


Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

## Software Impacts

journal homepage: [www.journals.elsevier.com/software-impacts](http://www.journals.elsevier.com/software-impacts)

Original Software Publication

gtrendsAPI: An R wrapper for the Google Trends API 

Ricardo A. Correia \*

*Biodiversity Unit, University of Turku, 20014 Turku, Finland**Helsinki Lab of Interdisciplinary Conservation Science (HELICS), Department of Geosciences and Geography, University of Helsinki, 00014 Helsinki, Finland**Helsinki Institute of Sustainability Science (HELSUS), University of Helsinki, 00014 Helsinki, Finland*

## ARTICLE INFO

## Keywords:

Google Trends  
Search engine  
Application programming interface  
Internet searches  
Data access

## ABSTRACT

Search engine data is a prime source of insights on the human information-seeking behaviour and such information is instrumental for the scientific study of human culture and behaviour. The gtrendsAPI R software package aims to facilitate programmatic access to data available from the Google Trends API. Here, I introduce the functions available through this software and provide worked examples of how to use it. I also discuss some the potential research applications and caveats of this software and the data available through it.

## Code metadata

Current code version	1.0.0
Permanent link to code/repository used for this code version	<a href="https://github.com/SoftwareImpacts/SIMPAC-2024-32">https://github.com/SoftwareImpacts/SIMPAC-2024-32</a>
Permanent link to reproducible capsule	
Legal code license	MIT license
Code versioning system used	Git
Software code languages, tools and services used	R
Compilation requirements, operating environments and dependencies	jsonlite (>= 1.5), httr (>= 1.4.1)
If available, link to developer documentation/manual	<a href="https://github.com/racorreia/gtrendsAPI/README.md">https://github.com/racorreia/gtrendsAPI/README.md</a>
Support email for questions	<a href="mailto:raheco@utu.fi">raheco@utu.fi</a>

## 1. Introduction

In an increasingly digitized world, the growing availability of data generated online is providing new opportunities for scientific research. Estimates suggest about two-thirds of the global population has access to the internet nowadays [1] and many people use digital online platforms such as search engines, social media and video streaming applications on a daily basis. These habits are providing a new wealth of data about people's interests, preferences and opinions that can generate unique insights on contemporary societies [2].

Search engine data is a particularly useful source of data in this context because it reflects people's information-seeking behaviour, and is thus associated with a conscious effort to acquire information about a topic that is considered of interest or for which there is a perceived information gap [3]. While there are multiple search engines available, and several that provide access to data, chief among all is Google's search engine. Google Search has an estimated 80%–90% share of the

global search engine market [4] and is the dominant search engine in the large majority of countries across the world, with the exception of a few countries such as Russia or China. The Google Trends platform provides access to summary data from searches carried out through Google Search. Google Trends data has been widely used for research in various fields of science, including for example medicine [e.g., [5,6]], economics [e.g., [7,8]], political studies [e.g., [9,10]], psychology [e.g., [11,12]], environmental sciences [e.g., [13,14]], biodiversity conservation [e.g., [15,16]] and sustainability [17].

Data from Google Trends is most commonly accessed through a dedicated website (<https://trends.google.com/trends/>). This platform allows users to query, visually explore and extract search volume data. Global data is available from 2004 onwards and users can select the time period, region, and keyword or topic of interest, but only up to five keywords or topics can be queried simultaneously. This is a limitation for work that requires multiple queries for a large number of keywords

The code (and data) in this article has been certified as Reproducible by Code Ocean: (<https://codeocean.com/>). More information on the Reproducibility Badge Initiative is available at <https://www.elsevier.com/physical-sciences-and-engineering/computer-science/journals>.

\* Correspondence to: Biodiversity Unit, University of Turku, 20014 Turku, Finland.

E-mail address: [raheco@utu.fi](mailto:raheco@utu.fi).

<https://doi.org/10.1016/j.simpa.2024.100634>

Received 5 February 2024; Received in revised form 21 February 2024; Accepted 16 March 2024

2665-9638/© 2024 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## Schematic overview of functions included in gTrendsAPI

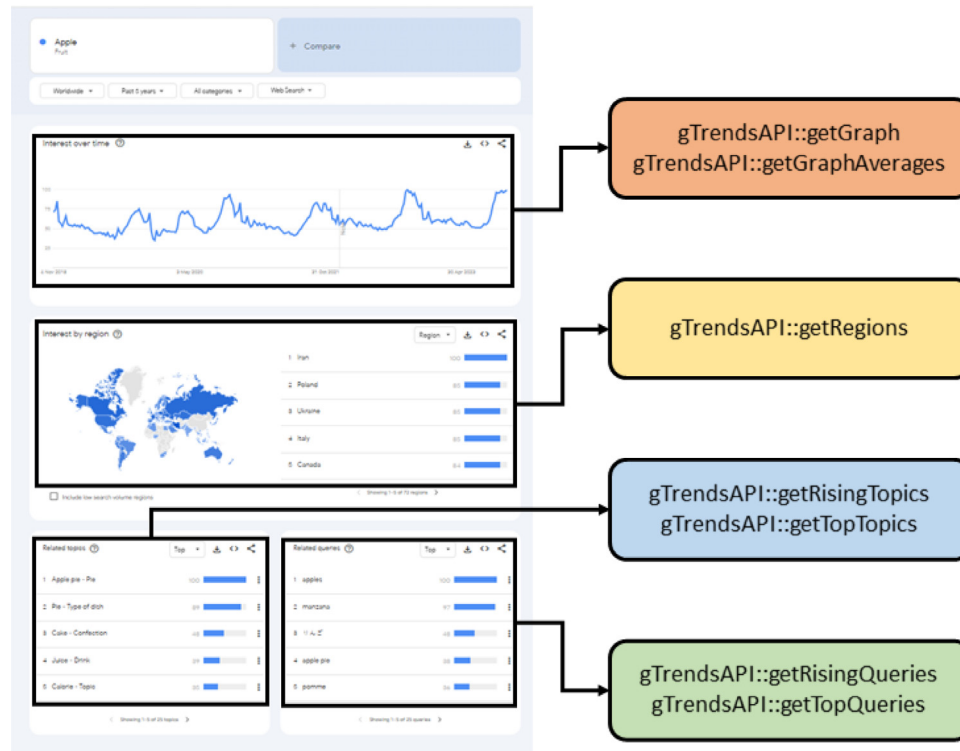


Fig. 1. Schematic overview of the gTrendsAPI software structure. The figure outlines the seven functions available in the software, each designed to access a specific API data node, and how they relate to the output available through the Google Trends website.

or topics, as implementing various combinations of searches manually can be straining and time consuming. One alternative is to access the data programmatically through an Application Programming Interface (API). It is possible to request access to the Google Trends API for research purposes, but there is a limited number of software packages allowing access to this data and they are restricted to a few programming languages. Here, I introduce the gTrendsAPI software package for R programming language which aims to facilitate access to Google Trends data through the official API. This programming language is widely used by scientists in ecology, environmental sciences, and other affiliated areas, and the software may be particularly useful to scientists engaging in culturomics [18] or iEcology [19] research where these data are frequently used.

## 2. Software description

The gTrendsAPI software package contains seven functions, each designed to interact with a specific Google Trends API data node (Fig. 1). The software includes functions to access a time series of relative search volume data for a given query (*getGraph* function) or the average search volume for the terms included in the query (*getGraphAverages* function) over a time period of interest. It also includes a function to extract relative search volume data at the country or region level based on a query of interest (*getRegions* function). Finally, it also includes functions to extract the top (*getTopTopics* function) or rising (*getRisingTopics* function) topics associated, and the top (*getTopQueries* function) or rising (*getRisingQueries* function) queries associated with a query of interest. As depicted in Fig. 1, each function provides access to data that is available through the Google Trends dedicated website and thus it is possible to collect the same data using either approach.

Each function in the gTrendsAPI package contains the same set of key parameters that users need to define to query the API. The ‘terms’ parameter identifies the keywords or topics to be queried. The

Google Trends API accepts both keyword-based queries (e.g., “apple”) or topic-based queries using Google Knowledge Graph identifiers (e.g., “/m/014j1m” is the identifier used for the fruit apple). Users can query the Google Knowledge Graph API, for example through the *gkgraph* R package [20] to identify the identifiers of their topics of interest. The ‘geo’ parameter identifies a country or region of interest for the query. This parameter is set to NULL by default, which queries the API for worldwide data, but it is also possible to use country or region codes to define the scope of the search. A list of valid country and region codes is available in the gTrendsAPI package (see Section 3. Illustrated examples for details). The ‘startDate’ and ‘endDate’ parameters define the temporal scope of the query. These parameters take year and month information as input, in the form of a “YYYY-MM” string, and both parameters are set to NULL by default representing January 2004 and the start date and the current month and year as end date. Finally, the ‘api.key’ parameter defines the user’s API key necessary to obtain access to the API. Use of the API is currently restricted to users with approved access, and at the time of writing access can be requested for research purposes through an online form (available here: [https://support.google.com/trends/contact/trends\\_api](https://support.google.com/trends/contact/trends_api)). All functions contain also the ‘property’ and ‘category’ parameters that users can use to identify the property of interest or filter results to a specific category. The ‘property’ parameter defaults to web searches, and the ‘category’ parameters defaults to an unfiltered query [as recommended; see [21]].

Using any of the functions once is equivalent to one API call and thus consumes one API quota. While it is possible to implement multiple calls in succession, thus allowing for swift and automated data extraction, users should be mindful of their respective daily and monthly API quota limits as these will delimit data access.

## 3. Illustrative example

To use the gTrendsAPI package, users must first install the package from its Github repository. This is possible to do directly from within

**Table 1**

Example of the summary output obtained by calling function *getGraph*. The output includes information of the relative search volume (column 'value') for a given search term (column 'keyword'), time (column 'date') and geography (column 'geo'). The remaining columns hold information about the scope of the call, including the requested time period of interest (column 'time'), property (column 'gprop') and category (column 'category'). Note that the figures representing search volume in the example below were randomly generated and do not represent actual data from Google Trends.

Value	Date	Geo	Time	Keyword	Gprop	Category
31	2004-01-01	world	2004-01 2024-01	apple	web	All categories
71	2004-02-01	world	2004-01 2024-01	apple	web	All categories
32	2004-03-01	world	2004-01 2024-01	apple	web	All categories
69	2004-04-01	world	2004-01 2024-01	apple	web	All categories
95	2004-05-01	world	2004-01 2024-01	apple	web	All categories

the R environment by running the following code (please not that you may need to install also the R package *devtools* beforehand):

```
R> # install.packages("devtools")
R> devtools::install_github("racorreia/gtrendsAPI", build_vignettes = T)
It is then possible to load the package by calling the following code:
R> library(gtrendsAPI)
```

After loading the package, users will also need to record their Google Trends API key to allow access to the API. This is possible to do so directly in the code by creating an object with the key, or by creating an environmental variable in the system. The latter approach is advised as it allows the user to share their code without revealing the API key. These approaches are exemplified below:

```
R> # Create Object with API key
R> key <- "YOUR_API_KEY"
R> # Create an environmental variable with the API key
R> Sys.setenv(GTkey = "YOUR_API_KEY")
R> # Call the key from the set of environmental variables and store it
in an object
R> key <- Sys.getenv("GTkey")
```

It is then possible to call any of the functions in the *gtrendsAPI* package by simply defining the search terms and API key. For example, if a user was interested in getting a time series of relative search volume for the term "apple" it would be possible to do so using the following code:

```
R> # Get time series of search volume for the keyword "apple" and
store in an object
R> apple_keyword <- getGraph(terms = "apple", api.key = key)
R> # Check the output of the call
R> head(apple_keyword)
```

The above code would print the first lines of the function output as stored in the object, which is exemplified below in [Table 1](#):

It is also possible to call the function using an identifier, or a list of up to five keywords and/or identifiers, and to define additional parameters as exemplified below:

```
R> # Get time series of search volume for the topic "apple (fruit)" for
the year 2020 and store in an object
R> apple_topic <- getGraph(terms = "/m/014j1m", startDate = "2020-
01", endDate = "2020-12", api.key = key)
R> # Get time series of search volume for the keywords "apple" and
"pear" targeting news and store in an object
R> apple_vs_pear <- getGraph(terms = c("apple", "pear"), property =
"news", api.key = key)
```

The same approach can be used to customize the call of any other function included in the *gtrendsAPI* R package.

#### 4. Impact

The use of Google Trends data in scientific research has grown steadily since the platform first became available, and has expanded to numerous fields of inquiry covering the natural and social sciences [22]. It has for example been widely used after the onset of the COVID-19 pandemic to explore disease incidence dynamics [5,6] and the impact of the pandemic on people's interests and well-being [12,23,24]. This package aims to further stimulate its uptake by facilitating

programmatic access to the Google Trends API, thereby streamlining data access and enabling the integration of data extraction and analysis pipelines. It contains a set of custom functions that provide access to the existing Google Trends API nodes and allow access to data pertaining to temporal and spatial patterns of search interest for a query of interest, and to topics and terms associated with that query. These features are a novel functionality in R software as currently there seems to be no other software package available that is designed to interact directly with the Google Trends API, but they parallel the functionalities of software available for other programming languages such as Python. Some of the functions available in the *gtrendsAPI* package have already been to support scientific research, for example to obtain data on search interest towards hundreds of animal and plant species [e.g., [25,26]].

Still, there are a few limitations that users of this software should take into account. Unlike other software packages that exploit the API that was designed for the Google Trends website, this package was designed to interact with a separate and dedicated API but is only available to users that have secured access to it. Users of this software should therefore abide by the terms of use of the Google Trends API and must be mindful of any limitations they impose, for example with regards to commercial applications or sharing any data that were directly derived from the API. Time series of relative search volume data are particularly suited to the analysis of temporal trends in search interest, but users should take care in validating the data represent their topic of interest [27] and in interpreting the observed trends [28,29]. It should also be noted that Google supports and maintains a set of functions designed to interact with the Google Trends API in Python and that users should refer to this software for future updates. Finally, this package is not designed to provide access only to the Google Trends API and not to the Google Health Trends API which provides access to a different dataset. Users interested in the latter API can see for example [30].

#### 5. Conclusions

As access to the internet continues to grow worldwide, search engine data is likely to remain a prime source of insights on the interests, opinions, and preferences of people from across the world as reflected through their information-seeking behaviour. This information can be of great utility to a broad range of research areas, particularly those related to the scientific study of human culture and behaviour [31]. The *gtrendsAPI* package aims to facilitate and further stimulate research in these fields by providing programmatic access to data available from the Google Trends API using the R programming language. These data represent searches carried out through Google Search, the most widely used search engine across the world, and are thus uniquely suited to explore large-scale patterns of information-seeking behaviour.

#### CRedit authorship contribution statement

**Ricardo A. Correia:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Funding acquisition, Conceptualization.

## 1 Declaration of competing interest

2 The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## 5 Acknowledgements

6 RAC acknowledges support from the Academy of Finland (Grant agreement #348352) and the KONE Foundation (Grant agreement #202101976). The author would also like to thank Google for making the Google Trends tool and API available to researchers.

## 10 References

- 11 [1] International Communication Union, Measuring Digital Development, Facts and  
12 Figures 2022, International Communication Union, 2022.
- 13 [2] R.A. Correia, et al., Digital data sources and methods for conservation  
14 culturomics, *Conserv. Biol.* 35 (2021) 398–411.
- 15 [3] D.O. Case, L.M. Given, Looking for Information: A Survey of Research on  
16 Information Seeking, Needs, and Behavior, fourth ed., Emerald, 2016.
- 17 [4] StatCounter, Market share of leading desktop search engines worldwide  
18 from 2015 to 2023, 2023, Statista. [Online]. Available: <https://www.statista.com/statistics/216573/worldwide-market-share-of-search-engines/>. (Accessed 2  
19 January 2024).
- 20 [5] L.B. Amusa, et al., Modeling COVID-19 incidence with Google Trends, *Front.  
21 Res. Metr. Anal.* 7 (2022).
- 22 [6] A. Mavragani, K. Gkillas, COVID-19 predictability in the United States using  
23 Google Trends time series, *Sci. Rep.* 10 (20693) (2020).
- 24 [7] D. Borup, E.C.M. Schütte, In search of a job: Forecasting employment growth  
25 using google trends, *J. Bus. Econom. Statist.* 40 (2022) 186–200.
- 26 [8] H. Choi, H. Varian, Predicting the present with google trends, *Econ. Rec.* 88  
27 (2012) 2–9.
- 28 [9] J. Mellon, Internet search data and issue salience: The properties of google trends  
29 as a measure of issue salience, *J. Elect. Public Opin. Parties* 24 (2014) 45–72.
- 30 [10] J. Mellon, Where and when can we use google trends to measure issue  
31 salience? *PS: Political Sci. Politics* 46 (2013) 280–290.
- 32 [11] R.A. Correia, S. Mammola, The searchscape of fear: A global analysis of internet  
33 search trends for biophobias, *People Nat.* (2023).
- 34 [12] E.A. Halford, et al., Google searches for suicide and suicide risk factors in the  
35 early stages of the COVID-19 pandemic, *PLoS One* 15 (2020) e0236777.
- 36 [13] W.R.L. Anderegg, G.R. Goldsmith, Public interest in climate change over the  
37 past decade and the effects of the ‘climategate’ media event, *Environ. Res. Lett.*  
38 9 (2014) 054005.
- 39 [14] J. Kam, et al., Monitoring of drought awareness from google trends: A case study  
40 of the 2011–17 california drought, *Weather Clim. Soc.* 11 (2019) 419–429.
- 41 [15] G.H. de Oliveira Caetano, et al., Evaluating global interest in biodiversity and  
42 conservation, *Conserv. Biol.* 37 (2023) e14100.
- 43 [16] L.T.P. Nghiem, et al., Analysis of the capacity of google trends to measure  
44 interest in conservation topics and the role of online news, *PLoS One* 11 (2016)  
45 e0152802.
- 46 [17] R.A. Correia, E. Di Minin, Tracking worldwide interest in sustainable  
47 development goals using culturomics, *PLoS Sustain. Transf.* 2 (2023) e0000070.
- 48 [18] R.J. Ladle, et al., Conservation culturomics, *Front. Ecol. Environ.* 14 (2016)  
49 269–275.
- 50 [19] Jarić I., et al., iEcology: Harnessing large online resources to generate ecological  
51 insights, *Trends Ecol. Evol.* (2020) <http://dx.doi.org/10.1016/j.tree.2020.03.003>.
- 52 [20] R. Correia, gkgraphR: Accessing the official google knowledge graph API, in: R  
53 Package Version 1.0.2, Zenodo, 2021.
- 54 [21] A. Van Huynh, Use of google trends categories in conservation culturomics,  
55 *Conserv. Biol.* 37 (2023) e14103.
- 56 [22] S.-P. Jun, et al., Ten years of research change using Google Trends: From the  
57 perspective of big data utilizations and applications, *Technol. Forecast. Soc.  
58 Change* 130 (2018) 69–87.
- 59 [23] A. Brodeur, et al., COVID-19, lockdowns and well-being: Evidence from Google  
60 Trends, *J. Public Econ.* 193 (2021) 104346.
- 61 [24] C.N. Souza, et al., No visit, no interest: How COVID-19 has affected public  
62 interest in world’s national parks, *Biol. Cons.* 256 (2021) 109015.
- 63 [25] M. Adamo, et al., Dimension and impact of biases in funding for species and  
64 habitat conservation, *Biol. Cons.* 272 (2022) 109636.
- 65 [26] S. Mammola, et al., Towards a taxonomically unbiased European union  
66 biodiversity strategy for 2030, *Proc. R. Soc. Lond. [Biol]* 287 (2020) 20202166.
- 67 [27] R.A. Correia, Google trends data need validation: Comment on Durmuşoğlu  
68 (2017), *Hum. Ecol. Risk Assess.* 25 (2019) 787–790.
- 69 [28] R.A. Correia, et al., Inferring public interest from search engine data requires  
70 caution, *Front. Ecol. Environ.* 17 (2019) 254–255.
- 71 [29] G.F. Ficetola, Is interest toward the environment really declining? The complexity  
72 of analysing trends using internet search data, *Biodivers. Conserv.* 22 (2013)  
73 2983–2988.
- 74 [30] J.E. Raubenheimer, Google trends extraction tool for google trends extended for  
75 health data, *Softw. Impacts* 8 (2021) 100060.
- 76 [31] Albuquerque U.P., et al., Exploring large digital bodies for the study of human  
77 behavior, *Evol. Psychol. Sci.* 9 (2023) 385–394.
- 78