the Google Trends Web graph.
Because the data are not scaled to 100, it is also easier to extract multiple samplings, which can be built into complete samples, allowing for a better sampling adequacy and more precise estimation.
Another problem with the scaling to 100 of the Google Trends Web data is that data from successive time frames cannot be appended to form a complete set for the total time frame, since each individual time frame will have its highest value scaled to 100, while in reality there may be massive differences between the highest points of various of the smaller time frames. One is thus restricted to the time resolution allowed for the full time specification of interest when using Google Trends Web data, whereas with Google Trends Extended Health data, one can, if desired, request the finest resolution (daily data) for successive time frames, and append them.
Furthermore, the Google Trends Extended Health data are not rounded to the nearest integer (which is the case with Google Trends Web data), which allows for more precise estimation.

3.1 Date specification

Google Trends Extended Health data can be requested for any period from the start of 2004 to two days before the present, and can be returned as daily, weekly, monthly, or yearly search volume date. The settings to determine in what form these data will be returned are in the Date range specification (Figure 22).

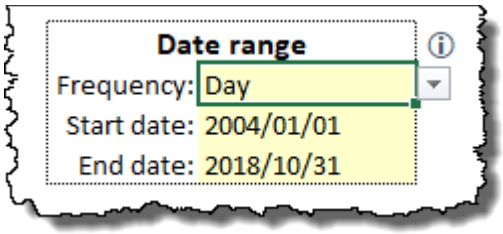


Figure 22 Google Trends Extended for Health Date range specification

The first task in choosing a date range is to choose the date resolution needed for one's analysis in the Frequency section. Data will be aggregated by Google according to the resolution chosen, being one of Day, Week, Month or Year (Figure 23). While the values stored in the start- and end date cells are always exact dates, the formatting of the cells are changed to reflect either an exact date (for Day/Week), a year and month (for Month) or a year only (for Year). When the user changes the Frequency to anything other than Day, the tool should (if all is working as planned) attempt to change the start and end dates to the start and end of the relative period (i.e., the start date should be set to the first day of the week/month/year currently input, and the end date to the last day of the week/month/year currently input). If it fails to do this, an error message will be displayed in the message area (cf. section 2.7—User Interface error messages). The start and end dates must be set to the start and end of the respective periods before an extraction can be done, so that the corresponding first and last periods of the entire time period will not contain incomplete data.

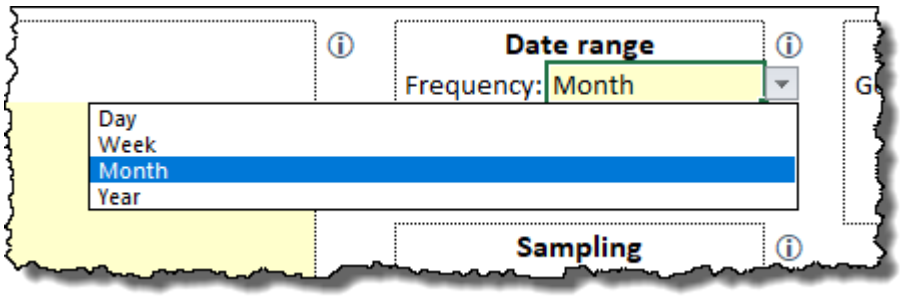


Figure 23 Setting the date resolution

Unlike with the Google Trends Web interface, the start- and end dates cannot be left blank. The user must provide a start date equal to, or later than, 1 January 2004, and an end date, later than the start date, but earlier than, or equal to, three days before the current date (Google Trends Extended Health data are only available up to two days before the present day). While dates can be input manually, double clicking on either of the date cells will launch one of two forms for date selection.

3.1.1 Date specification via the calendar form

If the Frequency is set to Day or Week, then double clicking on the start- or end date cells will open the calendar form. The calendar form has buttons to move to the previous or next month of the currently displayed month, but you can more easily jump to a specific year and/or month by clicking on the year, or the month, to display a dropdown list of the available options (Figure 10). Clicking on any particular date will close the form and load that date to the relevant start- or end date cell.

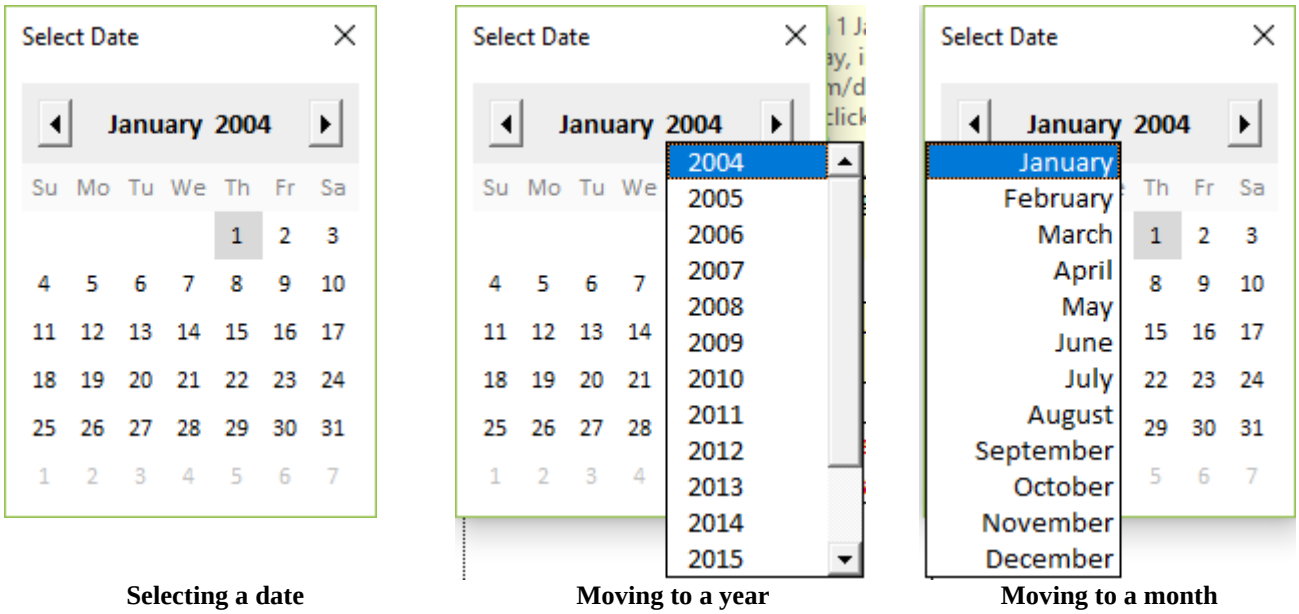
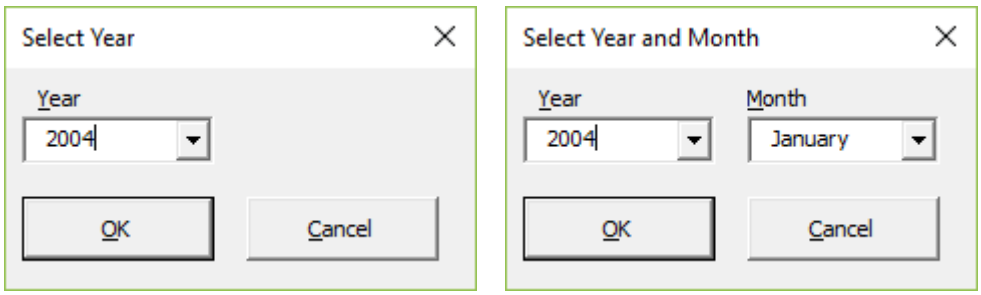


Figure 24 Using the Calendar form

3.1.2 Date specification via the year/month selector

If the Frequency is set to Month or Year, then double clicking on the start- or end date cells will open the Year/Month selector-for a Frequency of Month, both Year and Month drop-downs will be displayed, and for a Frequency of Year, only a Year dropdown will be displayed (Figure 25). When the OK button is clicked on the Year/Month selector, it will add the date to the relevant cell (i.e., start- or end date) as, respectively, the first or last date of that period (year or month).



Year only **Year and Month**

Figure 25 Google Trends Extended Health date specification: Year/Month selector

3.2 Sampling

One of the key innovations of this tool is the strategy it uses to extract and combine multiple samples of data from the Google data store, to obtain more precise estimates of the trend for the specified query terms. A method was devised which limits the number of calls needed (thus simultaneously favouring the Google requirement of not placing undue load on their servers, and minimising the usage of a researcher's daily quote), but still allows researchers to obtain multiple estimates without having to intersperse each extraction with an interval of 24 hours, or having to obtain multiple API keys[†]. There is, in theory, one approach that can reduce the number of extractions even more (about half the current process), but the 2000-point limit makes the implementation of this alternative so restrictive, that is was considered unsuitable for most use cases, and was not included in the tool. However, the limitation of the method which was implemented is that no more samples can be extracted than the number of periods specified by the date selection. This is itself dependent on the date resolution (e.g., 1 January 2004 to 31 December 2018 is 5479 days).

More details of the calculations and strategies behind the multiple sampling approach were published in:

Raubenheimer, J. E. (2022). A practical algorithm for extracting multiple data samples from Google Trends extended for Health. *American Journal of Epidemiology*, *191*(9), 1666–1669. https://doi.org/10.1093/aje/kwac088

Researchers will need to know how many samples to extract to obtain a good estimate of the values in their time series. The suggested approach is to download a small number of samples (e.g., 50), and then use the information calculated from that to assess the adequacy of the sampling, and whether additional samples are needed. See section 3.6.3 (p. 30) for more details.

When a number of samples are requested, the tool will send through a number of requests to the Google servers equal to one less than double the number of requested samples.

[†] As an example, Matsa et al.[11] used four API keys to access multiple samples of data.

N.B.! Shortly after the publication above appeared, the multiple sampling seemed to stop working—multiple requests for different date ranges all return the same value for the same date within those different ranges. The practical implication is that the multiple-sampling-within-a-day feature is broken by changes to the API, and researchers wanting to do due diligence are back to one sample per day. *Sigh*! I have no idea whether this is related to my publication or not (I am *not* accusing Google of changing to API to block my approach), or whether it might perhaps be related to the changes discussed in this must-read paper:

Hermanus Myburgh, P. (2022). Infodemiologists Beware: Recent Changes to the Google Health Trends API Result in Incomparable Data as of 1 January 2022. *International Journal of Environmental Research and Public Health 2022, Vol. 19, Page 15396*, *19*(22), 15396. https://doi.org/10.3390/IJERPH192215396

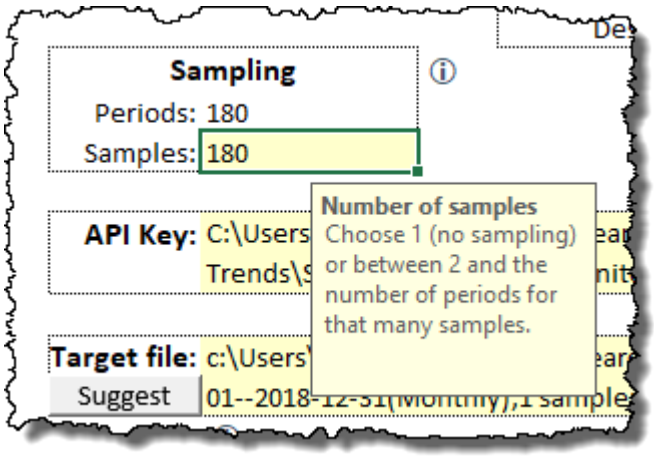


Figure 26 Setting the number of samples to be extracted for Google Extended Trends for Health data

3.3 Multi-term requests

The Google Trends Extended Health API allows up to 30 query terms to be submitted, although it is constrained by the 2000-point limit. This is useful for comparison of various trends, and when extracting monthly data, depending on the date range, ten terms will easily be comparable in this manner for the foreseeable future (up to September 2020).

A multi-term request is specified by simply entering each query term phrase in the Query terms input area. As each term is added, the conditional formatting will make the next SearchTerm phrase visible. Conditional formatting will also highlight duplicate query terms, as these are obviously nonsensical.

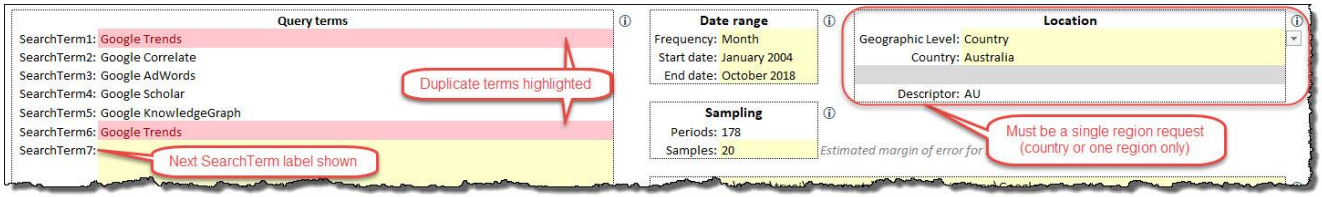


Figure 27 Specifying a multi-term request

When a multi-term request is submitted, a pair of Samplings- and Summary sheets are created for each term—see Figure 30 and Figure 32 for similar output from a multi-region request, and 3.6–Output below.

3.4 Multi-region requests

Because of the 2000-point limit, it makes no sense to request a comparison on even a single term for all countries. The user will thus have to extract data for the countries which they wish to compare, and for

the exact same time frame using the same time resolution. It should be noted that query term specifications are often country-specific, so the validity of inter-country comparisons should be carefully considered.

Having said this, the tool allows for a multi-region request *within* a country to be made (note that the 2000-point limit may still make this unviable for countries with a large number of regions). To specify a multi-region request, provide only one query term specification (multi-region, multi-term comparisons are not provided by the tool), and then in the Location input area, specify Region for Geographic Level, select the country, and leave the Region input blank (Figure 28).

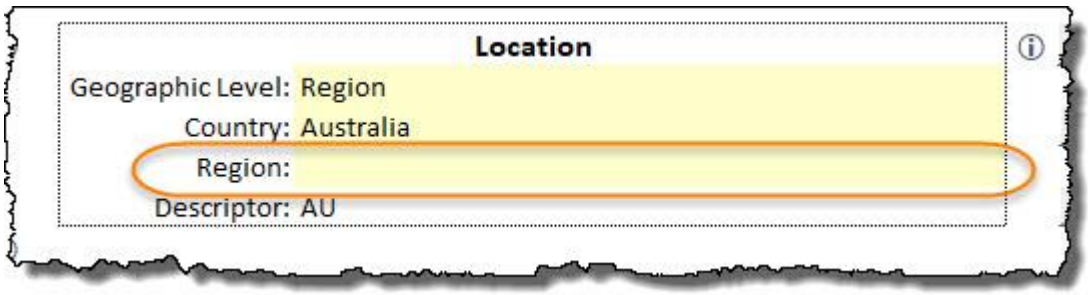


Figure 28 Specifying a multi-region request

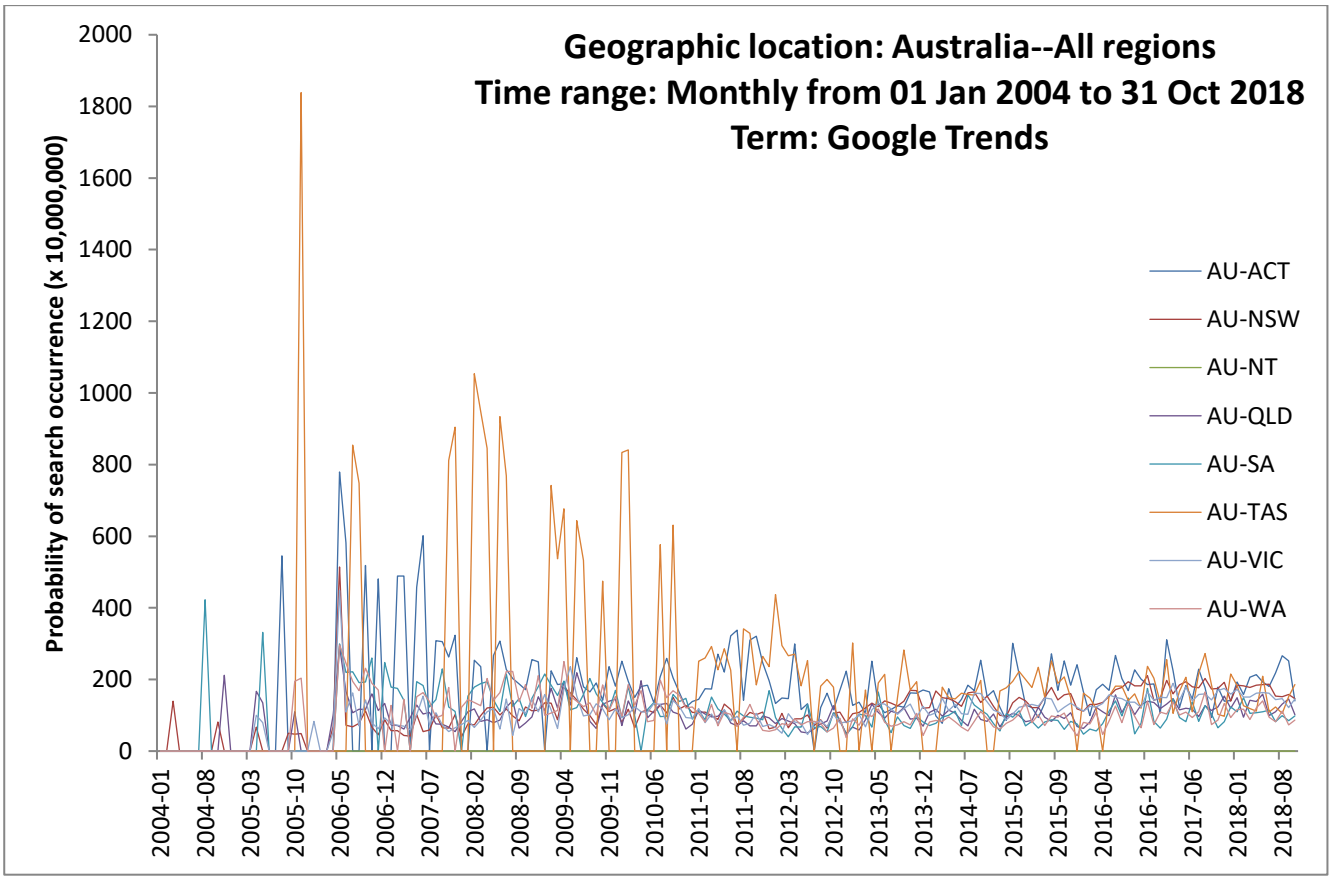


Figure 29 Example of a plot from a multi-region request

3.5 Extraction process

Because the method used by this tool issues a number of http requests to the Google servers equal to one less than double the number of samples required, it should be noted that the number of eventual requests will be quite large, and the progress of the tool is quite slow. There are two reasons for this. First, http requests are inherently slow—this is similar to a request to download a small web page for each sampling (i.e., each http request). Second, because of the rate limit referred to above (p. 2), the tool actually

monitors the speed of the requests, and if the requests appear to be in danger of exceeding the rate limit, will throttle back the requests to keep them within the rate limit (my experience of using this tool has been that this is not necessary, as the speed of the http requests is significantly slower than the rate limit). All of this means that the process is very slow, and depending on network and server traffic, a query with 500 samples (requiring 999 http requests) will take upwards of 20 minutes to complete, during which time you will not be able to use Excel for anything else. And if it is a multi-region request, then the number of samplings will be multiplied by the number of regions! The tool will reflect its progress in the Excel status bar, although Excel may not update this as regularly as it should. Also, if you see a "(not responding)" message in the Excel title bar, relax, and be assured that it is slowly chugging away at getting your data.

3.6 Output

The Google Trends Extended Health output always contains three types of worksheets, apart from the Query Specification sheet shown in Figure 20 (p. 20). Each of these classes of sheets will be discussed below.

Note that, when the API call requests data for a single query term for a single region, one Samplings sheet (cf. section 3.6.1 below) is created and one Summary sheet (cf. section 3.6.2 below) is created. However, for either multi-term requests (cf. section 3.3—Multi-term requests) or multi-region requests (cf. section 3.4—Multi-region requests), one Samplings sheet (cf. section 3.6.1 below) is created for each query term/region, and one Summary sheet (cf. section 3.6.2 below) is created for each query term/region, and in addition a Master Summary sheet (cf. section 3.6.3 below) is created which pulls together the Mean columns from all the individual summary sheets.

3.6.1 Samplings (Data) sheets

These sheets contain the raw data returned by the Google servers in response to the API request. Each sheet will always have a Date column as the first column, containing all the dates from the specified Start date to the specified End date, with an interval equal to the date resolution set in the Frequency drop down (cf. section 3.1—Date specification). Even though exact dates are used, the dates are formatted according to the selected date resolution (e.g.,. if yearly data are requested from 2004 to 2018, the dates will be 2004/01/01, 2005/01/01…, 2018/01/01, although Excel will display only the years 2004, 2005…, 2018).

Next, these sheets will contain one column for every sampling requested from Google needed to make up the number of samples specified in the Samples input (cf. section 3.2—Sampling), which will be one less than double the number of specified samples. Note that the first sampling will always contain values for the whole date range, and that successive samplings will then contain either values from the first date up to some point short of half the date range (all the odd columns), or values from some point after half the date range, to the end of the date range (this is a simple explanation of the sampling strategy I devised to extract multiple samples from the Google Trends API).

If some cells are encountered which contain zero values (cf. section 1.12—Data scaling and zero values, p. 4), these cells are deleted and marked with yellow cell shading. These are different to the cells omitted from more than half the date range due to the sampling strategy. If cells containing zero have been deleted, a cell comment is added to cell A1 of each Samplings worksheet to indicate the number of cells affected. Figure 30 shows the Samplings sheet from one region of a multi-region request. The cell comment indicator in cell A1 is visible, as are three deleted cells (which would have contained zero

values)—highlighted in yellow. Note also the blank sections of all the even-numbered columns, which is a result of the strategy used to obtain multiple samples from the API.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	Date	Sampling 1	Sampling 2	Sampling 3	Sampling 4	Sampling 5	Sampling 6	Sampling 7	Sampling 8	Sampling 9	Sampling 10	Sampling 11	Sampling 12	Sampling 13	Sampling 14	Sampling 15	Sampling 16	Sampling 17	Sampling 1
154	2016-09	173.509455		215.9699181		216.5170887		433.1848966		173.5691438		345.1518991		172.7621456		301.2207765		260.5072128	
155	2016-10	209.8552456		166.3296108		82.94366681		250.9633269		209.414273		166.6672435		208.6608936		290.912367		208.4436321	
156	2016-11	168.4919825		210.0945396		124.7158495		167.1499852		249.8492091		126.020996		125.8132513		251.7490927		166.8013593	
157	2016-12	240.28309		95.20696379		191.4713019		239.9749095		192.4257066		190.5832303		288.5449559		192.4753342		96.15976089	
158	2017-01			85.42944043		85.68997203				85.59222047		84.82512005		85.48265564		170.1070113		170.4777752	
159	2017-02	286.3185339		328.1532092		285.8560746		244.3583373		203.7319049		245.3199763		327.0779602		410.4119108		329.2699405	
160	2017-03	137.3700502		240.3189366		238.7877449		205.110064		240.7285133		272.9485111		275.0411675		205.3135922		238.0505676	
161	2017-04	80.01371002		159.9786344		159.6048207		119.6226988		158.6010008		79.51556819		79.91037217		80.21543744		119.5498939	
162	2017-05	147.136128		222.7635531		111.3159077		110.5816038		294.7997198		147.6094771		221.5881699		184.2823833		147.1042897	
163	2017-06	166.8241817		168.0215766		125.394292		125.296285		125.0597205		125.0407893		251.3665089		167.5066391		125.63207	
164	2017-07	328.6710677		287.7592459		204.20803		164.5687089		164.6710154		206.104258		205.7595272		163.8590278		326.542549	
165	2017-08	146.2165448		181.8628754		144.8251266		145.8747481		146.3821385		145.2167223		218.3152574		109.2967683		145.1142103	
166	2017-09	80.00069788		78.45432104		78.60503185		79.47479413		80.24662323		79.04813169		78.59438288		78.67130489		79.50233828	
167	2017-10	181.6021401		181.449666		217.7340941		182.5232083		108.9961123		109.2722286		145.4419456		145.2696539		145.5849932	
168	2017-11	75.94261636		114.4068162		229.5021593		151.7703705		191.6211424		76.3636244		114.1676639		151.4923746		114.7053964	
169	2017-12	179.927284				90.53793366		225.898264		90.70137529		91.10886392		180.8329472		89.58856618		181.6989531	
170	2018-01	163.2619586		162.8301454		243.7149813		161.7618257		160.6769319		242.3857952		122.2797319		121.7916255		321.5586	
171	2018-02	203.5337798		80.061609		162.102858		121.4463353		162.0913352		201.9722414		161.7729408		120.8850508		162.2718505	
172	2018-03	181.0435421		218.5541433		216.6936476		144.9152591		181.1763664		179.7120684		144.2262819		145.0601774		252.2517922	
173	2018-04	187.17914		149.7950709		112.6189721		112.5420876		112.8778166		112.259202		335.7244299		261.7726144		225.6851149	
174	2018-05	142.231343		213.5303859		283.8239969		212.9205497		213.0034747		177.9217367		282.8606844		176.7239166		176.9058218	
175	2018-06	157.3712986		156.1283333		235.274573		157.4000169		196.1323692		156.5724735		196.149012		195.297334		195.0500546	
176	2018-07	268.806473		230.7745195		191.2641889		192.7102049		232.3754484		154.2564548		115.1478502		153.9092548		230.6477434	
177	2018-08	180.5737055		144.6946988		322.6829452		216.5890966		180.5570365		287.2388985		288.8546685		324.3906149		287.3114913	
178	2018-09	279.4631297		281.3563422		200.1398394		281.0055492		399.7242896		238.5063225		237.288519		278.3090386		240.7871047	
179	2018-10	111.9315317		148.4087241		148.029576		185.816647		149.5592797		112.1389365		112.0825905		112.3647814		148.8368797	

Plot of Regions | Query specification | Means of all Region series | Summary-AU-ACT | Samplings-AU-ACT | Summary ...

Figure 30 Samplings sheets from a multi-region request

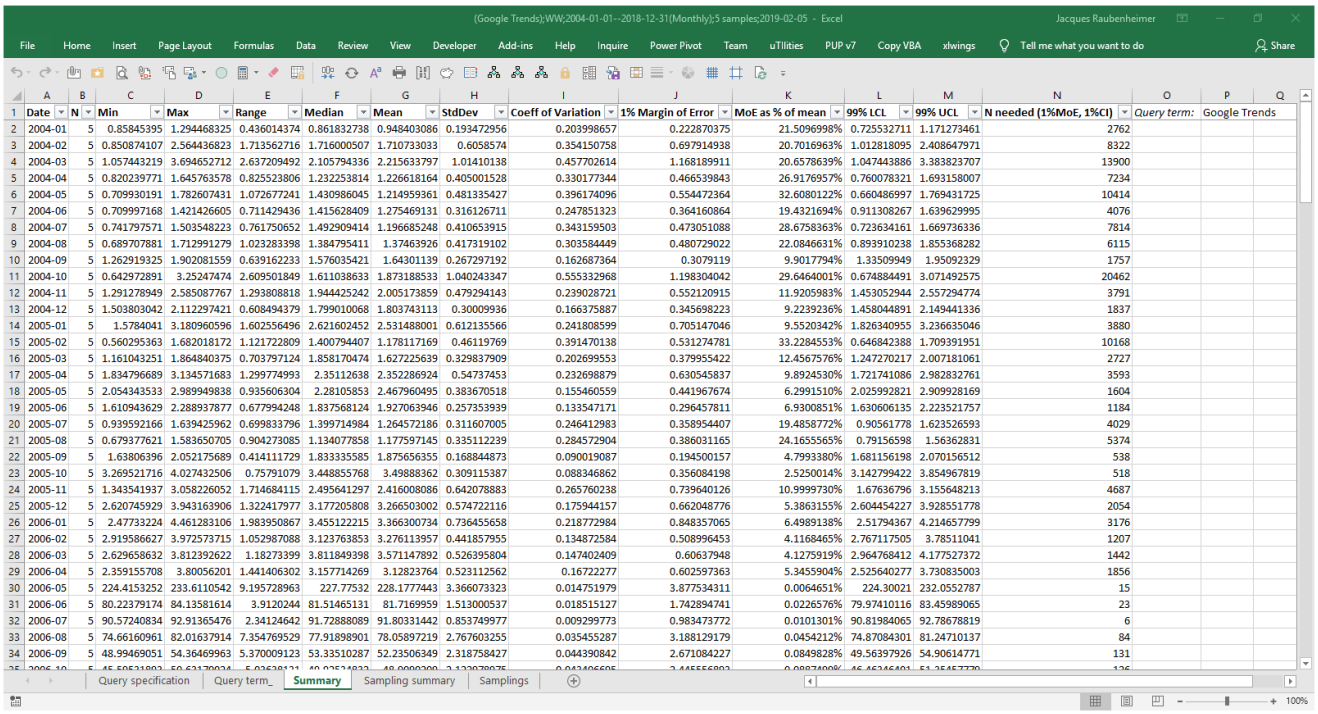
3.6.2 Summary sheets

The summary sheets (Figure 31) essentially serve to merge the information from the Samplings sheets. They contain the following self-explanatory columns:

Table 5 Summary sheet columns

Column	Information
A	Date
B	N
C	Min
D	Max
E	Range
F	Median
G	Mean
H	StdDev
I	Coeff of Variation
J	1% Margin of Error
K	MoE as % of mean
L	99% LCL
M	99% UCL
N	N needed (1%MoE, 1%CI) *[more about this in 3.6.3 below]*
O–P	Query term

In addition, the query term for which the sheet is summarising data is listed in cell P1. This is useful when multiple Summary sheets exists.

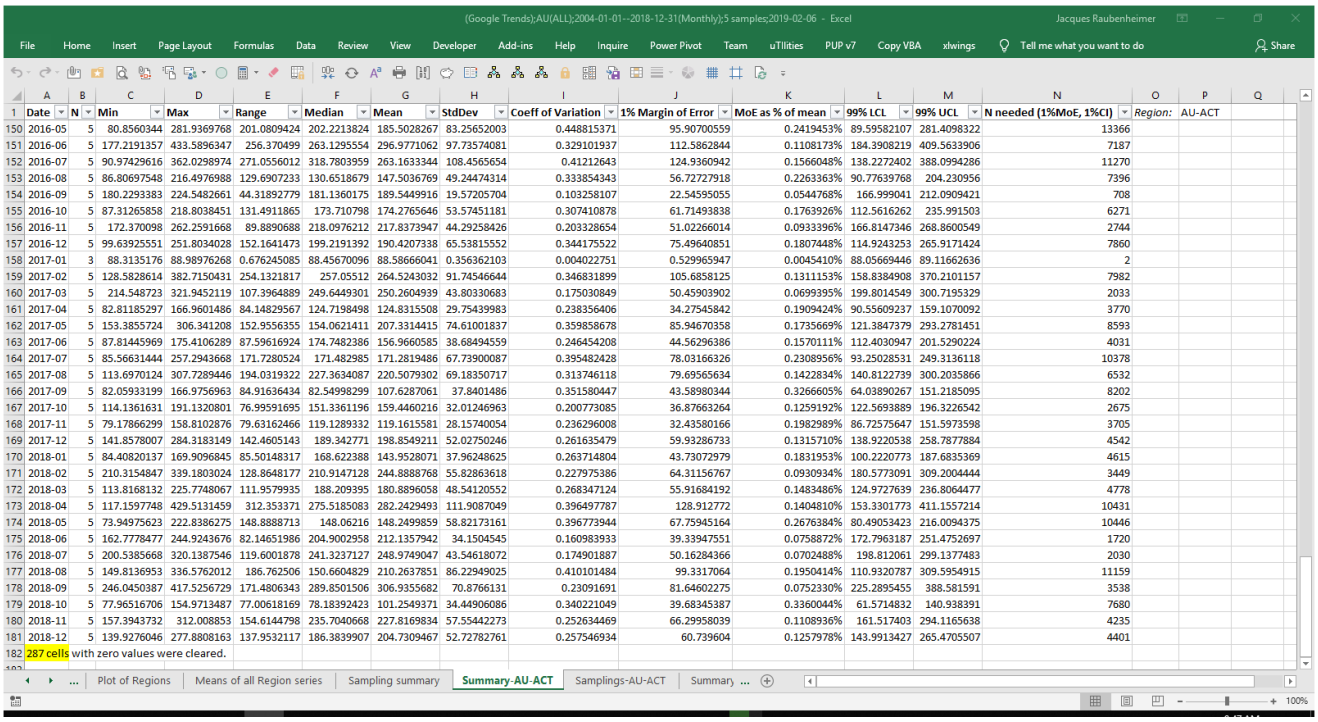


Date	N	Min	Max	Range	Median	Mean	StdDev	Coeff of Variation	1% Margin of Error	MoE as % of mean	99% LCL	99% UCL	N needed (1%MoE, 1%CI)	Query term:	Google Trends
2004-01	5	0.85845395	1.294468325	0.436014374	0.861832738	0.948403086	0.193472956	0.203998657	0.222870375	21.5096998%	0.725532711	1.171273461	2762		
2004-02	5	0.850874107	2.564436823	1.713562716	1.716000507	1.710733033	0.6058574	0.354150758	0.697914938	20.7016963%	1.012818095	2.408647971	8322		
2004-03	5	1.057443219	3.694652712	2.637209492	2.105794336	2.215633797	1.01410138	0.457702614	1.168189911	20.6578639%	1.047443886	3.383823707	13900		
2004-04	5	0.820239771	1.645763578	0.825523806	1.232253814	1.226618164	0.405001528	0.330177344	0.466539843	26.9176957%	0.760078321	1.693158007	7234		
2004-05	5	0.709930191	1.782607431	1.072677241	1.430986045	1.214959361	0.481335427	0.396174096	0.554472364	32.6080122%	0.660486997	1.769431725	10414		
2004-06	5	0.709997168	1.421426605	0.711429436	1.415628409	1.275469131	0.316126711	0.247851323	0.364160864	19.4321694%	0.911308267	1.639629995	4076		
2004-07	5	0.741797571	1.503548223	0.761750652	1.492909414	1.196685248	0.410653915	0.343159503	0.473051088	28.6758363%	0.723634161	1.669736336	7814		
2004-08	5	0.689707881	1.712991279	1.023283398	1.384795411	1.37463926	0.417319102	0.303584449	0.480729022	22.0846631%	0.893910238	1.855368282	6115		
2004-09	5	1.262919325	1.902081559	0.639162233	1.576035421	1.64301139	0.267297192	0.162687364	0.3079119	9.9017794%	1.33509949	1.95092329	1757		
2004-10	5	0.642972891	3.25247474	2.609501849	1.611038633	1.873188533	1.040243347	0.555332968	1.198304042	29.6464001%	0.674884491	3.071492575	20462		
2004-11	5	1.291278949	2.585087767	1.293808818	1.944425242	2.005173859	0.479294143	0.239028721	0.552120915	11.9205983%	1.453052944	2.557294774	3791		
2004-12	5	1.503803042	2.112297421	0.608494379	1.799010068	1.803743113	0.30009936	0.166375887	0.345698223	9.2239236%	1.458044891	2.149441336	1837		
2005-01	5	1.5784041	3.180960596	1.602556496	2.621602452	2.531488001	0.612135566	0.241808599	0.705147046	9.5520342%	1.826340955	3.236635046	3880		
2005-02	5	0.560295363	1.682018172	1.121722809	1.400794407	1.178117169	0.46119769	0.391470138	0.531274781	33.2284553%	0.646842388	1.709391951	10168		
2005-03	5	1.161043251	1.864840375	0.703797124	1.858170474	1.627225639	0.329837909	0.202699553	0.379955422	12.4567576%	1.247270217	2.007181061	2727		
2005-04	5	1.834796689	3.134571683	1.299774993	2.35112638	2.352286924	0.54737453	0.232698879	0.630545837	9.8924530%	1.721741086	2.982832761	3593		
2005-05	5	2.054343533	2.989949838	0.935606304	2.28105853	2.467960495	0.383670518	0.155460559	0.441967674	6.2991510%	2.025992821	2.909928169	1604		
2005-06	5	1.610943629	2.288937877	0.677994248	1.837568124	1.927063946	0.257353939	0.133547171	0.296457811	6.9300851%	1.630606135	2.223521757	1184		
2005-07	5	0.939592166	1.639425962	0.699833796	1.399714984	1.264572186	0.311607005	0.246412983	0.358954407	19.4858772%	0.90561778	1.623526593	4029		
2005-08	5	0.679377621	1.583650705	0.904273085	1.134077858	1.177597145	0.335112239	0.284572904	0.386031165	24.1655565%	0.79156598	1.56362831	5374		
2005-09	5	1.63806396	2.052175689	0.414111729	1.833335585	1.875656355	0.168844873	0.090019087	0.194500157	4.7993380%	1.681156198	2.070156512	538		
2005-10	5	3.269521716	4.027432506	0.75791079	3.448855768	3.49888362	0.309115387	0.088346862	0.356084198	2.5250014%	3.142799422	3.854967819	518		
2005-11	5	1.343541937	3.058226052	1.714684115	2.495641297	2.416008086	0.642078883	0.265760238	0.739640126	10.9999730%	1.67636796	3.155648213	4687		
2005-12	5	2.620745929	3.943163906	1.322417977	3.177205808	3.266503002	0.574722116	0.175944157	0.662048776	5.3863155%	2.604454227	3.928551778	2054		
2006-01	5	2.47733224	4.461283106	1.983950867	3.455122215	3.366300734	0.736455658	0.218772984	0.848357065	6.4989138%	2.51794367	4.214657799	3176		
2006-02	5	2.919586627	3.972573715	1.052987088	3.123763853	3.276113957	0.441857955	0.134872584	0.508996453	4.1168465%	2.767117505	3.78511041	1207		
2006-03	5	2.629658632	3.812392622	1.18273399	3.811849398	3.571147892	0.526395804	0.147402409	0.60637948	4.1275919%	2.964768412	4.177527372	1442		
2006-04	5	2.359155708	3.80056201	1.441406302	3.157714269	3.12823764	0.523112562	0.16722277	0.602597363	5.3455904%	2.525640277	3.730835003	1856		
2006-05	5	224.4153252	233.6110542	9.195728963	227.77532	228.1777443	3.366073323	0.014751979	3.877534311	0.0064651%	224.30021	232.0552787	15		
2006-06	5	80.22379174	84.13581614	3.9120244	81.51465131	81.7169959	1.513000537	0.018515127	1.742894741	0.0226576%	79.97410116	83.45989065	23		
2006-07	5	90.57240834	92.91365476	2.34124642	91.72888089	91.80331442	0.853749977	0.009299773	0.983473772	0.0101301%	90.81984065	92.78678819	6		
2006-08	5	74.66160961	82.01637914	7.354769529	77.91898901	78.05897219	2.767603255	0.035455287	3.188129179	0.0454212%	74.87084301	81.24710137	84		
2006-09	5	48.99469051	54.36469963	5.370009123	53.33510287	52.23506349	2.318758427	0.044390842	2.671084227	0.0849828%	49.56397926	54.90614771	131		

Figure 31 Summary sheet for a single-term, single region request

Note that the N column (column B of the worksheet) will reflect the number of samples effectively drawn across all the multiple samplings, for each date interval. If no zero values (cf. section 1.12—Data scaling and zero values, p. 4) are returned across all samplings for a particular date interval, then N will be equal to the number of samples requested in the Samples input (cf. section 3.2 above).

Also, if zero values were found and removed, then a note to that effect is made in the row directly beneath the data, in column A (i.e., directly below the last date value). This can be seen in Figure 32 (note the rows where N<20).



Date	N	Min	Max	Range	Median	Mean	StdDev	Coeff of Variation	1% Margin of Error	MoE as % of mean	99% LCL	99% UCL	N needed (1%MoE, 1%CI)	Region:	AU-ACT
2016-05	5	80.8560344	281.9369768	201.0809424	202.2213824	185.5028267	83.25652003	0.448815371	95.90700559	0.2419453%	89.59582107	281.4098322	13366		
2016-06	5	177.2191357	433.5896347	256.370499	263.1295554	296.9771062	97.73574081	0.329101937	112.5862844	0.1108173%	184.3908219	409.5633906	7187		
2016-07	5	90.97429616	362.0298974	271.0556012	318.7803959	263.1633344	108.4565654	0.41212643	124.9360942	0.1566048%	138.2272402	388.0994286	11270		
2016-08	5	86.80697548	216.4976988	129.6907233	130.6518679	147.5036769	49.24474314	0.333854343	56.72727918	0.2263363%	90.77639768	204.230956	7396		
2016-09	5	180.2293383	224.5482661	44.31892779	181.1360175	189.5449916	19.57205704	0.103258107	22.54595055	0.0544768%	166.999041	212.0909421	708		
2016-10	5	87.31265858	218.8038451	131.4911865	173.710798	174.2765646	53.57451181	0.307410878	61.71493838	0.1763926%	112.5616262	235.991503	6271		
2016-11	5	172.370098	262.2591668	89.8890688	218.0976212	217.8373947	44.29258426	0.203328654	51.02266014	0.0933396%	166.8147346	268.8600549	2744		
2016-12	5	99.63925551	251.8034028	152.1641473	199.2191392	190.4207338	65.53815552	0.344175522	75.49640851	0.1807448%	114.9243253	265.9171424	7860		
2017-01	3	88.3135176	88.98976268	0.676245085	88.45670096	88.58666041	0.356362103	0.004022751	0.529965947	0.0045410%	88.05669446	89.11662636	2		
2017-02	5	128.5828614	382.7150431	254.1321817	257.05512	264.5243032	91.74546644	0.346831899	105.6858125	0.1311153%	158.8384908	370.2101157	7982		
2017-03	5	214.548723	321.9452119	107.3964889	249.6449301	250.2604939	43.80330683	0.175030849	50.45903902	0.0699395%	199.8014549	300.7195329	2033		
2017-04	5	82.81185297	166.9601486	84.14829567	124.7198498	124.8315508	29.75439983	0.238356406	34.27545842	0.1909424%	90.55609237	159.1070092	3770		
2017-05	5	153.3855724	306.341208	152.9556355	154.0621411	207.3314415	74.61001837	0.359858678	85.94670358	0.1735669%	121.3847379	293.2781451	8593		
2017-06	5	87.81445969	175.4106289	87.59616924	174.7482386	156.9660585	38.68494559	0.246454208	44.56296386	0.1570111%	112.4030947	201.5290224	4031		
2017-07	5	85.56631444	257.2943668	171.7280524	171.482985	171.2819486	67.73900087	0.395482428	78.03166326	0.2308956%	93.25028531	249.3136118	10378		
2017-08	5	113.6970124	307.7289446	194.0319322	227.3634087	220.5079302	69.18350717	0.313746118	79.69565634	0.1422834%	140.8122739	300.2035866	6532		
2017-09	5	82.05933199	166.9756963	84.91636434	82.54998299	107.6287061	37.8401486	0.351580447	43.58980344	0.3266605%	64.03890267	151.2185095	8202		
2017-10	5	114.1361631	191.1320801	76.99591695	151.3361196	159.4460216	32.01246963	0.200773085	36.87663264	0.1259192%	122.5693889	196.3226542	2675		
2017-11	5	79.17866299	158.8102876	79.63162466	119.1289332	119.1615581	28.15740054	0.236296008	32.43580166	0.1982989%	86.72575647	151.5973598	3705		
2017-12	5	141.8578007	284.3183149	142.4605143	189.342771	198.8549211	52.02750246	0.261635479	59.93286733	0.1315710%	138.9220538	258.7877884	4542		
2018-01	5	84.40820137	169.9096845	85.50148317	168.622388	143.9528071	37.96248625	0.263714804	43.73072979	0.1831953%	100.2220773	187.6835369	4615		
2018-02	5	210.3154847	339.1803024	128.8648177	210.9147128	244.8888768	55.82863618	0.227975386	64.31156767	0.0930934%	180.5773091	309.2004444	3449		
2018-03	5	113.8168132	225.7748067	111.9579935	188.209395	180.8896058	48.54120552	0.268347124	55.91684192	0.1483486%	124.9727639	236.8064477	4778		
2018-04	5	117.1597748	429.5131459	312.353371	275.5185083	282.2429493	111.9087049	0.396497787	128.912772	0.1404810%	153.3301773	411.1557214	10431		
2018-05	5	73.94975623	222.8386275	148.8888713	148.06216	148.2499859	58.82173161	0.396773944	67.75945164	0.2676384%	80.49053423	216.0094375	10446		
2018-06	5	162.7778477	244.9243676	82.14651986	204.9002958	212.1357942	34.1504545	0.160983933	39.33947551	0.0758872%	172.7963187	251.4752697	1720		
2018-07	5	200.5385668	320.1387546	119.6008178	241.3237127	248.9749047	43.54618072	0.174901887	50.16284366	0.0702488%	198.812061	299.1377483	2030		
2018-08	5	149.8136953	336.5762012	186.762506	150.6604829	210.2637851	86.22949025	0.410101484	99.3317064	0.1950414%	110.9320787	309.5954915	11159		
2018-09	5	246.0450387	417.5256729	171.4806343	289.8501506	306.9355682	70.8766131	0.23091691	81.64602275	0.0752330%	225.2895455	388.581591	3538		
2018-10	5	77.96516706	154.9713487	77.00618169	78.1839243	101.2549371	34.44906086	0.340221049	39.68345387	0.3360044%	61.5714832	140.938391	7680		
2018-11	5	157.3943732	312.008853	154.6144798	235.7040668	227.8169834	57.55442273	0.252634469	66.29958039	0.1108936%	161.517403	294.1165638	4235		
2018-12	5	139.9276046	277.8808163	137.9532117	186.3839907	204.7309467	52.72782761	0.257546934	60.739604	0.1257978%	143.9913427	265.4705507	4401		

287 cells with zero values were cleared.

Figure 32 Summary sheet for one region of a multi-region request

3.6.3 Sampling Summary

As of Version 2, one additional worksheet is added which contains a summary of the CV and N needed columns from the summary sheets, as well as, at the top, the actual number of samples these estimates are based on. The Summary sheet for each query term expresses the coefficient of variation for each time point, as well as an estimation of the number of samples needed to accurately estimate the true value for that time point, assuming a two-sided test for $\alpha=0.01$ and a 1% margin of error. Of course, each time point is a sample from Google's data in and of itself, and thus the N needed for each time point will be different. In certain instances, the N needed will vary quite considerably, and could run into the thousands. Researchers will have to decide what number of samples is practical, and what proportion of time points in the series will be adequately estimated using that number of samples. This problem is further complicated when multiple query terms are used, as each time point in each query term's time series will have its own variability.

The information on this worksheet is intended to help researchers make this choice. Two columns are created for each query term (Figure 33).

	A	B	C	D	E	F
1		Google Trends		Google AdSense		
2	*Samples used for estimations*					
3	*Time periods*	180		180		
4	*N requested*	100		100		
5	*Min N achieved*	95		100		
6	*Max N achieved*	100		100		
7	*Median N achieved*	100		100		
8	*Mean N achieved*	99.9		100.0		
9	*Summary of Coeff of Var*	*CV*	*N*	*CV*	*N*	
10	*Min*	*.008*	1	*.007*	1	
11		*.055*	151	*.017*	129	
12		*.102*	0	*.027*	19	
13		*.150*	1	*.037*	12	
14		*.197*	4	*.047*	3	
15		*.244*	6	*.058*	2	
16		*.292*	5	*.068*	2	
17		*.339*	4	*.078*	4	
18		*.386*	3	*.088*	2	
19	*Max*	*.434*	5	*.099*	6	
20	*Median*	*.014*	*90*	*.011*	*90*	
21	*Mean*	*.057*	*152*	*.019*	*139*	
22	*Modes of N-needed*	*Mode*	*N*	*Mode*	*N*	
23		9	19	6	33	
24	*Summary of N-needed Percentiles*	*Percentile*	*N needed*	*Percentile*	*N needed*	
25		0%	4	0%	3	
26		1%	5	1%	3.79	
27		5%	6	5%	5	
28		25%	8	25%	6	
29		50%	14	50%	8	
30		75%	42	75%	20	
31		95%	7362	95%	345.8	
32		99%	11808.76	99%	613.52	
33		100%	12469	100%	645	
34	*Calculate Percentile for N-Needed*	**46.30%**	14	**67.00%**	14	

... Sampling summary | Summary-Google Trends ...

Figure 33 Sampling Summary sheet

3.6.3.1 Summary of coefficient of variation

The first ten data rows on the Sampling Summary sheet show the full range of Coefficient of Variation (CV) values for each all the time points in the time series of each query term. The minimum, maximum, and eight additional evenly interspersed values are calculated. These can be seen as histogram bins. Additionally, the Excel FREQUENCY function is used to reflect the number of time points within each bin.

Two more rows are provided, showing the mean and median CV values, as well as the number of time points less than or equal to these respective values.

In both these sets (i.e., the top ten rows, and the next two rows), the total should be equal to or less than the number of periods in the time specification (less when some time points have no data).

3.6.3.2 Modes of N needed

Each summary sheet also contains a column showing the number of samples needed to estimate each time point in that series (i.e., the time series for that query term). The mode of the number of samples needed, as well as the number of time points for which that mode applies, are shown.

3.6.3.3 Summary of N-needed

Next, the N needed for each query term is summarised in terms of its percentiles, showing, somewhat arbitrarily, the following percentiles: 0% (i.e., minimum); 1%; 5%; 25%; 50% (median); 75%; 95%; 99%; 100% (i.e., maximum). The actual N needed for each percentile is also shown. This gives researchers a quick indication of what percentage of their time points will be adequately estimated using a particular number of samples, across all query terms.

3.6.3.4 Calculate Percentile for N-Needed

Researchers may, however, wish a more detailed estimate of the number of time points estimated adequately for a given number of samples (i.e., the reverse calculation to the above). This is provided in the last row of the Sampling summary sheet. The cells are formatted with the Input style, and are initially set such that the value for the first query term is the median of the N needed values, and all following query terms have the same value, provided that it is not more than the maximum N needed for that query term. Researchers can change the N needed for the first query term as they wish, and it should update across all other query terms, although the N needed for each query term can be set individually.

3.6.4 Master Summary sheet (Means of all … series)

This sheet will only be created when either a multi-term request (cf. section 3.3—Multi-term requests) or a multi-region request (cf. section 3.4—Multi-region requests) is made. It will be titled “Means of all […] series” with the term in brackets being either ‘Region’ or ‘Query Terms,’ as required. Unlike the aforementioned Summary sheets, this sheet contains only the date range in the first column, and then the Mean column (column G) from every individual Summary sheet (for every term or every region) in order. The number of columns in this sheet will thus be one more than the number of query terms (for a multi-term request) or the number of regions (for a multi-region request). An example is displayed in Figure 34. Note that the column with missing values is not an error—all data returned for this region were zeros, and were deleted accordingly, thus not allowing a mean to be calculated.

	A	B	C	D	E	F	G	H	I	J
1	Date	AU-ACT	AU-NSW	AU-NT	AU-QLD	AU-SA	AU-TAS	AU-VIC	AU-WA	
154	2016-09	227.2594	181.96		101.8595	47.81736	159.7783	136.5635	97.25126	
155	2016-10	200.5617	182.4645		128.7353	77.96323	119.2014	122.2351	65.095	
156	2016-11	185.9521	210.8278		139.6649	175.2469	236.4733	165.0219	138.5926	
157	2016-12	187.1525	141.8861		136.0805	82.55362	203.2281	142.4034	73.84798	
158	2017-01	106.6324	148.2093		118.4584	64.05458	135.659	150.1443	109.2961	
159	2017-02	311.3221	197.2299		131.6858	89.32641	255.3199	149.1769	103.428	
160	2017-03	224.5447	159.5256		146.686	145.7136	105.4202	189.2968	136.5408	
161	2017-04	121.7603	184.9429		115.3517	94.02389	167.0273	142.337	100.8442	
162	2017-05	182.8912	193.8566		119.9354	82.40253	206.2996	179.325	108.5393	
163	2017-06	145.2323	179.5969		117.0971	137.3264	109.7081	139.9624	93.73429	
164	2017-07	228.7758	176.3941		96.07937	83.35242	216.9546	157.6578	109.9141	
165	2017-08	176.4526	202.6767		126.9669	127.2213	272.7901	160.2329	165.0361	
166	2017-09	93.69285	173.0057		113.5453	103.1839	184.6485	142.4625	74.49327	
167	2017-10	178.4355	174.9379		123.245	65.35724	102.5806	168.7578	86.41335	
168	2017-11	139.0115	191.7467		153.701	82.14257	97.52061	174.7371	132.9895	
169	2017-12	140.821	139.6856		119.9846	123.1129	215.2279	159.0621	93.30912	
170	2018-01	192.5512	183.5724		148.856	109.8161	182.9791	172.1769	116.2462	
171	2018-02	157.7681	181.9663		94.81364	149.2146	126.3791	151.178	117.2006	
172	2018-03	206.2757	177.6423		142.1975	105.3699	118.8299	151.0115	88.2108	
173	2018-04	213.6349	184.0519		138.8434	107.0939	111.3637	162.0658	120.6477	
174	2018-05	192.5487	185.4796		158.6466	109.4029	208.5925	163.4376	140.0616	
175	2018-06	180.2219	186.4158		98.64463	114.13	108.5404	161.4599	94.2359	
176	2018-07	217.5142	153.7278		109.6337	81.77799	122.8706	143.7731	82.49849	
177	2018-08	266.5588	151.7089		131.5687	99.37525	105.334	146.5613	111.8696	
178	2018-09	251.5852	158.123		145.2175	84.71665	152.0801	121.5416	72.9911	
179	2018-10	140.0352	146.0791		101.3389	98.54242	186.1239	145.2755	86.22164	
180										

Plot of Regions | Query specification | **Means of all Region series** | Summary-AU-ACT

Figure 34 Master summary sheet for a multi-region request

3.6.5 Chart

A Google Trends Extended Health extraction creates a line chart which plots either the values in the Mean column from the Summary sheet for a single term, single region request (Figure 35), or all the Mean columns from the Master Summary sheet for multi-term or multi-region requests, over time (cf. Figure 29, p. 26). The aim is to have a chart which is as close to presentation-ready as possible, although a little tweaking may be required from the user.

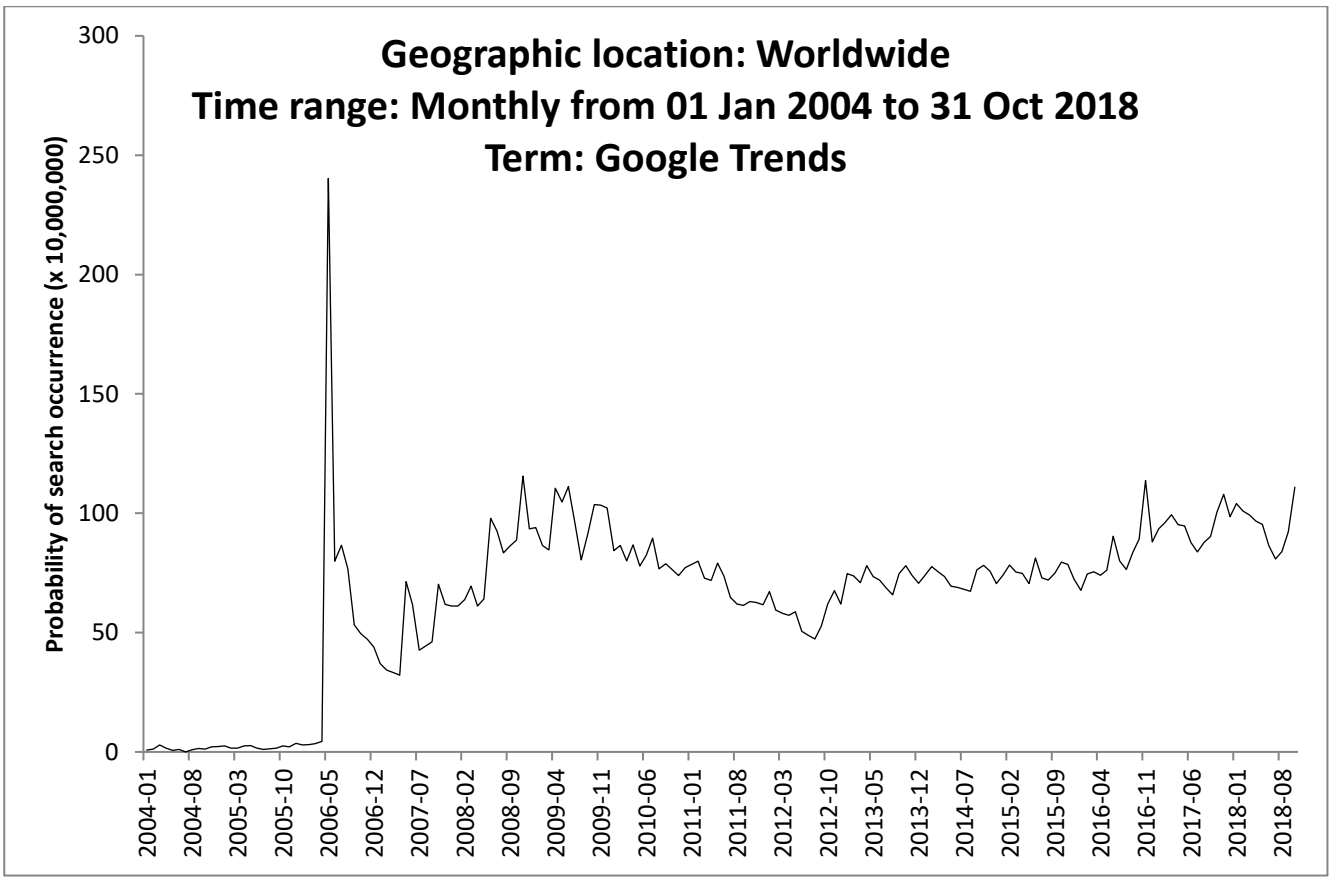


Figure 35 Line chart for Google Trends Extended Health data

Chapter 4 Google Trends Web

Google Trends Web data, despite their limitations, do offer some advantages over Google Trends Extended Health data. The biggest rea of advantage lies in the availability of real time data, down to a very fine time resolution. Whereas Google Trends Extended Health data are only available to two days before the current date, Google Trends Web data are available in real time. When downloading data for a very short time frame, data points are returned as frequently as every minute. The exact resolution used is dependent on the time frame selected, as illustrated in Table 6.

Table 6 Google Trends Web time resolutions (using default time settings available on Google Trends website)

Time scale	Results presented	Data source
Past hour	Per minute	Real time data
Past 4 hours	Per minute	
Past day	Every 8 minutes	
Past 7 days	Every hour	
Past 30 days	Every day	Archive
Past 90 days	Every day	
Past 12 months	Every week	
Past 5 years	Every week	
2004-Present	Every month	
Custom	Daily (1–269 days)	
	Weekly (270–1889 days)	
	Monthly (270 weeks–)	

4.1 Date specification

Because the date resolution is determined by the Google API for the Google Trends Web data, the date specification is much simpler than for the Google Trends Extended Health interface. Simply set a start date equal to, or later than, 1 January 2004, or leave the start date blank to default to the aforementioned date. For the end date, specify a date after the start date, or leave the end date blank to default to the present.

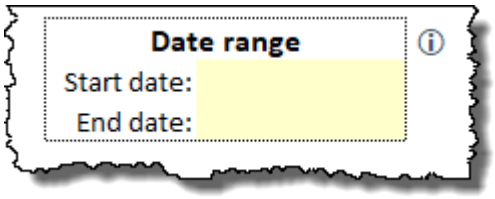


Figure 36 Google Trends Web Date range specification

Users can double click on either the start- or end date cells to open the calendar form and select a date. The use of the calendar form was summarised in section 3.1.1 Date specification via the calendar form.

4.2 Function

To understand the importance of the various Functions available from the Google Trends Web interface, it is useful to visit the Google API explorer at https://developers.google.com/apis-explorer/#p/trends/v1beta/. All of the API calls from Trends API v1beta (2018-11) are listed in Table 7. As is evident from this table, the entire Google Trends Extended for Health interface is devoted to the one API call trends.getTimelinesForHealth. The remaining seven API calls are all accessed via the Google Trends Web interface, and then by selecting the corresponding function from the Function dropdown (Figure 37).

Table 7 Google Trends Web interface Functions Plotted Against List of API calls from https://developers.google.com/apis-explorer/#p/trends/v1beta/

Function	API call	Explanation
Graph	trends.getGraph	Returns a Graph of search volume per time points, normalized. For better insights, one could provide restrictions for time range, geographic region, etc.
Graph Averages	trends.getGraphAverages	Returns the averages of normalized search volume for the given terms. For better insights, one could provide restrictions for time range, geographic region, etc.
Rising Queries	trends.getRisingQueries	Get a list of rising queries that were searched along with the requested term, under the given restrictions.
Rising Topics	trends.getRisingTopics	Get a list of rising topics that were searched along with the requested term, under the given restrictions.
Google Trends Extended for Health interface	trends.getTimelinesForHealth	For health research only, fetches a graph of search volumes per time within a set of restrictions. Note the data is sampled and Google can't guarantee the accuracy of the numbers.
Top Queries	trends.getTopQueries	Get a list of top queries that were searched along with the requested term, under the given restrictions.
Top Topics	trends.getTopTopics	Get a list of top topics that were searched along with the requested term, under the given restrictions.
List of Region Values	trends.regions.list	This would be the data behind the map seen in Regional Interest in http://www.google.com/trends/explore

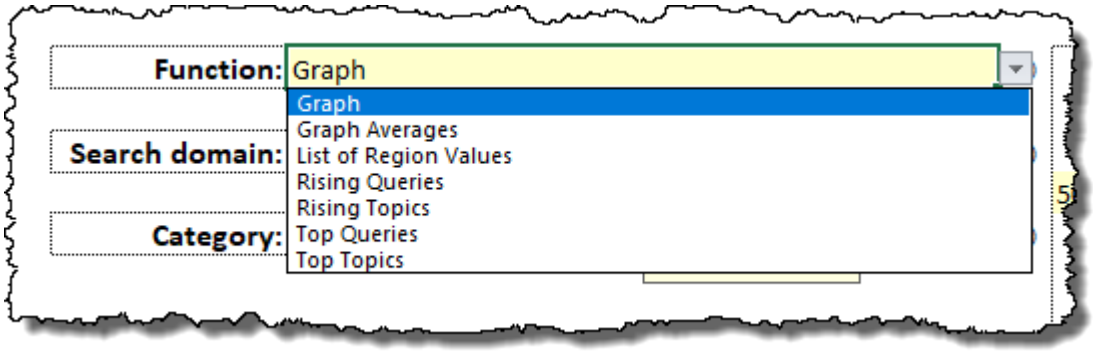


Figure 37 Trends API calls available via the Function selection

Apart from the explanations provided by Google on the API explorer page (https://developers.google.com/apis-explorer/#p/trends/v1beta/), and listed in Table 7, it helps to understand what these functions (i.e., each different API call) do by relating them to the Google Trends website (https://trends.google.com) itself (Figure 38).

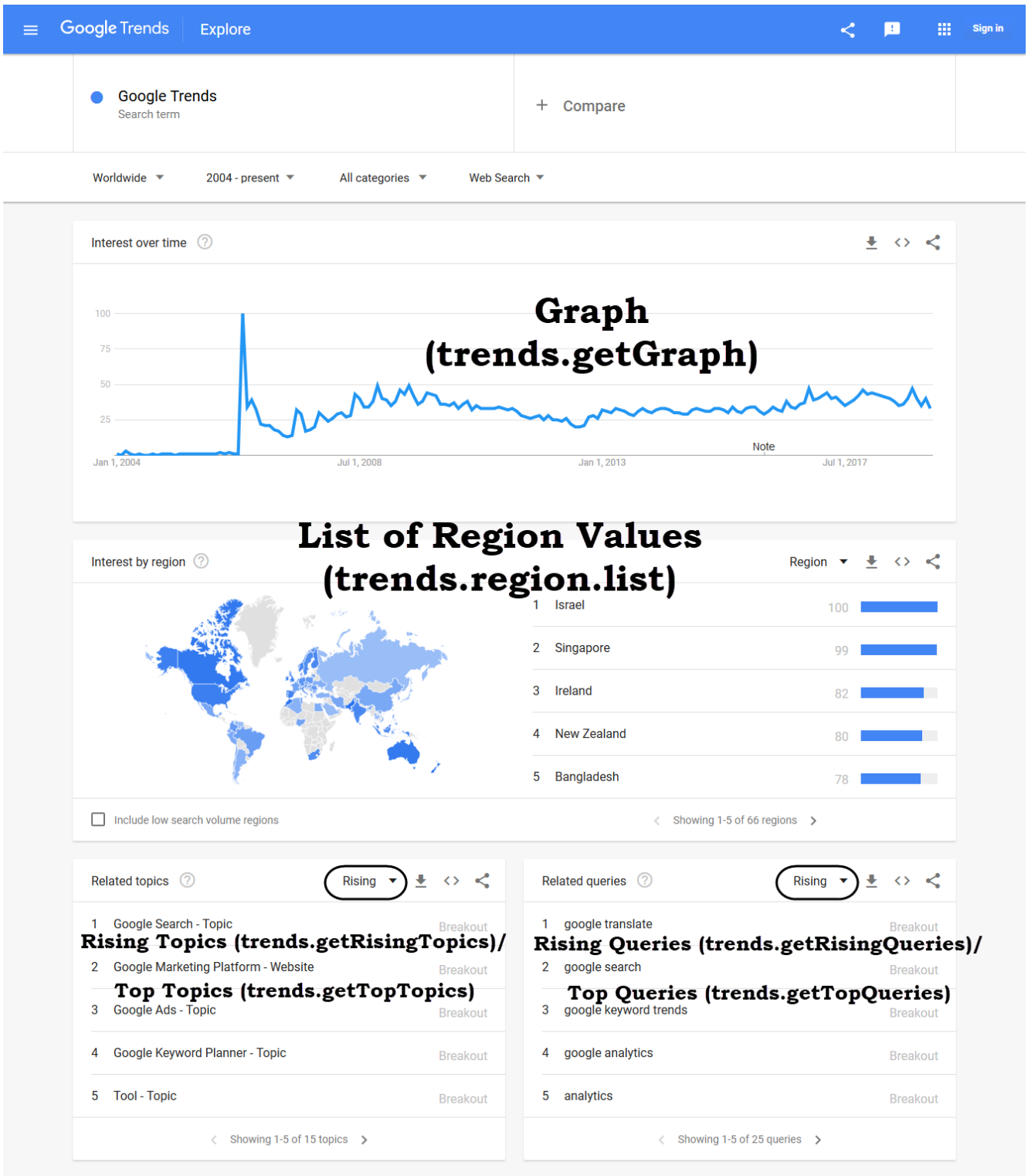


Figure 38 Functions/API calls related to Google Trends website content

As can be seen, each API call lies behind the data for one part of the results presented on the website, with these slight exceptions: First, both Rising And Top Topics (and also Queries) are presented as a unit on the website, and one switches between Rising/Top with the highlighted selectors. When the data are downloaded from the website, the download contains both the rising and the top values. Second, Graph Averages (trends.getGraphAverages) has no equivalent on the web page. It reveals what the averaged differences would be for various terms for a specific region. Thus, when using this function, more than one Query term should be used, and only one region can be specified.

4.3 Search domain

Google Trends Extended Health data return only Google searches, which is, nonetheless, most Google searching (although not verified by Google, as far as this developer could ascertain, that includes all searches from google.com or its various country-specific alternatives, cell phone searches, Google Home searches, and searches from other sources such as websites which employ the Google search widget, etc. Nonetheless, Google Trends Web data do allow users to specify either the Google search option, or to look at corresponding searches in Google image search, Google shopping search (a.k.a. Froogle), Google News, or YouTube. This can be specified in the Search domain selection (Figure 39).

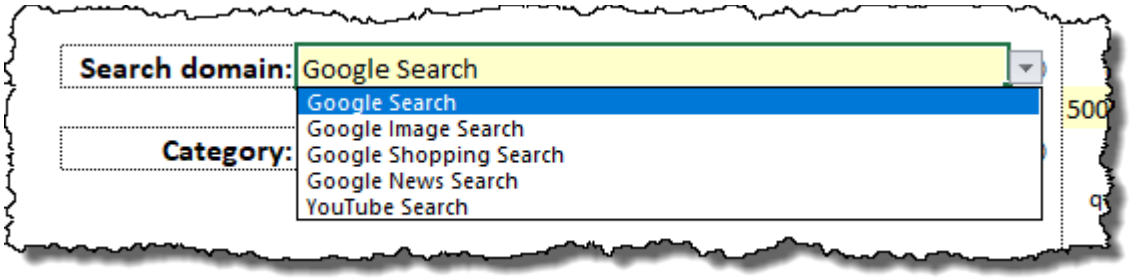


Figure 39 Search domain settings

4.4 Category

Google allows data extraction to be restricted to searches classified in certain domains only. The details of how Google classifies searches into these categories are not known‡, so the accuracy is dependent on the adequacy of their algorithms.

Generally, most Google Trends Web data extractions can be done from all categories, as restricting an extraction to a single category may be too restrictive, but in certain instances, the classification system provided by Google can be used to help eliminate ambiguity in the use of a term (e.g. restricting a query for 'Apple' to *Food & Drink: 71* when searching for interest in the fruit). This classification system is included in the data extraction tool. This developer used a list posted on GitHub by Trasborg[9], which at present provides 25 top level categories (26 if one adds the all-inclusive *All categories: 0*), 288 second level categories, 702 tertiary levels, 271 4th levels, 43 5th levels, and 4 6th levels, for a total of 1427 possibilities across all six levels (which also includes *All categories: 0*). This is somewhat different from Choi and Varian's rough estimate noted in footnote ‡ above.

To choose a category, one can either use the category selector, or one can browse the categories to get a better perspective of the whole system.

4.4.1 Category selector

To go to the category selector, simply double click on the category cell on the Google Trends Web interface.

‡ Choi and Varian[12] note that "Google classifies search queries into approximately 30 categories at the top level and approximately 250 categories at the second level using a natural language classification engine."

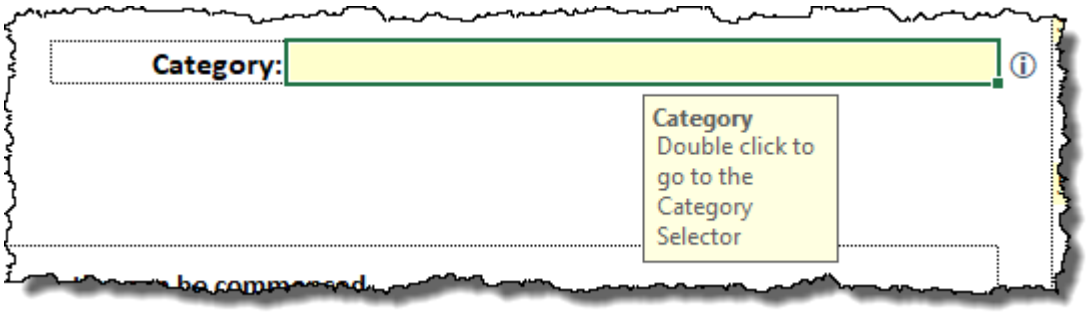


Figure 40 Category field (Google Trends Web)

The category selector has six levels, as this is the lowest level to which some categories are broken down, although not all categories contain that many levels. The tool attempts to intelligently guide the user by using conditional formatting to show whether additional lower levels may exist. Note, though, that this varies based on the selection at the current level. Thus, as an example, some third level selections will still be broken down into fourth levels, while others in the same category grouping will not, and so for various other levels.

In each instance, use the dropdown values to select the desired (sub-) category at the level you are working with. The selector will always show the lowest level selected as the category that is to be used. The example shown in Figure 41 shows both of the aforementioned aspects. 'Film & TV Industry' is the lowest category level selected (Level 4 being blank) and is thus the category that will be used in the Google Trends Web tool if the **Use this category** button is clicked. Also, the grey conditional formatting shows that there are no Levels 5 & 6 sublevels for 'Film & TV Industry,' while the lack of grey formatting shows that there are Level 4 possibilities for this category.

The buttons on this sheet are self-explanatory. With these buttons, you can **Clear** the category selections, **Cancel** the tool (i.e., return to the Google Trends Web interface without using the selected category), or **Use** the selected category in the Google Trends Web interface.

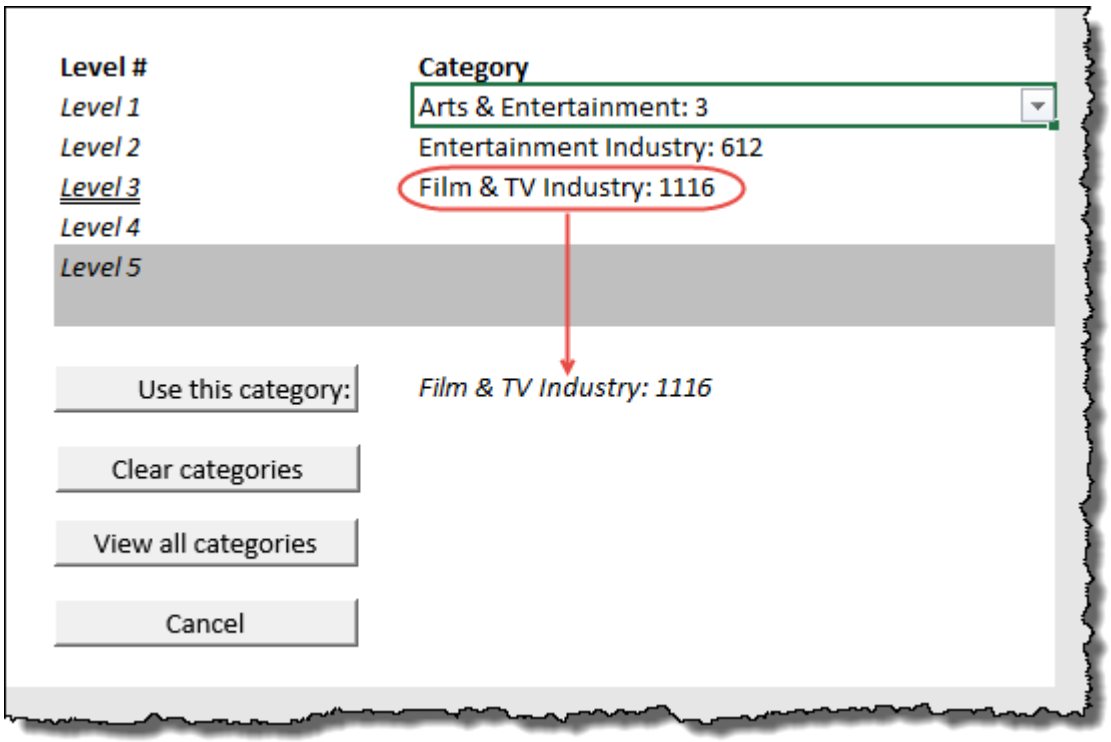


Figure 41 Category selector

4.4.2 Browsing categories

To browse the complete list of categories, click on the **View all categories** button from the category selector. This will take you to the Categories worksheet, from where you can navigate all the categories (including using the Excel filter to narrow down a selection). If you wish to use a category from this sheet, just double click on the cell containing that entry, and it will be loaded to the category selector tool. From there, click on the Use this category button to load it to the Google Trends Web interface.
If you wish to return to the category selector, just click on its sheet tab.

Level 1	Level 2	Level 3	Level 4	Level 5
All categories: 0				
Arts & Entertainment: 3				
Arts & Entertainment: 3	Celebrities & Entertainment News: 184			
Arts & Entertainment: 3	Comics & Animation: 316			
Arts & Entertainment: 3	Comics & Animation: 316	Animated Films: 1104		
Arts & Entertainment: 3	Comics & Animation: 316	Anime & Manga: 317		
Arts & Entertainment: 3	Comics & Animation: 316	Cartoons: 319		
Arts & Entertainment: 3	Comics & Animation: 316	Comics: 318		
Arts & Entertainment: 3	Entertainment Industry: 612			
Arts & Entertainment: 3	Entertainment Industry: 612	Film & TV Industry: 1116		
Arts & Entertainment: 3	Entertainment Industry: 612	Film & TV Industry: 1116	Film & TV Awards: 1108	
Arts & Entertainment: 3	Entertainment Industry: 612	Film & TV Industry: 1116	Film & TV Production: 1117	
Arts & Entertainment: 3	Entertainment Industry: 612	Recording Industry: 1115		
Arts & Entertainment: 3	Entertainment Industry: 612	Recording Industry: 1115	Music Awards: 1113	
Arts & Entertainment: 3	Entertainment Industry: 612	Recording Industry: 1115	Record Labels: 1114	
Arts & Entertainment: 3	Events & Listings: 569			
Arts & Entertainment: 3	Events & Listings: 569	Clubs & Nightlife: 188		
Arts & Entertainment: 3	Events & Listings: 569	Concerts & Music Festivals: 891		
Arts & Entertainment: 3	Events & Listings: 569	Film Festivals: 1086		
Arts & Entertainment: 3	Events & Listings: 569	Live Sporting Events: 1273		
Arts & Entertainment: 3	Events & Listings: 569	Movie Listings & Theater Showtimes: 1085		
Arts & Entertainment: 3	Events & Listings: 569	Ticket Sales: 614		
Arts & Entertainment: 3	Fun & Trivia: 539			
Arts & Entertainment: 3	Fun & Trivia: 539	Dress-Up & Fashion Games: 1173		
Arts & Entertainment: 3	Fun & Trivia: 539	Flash-Based Entertainment: 447		
Arts & Entertainment: 3	Fun & Trivia: 539	Fun Tests & Silly Surveys: 1174		
Arts & Entertainment: 3	Humor: 182			

Figure 42 Browsing categories

4.5 Output

The outputs created by the Google Trends Web interface vary depending on which function was used (cf. section 4.2—Function, p. 35). Each output is essentially the data returned by the API, formatted on a spreadsheet. Because a number of the functions do not contain charts, all there is to interpret are the data. Since the data need to be interpreted in the light of the date range for which the extraction was done, the date range is added to every output sheet created by the Google Trends Web tool. Sample outputs are shown for each Google Trends Web function (in the order listed in Table 6), in Figure 43–Figure 48 below.

4.5.1 Graph

The graph output (Figure 43) is the same as the main display (Interest over time) on the Google Trends website (https://trends.google.com—cf. Figure 38, p. 37). Since no multiple sampling strategy is employed when using the Google Trends Web interface (and no zero values removed—cf. section 1.12—Data scaling and zero values, p. 4), only one worksheet is created to store the data. This worksheet will have the date range specified in the first column, reflecting the date resolution returned by the Google Trends API for that date specification. Thereafter, between one and five columns of values—one for every Query term specified.

	A	B	C	D	E	F	G	H
1	Date	Google Trends	Google Correlate	Google Scholar	Google AdWords		Date range	
2	2010/01/01	7	0	39	15		Start date:	2010/01/01
3	2010/02/01	8	0	47	16		End date:	2017/12/31
4	2010/03/01	7	0	55	17			
5	2010/04/01	8	0	54	16			
6	2010/05/01	7	0	46	17			
7	2010/06/01	7	0	35	16			
8	2010/07/01	8	0	30	15			
9	2010/08/01	7	0	32	14			

Figure 43 Sample output: Graph (Four query terms—Google Trends, Google Correlate, Google Scholar, Google AdWords)

Additionally, a line chart is created (cf. Figure 44) plotting all of the specified query terms, which is similar to that created by the Google Trends Extended Health tool (compare Figure 44 with Figure 35), although the labels on this line chart are slightly different to those created by the Google Trends Extended Health interface to reflect the different nature of the data returned by the two different API calls.

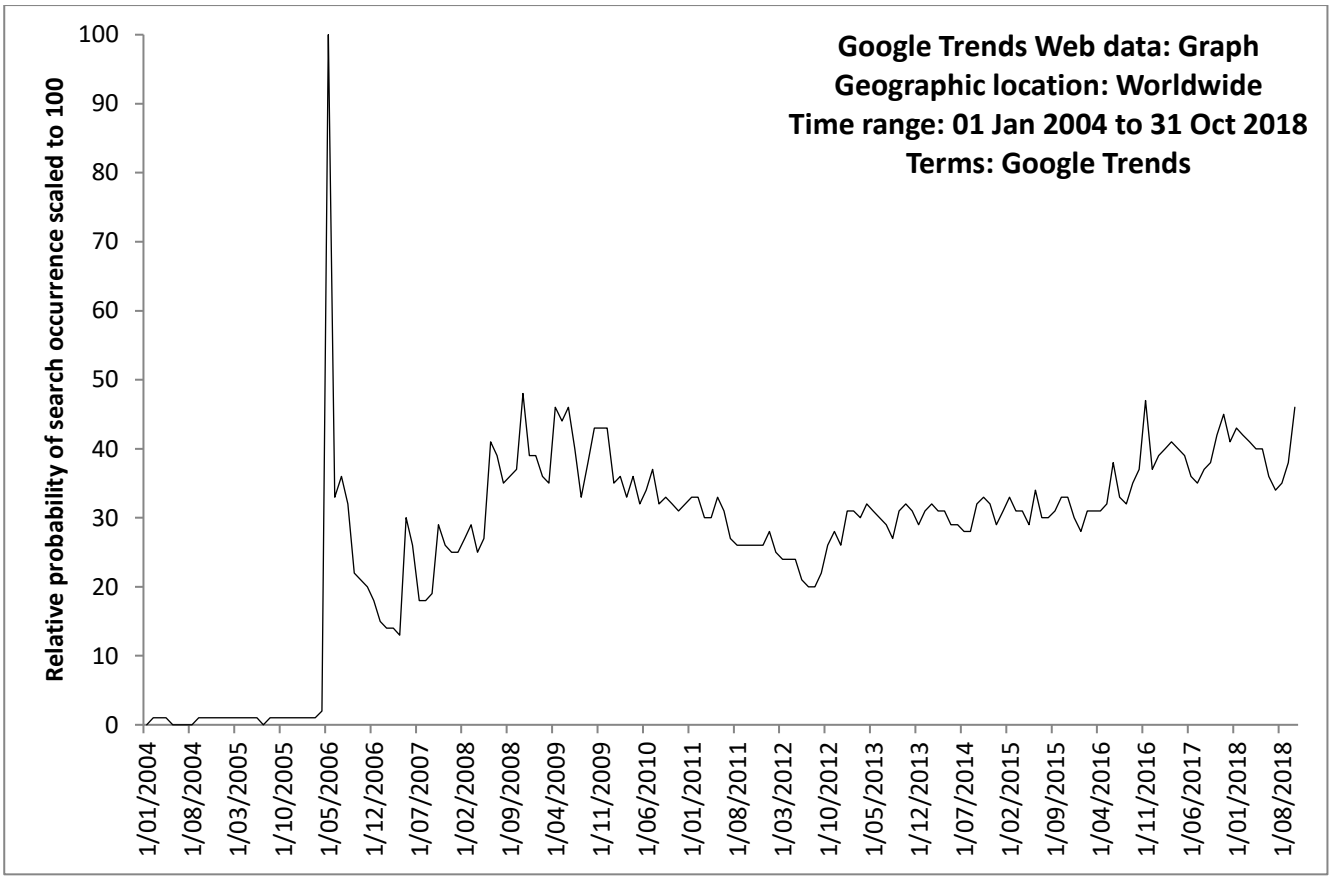


Figure 44 Line chart for Google Trends Web data

4.5.2 Graph Averages

The Graph Averages API call has no direct equivalent on the Google Trends website (https://trends.google.com). All that is returned is two columns of values, one containing the specified Query terms (Graph averages must compare at least two terms), and the second, their relative values (Figure 45).

	A	B	C	D	E
1	term	value		Date range	
2	Google Trends	7		Start date:	2010/01/01
3	Google Correlate	0		End date:	2017/12/31
4	Google Scholar	58			
5	Google AdWords	16			
6					
7					

Figure 45 Sample output: Graph Averages

4.5.3 Rising Queries and Top Queries

Rising Queries and Top Queries correspond to the Related Queries section on the Google Trends website (https://trends.google.com—cf. Figure 38, p. 37). These show other popular queries which are related to the Query term specified.

4.5.3.1 Top queries

Top queries will compare the various query terms with each other, with the most popular term again scaled to 100, and all other terms shown relative to that top term. This does not mean that the top term is wildly popular, but only that it is the most common of all the related terms, even though all of them together may be used very seldom.

4.5.3.2 Rising queries

Rising queries will show the same query terms as Top queries, with the values showing, not comparison to other terms, but percentage-wise comparison to searches in the period prior to the specified period (i.e., it will show growth in searches using each term over the specified period). Users should always remember that the comparison is with the specified period and the immediately preceding period of the same duration. As examples, if the specified period is 2018, these percentages will reflect growth from 2017 to 2018, or if the specified period is February 2018, then these percentages will reflect growth from January 2018 to February 2018. Again, this should be seen in context, since it is easy for a term that is used very little, to double in volume. Rising queries adds one further column, which is the isBreakout column (this information is also shown on the on the Google Trends website (https://trends.google .com—cf. Figure 38, p. 37). This is true when the searches for the relevant term have grown by more than 5000%[10] from the previous period to the current period. Bearing this in mind, it should be realised that it really makes no sense to extract Rising Queries or Rising Topics (see below) for the entire period of Google Trends data (i.e., from 2004), although Top Queries or Top Topics is sensible for that duration.

Top Queries

	A	B	C	D	E
1	title	value		Date range	
2	google translate	100		Start date:	2018/01/01
3	google analytics	42		End date:	2018/10/31
4	google search trends	41			
5	google scholar	27			
6					

Rising Queries

	A	B	C	D	E	F
1	title	value	isBreakout		Date range	
2	google translate	12.5	false		Start date:	2010/01/01
3	google search trends	0.8	false		End date:	2017/12/31
4	google analytics	0.4	false			
5						

Figure 46 Sample output: Top Queries and Rising Queries

4.5.4 Rising Topics and Top Topics

Rising Topics and Top Topics correspond to the Related Topics section on the Google Trends website (https://trends.google.com—cf. Figure 38, p. 37). This information is indicative of the Knowledge Graph entities Google's algorithms have associated with the query term specified (cf. section 2.2.5—Freebase IDs, p. 10), which means that these reports may be blank when Google's algorithms do not identify a related Knowledge Graph entity.

As with Rising/Top Queries, these API calls show other popular topics which are related to the topic with which the specified Query term is associated. In terms of what is displayed and how to interpret it, consult section 4.5.3 above, with the only difference that topics (i.e., Knowledge Graph entities) now replace actual terms.

One additional column is added to these results which is quite useful. It was mentioned in section 2.2.5—Freebase IDs, p. 10) that this tool does not incorporate the Google Knowledge Graph API to return Freebase IDs of terms, but these results (Rising Topics/Top Topics) contain an additional column ("mid") which shows the Freebase ID for each topic. This could, in certain cases, be a quicker way of finding a Freebase ID than searching for the specified term on the Wikidata site.

Top Topics

	A	B	C	D	E	F	G
1	title	mid	value		Date range		
2	Google	/m/045c7b	100		Start date:	2018/01/01	
3	Google Trends	/m/0d2p9p	96		End date:	2018/10/31	
4	Google Search	/m/0387r	9				
5	Google Ads	/m/03l68z	2				
6	Google Analytics	/m/08w9_f	2				
7							
8							

Rising Topics

	A	B	C	D	E	F	G	H
1	title	mid	value	isBreakout		Date range		
2	Google Translate	/m/025sndk	900	false		Start date:	2010/01/01	
3	Google Ads	/m/03l68z	100	false		End date:	2017/12/31	
4								

Figure 47 Sample output: Top Topics and Rising Topics

4.5.5 List of Region Values

The List of Region Values data (Figure 48) contains the same underlying information as that shown in the Interest by region section on the Google Trends website (https://trends.google.com—cf. Figure 38, p. 37).

	A	B	C	D	E	F
1	regionCode	regionName	value		Date range	
2	IL	Israel	100		Start date:	2010/01/01
3	SG	Singapore	99		End date:	2017/12/31
4	IS	Iceland	89			
5	CU	Cuba	89			
6	BD	Bangladesh	84			
7	MT	Malta	83			
8	PG	Papua New Guinea	80			
9	IE	Ireland	79			
10	PK	Pakistan	76			
11	NZ	New Zealand	75			
12	CA	Canada	74			
13	MA	Morocco	73			
14	AU	Australia	73			
15	MY	Malaysia	72			
16	XK	Kosovo	70			
17	US	United States	69			
18	NL	Netherlands	68			
19	SH	St. Helena	68			
20	GB	United Kingdom	67			
21	ES	Spain	66			
22	EE	Estonia	65			
23	IN	India	62			
24	ID	Indonesia	59			
25	BE	Belgium	59			
26	MK	Macedonia (FYROM)	59			
27	NO	Norway	59			
28	FI	Finland	59			
29	CH	Switzerland	58			
30	SE	Sweden	58			
31	PH	Philippines	57			

Query specification | Plot of Region values | Google Trends Web data

Figure 48 Sample output: List of Region Values

The List of Region Values function produces a bar chart showing the relative interest for each region (Figure 49).

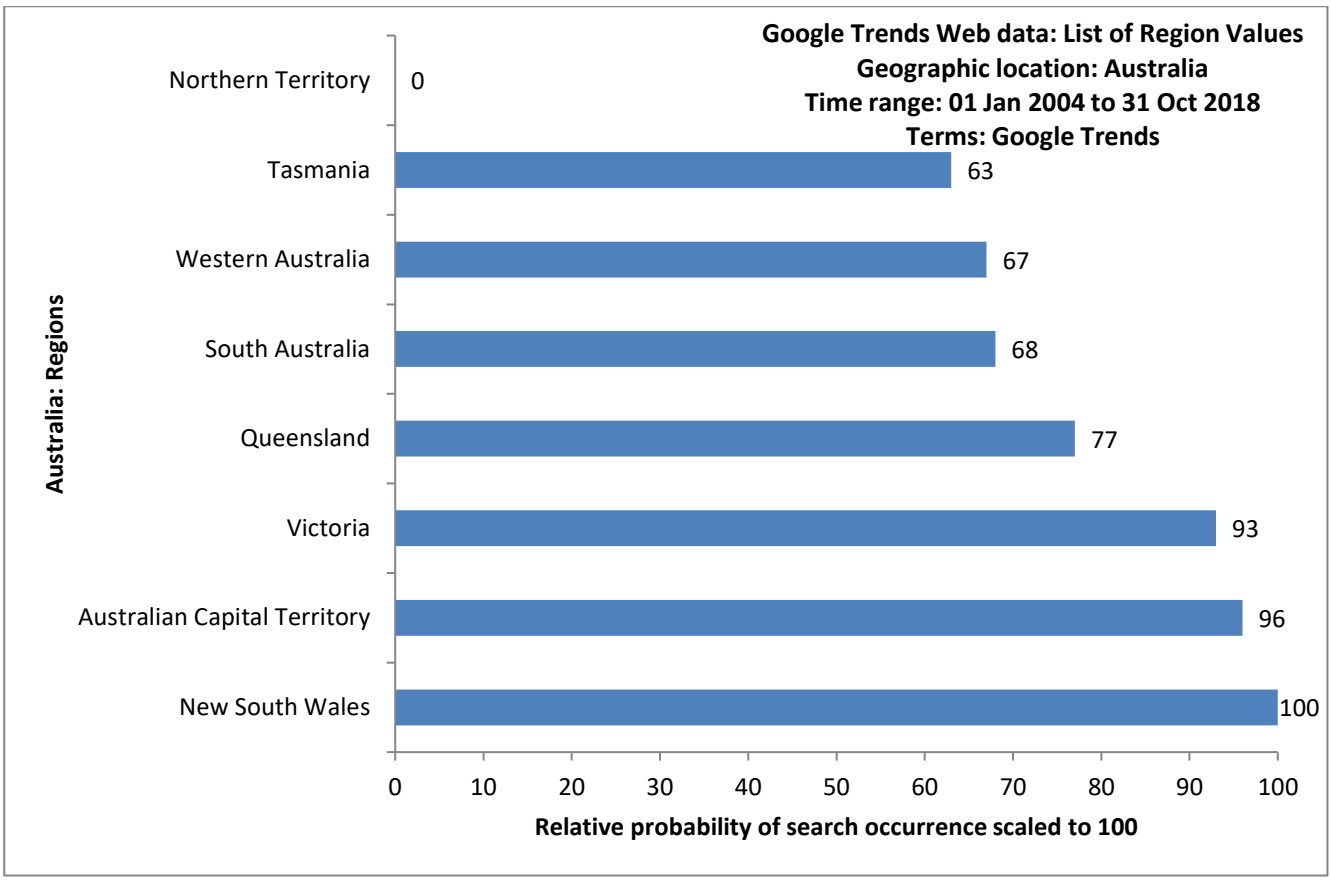


Figure 49 Bar chart for List of Region Values data from Google Trends Web interface

Even though the tool only provides a simple bar chart, enterprising users with the correct version of Excel (Excel 2016 or later, Excel 365) can easily use the Excel 3D map tool[§] to create a similar plot to that shown on the Google Trends website. For example, Figure 50 below shows the same data as Figure 49, transformed using Excel's 3D map feature.

[§] For a short description of the history and use of the tool, see Raubenheimer, 2017[13].

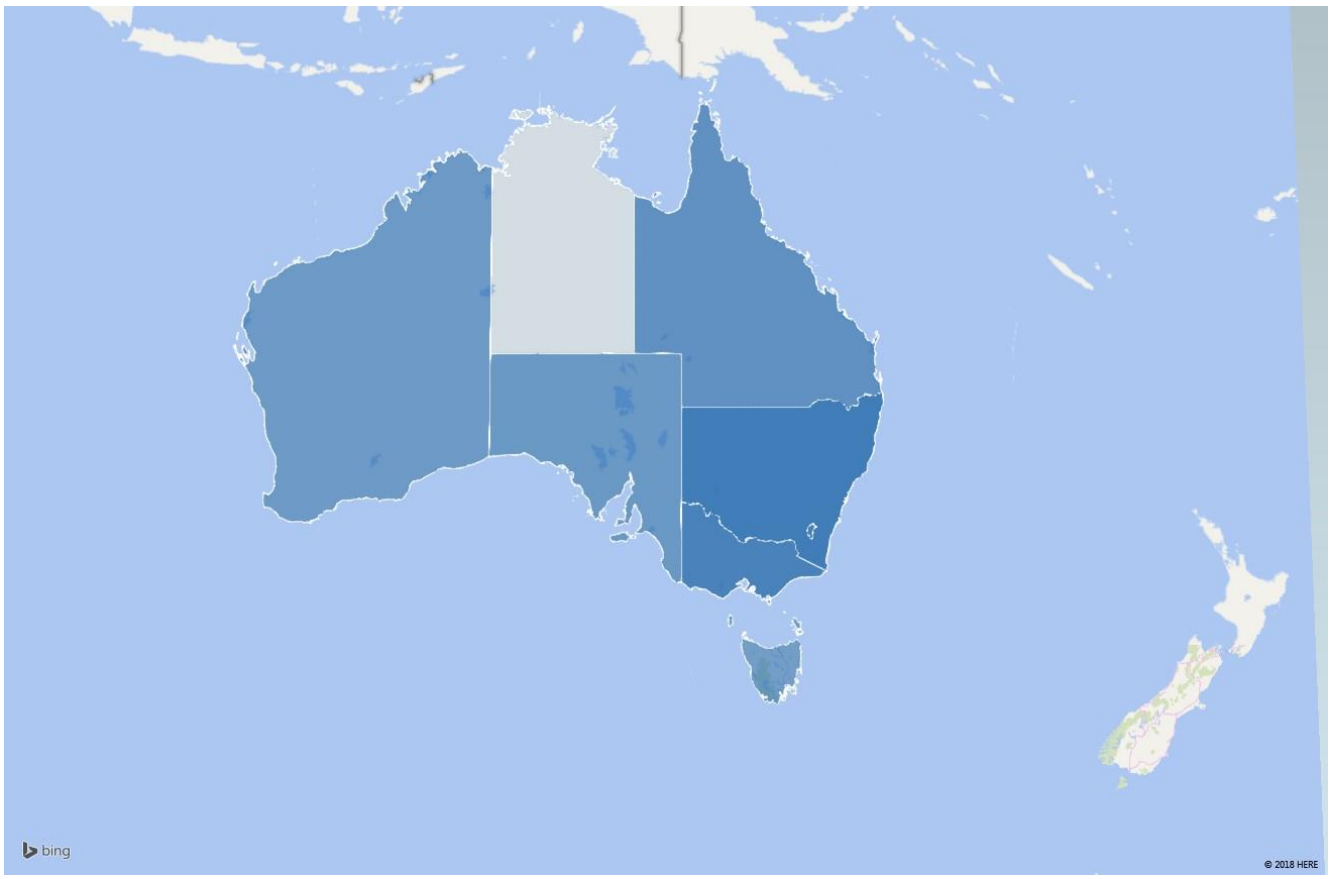


Figure 50 Map chart for List of Region Values data from Google Trends Web interface

Chapter 5 Publications and Citation

If you make use of this tool for a publication, then please cite it, in whatever format your citations need to be (APA format shown):

Raubenheimer, J. E. (2021). Google Trends Extraction Tool for Google Trends Extended for Health data. *Software Impacts*, 8, 100060. https://doi.org/10.1016/j.simpa.2021.100060

I have explained the algorithm I developed to extract multiple samples from Google Trends Extended for Health (cf. section 3.2—Sampling, p. 24) in this publication:
Raubenheimer, J. E. (2022). A practical algorithm for extracting multiple data samples from Google Trends extended for Health. *American Journal of Epidemiology*, *191*(9), 1666–1669. https://doi.org/10.1093/aje/kwac088
However, please note that shortly after this publication appeared, I noticed that this application was returning the same samples each time. I am still trying to get to the bottom of that.
Several further publications detailing certain aspects of the development and use of this tool are currently being prepared. When these have been published, details will be listed in this document.
I have used results obtained using this tool in several conference presentations, as well as this list of publications:

Raubenheimer, J. E., Riordan, B. C., Merrill, J. E., Winter, T., Ward, R. M., Scarf, D., & Buckley, N. A. (2021). Hey Google! Will New Zealand vote to legalise cannabis? Using Google Trends data to predict the outcome of the 2020 New Zealand Cannabis referendum. *International Journal of Drug Policy*, 90. https://doi.org/10.1016/j.drugpo.2020.103083

Chiew, Angela, L., Raubenheimer, J. E., Berling, I., Buckley, N. A., Becker, T., Chan, B., & Brett, J. (2022). Just "nanging" around – harmful nitrous oxide use. A retrospective case series and review of internet searches, social media posts and the coroner's database. *Internal Medicine Journal*, *52*, 1724–1732. https://doi.org/10.1111/IMJ.15391

Bhagavathula, A. S., & Raubenheimer, J. E. (2023). A Real-Time Infodemiology Study on Public Interest in Mpox (Monkeypox) following the World Health Organization Global Public Health Emergency Declaration. *Information (Switzerland)*, *14*(1), 5. https://doi.org/10.3390/INFO14010005/S1

Furthermore, if you do choose to use this tool, it will be appreciated that you notify this developer about what research you used it for, and what publications resulted from it. Regardless of this, as indicated in section 2.5—API key (p. 14), you will need to apply to Google for API key before you can make use of this tool, and Google also require that you keep them up to date with all the research done using that key.

Chapter 6 Version history

The following versions of the tool have been released:

Table 8 Change history for Google Trends Data Extraction Tool

Date	Version	Changes
2018/11/20	1.0	First release version
2018/11/26	1.0.2	Fixed bug in message box reporting of file loaded when using the Load Specifications button. Added code to increment build versions whenever the SetupWorkbook procedure is run.
2018/12/04	1.0.5	Fixed bug in checking dates in JSON string when yearly data are requested (added bDateCheck variable).

2019/02/06	2.0.0	1. Changed the license from Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License to the more appropriate GPLv3 (see https://creativecommons.org/faq/#can-i-apply-a-creative-commons-license-to-software). Since no-one has downloaded the previous versions, this change is not seen to be problematic, even though the by-nc-sa-4.0 is irrevocable. A License module was added, and license code added to the start of each existing module, as specified in https://opensource.org/licenses/GPL-3.0. 2. Added the Suggest button and underlying code: - Sub SuggestDataTarget - Function fGenerateSuggestedName - Sub RemoveIllegalCharsFromFileName As well as adding and modifying named ranges for error messages to allow this to work properly. 3. Updated code to report number of samplings done when all samplings could not be completed. 4. Removed the estimated margin of error calculation from the GTe worksheet interface and added columns to the summary sheets to show 99% margin of error with confidence levels, Coefficient of Variation, Margin of Error as percentage of mean. 5. Added a summary of the Coefficient of Variation and N Needed values This is now in the Sampling Summary sheet, discussed in section 3.6.3. 6. Modified data validation and formula checking for end date to be today-3, instead of today-2. 7. Set Category Selector sheet to only be made xlSheetVeryHidden if bProductionVersion=True. 8. Tweaked mMultiPurposeProcedures.fBreakPathForMsgBox to prevent infinite looping. 9. Added Public Const sAccountURL to mPublicDeclarations and added the View Account button and code. 10. Added module mButtonCode and put all button code on that module, renaming some assigned macros in the process [with the exception that all buttons on the 'Category Selector' worksheet still have their code on that sheet module (Sheet8)]. 11. Changed the *.key Account file to contain two lines, so that account name is stored in second line, added View Account button, and Sub ViewAccount(). 12. Added module mLog, which contain one public enumeration and code to create a log of events (opening, closing, GTe extractions, and GTweb extractions. The entries for each of these four are comma-separated, although they are not compatible. Nonetheless, the log shows all entries in chronological order, not in categories for each type of event. It would be a relatively simple exercise to dump the log file to Excel, filter the event type one desires, and then separate into columns from there.

Date	Version	Changes
2019/02/12	2.1.0	Added the summary of the samples used section (title “Samples used for estimation) to the Sampling summary sheet.
2019/02/27	2.1.1	A correction to the calculation of the MoE as % of mean column on the Summary worksheet (incorrectly calculating CV as % of mean in previous iterations). Added a line of code to SaveAsCopy the complete and cleaned workbook to the export directory when creating a new production version from the mWorkbookSetup module.
	2.1.2	Bug fixes: 1. Declared sPS in mWorkbookSetup.CreateReleaseVersion 2. Added ActiveSheet.Calculate to mMultiPurposeProcedures.SuggestDataTarget 3. Modified the end of: mGTWebFunctions.GetGTWeb mGoogleTrendsInfoExtraction.CompleteReporting mGoogleTrendsInfoExtraction.DrawSample to better accommodate external automation (preventing the EndGracefully call, and closing the created file for the first two procedures)
2019-03-27	2.1.3	Bug fixes: 1. Protection against chart title getting too long with complex queries 2. Fixed Google Trends Web worksheet Start- and End dates being set to InputYearMonth style when using Clear Specifications (now set to InputFullDate style). 3. Removed Shrink to Fit from cell formatting of query specifications. 4. Added sub DialBackColumnWidths to prevent the Columns.Autofit method from making columns too wide when a long query is used.
2019-08-02	2.1.4	Bug fixes: 1. AddSeries Sub line .Name = rngNameAddress places the actual content of the cell into the series name (.value is the default) instead of linking the series name to the cell. This is now corrected. 2. Added “LookAt:=xlPart” to the code in Sub CompleteReporting which clears the Find and Replace settings.

Date	Version	Changes
2020-08-11–2020-12-02	2.2.0	**Changes:** 1. Google tweaked the API, breaking the JSON parsing. I incorporated Daniel Ferry's JSON parser (details in Table 2) to make it more robust against this type of event in the future. The parser made some of my manual parsing code redundant, although that is still included, but commented out. The parser now handles all of the JSON strings returned for all the API calls, and then my reporting code takes the result from there. 2. Modified fGenerateSuggestedName by adding an "If bShowCompletionMsgBoxes Then _" statement so that shortening the name under which the output file will be saved will not generate a message when the process is being automated across multiple runs. 3. Modified fGetLogFile by adding an "If bShowCompletionMsgBoxes Then _" statement so that errors in writing to the log file will not generate a message when the process is being automated across multiple runs. 4. Imported the late Chip Pearson's modArraySupport module (details in Table 2) and replaced most instances of Application.Transpose with his TransposeArray and a modified Range.Resize statement, due to problems with Application.Transpose. **Changes which only apply to my own testing (i.e., if not bProductionVersion then…):** 1. Updated the fWriteToFile procedure so that, when logging json strings returned from the API using my multiple sampling strategy, each new returned string is appended to the existing file, instead of only the first string (which contains values for all dates under my implementation). Technically, this means the json can be imported anywhere else and all the samples reconstituted. 2. Added a line of code to append each query string for each sample to an output file (only in my own testing, i.e., if not bProductionVersion then…). Each query can be used in a browser as a URL. Note that the query strings contain the API account key! **Bug fixes:** 1. Corrected a bug in the updating of the queries used per session for Google Trends Extended for Health queries
2020-12-09–2021-07-01	2.2.1	When TopQueries returns an empty JSON string, still build the output workbook showing that an empty string was returned. Modified the bJSONEmptyString handling so that, when automated from another VBA process, it is reset on each run (to prevent the program from thinking the returned string is empty when it is not). Added error checking for possible "uninterpretable" dates in the JSON string.

Date	Version	Changes
2023-03-07	2.2.2	**Changes:** Modified the double click on the Target File cells so that the either the directory already specified, or if blank, the current working directory, is loaded in the dialog. **Bug fixes:** 2021-01-20: Fixed a bug when summary worksheets contained single quotes which caused an error during the creation of some of the other worksheets (e.g., the 'Sampling Summary' worksheet). 2022-02-23: Fixed a bug with the updated JSON strings returned by the API (cf. v2.2.0)—the year date for yearly data is passed as the integer year, which causes EDate to convert to the wrong date.

References

1. Google. Google Extended Trends API for Health - Getting Started Guide. Google;
2. Johansson E. How to get daily Google Trends data for more than 90 days (Excel) [Internet]. Global Payment Trends. 2013 [cited 2018 Nov 5]. Available from: http://erikjohansson.blogspot.com/2013/04/how-to-get-daily-google-trends-data-for.html
3. Raubenheimer JE, Cairns R, Dawson AH. Using Google Trends for drug interest monitoring—correlating Google Trends data with Poison Centre calls for Nitrous Oxide. In: Asia Pacific Association of Medical Toxicology 17th Annual Scientific Congress 2018. Bali, Indonesia; 2018.
4. Wikipedia. Knowledge Graph. In: Wikipedia [Internet]. 2018 [cited 2018 Nov 5]. Available from: https://en.wikipedia.org/wiki/Knowledge_Graph
5. Wikipedia. Freebase. In: Wikipedia [Internet]. 2018 [cited 2018 Nov 5]. Available from: https://en.wikipedia.org/wiki/Freebase
6. Google. Google Knowledge Graph Search API [Internet]. Google; 2018. Available from: https://developers.google.com/knowledge-graph/
7. Wikidata. Freebase ID. In: Wikidata [Internet]. Wikidata; 2018 [cited 2018 Nov 5]. Available from: https://www.wikidata.org/wiki/Property:P646
8. Google. Google Trends API - Getting Started Guide. Google;
9. Trasborg P. Google Trends Categories [Internet]. GitHub. 2017 [cited 2018 Sep 7]. Available from: https://github.com/pat310/google-trends-api/wiki/Google-Trends-Categories
10. Google. Find related searches [Internet]. Google Trends Help. 2018 [cited 2018 Mar 5]. Available from: https://support.google.com/trends/answer/4355000?hl=en&ref_topic=4365530
11. Matsa KE, Mitchell A, Stocking G. Methodology [Internet]. Pew Research Center. 2017 [cited 2018 Feb 26]. Available from: http://www.journalism.org/2017/04/27/google-flint-methodology/
12. Choi H, Varian H. Predicting the present with Google Trends. Econ Rec. 2012;88(SUPPL.1):2–9.
13. Raubenheimer JE. *Excel*-lence in Data Visualization?: The Use of Microsoft Excel for Data Visualization and the Analysis of Big Data. In: Prodromou T, editor. Data Visualization and Statistical Literacy for Open and Big Data [Internet]. Hershey: IGI Global; 2017. p. 153–93. Available from: http://www.igi-global.com/book/data-visualization-statistical-literacy-open/176342