Proceedings of the Twentieth International AAAI Conference on Web and Social Media (ICWSM 2026)

GoogleTrendArchive: A Year-Long Archive of Real-Time Web Search Trends Worldwide

Aleksandra Urman[1], Anikó Hannák[1], Joachim Baumann[2]
[1]Social Computing Group, Department of Informatics, University of Zurich, Switzerland
[2]Stanford NLP Group, Computer Science Department, Stanford University, USA
urman@ifi.uzh.ch, hannak@ifi.uzh.ch, joachimbaumann@stanford.edu

Abstract

`GoogleTrendArchive` is a comprehensive archive of Google Trending Now data spanning over one year (from November 28, 2024 to January 3, 2026) across 125 countries and 1,358 locations. Unlike Google Trends, which requires specifying search terms in advance, Trending Now captures search queries experiencing real-time surges, offering a way to inductively discover trending patterns across regions for studying collective attention dynamics. However, Google does not provide historical access to this data beyond seven days. Our dataset addresses this gap by presenting an archive of Trending Now data. The dataset contains over 7.6 million trend episodes. Each record includes the trend identifier, search volume bucket, precise timestamps, duration, geographic location, and related query clusters. This dataset, among other, enables systematic studies of information diffusion patterns, cross-cultural attention dynamics, crisis responses, and the temporal evolution of collective information-seeking at a global scale. The comprehensive geographic coverage facilitates fine-grained cross-country or cross-regional comparative analyses.

Code — https://github.com/aurman21/googletrendarchive
Dataset — https://doi.org/10.57967/hf/7531

1 Introduction

Search engines are among the primary gateways to online information and the key tool users turn to for information-seeking tasks (Gleason et al. 2023; Nielsen 2016). Due to this, web search behavior can be used to evaluate what information internet users care about at a given moment. For this reason, search trends have proven valuable for studying diverse phenomena including: (mental) health-related topics and epidemics (Ginsberg et al. 2009; Lazer et al. 2014; Nuti et al. 2014; Zhao et al. 2022; Hu et al. 2026); financial and cryptocurrency markets (Aslanidis, Bariviera, and López 2022; Carrière-Swallow and Labbé 2013; Preis, Moat, and Stanley 2013); economic uncertainty (Castelnuovo and Tran 2017); unemployment (Adu, Appiahene, and Afrifa 2023); energy consumption (Fu and Miller 2022); popularity of specific athletes (Malagón-Selma, Debón, and Domenech 2023); fashion trends (Skenderi et al. 2024), and a multitude of other topics (Jun, Yoo, and Choi 2018). However, existing publicly available resources for studying trending online

Figure 1: **We release the GoogleTrendArchive dataset**. We collected daily exports of Google's Trending Now data across **125 countries** and **1,358 total geographic locations**, creating a **multilingual** archive of **7.6M+ trend episodes** spanning **over one year** (Nov 2024–Jan 2026). Unlike Google Trends, which requires pre-specified queries, **this dataset enables inductive discovery of what captured collective attention globally**. This supports research on cross-national information diffusion, crisis communication, and cultural variation in real-time search behavior.

searches such as Google Trends have a fundamental limitation: researchers must specify in advance what to look for. Specifically, Google Trends allows querying predetermined terms retrospectively, but provides no way to *inductively discover* what actually captured attention across populations without pre-specifying search terms the way Trending Now does.

Addressing this limitation, we present a historical dataset of Google's **Trending Now** data collected daily over more than one year across all available geographic locations. Trending Now, rolled out to over 100 countries in August 2024 (Google 2024), displays search queries experiencing rapid surges in interest—as determined by Google compared to the predicted baseline—in the current moment, refreshing approximately every 10 minutes (Google 2024). While anyone can view *current* trends on Google's Trending Now website, the company provides *no historical archive* of such data beyond the last 7 days. Our dataset makes such historical information accessible for researchers, documenting what search queries trended, when the trends started and ended, where they appeared, and how large the surges were. With minimal coverage gaps (approximately 14 days total, concentrated in September 2025, due to technical collection issues), this data enables systematic analysis of collective attention patterns that are otherwise observable only in the moment. Existing archives like the Wayback Machine do not facilitate access to such past data. While Google Trends allows researchers to verify whether specific events they already know about generated search interest, this dataset enables the discovery of what events, topics, and information needs actually emerged as trends across different populations without requiring advance specification. Furthermore, the dataset's comprehensive geographic coverage addresses a critical limitation in computational social science and web search-focused research, where English-language and Western-centric data sources have historically dominated (Septiandri, Constantinides, and Quercia 2024). By capturing trending searches across 125 countries (and a total of 1358 locations, as some countries include data on a regional level), this dataset supports more globally representative studies of collective attention and information diffusion.

This dataset **facilitates several directions of research**. First, it allows for comparative analyses of information consumption patterns across different geographic contexts, for example examining how attention to global events spreads through various regions or which local phenomena capture attention within specific locations. Second, the high temporal resolution permits temporally fine-grained investigation of the trend lifecycles. Third, researchers can examine the interplay between trending topics and external events across time and locations. Finally, the data also captures the actual search queries people use when seeking information about breaking events, enabling cross-linguistic and cross-geographic comparative analyses of query formulations about the same events.

In the following sections, we first document the dataset structure, the data collection and preprocessing procedures, and then provide a high-level analysis of the data.

2 `GoogleTrendArchive` Collection Pipeline

We developed an automated scraper using Playwright in Python (Microsoft 2026) to systematically collect Trending Now data from Google's website. The scraper navigates to the Trending Now interface for each location, corresponding to the "Past 24 hours" timeframe, and downloads the data via the built-in CSV export function. The collection script and location list are released alongside the dataset to enable verification and possible extension of the data collection.

Data collection began on November 28, 2024, and is ongoing. The scraper initiates daily at 9:00 AM Central European Time and proceeds sequentially through all available locations. Due to processing time variations, the exact collection timestamp for any given location may vary by several minutes from day to day, but all locations are collected within the same daily cycle. The currently released dataset includes all data collected through January 3, 2026, with regular updates planned thereafter, subject to continued infrastructure availability.

We identified 1,358 available geographic locations through manual inspection of the Trending Now interface prior to the data collection. These locations comprise 125 countries plus sub-national regions for select countries—that is, all national and sub-national locations for which Google provides Trending Now data. The complete location list with official codes and names is released along with the dataset (see "Trends_LocationList.csv"). As location coverage corresponds to Google's geographic taxonomy, it may change over time as Google adds or removes regions from the Trending Now service. Our data collection has not been affected by such changes so far.

The dataset experienced coverage gaps totaling approximately 14 days due to server outages during the collection period. These gaps occurred when the infrastructure hosting the collection script became unavailable, and outages were not immediately detected. Data validation during collection included verification that the saved CSV files matched the expected format and were non-empty. For sub-national regions with lower search volumes, some daily files are legitimately empty when no trends met Google's detection threshold for that location on that day. We verified this behavior through manual spot-checks comparing empty files against the Trending Now website, confirming that empty files accurately reflected the absence of detected trends rather than collection errors.

It is worth noting that on the July 24, 2025—7 months after our data collection began—Google launched a Google Trends API (Google 2026) which facilitates both, access to the historical trends and also the currently trending data. However, the API is available only to a "very limited number of testers" (Google 2025). At the time of writing, we do not have access to this API. For this reason, we have been continuing the data collection using the original Playwright script. While API access would allow others to systematically archive daily trending data, the only such archive known to us (Aluoch 2026) has two main limitations compared to our dataset. First, the data collected by Aluoch

(2026) comprises only the US. Second, this dataset only contains the data starting 19th of September 2025.

3 `GoogleTrendArchive` structure and content

3.1 Raw Data Collection

The raw dataset consists of daily snapshots collected across all available locations in Trending Now, spanning over one year. Each snapshot captures trends active at or finished within 24 hours of the time of the collection.

Location information and the **collection date** are embedded in the file structure, with separate files for each geographic region following the format `trending_{LOCATION_CODE}_{TIMEFRAME}_{TIMESTAMP}.csv`. Location codes correspond to Google's geographic taxonomy, covering countries and, in some cases, sub-national regions (e.g., US states). The daily files for each location are organized in separate folders, with one folder per location.

Each raw CSV file contains six fields provided by Google's Trending Now system:

The **(1) trend identifier** represents the search query or cluster of related queries that experienced a surge in search interest. This may be a single query or a representative term—as determined by Google—for a cluster of related searches that Google's system identified as referring to the same underlying topic. For example, the following comma-separated queries "*man united vs bodø/glimt, manchester united - bodø/glimt, manchester united, man utd, man united, man u, bodo glimt, uefa, manchester united f.c. vs bodø/glimt lineups, where to watch manchester united f.c. vs bodø/glimt, rasmus højlund, manchester united f.c. vs bodø/glimt*" correspond to the trend identifier "*man united vs bodø/glimt*" and represent different ways in which users in the US on November 28, 2024, searched for information about a UEFA Europa League match on that day.

The **(2) search volume field** provides an approximate number of actual Google searches. Google describes this as "bucketed traffic for the trend within the selected timeframe" (Google 2025). These categories appear as ranges such as "100+", "200+", "500+", "50K+", "100K+", "200K+", "500K+", "2M+", and so forth. Google groups actual traffic into these predetermined buckets rather than providing exact numbers.

Precise temporal information is captured through **(3) start and (4) end timestamps, recorded with timezone information**. These timestamps indicate when Google's forecasting engine first flagged the trend as emerging and when it determined the trend had ended, returning to baseline search levels.

The **(5) trend breakdown field** lists the specific search queries that compose the trend cluster. When multiple related queries surge simultaneously, Google's system groups them together, and this field lists all the variants of the related queries as displayed in Trending Now—as shown in the "*man united vs bodø/glimt*" trend example above. It is worth noting that we found no publicly available information on how the queries are grouped or otherwise processed by Google—a limitation necessary to take into account when interpreting the query data.

Each record includes an **(6) explore link** pointing to the corresponding Google Trends page for the primary query, enabling verification and further investigation of individual trends through Google's standard Trends interface, including retrospective analysis of historical data.

3.2 Processed Dataset Release

To facilitate analysis, we provide a consolidated, processed version of the dataset as a single CSV file. This version addresses several challenges inherent in the raw daily snapshots: (1) the same trending topic may appear in multiple consecutive daily files if it remained active across days, requiring deduplication; (2) timestamp parsing and standardization across timezones; (3) calculation of trend durations; and (4) correction of data quality issues. The complete processing pipeline is described below and the preprocessing code is released along with the dataset.

Data Loading and Standardization We loaded all CSV files across all geographic locations, extracting the collection date from each filename (format: YYYYMMDD) and standardizing column names to lowercase with underscores. Each record was tagged with its source location and collection date to preserve provenance. The combined raw data contained 7,691,790 trend records.

Field Parsing We parsed three key data types from the raw files:

- **Search volume buckets**: We extracted the numeric lower bound of each categorical range (e.g., "500+" $\rightarrow$ 500, "50K+" $\rightarrow$ 50,000, "2M+" $\rightarrow$ 2,000,000) to enable quantitative analysis, while retaining the original categorical values.
- **Timestamps**: We removed timezone suffixes (e.g., "UTC+1", "UTC-5") and parsed all timestamps to UTC using the format "Month day, year at hour:minute:second AM/PM" (e.g., "January 1, 2025 at 2:00:00 AM"). This standardization ensures temporal consistency across all locations.
- **Query counts**: We counted the number of individual search queries in each trend cluster by parsing the comma-separated values in the trend breakdown field.

Episode-Based Deduplication Because we collected daily snapshots, the same trending topic could appear in multiple consecutive files if it remained active across days. Simply concatenating all files would artificially inflate trend counts. To address this, we implemented episode-based deduplication:

Episode identification: For each unique combination of trend identifier and location, we sorted all occurrences chronologically by start time. We defined a new "episode" when the start time differed by more than 1 hour from the previous occurrence's start time. This approach correctly handles cases where the same trend appears multiple times in daily snapshots (e.g., once with a missing end time, once with a complete timestamp), ensuring they are merged into a single episode. The 1-hour threshold distinguishes genuinely distinct trending episodes from duplicate observations of the same continuous trend.

Episode aggregation: For each episode, we collapsed multiple daily observations into a single record by selecting:

- Earliest start time across all occurrences (when the trend first emerged)
- Latest end time across all occurrences (when the trend finally ended)
- Maximum search volume bucket observed during the episode
- First observed trend breakdown and explore link
- Metadata: number of distinct days observed, total raw occurrences, date range

This approach preserves the true temporal extent of each trending episode while eliminating duplicate counting. For example, if a major news event trended continuously from Monday through Wednesday, appearing in three daily snapshot files, the processed dataset contains a single episode record spanning those three days.

Duration Calculation and Quality Corrections We calculated trend duration as the time difference between start and end timestamps. However, inspection of the data revealed two quality issues requiring correction:

Reversed timestamps: In approximately 0.001% of records (76 cases), the end timestamp preceded the start timestamp, likely due to timezone handling inconsistencies in Google's system. We corrected these by swapping the start and end timestamps, flagging such records with the boolean field `times_were_swapped`.

Missing end timestamps: Approximately 14.8% of trend episodes lacked end timestamps, typically representing ongoing trends at the time of collection or cases where Google's system did not record a clear end point. For these cases, we estimated the end time as 23:59:59 UTC on that day (assuming the trend ended by day's end, as otherwise it would be recorded in the next daily snapshot).

All estimated durations are flagged with the boolean field `duration_is_estimate` to enable sensitivity analyses. Researchers can choose to exclude estimated durations, use them as upper bounds of trend duration, or employ alternative estimation strategies.

3.3 Dataset Summary Statistics

The **final processed dataset contains 7,639,695 unique trend episodes across 1,339 geographic locations, spanning November 28, 2024 to January 3, 2026**. Table 1 presents key statistics characterizing the dataset.

The median trend duration of 3.3 hours reflects the typical lifecycle of trending topics, while the maximum duration of 47.8 hours indicates that some topics (such as major ongoing news events) can sustain elevated search interest over multiple days. The 14.8% of episodes with estimated durations represents trends that were ongoing at the time of collection or lacked explicit end timestamps from Google's system.

Metric	**Value**
Coverage	
Total trend episodes	7,639,695
Unique locations*	1,339
Date range	Nov 28, 2024 – Jan 3, 2026
Total N of queries	15,394,099
N of unique queries	1,654,056
Duration statistics (hours)	
Minimum	0.2
Maximum	47.8
Mean	7.2
Median	3.3
Data quality	
Episodes with estimated durations	1,127,320 (14.8%)
Episodes with corrected timestamps	76

*19 small regional locations with empty files only (low search volume) excluded.

Table 1: Summary statistics for the processed dataset.

3.4 Example Record

Table 2 shows an example record from the processed dataset. This trend represents search activity around a UEFA Europa League match, captured in the United States on November 28, 2024. The trend breakdown has 12 distinct query variants (from team abbreviations like "man utd" to specific searches like "where to watch manchester united f.c. vs bodø/glimt"), and the 11.5-hour duration indicates sustained public interest from pre-match through post-match discussion. The `total_occurrences` value of 2 indicates this episode was observed in two daily snapshot files but merged into a single episode as both snapshots corresponded to the same trend, based on the query pattern, location, and trend start time matching. **The Final Dataset Schema detailing the descriptions of each field in the dataset is available in the Appendix A.**

Field	**Value**
trends	man united vs bodø/glimt
search_volume	100K+
search_volume_lower	100000
start_time	2024-11-28 20:00:00 UTC
end_time	2024-11-29 07:30:00 UTC
duration_hours	11.5
n_queries	12
trend_breakdown	man united vs bodø/glimt, manchester united - bodø/glimt, manchester united, man utd, man united, man u, bodo glimt, uefa, [...]
location	US
collection_date	2024-11-29
n_days_observed	1
total_occurrences	2
duration_is_estimate	FALSE
times_were_swapped	FALSE

Table 2: Example trend record from the processed dataset.

3.5 Data Availability

Both the raw daily snapshot files and the processed consolidated dataset are available for download at https://doi.org/10.57967/hf/7531. The raw files preserve the original structure and content as captured from Google's Trending Now system, while the processed dataset provides a ready-to-analyze format. We also provide the complete data processing (R) and scraping (Python) code to ensure full reproducibility and to enable researchers to implement alternative processing strategies if desired. In addition, we provide a Datasheet (Gebru et al. 2021) accompanying the dataset, available at the same DOI.

4 High-level Analysis of the Data

In this section, we present high-level analyses of the data. All analyses are based on the processed—as described above—dataset of 7,639,695 trend episodes, unless specified otherwise.

4.1 Descriptive Statistics: Number of Trends, Queries per Trend, Trend Duration, and Average Search Volume per Trend

Table 4 (Appendix B) presents country-level descriptive statistics for the top 30 countries. The United States showed the highest trending activity with 175,143 total trends and 67,779 unique trends, followed by Japan (146,587 total trends) and the United Kingdom (116,113 total trends). The average number of queries per trend varies from 1.70 (South Korea) to 3.2 (Egypt), with most countries clustering between 2-3 queries per trend. Search volume averages are highest in India (11,774), the United States (10,957), and Brazil (9,239). The cross-national variation observed across variables may reflect differences in population size, internet penetration, search behavior patterns, topic diversity, or Google's trend detection thresholds across markets, though we cannot definitively attribute these patterns to any single factor without additional data on Google's algorithmic parameters by region.

The systematic difference between actual (3-6 hours) and estimated (20-26 hours) trend durations is an artifact of our estimation methodology. For trends lacking observed end times, we conservatively assigned 23:59:59 on the last observation day as the end time. This mechanically produces longer durations: a trend first observed at any time on day N and last seen (still ongoing) on day N receives an estimated duration spanning from its start time to end of day N. The tight clustering of estimated durations around 20-26 hours across all countries—with minimal cross-national variation compared to actual durations—confirms this is methodological rather than substantive. These estimated values represent upper bounds (maximum possible durations): trends ended sometime between their last observation and end of that day, but we cannot determine precisely when. **Researchers should interpret estimated durations as "lasted at most X hours"** and consider sensitivity analyses excluding estimated durations for duration-dependent analyses.

4.2 Topical Distribution of Trends

Top Trends in 2025. An overview of the top 30 trends of 2025 (based on country-level data only, see Figure 4 in the Appendix B) shows that **most top trends are related to sports events or weather**. This pattern is consistent for most regions.

Top Trending Topics by Region. *Topic Annotation:* To characterize the substantive content of trending searches, we manually annotated the top 50 trends from a subset of 20 countries, representing diverse geographic regions and political systems. The authors classified each trend into 15 mutually exclusive topic categories: sports, weather, news, passings (deaths of notable figures), holidays, celebrity scandals, politics, religion, entertainment, gaming, AI, education, food, gambling, and traveling.

For languages not spoken by any of the authors (i.e., Japanese, Arabic, Hebrew, Indonesian, Portuguese), trend titles and associated queries were automatically translated to English using machine translation. When translations were ambiguous or culturally specific, we employed web search to clarify the nature and context of trending topics. This hybrid approach enabled us to conduct reliable classification across linguistically, geographically, culturally and politically diverse contexts.

Results: Figure 2 reveals that sports-related searches dominate trending topics across most countries, often comprising 50-80% of top trends in most of the selected countries except Austria, Belarus, Israel, Japan, South Korea, and Ukraine. These predominantly relate to ongoing sporting events, with specific sports varying by region: cricket features prominently in India, American football (NFL) dominates in the United States, while football (soccer) is the most common sport searched for in other contexts. Given this dominance, researchers using this dataset for purposes unrelated to sports may wish to filter out sports-related trends to focus on the remaining content categories, which still provide substantial volume and diversity for analysis.

Some exceptions to sports dominance in top trends occur in countries experiencing crises, armed conflicts or natural disasters. In Japan, weather-related trends (largely tsunami warnings) constituted a substantial portion of searches. Notably, in South Korea where the majority of trends also corresponded to weather, they referred to "mundane" weather searches, not related to natural disasters. Israel showed elevated news-related searches, predominantly concerning the ongoing armed conflict and the lives and release of hostages. Ukraine also showed heightened attention to news topics, particularly air raid alerts and electricity shutdown notifications. However, otherwise, political topics remain relatively modest in most countries' trending searches, suggesting that while politically significant events may trend, they do not dominate the overall landscape of real-time search behavior. Entertainment, celebrity scandals, and gaming show more variable patterns across countries, possibly reflecting the cultural differences in media consumption patterns.

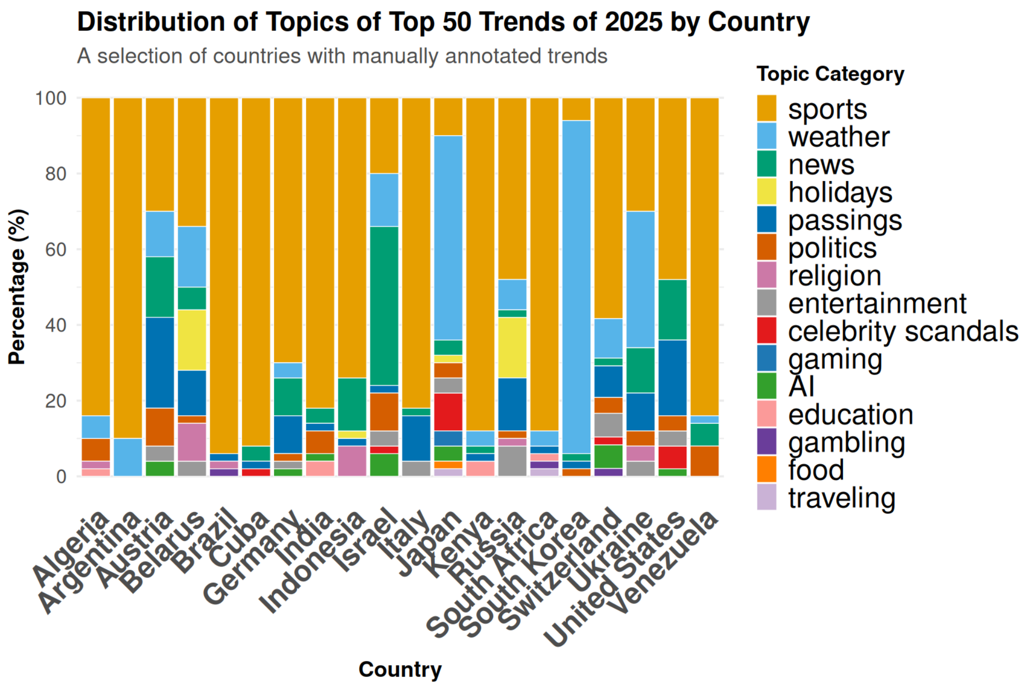

Figure 2: **Distribution of topics** among the top 50 trending searches in 2025 for 20 randomly selected countries. To avoid risks of large-scale automated annotations (Baumann et al. 2025), topics were manually annotated by the authors based on trend titles and associated queries. For languages not spoken by the authors, trend titles were automatically translated to English, with web search used to clarify ambiguous terms or culturally specific references.

4.3 Multi-Country Trends and Global Information Diffusion Patterns

Table 3 shows top 30 global (i.e., having appeared in multiple countries) trends, ordered by the number of countries they appeared in. For this analysis, we utilized only country-level (not regional) data. As seen in Table 3, among the top 30 trends that appeared in multiple countries, most correspond to various sporting events and topics. Beyond those, trends related to generative AI tools—ChatGPT, Deepseek, Gemini—stand out. ChatGPT was the only trend having appeared in all 125 countries in the dataset. Additionally, several political trends appeared globally such as those related to Donald Trump or Iran, and trends related to celebrations and holidays such as New Year or Earth Day.

To characterize the diffusion patterns of trending searches across countries, we analyzed the relationship between geographic spread (number of countries where a trend appears), temporal persistence (time span between first and last appearance), and search volume. For this analysis, we utilized only trends with actual—i.e., non-estimated,—durations to reduce the noise level in the data. Figure 3 presents four complementary views of these dynamics. The distribution of geographic spread (Panel A) follows a power-law-like pattern where 95.1% of trends remain localized to fewer than 20 countries, 4.8% of trends reach 20-99 countries, and only 0.01% achieve truly global reach (100+ countries). The temporal analysis (Panel B) of trends appearing in 10+ countries shows that most cross-border diffusion occurs within 72 hours, suggesting rapid information cascades across national boundaries. Specifically, 91% of trends that transcend country boundaries achieve that in less than 24 hours; 5.1% take 24-72 hours; 3.8% take longer than 72 hours. Furthermore, we observe a positive correlation (Spearman $\rho = 0.43$, $p < 0.001$, $N = 99{,}167$) between geographic spread and search volume (Panel C): trends reaching more countries tend to generate higher search volumes, though substantial variance exists at all levels of spread. The relationship between geographic spread and time span (Panel D) shows a non-monotonic pattern: trends spreading to 20-40 countries show relatively short durations (often under 20 hours), suggesting rapid but short-lived viral diffusion, while trends achieving broader geographic reach (60+ countries) demonstrate more sustained temporal persistence. This pattern suggests that universal relevance rather than mere virality drives lasting global attention. These high-level findings showcase how our dataset might be relevant for research on cross-national or cross-regional (if regional data was to be used) information diffusion, agenda-setting across borders, and the identification of short-lived “viral” versus “sustained” global search trends.

5 Research Directions Enabled by `GoogleTrendArchive`

This dataset enables a wide range of research questions across multiple domains of computational social science, information science, and communication science research. We

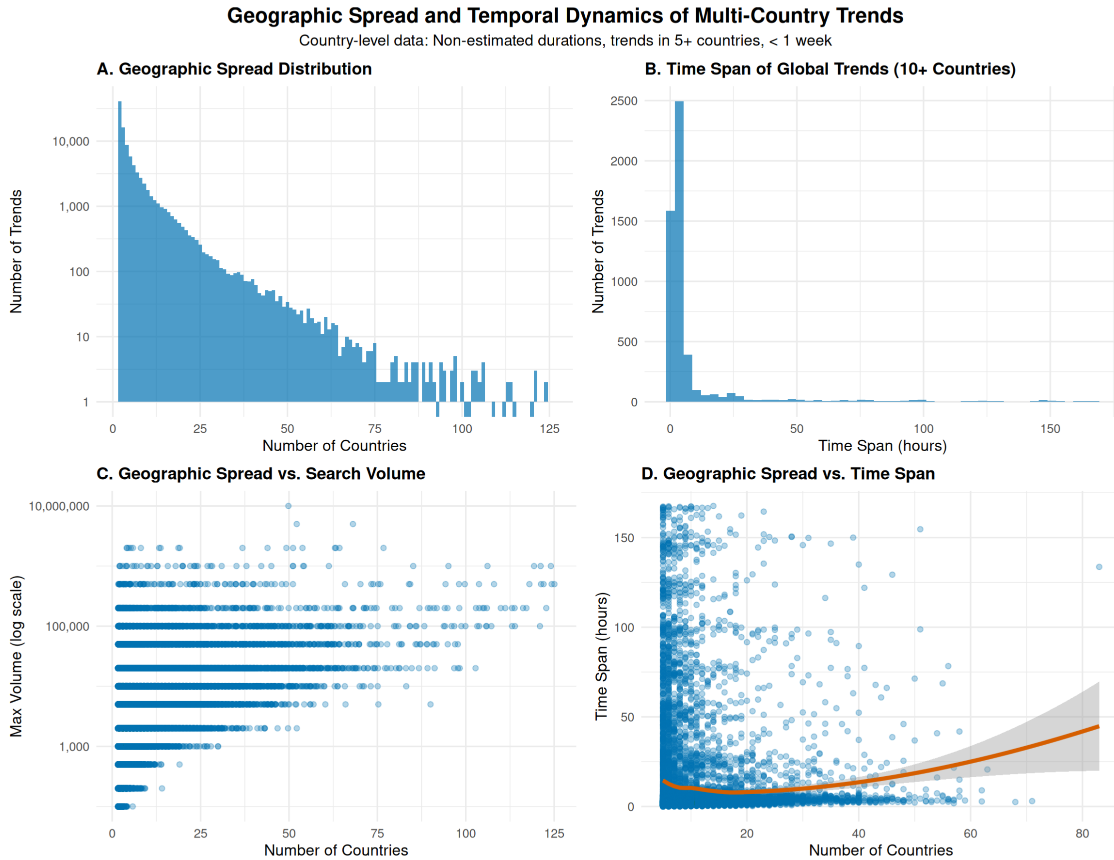

Figure 3: **Geographic spread and temporal dynamics of multi-country trends**; only trend episodes with actual—i.e., not estimated—durations are included. (A) Distribution of geographic spread showing power-law pattern. (B) Time span distribution for trends in 10+ countries. (C) Positive correlation between geographic spread and search volume (Spearman $\rho = 0.43$, p < 0.001). (D) Non-monotonic relationship between geographic spread and time span, with medium-spread trends (20-40 countries) showing shorter durations than broadly global trends (60+ countries). Orange lines: LOESS smoothing; gray shading: 95% CI. Country-level trends with 5+ countries, $<$ 1 week.

identify several key areas where we believe search data can be particularly useful, as partially exemplified by prior work on similar topics done using Google Trends data.

5.1 Information Diffusion and Attention Patterns

The temporal granularity and geographic coverage of this dataset make it particularly suited for studying how information spreads across contexts. Researchers can track how attention to breaking news events spreads across countries/regions, examining the speed and pathways of information diffusion. The dataset enables investigation of which events attract attention globally (capturing attention simultaneously across locations) versus those that remain geographically localized. By analyzing trend lifecycles—from emergence through peak to decay—scholars can identify temporal diffusion and attention patterns associated with different event types: natural disasters, political developments, entertainment news, or sporting events might exhibit distinct attention patterns that can be characterized and modeled (Nghiem et al. 2016; Timoneda and Wibbels 2022; Mellon 2014).

5.2 Crisis Communication and Public Response Patterns

The real-time and temporally fine-grained nature of Trending Now data makes it valuable for studying how publics respond to crises and emergencies. By using the queries associated with the crises-related search trends, researchers can analyze the immediate information needs that emerge during natural disasters, public health events, or political crises. The dataset allows for comparison of crisis response patterns across contexts, identification of information gaps that drive search behavior, and tracking of how attention shifts as crises evolve (Oswald 2026; Rovetta 2021; Timoneda and Wibbels 2022).

5.3 Comparative Cultural, Socio-Political, and Media Analyses

The comprehensive geographic coverage enables systematic comparisons of collective attention patterns across cultural, linguistic, and political contexts. Researchers can examine cross-national or cross-regional attention patterns to spe-

Trend	N Countries	Avg Volume
chatgpt	125	18,324
psg	124	6,841
real madrid	124	7,615
la liga	123	2,653
barcelona	122	5,922
champions league	121	5,514
uefa champions league	121	2,341
weather	121	7,207
new year's eve	120	28,668
diogo jota	119	34,315
weather tomorrow	119	11,841
chelsea	118	3,672
new year's day	118	22,729
premier league	118	4,098
deepseek	117	5,916
gemini	116	4,644
liverpool	116	5,278
earth day	115	15,816
hulk hogan	115	34,663
psg vs inter miami	115	2,241
real madrid vs barcelona	115	6,383
barcelona vs real madrid	114	21,563
copa del rey	114	2,572
arsenal	113	5,827
serie a	113	2,344
europa league	112	4,052
donald trump	111	1,383
iran	111	4,313
ufc	111	9,858
nba	110	2,581

Table 3: Top 30 Multi-Country Trends by the Total N of Countries They Appeared In.

cific topics, analyze whether certain issues resonate universally or remain culturally specific, and how political systems shape information-seeking behavior. For instance, scholars might examine how authoritarian and democratic contexts differ in their patterns of political information-seeking, or whether patterns of public attention to certain issues are related to the local media system structures (Dancy and Fariss 2024; Kupfer and Puhr 2022). The dataset also permits the analysis of linguistic and cultural boundaries in information flows—i.e., whether and how different topics spread across these divides. Furthermore, by comparing search trends with social media trends, news coverage, and other information sources, researchers can study how different platforms and media types interact within the broader information ecosystem (Nghiem et al. 2016). Questions such as whether search leads or follows social media trends, how traditional media events translate into search behavior, and whether certain topics dominate particular platforms can be systematically investigated.

6 Limitations and Ethical Considerations

While this dataset offers substantial research potential, users should be aware of several limitations and ethical considerations.

Original Data Characteristics. Search volume is provided by Google in buckets rather than exact counts, limiting precision in quantitative analyses. Google's trend identification algorithm is proprietary, meaning the exact clustering of related queries is not transparent.

Coverage Gaps. Technical issues resulted in approximately 14 days of missing data, which may have excluded some events. Geographic availability is determined by Google's system and may change over time. Certain locations may have different data quality or coverage patterns due to how Google's proprietary trend identification algorithms work.

Representativeness. Search behavior on Google does not represent entire populations uniformly. Digital divides, device access, and search engine preferences mean that Google Trends reflects particular demographic segments more than others. Cross-country comparisons must account for varying internet penetration rates and search engine market share. For instance, while Google holds over 80% market share in most countries where it is accessible and corresponding Trending Now data is thus present—which excludes, e.g., China,—there are prominent exceptions to this. For instance, in Russia and South Korea Google is currently less popular than local search engines, Yandex and Naver, respectively (Statbase 2026). We suggest the researchers take the market share of Google into account when using the dataset and, where applicable and possible, complement the Google data released by us with similar data from locally relevant search engines such as Yandex Wordstat (Yandex 2026) or Naver SearchTrend (Naver 2026).

Privacy and Ethical Considerations. While trending searches are aggregated public data, researchers should consider the privacy implications of analyzing collective search behavior. Patterns in health-related searches, for instance, might reveal sensitive information about populations. Studies using this data should follow ethical research practices and consider potential harms from publicizing certain attention patterns.

Lessons from Google Flu Trends. Google Flu Trends famously attempted to forecast influenza prevalence based on search query volumes but produced increasingly inaccurate estimates over time (Butler 2013; Olson et al. 2013). The core lessons were that search data should supplement rather than substitute for traditional measurement, and that the relationship between search behavior and real-world phenomena is unstable due to both platform-side changes and shifts in user behavior (Lazer et al. 2014). These concerns are also relevant for our dataset. Google may change how it detects or clusters trending queries at any time. Furthermore, there is a risk of feedback loops where users search for topics because they are listed as trending (Pagan et al. 2023), for example through search engine optimization driven by the public visibility of trends.

7 FAIR Principles Compliance

Our dataset is compliant with the FAIR principles in the following ways:

Findable & Accessible. The dataset is freely available and has a persistent DOI via Hugging Face (https://doi.org/10.57967/hf/7531).

Interoperable. The dataset consists of CSV files; we also release an accompanying preprocessing R script. No proprietary formats or tools are necessary to access or work with the dataset.

Reusable. The dataset is released under CC-BY-4.0 with complete methodology documentation, datasheet, and preprocessing code.

8 Conclusion

We present `GoogleTrendArchive`, a year-long archive of Google Trending Now data across 125 countries and 1,358 total locations. This resource enables systematic study of collective attention and real-time information-seeking at global scale. Unlike aggregated retrospective tools, Trending Now captures surges as they happen, offering unprecedented temporal resolution for studying attention dynamics.

The dataset addresses critical limitations in existing web search resources. The comprehensive geographic coverage enables cross-cultural comparisons, and near-complete temporal coverage supports robust time series analysis. We envision applications spanning information diffusion modeling, crisis communication, comparative cultural studies, and computational social science methodology more broadly.

Acknowledgments

This work was supported by Swiss National Science Foundation (SNSF) grant P500-2_235328 (JB) and SNSF Project Grant 215354 (AU and AH).

The authors acknowledge using Claude to generate parts of the code for Figures 2 and 3 and add comments to the data preprocessing code. The authors have verified no errors were introduced. The authors have not used AI-based tools for other parts of the research processes, including the writing of the manuscript.

References

Adu, W. K.; Appiahene, P.; and Afrifa, S. 2023. VAR, ARIMAX and ARIMA models for nowcasting unemployment rate in Ghana using Google trends. *Journal of Electrical Systems and Information Technology*, 10(1): 12.

Aluoch, A. J. 2026. Daily Google Search Trends. https://www.kaggle.com/datasets/adrianjuliusaluoch/daily-google-search-trends-us.

Aslanidis, N.; Bariviera, A. F.; and López, Ó. G. 2022. The link between cryptocurrencies and Google Trends attention. *Finance Research Letters*, 47: 102654.

Baumann, J.; Röttger, P.; Urman, A.; Wendsjö, A.; Plaza-del Arco, F. M.; Gruber, J. B.; and Hovy, D. 2025. Large language model hacking: Quantifying the hidden risks of using llms for text annotation. *arXiv preprint arXiv:2509.08825*.

Butler, D. 2013. When Google got flu wrong: US outbreak foxes a leading web-based method for tracking seasonal flu. *Nature*, 494(7436): 155–157.

Carrière-Swallow, Y.; and Labbé, F. 2013. Nowcasting with Google Trends in an Emerging Market. *Journal of Forecasting*, 32(4): 289–298. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/for.1252.

Castelnuovo, E.; and Tran, T. D. 2017. Google It Up! A Google Trends-based Uncertainty index for the United States and Australia. *Economics Letters*, 161: 149–153.

Dancy, G.; and Fariss, C. J. 2024. The Global Resonance of Human Rights: What Google Trends Can Tell Us. *American Political Science Review*, 118(1): 252–273.

FORCE11. 2020. The FAIR Data principles. https://force11.org/info/the-fair-data-principles/.

Fu, C.; and Miller, C. 2022. Using Google Trends as a proxy for occupant behavior to predict building energy consumption. *Applied Energy*, 310: 118343.

Gebru, T.; Morgenstern, J.; Vecchione, B.; Vaughan, J. W.; Wallach, H.; Iii, H. D.; and Crawford, K. 2021. Datasheets for datasets. *Communications of the ACM*, 64(12): 86–92.

Ginsberg, J.; Mohebbi, M. H.; Patel, R. S.; Brammer, L.; Smolinski, M. S.; and Brilliant, L. 2009. Detecting influenza epidemics using search engine query data. *Nature*, 457(7232): 1012–1014. Publisher: Nature Publishing Group.

Gleason, J.; Hu, D.; Robertson, R. E.; and Wilson, C. 2023. Google the Gatekeeper: How Search Components Affect Clicks and Attention. *Proceedings of the International AAAI Conference on Web and Social Media*, 17: 245–256.

Google. 2024. 3 Trending Now updates to help you keep up with the latest trends. https://blog.google/products-and-platforms/products/search/google-trends-trending-now-update/.

Google. 2025. Google Trends Help: Explore the searches that are trending now. https://support.google.com/trends/answer/3076011.

Google. 2025. Introducing the Google Trends API (alpha): a new way to access Search Trends data | Google Search Central Blog. https://developers.google.com/search/blog/2025/07/trends-api.

Google. 2026. Google Trends API Alpha | Google Search Central | Documentation. https://developers.google.com/search/apis/trends.

Hu, D.; Baumann, J.; Urman, A.; Lichtenegger, E.; Forsberg, R.; Hannak, A.; and Wilson, C. 2026. Auditing Google's AI Overviews and Featured Snippets: A Case Study on Baby Care and Pregnancy. In *Proceedings of the International AAAI Conference on Web and Social Media*.

Jun, S.-P.; Yoo, H. S.; and Choi, S. 2018. Ten years of research change using Google Trends: From the perspective of big data utilizations and applications. *Technological Forecasting and Social Change*, 130: 69–87.

Kupfer, A.; and Puhr, H. 2022. The Russian View on the War in Ukraine: Insights from Google Trends. https://papers.ssrn.com/abstract=4063194.

Lazer, D.; Kennedy, R.; King, G.; and Vespignani, A. 2014. The Parable of Google Flu: Traps in Big Data Analysis. *Science*, 343(6176): 1203–1205. Publisher: American Association for the Advancement of Science.

Malagón-Selma, P.; Debón, A.; and Domenech, J. 2023. Measuring the popularity of football players with Google Trends. *PLOS ONE*, 18(8): e0289213. Publisher: Public Library of Science.

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. Publisher: Routledge _eprint: https://doi.org/10.1080/17457289.2013.846346.

Microsoft. 2026. microsoft/playwright-python. https://github.com/microsoft/playwright-python. Original-date: 2020-07-01T15:28:13Z.

Naver. 2026. Naver Data Lab: Search Trends. https://datalab.naver.com/keyword/trendSearch.naver.

Nghiem, L. T. P.; Papworth, S. K.; Lim, F. K. S.; and Carrasco, L. R. 2016. Analysis of the Capacity of Google Trends to Measure Interest in Conservation Topics and the Role of Online News. *PLOS ONE*, 11(3): e0152802. Publisher: Public Library of Science.

Nielsen, R. K. 2016. News media, search engines and social networking SITES as varieties of online gatekeepers. In *Rethinking Journalism Again*. Routledge. Num Pages: 16.

Nuti, S. V.; Wayda, B.; Ranasinghe, I.; Wang, S.; Dreyer, R. P.; Chen, S. I.; and Murugiah, K. 2014. The Use of Google Trends in Health Care Research: A Systematic Review. *PLOS ONE*, 9(10): e109583. Publisher: Public Library of Science.

Olson, D. R.; Konty, K. J.; Paladini, M.; Viboud, C.; and Simonsen, L. 2013. Reassessing Google Flu Trends Data for Detection of Seasonal and Pandemic Influenza: A Comparative Epidemiological Study at Three Geographic Scales. *PLOS Computational Biology*, 9(10): 1–11.

Oswald, C. 2026. I Still Haven't Found what I'm Looking for: Predicting Security-Related Incidents and Conflict Fatalities with Google Trends and Wikipedia Data. *Journal of Conflict Resolution*, 70(2-3): 499–524. Publisher: SAGE Publications Inc.

Pagan, N.; Baumann, J.; Elokda, E.; De Pasquale, G.; Bolognani, S.; and Hannák, A. 2023. A Classification of Feedback Loops and Their Relation to Biases in Automated Decision-Making Systems. In *Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*, EAAMO '23. New York, NY, USA: Association for Computing Machinery.

Preis, T.; Moat, H. S.; and Stanley, H. E. 2013. Quantifying Trading Behavior in Financial Markets Using Google Trends. *Scientific Reports*, 3(1): 1684. Publisher: Nature Publishing Group.

Rovetta, A. 2021. Reliability of Google Trends: Analysis of the Limits and Potential of Web Infoveillance During COVID-19 Pandemic and for Future Research. *Frontiers in Research Metrics and Analytics*, 6. Publisher: Frontiers.

Septiandri, A. A.; Constantinides, M.; and Quercia, D. 2024. WEIRD ICWSM: How Western, Educated, Industrialized, Rich, and Democratic is Social Computing Research? *arXiv preprint arXiv:2406.02090*.

Skenderi, G.; Joppi, C.; Denitto, M.; and Cristani, M. 2024. Well googled is half done: Multimodal forecasting of new fashion product sales with image-based google trends. *Journal of Forecasting*, 43(6): 1982–1997. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/for.3104.

Statbase. 2026. Search engine market share by country. Data by Countries from 2024 to 2025. https://statbase.org/datasets/internet-and-technology/search-engine-market-share-by-country/.

Timoneda, J. C.; and Wibbels, E. 2022. Spikes and Variance: Using Google Trends to Detect and Forecast Protests. *Political Analysis*, 30(1): 1–18.

Yandex. 2026. Yandex Wordstat. https://wordstat.yandex.ru/.

Zhao, X.; Coxe, S. J.; Timmons, A. C.; and Frazier, S. L. 2022. Mental Health Information Seeking Online: A Google Trends Analysis of ADHD. *Administration and Policy in Mental Health and Mental Health Services Research*, 49(3): 357–373.

Paper Checklist

- For most authors...
  - Would answering this research question advance science without violating social contracts, such as violating privacy norms, perpetuating unfair profiling, exacerbating the socio-economic divide, or implying disrespect to societies or cultures? Yes, this dataset paper presents aggregated, publicly available search trend data. The data contains no personally identifiable information and represents collective attention patterns rather than individual behavior.
  - Do your main claims in the abstract and introduction accurately reflect the paper's contributions and scope? Yes, the abstract and introduction clearly state that we present a year-long dataset of Google Trending Now data across all geographic locations, describe the data collection methodology, and outline potential research applications. We do not overclaim the dataset's capabilities and explicitly acknowledge limitations including coverage gaps in the main text.
  - Do you clarify how the proposed methodological approach is appropriate for the claims made? Yes, we describe our data collection and preprocessing methodology in detail in the relevant sections. We explain how our methodology enables the temporal and geographic coverage necessary for the research applications we identify.

- Do you clarify what are possible artifacts in the data used, given population-specific distributions? Yes, we mention in the Limitations section that search behavior does not uniformly represent entire populations due to digital divides, varying internet penetration rates, and search engine market share differences across countries.
- Did you describe the limitations of your work? Yes, we explicate the limitations of the dataset in a dedicated section.
- Did you discuss any potential negative societal impacts of your work? Yes, we explicitly discuss privacy implications of analyzing collective search behavior, particularly for sensitive topics like health-related searches. We note that publicizing certain attention patterns could potentially cause harm and emphasize that researchers should follow ethical practices when using this data.
- Did you discuss any potential misuse of your work? Yes, we note that while the data is aggregated and public, researchers should consider privacy implications and potential harms from publicizing search patterns.
- Did you describe steps taken to prevent or mitigate potential negative outcomes of the research, such as data and model documentation, data anonymization, responsible release, access control, and the reproducibility of findings? Yes, we provide comprehensive documentation including data schema, collection methodology, and known limitations. The data is already aggregated at the collective level with no individual-level information.
- Have you read the ethics review guidelines and ensured that your paper conforms to them? Yes, we have.

• Additionally, if your study involves hypotheses testing...
  - Did you clearly state the assumptions underlying all theoretical results? NA
  - Have you provided justifications for all theoretical results? NA
  - Did you discuss competing hypotheses or theories that might challenge or complement your theoretical results? NA
  - Have you considered alternative mechanisms or explanations that might account for the same outcomes observed in your study? NA
  - Did you address potential biases or limitations in your theoretical framework? NA
  - Have you related your theoretical results to the existing literature in social science? NA
  - Did you discuss the implications of your theoretical results for policy, practice, or further research in the social science domain? NA

• Additionally, if you are including theoretical proofs...
  - Did you state the full set of assumptions of all theoretical results? NA
  - Did you include complete proofs of all theoretical results? NA

• Additionally, if you ran machine learning experiments...
  - Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? NA
  - Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? NA
  - Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? NA
  - Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? NA
  - Do you justify how the proposed evaluation is sufficient and appropriate to the claims made? NA
  - Do you discuss what is “the cost“ of misclassification and fault (in)tolerance? NA

• Additionally, if you are using existing assets (e.g., code, data, models) or curating/releasing new assets, **without compromising anonymity**...
  - If your work uses existing assets, did you cite the creators? NA
  - Did you mention the license of the assets? Yes, the dataset is released under CC-BY-4.0.
  - Did you include any new assets in the supplemental material or as a URL? Yes, we provide the links to the dataset repository and the corresponding DOI https://doi.org/10.57967/hf/7531
  - Did you discuss whether and how consent was obtained from people whose data you're using/curating? Yes, we discuss this at length in the Datasheet for the dataset (Gebru et al. 2021) released alongside the dataset in the dataset repository.
  - Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? Yes, we discuss this at length in the Datasheet for the dataset (Gebru et al. 2021) released alongside the dataset in the dataset repository.
  - If you are curating or releasing new datasets, did you discuss how you intend to make your datasets FAIR (see FORCE11 (2020))? Yes, we describe how our dataset release is in line with the FAIR principles in a dedicated section.
  - If you are curating or releasing new datasets, did you create a Datasheet for the Dataset (see Gebru et al. (2021))? Yes, the Datasheet is available in the corresponding dataset repository.

• Additionally, if you used crowdsourcing or conducted research with human subjects, **without compromising anonymity**...
  - Did you include the full text of instructions given to participants and screenshots? NA
  - Did you describe any potential participant risks, with mentions of Institutional Review Board (IRB) approvals? NA

– Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? NA
– Did you discuss how data is stored, shared, and deidentified? NA

A Appendix: Detailed Dataset Schema

The processed dataset contains the following fields:

Core trend information

- `trends`: Trend identifier (representative query or cluster name)
- `trend_breakdown`: Comma-separated list of all queries in the cluster
- `n_queries`: Number of queries in the trend breakdown
- `explore_link`: URL to Google Trends page for this trend

Search volume

- `search_volume`: Original categorical bucket (e.g., "50K+")
- `search_volume_lower`: Numeric lower bound of the bucket

Temporal information

- `start_time`: When Google first detected the trend (UTC)
- `end_time`: When Google determined the trend ended (UTC)
- `duration_minutes`: Calculated duration in minutes
- `duration_hours`: Calculated duration in hours
- `collection_date`: Date when first observed in our collection
- `first_collection_date`: First date the episode appeared (if multi-day)
- `last_collection_date`: Last date the episode appeared (if multi-day)
- `year`, `month`, `weekday`: Extracted date components

Geographic information

- `location`: Geographic location code (e.g., "US", "DE", "BR")

Episode metadata

- `episode_id`: Unique identifier within each trend-location combination
- `n_days_observed`: Number of distinct days the trend appeared
- `total_occurrences`: Number of raw daily records collapsed into this episode

Data quality flags

- `duration_is_estimate`: Boolean indicating whether duration is estimated (missing end time)
- `times_were_swapped`: Boolean indicating timestamp correction (reversed start/end)

B Appendix: Top 30 Trends of 2025 by Total Search Volume

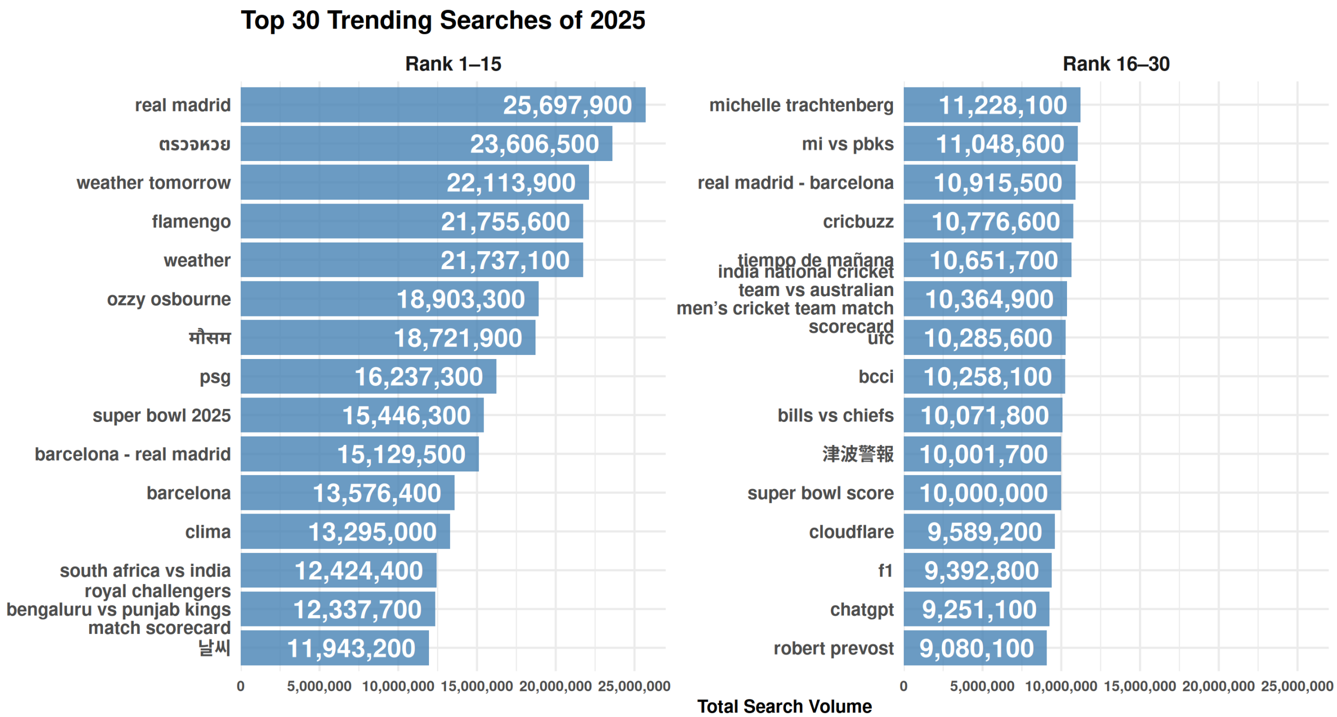

Figure 4: **Top 30 Trends** of 2025 (Country-level data only) by Total Search Volume; Country-Level Statistics

Rank	Country	Total Trends	Unique Trends	Total Queries	Avg Queries/Trend	Avg Dur Actual (hrs)	Avg Dur Est (hrs)	Avg Volume
1	United States	175,143	67,779	468,912	2.7	3.4	22.5	10,957
2	Japan	146,587	47,217	264,493	1.8	3.5	22.0	2,963
3	United Kingdom	116,113	48,205	261,915	2.3	3.6	21.6	3,840
4	Germany	99,788	44,584	198,577	2.0	3.6	24.1	4,028
5	Brazil	94,391	39,329	285,936	3.0	3.5	25.2	9,239
6	Canada	90,088	41,013	193,403	2.1	3.9	23.5	2,522
7	France	87,901	37,747	183,739	2.1	3.3	23.9	3,785
8	India	85,897	36,437	223,563	2.6	4.0	24.2	11,774
9	Italy	85,198	37,532	200,316	2.4	3.4	24.2	4,429
10	Spain	66,189	30,498	149,606	2.3	3.5	23.1	4,465
11	Türkiye	65,513	29,352	179,402	2.7	4.0	22.6	6,851
12	Indonesia	59,298	26,904	137,912	2.3	4.1	25.1	4,619
13	Mexico	54,931	26,968	134,024	2.4	3.8	23.4	6,609
14	Russia	50,314	24,113	98,423	2.0	5.7	23.8	2,064
15	Australia	47,956	25,045	98,987	2.1	3.8	22.9	2,325
16	Argentina	44,175	20,895	106,996	2.4	3.6	24.8	5,729
17	Netherlands	41,598	21,575	79,095	1.9	4.6	23.2	2,153
18	Poland	39,776	19,408	81,615	2.1	3.7	23.1	3,311
19	Taiwan	38,294	17,486	72,783	1.9	4.9	22.2	1,879
20	Iran	33,613	14,663	74,626	2.2	4.0	23.9	2,664
21	South Korea	33,054	13,961	55,735	1.7	4.9	23.4	2,507
22	Colombia	31,382	16,139	93,121	3.0	4.0	24.6	6,983
23	Saudi Arabia	31,082	16,163	71,167	2.3	5.8	24.7	3,205
24	South Africa	30,348	15,834	70,025	2.3	4.8	25.7	3,101
25	Chile	29,700	15,006	70,594	2.4	3.9	24.4	3,754
26	Egypt	29,408	15,755	93,265	3.2	5.8	27.0	4,349
27	Peru	28,171	14,482	77,526	2.8	4.4	24.7	5,618
28	Bangladesh	27,056	14,458	53,296	2.0	5.5	26.0	3,958
29	Vietnam	26,503	13,169	76,078	2.9	4.3	20.7	5,116
30	Philippines	25,950	12,767	53,952	2.1	3.9	24.0	2,351

Table 4: Country-level descriptive statistics (top 30 countries by the N of total trends)