

Manuscript accepted for publication in
Advances in Methods and Practices in Psychological Science.

Google Search Data for Psychological Scientists: A Tutorial and Best Practices

Jordan W. Moon ^{*1,2,3,4} & Michael Barlev ⁴

¹ School of Psychology, University of Southampton, Southampton, United Kingdom

² Centre for Culture and Evolution, Brunel University of London, Uxbridge, United Kingdom

³ Institute for Advanced Study in Toulouse, Toulouse School of Economics, and Toulouse Capitole University, Toulouse, France

⁴ Department of Psychology, Arizona State University, Tempe, Arizona, USA

* Correspondence should be addressed to Jordan W. Moon, School of Psychology, University of Southampton, B44 University Rd, Southampton, United Kingdom, SO17 1BJ.

Email: j.w.moon@soton.ac.uk

Conflict of Interest Statement: The authors declared no conflicts of interest with respect to the authorship or the publication of this article.

Funding Statement: The writing of this manuscript was supported by the Templeton World Charity Foundation, Inc. (funder DOI 501100011730) through grant #30290 (<https://doi.org/10.54224/30290>). JWM also acknowledges funding from the French Agence Nationale de la Recherche (under the Investissement d'Avenir program, ANR-17-EURE-0010).

Ethical Statement: This manuscript does not involve human participants, personal data, or experimental procedures requiring ethical approval. All data included are publicly available and no identifiable individual information is used.

Abstract

Google searches have been described as the most important dataset on the human psyche ever assembled. Google search data—accessible through a tool called Google Trends—can provide new insights on topics as varied as stereotypes and prejudices, political attitudes, religious identity and belief, personality, motivations, psychological well-being, mental health, and culture. Google Trends can generate highly customized datasets: Users can compare the popularity of search terms across most of the world, or access longitudinal data as far back as 2004, and they can do so with high geographical and temporal granularity. Notwithstanding these opportunities, Google Trends has significant limitations. Without appropriate caution, users can easily rely on data that are not meaningful or draw mistaken conclusions. We provide a comprehensive overview and tutorial, covering (a) opportunities of Google Trends for psychological scientists; (b) how Google Trends scores are calculated, how reliable they are, and why some queries might yield low-quality data; (c) instructions with accompanying R code for creating custom datasets beyond what Google Trends provides by default; (d) example analyses for studies that could be done using Google Trends data; (e) an overview of common pitfalls; and (f) recommendations for safeguarding data quality and their interpretation.

Keywords: big data, archival data, internet search volume, Google Trends, longitudinal data

Google Search Data for Psychological Scientists: A Tutorial and Best Practices

After the 2008 U.S. presidential election, when Democrat Barack Obama defeated Republican John McCain, most pundits and journalists concluded—largely based on national polls—that race did not play a large role in how Americans voted. However, internet search data told a different story: U.S. regions that more frequently searched for racist slurs—suggesting racism—voted less for Obama than would be expected based on prior election results. By this estimate, racism cost Obama about 4% of the popular vote (Stephens-Davidowitz, 2014). Although traditional self-report methods largely missed this race penalty, it was revealed not by what people said to pollsters, but by their online behavior.

Additionally, sometimes search behavior may be the only reliable data source available. Consider public opinion in Russia. After the launch of the invasion of Ukraine in 2022, public opinion polls in Russia claimed that the populace has rallied behind the war effort, was uniformly supportive of the Kremlin, and had a soaring overall life satisfaction. Such claims would be difficult to refute using traditional surveys, given Russian governmental interference and self-censorship by Russian citizens. Using Google search data, researchers found evidence of widespread political dissent among Russians, increased financial stress, and declining psychological well-being (Foa & Nezi, 2023).

Google—which handles more than 90% of internet searches worldwide¹—provides a free tool called Google Trends (<https://trends.google.com>), which allows users to compare the popularities of search terms and explore how popularities fluctuate over time and across

¹ As of February 2024. Google has a majority of the market share in most countries. Notable exceptions are Russia (where a search engine called Yandex is more popular) and China (where Google search is blocked and search engines like Baidu are used instead; <https://gs.statcounter.com/search-engine-market-share>).

geographical regions. Google search data have massive—and largely untapped—potential for psychologists and social scientists. However, the user-friendly Google Trends interface belies many complexities in how these data are to be interpreted. Without being aware of such complexities and nuances, researchers risk being led astray.

We provide a comprehensive overview and tutorial, covering (a) the opportunities afforded by Google Trends data, (b) how Google Trends scores are calculated, how reliable they are, and why some queries might yield low-quality data, (c) instructions with accompanying R code for creating custom datasets beyond what Google Trends provides by default, (d) example analyses of Google Trends data (i.e., potential studies that could be done using Google Trends data), (e) an overview of common pitfalls, and (f) recommendations for best practices.

To be clear, while we view Google Trends as a useful resource for psychological scientists, this tutorial should not be taken as unreserved enthusiasm for using Google Trends to answer any psychological question. Our position is that Google Trends data should be treated with healthy skepticism and supplemented with additional evidence wherever possible.

Why Google Search Data Are Useful to Psychologists

Google search data have been described as the most important dataset on the human psyche ever assembled (Stephens-Davidowitz, 2017). Of course, Google Trends differs greatly in approach from self-report surveys or psychological experiments, where researchers have close control over question wordings and stimuli design. Google Trends is similar to other sources of “big data” such as social media data that are increasingly popular in psychology (Chen & Wojcik, 2016; Harlow & Oswald, 2016).

When using big data, researchers are rarely able to customize exactly what data are available to them; nonetheless, such data allows researchers to test hypotheses across many

cultures and even over large spans of history. For example, researchers can glean information about constructs of interest by examining language usage over time and across space in large corpuses of books or transcriptions of speech (Jackson et al., 2022; Michel et al., 2011; Wang & Inbar, 2021); by analyzing posts on social media sites like Instagram and Twitter/X (Blake et al., 2018; Brady et al., 2017; Brooks et al., 2022; Kern et al., 2016) and online forums such as Reddit (Proferes et al., 2021); or by examining Wikipedia entries (Baumard et al., 2022).

Google Trends is a particularly promising source of data for psychologists, as it allows users to create datasets neatly tailored to their specific research question. Table 1 presents several previously-published and possible uses of Google Trends across several subfields of psychology. We believe that, through creative thinking, researchers of all types will be able to find Google Trends useful.

Table 1. Previously-published and possible uses of Google Trends across subfields of psychology.

Subfield	Construct(s) of Interest	Examples
Clinical	Mental health and psychological well-being can be operationalized using negative emotion topics.	Foa et al. (2022) validated several negative emotion topics (e.g., <i>Sadness, Stress, Fear, Boredom, Apathy</i>) against UK national polling data, and created a composite well-being measure, which they used to track how negative affect in the UK changed during the COVID-19 pandemic. Other researchers have validated measures of psychological well-being or related constructs in other countries (Algan et al., 2019; Greyling & Rossouw, 2025).
	Suicide risk can be assessed by phrases related to suicide.	Ma-Kellams et al. (2016) found that a composite measure, using searches containing words like <i>how to suicide</i> and <i>painless suicide</i> , correlated strongly with completed suicides at the state level.
Cognitive	Long-term memory could be proxied by considering the types of searches people who are having memory difficulties might make.	Users could proxy long-term memory using searches containing words such as <i>what is called</i> or <i>what is name</i> . Compatible with this proposal, Sparrow et al. (2011) suggested people use the internet as an “external memory.”
Cultural	Cultural tightness —the extent to which a culture has strong norms and is intolerant of deviant behavior (Gelfand et al., 2011)—	Ma et al. (2024) created and validated an index of cultural tightness using searches for topics such as <i>Conformity</i> and <i>Morality</i> , and an index of cultural looseness using

	can be operationalized using several topics related to norms and values.	topics such as <i>Autonomy</i> and <i>Creativity</i> . These indices correlated highly with previously-published measures.
	Cultural religiosity (Gebauer & Sedikides, 2021) can be operationalized using searches related to God/gods and places of worship.	Adamczyk et al. (2022) estimated cross-national religiosity scores using searches for the major holidays of the dominant religions across countries. These scores predicted several key outcomes.
Developmental	Parent-child relationships could be explored using the top related searches for phrases such as <i>is my child...</i>	Stephens-Davidowitz (2017) found a greater search volume for <i>is my son gifted</i> than <i>is my daughter gifted</i> , suggesting that parents are likelier to associate exceptional intelligence with sons rather than daughters, potentially reflecting—and reinforcing—existing stereotypes.
Personality	Personality traits could be operationalized by considering the types of searches people with different personality profiles might make.	Using the Big Five: Users in regions with greater openness to experience might search more for topics related to <i>travel</i> ; users higher in neuroticism might be more likely to search for terms such as <i>symptoms</i> ; users higher in agreeableness might be more likely to search for topics related to <i>empathy</i> .
Political	The popularity and spread of social movements could be proxies using searches for movement-specific slogans.	Users could assess the relative popularity of political movements (which might be related to average attitudes in a region) by comparing the searches including phrases such as <i>“back the blue”</i> vs. <i>“defund the police.”</i>
Social	Racism can be operationalized using taboo slurs or phrases (unlikely to be used by people without at least some prejudice).	Stephens-Davidowitz (2014) found that US districts that searched more for anti-Black slurs were less likely to vote for Obama than for previous Democratic presidential candidates. See also Connor et al. (2019) who tested hypotheses about racism and income inequality.
	Status motivations can be measured by interest in high-status consumer products.	Walasek and Brown (2016) used interest in luxury brands such as Prada, Louis Vuitton, and Chanel as a proxy for status motivations, showing that interest is higher in countries with more income inequality.

Overview

We build on existing resources, most notably the now decade-old guide by Stephens-Davidowitz and Varian (2014), tutorials provided by Google (beginning with *Basics of Google Trends*, 2024), and key articles and posts overviewing or scrutinizing uses of Google search data (Arora et al., 2019; Cebrián & Domenech, 2023; Choi & Varian, 2012; Mavragani & Ochoa, 2019; Mellon, 2014; Nuti et al., 2014; Rogers, 2016). For a review of several clever uses of

Google Trends data, we recommend Stephens-Davidowitz’s (2017) book *Everybody Lies*. Our article goes far beyond existing resources, providing new analyses and simulations demonstrating when and why Google Trends yields high quality data—and when it does not. Additionally, we cover many possible pitfalls (and solutions to them) which have not been extensively discussed elsewhere.

The examples used throughout this article can be replicated using the basic R code provided in the main text. All data and code are available on OSF: <https://osf.io/74v8p/>. Additional information, examples, and figures are available in the Supplementary Materials.

Table 2. Defining Key Terms.	
Term	Definition
<i>Search</i>	“Search” refers to people typing words and phrases into Google.
<i>Search volume; search popularity</i>	Generic terms referring to the quantity of searches people make on Google.
<i>Query</i>	“Query” refers to Google Trends users accessing data on Google searches—it does not distinguish between keywords and topics
<i>Keyword</i>	A Google Trends query for a specific word(s) or phrase. We italicize and lower-case topics (e.g., <i>weather</i>), proper nouns excepted.
<i>Topic</i>	Google Trends calculates topics using multiple related keywords (although it is not forthcoming with which keywords it uses nor how they are used to calculate topics). Topics allegedly include synonyms, misspellings, and translations to other languages—but see Pitfalls section. We italicize and capitalize topics (e.g., <i>Weather</i>).
<i>Google Trends Explorer</i>	The Google Trends online console (https://trends.google.com/trends/explore).
<i>Retrieval day</i>	The day on which a Google Trends user makes a query (as distinguished from “day” as a time period in a dataset).

Opportunities Afforded by Google Trends

Opportunity 1: Google Trends Equalizes Access to Data

Google Trends is free and easy to use, which equalizes access to data—this particularly benefits early career researchers and researchers from institutions and countries with fewer resources such as grant funding. Whereas cross-cultural research using traditional methods often requires substantial investments of time and resources, including the building and maintaining of large collaborative networks, Google Trends gives researchers instant access to data from around the globe.

Additionally, in contrast to other open-access datasets, Google search data are highly customizable, meaning that they can be used to investigate a wide range of questions, with researchers not constrained by the choices of other researchers (e.g., the specific wording of a survey item in a publicly accessible dataset).

Opportunity 2: Google Search Data Are Available in “Real Time”

Traditional systems that track social, economic, or epidemiological trends typically lag by days or weeks; yet, for many applications, such as tracking the spread of infectious diseases, the sooner data are available, the better and more effective interventions can be. Google Trends data are made available in real time, and therefore hold the promise of identifying what is happening and where much more quickly than traditional tracking systems (Choi & Varian, 2012; Ginsberg et al., 2009).

Perhaps the most famous use of Google search data, “Google Flu Trends,” was designed to predict influenza outbreaks before traditional trackers (Ginsberg et al., 2009). When people have health problems, they often turn to Google, searching for information about their symptoms. If, in a given geographical region, searches such as *fever*, *muscle aches*, and/or *flu*

symptoms increase, it might mean that the flu or flu-like infections are spreading in that region (perhaps even before many are diagnosed by doctors).

Ginsberg et al. (2009) used an automated algorithm to build a model, based on 45 Google searches, to predict the spread of influenza from real-time Google Trends data. This model was validated against more traditional national influenza tracking systems used by the Centers for Disease Control and Prevention, such as weekly reports from US laboratories. Importantly, Google Flu Trends predictions were available much earlier than reports based on traditional tracking systems, thereby allowing medical professionals to take preemptive steps to face influenza outbreaks before they became severe. (However, see Pitfall 3 for important caveats.)

Beyond such epidemiological uses, real-time data have many applications, such as “nowcasting” market behavior (Choi & Varian, 2012; Eichenauer et al., 2022), including tracking the very recent past and predicting the near future (Carrière-Swallow & Labbé, 2013; Nagao et al., 2019) and mapping how thoughts and feelings change in response to real-world events. For example, Google data revealed that searches associated with Islamophobia (e.g., *kill Muslims*) rose during and soon after a speech by Barack Obama following a 2015 mass shooting in San Bernardino, California (Stephens-Davidowitz, 2017).

In sum, Google Trends can provide “real-time,” and temporally fine-grained, insights into potentially consequential changes in internet search behavior. Although to date this has primarily been used to complement and improve on traditional tracking systems of epidemiological, social, and economic shifts, the range of possible applications to psychology is much broader.

Opportunity 3: Google Trends Data Are Not Susceptible to Socially Desirable Responding

Many people harbor attitudes or desires that they are reluctant to share with others—even anonymously on surveys. However, some of these attitudes or desires might be revealed through patterns of Google searches.

Stephens-Davidowitz (2017) provides myriad such examples: What people think of their sons vs. daughters (by comparing the popularity of searches for *son gifted* vs *daughter gifted*), whether they regret having children (by comparing the popularity of searches for *regret having children* vs. *regret not having children*), what stereotypes they hold about different groups (by exploring what words most commonly follow “*Why are [African-Americans/Muslims/Jews]* . . .”). As discussed earlier, such Google searches are associated with real-world behavior—searches for racial epithets were associated with lower-than-expected popular vote shares for Barack Obama in the 2014 US presidential elections (Stephens-Davidowitz, 2014), and searches for *kill Muslims* were associated with hate crimes toward Muslims (Stephens-Davidowitz, 2017).

In short, Google search data represent the information people seek, often in the privacy of their own home and when no one else is watching—this is a unique advantage over alternative research tools and datasets, including other sources of big data.

Opportunity 4: Google Search Data Includes Many People Otherwise Inaccessible to Research Studies

Google Trends data might reach people at times when they are otherwise unlikely to be included in research data. In the US, 63% of Google searches are now made from mobile phones (Statista, 2023), and this includes searches at all hours of the day. This means that, unlike traditional research methods that try to overcome such place-time-situation limitations, such as experience sampling, it is possible to examine Google search trends at any time of day. For

example, Stephens-Davidowitz (2017) found that late at night is when people are more likely to search for information about suicide and meaning in life.

Additionally, Google search data can include people at locations around the globe where they would be hard to reach by researchers. In recent years, psychologists have increasingly understood the importance of including more diverse populations in their research. Yet, obtaining cross-cultural data typically requires substantial investments of time and resources, leaving such populations out of reach to most. Google Trends includes most countries around the globe, including many cultures that are poorly represented in the psychological literature (that is, non “WEIRD” cultures; Henrich et al., 2010), or that are difficult to reach due to governmental censorship (Foa & Nezi, 2023).

Opportunity 5: Google Search Data Can Be Used Retroactively with Natural Experiments

After important world events (e.g., the COVID-19 pandemic, the 2022 Russian invasion of Ukraine), researchers often look for data that would allow them to interpret the effects of those events. For example, social scientists might look for survey data serendipitously collected before events to use as a baseline which with to compare data they collect after those events. With Google Trends, social scientists do not have to rely on serendipity—they can access baseline data before any event in the past two decades, for many locations around the globe, which they can then compare to post data, either from Google Trends or collected by more traditional methods. See Grosz et al. (2024) for an introduction to natural experiments and their potential benefits for psychological science.

Basics of Google Trends

Downloading Google Trends Data

Users can download Google search data using the Google Trends explorer (<https://trends.google.com/trends/explore>). In this tutorial, all data were accessed via the explorer, and this is what we assume users will do.

Although there are tools to automate downloads—most notably, the Python and R packages `pytrends` and `gtrendsR` (Hogue & DeWilde, 2023; Massicotte & Eddelbuettel, 2022)—they are not supported by Google, making them *pseudo-APIs*. Because of this, we do not discuss them further.

Google does support an official Google Trends API, sometimes called Google Extended Trends for Health (Raubenheimer, 2021). Users must apply for access to this tool (https://support.google.com/trends/contact/trends_api) and provide information about their proposed research, and agree to conditions of use (which prohibit sharing of data). If approved, users can use Python to download Google search data, using an API key. Data from this tool are scaled in an easier-to-interpret way than standard Google Trends data².

Because most users will use standard Google Trends data, and because the guidelines presented also apply to Google Trends API (except that some of our techniques for creating custom datasets are not necessary), we do not discuss this tool further. (See Raubenheimer, 2021, for more information on this API.)

Types of Data Google Trends Provides

² Specifically, whereas Google Trends data are scaled such that the highest score is 100, and all scores are rounded to the nearest integer, scores from the Google Trends API represent the raw search probability (i.e., the probability that the selected search term was included in a search within the selected time period and region) multiplied by 10^7 (for interpretability). As the data are not rescaled for each search, this means that users do not need to use rescaling techniques to create custom datasets. Additionally, the Google Trends API allows greater customization, allowing users to select the temporal granularity (i.e., daily, weekly, monthly) across the desired period.

Google Trends provides multiple types of data. We focus on longitudinal data (“interest over time” panel in the Google Trends explorer) and cross-sectional data (“interest by region” and, where multiple terms are queries at once, “compared breakdown by region”). We explain how Google Trends calculates these and how to interpret them.

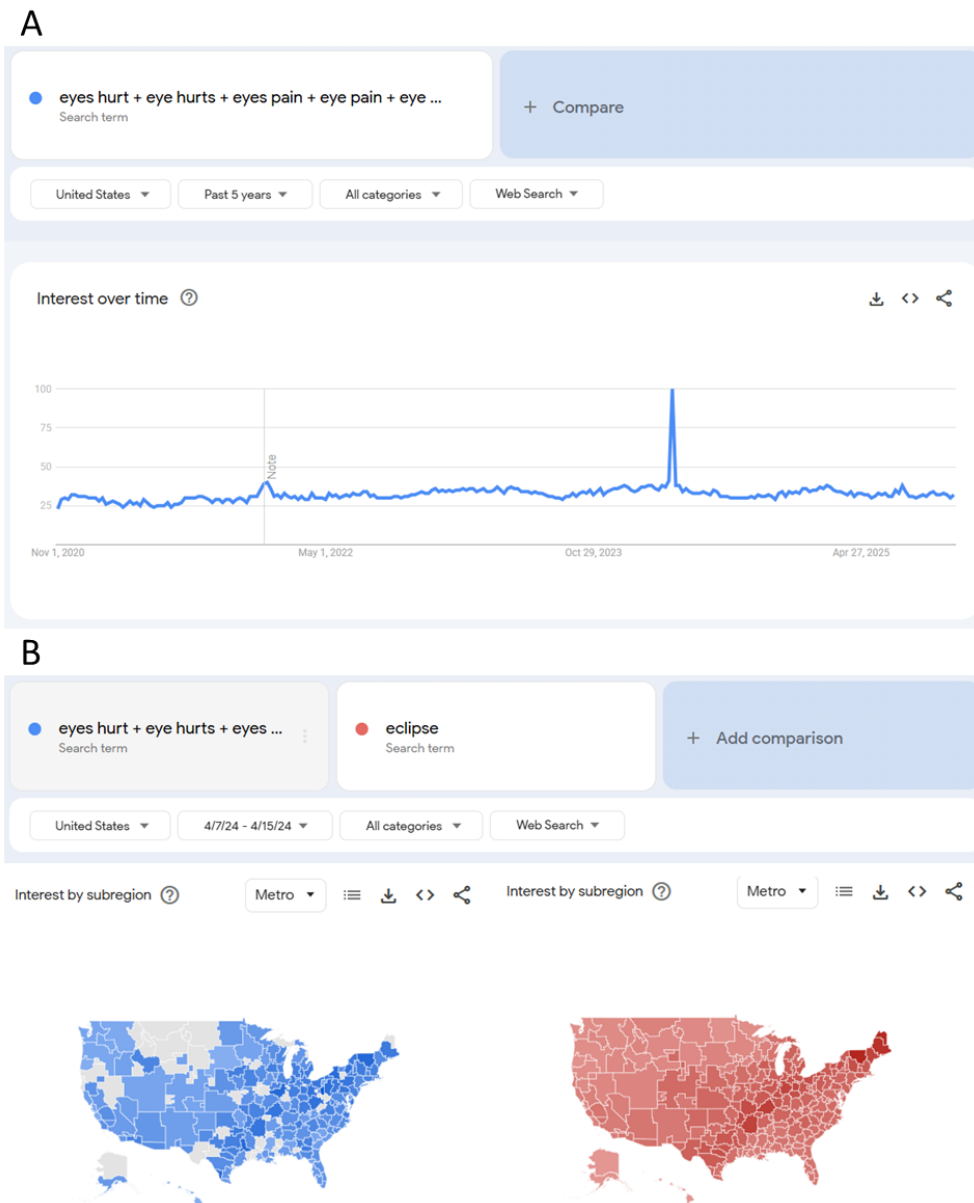


Figure 1. Examples of (A) interest-over-time and (B) interest-by-(sub)region on the Google Trends explorer. In Figure 1A, higher scores represent greater volume of searches related to eye

pain (*eyes hurt + eye hurts + eyes pain + eye pain + eye damage + eyes damage + eye symptoms*). The large spike in interest corresponds to the week of the solar eclipse on April 8, 2024, where many people injured their eyes by looking directly at the eclipse (Waisberg et al., 2024). Figure 1B shows metro-level interest in the same searches related to eye pain (blue) and searches including the term *eclipse* (red) during the week following the eclipse. Note the close correspondence between the eclipse path and search volume related to eye pain. Darker colors represent regions with more search volume. Source: Google Trends (www.google.com/trends).

Interest over Time

The first panel users will notice when querying Google Trends explorer is “interest over time”—it maps temporal variation in the popularity of a term. Figure 1A displays a query for searches containing searches related to eye pain across five years in the US. Note the large spike that corresponds to the solar eclipse on April 8, 2024.

The scores along the y-axis are calculated somewhat unintuitively—they are *scaled proportions* or *proportions of proportions*, not raw counts of searches (i.e., absolute search volumes). The timepoint scored as 100 in any trend is the timepoint in the query where *relative interest* was the highest—that is, the timepoint when the queried term was most likely to be searched for *as a proportion of all Google searches in that time period*. All other scores in a Google Trends query will be proportional to that score. Note that because the total raw number of Google searches varies across days, it is possible for one day to have a lower score for a given term than another, even if the raw number of searches for that term is higher on that day.

We flag two cautions with how Google Trends calculates scores, which we return to later:

(1) The lowest possible relative interest is represented as <1 , not 0. A timepoint with a zero

means that there were not enough *raw searches* for that term to pass a pre-specified privacy threshold in that timepoint (*FAQ about Google Trends Data*, 2024). (2) Scores are rounded to the nearest integer. When there is very large variation in scores, rounding can reduce precision—in the above example, if there was a day where *eyes hurt* searches were 1000 times more frequent than in the remaining 29 days, then those remaining days would all be scored as <1, losing all the variation therein.

Daily, Weekly, and Monthly Data. Users can specify different date ranges, and these yield different temporal granularities: daily data for queries of up to 270 days; weekly data for queries of up to five years; and monthly data for queries of wider time windows.

Real-Time Data. If users query the past hour or four hours, Google Trends provides by-the-minute data; past day queries yield 8-minute increments; past-week queries yield hourly increments. “Real-time” data, as these are called, come from separate cached samples of the Google database of searches than queries of older searches (*FAQ about Google Trends Data*, 2024). (We explain caching in the section “How Reliable Are Google Trends Data?”) Real-time data are not stored long-term and can only be accessed by querying the past hour/four hours/day/week. That is, it is not possible to specify a 7-day period in the past and obtain real-time data for it—such queries will return daily scores.

With real-time data, users must consider time-zones. Queries at these temporally granular scales are shown in the time zone of the user, rather than those of the region(s) explored. If a user in France wants to know what time Californians are searching for *Suicide (Topic)*, Google Trends will seem to show rises in (relative) popularity in late mornings, between 10-11am. However, when adjusting for the time difference between France and California, it becomes clear that such searches actually rise in popularity late at night.

Linking Interest-Over-Time Scores to Other Data. Researchers wishing to link interest-over-time scores to other datasets must match their temporal granularities. For example, datasets with monthly data should be matched to monthly Google Trends queries. The “sample size” for interest-over-time data is simply one per time point (e.g., for weekly data, there are 52 [or 53] time points per year).

Google Trends allows users to access up to five interest-over-time trends at once—these can be different terms in the same region or the same terms in different regions³ (or even different terms in different regions, if desired). Thus, a researcher who needs monthly data on suicide rates across states could retrieve trends for up to five states at once. On how to compare trends in more than five states at once, see the Creating Custom Datasets section.

Cross-Sectional Scores: Interest by Region

The Google Trends explorer also displays an “interest by [region/subregion/metro/city]” panel for each term in a query. The geographical scale (and label) of this panel varies depending on the selected region. If the broadest region—“worldwide”—is selected, this panel compares interest at the level of countries; for US queries, it compares all US states; UK queries compare the four member-countries; for France, the subregions will be the (pre-2016) regions of France, and so on.

Interest-by-region scores are scaled in the same way as interest-over-time scores, such that the region with the highest relative interest (i.e., the highest likelihood of the keyword or topic being searched *across the entire specified time period*) is scored as 100, and scores in other regions are proportional to it. Figure 1B displays interest-by-region (metro) for phrases including

³ To download multiple regions in the explorer, first create searches for multiple keywords; then, by clicking on the ellipses next to the keyword/topic, use the “Change filters” tab to change the region.

keywords related to eye pain (blue) and the keyword *eclipse* (red) in the US during the week following the eclipse. Note that regions where the number of searches does not exceed the privacy threshold are shown in gray.

Linking Interest-By-Region Scores to Other Data. Here, researchers again need to match the granularities of their datasets. The “sample size” of interest-by-region scores differ depending on the location set by the user. Setting the region to “Worldwide” will yield a single region score for each country in the dataset; setting the region to United States will yield a subregion score for each state (and the District of Columbia); setting the region to a single US state will yield a single score for each metro area. When searching for a single country, it is also possible to select more granular regions for that country. For example, users can query all metro areas in the US at once.

Cross-Sectional Comparisons of Multiple Terms: Compared Breakdown by (Sub)Region

Suppose you want to compare interest in music tastes across US states. Google Trends allows up to five terms at once. For example, one can compare the popularity, over a user-specified time period, of *Rock and roll*, *Hip-hop*, *Blues*, *Pop music* (topics) (see Figure S1). In addition to interest over time and interest by subregion for each genre, you will find a panel titled “compared breakdown by subregion”. Here, you can see the most popular genre in each state (with darker colors representing stronger relative interest), and, for each state, what percent of searches was for each genre (see Figure 2).

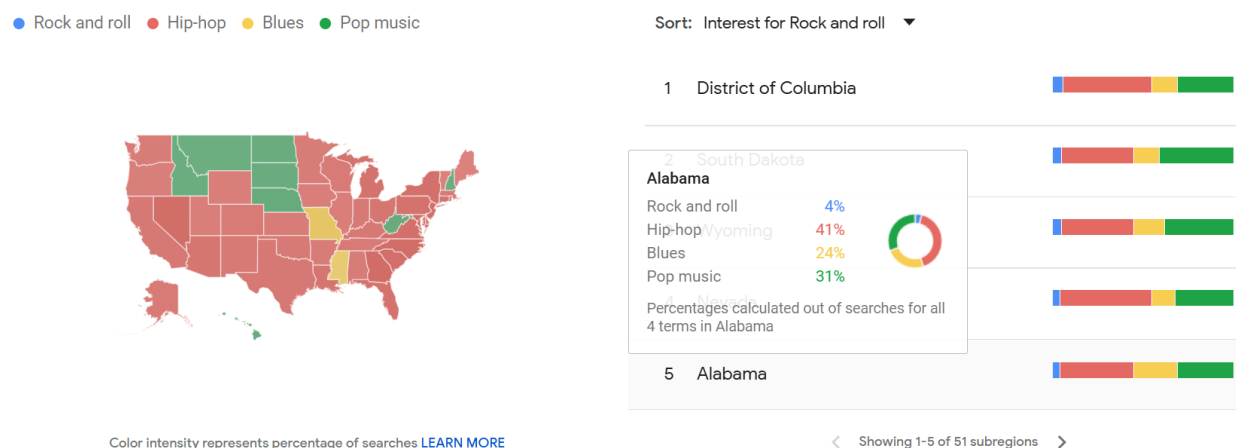


Figure 2. Example of the “compared breakdown by subregion” panel, from a query comparing interest in four different music genres (topics)—*Rock and roll*, *Hip-hop*, *Blues*, *Pop music*—in the US. Searches for the four genres add up to 100% within each state, as illustrated in the box with scores for Alabama. Source: Google Trends (www.google.com/trends).

Generating and Interpreting Data in Google Trends

In this section, we demonstrate how to access Google search data using Google Trends to specify which searches are included) and how to interpret these data. We also evaluate the quality of Google Trends data, showing when and why some queries produce lower quality data.

Operators in Searches

Keyword queries can be customized using operators (queries including *topics* cannot—see below). We describe them below and summarize in Table 3. For example, a soccer fan exploring interest in the FIFA World Cup might query *world cup*. A space between words represents an AND command and will return all Google searches that include both the words *world* and *cup*, in any order (including words in between *world* and *cup*). This soccer fan can restrict the order of the words and exclude words between *world* and *cup* using parentheses, by querying “*world cup*”.

Imagine, instead, someone more interested in cups than in soccer. The popularity of the sporting tournament might make it difficult to keep tabs on cup-related trends. To remedy this, Google Trends users can exclude words using the minus sign (or hyphen): the query *cup - world* will give this user the desired data. (Be sure to include spaces with this operator, or Google Trends will read it as a single hyphenated word.) As one would expect, *cup - world* does not show a spike around FIFA World Cup events, in contrast to *world cup*, *cup world*, and “*world cup*”.

Finally, for less selective queries, the + operator represent an OR command. Users interested in Google searches for either *world* or *cup* can query *world + cup* (this will include searches for the *FIFA World Cup*, *Disney World*, etc.). The + operator is especially useful for users needing to include synonyms (e.g., *bunny + rabbit*), plurals (e.g., *bunny + bunnies*), alternative spellings (e.g., *color + colour*), common misspellings (e.g., *calendar + calender*), or, in applicable languages, versions of words without special characters (e.g., the French word for weather is *météo*, but in France the easier-to-type *meteo*, without the accent marks, is slightly more popular; Google Trends differentiates the two versions).

Table 3. Operators for Use with Keywords.

Operator	Meaning	Example	Results
Space	AND	<i>world cup</i>	Returns all searches containing both <i>world</i> and <i>cup</i> , in any order
Quotation marks	Exact phrase	<i>“world cup”</i>	Returns only searches containing the phrase <i>world cup</i> ; excludes searches like <i>cup world</i> and <i>world football cup</i>
+	OR	<i>world + cup</i> (or <i>world+cup</i>)	Returns all searches containing <i>world</i> as well as all searches containing <i>cup</i>
–	NOT	<i>world - cup</i> (<u>not</u> <i>world-cup</i>)	Returns searches for <i>world</i> that do not include <i>cup</i>

Keywords vs. Topics

When a user types *weather* into the Google Trends explorer, it will make several suggestions, including *weather* (*Search term*) and *Weather* (*Topic*)⁴. The former, what we call keywords, are queries for specific words, exactly as specified by users (including operators)⁵. Keywords give the user complete control over which words are included in a query.

A *topic*, in contrast, is a composite of multiple Google searches related to the concept of interest. “Topics are generally considered to be more reliable for Google Trends data. They pull

⁴ Note that not all topics are listed with “Topic” in the Google Trends explorer; many have more descriptive labels such as “Book” or “16th President of the United States.” Anything that is not a “Search term” is a topic.

⁵ Queries are officially *not* case sensitive (Mavragani & Ochoa, 2019). However, we have been able to find instances where there are (usually minor) discrepancies between capitalized versus lowercase queries (e.g., “*WORLD CUP*” vs. “*world cup*” or *WORLD CUP* vs. *world cup*). These discrepancies seem to occur only in highly granular data (e.g., by-the-minute), especially when there are many zeros.

in the exact phrase as well as misspellings and acronyms, and cover all languages. This is more useful, particularly when looking at world data” (*Basics of Google Trends*, 2024).

The topics available correspond to the topic codes used in the Freebase database, which is now found in Wikidata (<https://www.wikidata.org/>)⁶. Google Trends does not publish exactly how searches are categorized into topics; some suggest that keywords are categorized into topics by considering the links people click on after making a Google search (Woloszko, 2020). For example, if someone searches for *apple* and then clicks on the corporation’s website, it could be inferred that the person intended to search for the technology company, rather than the fruit. Thus, this search might be categorized under the *Apple (Technology Company)* topic.

Topics have several benefits: (1) Topics allow users to disambiguate words that can refer to multiple constructs (e.g., *apple*). Fortunately, there are topics for both *Apple (Technology Company)* and *Apple (Fruit)*. (2) Topics pick up on synonyms, plurals, alternative spellings, and common misspellings. By looking at the *related keywords* panel in the Google Trends explorer, users can verify that topics are related to (and presumably include) relevant words—the topic *Rabbit* shows the related keywords *rabbit*, *bunny*, *rabbits*, and *bunnies*; *Color* shows the related keywords *color* and *colour*; *Calendar* includes the misspelled *calender*. (3) Topics are claimed to be language-independent, allowing comparison of constructs globally.

⁶ For example, searches related to sadness would be included in the topic coded `/m/02y_3dj` (the Freebase code can be found here: <https://www.wikidata.org/wiki/Q169251>). These Freebase codes are necessary when downloading trends for topics using R or Python packages (including Google Extended Trends for Health). Note that it is impossible to use operators with topics, even when using these codes—a query for the topics *Sadness* and *Psychological Stress* using their Freebase codes and with the `+` operator (i.e., `/m/02y_3dj + /m/0glpkn8`) does *not* return combined searches for the two topics.

Though these are impressive promises, topics often fall short of them—especially the promise of covering all languages. We review several important concerns with topics in Pitfall 4 and discuss this further in the Recommendations section.

Categories

Google Trends uses Natural Language Processing to organize terms into categories (Arora et al., 2019). On the Google Trends explorer, users can select from a list of 1200 categories (Woloszko, 2020), which follow a hierarchical nested structure. Users can query categories without specifying a keyword or topic (for an example, see Figure S2).

As with topics, categories can be useful for disambiguating keywords that refer to multiple constructs. For example, the keyword *bush* will return shrub-related searches, but also searches related to two former US presidents. When we query for *bush* in the US across the past five years, we get a largely cyclical pattern; when specifying the category “Politics”, however, we instead see one major spike in November 2020 (potentially because George W. Bush congratulated Joe Biden on winning the presidential election on November 8). Still, like topics, how categories are created is opaque, and it is unlikely that they perfectly filter out unrelated searches. We therefore urge caution when using categories.

Comparing Interest of Different Terms Over Time

Google Trends allows users to enter up to five terms per query. When downloaded alongside other terms, trendlines are rescaled relative to each other (so, except for the term with the highest relative interest, terms will be proportions of their original scores). Thus, whereas when downloaded separately, each trend will have a peak of 100, when downloaded concurrently, only the trend(s) with the highest relative interest will have a score of 100. An illustration of this rescaling is shown in Figure S3.

Does Google Trends Yield High Quality Data?

In this section, we first explore the implications of “caching” by estimating the reliability of different types of searches across retrieval days. We then discuss the implications of zeros in Google Trends results, and show using a simple simulation why researchers should avoid datasets with zeros.

How Reliable Are Google Trends Data?

Google logs billions of searches daily, and many trillions have been made in the years Google has existed (Cebrián & Domenech, 2024; *FAQ about Google Trends Data*, 2024; Rogers, 2016). Google Trends queries are not made on this entire database of searches—this would be very computationally expensive. Rather, each day, Google Trends pulls a random sample from this database⁷, and all queries that day are made on that sample. This sample is renewed (“cached”) each day, so queries made on different retrieval days use different samples.

If Google Trends data come from new cached samples each day, how consistent (statistically reliable) are queries across retrieval days? As Cebrián and Domenech (2024) note, previous sources have given conflicting answers: Some have suggested reliability to be high across downloads (D’Amuri & Marcucci, 2017; Stephens-Davidowitz & Varian, 2014), whereas others have reported more modest estimates (Cebrián & Domenech, 2023).

We tested whether three factors influence the reliability of Google Trends data: (a) relative search frequency (the very specific query of searches including the words *I’m sad* vs. the broader topic *Sadness [Topic]*), (b) population size (we retrieved data for the entire US and also data for all US states separately), and (c) granularity of time (daily vs weekly data). Each of

⁷ Google Trends excludes some searches, such as repeat searches by the same person within a short period of time (*FAQ about Google Trends Data*, 2024).

these factors can be thought of as contributing to the effective “sample size”—all else equal, there are fewer searches for unpopular (vs. popular) terms (cf. Cebrián & Domenech, 2024), in highly (vs. sparsely) populated regions (cf. Eichenauer et al., 2022), and in longer (vs. more specific) time periods. See Supplemental Materials for details on this analysis.

As shown in Figure 3, all three factors—population size, search frequency, and temporal granularity—systematically influenced data reliability across retrieval days: Queries for the specific keywords *I’m sad* yield less reliable data than the broader *Sadness (Topic)*; daily data yields less reliable data than weekly data; and queries restricted to small populations yield less reliable data than those using large populations.

Are Google Trends data reliable? Yes—as long as there is a large enough search volume. The best heuristic of high reliability in Google Trends data is the lack of zeros—datasets without any zeros had excellent reliability (average ICC = .95), whereas datasets with even up to 10% zeros had substantially poorer reliability (average ICC = .73). Datasets with over 50% zeros—such as many of the *I’m sad* queries in Figure 3—yield essentially useless trends (average ICC = .06). We discuss zeros next.

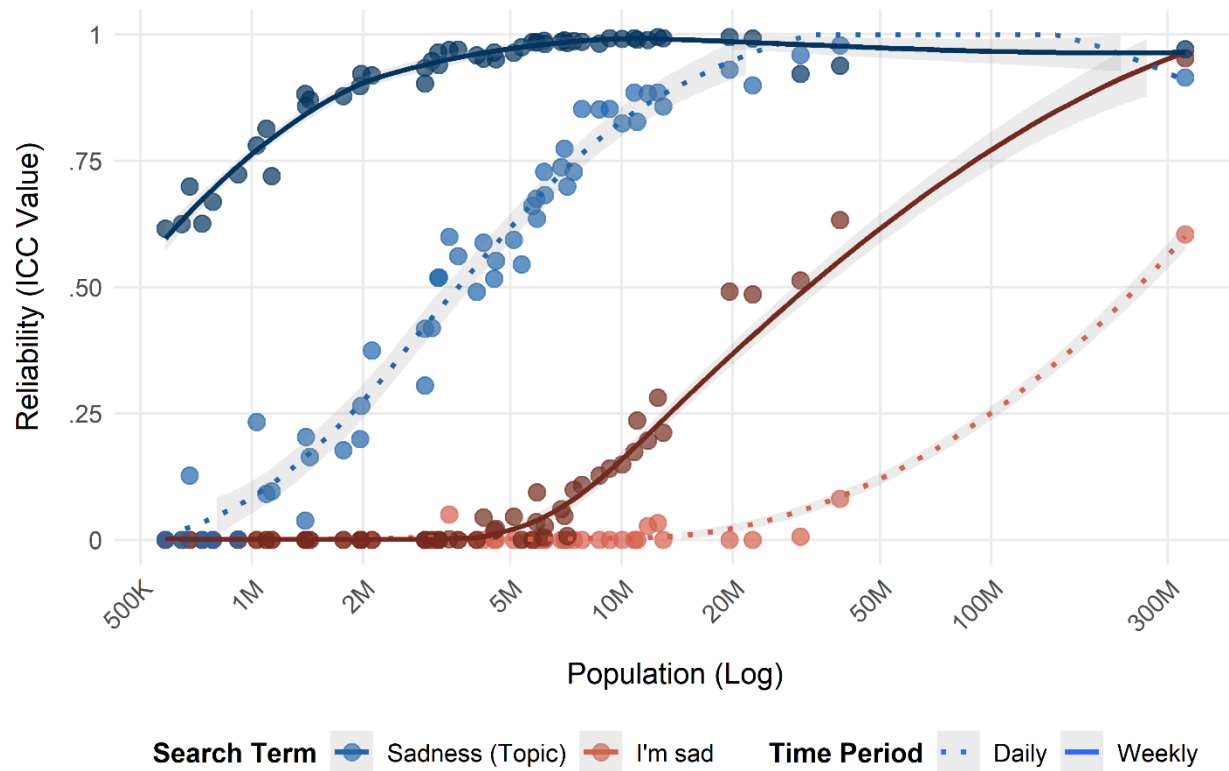


Figure 3. Reliability estimates (ICC) Google Trends samples across six retrieval days. On each day, we accessed data for a less common (*I'm sad*) and more common (*Sadness [Topic]*) search, at a more granular (daily) and less granular (weekly) time period, and for all US states separately as well as the entire US. Loess lines show trends across population size.

Zeros are Missing Data: Google Trend's Privacy Threshold

Why do zeros threaten reliability? At first blush, it might seem that datasets with many zeros might be especially reliable, as long as those zeros appear consistently across retrieval days. However, *zeros do not necessarily represent extremely low interest*. Zeros represent missing data—they are displayed when the raw number of searches for a keyword or topic are below an (undisclosed) privacy threshold set by Google Trends (Cebrián & Domenech, 2023; *FAQ about Google Trends Data*, 2024; Mavragani & Ochoa, 2019; Stephens-Davidowitz &

Varian, 2014). Trends with exceptionally low interest (but with enough search volume to exceed the privacy threshold) are displayed as “<1”.

To demonstrate how the privacy threshold functions, and its implications for data quality, we simulated one year of Google Trends data for two terms differing in popularity. (See Supplemental Materials for details on this simulation.) Results of this simulation are shown in Figure 4. These results demonstrate that, when the average search interest for a term hovers around the privacy threshold, scores for that trend are often unusable, with many timepoints coded as zeroes across retrieval days. Especially troubling is that zeros are not consistent across retrieval days. A given timepoint might be coded zero when accessed on one day, but have a different score—even a high score—when accessed on a different day.

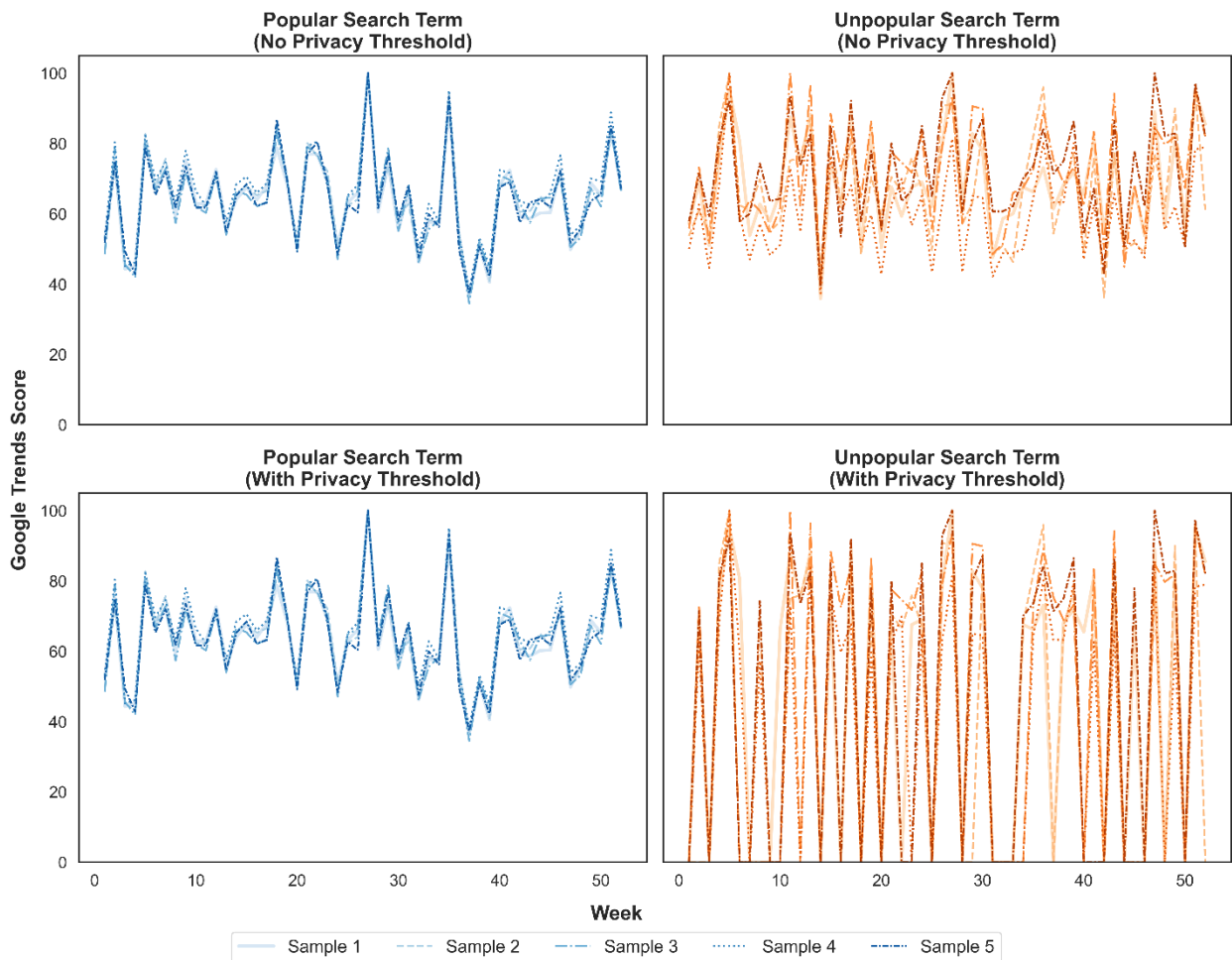


Figure 4. Results of a simulation to illustrate how the privacy threshold functions and why Google Trends datasets with zeroes are unreliable. The top panels show the “true” Google Trends scores as they would be calculated from samples of *Popular* (blue) and *Unpopular* (orange) searches. Each line represents a Google Trends score calculated from one of five different samples (as happens when users retrieve Google Trends data on different days). The bottom panels show results when a privacy threshold is imposed. When raw searches fall below this threshold, Google Trends returns a zero for that timepoint, rather than the true Google Trends score. As shown in the bottom right panel, scores can be coded as zero inconsistently across samples (i.e., when looking at data retrieved on different days), and they can be coded as zero even if the true score is not much lower than the scores on other days.

The findings of this simulation are not just a theoretical possibility—Google Trends results often resemble the unusable *Unpopular* trend (bottom-right panel) from Figure 4. For example, Franzén (2023) accessed the same Google Trends query across days, reporting data that look much like this panel, and concluded that researchers should refrain from using Google Trends altogether.

In reality, Franzén’s query—for a relatively uncommon search—simply does not have a sufficiently high search volume to consistently exceed the privacy threshold (cf. Raubenheimer, 2024). Each day Franzén repeated this query, different time points failed to exceed the privacy threshold, yielding zeros. Because the maximum peak interest score is always coded as 100, this caused the trend to look like it has massive fluctuations when, in reality, the variation in interest might have been quite small.

We echo the advice of Stephens-Davidowitz and Varian (2014) to exercise caution with datasets that include zeros. We return to this issue in the Recommendations section.

Creating Custom Datasets

Here, we describe how to create custom datasets that extend beyond the datasets Google Trends supplies by default. The three types of custom datasets we describe all rely on the same principle: in a given trend, datapoints are scaled proportional to the most popular datapoint, so it is possible to extend default datasets by making multiple overlapping queries and harmonizing them.

We cover three types of custom datasets (see Table 4): Longitudinal (extending interest-over-time, thereby allowing for higher temporal granularities), cross-sectional (comparing regional scores across more regions than allowed by default), and panel data (synchronizing many longitudinal trends to be on the same scale). We illustrate each with an example presented with accompanying R code.

Recall that Google Trends scores are rounded to the nearest integer. This has implications for users creating custom datasets. Imagine a query that yields five scores: 100, 50, 49, 2, and 1. Although rounding is less consequential for the larger scores 50 and 49 (i.e., which could be as different as 50.4 and 48.6 or as similar as 49.6 and 49.4), it substantially reduces precision for the smaller 1 and 2—the difference between their real values can range from large (0.6 vs. 2.4—the latter representing three times the relative interest of the former) to relatively small (1.4 vs. 1.6—where relative interest is nearly identical).

Table 4. Three Types of Custom Datasets

Type	Goal	Type of Data Used	Example
Longitudinal	Extend a single daily trend beyond 270 days, or a weekly trend beyond 5 years	Interest over time (over multiple time periods)	Weekly data for <i>Depression (Topic)</i> in Egypt across 13 years
Cross-Sectional	Compare cross-sectional interest for more than 5 keywords/topics	Breakdown by (sub)region (for many terms)	Interest in 17 religious denominations in each US state
Panel	Synchronize more than 5 longitudinal trends to be on the same scale	Interest over time (for many locations)	Longitudinal scores for <i>Abortion (Topic)</i> for all 50 US states

Longitudinal Datasets: Interest over Time at High Temporal Granularity

Google Trends provides daily data for interest-over-time queries of up to 270 days and weekly data for queries of up to five years. However, by knowing how trends are scaled, it is possible to create custom datasets that go beyond these limits. To demonstrate this, we created a trend of weekly data for *Depression (Topic)* in Egypt from 2011 to 2023. We downloaded three 5-year datasets of interest in *Depression* in Egypt, each overlapping with the next by one year (Trend 1 = 2011-2015; Trend 2 = 2015-2019; Trend 3 = 2019-2023). If Google Trends data were not rounded to the nearest integer, then it would be sufficient to have only a single overlapping week. We use a full year to obtain a better estimate of the ratio between the trends.

We first load these separate trends into R, calling them `eg1`, `eg2`, and `eg3`, and ensure that the `Week` column is formatted as a date, then rename the columns to make them easier to

work with. (We show the code for eg1 only; for eg2 and eg3 simply run the same code, replacing the numbers.) Note that the “skip = 2” syntax is necessary when using files downloaded straight from Google Trends, as the first two rows are blank.

```
eg1 <- read.csv("EG1.csv", skip = 2)
eg1$Week <- as.Date(eg1$Week)
names(eg1) <- c("Week", "Depression1")
```

We can visualize the three trends using the `ggplot2` package (Wickham, 2016). This code will plot each of the lines (i.e., Trends 1-3) across time in different colors. As shown in Figure 5A, Trends 2 and 3 share a peak (on 2019-12-01) so they are already on the same scale (and thus perfectly overlap on the year they share). However, Trend 1 is on a different scale. Compare Trends 1 and 2 on the year they share—they rise and fall together (i.e., they are perfectly correlated—or would be without rounding), but Trend 1 is shifted upwards compared to Trend 2.

```
library(ggplot2)

p1 <- ggplot() +

  geom_line(data = eg1, aes(x = Week, y = Depression1), color =
"lightblue", size = .75, alpha = 0.8) +

  geom_line(data = eg2, aes(x = Week, y = Depression2), color =
"lightgreen", size = .75, alpha = 0.8) +

  geom_line(data = eg3, aes(x = Week, y = Depression3), color =
"lightpink", size = .75, alpha = 0.8) +

  ggtitle("Depression (Topic) in Egypt (2011-2023)") +

  ylab("Google Trends Score") +

  ylim(20,100) +
```

```
scale_x_date(date_breaks = "2 year", date_labels = "%Y") +
theme_minimal()
```

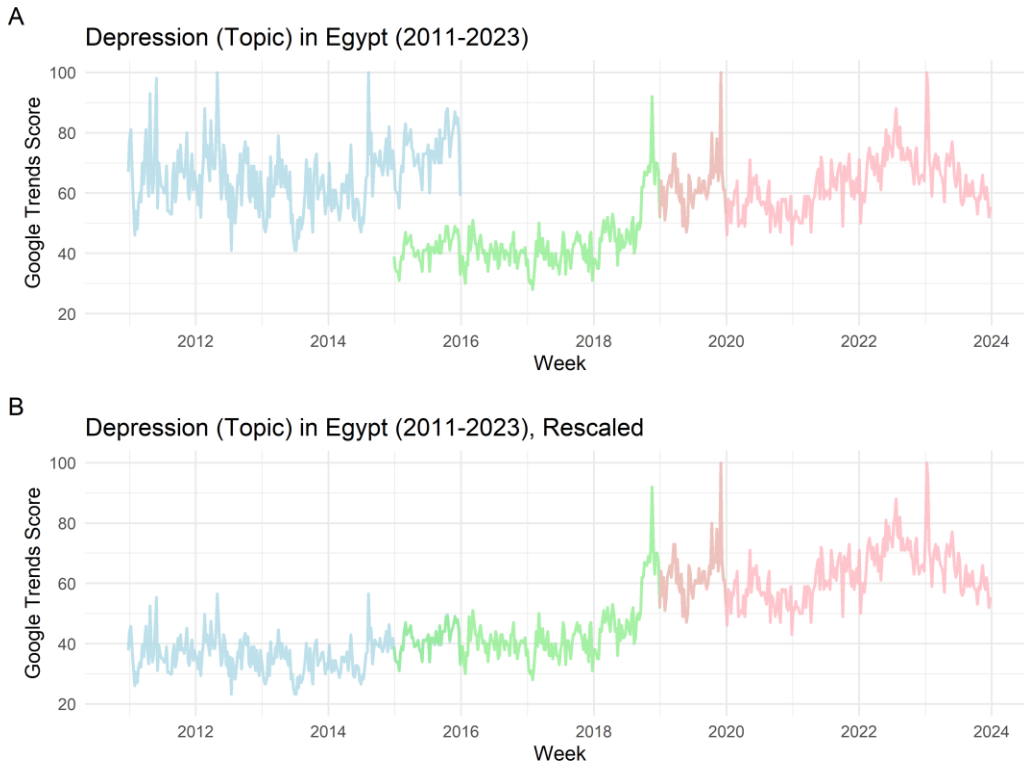


Figure 5. Google searches for *Depression (Topic)* in Egypt, 2011-2023. We downloaded three 5-year datasets (yielding weekly data), each overlapping with the next by one year. Trend 1: 2011-2015, Trend 2: 2015-2019, Trend 3: 2019-2023. Panel A shows the three trends unscaled. Panel B shows the three trends with Trend 1 adjusted to be on the same scale as Trends 2 and 3.

Because Trends 1 and 2 are proportional to each other, we can place them onto the same scale if we compute the correct scaling factor. To compute this proportion, we first create a dataframe that contains only the overlapping periods: The dataframe “overlap_1_2” includes only the year 2015, and the dataframe “overlap_2_3” includes only 2019.

```
overlap_1_2 <- merge(eg1, eg2, by = "Week")
```

```
overlap_2_3 <- merge(eg2, eg3, by = "Week")
```

Next, we divide the trends by each other (which will yield data only for the weeks with values in the overlapping years) to obtain a scaling factor for each week. We then estimate the scaling factor using the code below. Rather than taking the average ratio for all weeks, we compute a *weighted average*, which gives more weight to higher numbers. A simulation (provided on our OSF page) shows that this generates more accurate estimates of the true ratio between trends⁸.

```
ratios_1_to_2 <- overlap_1_2$Depression2 /
overlap_1_2$Depression1
```

```
ratios_2_to_3 <- overlap_2_3$Depression3 /
overlap_2_3$Depression2
```

```
scaling_factor_1_to_2 <- sum(ratios_1_to_2 *
overlap_1_2$Depression2) / sum(overlap_1_2$Depression2)
scaling_factor_2_to_3 <- sum(ratios_2_to_3 *
overlap_2_3$Depression3) / sum(overlap_2_3$Depression3)
```

Using this code reveals that the scaling factor between Trends 1 and 2 is approximately 0.56—that is, Trend 1 needs to be multiplied by 0.56 to be on the same scale as Trend 2. The scaling factor for Trends 2 and 3 is 1, because these two trends were already on the same scale

⁸ Small numbers are affected more by rounding (e.g., rounding 1.6 to 2 is more consequential than rounding 96.6 to 97), so large numbers provide more precise estimates of the true ratio. We conducted simulations (where we know the true ratio) to test which method (average vs. weighted average) estimates the true ratio better (see the OSF page). The weighted average ratio typically does better, especially when there is a large discrepancy in magnitude between trends (e.g., a ratio of 0.25).

(the overlapping timepoints between these two trends included the highest score). We can rescale Trend 1 by multiplying the scores in `eg1` by the scaling factor:

```
eg1$Depression.Rescaled <- eg1$Depression1 *
scaling_factor_1_to_2
```

By multiplying Trend 1 scores by this constant, we shift them downwards to the same scale as Trend 2 (and Trend 3). All three trends now match (except for minor error due to rounding), though they remain in separate columns. We have named the new version of Trend 1 `Depression1.Rescaled`. We can plot the three trends with the re-scaled version of Trend 1 (i.e., simply switching `Depression1` with `Depression1.Rescaled`). See Figure 5B.

To create a single column with our new rescaled trend, we merge the three dataframes, then take the average of the three trends, ignoring NA values. This means that the overlapping periods are the mean of the two trends—that is, 2015 scores are the average of Trend 2 and the rescaled Trend 1. As these two trends differ only due to rounding, these changes are minimal.

First, we merge the datasets by the `Week` column (beginning by merging `eg1` and `eg2` then adding `eg3`):

```
merged_df <- merge(eg1[, c("Week", "Depression1.Rescaled")],
                  eg2[, c("Week", "Depression2")],
                  by = "Week", all = TRUE)
merged_df <- merge(merged_df,
                  eg3[, c("Week", "Depression3")],
                  by = "Week", all = TRUE)
```

We then calculate the average, ignoring NAs:

```
merged_df$Depression_combined <- rowMeans(
```

```
merged_df[, c("Depression1.Rescaled", "Depression2",
"Depression3")], na.rm = TRUE)
```

This code creates, in the dataframe `merged_df`, a column called `Depression_combined`, which has weekly data for *Depression (Topic)* in Egypt from 2011 to 2023.

The same principles can be used to extend other longitudinal trends, such as daily datasets, or sets of more than three weekly datasets. As long as researchers include overlapping periods to estimate the scaling factor, they can use these scaling techniques to create their desired datasets.

Extending Cross-Sectional Comparisons: Breakdown by (Sub)Region

Google Trends allows comparisons of up to five different topics/keywords in a single query. The same principles described above—download multiple and partially overlapping queries, and adjust to put them on the same scale—can be used to bypass this limit.

One approach is to use the average score for each term across the user-specified time periods (these values appear at the left side of the interest over time panel). Holmes and colleagues (2022) used this approach to compare the popularity of 100 animal species (all topics). First, they found the most popular species through the process of comparison and elimination; in their list of species this was *Lion*. Then, they included *Lion* in every query (so, four unique species plus *Lion*), meaning, the scores for all 100 species in their dataset were already set to the same scale, proportional to *Lion*. Finally, for each query, they saved the average score for each species, which is a rough estimate of how much interest that species garnered over time. We say “rough estimate” because the time points in a date range do not all have the same total search volume, but taking an average gives equal weight to each time point. (As an analogy,

this is akin to calculating the average income in the US by averaging the average income across all US states, thereby giving them all equal weights notwithstanding differences in their population sizes.)

A more precise approach is to use breakdown by region scores. As shown earlier in Figure 2, breakdown by region provides percentage scores for each subregion (adding up to 100, rather than the highest interest score being scaled as 100). Because these scores are percentages, rather than the typical Google Trends scores, they require a different rescaling method.

We illustrate this approach by comparing, across US states, interest in 17 different religious denominations. We queried Google Trends multiple times, in sets of five, for interest in all of 17 denominations. This required 4 separate queries. We included one denomination (LDS/Mormon) in every query to ensure all denominations are on the same scale. With the “breakdown by region” data, we obtain percentage scores for each state (adding up to 100) for the five denominations in each query.

The logic of the rescaling process is as follows: Suppose you have two queries (so, each with a set of four unique denominations plus LDS). In a given state, for the first set of five, LDS made up 60% of searches, and Catholic made up 30% of searches (with the other three denominations making up the remaining 10% of searches); for the second query, LDS searches made up 40%, and Baptist searches made up 20% in that state (again, with the other denominations making up the remaining searches). Remember that the number of searches for LDS in this state is equal in both queries (as long as the timeframe is the same and the data are accessed on the same day). Because Catholic searches were half as common as LDS searches (60% / 30%), and Baptist searches were also half as common as LDS searches (40% / 20%), we can conclude that Catholic and Baptist searches were equally common in this state, even though

they were not downloaded in the same query. By putting them on an equal scale in this way, we can compare relative popularity of all selected denominations within each state.

To conduct this rescaling, we first read in the four separate datasets using the below code.

(Repeat for sets 2 through 4.)

```
religion_set1 <- read_csv("Religion 1.csv", skip = 2)
```

To make them easier to work with, we can also change the names of the columns. For example, we use LDS instead of the default heading, 'The Church of Jesus Christ of Latter-day Saints: (4/28/20 - 4/28/25)':

```
names(religion_set1)<-c("Region", "LDS", "Atheism", "Baptist",  
"Buddhist", "Catholic")
```

```
names(religion_set2)<-c("Region", "LDS", "Eastern Orthodox",  
"Episcopal", "Evangelical", "Hinduism")
```

```
names(religion_set3)<-c("Region", "LDS", "Islam", "JW",  
"Lutheranism", "Judaism")
```

```
names(religion_set4)<-c("Region", "LDS", "Methodism",  
"Presbyterianism", "Seventh-Day Adventism", "Unitarian  
Universalism")
```

For each of these datasets, the set of five religions in that dataset add up to 100 within each state. To put each dataset on the same scale, we need to compare all religions to the same benchmark—in this case, percent interest in the LDS Church. By multiplying each row so that interest in the LDS Church is 100, we place all denominations on the same scale across the four datasets. Table 5 shows what this process looks like in the first set, before and after rescaling (showing only 6 of the 50 states).

Table 5. Sample of Cross-Sectional Interest in Different Religious Denominations Before (Left) and After (Right) Scaling.

Region	Raw Data					Data After Rescaling				
	LDS	Atheism	Baptist	Buddhist	Catholic	LDS	Atheism	Baptist	Buddhist	Catholic
Utah	94%	<1%	<1%	2%	4%	100	0.00	0.00	2.13	4.26
Idaho	84%	1%	1%	3%	11%	100	1.19	1.19	3.57	13.10
Wyoming	63%	1%	3%	5%	28%	100	1.59	4.76	7.94	44.44
Arizona	47%	3%	3%	10%	37%	100	6.38	6.38	21.28	78.72
Hawaii	43%	2%	2%	20%	33%	100	4.65	4.65	46.51	76.74
Montana	45%	2%	3%	10%	40%	100	4.44	6.67	22.22	88.89

Note: This table comes from one query of interest in five religious denominations. Only six states in the dataset are shown for demonstration. Scores of “<1” were rounded to zero to be included in the analyses.

We achieved this rescaling using the below code. This code (a) changes all instances of “<1” to zero, (b) removes “%” and converts all values to numeric, and (c) divides each row by the value of the LDS column, then multiplies by 100, thus making all values proportional to an LDS score of 100 in each state. This code rescales the first dataset; for the other datasets, simply substitute in the name of the dataset (e.g., change `religion_set1` to `religion_set2`).

```
religion_set1 <- religion_set1 %>%
  mutate(across(-Region, ~ gsub("<1", "0", .))) %>%
  mutate(across(-Region, ~ as.numeric(gsub("%", "", .)))) %>%
  mutate(across(-Region, ~ . / LDS * 100))
```

After rescaling all four datasets with the above code, we merge them:

```
religion_merged <- full_join(religion_set1, religion_set2, by =
"Region")
religion_merged <- full_join(religion_merged, religion_set3, by
= "Region")
```

```
religion_merged <- full_join(religion_merged, religion_set4, by
= "Region")
```

We can now remove the duplicate columns (because there was an LDS column in each dataset, and the value for every row in each of these columns is 100, all of these columns except one are redundant):

```
lds_columns <- grep("LDS", names(religion_merged), fixed = TRUE)
religion_merged <- religion_merged[, -lds_columns[-1]]
```

Finally, because of our scaling process, the data no longer represent percentages. We can convert them back to percentages by simply ensuring that the sum of each column is 100. We do this with the following code, which ensures that the sum of all numeric columns is 100:

```
religion_percentages <- religion_merged %>%
  rowwise() %>%
  mutate(across(-Region, ~ . / sum(c_across(-Region), na.rm =
TRUE) * 100)) %>%
  ungroup()
```

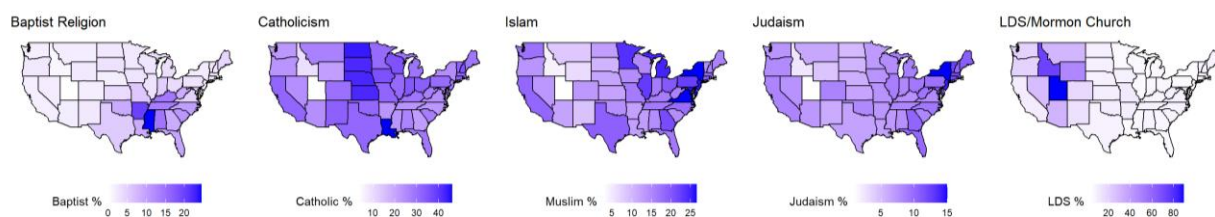


Figure 6. The relative popularity (in percentages) of five of the 17 religious denominations. Each map represents, for each state, the percentage of searches for the given denomination out of all 17 denominations. Darker colors reflect more interest in the denomination. Note that different denominations have different peak interests (and scales differ on the legends).

Figure 6 visualizes the relative interest in five religious denominations—Baptist Religion, Catholicism, Islam, Judaism, LDS/Mormon Church. These numbers can be interpreted as percentages of all included religion searches that were for each specific denomination. For example, the score of 41 for the Catholic Church in Rhode Island means that, out of all searches for all 17 denominations, 41% of the searches in that state were for the Catholic Church.

The resulting dataset reflects interest in different religious affiliations across states, and in some cases—as highlighted in Figure 6—also nicely tracks the distribution of actual religious affiliations. See Adamczyk et al. (2022), for a more sophisticated operationalization of religious activity using specific searches such as for places of worship.

Cross-Sectional Data at Different Levels

The above example used breakdown by subregion scores in the US to obtain state-level data. Users can use the same approach to compare subregions in whichever geographic location they choose. If users wish to compare such data at the country level, they can set the location to Worldwide (so that the regions are countries). Because percentages add up to 100 within each region (i.e., they are not scaled to other regions), users can simply use data from the regions of interest. As subregions are required, it is impossible to obtain such percentage data at the worldwide level.

Panel Data: Harmonizing Interest Over Time Across Many Locations

Panel datasets include longitudinal data across multiple cases (in this case, regions). Users can create panel datasets with Google Trends using the principles of re-scaling presented above.

We demonstrate by creating a panel dataset with trends for *Abortion (Topic)* across all 50 US states. This required 13 queries, each with Arizona plus four additional states. By including

Arizona in each query, and by scaling all results within each query such that the maximum score for Arizona is 100, we are able to create a dataset with a timeseries for each state, all on the same scale (i.e., the Google Trends scores are comparable across states).

To do so, we first load in all 13 datasets using the below code (we provide code for the first dataset only; the code for the others differs only in changing `Panel1` to `Panel2`, etc.) This code (a) loads the datasets, then (b) cleans up the column names so that only the state name remains (in the raw file, column names list the topic and state, such as `Abortion:`

```
(Arizona)).
```

```
library(readr)
```

```
library(tidyverse)
```

```
panel1 <- read_csv("Panel1.csv", skip = 2, show_col_types =  
FALSE)
```

```
names(panel1) <- gsub("Abortion: \\((.+?)\\)", "\\1",  
names(panel1))
```

We now have 13 datasets, each with a column called `Week` and five columns whose names represent states. Because we included Arizona in every set of five, each dataset includes Arizona. To ensure that all 13 datasets are on the same scale, we need to multiply all scores in the dataset by a factor that ensures that the highest score for Arizona is 100. In datasets where the peak interest in Abortion (i.e., the week with the highest relative interest score out of all five states) is already in Arizona, scores will not change at all.

We can do this with the following code, which multiplies all columns (except `Week`) by 100, then divides them by the maximum score for `Arizona`. Thus, if Arizona already has a maximum score of 100, all scores will remain the same; if Arizona had a maximum score of 50,

then all scores would be doubled. Again, we show code for the first of the 13 datasets—the only required change is `panel1` to `panel2`, etc.

```
library(dplyr)

panel1_scaled <- panel1 %>%
  mutate(across(-Week, ~ . * 100 / max(Arizona, na.rm = TRUE)))
```

Next, we can merge all the dataframes, including only one of the Arizona columns. The below code does this by (a) creating a dataframe called `all_panels`, which initially is simply a copy of `panel1_scaled`; (b) one at a time, it adds columns from subsequent dataframes, but does not add the columns `Week` and `Arizona`, as these columns are already present in `panel1_scaled`. We again provide the first lines of code; subsequent lines need only to change the name of the dataframe (e.g., `panel3_scaled` to `panel4_scaled`, etc.)

```
all_panels <- panel1_scaled

all_panels <- cbind(all_panels, panel2_scaled[,
!names(panel2_scaled) %in% c("Week", "Arizona")])

all_panels <- cbind(all_panels, panel3_scaled[,
!names(panel3_scaled) %in% c("Week", "Arizona")])
```

We now have all of the scores on the same scale—except they are not on the original Google Trends scale, where the maximum value is 100. To convert scores back to the original scale, we can use this code, which finds the global maximum, computes the scaling factor to convert the maximum to 100, then multiplies all values by that scaling factor:

```
global_max <- max(as.matrix(all_panels[, -1]), na.rm = TRUE)

all_panels_GTscale <- all_panels

scaling_factor <- 100 / global_max
```

```
all_panels_GTscale[, -1] <- all_panels_GTscale[, -1] *
scaling_factor
```

We now have a panel dataset with longitudinal scores for each US state's interest in *Abortion*, and these scores are on the original Google Trends scale. The resulting data are visualized as a heatmap in Figure 7. We have also marked several key events, and spikes in interest in abortion can be seen in response to these events: A spike in interest in Alabama during May 2019 corresponds to a bill banning abortion in Alabama. A spike across all states in May 2022, then an even larger spike in August 2022. The former corresponds to the *Dobbs vs. Jackson Women's Health Organization* case, which overturned the constitutional right to abortion. The decision, which was made official in June 2022, was leaked in May 2022. The latter (which was also the largest—it was scored as 100) was in Kansas in August 2022, and corresponds to a Kansas referendum on abortion legality.

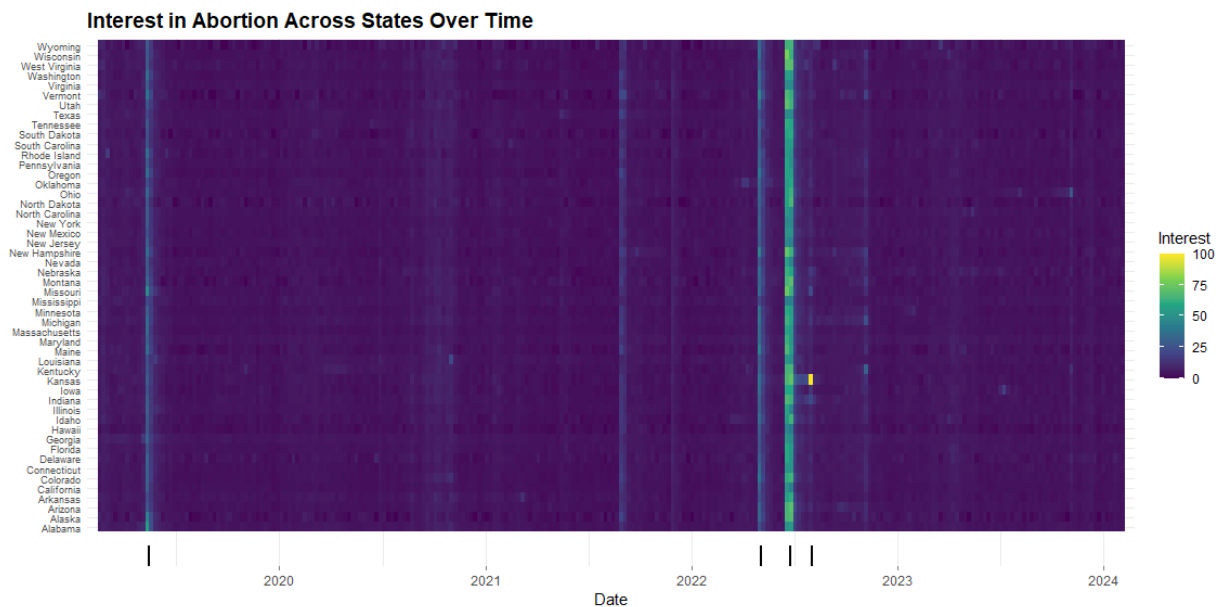


Figure 7. Interest in *Abortion (Topic)* in all 50 US states across time, displayed as a heatmap.

Tick marks along the x-axis denote key events related to abortion laws in the US.

It is worth noting that, if Arizona had the highest peak of all states (i.e., had the score of 100 been the same datapoint in every set of queries), or if we had instead included Kansas in every query, then no rescaling would have been necessary—all queries would have been scaled relative to the peak interest. Researchers who prefer not to rescale could first ascertain which region has the highest peak interest score, and simply include that region in every query.

Study Examples

In this section, we walk through a full analysis that could be done by linking Google Trends data to external datasets. We provide three examples: one in the main text and two in the Supplemental Materials. These examples were chosen to demonstrate how different types of hypotheses can be tested using Google Trends data. Example 1 tests the effect of a recurrent yearly event on Google search patterns. Example 2 estimates the relationship between a monthly trend of Google Trends data with monthly data for unemployment in U.S. states. Example 3 estimates the effect of non-recurring events on Google search patterns.

In all three examples, we use the `feols()` function via the `fixest` package in R. This is more appropriate for longitudinal data than standard `lm()` models, which assume that observations are independent. With longitudinal Google Trends data, there is likely temporal autocorrelation (i.e., observations within the same time series tend to be correlated). The `feols()` function makes it straightforward to apply clustered standard errors, which produces more accurate standard errors (Cameron & Miller, 2015). Additionally, we can easily incorporate fixed effects for time to control for seasonal patterns in search behavior.

Example 1: The Effects of Recurring Events

One possible use of Google Trends is to estimate the effects of events, whether unique or recurring. In ongoing work, we are investigating the effects of religious fasts on well-being

(Moon et al., 2025). Following Foa et al. (2022), we created composite measures of mental health and well-being using topics related to negative emotions (such that lower search volume for negative emotions indicates greater well-being). We demonstrate using a small subset of this analysis, examining whether searches related to *Depression (Topic)* are lower during Ramadan fasting periods in Egypt.

Above, we created a 13-year trend of weekly Google Trends data for *Depression* in Egypt. Egypt has a strong Muslim majority (about 90%), so it can be assumed that most of its inhabitants fast during the Islamic month of Ramadan. We can code which weeks of the year include Ramadan, and test whether depression-related searches decline in Egypt during those weeks.

First, we create a vector with weeks coded as Ramadan. We selected weeks with at least four days of Ramadan (i.e., majority-Ramadan weeks):

```
ramadan_weeks <- as.Date(c("2011-07-31", "2011-08-07", "2011-08-14", "2011-08-21", "2012-07-22", "2012-07-29", "2012-08-05", "2012-08-12", "2013-07-07", "2013-07-14", "2013-07-21", "2013-07-28", "2014-06-29", "2014-07-06", "2014-07-13", "2014-07-20", "2015-06-14", "2015-06-21", "2015-06-28", "2015-07-05", "2015-07-12", "2016-06-05", "2016-06-12", "2016-06-19", "2016-06-26", "2017-05-28", "2017-06-04", "2017-06-11", "2017-06-18", "2018-05-13", "2018-05-20", "2018-05-27", "2018-06-03", "2018-06-10", "2019-05-05", "2019-05-12", "2019-05-19", "2019-05-26", "2020-04-26", "2020-05-03", "2020-05-10", "2020-05-17", "2021-04-11", "2021-04-18", "2021-04-25", "2021-05-02", "2022-04-03", "2022-
```

```
04-10", "2022-04-17", "2022-04-24", "2023-03-26", "2023-04-02",
"2023-04-09", "2023-04-16"))
```

Next, we create a dummy-coded Ramadan variable for the 13-year weekly trend we created earlier (`final_df`), coded as 1 during Ramadan weeks and 0 otherwise:

```
final_df$Ramadan<-ifelse(final_df$Week %in% ramadan_weeks, 1, 0)
```

Next, we create variables indicating the year and month (to control for year-by-year and seasonal fluctuations in depression-related searches):

```
final_df$Year <- format(final_df$Week, "%Y")
final_df$month <- format(final_df$Week, "%m")
```

Finally, we run a fixed effects model, using the `fixest` package. This fixed effects regression controls for year-to-year differences in searches and accounts for normal seasonal variation in searches (by controlling for month). We also cluster standard errors by year, as weekly observations within the same year are likely correlated:

```
library(fixest)
fixed_effect_regression <- feols(Depression.Complete ~ Ramadan |
year + month, vcov = ~year, data = final_df)
```

We find that depression-related searches are significantly lower during Ramadan (vs. non-Ramadan periods) in Egypt, $b = -4.53$, $SE = 1.08$, $t = -4.19$, $p = .001$.

This basic analysis can be scaled up. In the full study (Moon et al., 2025), we included all countries that (a) had sufficient data (i.e., datasets without zeros—see Recommendations) and (b) have populations that are at least 75% Muslim. We also use a wider range of topics (*Depression*, *Sadness*, *Antidepressant*, *Psychological Stress*, *Apathy*) to create composite measures of well-being and to test robustness of the effects.

Potential Pitfalls in Google Trends Data

As with any tool, researchers need to understand how to use Google Trends responsibly. We overview common pitfalls researchers should be aware of.

Pitfall 1: Data Quality Is Not Guaranteed

Google's search algorithms are constantly being updated to improve service to clients and users, and these changes likely influence Google Trends data in ways unpredictable to users (Lazer et al., 2014). Some major changes that might influence Google Trends data are marked in the Google Trends explorer (e.g., at the start of 2011, 2016, 2017, and 2022). However, the specifics of these changes are not publicized.

This means that the data quality can change over time—some changes might improve the quality of data along the availability within the range of dates offered by Google Trends (e.g., the data seem to have improved after some of the above algorithm shifts). Some changes might also affect prior data. For example, Algan et al. (2019) found large jumps in several searches (e.g., *pregnancy*) at the beginning of 2011 (following a major, marked algorithm change), however, when we tried to replicate those query, we no longer saw those jumps, suggesting that more recent algorithm changes have altered previous data.

Additionally, we have occasionally noticed patterns that seemed highly implausible. For example, on November 6, 2023, we noticed that, on October 30, 2023, there were worldwide anomalies on nearly every term we examined—some spiked, whereas others dipped to near <1 (see Figure S4). When we repeated these queries on subsequent days, these anomalies disappeared, suggesting that they represented a quirk in Google Trends, rather than drastic and hard-to-explain shifts in worldwide internet search behavior. Our advice for researchers noticing such aberrant patterns is to check whether they disappear when the data are accessed on subsequent days, and, if they do not, to avoid using data with such patterns.

In sum, researchers using Google Trends cannot audit the specific ways in which Google Trends data are collected or aggregated. Google Trends was not created specifically for researchers and does not make guarantees about data quality. Though we suspect and hope that the quality of Google Trends data will improve over time, this is not a certainty.

Pitfall 2: Google Search Data Are Highly Representative, But Not Fully

In many countries, an examination of Google search data seems to suggest that interest in science has declined over the last 20 years. However, as Stephens-Davidowitz and Varian (2014) suggest, an alternative possibility is that the demographics of Google users—in the US and likely elsewhere—has changed: early adopters tended to be more computer-literate, more interested in science and technology than the average person (Mellon, 2013). Thus, it might not be that general interest in science declined, but that people who are less interested in science have gained more access to the internet. (It is also the case that the menu of offerings available on the internet has increased over time.)

Although nowadays nearly the entire US population uses Google, this is not the case for developing countries (or countries where Google is not the search engine of choice, such as China and Russia where Baidu and Yanex, respectively, are more popular). Thus, researchers should be aware that Google search data may not be representative of the entire world population, both across time and geographical regions.

Another reason Google search data may not be representative stems from what people use Google for. There may be different rates of technology use by cultural background or by socioeconomic status (Pew Research Center, 2021). Additionally, different populations might vary in *how* they use Google—it could be the case that search behavior is a valid indicator of some underlying psychological trait in one population, but not in another.

In sum, although Google handles most of the world's internet searches, it is not a random and fully representative sample of the world's population (although it is likely more representative than traditional research methods such as surveys on university undergraduate or on samples recruited through crowdsourcing platforms such as Amazon's Mechanical Turk and Prolific Academic). Around the globe, there are still wide swaths of people without stable internet access, and even among internet users, not all people use Google in the same way.

Pitfall 3: Construct Validity Can Be Difficult to Establish

Google Flu Trends showed much early promise, as we described above, but it has become a cautionary tale about using big data for prediction (Lazer et al., 2014). The accuracy of the predictive model declined over time, especially during the H1B1 (swine flu) outbreak in 2013. In part, this decline in accuracy stemmed from one potential pitfall with Google search data: people might make the same Google search but for very different reasons. During the H1B1 outbreak, news of the flu spread fast, many people became worried and curious about this new flu variant, and searched for H1B1 symptoms without being sick (see Milinovich et al., 2014)⁹.

The case of Google Flu Trends underscores the fact that construct validity—whether or not Google Trends scores are an appropriate proxy for the construct of interest—can be difficult to establish and might even change over time. The link between a search and a construct is more straightforward the fewer inferential steps a researcher takes—it seems reasonable to infer that people searching for *wordle answer* and *today wordle* want to know the daily answer to the popular New York Times brainteaser (Wormley & Cohen, 2023). However, it is less clear whether such searches are indicators of constructs that are multiple inferential steps away, such

⁹ Lazer et al. (2014) proposed another possible explanation for this decline in accuracy of Google Flu Trends, having to do with changes to the Google search algorithm.

as being more likely to be a cheater in general. The more inferentially removed a construct is from the search(es) that proxy it, the more caution is required.

Thus, researchers must consider tradeoffs between query specificity and construct validity. In many cases, it is possible to construct queries with very high face-validity for some variable of interest. For example, searches for anti-Black epithets work especially well as a proxy for racism because they are so taboo that it is difficult to imagine using them without harboring some level of prejudice. Similarly, comparing frequencies of *son gifted* vs. *daughter gifted* seems like a straightforward way to assess how readily people look for talents in boys vs. girls (Stephens-Davidowitz, 2017). However, such specific queries may not yield sufficient data. For example, there are fewer people who search “*I want to kill myself*” than there are that search for *suicide*. Although the former has higher face validity, the latter is more likely to yield usable data (especially with more temporal or geographical granularity). (See Figure 3 where we show this with queries for *I’m sad* vs. *Sadness [Topic]*.)

Further, as mentioned under Pitfall 2, people likely use Google in different ways. Thus, even if a query serves as a valid proxy in one geographical location or at one point in time, this may not generalize. (See also our discussion of topics in Pitfall 4.)

Exogenous Factors and Fluctuations in Construct Validity

Foa et al. (2022) found that in the UK, *Boredom (Topic)* searches closely tracked reported boredom from YouGov surveys ($r = .85$). When attempting to replicate these analyses with more recent dates, we noticed a spike—more than 10 standard deviations above the mean—in relative interest in *Boredom (Topic)* on June 5, 2023 (see Figure S5). Upon closer inspection, the top related search on that day was “ennui”, and, more revealing, the related search showing the highest rise in popularity on that day was *ennui wordle*—it was the answer to the New York

Times Wordle puzzle that day. That is, this spike in Google searches was simply due to an exogenous influence, rather than a real increase in boredom.

In contrast, there was another large (but less extreme) spike in the popularity of *Boredom (Topic)* in March 2020. This spike corresponded to the first UK COVID-19 lockdown, and related searches in this period include *bored during lockdown*; this spike, therefore, seems to represent an actual increase in boredom.

Although both events represent spikes, and both can be attributed to exogenous factors (either the Wordle answer of the day or a government-imposed lockdown), we would consider only the latter to represent a genuine shift in felt boredom. (Consistent with this, YouGov data also show a spike in self-reported boredom during this time.) Without YouGov survey data for comparison, one tool to determine which spikes are “genuine” are the related keywords/topics—for large spikes, these can provide clues to alternative reasons why a keyword or topic are being searched. Users will need to use their best judgment to determine which patterns they treat as genuine.

Google Trends Data Are Not Individual-Level Data

Google Trends data are society-level data. That is, they represent the search behavior of the population specified in the query. It is important to note that correlations between variables at the individual level are not always the same as correlations at the society level—a phenomenon known as ecological correlations (Robinson, 1950), or the “ecological fallacy.”

For example, consider the example we began this paper with, whereby more racist counties in the U.S. (i.e., counties where people made more searches for anti-Black racial slurs) had lower vote shares for Barrack Obama in the 2008 U.S. presidential election than expected, relative to previous elections (Stephens-Davidowitz, 2014). Although it is tempting to conclude

that racist *individuals* were less likely to vote for Obama, we cannot conclude this with absolute certainty, only that racist *counties* were less likely to vote for Obama. Additionally, aggregate-level correlations obscure important within-group variability—for instance, a county-level correlation between search terms and voting patterns does not necessarily tell us much about the distribution of these behaviors within counties.

What does this mean for users? First, users should use precise language about the levels of analysis in their claims (e.g., “counties with more racist searches” or even “racist counties” rather than “racist people”). Second, where possible, users can triangulate Google Trends findings with additional evidence. A study that establishes an effect using individual-level data, then conceptually replicates that effect (perhaps in different regions or across time) using Google Trends would be especially compelling. When such triangulation is not possible, it is important to acknowledge this limitation explicitly.

Finally, it is important to consider potential confounding factors that might create spurious society-level correlations. For instance, shared demographic, economic, or other cultural factors might influence both search behavior and the outcome (or predictor) of interest.

Effect Size Considerations

We have suggested several threats to construct validity above; however, whereas some of these threats (e.g., large spikes) are easily detected and accounted for, there are subtler threats that are less easy to detect: small or gradual changes in the demographic composition of a region, evolution in how words are used, or shifts in how Google Trends calculates topics. These subtle factors can add up, making Google Trends data noisy.

In our view, one implication of this is that users should be particularly cautious with research questions that deal with small effect sizes. Even in the best-case scenario, Google

Trends proxies are not precise indicators of psychological constructs but include measurement error. Because of the possibility of substantial noise, researchers likely need a relatively large signal (or effect size) to cut through it.

Pitfall 4: Issues with Topics

Topics have some impressive benefits. Yet there are several reasons to use them with caution. First, Google Trends does not report how they are created. Although related keywords/topics can be useful checks of the constructs that are associated with (and potentially included in) a topic, users cannot uncover the full list of constructs that make up a topic, nor the precise manner in which they are combined. As such, both quality control and making specific inferences are more difficult with topics than keywords.

Second, although they supposedly “cover all languages” (*Basics of Google Trends*, 2024), topics do not include terms from all languages equally (Arora et al., 2019; Woloszko, 2020). Consider *Weather (Topic)*. In the U.S., when we compare queries for *Weather (Topic)* and the *weather* keyword, they return nearly identical relative search volumes (suggesting that in English-speaking countries this topic is almost exclusively based on the keyword *weather*). Elsewhere, as expected, *Weather (Topic)* tracks the words people use in local languages to search for the weather forecast—*météo* in France, *מזג אוויר* in Israel, etc.

However, at the time of this writing, *Weather (Topic)* does not seem to include the Spanish words for weather (*tiempo* or *clima*). When we queried *Weather (Topic)* worldwide, interest-by-region scores were missing (i.e., zeros) for all Spanish-speaking countries, except for Spain. Looking at the related keywords within Spanish-speaking countries, the Spanish words for weather were not listed; Spain’s present but low score seemed to be driven by sufficient search volume for the English and Catalan terms.

Thus, queries for *Weather (Topic)* might erroneously lead one to conclude that Spanish-speaking countries are not interested in the weather. Other topics lack coverage in other languages (Woloszko, 2020). Thus, when using topics, users should explore the associated topics and keywords in each of the languages of their regions of interest to make sure that the topic is valid in that language and region, and across the entire selected date range.

Third, although topics are supposed to isolate constructs of interest (e.g., differentiating *Apple (Fruit)* from *Apple (Technology Company)*), they sometimes fall short. Consider the example of Holmes et al. (2022), discussed above, who compared interest in 100 animal species. Holmes and colleagues concluded that *Lion* was the most popular animal topic. However, when replicating their analysis, we noticed a spike for *Lion* on July 21, 2011 (Figure S6). Upon closer inspection, the top related search keywords (after *lion* itself) on that day were *os lion* and *lion mac*—the new operating system for Mac computers released the day before.

When we limit the *Lion (Animal)* query to the range of dates before the new Mac operating system was released (2004-June 2011, in contrast to Holmes et al. (2022), who looked at 2004-2020), we find that it was *Tiger (Animal)*, not *Lion (Animal)*, that was most popular¹⁰.

Fourth, we have sometimes observed inexplicable shifts in some topic scores. For example, the topic *Psychological Stress (Illness)* is a decent proxy for self-reported stress (Foa et al., 2022). However, this topic inexplicably dips around the end of 2016, recovering toward the end of 2018, worldwide. The query *stress + stressed* does not show a similar dip, suggesting this dip might be a quirk of the topic, rather than a broad mental health shift during this time period

¹⁰ *Tiger (Animal)* itself seems to have been boosted by an external event: in early April, 2020, national attention was captured by a tiger at the Bronx Zoo who tested positive for COVID-19. Although searches for the COVID-19 infected tiger are still for tiger the animal (not, for instance, a new operating system), it is up to researchers whether to consider these to be an unwanted extraneous event or a reflection of interest in tigers.

(see Figure S8). The cause of this dip is not obvious to us, and we can only recommend that *Psychological Stress (Illness)* be used with caution during this time period.

Recommendations

We close with recommendations for the responsible use of Google Trends, including choosing appropriate data, best practices to safeguard data quality, and recommendations for open science and transparent reporting.

Recommendation 1: Creating Proxies for Psychological Variables

Perhaps the most straightforward use of Google Trends—in terms of face validity, or the small inferential gap between Google search data and constructs—is to explore how much information people seek about different constructs, including because of how salient different constructs are (what political scientists term “issue salience”). For example, political scientists have explored interest in corruption, world trade, or human rights (Ajzenman, 2021; Dancy & Fariss, 2023; Mellon, 2013, 2014; Pelc, 2013). With such uses, it does not require a large inferential leap to conclude that people in countries with for example high search volumes for *human rights* are currently thinking more about human rights.

However, psychologists often want to go beyond such simple (and face valid) uses, to explore constructs such as thoughts, feelings, and behaviors. Google search data are more valuable to the extent that they can be used as proxies for such constructs. As noted earlier, it is possible to create proxies for many psychological constructs (see the examples in Table 1), but doing so takes some additional care—and creativity. We have four recommendations for researchers building proxy variables.

Do not Overlook Face-Valid Terms

In our opinion, some of the most interesting uses of Google search data rely on face-valid queries—it stands to reason that people who search for taboo racist slurs or phrases containing the words *how to suicide* harbor at least some racism or have at least some suicidal ideations, respectively. If you can find a face valid keyword(s) or topic with a sufficiently high search volume, that is likely an excellent choice.

Consider the Point of View of Google Users

People often turn to Google to solve problems. Considering the types of problems or needs people have, and the searches they are likely to type into Google to solve these problems or fulfill those needs, can suggest reasonable proxies. For instance, social and clinical psychologists might want proxies for constructs related to mental health and well-being, such as happiness. Google Trends has a *Happiness* topic. However, it is not clear that people who are happy are more likely to search for happiness-related words (indeed, *Happiness (Topic)* seems to be more closely related to *Birthday (Topic)* and *happy birthday*).

A more valid proxy might be the inverse of happiness—*Sadness (Topic)* or even *Depression (Mood)*, or the concrete solutions people who are sad might seek such as *antidepressants* or *therapists near me*. In contrast to happiness, when people have issues, such as a *lack* of happiness, it does seem that they are more likely to turn to Google for information or advice.

Consistent with this notion, previous research has found that searches for negative—not positive—emotions are more predictive of well-being. Foa et al. (2022) showed that searches for most negative emotion topics (e.g., *Sadness*, *Boredom*) are adequate proxies for these emotions (i.e., Google Trends scores for these negative emotion topics closely tracked these self-reported emotions in YouGov surveys); none of the positive emotion topics (such as *Happiness* or

Optimism) were adequate proxies. Similarly, Greyling and Rossouw (2025) used machine learning to test which Google Trends searches are most closely related to survey data on subjective well-being, finding that “sad,” “headache,” and “depressed” were the three most important queries.

To summarize, when constructing a proxy, it is helpful to think not about the topics/keywords that describe the construct of interest, but about what kind of information this construct will prompt people to seek. In many cases, it might also be useful to consider the opposite of the construct of interest—what information do people seek when severely lacking the construct of interest?

Consider Creating Composite Indices

Many of the uses of Google Trends shown in Table 1 incorporate several keywords or topics to create composite scores or indices of a given construct. For example, Foa et al. (2022) created a measure of well-being that incorporated several negative emotion topics. Adamczyk et al. (2022) created cross-cultural religiosity scores by using several queries related to religiosity.

One strength of these indices is that they apply appropriate weighting to multiple keywords or topics. This means that these indices are less susceptible to idiosyncratic shifts in a single variable. For example, an exogenous influence that causes a large spike in a single keyword will be less consequential if that keyword is one of several in a composite measure.

Validate Proxies when Possible

Ideally, users will be able to validate their proxies by showing that they correlate with more standard indices of the underlying construct. For example, Foa et al. (2022) created a measure of well-being using negative emotions topics such as *Sadness*, and showed that, in the UK, weekly searches for these negative emotions correlate highly with weekly YouGov survey

data. The correlations between temporal fluctuations in the self-reported emotions and Google searches related to these emotions suggested that searches are at least somewhat indicative of the actual experience of these emotions. To avoid overfitting, Foa et al. also ensured that their measure was similarly predictive in an “out-of-sample” period.

When such validation data are not available, researchers might validate their proxies by checking whether they correlate with variables that are known to strongly vary with the underlying construct. For example, happiness—and its proxies—might rise on holidays like Christmas in countries that celebrate it (one of the happiest days of the year; Stephens-Davidowitz, 2017), and proxies for religiosity should predict outcomes typically associated with religiosity (Adamczyk et al., 2022).

Recommendation 2: Safeguarding Data Quality

Aim for Datasets without Zeros

We recommend avoiding datasets with zeros whenever possible. As shown above, datasets without zeros typically have high reliability across downloads (average ICC = .95, minimum ICC = .82). While the minimum ICC of .82 represents good reliability, researchers who conduct analyses where even modest measurement error is problematic (e.g., testing small effect sizes) may wish to download the relevant data multiple times to verify that the ICC is higher than this minimum, or if it is not, to average scores across downloads. (However, we reiterate our caution that Google Trends is inherently a noisy data source, and as such not likely well-suited to detect small effects.)

In general, for most applications, a single download of a dataset will be sufficient as long as it does not contain zeros.

What to Do if There Are Zeros

Option 1: Adjust Your Query. If your data have zeros, the most ideal solution is to adjust the query so that scores more consistently rise above the privacy threshold. We describe adjustments that can help.

Find Keywords or Topics with Higher Search Volume. One way to increase the search volume of a query is to find more popular keywords or topics, such as a synonym or a related term. With keywords, you could use the + operator to make the query broader (for example, if *rabbit* is too rare a search, one could change the query to *rabbit + rabbits + bunny + bunnies*, which might more reliably surpass the privacy threshold). As mentioned above, there is a tradeoff between the specificity of a query and the likelihood that it will yield consistent data.

Reconsider Your Temporal and Geographical Granularities. As shown earlier, more granular queries yield less consistent results (and are more likely to yield zeros). If an interest-over-time trend includes zeros, consider using weekly (instead of daily) or monthly (instead of weekly) data (do this simply by making the query cover a longer period, even if the entire period is not needed); if subregions yield inconsistent data, consider zooming out to the country level.

Consider Your Time Period. Older Google Trends data (2004-2008) tend to be more variable and are also likelier to contain zeroes (Raubenheimer, 2024). In part this may be due to fewer people using Google during that period (i.e., a smaller sample size). When older data are not necessary, we suggest 2011 as a sensible starting point for collecting higher quality data. If earlier data are necessary, be aware that those periods may be somewhat more variable.

Option 2: Consider Accessing the Data Across Multiple Retrieval Day. If changing the search terms or reducing the query granularity does not eliminate zeros, or is not feasible for your research question, consider downloading the data multiple times and combining the downloads.

If there are a few zeros throughout a trend on one retrieval day, it is possible that those periods will have non-zero data on subsequent retrieval days. After accessing the data on multiple retrieval days, you can average together all of the non-zero scores (or take the median; Raubenheimer, 2024). The amount of downloads needed will vary depending on the nature of the data—if there are a small amount of zeros, only a few downloads will likely suffice.

Note that if your query contains many zeroes, it is likely not possible to obtain quality data (cf. Raubenheimer, 2024).

What to Do if You Need to Deal with Zeros. If you cannot avoid using datasets with zeros, we recommend coding zeros as missing, as they do not necessarily represent extremely low interest. We recommend extreme caution when using datasets with zeros. Furthermore, even with zeros coded as missing, the rest of the data are likely to have greater measurement error than datasets without zeros.

Other Threats to Data Quality

Rounding. As noted earlier, Google Trends rounds scores to the nearest integer. While this is not a problem with large numbers, it can drastically reduce precision when working with small numbers. As an extreme example, imagine a trend such as *Depression* that has a small fraction (e.g., < 1%) as much research volume as a trend such as *Weather*. In such cases, if the two trends are downloaded together, then all (or nearly all) of the less popular trend scores will be rounded to 1 (or “<1”). To avoid removing all variation in the less popular trend scores, it would be necessary to download the two trends separately.

However, rounding can also reduce precision within a single trend if one time point has exponentially more search volume than other time points. In this case, one solution would be to download data for the desired time periods *around* the large spike. For example, suppose a trend

is composed of weekly scores below 10, with only a single large spike at 100. To obtain greater precision within the non-100 weeks, you could download weekly data spanning up until just before the week coded as 100, and again for all the weeks following that week. Although this would put the before and after week on different scales, it would make rounding less consequential; researchers could then harmonize the two sets using the principles discussed in the Creating Custom Datasets section.

Outliers. We suggest checking for outliers, both by visually examining the trendline and mathematically searching for values high above or below the mean (we avoid making specific recommendations about outlier cut-offs). This is particularly important if your analysis relies on a single trend (rather than a composite, where a single aberration would be less influential).

Outliers can threaten the quality of the data in at least two ways. First, they might represent threats to construct validity. As discussed above, one example of this was searches for *Boredom (Topic)* peaking because the Wordle puzzle of the day was *ennui*. To inspect spikes, we recommend zooming in on the affected period (e.g., the week when the spike occurs) and examining related keywords and topics (both top and rising). In this case, we found related keywords that included *wordle*, leading us to verify the Wordle answer for that day and to conclude that this spike does not represent a genuine increase in boredom.

Researchers must then decide what to do with spikes that seem to stem from unrelated events—in some cases, this spike is short lived, and could be reasonably coded as missing or imputed using standard approaches to missing data. In other cases, such as the *Lion* case discussed earlier, there seemed to be sustained interest in *lion os*—considering this, there is perhaps no straightforward way to salvage these search periods.

Second, even “true” outliers can pose a problem by making the data less precise—see the discussion of rounding above.

Algorithm Updates. As described above (Pitfall 1), Google Trends notes several improvements to its algorithm (e.g., to its geographical assignment in 2011, to its data collection system in 2016, 2017, and 2022). In some cases, there might be some shifts in the data quality following these periods.

If researchers spot such anomalies following improvement timepoints, they might consider correcting for them, either by adjusting the data (Algan et al., 2019), or by adding a dummy-code for before/after the timepoints. Whether an adjustment is necessary (and what the correct adjustment should be) will depend on the nature of the dataset and the intended analyses.

Recommendation 3: Best Practices in Open Science and Reporting

Preregistration

Pre-registration with Google Trends is less straightforward than with most studies using primary data. Unless researchers make predictions about future events (e.g., a hypothesis about how an upcoming holiday will alter search behavior), then pre-registration will not precede the creation of the data.

Nevertheless, it can still be useful to preregister secondary data analysis (Weston et al., 2019). We recommend preregistering a research plan, including (a) the exact keywords or topics to be used, (b) the geographic scope, (c) the temporal scope (e.g., whether daily or weekly data will be used, and across what range), and (d) the full analysis plan.

Of course, researchers who preregister their plan before accessing the data cannot be sure that the data they plan to use will be high quality. Thus, we recommend preregistering decision rules, such as what you will do if a dataset has zeros and how you will treat outliers. For

example, you might preregister an analysis using daily data, but specify that if these datasets include zeros, then you will use weekly data instead. You might also specify contingency plans in the search terms or keywords you will use (e.g., “We will examine whether searches for “*I’m sad*” decline over time, but if this trend includes zeros, then we will instead examine the topic *Sadness*.”)

Reporting

To minimize ambiguity, always specify the exact queries used. When described, it should be clear whether a query used keywords or topics, and with keywords, whether any operators (e.g., quotation marks) were used. We introduced and followed throughout this article several conventions described in Table 2 (e.g., italicizing queries).

Always specify the exact range of dates, temporal granularity, and geographical regions queried. When reporting cross-sectional (interest by region) scores, it may seem that the date range is unimportant, but terms that are more popular during one time period are not necessarily popular during all time periods.

We recommend including Google Trends links (for downloads from the explorer) or code (if downloaded using the API or a pseudo-API) showing exact downloads (which would include the exact range of dates and geographical region). Results will not replicate exactly (because of daily cached samples), but for downloads of trends with high enough search volume, these differences will be negligible.

Conclusion

Google search data are a potential gold mine, holding promise for tackling many psychological and social scientific research questions. Google Trends is a free, easy-to-use, and highly adaptable tool; it makes available both longitudinal and cross-regional data across much

of the internet-using world; data are available nearly immediately, and can range from real-time (by-the-minute) data to monthly data going back years; it circumvents many of the shortcomings of traditional research methods (e.g., desirability biases); it reaches people at times and places where traditional methods are unlikely to reach them; and it is highly amenable to taking advantage of natural experiments and sophisticated research designs that are underused in psychology.

Nevertheless, Google Trends data can be messy. Without proper care, it is easy to download datasets full of low quality data; researchers are unable to audit the algorithms that generate the scores they download, and these data may sometimes be unreliable or invalid; Google Trends relies on users of Google's search engine, which may not be representative of the entire population; it is sometimes difficult to establish construct validity, and construct validity can be compromised by changes across time or by exogenous events; and topics—which are offered as a language-independent alternative to keywords—frequently fall short of their promised function. We hope this tutorial will provide the knowledge necessary to exercise sufficient caution to understand and make use of this tool.

References

- Adamczyk, A., Scott, J., & Hitlin, S. (2022). Using internet-derived data to measure religion: Understanding how Google can provide insight into cross-national religious differences. *Sociology of Religion*, 83(2), 222–251. <https://doi.org/10.1093/socrel/srab034>
- Ajzenman, N. (2021). The power of example: Corruption spurs corruption. *Applied Economics*, 13(2), 230–257. <https://doi.org/10.1257/app.20180612>
- Algan, Y., Murtin, F., Beasley, E., Higa, K., & Senik, C. (2019). Well-being through the lens of the internet. *PLOS ONE*, 14(1), e0209562. <https://doi.org/10.1371/journal.pone.0209562>
- Arora, V. S., McKee, M., & Stuckler, D. (2019). Google Trends: Opportunities and limitations in health and health policy research. *Health Policy*, 123(3), 338–341. <https://doi.org/10.1016/j.healthpol.2019.01.001>
- Basics of Google Trends*. (2024). <https://newsinitiative.withgoogle.com/en-gb/resources/trainings/google-trends/basics-of-google-trends/>
- Baumard, N., Huillery, E., Hyafil, A., & Safra, L. (2022). The cultural evolution of love in literary history. *Nature Human Behaviour*, 6(4), 506–522. <https://doi.org/10.1038/s41562-022-01292-z>
- Blake, K. R., Bastian, B., Denson, T. F., Grosjean, P., & Brooks, R. C. (2018). Income inequality not gender inequality positively covaries with female sexualization on social media. *Proceedings of the National Academy of Sciences*, 115(35), 8722–8727. <https://doi.org/10.1073/pnas.1717959115>
- Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., & Van Bavel, J. J. (2017). Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences*, 114(28), 7313–7318. <https://doi.org/10.1073/pnas.1618923114>

- Brooks, R. C., Russo-Batterham, D., & Blake, K. R. (2022). Incel activity on social media linked to local mating ecology. *Psychological Science*, 33(2), 249–258.
<https://doi.org/10.1177/09567976211036065>
- Cameron, A. C., & Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources*, 50(2), 317–372. <https://doi.org/10.3368/jhr.50.2.317>
- Carrière-Swallow, Y., & Labbé, F. (2013). Nowcasting with Google Trends in an emerging market. *Journal of Forecasting*, 32(4), 289–298. <https://doi.org/10.1002/for.1252>
- Cebrián, E., & Domenech, J. (2023). Is Google Trends a quality data source? *Applied Economics Letters*, 30(6), 811–815. <https://doi.org/10.1080/13504851.2021.2023088>
- Cebrián, E., & Domenech, J. (2024). Addressing Google Trends inconsistencies. *Technological Forecasting and Social Change*, 202, 123318.
<https://doi.org/10.1016/j.techfore.2024.123318>
- Chen, E. E., & Wojcik, S. P. (2016). A practical guide to big data research in psychology. *Psychological Methods*, 21(4), 458–474. <https://doi.org/10.1037/met0000111>
- Choi, H., & Varian, H. (2012). Predicting the present with Google Trends. *Economic Record*, 88(s1), 2–9. <https://doi.org/10.1111/j.1475-4932.2012.00809.x>
- Connor, P., Sarafidis, V., Zyphur, M. J., Keltner, D., & Chen, S. (2019). Income inequality and White-on-Black racial bias in the United States: Evidence from Project Implicit and Google Trends. *Psychological Science*, 30(2), 205–222.
<https://doi.org/10.1177/0956797618815441>
- D'Amuri, F., & Marcucci, J. (2017). The predictive power of Google searches in forecasting US unemployment. *International Journal of Forecasting*, 33(4), 801–816.
<https://doi.org/10.1016/j.ijforecast.2017.03.004>

Dancy, G., & Fariss, C. J. (2023). The global resonance of human rights: What Google Trends can tell us. *American Political Science Review*, 1–22.

<https://doi.org/10.1017/S0003055423000199>

Eichenauer, V. Z., Indergand, R., Martínez, I. Z., & Sax, C. (2022). Obtaining consistent time series from Google Trends. *Economic Inquiry*, 60(2), 694–705.

<https://doi.org/10.1111/ecin.13049>

FAQ about Google Trends data. (2024).

https://support.google.com/trends/answer/4365533?hl=en&ref_topic=6248052

Foa, R. S., Fabian, M., & Gilbert, S. (2022). Subjective well-being during the 2020–21 global coronavirus pandemic: Evidence from high frequency time series data. *PLoS ONE*, 17(2), e0263570. <https://doi.org/10.1371/journal.pone.0263570>

Foa, R. S., & Nezi, R. (2023). *Piercing the fog of war: Measuring Russian public opinion via online search data*. Bennett Institute for Public Policy.

<https://api.repository.cam.ac.uk/server/api/core/bitstreams/0d2e9636-78fb-4d8f-8b56-2bb9e2ef68d8/content>

Franzén, A. (2023). Big data, big problems: Why scientists should refrain from using Google Trends. *Acta Sociologica*, 66(3), 343–347. <https://doi.org/10.1177/00016993221151118>

Gebauer, J. E., & Sedikides, C. (2021). Cultural religiosity: A neglected but powerful dimension of culture. *Current Opinion in Psychology*, 40, 73–78.

<https://doi.org/10.1016/j.copsyc.2020.08.027>

Gelfand, M. J., Raver, J. L., Nishii, L., Leslie, L. M., Lun, J., Lim, B. C., Duan, L., Almaliach, A., Ang, S., Arnadottir, J., Aycan, Z., Boehnke, K., Boski, P., Cabecinhas, R., Chan, D., Chhokar, J., D’Amato, A., Subirats Ferrer, M., Fischlmayr, I. C., ... Yamaguchi, S.

- (2011). Differences between tight and loose cultures: A 33-nation study. *Science*, 332(6033), 1100–1104. <https://doi.org/10.1126/science.1197754>
- Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature*, 457(7232), 1012–1014. <https://doi.org/10.1038/nature07634>
- Greyling, T., & Rossouw, S. (2025). Development and validation of a real-time happiness index using Google Trends. *Journal of Happiness Studies*, 26(3), 39. <https://doi.org/10.1007/s10902-025-00881-9>
- Grosz, M. P., Ayaita, A., Arslan, R. C., Buecker, S., Ebert, T., Hünermund, P., Müller, S. R., Rieger, S., Zapko-Willmes, A., & Rohrer, J. M. (2024). Natural experiments: Missed opportunities for causal inference in psychology. *Advances in Methods and Practices in Psychological Science*, 7(1), 1–15. <https://doi.org/10.1177/251524592312>
- Harlow, L. L., & Oswald, F. L. (2016). Big data in psychology: Introduction to the special issue. *Psychological Methods*, 21(4), 447–457. <https://doi.org/10.1037/met0000120>
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33(2–3), 61–135. <https://doi.org/10.1017/S0140525X0999152X>
- Hogue, J., & DeWilde, B. (2023). *Pytrends 4.9.2* [Computer software]. <https://pypi.org/project/pytrends/>
- Holmes, B., Strzelecki, A., Springer, S., & Zieger, M. (2022). Google Trends data reveal a sharp trend: Teeth and claws attract more interest than feathers, hooves or fins. *Environmental Conservation*, 49(2), 65–73. <https://doi.org/10.1017/S037689292200011X>
- Jackson, J. C., Watts, J., List, J.-M., Puryear, C., Drabble, R., & Lindquist, K. A. (2022). From text to thought: How analyzing language can advance psychological science.

- Perspectives on Psychological Science*, 17(3), 805–826.
<https://doi.org/10.1177/17456916211004899>
- Kern, M. L., Park, G., Eichstaedt, J. C., Schwartz, H. A., Sap, M., Smith, L. K., & Ungar, L. H. (2016). Gaining insights from social media language: Methodologies and challenges. *Psychological Methods*, 21(4), 507–525. <https://doi.org/10.1037/met0000091>
- Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of Google Flu: Traps in big data analysis. *Science*, 343(6176), 1203–1205.
<https://doi.org/10.1126/science.1248506>
- Ma, M. Z., Chen, S. X., & Wang, X. (2024). Looking beyond vaccines: Cultural tightness–looseness moderates the relationship between immunization coverage and disease prevention vigilance. *Applied Psychology: Health and Well-Being*, 16(3), 1046–1072.
<https://doi.org/10.1111/aphw.12519>
- Ma-Kellams, C., Or, F., Baek, J. H., & Kawachi, I. (2016). Rethinking suicide surveillance: Google search data and self-reported suicidality differentially estimate completed suicide risk. *Clinical Psychological Science*, 4(3), 480–484.
<https://doi.org/10.1177/2167702615593475>
- Massicotte, P., & Eddelbuettel, D. (2022). *gtrendsR: Perform and display Google Trends queries* (Version 1.5.1) [Computer software]. <https://cran.r-project.org/web/packages/gtrendsR/gtrendsR.pdf>
- Mavragani, A., & Ochoa, G. (2019). Google Trends in infodemiology and infoveillance: Methodology framework. *JMIR Public Health and Surveillance*, 5(2), e13439.
<https://doi.org/10.2196/13439>

Mellon, J. (2013). Where and when can we use Google Trends to measure issue salience? *PS:*

Political Science & Politics, 46(02), 280–290.

<https://doi.org/10.1017/S1049096513000279>

Mellon, J. (2014). Internet search data and issue salience: The properties of Google Trends as a measure of issue salience. *Journal of Elections, Public Opinion and Parties*, 24(1), 45–

72. <https://doi.org/10.1080/17457289.2013.846346>

Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Google Books Team, T. G. B.,

Pickett, J. P., Hoiberg, D., Clancy, D., Norvig, P., Orwant, J., Pinker, S., Nowak, M. A., &

Aiden, E. L. (2011). Quantitative analysis of culture using millions of digitized books.

Science (New York, N.Y.), 331(6014), 176–182. <https://doi.org/10.1126/science.1199644>

Milinovich, G. J., Williams, G. M., Clements, A. C. A., & Hu, W. (2014). Internet-based

surveillance systems for monitoring emerging infectious diseases. *The Lancet Infectious*

Diseases, 14(2), 160–168. [https://doi.org/10.1016/S1473-3099\(13\)70244-5](https://doi.org/10.1016/S1473-3099(13)70244-5)

Moon, J. W., Barlev, M., Chen, D. L., & Neuberg, S. L. (2025). *A cross-cultural analysis of*

religious fasting and well-being using 14 years of global internet search data. Manuscript in preparation.

Nagao, S., Takeda, F., & Tanaka, R. (2019). Nowcasting of the U.S. unemployment rate using

Google Trends. *Finance Research Letters*, 30, 103–109.

<https://doi.org/10.1016/j.frl.2019.04.005>

Nuti, S. V., Wayda, B., Ranasinghe, I., Wang, S., Dreyer, R. P., Chen, S. I., & Murugiah, K.

(2014). The use of Google Trends in health care research: A systematic review. *PLoS*

ONE, 9(10), e109583. <https://doi.org/10.1371/journal.pone.0109583>

- Pelc, K. J. (2013). Googling the WTO: What search-engine data tell us about the political economy of institutions. *International Organization*, 67(3), 629–655.
<https://doi.org/10.1017/S0020818313000179>
- Pew Research Center. (2021). *Home broadband adoption, computer ownership vary by race, ethnicity in the U.S.* <https://pewrsr.ch/3Bdn6tW>
- Proferes, N., Jones, N., Gilbert, S., Fiesler, C., & Zimmer, M. (2021). Studying Reddit: A systematic overview of disciplines, approaches, methods, and ethics. *Social Media + Society*, 7(2), 20563051211019004. <https://doi.org/10.1177/20563051211019004>
- 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>
- Raubenheimer, J. E. (2024). Of babies, bathwater, and big data: Going beneath the surface of Franzén's (2023) Google Trends recommendations. *Acta Sociologica*, 67(2), 251–256.
<https://doi.org/10.1177/00016993231187489>
- Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review*, 15(3), 351–357.
- Rogers, S. (2016). What is Google Trends data—And what does it mean? *Medium*.
<https://medium.com/google-news-lab/what-is-google-trends-data-and-what-does-it-mean-b48f07342ee8>
- Sparrow, B., Liu, J., & Wegner, D. M. (2011). Google effects on memory: Cognitive consequences of having information at our fingertips. *Science*, 333(6043), 776–778.
<https://doi.org/10.1126/science.1207745>

- Statista. (2023). *Mobile share of organic search engine visits in the United States from 3rd quarter 2013 to 4th quarter 2021*. <https://www.statista.com/statistics/297137/mobile-share-of-us-organic-search-engine-visits/>
- Stephens-Davidowitz, S. (2014). The cost of racial animus on a black candidate: Evidence using Google search data. *Journal of Public Economics*, 118, 26–40.
<https://doi.org/10.1016/j.jpubeco.2014.04.010>
- Stephens-Davidowitz, S. (2017). *Everybody lies: Big data, new data, and what the internet can tell us about who we really are*. Bloomsbury.
- Stephens-Davidowitz, S., & Varian, H. (2014). *A hands-on guide to Google data*.
<https://people.ischool.berkeley.edu/~hal/Papers/2015/primer.pdf>
- Waisberg, E., Ong, J., & Lee, A. G. (2024). Google search spike of “My Eyes Hurt” in United States after solar eclipse: An analysis and future prevention. *Eye*, 38(14), 2677–2679.
<https://doi.org/10.1038/s41433-024-03145-7>
- Walasek, L., & Brown, G. D. A. (2016). Income inequality, income, and internet searches for status goods: A cross-national study of the association between inequality and well-being. *Social Indicators Research*, 129(3), 1001–1014. <https://doi.org/10.1007/s11205-015-1158-4>
- Wang, S.-Y. N., & Inbar, Y. (2021). Moral-Language Use by U.S. Political Elites. *Psychological Science*, 32(1), 14–26. <https://doi.org/10.1177/09567976209603>
- Weston, S. J., Ritchie, S. J., Rohrer, J. M., & Przybylski, A. K. (2019). Recommendations for Increasing the Transparency of Analysis of Preexisting Data Sets. *Advances in Methods and Practices in Psychological Science*, 2(3), 214–227.
<https://doi.org/10.1177/2515245919848684>

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag.

<https://ggplot2.tidyverse.org>

Woloszko, N. (2020). Tracking activity in real time with Google Trends. *OECD Working Papers*,

1634, No. 1634. <https://doi.org/10.1787/6b9c7518-en>

Wormley, A. S., & Cohen, A. B. (2023). C-H-E-A-T: Wordle cheating is related to religiosity and cultural tightness. *Perspectives on Psychological Science*, 18(3), 702–709.

<https://doi.org/10.1177/17456916221113759>