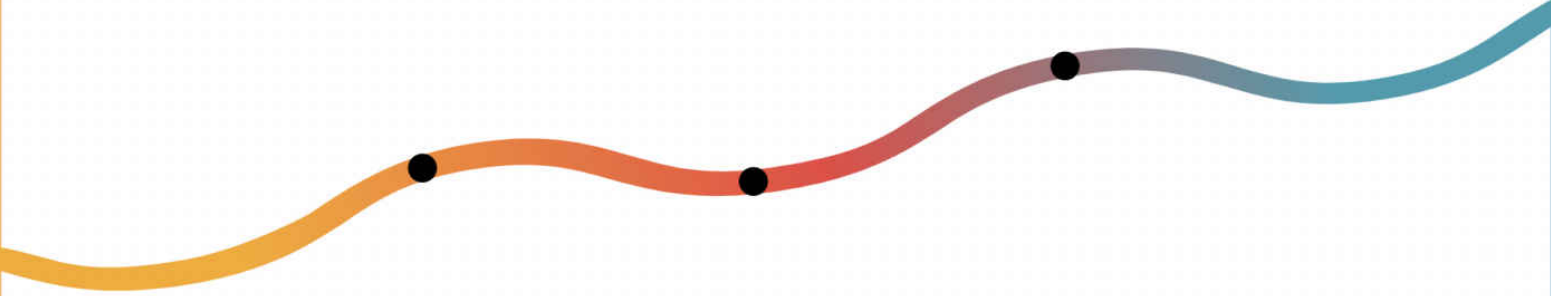


GOOGLE ANALYTICS ALTERNATIVES



A Guide to Navigating the
World of Options Beyond Google

JASON PACKER

Contents

Part One - Background

[Chapter 1 - Introduction](#)

[Chapter 2 - The Product Space](#)

[Chapter 3 - The Google Ecosystem](#)

[Chapter 4 - It's Time to Consider Alternatives](#)

[Chapter 5 - Types of Tools](#)

[Chapter 6 - Pricing](#)

[Chapter 7 - Event Tracking Methodology](#)

[Chapter 8 - Support and Community](#)

[Chapter 9 - Should I Self-Host?](#)

[Chapter 10 - Open Source](#)

[Chapter 11 - How Important Is Tracker Speed?](#)

[Chapter 12 - Advanced Features](#)

[Chapter 13 - Cookies & "Future-Proofing"](#)

[Chapter 14 - Privacy & Compliance](#)

Part Two - Product Evaluations

[Evaluation Methodology](#)

[Tools Index](#)

[Traditional Web Analytics](#)

[Chapter 15 - Matomo Cloud](#)

[Chapter 16 - Piwik PRO](#)

[Chapter 17 - Clicky](#)

[Simplified Web Analytics](#)

[Chapter 18 - Cloudflare Web Analytics](#)

[Chapter 19 - Statcounter](#)

[Chapter 20 - Chartbeat](#)

[Chapter 21 - Fathom](#)

[Chapter 22 - Plausible Analytics](#)

[Chapter 23 - Visitor Analytics](#)

[Product Analytics](#)

[Chapter 24 - Google Analytics 4](#)

[Chapter 25 - Mixpanel](#)

[Chapter 26 - Snowplow](#)

[Chapter 27 - Amplitude Analytics](#)

[Chapter 28 - Heap](#)

[Chapter 29 - PostHog](#)

[Chapter 30 - Conclusion](#)

[Thanks and Acknowledgments](#)

[Appendix](#)

PART ONE

BACKGROUND

CHAPTER 1 - INTRODUCTION

IT'S AN INTERESTING TIME to work in analytics. As web and digital analytics have matured and attracted more attention over the last decade, there's also been significant technological and regulatory change. I think of web analytics as an accidental industry, one that began as a generally ignored offshoot of IT that's grown to be an important part of marketing, product, and development.

Web analytics remains new and ever-changing, evidenced by the lack of a consistently used name or definition within the field. What kind of analytics is it? Web? Digital? Product? Mobile App?

Google Analytics (GA) has been the predominant analytics tool since its launch in 2005 and is so entrenched, it's become synonymous with the term "web analytics." Google has also helped move the field from its roots in IT towards marketing — integrating GA into its larger advertising technology stack and focusing functionality on marketing analytics.

If the first age of web analytics was the era of log file processing and Webtrends, the second has been client-side JavaScript and Google Analytics.

However, on July 1, 2023, this era will come to an end, as Google Analytics Universal Analytics (UA) will no longer process new data. Six months after that, access to historical data within UA will also disappear. This abrupt end to 18 years of contiguous data has upset many people and brought into question Google's long-standing dominance in the field.

Era	Predominant Tools	Technologies
1 st : 1995- 2004	Webtrends, Analog, AWStats, Omniture	Logfile Analysis, Hit Counters
2 nd : 2005- 2022	Google Analytics	JavaScript Tags, Third Party Web Data Collection
3 rd : 2023-	Google Analytics 4 and ?	Hybrid Client+Server, First Party Data, Multi-platform Data Collection

In this guide, we'll look at some of the **most popular alternatives to Google Analytics** and create a decision-making framework to select a new solution. Instead of declaring a single best solution, it'll inform you of the available options and how to choose between them.

Even though this is a guide to GA alternatives, we will approach GA both as a baseline for comparison and as a potential solution itself. You may find GA4 is the right solution for you, but my goal is to help you make that choice in an informed way.

I come at analytics from a technical perspective, so this guide will be oriented towards technical details more than business concerns.

You also won't find any big lists of feature comparisons. Instead, I look at the point of view that each tool comes from and its most appropriate use cases.

Giant feature comparison matrices are helpful, but they can also be misleading and cause choice paralysis. Just because a feature is available doesn't guarantee it'll be able to solve the particular issue you have. The vendor's idea of what checks that feature box may not align with how you envision that feature. Additionally, those comparison lists become outdated very quickly, with features and pricing changing even from week to week.

We'll go through a series of Universal Analytics alternatives, review their pros and cons, and compare them to each other. I'll do my best to highlight features that I think are most representative, but we will not go deep into implementation details. I also treat GA4 as one of these alternatives, endeavoring to look at it with fresh eyes rather than in the shadow of UA.

The choice at hand is between GA4 and its competitors, not UA. UA can provide us with context for our comparisons, but that's all. Many people in the industry (myself included) hoped Google would somehow find a way to continue the functionality of UA, but they've given no indication of this, so we have to proceed assuming UA's days are numbered.

Tool choice is only part of the equation of a successful analytics implementation. No solution will meet your needs perfectly — it's about finding the best fit and then putting in the time to learn, customize, and implement the tool. All too commonly, companies abandon Tool A for Tool B because they think the former isn't right for them, but in reality, they simply failed to use it well.

The background context we will go over in Part One of this guide is not only to help with tool selection, but to provide a better understanding of the paradigm the tool works within. It's my belief that this deeper understanding will also help to better utilize whichever tool you pick.

CHAPTER 2 - THE PRODUCT SPACE

OUR FIRST STEP IS to list the potential Google Analytics alternatives to evaluate. We seem to be moving into an age where no single tool will dominate the market, so it's important to identify all the possibilities rather than blindly relying upon one market leader.

This is a difficult task, because the web analytics space is incredibly crowded with tools all vying for your attention and a spot in your tag manager. GA is currently the most popular solution, but there are a multitude of other options — so many, it's impossible for me to evaluate them all (sorry to disappoint if you were hoping for a 1,000-page book).



The famous ChiefMartec industry map (available at <https://martechmap.com>), which you now need an electron microscope to read, lists 120 vendors in the "mobile & web analytics" category, but there's certainly more than that! Many lesser-known options (such as PostHog, which we'll evaluate later) are not included on the map. Plus, solutions that should be in that category aren't (e.g., Simple Analytics is a web analytics tool but is listed under "Dashboards & Data Viz").

We're starting with more than 120 possible solutions. But with seemingly every article and product review site listing a different group of purported alternatives, how do we proceed?

I imagine the first thing many people (including myself) would do is Google Search for "Google Analytics Alternatives," implicitly trusting Google to give us an unbiased opinion on how to replace their product. As is typical with this type of search, there are some good results, along with plenty of junk written solely for SEO.

You'll see a number of vendors (e.g., Semrush, Hubspot, and Leadfeeder) with their own lists that just happen to feature their product at number one. Hat tip to their content teams, but those tools aren't GA alternatives. Some lists also include items that are in the same space but are not web analytics tools (e.g., session capturing, SEO tools, etc.).

So, before we identify possible alternatives to GA, we need to define what GA actually is and the space it occupies. GA has a ton of features that allow it to be compared to a number of different types of tools. However, just because GA can do rudimentary heat-mapping doesn't mean we should consider a heat-mapping tool like CrazyEgg or Hotjar a GA alternative. Similarly, GA can monitor site speed via real user monitoring (RUM), but that doesn't make a web performance tool like SpeedCurve or New Relic a GA alternative.

Many of GA's secondary features are limited in scope and functionality. GA4 attempts to improve the tool's focus by jettisoning many of the poorly implemented legacy features. This narrows what GA4 can do, but in a good way, I believe.

By our definition, an "alternative" is something that **can replace a significant amount of the GA core functionality**. That constraint alone doesn't cut enough from our list of 120+ though. To narrow it down further, let's try to analyze what products actually compete with GA within the marketplace.

Who are these competitors? When measured in terms of overall installation base, no tool or software can rival GA. Nonetheless, looking at alternatives sorted by popularity is still an effective way to filter down our options.

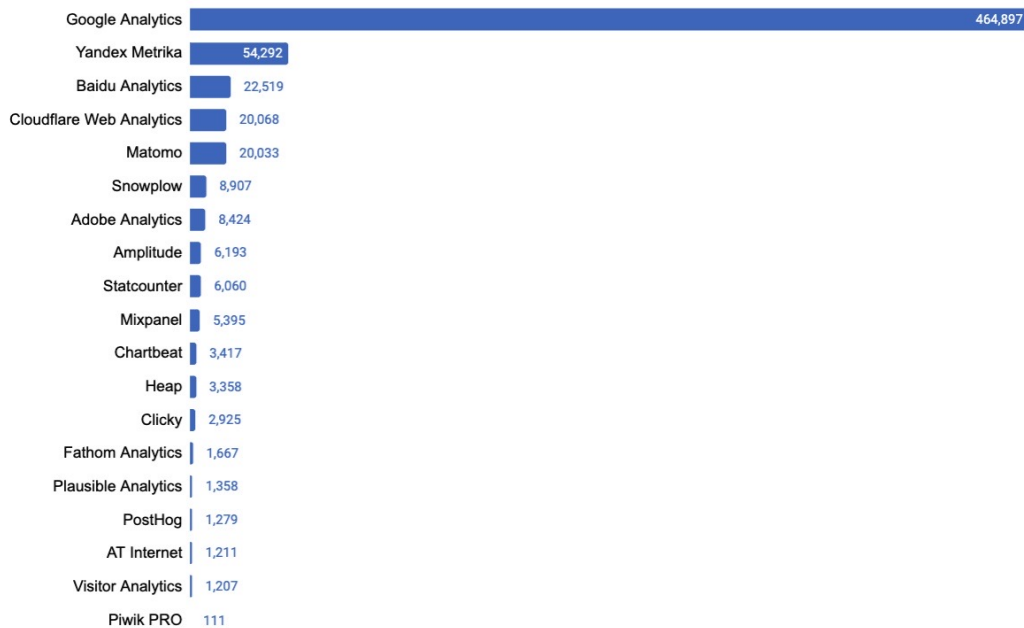
Market research tool BuiltWith (<https://builtwith.com>) crawls millions of sites to track what tags are being used. From this data collection, BuiltWith reports that GA holds approximately 85% market share in the "visitor count tracking" category and is deployed on around 50% of the top one million websites.

We'll frequently reference the BuiltWith numbers (with permission) throughout this guide. Other services, such as W3Techs, generally agree with BuiltWith's statistics (<https://w3techs.com/technologies/details/tag-googleanalytics>). The BuiltWith data we use was requested in January 2023, with data through November 2022.

The only other analytics tracking tool that comes close to GA's market share is the Meta Pixel (formerly the Facebook Pixel), which is deployed on 15% of the top one million websites. Although the Meta Pixel does include analytics reporting capabilities, I would not call it an alternative to GA based upon functionality and intent.

Here's a list of the number of deployments BuiltWith detected, limited to the top one million websites. I've introduced this restriction to reduce noise in the numbers and focus on websites that are more likely to actively manage their analytics and really use the tools in question.

Deployments on Top 1M Sites



*BuiltWith still labels some installs as Omniture SiteCatalyst, which may result in a slight undercount of Adobe's overall installation base.

†BuiltWith frequently detects Piwik PRO as being Matomo and therefore significantly underreports the tool. The tools have a shared history, covered in Chapter 15.

In evaluating alternatives, I drew a line in terms of popularity at 0.1% out of the top one million sites. This means we'll skip over alternatives with less than roughly 1,000 deployments in that dataset. This excludes some relatively well-known products like [Simple Analytics](#), [Woopra](#), [Gauges](#), and many region-specific options. I exclude tools such as [Pendo](#) and [FullStory](#) that are focused on user experience and include minimal traditional web analytics functionality.

Is your perfect alternative out there but not yet popular enough to be on our list? It's possible, but we are probably covering a tool that is similar — and there are downsides to picking something without a significant user base, as we'll talk about later.

Another requirement for inclusion in this guide is having a self-service trial. Since I only intended to perform product research, it would've been a waste of time for myself and the vendor to go through an extended sales

process with no hope of an actual sale. If there was no trial readily available, I skipped the product, which excluded Adobe Analytics and AT Internet (Piano) from our list.

I made an exception to the 0.1% rule for Piwik PRO, which is underreported in BuiltWith and may have over 1,000 deployments. Piwik PRO self-reports 250+ clients as well as 800-1000 new free accounts per month, likely putting them well above our threshold.

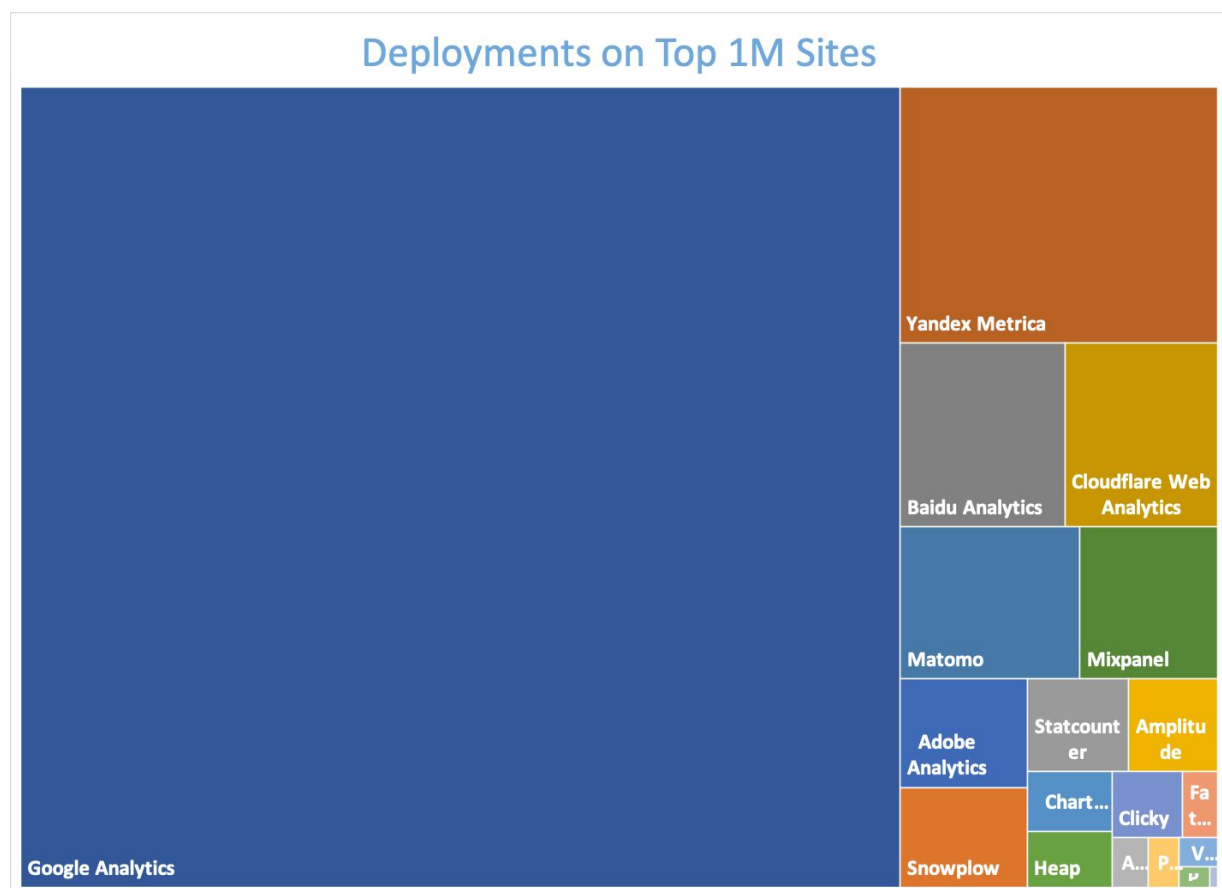
If you're looking at some of these numbers and thinking they seem low, consider the following:

1) Our dataset is just the top one million sites. BuiltWith detects 37.5 million total sites with GA, so we're working with a small subset (1.5%) of all installations. Additionally, the "top one million sites" includes many domains that would never run analytics: CDNs like akamaiedge.net, advertising domains like doubleclick.net, DNS domains like gtld-servers.net, redirection domains like t.co, etc. BuiltWith bases its list in part on the Tranco list (<https://tranco-list.eu>) which itself is a meta list of rankings. This methodology means we can accurately compare the popularity of individual products, but not market share per se.

2) False negatives are more likely than false positives. When BuiltWith scans a site, it's unlikely to identify GA when it's not there. However, the scan may miss cases of unusual implementations, sites that block their crawler with anti-bot protections, sites that require login, etc. Meaning our alternatives probably have more installations than BuiltWith can detect. For example, Statcounter claims to be on more than 1.5 million sites on their FAQ page, but BuiltWith only detects approximately 500,000 in their full dataset.

Additionally, the more customizable the tool, the more likely BuiltWith is to miss it. Tools like Snowplow that offer greater customization than GA have a higher chance of being undercounted. Finally, BuiltWith is also likely to miss many installations set up to bypass ad blockers, since those have been tailored to avoid programmatic detection.

I'm not exactly divulging industry secrets by saying GA is ubiquitous, though it is surprising the extent to which GA is so far ahead of its competitors and remains so in nearly all market segments. GA is installed on roughly twice as many of the top one million websites than all GA alternatives *combined*. For additional context, I think of Adobe Analytics as being a top rival to GA, but GA is installed on 263 times the total number of sites as Adobe Analytics.



In digital analytics circles, tools like Adobe Analytics, Matomo, Amplitude, etc. are considered direct competitors to GA. For many years, GA vs. Adobe was thought of as the Coke vs. Pepsi of the analytics world (at least in the U.S.). Measured in terms of overall deployment figures, Adobe doesn't appear to be anywhere near GA. However, number of installations is only one measure of market share, and GA does not have the

same kind of overwhelming lead in mindshare — which is much harder to quantify.

The obvious reason for this install base discrepancy is we're comparing the paid and free versions of GA together against only the paid version of Adobe (there's no free version). Scanning a website externally can't reliably determine if it runs GA360 or free GA, so a true apples-to-apples comparison of just paid versions is impossible. If we segment our BuiltWith numbers to the top 10,000 sites, we see a noticeably different market picture, with GA on 61% of sites and Adobe on 8.5%. **GA is still far ahead, but Adobe is relatively more popular among the top sites.** It's not just free vs. paid though; GA has a huge market share lead on everyone.

This highlights how many different types of markets GA operates in. It's run on sites with huge amounts of traffic at a frequency similar to tiny sites with virtually no traffic. Additionally, many GA installs are simply not used, no matter the size of the site. By focusing on the top one million I am trying to focus on those that are actively maintained, but I've found untended GA installs at nearly every level of site traffic.

As mentioned, **we won't evaluate Adobe as a GA alternative.** In addition to the lack of self-service trials, Adobe is already well known, with many existing comparisons between it and GA, like this one:

<https://www.qa2l.com/blog/202-google-analytics-vs-adobe-analytics-data-collection>.

As we can see from the deployment stats, the two biggest GA market share competitors are Baidu (Tongji) Analytics and Yandex Metrica. Both are very similar to UA; think of them as the Chinese and Russian versions of GA. However, **we will not be evaluating Baidu Analytics or Yandex Metrica as possible alternatives for political and privacy reasons.**

Yandex Metrica has gained significant market share in the U.S. in recent years, but companies should avoid it due to concerns about data sharing with the Russian government. In March 2022 the *Financial Times* reported that, in the first month of the Russian invasion of Ukraine, more than 2,000 apps added Yandex tracking, including apps specifically designed to target

Ukrainian users and VPNs. (<https://www.ft.com/content/c02083b5-8a0a-48e5-b850-831a3e6406bb>, archived at <https://archive.ph/HQZzq>).

BuiltWith lists nearly 300,000 sites in the U.S. running Yandex Metrika, making it more popular than most of the alternatives we'll discuss. I encourage those sites to remove the Yandex tracker.

After excluding Adobe, AT Internet (Piano), Yandex, and Baidu, that leaves us 14 alternatives to evaluate (15 counting GA4).

We'll evaluate all of these tools individually in the second part of this guide, but first, we'll dig into the technical details of the questions we might ask as part of our evaluation process.

CHAPTER 3 - THE GOOGLE ECOSYSTEM

BEFORE WE TALK ABOUT alternatives, let's delve into what we're leaving behind.

Universal Analytics is a multifaceted, extensible product that can be tweaked to perform a wide range of tracking and reporting functions.

We should also remember we're not the only users of the analytics toolset, and that the complexity of UA can be daunting to new users. Although we may be comfortable with UA and know how to make the most of it, other users may not feel the same way. After working with the product for so long, it's natural that we lose perspective on it, but not all users have that familiarity and long history.

When we think about an alternative, we need to take into account all of its potential users and data, not just our personal use cases.

Imagine you're a small business owner who just launched a WordPress site. Your developers set up Google Analytics and then sent a product invite with no further instructions. You log in to GA and are greeted by 115 standard reports —with no idea where to start, what's useful, or how anything works. Simply finding the data you are looking for can be an insurmountable challenge.

Now put yourself in the shoes of a product manager for a large ecommerce site with millions of pageviews per month. You ask your analytics team to use GA to tell you about the performance of certain product bundles, segmented by campaign source and active Google Optimize experiment. You're running into sampling issues and frustrations

with GA data imports for calculating marketing ROI. You're outgrowing the same product that was much too big for the small business owner.

These two groups both use the same version of the same product (free GA), even though they have different requirements and pain points. The large ecommerce site likely has the budget for an enterprise-level analytics tool, but isn't paying for one because they can mostly get what they need from free GA. This is how we ended up with a tool that isn't ideal for either use case. It's also market-distorting, causing users to think something as powerful as UA *should* be free.

Analytics tooling shouldn't be a one-size-fits-all scenario, nor should we pick and deploy a tool before we know what we want it to do. Also, it probably shouldn't be free at levels beyond the most basic installations.

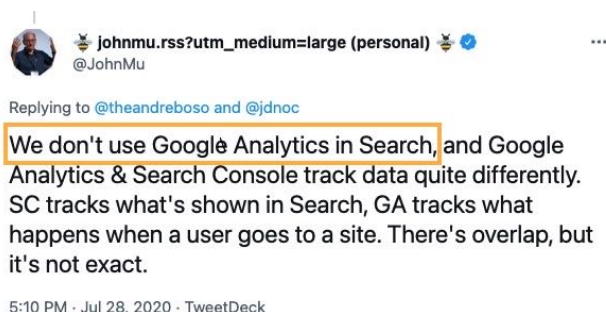
Looking outside GA can feel scary because of how deeply embedded the Google ecosystem is in many organizations. This is the biggest competitive advantage of GA, and something many alternatives simply can't replicate.

GA itself is part of the larger "Google Marketing Platform" (GMP), including Looker Studio (dashboarding), Optimize (A/B testing), Surveys, Tag Manager, and the ad tools (Campaign Manager, Search Ads, Display & Video).

Additionally, GA can link to Google tools outside the GMP: Google Ads, Ad Exchange, Google Ad Manager, Search Console, AdSense, Google Play, Merchant Center, and BigQuery. That's a ton of integrations! It's probable that more than one of these is critical to your business, which is a huge advantage for Google.

You're not required to use different parts of the platform together, but there are many integration benefits if you do.

One point of integration that definitely doesn't exist is between GA and Google Search results rankings. In other words, **there is no SEO benefit from running GA**. There's long been a rumor that Google uses things like bounce rate from your GA to inform Google rankings, but Google has denied this many times. Here's a Tweet from Google's Search Advocate John Mueller making it very clear:



(<https://twitter.com/JohnMu/status/1288220268155076611>).

Using GA data in their search engine might seem like a smart move to improve Google search relevancy, but let's think about how poor of a SEO signal that would be. It would only apply to 50% of sites (those with GA installed), and the quality of the data would depend on the quality of the GA implementation. If they did use bounce rate from GA, every SEO professional would try to game that signal by adding an interaction event on every page that caused their bounce rate in GA to be 0%. This bounce rate manipulation could even occur unintentionally; plenty of implementations have artificially low bounce rates due to events.

The Yandex search engine code leak in January 2023 (an outstanding analysis by Michael King available here: <https://searchengineland.com/yandex-leak-learnings-392393>) indicates that **Yandex Search does use Yandex Metrika** data as a search engine ranking factor. The full list of over 20,000 ranking factors shows dozens based on on-site user activity from Metrika (<https://yandex-explorer.herokuapp.com/search?q=metrika&o=all>). Google is not Yandex, and has been subjected to many more attempts at reverse-engineering by creative SEOs over the years. If data from GA affected rankings in any significant way I expect this would have been uncovered years ago.

Another less obvious integration is the data integration with Google Signals & DoubleClick that enables the demographic reporting features within GA. This type of information is largely unavailable from other alternatives because they lack an advertising network and unifying login system necessary to generate demographic data. This kind of integration could be done with other analytics tools and third-party demographic information providers, but I advise against it. In addition to the expense of acquiring the data, it's a serious privacy liability and, ultimately, the reporting it drives is typically not actionable.

Adopting an alternative will make integration with the other parts of the Google ecosystem we mentioned more difficult. Many options offer API integrations to some Google services, but before committing to another tool, you should enumerate those points of integration and how mission-critical they actually are.

For example, several of the alternatives we look at offer the ability to pull in data from Google Search Console (GSC), but since that information is also available from GSC and can be pulled into dashboards directly from there, it may be unnecessary to do an integration. Compare this to the pricing data from Google Ads, which is useful for many users but is impossible to pull into other platforms with the same granularity that Google Analytics can achieve.

CHAPTER 4 - IT'S TIME TO CONSIDER ALTERNATIVES

AFTER ENUMERATING ALL THE functionality that is hard to replicate outside of GA, you may be thinking, do we really want or need an alternative? Can't we just upgrade to GA4 and be done with it?

Since you're reading this guide, you likely have some questions about GA4 and whether or not it's the right choice for you. As stated, one of the primary points of this guide is to **stop looking at GA as the default choice for all circumstances**. GA should be viewed as one potential solution among many other valid selections, not the default choice installed without consideration.

Is GA in trouble? Do we need to jump ship off of a sinking platform? Following the analytics news over the last couple years might lead one to believe there's a kind of mass migration happening from GA to something else.

Google Analytics declared illegal in many EU countries

<https://noyb.eu/en/update-further-eu-dpa-orders-stop-google-analytics>

A major GA4 backlash in the analytics & SEO communities

<https://www.searchenginejournal.com/google-analytics-4-backlash/411392/>

Questions about the accuracy of GA, from legitimate concerns about the quality of geolocation data in GA

<https://www.analytics-ninja.com/blog/2022/06/apples-safari-15-5-impact-on-geolocation-analytics.html>

to widely repeated false stories about how GA is blocked in Safari

<https://appleinsider.com/articles/20/06/22/safari-now-blocks-google-analytics-on-sites-new-privacy-report-feature-shows>

While many people in the market appear to be searching for something else, **GA's market share has not decreased in a major way**. It has shrunk slightly, which is a notable event: Due to its huge market footprint, while the percentage lost is not big (6% loss from peak according to BuiltWith, 3% according to W3Techs), it represents a market share as big as or bigger than any of the alternatives we'll evaluate.

For example, if GA lost 4% of the top one million websites, that'd be 40,000 sites — nearly double the market footprint of Cloudflare or Matomo, our two most widely deployed alternatives. So, while this may not end up being a major blow to Google Analytics, it's a huge opportunity for the rest of the market. It's also worth noting that we are only measuring the presence of the GA tag, so sites that have migrated away from GA but continue to run its tag for backup or legacy reporting would still be counted as GA users. We're not measuring how a site uses GA, just that the tag exists.

My experience with GA "alternatives" has usually meant another analytics platform run in parallel with UA that has some duplicative functionality rather than a total replacement of GA. For example, running both GA and Amplitude, with the marketing team using GA and the product team using Amplitude, or employing both GA and Adobe and having different reporting packages run against each tool. This is frequently necessary but, as anyone who has worked on sites with multiple "single sources of truth" can attest, it can also be a big pain point for an organization. So, if it's possible to replace GA completely, great, but running multiple platforms (especially during a transition) is a viable way to move forward as well.

To be a true "alternative," a tool should be able to replace GA in a significant percentage of use cases. That doesn't mean everyone using an alternative is going to dump GA — in fact, the majority of sites likely won't — but they *could*. As mentioned, in some lists of alternatives you'll see things like session capture tools (e.g., HotJar or Clarity), SEM tools (e.g.,

Ahrefs or Semrush), log analyzers (e.g., GoAccess or AWStats), or platform-specific analytics (e.g., WordPress stats or HubSpot Analytics). These are not alternatives to GA in any broad sense, and their inclusion in articles about GA alternatives further complicates an already complex situation. We're focusing on true alternatives, even though those other tools do cover functionality that exists now in GA.

Let's discuss the three major issues driving users away from GA. To better classify our options, it's important to understand the cause behind this search for alternatives.

Why Are Users Leaving Google Analytics?

Reason One: Google Analytics 4

This is the easiest reason to understand. Google is forcing all users off of Universal Analytics and onto GA4 by July 1, 2023. Paid GA4 users will have one additional year. After that date Universal Analytics will stop ingesting new data, and then six months or so later Google will remove access to historical data. Any forced upgrade will cause a user base to consider whether or not the product they have is right for them, but Google has made this more painful for themselves and their users than I think was necessary.

GA4 isn't backwards compatible, which means it'll be a big change for most. We'll evaluate GA4 more in the second part of this guide, but anyone who's looked at it can tell it isn't a simple upgrade.

Since this new version is GA "4," we should briefly review and identify the previous versions of GA. It can be confusing, because previous versions weren't given version numbers at the time. So, GA4 is the first to use a new data model and the first with a number in its name (although Universal Analytics is sometimes now referred to as "GA3"). "Universal Analytics" is a slightly more inclusive term that Google occasionally uses to mean "any

version of GA previous to 4", so we will prefer the term "UA" unless we are referring to a specific historical version.

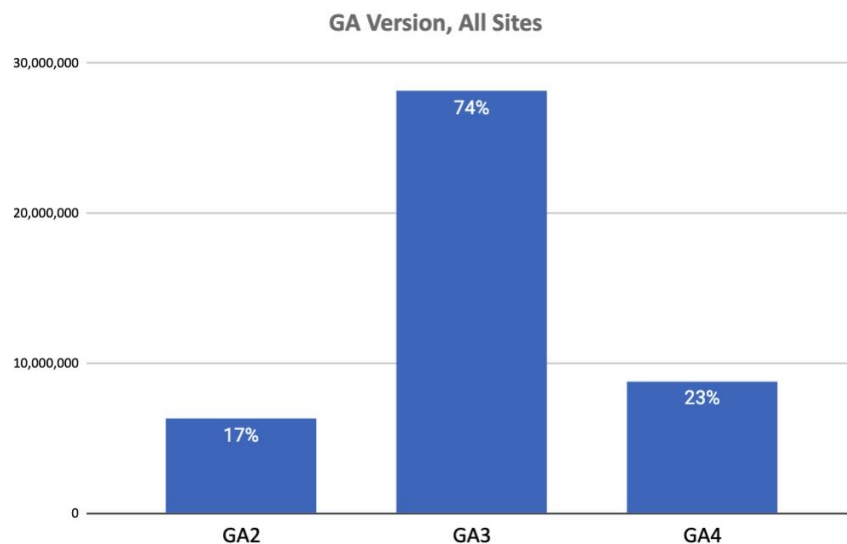
Version	Web Tracker	First Released
<i>Original Measurement Framework</i>		
"GA1" Urchin	urchin.js	2005
"GA2" Classic	ga.js	2007 (async in 2009)
"GA3" Universal	analytics.js gtag.js	2012 2017
<i>Next Gen Measurement Framework</i>		
GA4	gtag.js	2019 (as "App+Web")

Google has encouraged users to update tags with each new version, but previous upgrades were backwards compatible. This means you can still run urchin.js on your website and collect data into the same property as your modern gtag.js implementation. If you ran urchin.js you would miss out on a huge number of features, but it would still work, and the data would be contiguous. You could have a GA property running from 2005 to now with active data in the same single property since then.

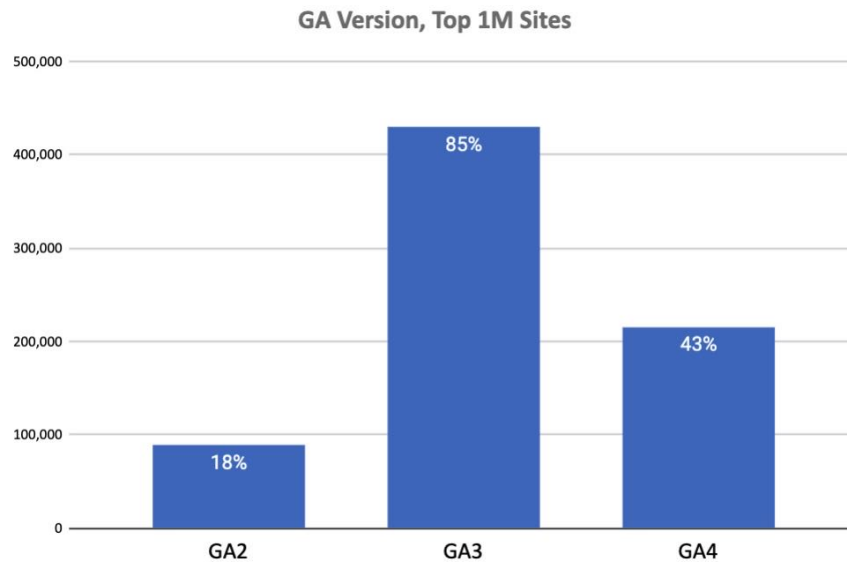
GA4 breaks this backwards compatibility and doesn't even import the old data into the new property. It's understandable why people who may have run GA consistently for 15 or more years are upset they can no longer do so in the same way.

However, it's also understandable why Google would want a blank slate that allows them to get rid of all that legacy technology and data. The end of UA does feel abrupt, but gtag.js, which is compatible with both GA3 and

GA4, was released five years ago. So, although Google has encouraged people to update their tagging for some time, many sites have not done so.



Looking at the sites BuiltWith detected as running GA, we can hone in on the version implemented. As of January 2023, GA3 is still by far the most deployed version of GA. There are even many GA2 implementations still lingering around, even though GA3 replaced GA2 a decade ago. The situation improves for GA4 when looking at the top one million sites, where 43% have implemented GA4.



We can also see from these numbers that dual-tagging with both GA3 and GA4 is common, which is how the numbers add up to greater than 100%. In March 2023 Google will start automatically creating GA4 properties for UA who have not yet done so (<https://support.google.com/analytics/answer/12938611>), though many sites that use analytics.js will still need to re-tag. **It's hard to convince people to upgrade, especially when the old version continues to function.**

Considering Google runs free data collection for roughly 50% of the internet with unlimited historical reporting, you can understand why they might want to reset that. With Universal Analytics, they have to provide that collection in a way that was not designed for their infrastructure. Not to mention the liability and compliance issues involved with sitting on that mountain of historical data.

Google made an attempt to reign in that data volume with their change in data retention policies for UA in 2018, which they have curtailed further with GA4. Full event data is now only retained for a max of 14 months in free GA4, though standard reports will still allow for unlimited lookback.

Universal Analytics traces its history back to the late '90s with the first versions of Urchin, the software that Google bought and turned into the first version of GA. I expect very little code and architecture from those early

versions of Urchin exist inside Universal Analytics, but the data structure and pageview-oriented reporting still work within the same paradigm that existed in the early 2000s with Urchin 4.

Anyone who's worked in software development for any amount of time knows how hard legacy and tech debt is to remove. This legacy stares us in the face every day with *UTM* parameters ("UTM" originally stood for "Urchin Traffic Monitor") and *UA-xxxx* property IDs ("UA" stood for "Urchin Account").

From an infrastructure perspective, we should also consider how much harder it must be for Google to store and access data with a legacy system retrofitted for Google's modern infrastructure, rather than a system specifically designed for Google Cloud Platform (like GA4).

I feel I can discuss Urchin with some confidence since I began using it in 2000 when it was still a log file analyzer and continued its use well after GA replaced it (I used Urchin until 2016; I'm not sure if I should be proud of that or not). If you're interested in the history of Urchin and how it became Google Analytics, there's a great history of the company written by one of Urchin's founders, Scott Crosby, here: <https://urchin.biz>.

Not only does GA4 make users start over with a new property and different paradigm, the rollout and timing of the forced migration has left GA users feeling stranded. GA4 is, in my opinion, a decently good piece of analytics software. If it was named something else (like Google Product Analytics) and not presented as an upgrade to UA with forced migration, it's likely people would think it was a good tool, or at least one that had a lot of potential.

So, enough complaining about GA4. For the purpose of evaluating Google Analytics alternatives, it's enough to know that GA4 is starkly different, with no automatic migration path.

Reason Two: Privacy

Another reason GA users are looking elsewhere is that people don't trust Google when it comes to privacy. This is sensible, as Google has an

astonishing amount of data about users across its broad platform of offerings.

That being said, GA as a standalone product has surprisingly limited data:

- GA works largely in aggregated data rather than individual users.
- GA doesn't permit personally identifiable information (PII) to be stored and will actively remove it if they find it (though they don't excel at this identification).
- GA doesn't provide access to user IPs and restricts storing IPs.
- GA doesn't capture data with high granularity. This means it doesn't support actions like screen capturing, mouse movement, dwell time, etc. It can do some of these if customized, but not by default.

This is less invasive than many other analytics products and far less so than most ad networks. The issue, of course, is that GA is part of Google, the largest ad network in the world. Google made up an estimated 44% of all digital ad spend in 2021. Not the utter market share dominance that GA has, but they're up against stiffer competition with Facebook and Amazon.

When you include Facebook and Amazon, those three combined make up 74% of digital ad spend (<https://digiday.com/marketing/the-rundown-google-meta-and-amazon-are-on-track-to-absorb-more-than-50-of-all-ad-money-in-2022/>).

It's easy to improve privacy on GA, though in my experience working with thousands of installations most do not. It's impossible to sever GA from Google completely, but turning off Google Signals and the advertising features in GA reduces the amount of data sharing. Anonymizing IPs is also a useful privacy option to turn on and is the default selection in GA4. Note though that anonymizing IPs doesn't hide a user's IP from Google.

It's also possible to sanitize some of the data sent via the GA JavaScript sensor before it's sent to Google. I've written about this previously (<https://www.quantable.com/analytics/do-not-track-for-google-analytics/>), describing which measurement protocol fields and cookies are potentially

problematic from a privacy perspective. Nevertheless, I don't recommend this approach, as heavily customizing GA opens up the possibility for errors and introduces significant ongoing maintenance.

So, we can somewhat lock down GA, but not completely. Combine this with Google's spotty privacy record and the number of regulatory actions against the company, and it's no wonder people would look for something that has privacy as a core feature.

Any discussion of privacy in analytics tools should also talk about price.

GA is free to a high level of usage — 10 million hits per month or more in UA, with indefinite data retention and reporting available. To compare, 10 million hits with Matomo Cloud would cost \$1,790 per month. Google has amazing infrastructure and economies of scale on their side, but GA still requires significant resources to support, and we should consider what they might get out of that investment.

The reason for this effort on Google's part is the benefits GA brings to their advertising business. For all their other activities, Google (and parent company Alphabet) is unquestionably an advertising company; they make 80% of their revenue from advertising (https://abc.xyz/investor/static/pdf/2022Q1_alphabet_earnings_release.pdf?cache=d9e9d97) for a total of over \$200B of ad-related revenue in 2021 (<https://www.statista.com/statistics/266249/advertising-revenue-of-google/>).

GA and the greater Google Marketing Platform (GMP) help to support the ad ecosystem. In fact, Google seems to consider GMP part of the advertising product line. Alphabet's 95 page investor report (10-K: https://abc.xyz/investor/static/pdf/20220202_alphabet_10K.pdf?cache=fc81690) doesn't even contain the word "Google Analytics," and the only reference to GMP is:

"Our customers generally purchase advertising inventory through Google Ads, Google Ad Manager, and **Google Marketing Platform**, among others."

I say this not to criticize Google, but to emphasize that Google is in the business of selling ads, not analytics software. You can certainly pay for GA, but that's not a significant revenue stream for Google. GA exists as a free product to help the industry spend more money on advertising in general and to help those ads perform better.

Everyone loves getting something for free, but when I think about free GA, I think of my site visitors' data as subsidizing the cost of the tool. Google, meanwhile, would look at it more in the context of improving Google Ads sales and efficiency.

Ultimately, Google has stifled competition in the web analytics market by offering so much functionality for free. This has made it incredibly difficult for any alternative (outside of the large enterprise space) to compete with GA. With UA, Google has gotten everyone used to an incredibly high level of functionality and performance at no cost.

If you want strong customer privacy to be part of your analytics platform, you may have to give up that subsidy and pay for your analytics tool.

Reason Three: Legality and Compliance

Several countries in the EU — including Austria, Denmark, France, and Italy (thus far) — have declared the use of GA to be in violation of the GDPR because of data transfers to the United States. At the time of writing this, these decisions lack any major penalties and give websites 90 days to comply (which means removing GA because there's no sure-fire way to make UA compliant). So, even if you run a website in one of those countries, you're probably not in immediate trouble.

BuiltWith tracks 823,571 websites in France running GA, but as of August 2022, the French data protection commission CNIL has explicitly named only three sites. Also, these rulings concerned UA, not GA4, so upgrading provides another level of insulation. However, GA4 has some of the same core issues that ran afoul of GDPR, so an upgrade likely will not

meet full compliance (read <https://cunderwood.dev/2022/04/22/eu-focused-privacy-controls-come-to-ga4/> for more information).

These compliance issues are convoluted and operate in an intricate space that overlaps both technical analytics engineering and international law. I find it unreasonable to expect the analysts and engineers responsible for tool evaluation and implementation to also evaluate the legality of their toolkit, especially considering how quickly the landscape is changing. Analysts aren't lawyers, and organizations shouldn't depend on our interpretation of the legal landscape for compliance. An organization's legal decisions should be informed by their analysts' expert knowledge, but the onus for compliance should remain on the legal team.

With the complexity involved to figure out where your particular jurisdiction stands on the legality of GA and how you should configure it, if compliance is of serious concern, then there is strong incentive to switch to something else. GA's competitors know this, and many have focused on privacy and compliance in their features and marketing.

So far, actions in the EU — stemming from the initiative of the privacy advocacy organization NOYB ("none of your business") — have focused on GA and the Meta Pixel, in particular their data transfers to the United States. This doesn't mean switching to Adobe Analytics or something else U.S.-based would make a site compliant (the Schrems II decision indicates it wouldn't), but it *does* mean alternative tools are not (currently) under such scrutiny. If there is a solution for U.S.-EU data transfers, then these particular compliance issues will go away. My hope is that, by the time you read this, there is some kind of agreement that makes this issue clearer and allows transfer of data. Unfortunately there are some serious differences in perspective that makes this agreement difficult. A new US-EU privacy shield agreement does not make GDPR compliance issues go away, but it does probably render the basis of these GA legal decisions moot.

However, concerns about compliance extend well beyond the EU. Although the EU is the leader in this arena, much of the rest of the world is following after in some fashion or another. We'll talk in more detail about compliance issues throughout this guide.

There are certainly reasons beyond the three mentioned that someone should consider GA alternatives, but these are the most commonly cited. Primarily, we should avoid viewing GA as the default and only start looking elsewhere upon facing an intractable issue with GA.

CHAPTER 5 - TYPES OF TOOLS

BEFORE WE DIG INTO individual products, I want to go over many of the types of questions we'll ask in the process of our evaluations. These questions and discussions will focus more on the technical architecture and paradigm that a tool fits into rather than specific functionality details. For example, we'll talk about implicit vs. explicit event tracking and what that means instead of how auto-tracking works within a particular tool.

The first and most important question to answer is what category of analytics tool we need.

As discussed, UA has features that span many product categories, including web analytics, marketing analytics, product analytics, app analytics, and application performance monitoring. GA4 has also pushed the product further toward the product and app analytics segments and away from traditional web analytics.

We're evaluating three categories of analytics tools, so let's define what those are.

1. Web Analytics

A) Traditional Web Analytics

Universal Analytics is not only the baseline example for this category — for better or worse, it defines it. Our definition of "web analytics" here is the aggregate measurement of websites, focused on both traffic acquisition and on-site behavior.

Traditional web analytics has focused on measuring sessions (based on cookies), identifying where that session came from, and pageviews

registered on the site. You're probably very familiar with this category already, so most of our other definitions will highlight differences compared to this category.

B) Simplified Web Analytics

This sub-category of web analytics represents a different approach than traditional web analytics, not just a more limited product. Good examples of tools in this category have made purposeful decisions on what to leave out. So what makes a tool fall under this as opposed to the traditional category? The simplified approach includes fewer metrics tracked, fewer options for customization, a more basic reporting interface, and less information stored about users.

This is a better choice as opposed to a traditional web analytics tool if you don't need all of the data a tool like UA provides. **Tracking more than necessary simply because you can is a poor approach to analytics.**

We all know UA tracks a lot of stuff, but did you know that it currently has 510 metrics and dimensions? (<https://ga-dev-tools.web.app/dimensions-metrics-explorer/>)

Many of these are specific to certain integrations, like metrics used in Google Ad Manager or Ad Exchange. Some are no longer being used in UA, such as the ISP dimensions (Network Domain and Service Provider). Others were simply ill-conceived or unnecessary to begin with, but once added, it's hard to take metrics away from the small number of users that might be using them. An approach that focuses on the essentials can trim this down to a much more reasonable number, which is advantageous for privacy, usability, and ease of implementation.

We see the complexity this depth of tracking options creates every time we log in to UA. Compare UA's 115 standard reports to a simplified tool like Statcounter's eight standard reports. This simplicity can be incredibly helpful for self-service users who aren't analytics experts.

Since tools in this category are generally much easier to create and manage, there are a plethora of options to choose from. We're analyzing six

tools in this group and, if none of them seem to fit your needs, there's likely another variation that does. Implementation in this category is relatively easy, and the costs are low, so you can test a couple of options and see what you like without a lot of effort.

2. Product Analytics

The other major analytics tool category is product analytics. This is the direction Google Analytics has taken with GA4. The difference between web and product analytics can be hard to pinpoint, because there's significant overlap, and many tools are not clearly in one category or the other. A tool like GA4, for example, has many product analytics-oriented features, but also has functionality that we'd generally think of as traditional web analytics. It's a continuum, but every tool we are evaluating can be fit into one of our defined categories well enough.

Here's a high-level comparison of what each type is most focused on.

Web Analytics	Product Analytics
Pageviews First	Events First
Sessions and Unique Visitors	Users, Customer Lifetime
Web-based	Cross-platform
Focused on inbound and marketing ROI	Focused on product usage and retention

You can certainly measure events with a web analytics tool or track mobile app data with your web analytics tool; it's a question of focus and perspective. Tools that label themselves "product analytics" such as Heap or Amplitude share functionality with traditional "web analytics" tools and, with the decommissioning of UA, have been incentivized to close that gap even further.

Are web analytics and product analytics converging under one umbrella? I think it is likely, especially since there is limited new

development happening within the traditional web analytics category and product analytics tools continue to add features.

3. Mobile App Analytics

While this category may seem self-explanatory, it can be a little confusing. Historically, mobile apps were measured with entirely separate tools than websites. You would run UA on your website and then something app-specific like Flurry, AppsFlyer, or Firebase on your app, and the two datasets would be completely separate.

Newer tools like Amplitude, Heap, or GA4 combine web and mobile app reporting into the same measurement and reporting framework. Since we're focusing on alternatives to UA, we won't look at mobile app analytics specifically. However, if you want to unify your mobile app and website analytics, you should consider a solution like many of the product analytics tools that are designed for both web and app.

What about "marketing analytics" or "digital marketing analytics"? Some people (including Adam Greco in a series of articles here: <https://amplitude.com/blog/marketing-analytics-vs-product-analytics-part-1>) prefer the term "marketing analytics" over "web analytics." I think of marketing analytics as a reporting function rather than a category of analytics tools. Calling UA or any other traditional web analytics tool a marketing analytics tool discounts its abilities in areas outside of marketing.

Ultimately, someone doing marketing is interested in a different kind of reporting than someone working on the product. These functions could in theory be separated cleanly between two departments, but more typically there is overlap. In smaller companies it can even be the same person doing both functions. These two separate functions might employ the same tool, but that tool has to be flexible if so.

If the overlap in features between web and product analytics is so confusing, why bother to categorize?

We need these top-level categories due to the differences in structure and usability based on who they are intended to be utilized by. In the previous chapter, we discussed how UA tries to be all things to all people, and the mess you end up with when using this approach. A tool geared towards marketing will unsurprisingly be easier and more conducive to accomplishing your marketing analytics goals, etc.

Additionally, many of the newer tools have opted to provide a blank slate and let the user create the reporting framework they need. This can be great if you have the team resources and desire for custom reporting and instrumentation, but it can be frustrating work if you have a simple website and are accustomed to UA's built-in reporting.

CHAPTER 6 - PRICING

AFTER FUNCTIONALITY, THE PRICE of a product is typically at the top of my (and many others') list of concerns. Due to the complexity of pricing and a general lack of transparency, it can be difficult to determine how much a product costs until you get pretty far along into an evaluation.

I've wasted significant time looking into "enterprise" products that don't publicly show prices only to discover an offering is totally out of my price range. Many vendors protect their rate cards like it's a trade secret, which frustrates and confuses potential buyers.

In this guide, I try to be as detailed about pricing as I can. I report any publicly available pricing and do my best to compare each tool's pricing model so you can understand how well they fit your budget and use case.

While it may seem like tools of a similar type should have comparable costs for the same use case, this is not always so and a tool's pricing model can make or break affordability.

Most pricing models are based upon a combination of traffic volume and feature limitations. However, the way that traffic is metered can make a big difference in how much the product costs depending on your site's pattern of usage.

Pricing Meter	Used in Category
Events	Product Analytics
Hits	Traditional Web Analytics

Pricing Meter	Used in Category
Pageviews	Simplified Analytics
Sessions or unique users	Less common

Hits and events are effectively the same in this context. But pageview-based or session-based pricing can lead to a significantly different monthly bill vs. hit- or event-based pricing.

Event and hit levels will vary as a function of the complexity of your setup. In other words, the same traffic can become more costly if you decide to track more on each page. Similarly, session or user-based pricing can vary based upon the usage patterns of your site. Thus the best matched pricing model for your use case depends upon how your users interact with your site and how you plan to instrument your tracking.

Here's a quick comparison to illustrate using the "Plus" plan from Amplitude (user-based meter) and Posthog (hit-based meter). Amplitude and Posthog are similar tools in terms of functionality, both falling into the product analytics category.

Site A, a B2B SaaS website with highly engaged users active on the site almost every day:

- 20k monthly users, average of 500 events per user (10 million total events)
- Amplitude base cost: \$240/mo. (user-based pricing)
- Posthog base cost: \$2,250/mo. (event-based pricing)

Site B, a B2C ecommerce site with significant traffic but lower user engagement:

- 100k monthly users, average of 20 events per user (two million total events)
- Amplitude base cost: \$960/mo. (user-based pricing)

- Posthog base cost: \$450/mo. (event-based pricing)

Both sites are in the same league in terms of overall traffic, but the best product choice from a budgetary perspective depends upon the way the site is used.

500 events per user might sound like a lot, but if you track multiple events for users who return frequently over a month, it's not an unusual usage pattern. Think about how many events per month a single user might create in their Facebook account — it could be thousands for an active individual.

If your traffic pattern is more "typical," picking a tool based on the pricing model may matter less, but it's still important to understand how the pricing model works and not focus solely on functionality.

Like most buyers, I'm a fan of transparent pricing and consider it a positive selling point. I don't want to schedule a call with an outreach associate, who then needs to loop in their regional manager, who has to sync up with their subject matter expert sales engineer to start working on calculating a price. I don't want to get stuck with overages because there's no clear rate card or a way to self-upgrade. I just want to know the price to see if it will fit my budget! But that number can be difficult to obtain, especially when there are many features and sub-components that may need to be licensed and sized independently.

I can understand how an enterprise installation quote might require time and effort to nail down a price. If you're implementing a tool across an organization with complex requirements and huge volume, then you'll need help from the vendor in figuring out what components you should have to meet these needs. But many times the pricing is kept private for reasons that solely benefit the vendor: acquiring a lead, being able to offer a "deal" to close a sale, being able to change pricing models without disturbing existing customers, etc. That's the vendor's prerogative, but it doesn't help me as a buyer.

In this guide, I refer to cases where the vendor does not publish the price but it is generally known as "reportedly" and mark the price in italics if it's embedded in a chart or table. For example, "Google Universal Analytics 360 reportedly starts at \$12,500 per month." As far as I know, Google doesn't publicize that price, but it's been widely circulated. I found prices on product review sites such as TrustRadius, GetApp, or Capiche — where users post the prices they are paying. I only report prices where I saw confirmation from multiple sources of a price range.

I report all prices in cost per month, regardless of whether or not a yearly contract is required. That means if something is only available yearly or has discounts for yearly agreements, that won't show in the prices I report. This may generate some inconsistencies between products that offer monthly discounts when purchased yearly, etc., so consider the prices listed in this guide as a starting point for informational purposes only.

CHAPTER 7 - EVENT TRACKING METHODOLOGY

OUR NEXT QUESTION TO tackle applies mostly to the product analytics category. If you're in the market for a simplified analytics tool, you can skip this chapter.

In terms of event tagging, there are two basic strategies tools follow: manual tagging and automatic event tracking. Sometimes this is called explicit vs. implicit, where explicit refers to manual tagging and implicit refers to automatic.

Which option sounds better to you?

Option 1: Manual Event Tracking (explicit)

Pro

You start with a blank slate: no events are tracked automatically (maybe not even pageviews). This allows you to construct the exact event structure you want and track only what you need. You avoid firing wasted events that nobody will use, which also improves reporting since only the metrics purposely constructed are present.

Con

It takes significant instrumentation work to gather "basic" data, then still more work when something new needs to be tracked. Getting over this initial implementation hump can be difficult and sometimes cripples a tool's adoption if the purchaser fails to grasp its necessary groundwork.

Option 2: Automatic Event Tracking (implicit)

Pro

Every event is already captured and ready to go. Each click, form submission, and pageview is already a tracked event waiting for reporting.

If the development team launches a new feature without letting you know, it's automatically reported with no new event coding needed. When a business stakeholder decides they want reporting on a certain element, you already have historical data and simply need to select which event to pull information from.

Con

It introduces a jumble of events you'll never need, making tracking more confusing. Since everything is supposedly automatic, you may come in without a plan and end up with something that doesn't answer the questions you need. If you are paying by event volume, you're paying for a bunch of events you don't use.

If you've used Google Analytics autotrack (an example of strong implicit tracking), you may already have a well-formed opinion on the subject. Personally, I strongly disliked autotrack, but some people liked it, at least for certain scenarios.

Even if you hated GA autotrack, you shouldn't immediately rule out all implicit tracking. Newer products like Heap and Posthog that are built around the concept of auto-tracking (rather than having it tacked on after, like GA autotrack) handle things in a more usable way. Click events are tracked and *available* to be "defined," but not used in reports until they're defined. So, there's no development work required to track something new, just setup within the tool UI.

This tracking approach can become an important deciding factor, especially when looking at product analytics tools that may have vastly different approaches.

Many of our options are not totally in one camp or another but rather somewhere in the middle of the explicit-implicit continuum. GA4, for example, can track a number of events automatically (pageviews, scroll, outbound clicks, site search, video engagement, file downloads) with the "enhanced measurement" configuration enabled in the web stream setup.

Product	Level of automatic event tracking
Google Analytics	low
Cloudflare Web Analytics	n/a (no events)
Matomo	low
Mixpanel	low
Snowplow	low
Statcounter	low
Amplitude	very low
Chartbeat	n/a
Heap	high
Clicky	low
Fathom Analytics	low
Plausible Analytics	low
Visitor Analytics	high (if auto-track is enabled)
Posthog	high
Piwik PRO	low
GA4	medium

CHAPTER 8 - SUPPORT AND COMMUNITY

AS ITS USERS CAN attest, GA customer support is nearly non-existent. GA does, however, have a tremendous community with knowledgeable users willing to offer solutions. If you pay for GA360, you'll likely receive excellent support from your GA reseller, but you still won't get much help directly from Google.

If you want to do something uncommon in GA, odds are you can find an article from someone else who has already done that precise thing and provided their solution back to the community.

When we move to a GA alternative, we lose that rich community. This is a major downside to consider when thinking about how you'll implement, troubleshoot, and support an alternative. Some alternatives have larger communities than others, but you shouldn't expect things like a vast library of how-to articles from third parties or lots of free Google Tag Manager (GTM) recipes.

Many of these alternatives do provide excellent support from the vendor. So, when evaluating products like Amplitude or Piwik PRO that have generous free tiers, you should consider what kind of support you'll need; you may want to pay for a support plan instead of counting on blogs and Stack Overflow.

In addition to losing the community, integrations with outside software platforms are scarcer outside of GA. Whether you are looking for integration for a chat widget on your site or pulling data via API into a third-party dashboard software, these connections are harder to do with a GA alternative and require more research.

Let's go through some examples of typical integrations.

Back-End Examples

When referring to "back-end" analytics integrations, I mean data pipelines to or from GA. Here are some example scenarios:

- 1) Analytics tool -> Funnel.io -> Google BigQuery -> Power BI
- 2) Analytics tool -> Looker Studio
- 3) Analytics tool -> Supermetrics -> Google Sheets

With GA, we take it for granted that nearly any third-party software or tool will have an integration. We can't make that assumption with an alternative.

When evaluating a GA alternative as part of a data flow pipeline, consider that we only need a connector from that alternative to the next step in the pipeline — although the alternative's data schema will affect the whole data flow. In our first data flow example above, we don't need our new tool to have a connector for with BigQuery and Power BI, just Funnel. But if we are replacing UA in an existing pipeline, we need to remember that the schema we get from the alternative probably won't match what we got from UA.

Let's look at how many of our alternatives can connect directly to these common third party tools with a easily available connector.

Funnel.io: 3 out of 15 (Mixpanel, Amplitude, GA4)

Looker Studio: 5 out of 15 (Matomo*, Mixpanel*, Plausible*, Amplitude*, GA4) \through a paid marketplace connector*

Supermetrics: 1 out of 15 (only GA4)

This may improve as other alternatives grow in popularity, but for now, the options are limited. It's likely you will be able to flow your data from a GA alternative into a data warehouse or visualization tool, it'll just be little harder. In general there is a lack of "out of the box" data connection functionality from third party tools compared to GA. For some alternatives this is irrelevant, because the alternative itself can connect to our data store.

Many of the product analytics tools connect to a wide range of back end databases directly. For example, PostHog can connect for free to: BigQuery, Postgres, Snowflake Redshift, S3, and others. This can obviate the need for a connector tool like Funnel or Supermetrics, but if your existing flow is dependent upon those tools that can still be a problem.

Any tool with an adequate API can connect to most destinations, and most of the alternatives in this guide do have APIs. Other third-party connector apps such as Zapier or Precog can function as additional links if you want a no-code solution. Also, tools like Funnel and Supermetrics will typically build custom connectors as an add-on service. Some of the options we're looking at (e.g., Cloudflare, Visitor Analytics) don't currently have a reporting API, which makes any sort of data pipelining impossible.

Front-End Examples

From WordPress plugins, to live chat integrations, to ticket-buying platforms — if it has an "enable third party stats" option, odds are it'll be Google Analytics. As an example, let's look at Gumroad, the platform on which I sell this very book. Gumroad allows me to add a GA property ID (for *either* UA or GA4). On the happy occasion where someone adds the book to their cart and checks out, it sends events for each step to GA, allowing me to see the whole checkout funnel.

Google Analytics Property ID

[Learn more](#)

(UA-1234567-12 or G-ABCD232DSE)

Google Analytics Primary Domains

yourdomain.com

Facebook Pixel

[Learn more](#)

9127380912836192

☐ Send 'Purchase' events for free (\$0) sales

Include other services' code snippets

[Learn more](#)

Choose products the code will be included for

Enter your analytics

Google Analytics Alternatives

Enter your analytics

Hold on, there's a slot for "other services" too! Does that mean I can track Gumroad traffic in my alternative of choice instead of GA? Not really, because that snippet is fired only upon a successful sale. This is useful for adding an ad platform conversion pixel, but it doesn't allow me to track pageviews, cart adds, etc.

I can't imagine Gumroad settings ever having a "select your analytics platform" dropdown with a bunch of tools listed. In a world where there's not one ubiquitous platform like UA, we'll need to rely on open standards more. Luckily there's already the start of one in the form of the data layer. The data layer is a JavaScript object that is part of a web page and acts as glue between that page and a tag manager. This had a start towards a standard in a W3C community group report from 2013 (<https://www.w3.org/community/custexpdata/2013/12/16/final-report-customer-experience-digital-data-layer-v1-0-ready/>), but was never turned into an official standard, and what's used today in GTM and elsewhere is a bit different from that original proposed standard.

If you've done much with Google Tag Manager you are likely familiar with the data layer, but it's not specific to GTM and could be used by any tag manager. If Gumroad had generic dataLayer event pushes built into their platform then we could implement any analytics platform we wanted via a tag management system (TMS) of our choosing.

This means your TMS could handle all of the tool-specific instrumentation work. You want to upgrade from GA3 to GA4? Redo the tags inside your TMS. You want to run both GA4 and Amplitude? Create new Amplitude tags in the TMS tied to the same preexisting events and triggers.

CHAPTER 9 - SHOULD I SELF-HOST?

GDPR, HIPAA, SOC2, CCPA, CPRA... OMG! These days, the analytics world is overrun with anxiety-inducing acronyms. Sometimes it makes you want to just host everything yourself and know for sure that you're in compliance rather than wondering if your vendor has the same interpretation of the rules as you.

Knowing the exact method and location of the data you store isn't always possible with cloud-hosted analytics, so hosting it yourself provides that control. There's no better way to be sure that you own your data than to not let it outside your infrastructure.

For example, historically GA made no assurances that EU user data collection would occur within the EU. As of April 2022, this has changed (at least with GA4) such that if you are in the EU, your data is collected in the EU. However, that data may or may not be processed or kept in the EU, making it likely inadequate for full GDPR compliance. Cory Underwood has covered this ongoing issue superbly in his blog here: <https://cunderwood.dev/2022/04/22/eu-focused-privacy-controls-come-to-ga4/>.

If we self-host, we know where the data is — it's on our servers. We aren't subject to the whims and vagaries of the vendor; we have control. The data may still be on a public cloud like Amazon or Google, but we're aware of its location.

There's a well-known saying that the cloud is "just someone else's computer." I have mixed feelings about this aphorism. On the one hand, I like it because it's a reminder that there's nothing magical involved in cloud hosting; it's all just a computer in a rack sitting in a data center somewhere.

However, I also dislike the saying because it neglects the differences in architecture and the massive benefits of cloud hosting.

Cloud hosting at its best implies your data will be hosted in a redundant way across multiple servers with transparent scalability and edge optimization (that is, doing as much as we can closest to the end user). So, if the end user is in Germany, we'd want them to connect to a computer in Germany and have the data processed as close to that point of collection as possible.

That's part of the cloud promise, when done well at least. For our purposes of comparison, the difference is who runs and pays for the servers: If you run the servers, it's "self-hosted"; if your vendor runs the servers, it's "cloud-based." An oversimplification, but accurate enough for our purposes.

Sometimes the tool vendor will use the term "on-premise" (or just "on-prem") rather than "self-hosted." When "on-premise" is used, there's an implication to me that the servers are *not* on a public cloud like AWS or GCP. I find this to be unnecessarily confusing terminology since, from the vendor's perspective, there's little difference.

In general, my recommendation is to let someone else manage the server unless there's good reason not to. Having full control over the tool is enticing, but typically more trouble than it's worth.

My perspective on self-managing may be colored by spending many years as the person who received the 3 a.m. texts when a server went down. Paying a little more to have that text go to someone at a different organization is money well spent.

So, what are the reasons someone *would* choose to be self-hosted?

Compliance

Sometimes you need complete control over the data to ensure you're following the rules and self-hosting is the only way to guarantee this.

Privacy and Security

Self-hosting doesn't necessarily mean better privacy, but it can. This is a double-edged sword, because when you self-host, you're also responsible for the security of these services (patching the servers, maintaining security groups, firewalls, etc.). There are certainly risks involved in trusting a vendor with your data but, unless your organization has a strong security team, the vendor will likely be more capable of handling security issues related to the hosting of their software.

Price

If you run a high-volume site, it can be significantly cheaper to self-host. However, do remember to consider the total cost of ownership (TCO), not just that of the servers alone (https://en.wikipedia.org/wiki/Total_cost_of_ownership#Computer_and_software_industries).

Let's look at a Matomo Cloud example for a site with 10 million hits per month.

	Matomo Cloud	Matomo Self-Hosted
Base monthly cost	\$1,790	\$0
Premium plugins (estimate)	\$0	\$100
Server cost	\$0	\$750
Total cost	\$1,790	\$850

In this case, for a reasonably high volume site, it seems we'd save more than 50% through self-hosting. Base Matomo costs us nothing, but there are paid plugins which add expenses (which ones you need will vary, so \$100 a month is just an estimate), as well as the cost of the servers. The \$750 monthly charge is based on a simple AWS estimate I made here:

<https://calculator.aws/#/estimate?id=0d725d270511e9d55d68f9f5e39ae5db4332661e> (where the majority of

the cost is in the MySQL RDS database and I exclude smaller costs like bandwidth).

Now considering the TCO, we can start with the fact that the \$850 monthly estimate doesn't include the cost of people. Who sets up the hosting? Who manages the backups? Who patches the servers and opens tickets with AWS when something goes wrong? With self-hosting, you take on plenty of additional responsibility. What looks like 50% cheaper may end up being just as expensive as cloud hosted when all costs are accounted for.

Control and Customization

Self-hosted means you can tweak elements of your installation in a way that most managed cloud hosting would not allow. Just like with privacy, this can come back to bite you if done incorrectly.

Additionally, the level of customization will depend on whether the product is open source or not, which we'll talk about in the next chapter.

Integration With Existing Infrastructure

This is a similar point to our last one about control and customization, but focuses more on communication and data transfer.

Do you need to tie something in your analytics stack to a core element of your infrastructure? With self-hosting, your analytics data can sit in the same network as your other data. This allows for faster (and cheaper) data transfers, safer, more secure data access, and other benefits. This kind of integration is not necessary for most installations, but when it is, self-hosting may be the best way to bring data sets close to each other.

CHAPTER 10 - OPEN SOURCE

I'VE LONG BEEN A proponent of open source tools. I've volunteered for open source projects and contributed code. However, open source isn't always the best option for your analytics tool. All things being equal, I recommend the open source option — but in many cases, there are other more important considerations. Moreover, many open source tools in this space are not *fully* open with the kind of community you may associate with famous open source projects like Linux, Apache, etc.

I encourage you to think of a tool being open source as an additional feature that may or may not be helpful to you. Supporting open source is a good thing, but tool selection is about more than picking the one that ticks the most boxes. How does a tool being open source help to meet your particular needs?

Let's dig into some of the common benefits associated with open source tools.

1. Price

Open source = no cost, right? Turns out this one is not as simple as it seems. We talked about server cost in Chapter 9, but servers are not the only additional expense. Many GA alternatives are "open core" software (https://en.wikipedia.org/wiki/Open-core_model), a sub-type of open source where the core of the product is open, but some features and add-ons (usually the ones you want) are closed source and commercial. On the business side, this has been a great way for open source vendors to make money and continue to improve their product. From the consumer's point of view, it means you need to be careful about what part of the product is free, both in terms of cost and access to the code.

2. Avoiding vendor lock-in

Some make the argument that "even if the vendor goes away, you won't lose access to the product" or "if the vendor drastically changes pricing model, we'll have options."

If we're talking about the major open source projects like Linux, they absolutely can outlast any vendor or contributor. However, the open source alternatives we look at in this guide — Snowplow, Matomo, Posthog, and Plausible — are all closely tied to their vendors, meaning vendor-sponsored developers perform the vast majority of work on the product. So, if the vendor goes away, the funding for the majority of developer time disappears as well.

3. "You can customize it to do anything you want"

Customization is much harder than it sounds. Contributing meaningful code to an open source project is no simple task. If you *don't* want to contribute back, forking your own version is a very bumpy road to go down. So, unless you have excess developer time and highly specific needs, customization alone is not a compelling reason to opt for open source.

Granted, there are other advantages to open source tools, including security, community, and verifiability. These are all "potential" benefits that may or may not manifest in any particular tool. A vibrant community of contributors *can* be an outgrowth of an open source project, but that doesn't guarantee it'll happen (in fact, it usually doesn't).

An Open Source Test Case: Fathom Versus Plausible

Let's analyze a case that is as near to "all other things being equal" as you're likely to see. Fathom and Plausible are very similar products, as we'll delve into more in the individual project evaluation chapters. While not exactly the same, they're probably the closest of any two products we evaluate, open source or not. One of the more significant differences is that Plausible is open source while Fathom is not.

You can download the Plausible source code from GitHub (<https://github.com/plausible/analytics>) and run it yourself totally separately

from Plausible. You could fork it and do whatever you want, short of using the code for your own commercial product.

How does this affect you as a user? For most use-cases, open source status won't impact you at all. Let's evaluate against our previously stated potential benefits:

Price - *Minimal difference*

As both products have a low price point, you wouldn't save much money and would take on some work if you decided to self-host your version. If you then modified your own version it would be *significantly* more work. In most cases it would cost you more to maintain a modified version considering the developer time required.

Vendor lock-in - *Slight advantage to Plausible*

Plausible's two co-founders are historically the main contributors to the tool's source code (<https://github.com/plausible/analytics/graphs/contributors?from=2019-09-01&to=2022-08-23&type=c>). Other individuals have committed code (including recently hired new employees), but it lacks a large community that could maintain the product in the absence of the vendor. If Plausible went out of business tomorrow, you'd be better off shifting to another product, as the Plausible code would likely no longer be actively developed. However, having access to the code would give you a few options to continue to run the product as is, though that's not without its own risks.

Customization - *No difference for most use cases*

Want to customize the product? Do you or your development team know Elixir? That's what the core of Plausible is written in, and it's an uncommon language. Elixir runs on the Erlang virtual machine, which is also not very widely used. These are, in my opinion, strong technologies to run an analytics platform on, but they make it difficult (and expensive) to find a developer. The project's language is not the only problem, but it's representative of the kind of issues you'll have when you customize the tool.

You'd also need to learn the structure and standards of the project, especially if you want to contribute your modifications back to the main

project. It's highly preferable to contribute patches back. Not only because it's good for the project, but it's also helps keep your version from diverging away from the main version. If you have diverged, it makes updates the main version gets harder to apply to your customized version.

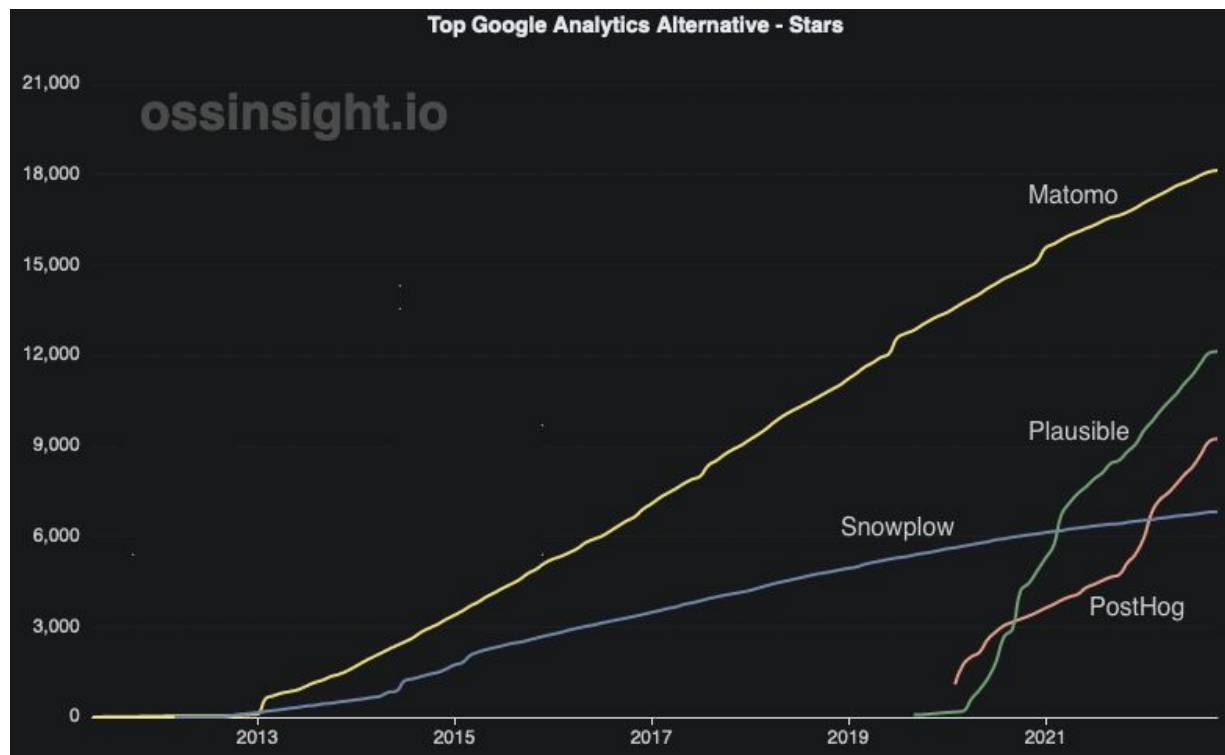
Whether or not this is an advantage depends on your use case. Let's say you run a large content-hosting platform and site builder, which we'll call "Trianglespace," and it has thousands of hosted sites. If you wanted to offer easy-to-use web analytics to all your customers on Trianglespace, you'd probably choose Plausible over Fathom so you could self-host at scale and customize it to your needs. The individual sites on Trianglespace might be small, but the operation overall would be large with a lot of technical capacity.

Most potential users will experience no benefit from Plausible's open source status compared to Fathom. I would put open source status even lower on the list of relevant differentiators when looking at Piwik PRO (closed source) versus Matomo (open source). Even though both products have a common ancestor in the Piwik open source project, they now have many differences in features and technology that take precedence over their open source status.

Since we've discussed install base as a measurement of market share, we should also talk about popularity within the open source community. This is a good way to determine if a project has stability and will continue to be maintained and updated.

The typical way to gauge popularity is through stars on GitHub, which shows how many users on that site have a project's repository saved to their favorites.

Matomo is the most popular, with a substantial lead over the other open source options we evaluate:



Graph from <https://ossinsight.io/collections/google-analytics-alternative/trends/?historical-trending=prs>. Although you can un-star something on GitHub, people generally don't, which gives a measurement advantage to older projects.

According to this graph, Plausible and Posthog are quickly gaining popularity, and both Snowplow and Matomo show steady growth.

These GitHub stats show another way to measure the popularity of a platform and help us gain a feel for the different projects. Snowplow has more deployments than Posthog or Plausible according to BuiltWith, but is less popular on GitHub, especially recently.

This difference is likely indicative of how Snowplow is used. It's an enterprise-oriented product, especially compared to Plausible, and individual enterprise users and deployments are less likely to interact with the product on GitHub. Additionally, Snowplow may be implemented globally on a large number of websites run from by a single organization. For example, about 50% of Snowplow's install base are sites that run on the

Weebly CMS platform. Developers employed by Weebly might be heavily engaged with Snowplow, but the individual Weebly sites may not even know they are running it.

As with the install base stats, popularity isn't everything, but it's something I look at when evaluating a product to predict how it may endure and improve over time.

CHAPTER 11 - HOW IMPORTANT IS TRACKER SPEED?

ONE OF THE ARGUMENTS you'll hear against GA is that removing it can speed up your site. The simplified tools in particular make this claim, sometimes implying their tool will improve your SEO. Since SEO is a very opaque practice, these kind of assertions are concerning and can be difficult to verify.

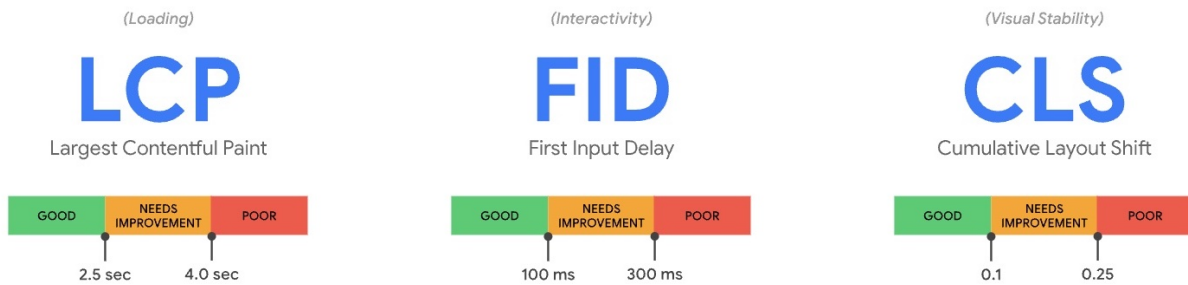
As we established in Chapter 3, using GA won't help your SEO — since Google doesn't use GA data for their search ranking algorithm. Conversely, GA won't *hurt* your SEO either, though the reason why is a more involved than "because Google says so".

The simplified tools may be faster than GA, but the amount of speed improvement (if any) is *very unlikely* to move the needle for your site with respect to SEO.

From a web performance perspective, **the metric Google Search claims to use for ranking is Core Web Vitals**. I imagine most of us are familiar with this term (if not the specifics of its definition), as it officially became part of the Google Search algorithm in 2021 (<https://developers.google.com/search/blog/2021/04/more-details-page-experience>).

Assuming it's implemented properly, Google Analytics has a near zero impact on Core Web Vitals.

Here's an overview of the components that make up Core Web Vitals, as provided by Google:



- Largest Contentful Paint (**LCP**): measures *loading* performance. To provide a good user experience, LCP should occur within 2.5 seconds of when the page first starts loading.
- First Input Delay (**FID**): measures *interactivity*. To provide a good user experience, pages should have a FID of 100 milliseconds or less.
- Cumulative Layout Shift (**CLS**): measures *visual stability*. To provide a good user experience, pages should maintain a CLS of 0.1 or less.

(from <https://web.dev/vitals>, content unchanged under CC license: <https://creativecommons.org/licenses/by/3.0/>).

If you look at the performance metrics Google chose for Core Web Vitals, you'll notice they don't include anything about the final completion time of a page. Google is focused on how long it takes users to see and start interacting with the page, not the total time for every element to finish loading.

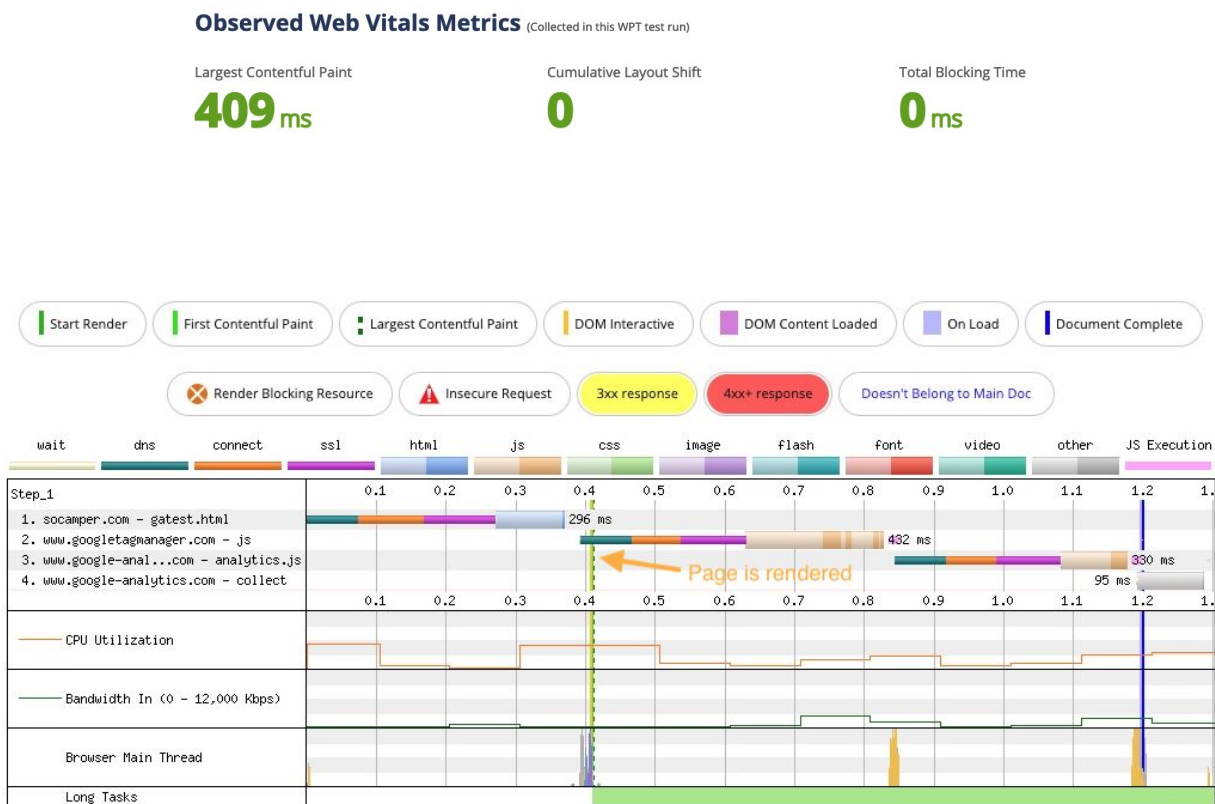
As long as a page renders its largest section and becomes interactive quickly, Google has no problem with additional late-loading content (provided this late-loading content doesn't shift the page around or eat up too much CPU). This is a reasonable way to measure things. However, it's also friendly to ads, which effectively leaves a hole in Google's rules

allowing ads to degrade the user experience without hurting Core Web Vitals scores.

In terms of the component metrics of Core Web Vitals:

- GA doesn't affect LCP because it's not a render blocking script.
- GA doesn't affect CLS because it doesn't insert any content into the page.
- GA *can* occasionally affect FID, but is typically limited to slow mobile devices or in conjunction with egregiously slow JavaScript.

Still not sure? It's easy to test. Create an HTML page with no components on it other than a stock UA tag. Then, run it through a performance testing tool such as [webpagetest.org](https://www.webpagetest.org). Here are the results:



with the full test results archived here:

https://www.webpagetest.org/vitals.php?test=221001_BiDcXK_617.

These performance waterfall charts can be tricky to read if you aren't familiar with them, but these are excellent results.

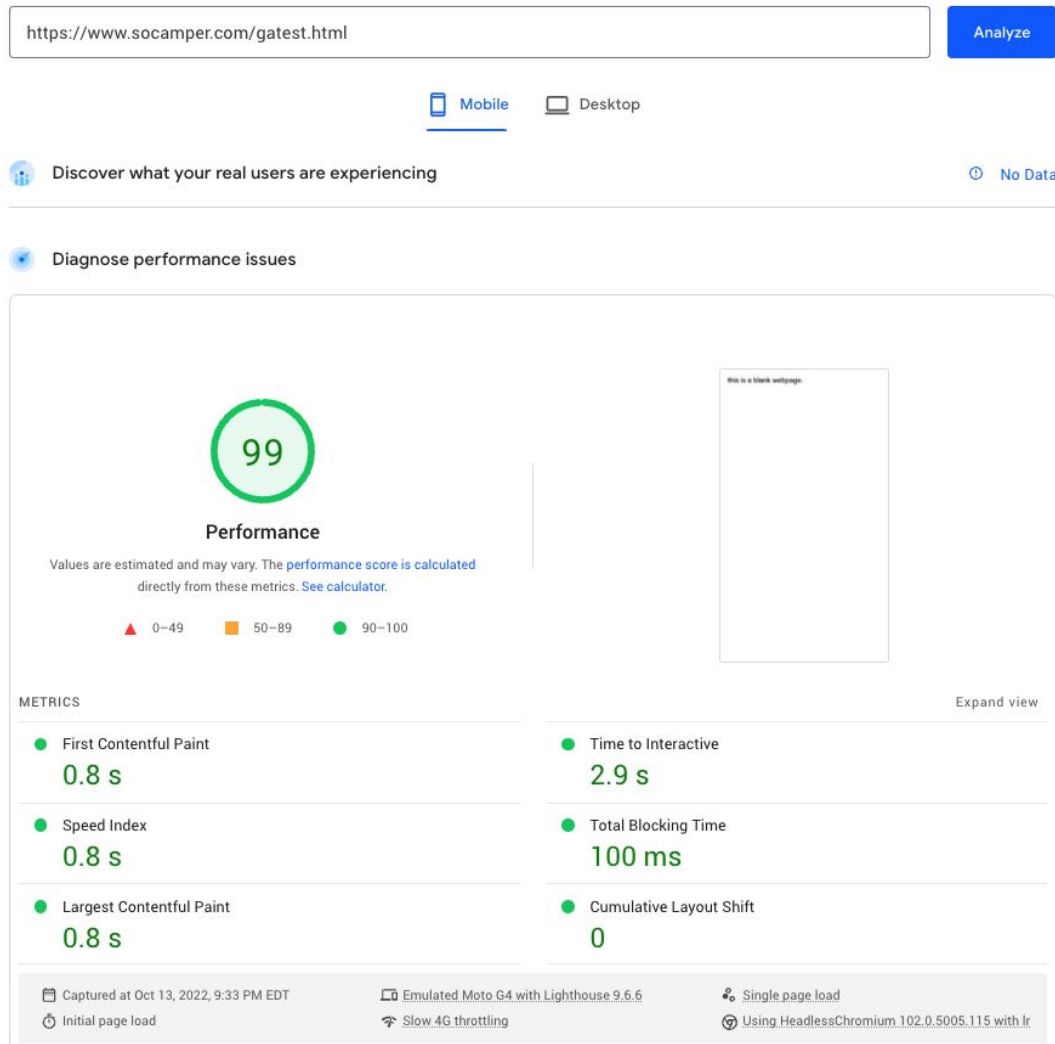
- LCP — 409ms (effectively the same as a page without the UA tag, which measured 417ms)
- CLS — 0 (meaning no content shifting)
- Total Blocking Time (TBT) — 0 (which means no FID is possible)

So, GA made no significant impact on Core Web Vitals. LCP (the dotted green line in the waterfall chart) came before the contents of gtm.js were seen by the browser, showing that the browser knew that nothing coming from that tag was going to be render-blocking.

Careful readers may have noticed we switched from talking about FID (First Input Delay) to TBT (Total Blocking Time). The two metrics measure the same thing, but in a different way. FID is a real-world measurement of how long an average user has to wait for the browser to realize they've done something (for example, if you're loading the amazon.com home page and start typing into the search box but have to wait to be able to type because the browser is still churning on page processing).

TBT is a measure of how long the browser has spent processing the page, unavailable to respond to a user interaction (<https://www.semrush.com/blog/fid-tbt-explained/>). There has to be blocking time in order for there to be input delay — so if there is zero TBT, there is no chance for a user to experience input delay. The critical point is that our trivial test showed no Core Web Vitals impact.

Then, why does Google PageSpeed Insights seem to cap my score at 99/100 if I use GA?



(from <https://pagespeed.web.dev>)

PageSpeed Insights defaults to testing via a simulated mobile browser with a very slow processor (an emulated Moto G4, a phone that came out in 2016 with a Geekbench CPU score of 112). My webpagetest.org performance tests were done with a simulated iPhone X (a newer phone with a Geekbench single CPU benchmark score of 907), a device with a processor eight times faster than what Google uses in their tool, and likely a more realistic representative for most sites. The most popular individual phone model GA currently reports for a typical US-based site is the Samsung Galaxy S21 (Geekbench single CPU score of 929). I picked the iPhone X because it was the closest to the Galaxy S21 CPU benchmark among the available options in the webpagetest.org tool. Average

smartphone lifespan is reported to be around 2.5 years in the U.S. (<https://www.statista.com/statistics/619788/average-smartphone-life/>), so a six year old phone is likely not representative.

The PageSpeed Insights lab tool takes a worst-case scenario approach by simulating a phone that's probably significantly slower than your site's average user. To be clear, the 0-100 score in PageSpeed Insights is not what Google uses as a ranking factor, and you can do well with a low score. For example, eBay.com's homepage scores a 38, but still passes the Core Web Vitals test and brings in plenty of Google Organic traffic.

Because Google bases Core Web Vitals on actual user measurements, your users' actual experiences are more important than a lab test. Assuming most of your users don't use very underpowered devices, the JavaScript from GA shouldn't impact FID. GA *can* exacerbate an existing JavaScript problem though, so if you have a ton of JS already, running GA adds that much more processing, but it's a marginal effect. It could be the straw that breaks the performance camel's back, but if that happens, you have bigger JavaScript issues than GA alone.

So, I give GA a clean bill of health for SEO. This doesn't hold true for all our alternatives (especially Visitor Analytics, whose heavy JavaScript payload could negatively affect SEO), but most of them will have little to no effect.

We may have busted another SEO myth, but we can't let analytics tools completely off the hook for performance.

Google's official measurement doesn't care about full page load times for SEO, but the end user can have a bad performance experience if we overwhelm the page with too many trackers.

Say you run 14 analytics tools all on the same page — not something I'd recommend, but that's what we're doing with our test sites for this guide. How will that affect web performance?

Here's another web page test with a simulated iPhone X:

https://www.webpagetest.org/result/220817_BiDc9M_9MS/1/details/#waterfall_view_step1

Observed Web Vitals Metrics (Collected in this WPT test run)

Largest Contentful Paint

470ms

Cumulative Layout Shift

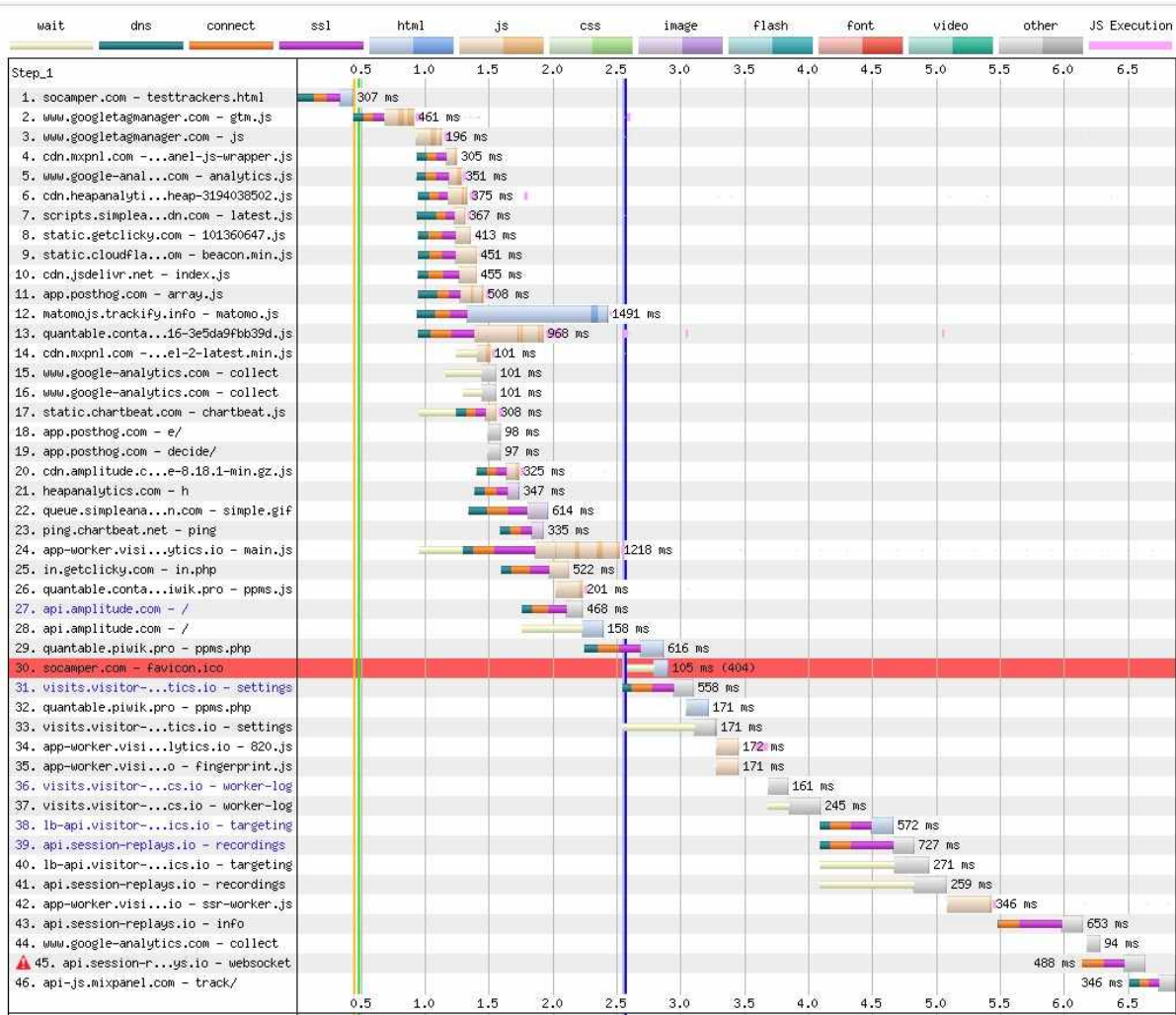
0

Total Blocking Time

149ms

We see that, despite being loaded down with 14 different trackers (including some that run session capture as well as web analytics), we haven't significantly affected LCP or CLS. We're still in the green for Core Web Vitals.

We've potentially altered FID, but it's not too bad. So, even with 14 trackers, we have a minimal SEO impact. The waterfall chart isn't so pretty though:



You can see here that the total time the page takes to load has skyrocketed: It now takes over 2.5s to render the full document and over 6.5s to finish everything. This can hurt the end user's overall experience.

So, don't run 14 analytics tools, duh. That might seem like an absurd number, but plenty of sites run *at least* that many trackers when you include all the tags on a site, especially ad tags with chaining. The average number of trackers per site is seven (according to Ghostery: <https://whotracks.me/websites.html>), but it's a wide distribution, with many sites running 14 or more.

In addition to limiting the total number of trackers, we should also make sure the ones we recommend perform well.

How do we evaluate tracker performance? The main aspects we should look at are:

1. Tracker size
2. Tracker complexity and chaining
3. Physical location of the servers

1. Tracker Size

UA is 50k (uncompressed), which is a reasonable starting benchmark. Tracker size can vary significantly based on what the tracker does and what features and plugins you enable.

Tracker	Uncompressed Size	CPU Time
Google Universal Analytics	50k	17ms
Matomo	201k	8ms
Cloudflare	14k	2ms
Clicky	15k	8ms
GA4	201k	47ms

In this era of 5M hero images on home pages, the 200k of Matomo or GA4 is a modest amount of data (and less than 200k will be transferred over the network since it should be compressed). A JavaScript payload of 200k does indicate the tracker contains a significant amount of scripting that may need to run on the browser. This means tracker size is a reasonable indicator of browser performance impact, which, for trackers, is typically more important than network size. In other words, when looking at something relatively small like a tracker, the main concern is how long the browser will take to run the JavaScript, rather than how long it'll take to download. This is why I've also included a measurement of CPU time in the chart above as well as size.

2. Tracker Complexity and Chaining

Once we've downloaded that 200k, the browser may have to perform many actions. As we can see from the amount of CPU time used, it's not a one-to-one relationship between size and CPU usage. Just because the tracker is bigger, it doesn't mean it'll do more. You also have to look out for trackers that chain in other trackers. For example, if you use gtag.js to run your UA, it'll have to load gtag.js and then chain in analytics.js, making using gtag.js slower with more overhead than calling analytics.js directly. Our CPU time measurements are from one sample load only, not a comprehensive test of JavaScript performance.

3. Server Location

The physical location of the servers involved in site analytics is of increasing importance, though much more for legal compliance than site speed. We'll talk about site speed impact here and compliance issues in Chapter 14.

From a site speed standpoint, we want the server recording the analytics measurement data to be as close to the end user as possible. If our end user is in Australia, we'd like the server that collects the measurement calls to be in or near to Australia. This is what I referenced earlier when talking about edge delivery and is the same concept as content delivery networks (CDNs) for things like images and CSS delivery.

If the end user is in Perth, Australia and our analytics measurement server is in Virginia, USA (AWS' largest data center), that means every tracked event click sends data across a vast distance of 18,500km (11,500 miles). That would take 123ms to do a round trip as the crow flies at the speed of light.

Our data packets bounce around much more than these imaginary crows capable of relativistic speeds, so in reality, we're looking at more like 300ms of latency added to every request. If I ping an IP in Perth from where I am now in Boston, USA, it takes 280ms, confirming our estimate.

If I ping analytics.google.com it only takes 30ms, sending me to a server that appears to be in the Philadelphia area (about 500km away from

Boston). This saves me 250ms and 18,000km of overhead per request compared to going all the way to Australia.

Extreme distance causes slowness and increases the chance that the measurement requests might not even arrive. 250ms may not sound like a lot, but that's just the overhead, adding a hefty tax onto server and browser processing time of every request.

How fast should a web request be? Shopify analytics tool Littledata reports an average server response time of 456ms among all the requests they saw from 5,822 websites (<https://www.littledata.io/average/server-response-time>). This is not a measurement of a full web page load, but just one component on the page — be it an analytics request, HTML source, or a CSS stylesheet. Adding 250ms to our measurement requests right off the bat is not setting ourselves up for good performance.

We want our analytics reporting system to have end points distributed around the world, transparently routing measurement requests to whatever endpoint can most quickly serve the user. Many analytics tools have taken into account this kind of edge routing, but some are better at it than others. If we're self-hosting, we need to consider this as well. (Note that the cost of a geographically distributed system was not included in our self-hosting example earlier.)

Then there's the compliance aspect of where the data is transferred to and stored, which is especially critical if you operate in the EU. When serving EU end users, you're required by GDPR to keep their personal data in the EU. So, if we have an end user in Germany, we'd need to keep all of the data that could be associated with that person inside the EU. This means having a measurement end point in Frankfurt, Germany and then transferring the raw data to Virginia, USA for processing would be verboten.

As you can imagine, sometimes these two requirements are at odds with each other. An easy solution to solve GDPR data storage issues would be to have all your analytics servers in the EU, but in doing so, you'd make the connection from people in Australia or the U.S. much slower than it could

be. For example, the Piwik PRO free tier sends everything to the EU. For an additional fee, you can establish points elsewhere, which I'd recommend if you have significant traffic outside the EU.

With GA, we've historically never had to think about this; Google handled endpoint routing automatically and transferred all data to the U.S. for processing. Since this is no longer an option for full GDPR compliance, we need to consider server location carefully for both legal reasons and performance.

CHAPTER 12 - ADVANCED FEATURES

IF YOU'VE WORKED WITH GA installations beyond basic content sites, you've surely dug into some of the more advanced functionality in UA. Features like cross-domain tracking, multi-channel funnel (MCF) reports and attribution, data imports, custom channel groupings, etc.

UA has many capabilities, and none of the alternatives we look at is a 100% match for its full functionality. Before tossing out an alternative for not having *all* the features you're used to, reevaluate your need for those particular features. In my experience, many of these advanced features have been implemented to address an issue or handle a feature request that wasn't mission critical or is no longer actively needed. Maybe you made a complex custom channel grouping to handle a series of unusual edge cases, only to find the people who requested the grouping never use it and simply default to source/medium. Or, you built a MCF custom model but failed to find adoption for it when the results didn't meet the business' expectations.

Granted, many advanced feature implementations are necessary, such as a customized enhanced ecommerce implementation for a shopping site. But many times, we enable functions in UA to solve issues and answer questions that aren't vital. Or, we'll throw more metrics at the problem instead of more deeply analyzing the core issues. Many GA implementations substitute breadth of tracking for genuine analysis.

Decreasing the amount of tracking tends to be a hard sell to your stakeholders. If the consumers of your reports are used to seeing a certain metric and then it disappears, they may ask for it back, even if it was incorrect or unnecessary. How should we proceed in that case? What do we do if the consumers of our reporting from UA are accustomed to a

particular metric like bounce rate and page value, but the replacement tool we are considering doesn't include those things?

We should attempt to return to the business questions underlying those metric requests and find another, hopefully better way to answer them. Not all stakeholders will be receptive to these alternative answers, but many will if you explain how the new reporting still answers their questions.

The primary point is not to let stakeholders and other report consumers drive low-level requirements for the new tool. Having a checklist of metrics you "need" to match your existing UA implementation is a trap that invites a repeat of previous mistakes. A silver lining of the discontinuation of Universal Analytics is that it's a good opportunity to reevaluate tracking scope and try to cut out some of the legacy cruft and technical debt that's built up over time.

If we do need to replicate the more advanced components of our UA implementation, then we need to look for either a platform that allows us to look at the problem in a similar way to UA (like Matomo and Piwik PRO), or one that has a lot of flexibility (like most of the tools in the product analytics category).

CHAPTER 13 - COOKIES & "FUTURE-PROOFING"

WHAT SHOULD WE DO about cookies?

Cookie banners and warnings, premature cookie expiration, a post-cookie world, the death of cookies — it all sounds quite morbid for something with such a comforting name.

Why the hyperbole? In part, "the death of cookies" sounds much catchier than "the near elimination of cookies set in a third-party context in conjunction with increasing limitations on first-party cookies dependent upon your browser and its settings." Another aspect is this change has happened relatively quickly, and much of it (especially with Safari) happened through unilateral action of browser makers with minimal advanced notice.

Cookies aren't going away, and the majority of the GA alternatives we look at are based on them. What is changing is how browsers restrict cookies, and unfortunately for us, each one is handling things a bit differently. This is causing plenty of headaches in the analytics world, but as of yet has not caused a sea change in how we perform web analytics.

Safari is notoriously strict, blocking all third-party cookies and restricting first-party cookies set with JavaScript to seven days, or 24 hours if the URL has a referrer from a known tracker with tracking info on the URL. For example, a page referred from Google with "?gclid=xxx" in the URL will have its cookie downgraded to 24 hours.

The most lenient browser is Chrome, which has no enforcement beyond HTTP standards. A GA cookie in Chrome will last up to 400 days,

compared to seven days in Safari. GA cookies are specified to last for two years, but that's no longer true, even in Chrome.

The most restrictive is Brave, which blocks all known ads and trackers (including GA) by default, so you don't have to worry about analytics cookie expiration, because there aren't any being set! As of September 2022, Brave now can even block consent management platform (CMP) cookie banners (<https://brave.com/privacy-updates/21-blocking-cookie-notices/>).

Chrome is scheduled eventually to start blocking third-party cookies as well, which will effectively end their usage. Google has pushed that date back multiple times, now scheduling it for the end of 2024. Cookiebot (a CMP provider) provides an informed overview of this here:

<https://www.cookiebot.com/en/google-third-party-cookies/>

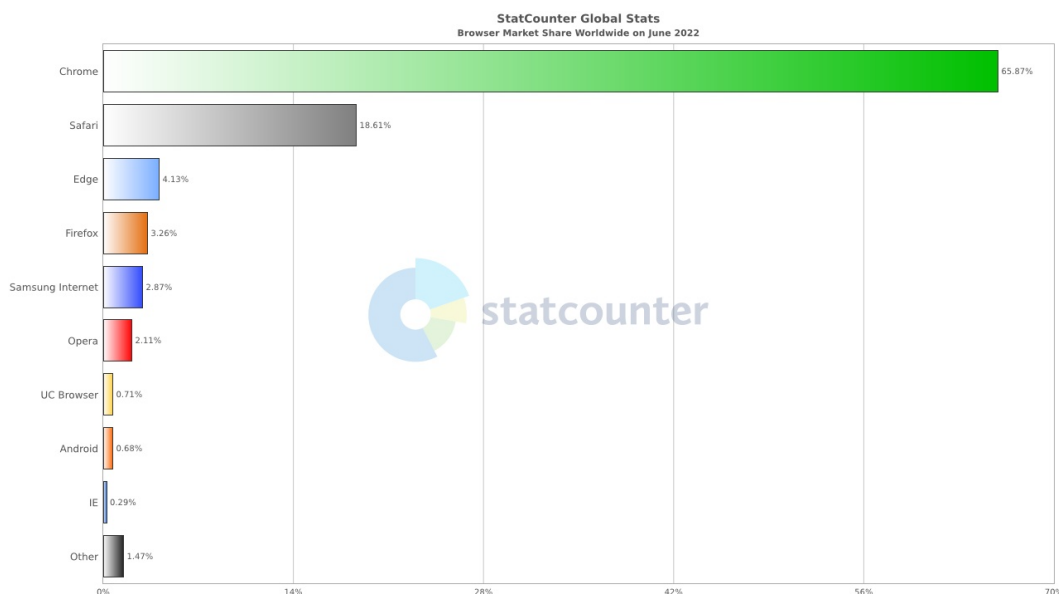
Browser	First-Party Analytics Cookies	Third-Party Advertising Cookies
Chrome	400 days	400 days (for now)
Safari	7 days	Blocked
Brave	Blocked before setting	Blocked before setting

How do you keep track of all these changes and how they affect your analytics? I recommend <https://www.cookiestatus.com>. It's an excellent resource maintained by the inimitable Simo Ahava that shows how six major browsers handle cookies and other privacy- and tracking-related items like localStorage and referrers. It provides great background information on tracking as well.

Each time there's been a major change in this list, people ask, "Will this break my (Google) analytics?" So far, the answer has mostly been, "Not too badly."

Analytics platforms don't (for the most part) use third-party cookies, so, while that event is big for the advertising industry, it's less so on the analytics side. The downgrading of first-party cookies to shorter time frames, however, has a bigger impact.

Everything Chrome does makes news because it dominates the browser market. With 66% of the market, it has a huge margin over the second-place competitor Safari:



(stats from Statcounter <https://gs.statcounter.com/browser-market-share>).

You might ask, "Hey wait, where's Brave?" Brave doesn't show up in the Statcounter benchmarks because it blocks Statcounter by default (more on this in Chapter 19 - Statcounter). Brave also doesn't declare itself in its user agent — it appears as generic Chromium — so the best source of how many Brave users there are is the company itself.

Brave reported 50.2 million monthly active users (MAU) at the end of 2021, and that number is growing rapidly: <https://brave.com/2021-recap/>. This is a lot of users, but it's still small in terms of market percentage (the market in question being all the internet users in the world).

Firefox also publicly reports MAU, showing around 210 million MAU at the end of 2021 (<https://data.firefox.com/dashboard/user-activity>), a little more than four times the number of Brave users.

If Brave has one quarter the market share of Firefox, then giving Brave an overall market share of about 0.8% based on Firefox's Statcounter numbers seems a reasonable estimate. Other sources suggest Brave has about 1%. I included Brave in my chart above rather than Edge or Firefox because it's also representative of any browser that has an ad blocker installed.

So, Brave takes a bite out of our aggregate analytics, but it's a much smaller bite than ad blockers overall, which I've previously estimated to be around 10% (<https://www.quantable.com/analytics/how-many-users-block-google-analytics/>). As I state in that study, the 10% is highly variable depending on the demographics of your site.

I prefer to lump Brave users in with other missing adblock users rather than considering Brave's inherent cookie restrictions, since the adblocker built into Brave short circuits analytics measurement before any sort of cookie restrictions.

The challenging, ever-changing landscape of analytics has convinced many people of the need to "future-proof" their analytics installations.

This is a worthy idea, but I recommend focusing first on existing issues rather than trying to figure out where the market is going. Otherwise, you may find yourself working on theoretical problems that never actually manifest.

We face plenty of analytics challenges today, so let's get ready for the future by solving what's in front of us now, rather than worrying too much about what may come.

Compliance, ITP, cookie instability, content blockers, and tracking proxies are some of the known issues in analytics today. How should we think about these problems when evaluating a GA alternative?

I suggest identifying the biggest weaknesses in your current implementations and how our possible alternatives could remedy them.

Does your site have a technologically oriented demographic and a high percentage of users who run ad blockers? Consider an alternative that mitigates those blockers in a privacy-friendly way. Dealing with a product that has a long sales cycle and a large number of mobile Safari users? Look into server-side analytics or modeled conversions (which I'm not a big fan of, but they can be a necessity).

The biggest "fix" here is server-side analytics. Server-side measurement solves many of the issues mentioned above, as it can:

- Bypass content blockers
- Avoid rules in ITP (and similar) that focus on cookies set by JavaScript

(server-side sets cookies with HTTP headers rather than `document.cookie()` in JavaScript)

- Sign and validate the measurement requests for improved data integrity. I've written about that previously here:

<https://www.quantable.com/analytics/improving-analytics-integrity/>

What server-side can't do is solve your compliance issues. In fact, you need to pay *more* attention to privacy compliance when using server-side because it circumvents anti-tracking tech like ITP. If you are running the data collection on your own domain and services, you are taking more responsibility for it and should act accordingly.

Server-side tracking is undoubtedly more work, but some alternatives ease this by designs that support it as a possible implementation method. For example, Piwik PRO is built with server-side in mind and has helpful

documentation available (<https://piwik.pro/blog/server-side-tracking-first-party-collector/>). Compare this to most of the options in the simplified category, which don't explicitly support server-side. Any client-side tool could be run through a proxy to make it server-side, but those like Piwik PRO that already have the documentation and design to allow this will make a server-side solution much easier to implement and more reliable in practice.

CHAPTER 14 - PRIVACY & COMPLIANCE

WHAT ARE THE PRIVACY concerns we need to address to make a more informed decision?

Throughout this guide, we talk about privacy and compliance, but I've consolidated that discussion here into a few specific questions.

1. Can I do consent-less tracking?

If you do any tracking that can persistently identify users, you need user consent in an increasing number of locales, if not universally.

This means, unless you are looking in the simplified category, consent is likely required for strict compliance. Even if you can find a solution that can be run consent-less, consider that you may need consent for other things on your site. Do you run cookies or track personal information in any other way that would require consent management? There's little point in doing the work for consent-less analytics tracking only to realize you still need consent for your A/B testing platform or advertising tools.

In my opinion, cookie consent boxes are a terrible solution to user privacy, especially for consents that lack an option for the user to say no. Additionally bringing in another party to keep track of a (possibly meaningless) user choice only decreases user privacy by exposing their information to yet another third party. Unfortunately these are the rules we're forced to work within.

We should also remember this issue extends beyond the EU. Obviously, the EU has taken the lead with GDPR, but data privacy and protection laws are spreading quickly throughout the world; as of December 2021, 137

countries have established such laws (<https://unctad.org/page/data-protection-and-privacy-legislation-worldwide>).

As of October 2022, only five US states (CA, CO, CT, VA, UT) have active data privacy and protection laws (<https://iapp.org/resources/article/us-state-privacy-legislation-tracker/>), leaving the US overall lagging behind much of the world. The California laws (CCPA/CPRA) are the most notable, with many other states basing their laws upon California's.

Picking a GA alternative like Cloudflare Web Analytics that doesn't require user consent is a great way to avoid having to worry much about those laws, but remember that only gets you off the hook with your analytics tool.

2. Assuming I need consent, how does my analytics tool work with my consent management platform (CMP) and my tag manager?

While a few alternatives include consent management tools, such as Piwik PRO (<https://piwik.pro/gdpr-consent-manager/>), most don't and consent will likely be managed via a third-party CMP. Similarly, some alternatives feature tag managers (Piwik PRO, Matomo), but it's more common to use a third-party TMS like Google Tag Manager.

So, when we implement an alternative, we need to consider our CMP and tag manager tools and how they work with our new solution.

3. What data can I collect and how can I store it?

Information security has a rule called the "principle of least privilege" (https://en.wikipedia.org/wiki/Principle_of_least_privilege) or the "principal of minimal privilege" (my preferred nomenclature). This rule states that each person or process should have access to the minimal amount of information and resources needed to do their function. Typically, this is used in the context of permissions in computer security to limit access.

For example, you may give an outside consultant read-only access at the GA view level to the only view they need to perform their analysis, instead of allowing edit access at the property or account level.

This same principle can and should apply to the data itself. What's the minimal amount of data I need to collect and store to accomplish the job at hand?

Introducing this limitation can easily become annoying, but we need to do our best to avoid taking the easy way out. In our GA access example above, if you're the analytics consultant receiving access and you discover you need to see a different view as well, or need to edit a filter or something, you'd have to go back and ask for more permissions, which is frustrating and time-consuming. Considering that user permissions are infrequently cleaned up, it's obvious what issues could arise from giving everyone admin access. A common problem is "privilege creep," where access levels *increase* over time rather than decrease.

Now think about how this rule would apply to data. If you don't need user IPs, don't store them. If you don't need historical access to raw data for more than 30 days, don't keep it. This approach aligns well with the "explicit" style of tracking we discussed previously. It's more work, but it's a substantial improvement for both compliance and user privacy. Minimal privileges is something that most projects intend to do, but few actually succeed in doing.

PART TWO

PRODUCT EVALUATIONS

EVALUATION METHODOLOGY

IN THE FIRST PART of this guide, we covered a lot of background and created a framework of questions to help pick a replacement for Universal Analytics. In part two, we'll apply this to 15 different products. Rather than scrutinize every detail of each tool, I'll focus on what makes a particular product unique and how it compares to similar alternatives.

To gain a better feel for these products, I installed them on websites with real traffic. I believe the only way to understand a product is to see it in action with real-world data.

A read-only demo account with test data can provide first-hand experience with the product, but every data and use-case scenario is unique. Plus, the sample data will be a perfect match for the tool and likely cleaner than your own. Additionally, you miss out on installing and instrumenting the tool, which is a big part of really learning it.

The two websites I used for testing had modest levels of traffic (approximately 300 sessions per month for one, and 1,200 for the other). These lower levels of traffic and relatively simple functionality made it impossible to delve deeply into some of the sophisticated tools available in the product analytics options, but they still provide enough information to point us in the right direction.

While I did do testing of migration tools where available, this was ad-hoc and only applicable to some tools.

In an attempt to reduce bias further between products, I haven't spoken at length with any vendor reps from these products in the process of doing these evaluations.

Vendor reps are obviously an excellent source of information about their products, and I recommend you talk to them as part of your product evaluation, but to make these comparisons as fair as possible I decided to universally avoid talking to the vendors. I also paid for the plans I used if I wasn't in a free tier or under trial rather than trying to receive any discounts.

That said, it's impossible to entirely avoid mistakes or omissions in the process of evaluation. It's also likely these products have changed or received upgrades (I conducted my evaluations primarily in July through September of 2022).

TOOLS INDEX

THE FOLLOWING IS A list of all the tools we'll evaluate, grouped by category and sorted by popularity within each section. Tools marked with a "☆" are my top picks.

Tool	Category	Default Tracking Method	Mobile SDK	Self-Hosted Option	Open Source
Matomo	Traditional	Cookies	✓	✓	✓
Piwik PRO ☆	Traditional	Cookies	✓	✓	✗
Clicky	Traditional	IP + User-Agent	✗	✗	✗
Cloudflare Web Analytics	Simplified	Referrer	✗	✗	✗
Statcounter	Simplified	Cookies	✗	✗	✗
Chartbeat	Simplified	Cookies	✗	✗	✗
Fathom	Simplified	IP + User-Agent	✗	✗	✗
Plausible Analytics ☆	Simplified	IP + User-Agent	✗	✓	✓
Visitor Analytics	Simplified	Fingerprinting	✗	✗	✗

Tool	Category	Default Tracking Method	Mobile SDK	Self-Hosted Option	Open Source
GA4	Product	Cookies	✓	⊗	⊗
Mixpanel	Product	Cookies	✓	⊗	⊗
Snowplow	Product	Cookies	✓	✓	✓
Amplitude ☆	Product	Cookies	✓	⊗	⊗
Heap	Product	Cookies	✓	✓	⊗
PostHog ☆	Product	Cookies	✓	✓	✓

Please see the Appendix for the full comparison chart.

TRADITIONAL WEB ANALYTICS

MATOMO CLOUD

Piwik PRO

Clicky

CHAPTER 15 - MATOMO CLOUD

Type	Traditional Web Analytics
Default Tracking Method	Cookies
Popularity	#5, 2% of top 1M sites
Price	\$39/mo. for 100k hits \$369/mo. for 2M \$4,390/mo. for 25M
Home page	https://matomo.org

Pros: Mature product, full suite of tools, open source, similar to UA in functionality

Cons: EU-hosting only, can be expensive at scale, limited ability for ad-hoc analysis (compared to UA), interface can be clunky

Within the world of Google Analytics alternatives, Matomo has long been one of the most well known; it's the first product that shows up in a Google search for "google analytics alternatives." According to BuiltWith, Matomo is also one of the most used product among the alternatives we're evaluating.

Matomo's popularity is well earned: It's been around since 2007, is open source, and is similar to UA in terms of design and functionality. Matomo also has a feature set that's far closer to UA than GA4. So, **if you're looking**

for something that most closely matches UA, then Matomo (or Piwik PRO) is likely your best option.

This similarity to UA makes for a straightforward migration to Matomo, with an easy learning curve. Matomo even offers automated migration tools to help transfer your UA setup and data to Matomo. Historical data import is a very helpful bonus feature, but full of caveats. The biggest one to watch out for is that data imported from UA is aggregated data only, not reprocessed raw data. This means there will be no detail available on this imported data beyond standard reports, and all the raw data was originally processed according to UA's rules, not Matomo's (<https://matomo.org/faq/general/limitations-when-importing-google-analytics-data/>). The result is that you'll inherently be comparing different types of data. If you do this import, I suggest doing it into a separate historical website property rather than into your current product website property (you also can't import into a pre-existing property).

The downside of this similarity is that some of the drawbacks of UA also exist in Matomo. Matomo is pageview-based rather than event-based like GA4, which can be a drawback when dealing with mobile app data. While it does have a mobile SDK, it doesn't provide the level of seamless integration between different platforms that other event-based tools in the product analytics category can.

Like Universal Analytics, it also carries baggage in the form of technical legacy that newer platforms do not. For example, Matomo uses PHP and MySQL, a classic combo used by WordPress and many others — but one that lacks some of the technical advantages of more modern architectures.

In particular, MySQL does not compare favorably at scale to databases like Google BigQuery or ClickHouse, which were designed with analytics in mind. As a result, if you're looking for a solution for sites that get more than 10 million hits a month or have large amounts of historical data actively used in reports, you may want to look elsewhere. If you have lower traffic levels, this will be less of an issue.

I mentioned Piwik at the start of this chapter, so let's clear up the Matomo vs. Piwik confusion. Piwik started as an open source project in 2007 and was renamed "Matomo" in 2018. There's no longer a product called Piwik, but there is "Piwik PRO," which began as a part of the Piwik project in 2013 before splitting off in 2016.

Matomo and Piwik PRO are fully separate products that share the same historical foundation. Piwik PRO no longer contributes back to the core Matomo project, and they've significantly diverged. We'll evaluate Matomo and Piwik PRO individually, but much of what I say will still apply to both. Piwik PRO isn't open source or used as widely as Matomo, but it does have some advantages over Matomo that we'll cover in the Piwik PRO chapter.

We're also looking at the cloud-hosted version of Matomo rather than the self-hosted version (discussed more in Chapter 9: Should I Self-Host). I've installed and run the self-hosted version in the past and found it works as expected and is not difficult to set up.

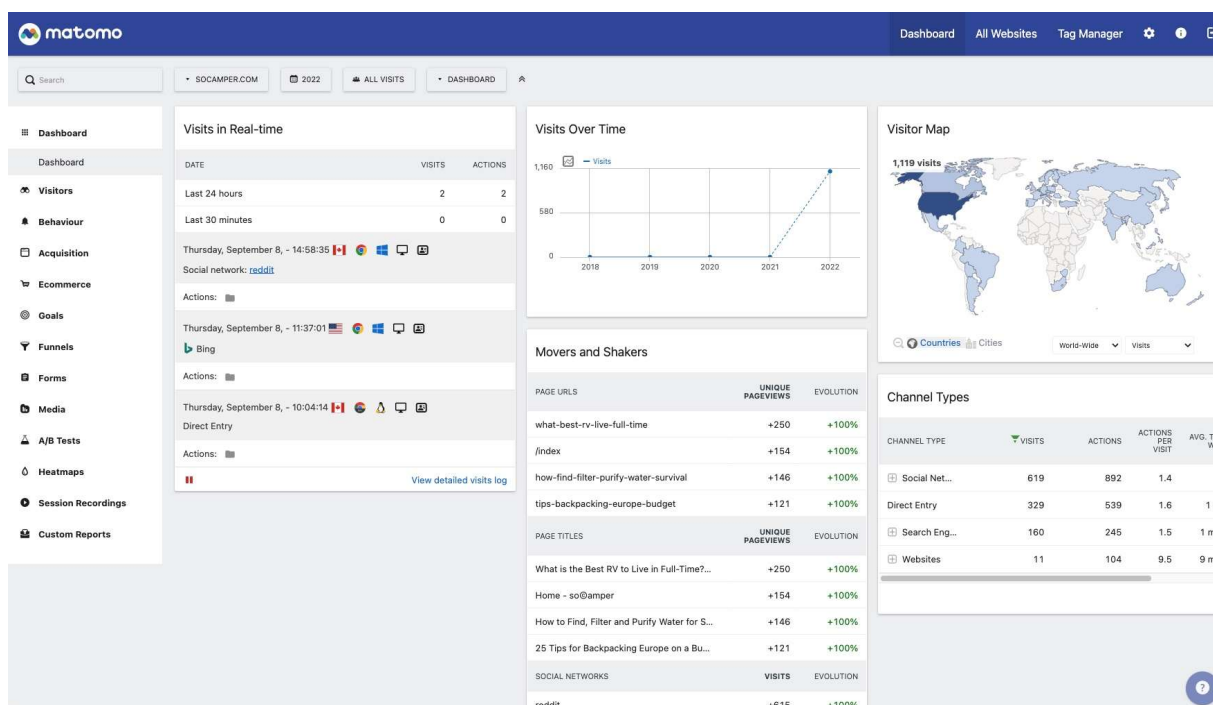
Like UA, Matomo is a comprehensive tool with a myriad of customization options. It might be more appropriate to compare Matomo to the Google Marketing Platform as a whole rather than GA alone, since Matomo includes the following features:

1. Analytics
2. Tag manager
3. A/B testing
4. Heatmaps
5. Session recording

Several of the alternatives we evaluate are components of suites that include functionality beyond analytics. I'll mention those additional functions when relevant, but we aren't explicitly evaluating those. Access to this full suite of tools as part of the base Matomo is a big advantage of the product.

One site speed issue I had with Matomo Cloud is that everything is hosted in the EU. Even though I'm in the U.S., every measurement request was sent to Germany. Please see Chapter 11 for a further discussion of why this is problematic.

As stated, the core reporting functionality is similar to UA. The main dashboard is more akin to some of the simplified analytics tools, but its reporting goes much deeper than anything from that category.



When you look at the left-hand navigation, you can see the similarities to Universal Analytics: Visitors, Behavior, Acquisition, etc., with submenus underneath. This isn't to say Matomo is a clone of UA — on the contrary, it's substantially different, but it works within the same paradigm.

Additionally, Matomo has improved upon a number of features from UA. For example, Matomo's goals system is similar to UA, but goals in Matomo can be deleted or set up to allow multiple conversions per visit. Other similar improvements in usability make it clear Matomo has listened to their users' feedback much more than Google.

Where Matomo falls short is its lack of efficient ad-hoc analysis. In other words, it can become slow and frustrating when you want to move beyond the standard reports. This likely stems from the product's technical underpinnings, specifically what we talked about in regards to the database.

We should try to temper our expectations when comparing Matomo to UA, since the latter is a product of one the biggest tech companies in the world, run for many years on tens of millions of sites.

For example, Matomo has a powerful segment builder similar to UA's:

The screenshot displays the Matomo segment builder interface. At the top, a tab is labeled "US MOBILE SAFARI USERS". Below this, a status bar indicates "THIS SEGMENT IS VISIBLE TO: ALL USERS" and "AND PROCESSED FOR THIS WEBSITE ONLY". The segment name "Name: US Mobile Safari Users" is shown with an "edit" link. The main area contains two conditions stacked vertically, connected by an "AND" operator. The first condition is "Browser" is "Mobile Safari". The second condition is "Country" is "United States". Each condition has an "OR" button and a "+Add OR condition" link. At the bottom, there is an "AND" button and a "+Add AND condition" link. The bottom bar includes icons for thumbs up/down, and buttons for "DELETE", "CLOSE", "TEST", and "SAVE & APPLY".

But unlike UA, segments are not immediately available, as they need to be processed first:

SOCAMPER.COM SEPTEMBER 2022 US MOBILE SAFARI USERS

These reports have no data, because the Segment you requested (US Mobile Safari Users) has not yet been processed by the system. Data for this Segment should become available in a few hours when processing completes. (If it does not, there may be a problem.)

Meanwhile you can use the Visitor Log to test whether your segment will match your users correctly by applying it there. When applied, you can see immediately which visits and actions were matched by your segment. This can help you confirm your Segment matches the users and actions you expect it to.

Pages

There is no data for this report.

So, once you create a segment, you may have to wait a few hours to view its data. With my smaller testing dataset, segments tended to be available within a few minutes of creation — so if you have a site with less traffic, this may only be a slight annoyance. However, for a site that has significant amounts of data, this drawback could be incredibly frustrating.

Another of Matomo's ad-hoc analysis limitations is evident in default reporting.

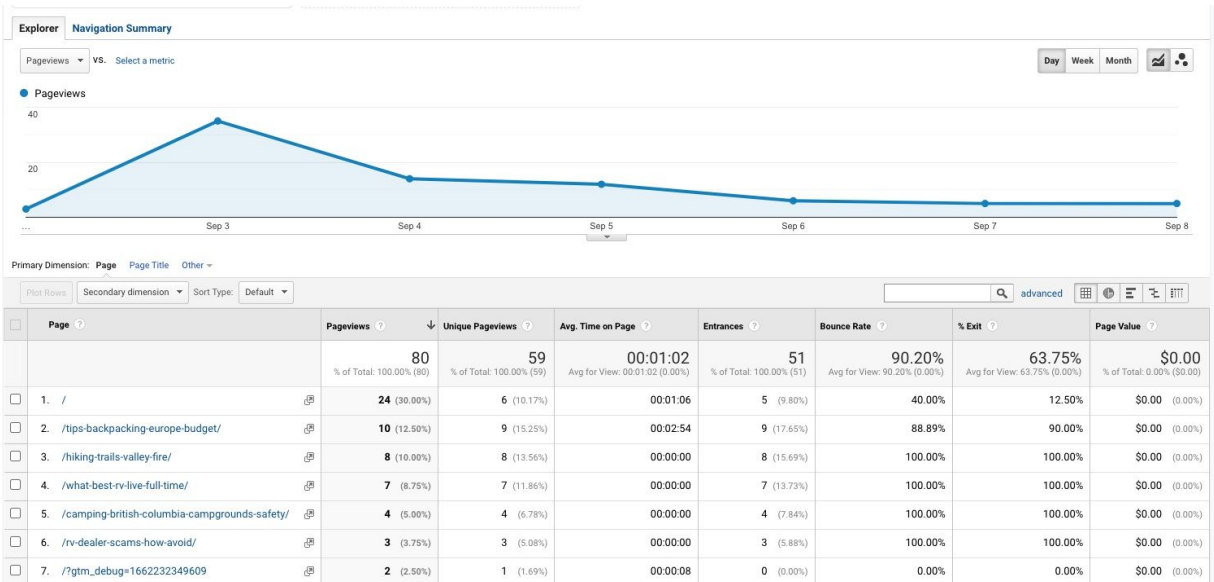
Compare the pages report in Matomo:

SOCAMPER.COM SEPTEMBER 2022 ALL VISITS

Pages

PAGE URL	PAGEVIEWS	UNIQUE PAGEVIEWS	BOUNCE RATE	AVG. TIME ON PAGE	EXIT RATE	AVG. PAGE LOAD TIME
/index	33.3% 21	4	50%	00:00:49	50%	1.84s
/render2?id=GTM-M4DJK9C	9.5% 6	6	100%	00:00:00	100%	0.08s
hiking-trails-valley-fire	7.9% 5	5	100%	00:00:00	100%	6.63s
what-best-rv-live-full-time	6.3% 4	4	25%	00:00:00	100%	4.36s
tips-backpacking-europe-budget	4.8% 3	3	67%	00:00:00	100%	3.15s
camping-british-columbia-campgrounds-safety	3.2% 2	2	0%	00:00:00	100%	3.61s
cast-iron-skillet-recipes-camping	3.2% 2	2	50%	00:00:00	100%	2.47s
rv-dealer-scams-how-avoid	3.2% 2	2	0%	00:00:00	100%	5.57s

to the equivalent in UA:



Although fundamentally similar, UA allows substantially greater ability to filter dynamically. We can use the advanced search box to include or exclude by dimension, leverage secondary dimensions, and even change primary dimensions within the same report. Matomo can't do this because of its limitations in dynamic reporting; rather, it has to generate its reports ahead of time. As someone that uses this functionality in UA almost every time I am in the tool, I found this a major deficiency.

Another downside of Matomo, particularly compared to tools in the product analytics category, is its user interface can feel slow, clunky, and outdated. I have no issues with older user interfaces per se; sometimes they're well-tuned, easy to use, and bulletproof from so many years of testing and customer feedback. In Matomo's case, it tends to slow down when dealing with lots of data or filtering. There's also a large number of different reports and it can be confusing and require a lot of clicks to find what you're looking for. UA has similar issues with complexity, even we're less of aware of them after many years of use.

Overall, Matomo is a solid replacement for many current UA users. It can be underwhelming if you expect it to be what you wish GA4 had been, but it's a good fit for the majority of smaller UA installations.

CHAPTER 16 - PIWIK PRO

☆ FAVORITE

Type	Traditional Web Analytics
Default Tracking Method	Cookies
Popularity	? (mixed in with Matomo)
Price	Free to 500k hits \$325/mo. for 1M hits (Core+) \$545/mo. for 2M hits (Core+) ~\$2k/mo. for 10M hits (Enterprise)
Home page	https://piwik.pro

Pros: Closest to UA, generous free tier, deep feature set

Cons: Relatively high entry price, no paid support or non-EU hosting until enterprise plans, lack of enhanced ecommerce tracking

Piwik PRO is my top recommendation for someone who wants to continue processes similar to UA on a new platform. Matomo typically garners more attention and has a larger market share, but Piwik PRO has a more technically advanced core and deeper functionality.

Matomo is broader and a good option for smaller sites with less advanced needs, but if you use more of the complex functionality in UA, you should look at Piwik PRO first. Matomo includes some ancillary

features which Piwik PRO does not such as session recordings and heatmaps, but Piwik PRO allows for more sophisticated analysis of web analytics data.

Piwik PRO has its origin in enterprise consulting for open source Piwik (before it changed its name to Matomo). Due to these enterprise roots, Piwik PRO has been targeted more at enterprise installs, with less transparent pricing than Matomo's. Matomo Cloud has a full public pricing page with pricing up to 100 million hits per month, whereas Piwik PRO only shows prices to those with a free account when they look to upgrade (currently listing up to two million hits per month).

As discussed in the Matomo chapter, both products are derived from the Piwik open source project, but Piwik PRO has branched off and is closed-source. You can still self-host with Piwik PRO, but only with enterprise plans.

Piwik PRO's free tier, called "Core," offers a generous level of hits with up to 500,000 per month, though it has some limitations on its features and hosting options. Some of these enterprise-only features include: US-based hosting, BigQuery access to raw data, longer data retention (limited to 14 months for the free tier). These limitations aren't severe, but anytime you're in a free plan, you risk running into a particular paid feature that you need, which forces you into a paid plan at a high cost.

At one million hits, this becomes the "Core+" plan, which is \$325 per month. This still has the same functionality as Core, just a higher hit limit. Core+ still has no paid support or US-based hosting — as those are only available once you get to the enterprise level. This lack of paid support is a serious downside of Core+.

This "freemium" model is much different from Matomo Cloud's pricing model, which offers no free level and is geared toward small to medium sites. This makes Matomo Cloud and Piwik PRO hard to compare pricing as they are targeting different markets with a much different pricing structure.

Max Hits (per month)	*Matomo Cloud\ (monthly)**	Piwik PRO† (monthly)
50k	\$23	\$0
100k	\$39	\$0
1M	\$189	\$325
2M	\$369	\$545
10M	\$1,790	~\$2,000
25M	\$4,390	~\$3,500
50M	\$8,590	~\$4,295

- Excludes Matomo Cloud Enterprise, which may be cheaper than Matomo Cloud at high scale.

† Piwik PRO levels above two million are unavailable for self-checkout, prices are estimates.

At levels above 10 million hits per month, Piwik PRO becomes cheaper than Matomo Cloud, eventually significantly so. As with all pricing in this guide, this is a snapshot taken in October 2022 and subject to change.

Piwik PRO maintains a detailed comparison versus Matomo: <https://piwik.pro/piwik-pro-vs-matomo-comparison/>.

Matomo has also published a comparison: <https://matomo.org/piwik-pro-vs-matomo>, which doesn't go into quite the same depth but is still pretty comprehensive.

We of course need to take vendor-published comparisons with a grain of salt — especially checklists without deeper context.

The most significant technological difference between the two products is their back-end database. Matomo uses MySQL, whereas Piwik PRO uses ClickHouse. MySQL is a traditional relational database that's existed since 1995, whereas ClickHouse is a newer column-oriented database, first

released as an independent product in 2016 and designed specifically for analytical tasks.

ClickHouse is significantly faster at scale than MySQL for analytical queries and, in my opinion, a far superior choice to power an enterprise-level analytics database. On the other hand, MySQL is much more widely used and supported, so if you plan to self-host, MySQL is easier to manage. Both solutions make sense for their respective target markets.

ClickHouse was developed by Yandex and used in their UA-style analytics platform Yandex Metrika. This means ClickHouse was specifically designed with a web analytics platform's needs in mind. ClickHouse has since spun out from Yandex completely and is now based in the U.S and the Netherlands. I've seen ClickHouse be compared to BigQuery, which isn't a perfect comparison, but it's close enough for our purposes.

Both ClickHouse and BigQuery are column-oriented databases (https://en.wikipedia.org/wiki/Column-oriented_DBMS) that can perform fast analytical queries on huge datasets. You can find a more detailed comparison between the three databases here: <https://db-engines.com/en/system/ClickHouse%3BGoogle+BigQuery%3BMySQL>.

Selecting a database like ClickHouse instead of a traditional RDBMS like MySQL is not a choice unique to Piwik PRO. Among our alternatives, many of them use analytical databases such as ClickHouse (used by Piwik PRO, Plausible, PostHog, Yandex, and likely Cloudflare), SingleStore (Fathom), Snowflake (Amplitude, Snowplow), or others. Some alternatives support multiple choices for back-end databases if you self-host.

Piwik PRO is also similar to UA, but supports more of the ad-hoc reporting that UA allows for. Let's return to the pages report example from the Matomo chapter. Matomo lacks dimension-based filtering, but Piwik offers that and more:

Pages Page titles Entry pages Exit pages Page speed

Nested dimensions: Page URL > Channel > Device type > Device model

Page URL Search Quick filters ...

Page URL matches regex Enter value...

AND

+ Add condition

Apply Cancel Clear filters

	Page URL	+ Page views	Unique page views	Entries	Page bounce rate	Average time on page	Exit rate
		46	28	27	55.56%	1m 13s	100%
☐	https://socamper.com/	12 26.09%	4 14.29%	4 14.81%	25%	2m 28s	100%
☐	https://socamper.com/what-best-rv-live-full-time/	9 19.57%	6 21.43%	6 22.22%	50%	18s	100%
☐	https://socamper.com/tips-backpacking-europe-budget/	7 15.22%	5 17.86%	5 18.52%	60%	1m 39s	100%
	https://socamper.com/hiking-trails-valley-fire/						

In Piwik PRO, we can perform advanced searches by dimension (even using regex) and string together multiple conditions, just like we did in UA. We can pick separate top-level dimensions (title, entry page, exit page, page speed), and select from available secondary dimensions under the "+" sign — which was not obvious to me at first.

Like Matomo, Piwik PRO has also made improvements to the UA interface. For example, Piwik PRO has added the label "nested dimensions" into the report above, which shows what you'll see when you drill down. UA allows drill down in the same way, but it doesn't tell you what dimension you'll get. That's a small UI improvement, but one that shows a good attention to real-world usage.

Piwik PRO has published a UA to Piwik PRO migration guide and cheat sheet showing where in Piwik PRO you can find the UA equivalent: <https://help.piwik.pro/support/questions/how-do-i-set-up-piwik-pro-after-using-google-universal-analytics-ga3-and-google-tag-manager/>.

Piwik PRO also has some of the best Google Ads and Google Search Console (GSC) integrations of any GA alternative. The Google Ads integration is incredibly easy to set up. The GSC integration takes a little more work and requires using the Google Cloud Platform to set up API access, not just a simple click-through OAuth.

Let's look at another specific example that helps elucidate the pros and cons of following closely in UA's footsteps.

One of the more powerful tools in UA is the Multi-Channel Funnels area and the Model Comparison Tool. GA4 completely abandons this, but Piwik PRO has a reproduction, with even more features than UA has available as custom conditions:

Manage custom models

Model name *

Baseline model

Last click Position based First click Time decay Linear

Visibility

☒ **Author**
jason@quantable.com (me)

☐ **All users**
All users with access to this site or app can see this custom model

☐ **All sites & apps**
This custom model will be available on all sites & apps

Credit conditions

If conditions are not set, 100 % of credit goes to the last interaction. If conditions are set, 100 % of credit goes to the last interaction that meets them. If conditions are set but no interaction meets them, 100 % of credit goes to any last interaction.

Search

Session

Session time

Page views in session

Unique page views in session

Session entry URL

Session entry title

Bounced session

User

Traffic

Device & platform

Location

Custom dimensions

Custom variables

Google Ads

This is great if you love this tool in UA and want to keep using it. The downside is that many analysts, including myself, have struggled to use and

justify these types of models. While these simple, rule-based models were popular for a while, they've since fallen out of favor. The industry has mostly moved to data-driven attribution and modeled attributions or other paradigms altogether, such as marketing mix modeling. Heuristics-based multi-touch attribution is something I believe should have died with UA, not replicated rule-for-rule in a new form.

One place Piwik PRO unfortunately does *not* follow UA is enhanced ecommerce tracking. Piwik PRO has basic ecommerce tracking, but not nearly so many options as UA. Measuring things such as promotions or promo codes has to be done with custom events, rather than being part of ecommerce measurement specifically.

Piwik PRO also feels rough around the edges in places. It has a more advanced UI than Matomo does, but it also has more features it has to fit into the UI. Functionally, it's more akin to enterprise tools like Adobe, Amplitude, or UA — but without the polish and consistency of those tools. For example, all of the charts in Piwik PRO have a line drawing animation on them that causes them to take a moment to render. This is the default behavior of the charting package they use called Recharts. If you are working with charts all day you don't want to have to wait for an animation to complete, even if it's only a second. This same charting package is also used in the popular SEO analytics tool Ahrefs, but they have wisely turned off the animations. A small annoyance, but the kind of UI detail Piwik PRO still needs to work on.

Note that I didn't evaluate the tag manager or consent manager; as with all the alternatives we're evaluating, I used Google Tag Manager (GTM). Piwik PRO's tag manager does look similar to GTM's previous interface and seems straightforward. However, it wouldn't make sense to have multiple tag managers in use. If you have a large existing GTM installation, there's no great incentive to shift to Piwik PRO's tag manager.

Piwik PRO is overall an excellent tool. It might not be UA, but it's the next best thing and improving.

CHAPTER 17 - CLICKY

Type	Traditional Web Analytics
Default Tracking Method	User-Agent + IP
Popularity	#13, 0.3% of top 1M sites
Price	\$10/mo. 30k daily pageviews \$20 for 100k \$188 for 800k
Home page	https://clicky.com

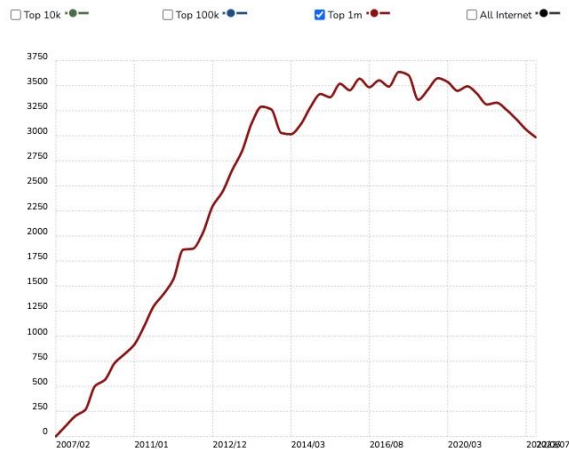
Pros: Mature product, price scales very well

Cons: Outdated UI, rudimentary add-on features

Clicky is another long-standing GA alternative. It launched in 2006, making it the second-oldest alternative we look at behind Statcounter. I first used the product in 2011 and felt it was a market leader at that time. Reviewing the product now, I found little has changed in subsequent years. The upside of this approach is that Clicky has remained consistent and slowly introduced new features and add-ons over time. Unfortunately, like UA, its core design has experienced few fundamental changes since its creation.

Per BuiltWith, Clicky's market share plateaued in 2015 and has started to decrease in recent years.

Clicky Usage Statistics



Much of the loss is to newer options like Cloudflare, Plausible, and Snowplow. Identifying the products Clicky is gaining and losing market share to is a good indicator of which tools are up and down in the industry overall.

Clicky is gaining users from ...	... and losing to ...
GA	Snowplow
Matomo	Yandex
Adobe	Cloudflare
Mixpanel	Plausible
	Piwik PRO

Clicky's pricing model scales very well. For a site with 800,000 daily pageviews (comparable to our 25 million a month large site benchmark), it's only \$188 per month. This makes Clicky the lowest priced option for larger sites, beating out Statcounter by a small amount. Clicky also supports a selectable number of sites, so you could run 500 different sites all under that same \$188 monthly plan, assuming they add up to less than 800,000 daily pageviews combined.

Pairing this advantageous price model with Clicky's white-labeling ability makes it a great option for running analytics on a large number of smaller sites. I used Clicky for that exact scenario in the early 2010s — white-labeled for hundreds of similar but unrelated sites all under one umbrella, and it proved very effective.

The reporting is similar to Statcounter and others in the simplified category, but it's more robust than the options in that group. This reporting, along with the wide range of options for customization, lands Clicky in the traditional web analytics category rather than the simplified, but it could have been placed in either.

Here's what Clicky's main dashboard page looks like:

Home

Spy

Visitors

Content

Searches

Links

Platforms

Locale

Goals

Campaigns

Uptime

Prefs

The Basics

Summary

Visitors

Actions

Uniques

Time

 Visitors	Expand	421	+999%
 Actions	Expand	515	+999%
 Average actions		1.2	+9%
 Total time		13:51:29	+999%
 Average time per visit		01:59	+91%
 Bounce rate		53%	+241%

Recent visitors

List

Most active

8:42p	 Microsoft Azure	/tips-planning-trip-hiking/	
8:41p	 Microsoft Azure	/rain-camping-tent-sleeping-...	
8:40p	 Microsoft Azure	/disclaimer/	
8:39p	 Microsoft Azure	/about/	
8:38p	 Microsoft Azure	/	
9:41a	 Googlebot	/	
8:39a	 Bell Canada	/lightweight-easy-hiking-bac...	
8:37a	 Google Wifi	/paleo-breakfast-camping-re...	
5:45a	 Googlebot	/	
3:39a	 Googlebot	/	
1:48a	 Googlebot	/	
9:50p	 Googlebot	/	
8:57p	 Googlebot	/	
1:52p	 Ameriprise Financial Servi...	/cast-iron-skillet-recipes-ca...	

Links

Incoming

Domains

Recent

Unique

Outgoing

reddit.com		69	+999%
out.reddit.com		44	+267%
chartbeat.com		13	0%
search.answers.com		1	0%
rootviewnetworks.com		1	0%

Searches

Searches

Keywords

Recent

Unique

Rankings

[secure search]	38	+999%
-----------------	----	-------

Visitors



Content

Pages

Entrance

Exit

Downloads

Events

Media

/		272	0%
Home - soCamper			
/how-find-filter-purify-water-survival/		25	+317%
How to Find, Filter and Purify Water for Survival - soCamper			
/what-best-rv-live-full-time/		24	+200%
What is the Best RV to Live In Full-Time? - soCamper			
/tips-backpacking-europe-budget/		23	+667%
25 Tips for Backpacking Europe on a Budget - soCamper			
/benefits-walking-barefoot-earth/		16	+999%
20 Benefits of Walking Barefoot on Earth - soCamper			
/cast-iron-skillet-recipes-camping/		13	+999%
10 Cast Iron Skillet Recipes for Camping - soCamper			
/paleo-breakfast-camping-recipes/		12	+999%
7 Paleo Breakfast Camping Recipes - soCamper			
/most-important-rv-features-you-need/		12	0%
The 15 Most Important RV Features that You Need - soCamper			
/about/		10	0%
About - soCamper			
/category/camping-gear/		9	0%
Camping Gear Archives - soCamper			

Locale

Countries

Cities

Languages

Org

Hostnames

 The United States	357	+999%
 Canada	34	+386%
 Belgium	5	0%
 Israel	2	0%
 Austria	2	0%
 Australia	2	0%
 The United Kingdom	2	+100%
 Bermuda	2	0%
 China	2	0%
 Ireland	1	0%

Traffic sources

 Direct	254	0%
 Social media	113	+528%
 Searches	38	+999%
 Links	15	0%
 Advertising	1	0%

As you can see, it uses small fonts, making for poor readability, and has an outdated look with an older icon set.

The font Clicky uses, Trebuchet MS, was one of the most-used fonts in 2009 (according to Wikipedia) but has dropped in popularity since. If you think Clicky looks old, this font choice might be part of the reason. It's important to point this out since it highlights how UI elements that are purely superficial can change our perceptions so easily even if they are irrelevant to a product's reports and data functionality.

It's admittedly hard to look past an outdated UI, since it also may correlate with outdated functionality. I don't find Clicky's functionality outdated, though it does operate in the traditional web analytics paradigm that many product analytics vendors would likely describe as outmoded.

If we drill into the "Content" report, we see a number of sub-reports, including some of the dimensions (like domains) that we don't typically get in the simplified tools.

Home Spy Visitors Content Searches Links Platforms Locale Goals Campaigns Uptime Prefs									
Pages Entrance Exit Downloads Events Media Domains									
Page	Filter results...	Views ▼	Visitors	Avg time	Total time	Exit			
☆ / Home - soCamper		272 53.3%	246	00:18	1:21:36	95%	0%		
☆ /how-find-filter-purify-water-survival/ How to Find, Filter and Purify Water for Survival - soCamper		25 4.9%	25	01:59	49:35	100%	+317%		
☆ /what-best-rv-live-full-time/ What is the Best RV to Live in Full-Time? - soCamper		24 4.7%	24	02:10	52:00	100%	+200%		
☆ /tips-backpacking-europe-budget/ 25 Tips for Backpacking Europe on a Budget - soCamper		23 4.5%	23	03:58	1:31:14	100%	+667%		
☆ /benefits-walking-barefoot-earth/ 20 Benefits of Walking Barefoot on Earth - soCamper		16 3.1%	16	02:15	36:00	100%	+999%		

You may look at their top-level navigation and think "Spy" sounds privacy-unfriendly, but it's just their real-time stats page. Overall, Clicky is

privacy-friendly and uses IP + User-Agent to identify a visitor rather than cookies by default, though it supports both. Clicky wasn't designed with privacy in mind the way Cloudflare, Fathom, and Plausible were, but it's been retrofitted with good privacy features.

As Clicky still allows cookie-based tracking and that was historically the default, it has more issues with tracker blockers than other tools such as Plausible or Fathom which utilize solely IP + User-Agent.

For example, Safari sends Clicky requests through the Apple iCloud Private Relay (<https://www.daasity.com/post/icloud-private-relay>) whether you use cookies or not. iCloud Private Relay functions as a VPN for those tracking requests, which changes the IP address of the end user and can cause the location reported in Clicky to be inaccurate. Fathom and Plausible were not affected by this in my testing. This is likely because Fathom and Plausible are not listed as invasive trackers on the DuckDuckGo Tracker Radar list (<https://github.com/duckduckgo/tracker-radar>) which seeds ITP, and the ITP machine learning algorithm didn't detect them as a cross-site tracking threat.

Private Relay is a future concern with all of our User Agent + IP tracking-based systems however, since that system changes a user's IP and may rotate those IPs randomly. In my testing, Private Relay didn't rotate the IPs quickly, so it only broke the location information, not the base visitor counting method.

Clicky also has additional features like uptime monitoring and SEO rank tracking integration with third-party service sheerseo.com. In most cases, I recommend separate tools for uptime monitoring and SEO rank tracking, but for simpler sites, it can be nice to have them all in the same place.

Unfortunately, Clicky's implementation of these features tends to be very limited. For example, their split testing feature only tracks which variation drove a conversion, similar to using a session-scoped custom dimension on a goal in UA and unlike what you get with Google Optimize (<https://clicky.com/help/faq/features/split-tests/how>).

My personal preference regarding IT solutions leans towards best-of-breed options rather than all-inclusive suites. For example, if I need an analytics tool and an A/B testing tool, I prefer to pick the best match for each, even if that means using two separate tools. It's a balancing act and another case where we need to investigate the details of what an alternative offers and not assume it's the same as what we're used to with Google.

If you choose Clicky thinking you can run robust A/B testing or powerful uptime monitoring, you may find their implementations insufficient. In Clicky's defense, it also offers A/B testing integrations with Visual Website Optimizer and Convert.com and doesn't try to sell its split testing solution as more than what it is.

Clicky may be a good option if you need more reporting than what you'd receive from a minimalist system like Cloudflare, but don't need the full depth of reporting and customization you had with Universal Analytics. That middle ground covers a lot of use cases that Clicky is well-positioned to handle, especially if they are able to modernize the product some more.

SIMPLIFIED WEB ANALYTICS

CLOUDFLARE

Statcounter

Fathom

Plausible Analytics

Visitor Analytics

CHAPTER 18 - CLOUDFLARE WEB ANALYTICS

Type	Simplified Web Analytics
Default Tracking Method	Referrer (no cookies)
Popularity	#4, 2% of top 1M sites
Price	Free
Home page	https://www.cloudflare.com/web-analytics/

Pros: Free, robust privacy, strong technical underpinning, unsampled site speed monitoring

Cons: Limited tracking abilities, no events, tied to complex platform

Cloudflare's analytics offerings can be confusing. We're evaluating their web analytics offering, which launched in 2020, but there are other Cloudflare solutions that look similar and are frequently mistaken for the web analytics tool.

Cloudflare Web Analytics is free with no volume restrictions, although there are some feature limits if you aren't on a regular Cloudflare plan (restrictions are described here: <https://developers.cloudflare.com/analytics/web-analytics/understanding-web-analytics/limits/>).

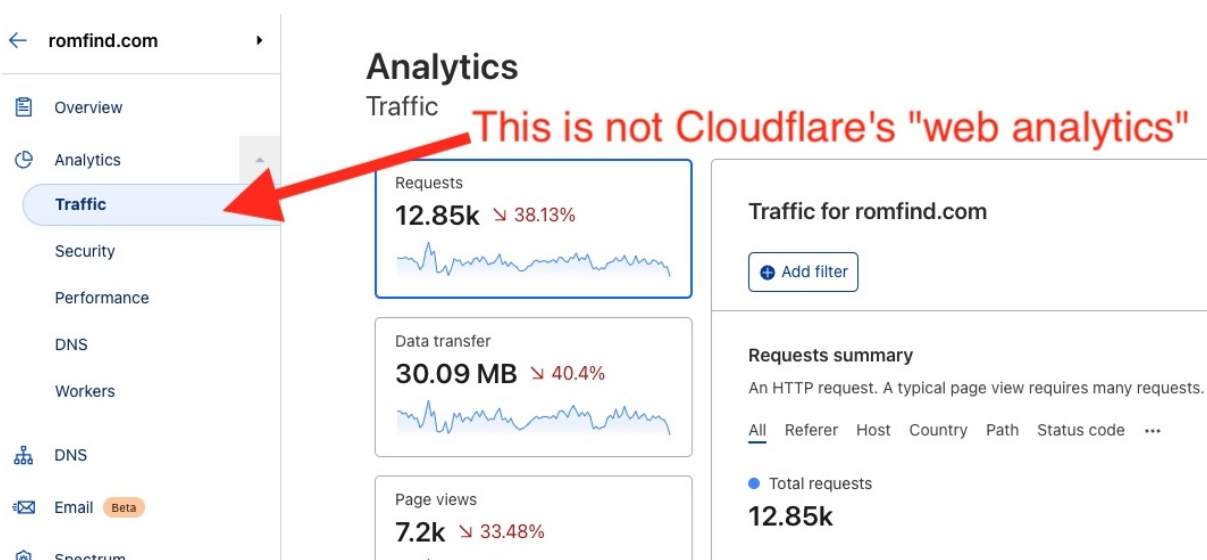
You can deploy Cloudflare Web Analytics as a JavaScript snippet on your site like any other analytics tag. Alternatively you can deploy it

automatically via the Cloudflare proxy if your site uses the regular Cloudflare service. The "regular" Cloudflare service is a CDN and reverse proxy service that sits in front of your website and provides a full suite of tools, from caching and security to image optimization.

More than 10 million sites use the regular Cloudflare service, but this service doesn't automatically include the web analytics tool. The web analytics tool can be incorporated either as part of their regular service or as a separate *à la carte* offering.

BuiltWith detects Cloudflare Web Analytics running on about 19,000 of the top one million sites (1.9% market share) vs. regular Cloudflare, which runs on 24% of the top one million, so only a fraction of their customers (less than 10%) appear to employ the web analytics tool.

The main point of confusion is that the regular Cloudflare service includes a "traffic analytics" offering available under the individual website options in their nav:



Traffic analytics reports metrics we associate with web analytics, like pageviews, but it's not their official "web analytics" tool. This traffic analytics tool is run automatically on all requests that come through the

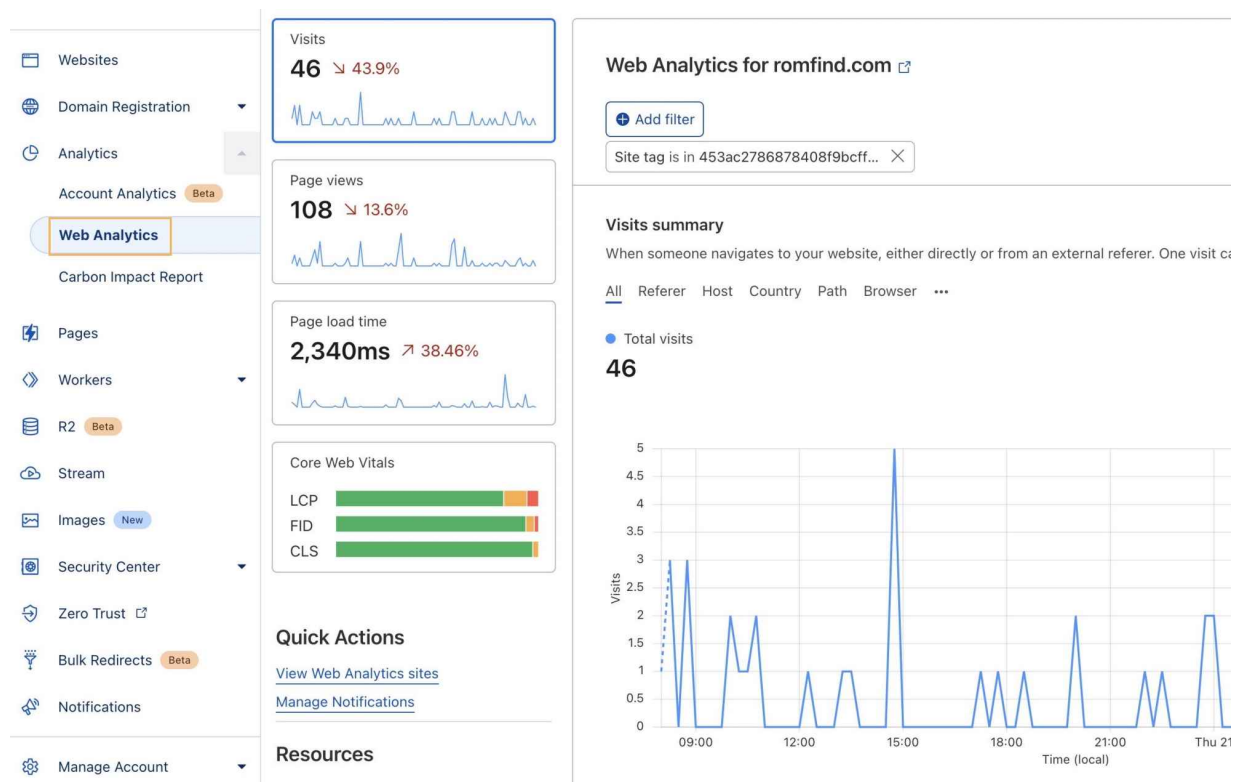
Cloudflare reverse proxy, so it intentionally includes lots of bot and other types of traffic that a web analytics platform generally wouldn't report. On our test site (which has significant bot traffic), the traffic analytics reported **100 times** the amount of users GA did. Also, these numbers are *not* free; you need to be on the \$20 monthly pro plan or higher to see them, compared to the web analytics numbers, which are free.

Cloudflare's supporting documentation doesn't do much to clear up the confusion. Under the "Cloudflare Analytics" section, it says, "It is possible that your Cloudflare metrics do not fully align with data for the same site as reported by other analytics sources, such as Google Analytics and web server logs." (<https://developers.cloudflare.com/fundamentals/data-products/about-analytics/>).

I'd say a hundred-fold traffic increase counts as "not fully aligned." There is information about the web analytics product under the "Types of analytics" link on that page, but it seems odd to mention data discrepancies with GA on their main analytics documentation page and yet not mention their tool that actually will line up reasonably well with GA.

I've spent a fair amount of this chapter telling you what Cloudflare Web Analytics is not, because it's a common point of confusion. I was flummoxed by this the first time I tried to use Cloudflare Web Analytics, and I've seen multiple online reviews that claim to be about Cloudflare Web Analytics but actually cover Cloudflare Traffic Analytics.

Let's look at the correct feature:



This is their web analytics tool. It looks and functions like their traffic analytics but is located at the account level rather than inside a particular website.

Similar to the other tools in the simplified category, this tool is limited by design. It doesn't use cookies or any session tracking technology at all, so it's incapable of tracking visitor paths through the site — no flow tracking, no goal tracking, not even events (although the FAQ says the company may add events in the future). I'd argue Cloudflare is the simplest of the simplified, probably too much for most use cases.

Unlike Plausible, Fathom, and Clicky, which employ hashed IP + User-Agent, Cloudflare uses no session identifier. It tracks visitors by looking at HTTP referrer headers. So, any time it sees a request without an outside referrer, it counts it as a new visitor:

- Arrive from google.com: new visitor

- Arrive directly (no referrer): new visitor
- Click from home page to about page: *not* a new visitor (internal referrer)

This is a clever way to count visits and isn't unique to Cloudflare. Cloudflare doesn't provide a detailed summary of how it applies this counting rule, but Simple Analytics (which we're not evaluating but is a similar platform to Plausible and Fathom) also uses referrers and explains it well:

<https://docs.simpleanalytics.com/explained/unique-visits>

In my testing, Cloudflare over-reported visitors by about 30% vs. UA. This is vastly more accurate than the one-hundred-fold over-report from Cloudflare traffic analytics, but it's still less accurate than the IP + User-Agent method, which itself is less accurate than cookies. The upside of this method is it's great for privacy and compliance. There's no cookie set, no IPs tracked, and no ability to identify individual users. For better or worse, these restrictions make it hard to learn much about your site visitors.

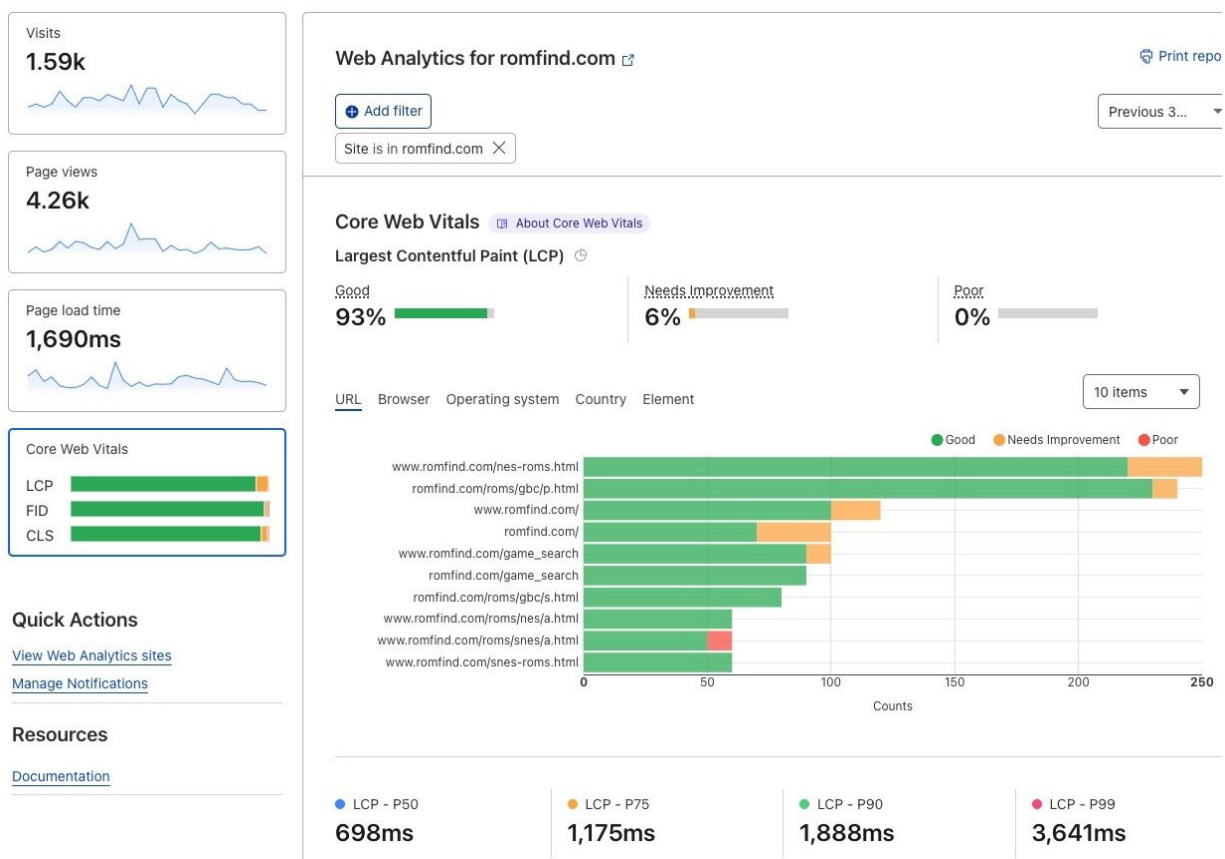
I consider this method to be superior to the method of IP + User-Agent from a privacy perspective. A potential privacy drawback though is you send traffic through Cloudflare, one of the largest collectors of data on the internet.

Cloudflare is U.S.-based and sees about **20% of all traffic on the internet** route through its systems (<https://twitter.com/AxelrodG/status/1447938954758705155>). Cloudflare Web Analytics offers no choice for data location, which could potentially be a compliance issue for EU-based companies based on the data transfers. There's no personally identifiable info tracked within Cloudflare, but the company does see the user's IP, which could be a GDPR violation.

Another Cloudflare feature that's unique among the evaluated alternatives is that they perform their web analytics measurement through the browser performance API (<https://developer.mozilla.org/en-US/docs/Web/API/Performance>) rather than a traditional web beacon. This

is a smart move because it allows them to provide detailed website speed data without additional calls on 100% of the data (i.e., no sampling of page speed tracking like in UA, Piwik PRO, and others).

Here's an example of Cloudflare's web performance reporting, based on real user measurement (RUM, field-based) rather than automated testing (lab-based). As previously discussed, RUM is more important for SEO as that's what Google uses for Core Web Vitals. A more detailed description of the differences between field-based and lab-based measurements is available from Google here: <https://web.dev/lab-and-field-data-differences>.



With Cloudflare, you can drill down into LCP, FID, and CLS in detail (see Chapter 11 for definitions of these). Because Cloudflare currently has no limits (that I'm aware of) on the volume of traffic, this could be a great free option versus tools like [Pingdom](#) that charge significant amounts of money at scale for RUM data. Cloudflare's depth and reporting flexibility are far less sophisticated than more robust paid RUM tools like [SpeedCurve](#)

or [Calibre](#), but it's a strong, free option to monitor Core Web Vitals numbers. Many sites are dependent upon the free Core Web Vitals monitoring that's part of the [Google CrUX](#) dataset, which is an excellent source of data but is delayed (released for the previous month on the second Tuesday of the month) and only includes data from Chrome.

If you already use Cloudflare for some other purpose, turning on their web analytics tool as well could be a fruitful supplement, especially if you want to monitor Core Web Vitals. However, as a standalone product, it's too basic to replace UA for many use cases.

CHAPTER 19 - STATCOUNTER

Type	Simplified Web Analytics
Default Tracking Method	Cookies
Popularity	#9, 0.6% of top 1M sites
Price	\$9/mo. for 100k pageviews \$59 for 2M \$249 for 50M
Home page	https://statcounter.com

Pros: Mature product, affordable, full-featured (within its category).

Cons: Weak on privacy and data ownership

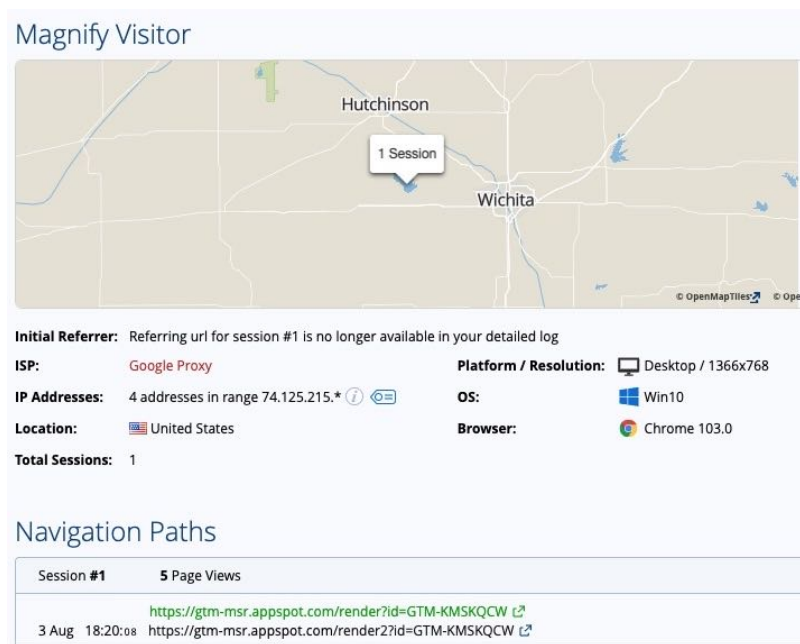
Statcounter is the longest-running option we are evaluating. It was established in 1999, well before Google Analytics existed. Like the other options we look at in the simplified category, it doesn't have an extensive feature set.

Unlike Cloudflare, Fathom, and Plausible — this more basic reporting isn't part of a strategic trade-off for improved privacy. Meaning Statcounter tracks significantly less than UA, but we haven't gained user privacy features in exchange.

Statcounter uses cookies by default (though they can be switched off), and full IP addresses are stored and available in reports. In comparison, UA doesn't make IPs available in the product and forbids the storage of IP in

custom fields. IPs are typically available in the product analytics tools we review, but not in the simplified analytics tools.

Here's an example of the level of detail provided an individual user in Statcounter:



This is much more information than GA returns, including the IP of the visit, other IPs that the same visitor came from, and the ISP information. (Note: for privacy reasons this screenshot is of bot traffic, not a real user.)

Since Statcounter isn't connected to the world's largest ad network, it's arguably still a more privacy-friendly choice than Google Analytics. However, in the context of the simplified alternatives we evaluate, it's one of the weakest on privacy.

On the plus side, Statcounter's pricing is attractive and scales well, charging only \$249 per month for a site with 50 million monthly pageviews. This makes it one of the cheapest options available for large-volume sites.

For a modest additional fee (starting at another \$9 a month), Statcounter also includes a session recording tool, similar to products like HotJar or Clarity. This session recording tool is integrated into Statcounter's standard reporting, meaning once you identify a visitor to investigate, you can see their session recordings attached to their regular analytics reporting. The functionality available in the session recording tool is minimal compared to tools like HotJar, but it's much cheaper and delivered via the same existing tag so there's nothing additional to install. This offering is not unique among the alternatives we assess, but it's still an attractive feature.

The pageview-based reporting will be familiar to users of any traditional web analytics product and has the basic reporting you'd expect:



"Custom Tags" are free-form key value pairs that can be used to tag and segment traffic. Though not as useful as a custom dimension in UA for reporting purposes, Statcounter's version is easier to set up since it's fully free-form (meaning you don't have to pre-define custom dimension slots like in UA).

Statcounter is perhaps best known for their browser and search engine industry stats numbers (<https://gs.statcounter.com>). These stats are widely cited, in fact I referenced them myself in Chapter 13. The stats come from the sites that run Statcounter, and, unlike Google Analytics Benchmarking, there's no obvious way to opt out. Meaning when you run Statcounter your

users contribute to those industry benchmarking numbers whether you like it or not. According to Statcounter, the stats come from 1.5 million sites, which is good for the quality of the benchmark data but problematic for the users of the analytics service who lack a clear opt-out choice.

In addition to the weak privacy features, there are also concerns about data ownership. The following statement appears in the Statcounter terms (<https://statcounter.com/about/legal/>):

"We both own the data regarding visitors to your Web site that we collect. You can use the data we provide for any legal purposes. We will use the data in compliance with our privacy policy."

My interpretation of that somewhat paradoxical statement is that Statcounter claims joint ownership of the data collected, which I find highly problematic.

This is a more laissez-faire approach to data ownership than Google's. Some claims (including one from Matomo here: <https://matomo.org/google-owns-your-data/>) assert Google owns your site's analytics data, but that's incorrect; in GA's terms (<https://marketingplatform.google.com/about/analytics/terms/us/>), Google says they're allowed to "retain and use" your data, but they don't claim ownership. They explicitly state you retain ownership here: <https://blog.google/around-the-globe/google-europe/google-analytics-facts/>

"[Organizations] retain ownership of the data they collect using Google Analytics"

These issues make me hesitant to recommend Statcounter for any site concerned about data privacy. I think the price and functionality is solid and appropriate for many content-oriented websites, but its privacy stance needs to be updated to reflect current standards.

CHAPTER 20 - CHARTBEAT

Type	Simplified Web Analytics (content publishing)
Default Tracking Method	Cookies
Popularity	#11, 0.3% of top 1M sites
Price	Starts at \$580/mo.
Home page	https://chartbeat.com

Pros: Ease of use, focus on content publishing

Cons: Price, limited feature set

Chartbeat is a difficult tool to categorize. I placed it in the simplified category because it doesn't have a lot of metrics or customization, but it does have some sophisticated segmentation and optimization tools beyond what is typical for the category. It's also aimed solely at high traffic sites, which no other option in the simplified category is.

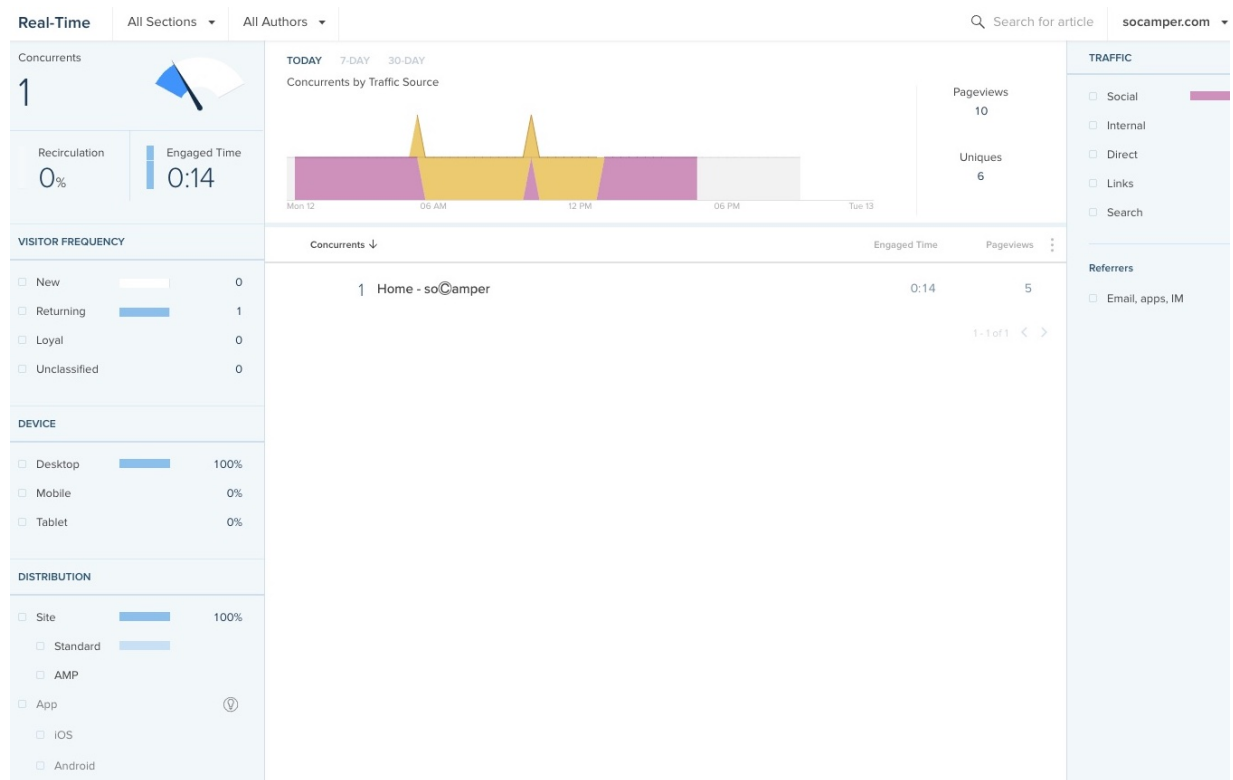
Chartbeat specifically targets content publishers like newspapers and magazines. If you operate in ecommerce, SaaS, or anything not focused on published content, then Chartbeat isn't a good match for your site. An industry-specific product like Chartbeat making our list based on installation volume is a testament to its dominance within its chosen market. It's also an indicator of how many of these type of sites have high

traffic. Chartbeat commands a respectable 0.3% of all sites in the top one million and a whopping 4.3% of the top 10,000 — almost as high in that segment as Cloudflare (4.6% of the top 10,000 and our most popular evaluated alternative overall).

If you run a high-traffic website focused on content and want to give your editorial team access to easy-to-understand analytics, Chartbeat might be the right solution. Because it's a niche product, this chapter will be less in-depth compared to others.

When Chartbeat launched in 2009, it was more of a direct Google Analytics alternative. Notably, Chartbeat offered real-time analytics, which GA didn't at the time. Chartbeat had a low introductory price in the 2010s (less than \$20 a month) and was marketed at bloggers and other content sites. Now, it has a much higher entry price (\$580 monthly) and targets large news media sites. Chartbeat does offer a three-month, low-pageview demo, which I used to test the product.

Real-time data remains the tool's focus. However, since my test sites have small amounts traffic, those real-time reports looked empty:



Concurrents ↓

Engaged Time

Pageviews

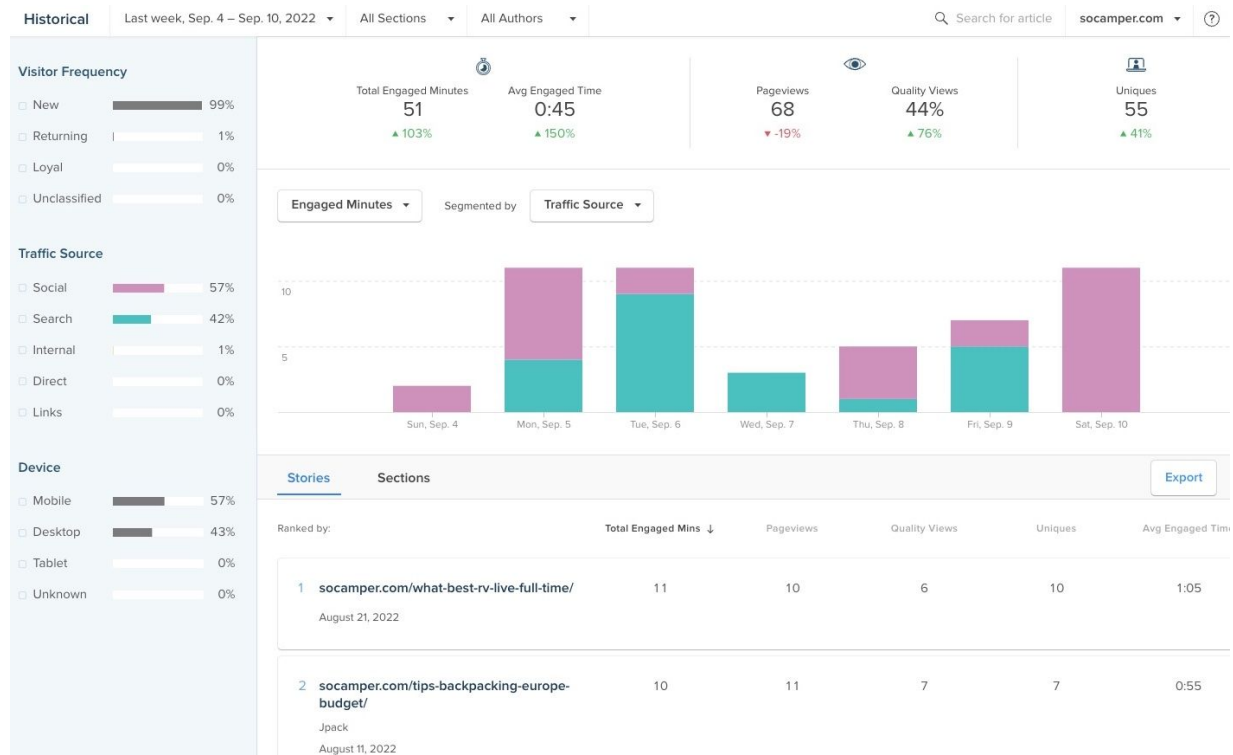
1 Home - socamper.com

0:14

5

1 - 1 of 1 < >

There's also historical reporting, and the metrics and reporting are similar to the real-time dashboard:



Chartbeat's well-designed dashboards focus on metrics that interest content teams, including user engagement with easy segmentation by content category (section), author, device type, and more.

Additionally, Chartbeat has tools for A/B testing content and advanced segmentation reporting. Neither of those are available in the trial, so I couldn't test them.

I can see how Chartbeat would be a popular tool with a content team or newsroom, as it's easy to use and has the functionality and point of view those teams want.

Its entry price is high, but that appears to be for five million pageviews, so the cost isn't exorbitant compared to other tools (outside the simplified category) for the same volume. Still, I see this as a targeted product to be used in conjunction with another analytics tool that could cover most of the same functionality. So, if you ran Chartbeat alongside another more

generalized analytics tool, you wouldn't gain significant additional functionality with Chartbeat.

CHAPTER 21 - FATHOM

Type	Simplified Web Analytics
Default Tracking Method	User-Agent + IP
Popularity	#14, 0.2% of top 1M sites
Price	\$14/mo. for 100k pageviews \$74/mo. for 2M pageviews
Home page	https://usefathom.com

Pros: Privacy, EU isolation mode, uptime monitoring

Cons: Limited reporting and filtering

Fathom is similar to Plausible in price and functionality, as discussed in Chapter 10 - Open Source. Fathom is currently slightly cheaper overall, but Plausible's plans start at a lower traffic level.

Pageviews/mo.	Fathom	Plausible
10k	-	\$9
100k	\$14	\$19
2M	\$74	\$89

As with many of the offerings in this category, you can see most of what the product can do from the main dashboard.



Device Types

Visitors

Desktop	114
Phone	90
Tablet	3

Q

Browsers

Visitors

Chrome	111
Safari	63
Firefox	15
Edge	13
Mozilla	3

Q

Countries

Visitors

United States of America	152
Canada	17
China	12
Australia	3
United Kingdom	3

Q

→

Events Manage events

Unique Completions

Completions

Conv. Rate

Value

Something	1	1	0.3%	\$0
-----------	---	---	------	-----

UTMs

Build Campaign

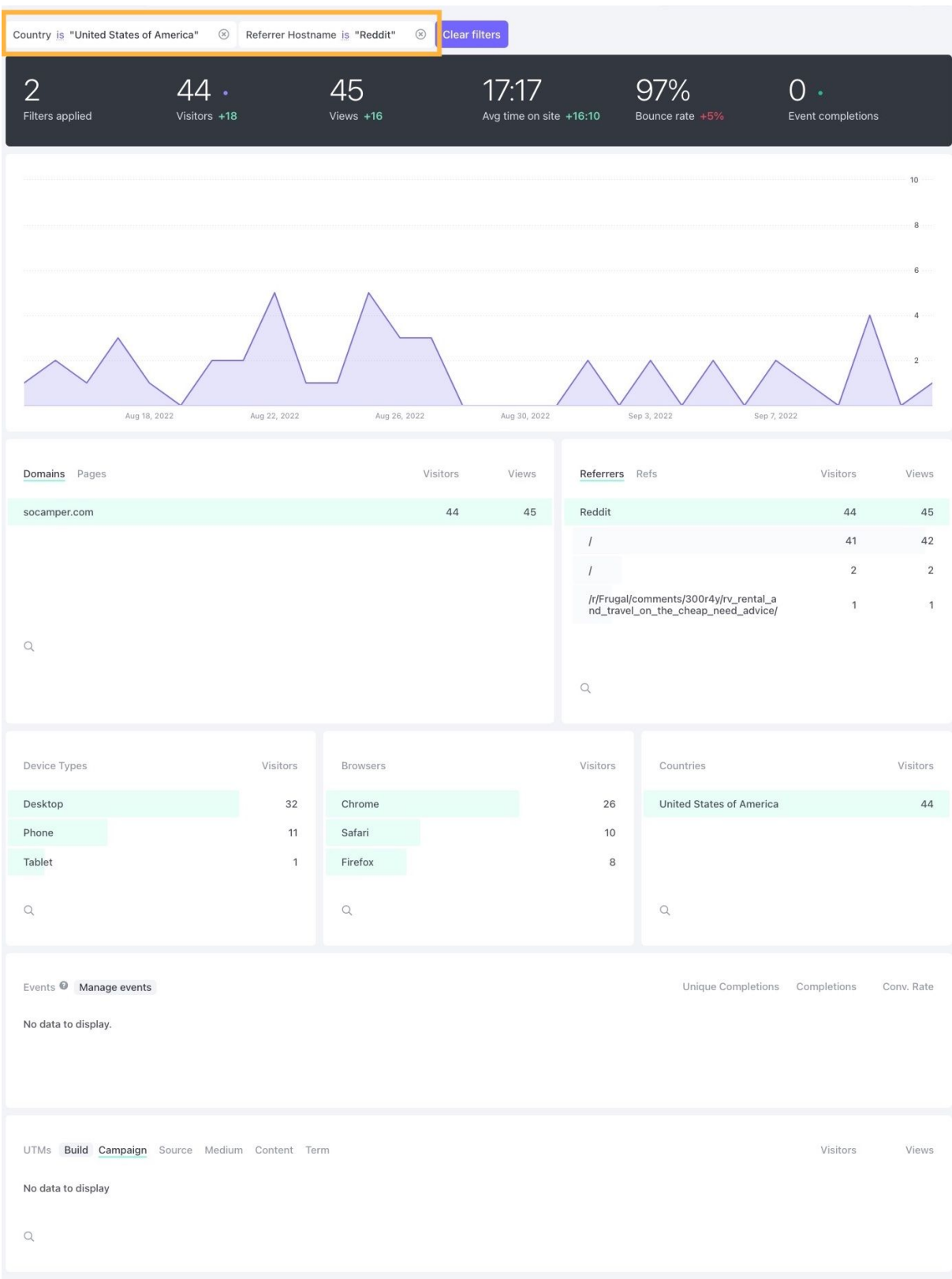
SourceMediumContentTerm

VisitorsViews

No data to display

Q

There's no navigation to drill down into, nor any hidden reports. Fathom is at the opposite end of the spectrum from UA's 100-plus default reports. It lacks custom report building, and there's no secondary metrics to add to our reports. The main action you can perform from the dashboard is filter to a particular metric. For example, if we wanted to see only the U.S. traffic from Reddit, we'd click on each of those in the dashboard to add them as reporting filters.



Device Types

Visitors

Desktop	32
Phone	11
Tablet	1

Q

Browsers

Visitors

Chrome	26
Safari	10
Firefox	8

Q

Countries

Visitors

United States of America	44
--------------------------	----

Q

Events Manage events

Unique Completions

Completions

Conv. Rate

No data to display.

UTMs

Build

Campaign

Source

Medium

Content

Term

Visitors

Views

No data to display

Q

In this second screenshot, you can see I added those filters at the top of the page. Fathom recently (November 7, 2022) launched the ability to filter based upon events as well, which was previously a concern as a long-standing missing basic feature (<https://usefathom.com/blog/event-filtering>).

If you feel this lack of focus on attribution reporting could be a deal breaker, consider that Fathom and Plausible have **no concept of returning visitors across multiple days**. Fathom's user identification method hashes IP addresses in such a way that the same user returning the next day from the same IP would receive a different hash — so, new day, new user, as described here: <https://usefathom.com/blog/anonymization>. This means you should avoid making judgements based on conversions per source, unless you have a very short sales cycle. This is the world of consent-less analytics: significantly reduced detail on users and a lack of reliable attribution.

One difficulty I ran into was that Fathom's default tag didn't work in Google Tag Manager. They provide a different tag in the documentation that *does* work with GTM, but this was unclear from the script settings page and caused me confusion during installation.

On the positive side, Fathom makes it simple to use a custom domain to mitigate data loss from ad blockers. In Chapter 13, I explained how server-side analytics was the best way to solve this issue. Custom domains are an easier but less effective solution. In the case of Fathom, custom domains are effective because the product doesn't use cookies; the custom domain requires DNS setup on your side, nothing more. Plausible offers a similar solution, though does not place it so prominently in the product UI.

If you decide to use this anti-ad block feature, don't deploy your tag via GTM, because many ad blockers block GTM, which would defeat the purpose of a custom domain. Other tag managers may also be subject to ad blocking, so if you use a tag manager for this it should be a self-hosted one.

Because Fathom doesn't use cookies, Safari ITP CNAME-cloaking prevention has no effect on it. If you're unfamiliar with CNAME-cloaking, it's essentially what we are doing in Fathom with this feature: creating a

DNS alias via a CNAME record that maps from a domain under a primary website domain to the analytics tool's domain (<https://webkit.org/blog/11338/cname-cloaking-and-bounce-tracking-defense/>). For example, if our main site was "oursite.com," we could call our tracking subdomain "potato-battery.oursite.com," and point it to Fathom's servers. ITP is designed to root out this kind of aliasing, but its enforcement applies to cookies, so it has no impact here. We shouldn't use a subdomain like "analytics.oursite.com" because ad blockers are more likely to block it.

Fathom's claim that this feature will restore data that would otherwise be lost to ad blocking is correct. In contrast, there are cookie-based solutions that use this kind of evasion method, but ITP downgrades their cookies.

As we mentioned in Chapter 17 - Clicky, Apple's iCloud Private Relay did not affect Fathom or Plausible, but might potentially in the future. The DuckDuckGo Tracker Radar list has Fathom listed as less invasive than Plausible currently, meaning that ITP and iCloud Private Relay are less likely to impact Fathom going forward (though neither were currently affected in my testing).

Like Clicky, Fathom offers site uptime monitoring but takes it a step further by providing notifications to Slack and SMS rather than email alone.

Fathom's "EU isolation" mode is intelligently designed to comply with EU privacy rules without sacrificing data or performance. As Fathom states here <https://usefathom.com/features/eu-isolation>, users in the EU never have any personal data transferred or stored outside the union. Fathom does store de-personalized data on U.S. servers, but that can be turned off for people who want to avoid storing data in the U.S.

Since Fathom and Plausible have both gained significant numbers of customers in recent years, I expect each product will continue to improve. As such, they're both likely to remain good options for cases where minimal tracking and reporting is needed.

CHAPTER 22 - PLAUSIBLE ANALYTICS

☆ FAVORITE

Type	Simplified Web Analytics
Default Tracking Method	User-Agent + IP
Popularity	#15, 0.1% of top 1M sites
Price	\$19/mo. for 100k pageviews \$89 for 2M
Home page	https://plausible.io

Pros: Open source, privacy, easy GA data import

Cons: Limited reporting

Plausible (and Fathom) are, in my mind, the quintessential representations of the "simplified web analytics" category. As stated previously, the two products are the closest in functionality of any of our alternatives. This makes the choice between the two difficult, and I can't make a blanket recommendation for one over the other, as their strengths and weaknesses are situational.

Plausible is currently slightly ahead on functionality. For example, at the time of publication, Plausible has a historical GA data importer and Fathom does not — however I expect Fathom's importer will launch soon. There is also a (paid) Looker Studio connector for Plausible but not for Fathom. On the other hand, Fathom offers site uptime monitoring, which

Plausible does not. I've given Plausible the choice as favorite due mainly to the GA historical data importer.

Why are both products so emblematic of the category?

- They're designed with privacy in mind.
- They have streamlined user interfaces that focus on popular pageview-oriented metrics.
- They're produced by a small team and target smaller sites.
- They have an intentionally limited feature set.

Plausible has limited capacity for customization (unless you want to fork your own version and code the features you need, which I don't recommend). It does have the ability to track events and goals, but, as with UA, you need to tag your events manually. Depending on your use case, this simplicity could be incredibly frustrating or exactly what you need. Less analytically adept users will find it easy to use, which is great in a scenario where you are handing an implementation to a team of non-specialists.

Here's the main reporting view:

UNIQUE VISITORS

246 ↑ 24%

TOTAL PAGEVIEWS

886 ↓ 30%

BOUNCE RATE

49% ↑ 9%

VISIT DURATION

2m 08s ↓ 34%



Top Sources

All Campaigns

Source	Visitors
Google	4
petalsearch.com	3
yandex.com.tr	2
wz6z6z6.ysepan.com	1
Yandex	1
miniature.io	1

DETAILS

Top Pages

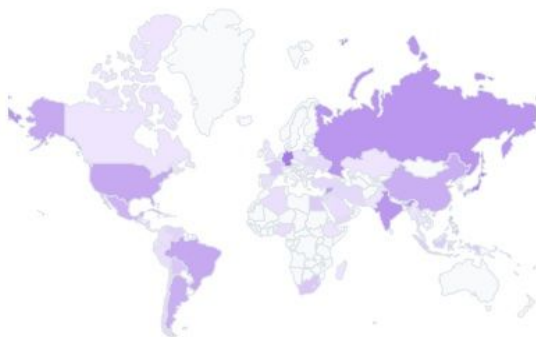
Top Pages Entry Pages Exit Pages

Page	Visitors
/nes-roms.html	99
/	61
/snes-roms.html	29
/game_search	28
/roms/nes/a.html	19
/roms/gbc/p.html	19
/roms/snes/a.html	14
/roms/genesis/a.html	8
/roms/gbc/s.html	5

DETAILS

Locations

Map Countries Regions Cities



DETAILS

Devices

Size Browser OS

Operating system	Visitors
Android	222 (91%)
Windows	15 (6%)
GNU/Linux	4 (2%)
Mac	2 (1%)
iOS	1 (0%)
iPadOS	1 (0%)

Goal Conversions

Goal	Uniques	Total	CR
Visit /game_search	28	54	11.4%

As you can see, Plausible provides a limited number of metrics and dimensions. So, unlike GA, there are only a few ways to perform a task, making reporting more intuitive. What you see is what you get.

Plausible is one of several analytics tools (mostly in the simplified category) we've included in this guide that doesn't use cookies. It relies on more old-school technology in the form of IP combined with User-Agent. The details are described here: <https://plausible.io/data-policy#how-we-count-unique-users-without-cookies>.

Interestingly, this is the same base method Urchin (UA's predecessor) employed before it switched to cookies set by JavaScript in the early 2000s. Some might call this IP + UA method browser fingerprinting, but I prefer to reserve that term for more invasive JavaScript-based solutions which use a wide variety of browser signals.

IP + User-Agent is in every browser request, so Plausible isn't dependent upon additional data from JavaScript browser inspection; the program only uses what's inherently part of each request. This is why IP + User-Agent was the method of choice in the old days: the HTTP header data was typically in the web logs ready for processing. Because Plausible is deployed via JavaScript and not a log analysis tool from yesteryear, it does report on some data that can't be obtained from basic web logs (such as screen size, though it doesn't use that for visitor identification). However, it's otherwise similar in core functionality to a classic log analyzer such as Urchin, AWStats, Analog, Webalizer, etc. If you used those tools back in the day and thought that amount of data was adequate to accomplish what you needed, then you may like Plausible.

Plausible claims this IP + User-Agent approach to visitor counting is reasonably close to Google Analytics' method of long-lived cookie IDs (within 10%). I was skeptical of this assertion, but in my test site, it actually came within 5% of the GA number, with only a slight undercount. However, this accuracy will vary significantly depending on the type of traffic you receive.

One such case where Plausible might significantly underreport would be if much of your traffic comes from shared IPs or VPNs. For example, if you ran a B2B website where a number of your corporate customers had the same browser installed company-wide and the corporate network only had a small number of outgoing IPs, then hundreds of different users might show in Plausible as only one or two.

Unsurprisingly, Apple is making our job harder by hard-coding some parts of the User-Agent. Apple has begun to insert "Intel Mac OS X 10_15_7" as the OS version in the User-Agent no matter what OS version the system runs, which limits the uniqueness in the IP + User-Agent identifier.

Plausible and Fathom both take this old method a step further for additional privacy by salting and hashing that data and rotating the salt every day. "Salting" is the process of appending additional random text to data like a password or user identifier before hashing the data to make it harder to break the hash to reveal its original data (<https://www.theguardian.com/technology/2016/dec/15/passwords-hacking-hashing-salting-sha-2>).

This is excellent for privacy because it means the tool itself doesn't store anything that might be construed as user-identifiable (such as IP address). Unfortunately, this process does have significant downsides for accuracy, which is why Urchin and others switched to cookie session IDs roughly 20 years ago. Also, the daily rotation of the salt means there's no concept of a return visitor or tracking a user over multiple days. This lack of sticky user identification is both a big obstacle for attribution and a boon for compliance.

Plausible's UA data importer works smoothly (although it would be nice if you could easily search through the dropdown that lists the views to import from). Unlike Matomo's importer, you do not even need to setup a GCP project and credentials — it can be done with basic OAuth alone in seconds. The data you import from UA functions no differently than if the data had been collected by Plausible directly. The additional data that UA might have that Plausible does not support is not migrated.

One final note: although Plausible uses a consent-less, privacy-aware implementation, ad blockers that use the EasyPrivacy list (such as Brave or uBlock Origin) still block it. The keepers of such lists have broad definitions of what a "tracker" is and are disinclined to make exceptions for every tool that claims to be privacy-friendly.

If you're looking for basic visitor counting, a reasonable price at scale, and easy compliance (thanks to eschewing cookies), then Plausible is an excellent option. Even at 10 million monthly pageviews, the tool charges just \$169 a month. On the other hand, there's no free tier, so, if you run a low-traffic site (less than 10,000 pageviews), the base \$9 plan may seem like a high price to pay each month if you're used to free UA.

CHAPTER 23 - VISITOR ANALYTICS

Type	Simplified Web Analytics
Default Tracking Method	User-Agent + IP
Popularity	#18, 0.1% of top 1M sites
Price	\$13/mo. for 10k visits \$70/mo. for 100k visits
Home page	https://www.visitor-analytics.io

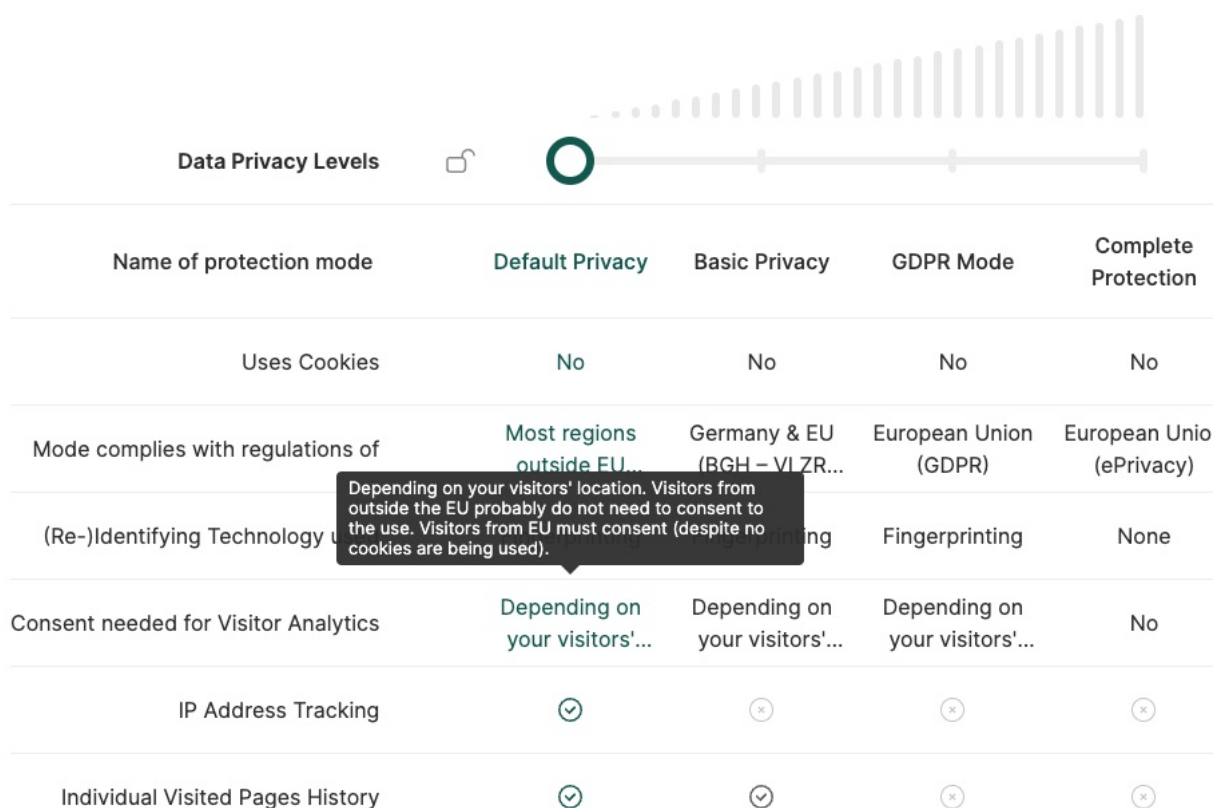
Pros: User-friendly, price, adjustable privacy levels, full suite of features

Cons: Privacy (fingerprinting), breadth over depth, site performance impact

Visitor Analytics is less popular than some of our other GA alternatives (1,207 installs in the top one million websites), but it's become broadly used among lower-traffic websites. The vendor claims more than 2.5 million total installs, and BuiltWith backs up that claim by detecting over 500,000 across all monitored websites. One year of Visitor Analytics is also included for free as part of most plans for the popular site builder Wix.com. Wix is used on 0.6% of the top one million sites, so much of Visitor Analytics' user base is from Wix sites.

I found the web analytics reporting to be user-friendly, with a complete, albeit basic, suite of reporting. Its web analytics capabilities are similar to most of the other simplified category tools. As such, you won't find features like secondary dimensions, custom reports, or advanced filtering. It does have a "competition" area, which operates like a simpler version of UA's benchmark reports, and is unique within this product category.

Manage Privacy-Settings ⓘ



This lets you adjust how the product works based on your location and privacy sensitivity. I love the idea of a simple slider-based adjustment with clear messaging on the privacy impact. My issues can be seen in the tooltip above; EU users must still consent at every level other than "complete production," even though Visitor Analytics sells itself as being consent-less. On their site, the company states (<https://www.visitor-analytics.io/en/why-visitor-analytics/cookieless-tracking/>):

Users don't need to opt-in when cookieless tracking is used and you don't need to worry about any consent banners for this. Cookieless tracking means no privacy regulation related headache anymore.

However, that tooltip indicates consent is still necessary in the EU unless the privacy setting is maxed out, which isn't the default. I agree that **consent is likely needed for fingerprint-based tracking**. Even their "Complete Protection" plan may need consent depending on your jurisdiction and interpretation of regulations. "Complete Protection" uses User-Agent + IP, the same base method used by Plausible and Fathom, but Visitor Analytics does not appear to do the same kind of salt rotation described in the previous chapter.

Another dubious marketing claim is that Visitor Analytics avoids issues with tracker blocking (from <https://www.visitor-analytics.io/en/why-visitor-analytics/why-switch-from-google-analytics-to-visitor-analytics/>):

No cookies means there is nothing to block. With Visitor Analytics you will still get to see 100% of your visits in your dashboard, regardless of the browser.

This is largely incorrect. Content blockers such as uBlock or Brave **do** block Visitor Analytics because their block lists (e.g., EasyPrivacy) contain the visitor-analytics.io domain and block the web request itself.

Firefox doesn't block Visitor Analytics because the domain isn't on the disconnect.me blocking lists at this time, but likely will be added if the service continues to grow. Safari, Brave, and others also contain anti-

fingerprinting techniques (<https://webkit.org/tracking-prevention/#anti-fingerprinting>) so the number of users invisible to your analytics will be more than 0% and likely similar to what you miss with GA.

As mentioned above, Visitor Analytics uses browser fingerprinting to create visitor IDs instead of cookies. They leverage fingerprintjs (<https://github.com/fingerprintjs/fingerprintjs>), a popular fingerprinting tool, to achieve this. Visitor Analytics asserts the advantage of this approach is that consent isn't required because it neither uses cookies nor stores personal information (<https://www.visitor-analytics.io/en/support/legal-data-privacy-certificates/gdpr-compliance/gdpr-compliance-overview/>).

I can't say if this is legally accurate (consult your legal counsel). From a technical perspective, though, my opinion is that fingerprinting is more invasive and less favorable to privacy than standard first-party analytics cookies, and I don't recommend their use for the following reasons:

1. Cookies can be deleted, but fingerprints remain constant until the browser itself is changed.
2. Fingerprints are the same across different sites, opening a greater possibility for cross-site tracking than first-party cookies (Visitor Analytics firmly states they don't do this, but they could, because your browser fingerprint across every site that has their tracker installed could return the same information). I've verified the visitor key the tool sends does vary per site, so they probably don't do cross-site tracking.
3. Fingerprinting can detect the use of incognito mode and maintain the same fingerprint between regular and incognito mode.
4. Fingerprinting looks at your browser in detail and creates a much more unique and durable identifier than any cookie. Visitor Analytics' fingerprinting code looks at these browser settings and more: installed fonts, browser plugins, type and number of CPUs (if available), amount of memory, and canvas detection. See <https://coveryourtracks EFF.org> from the Electronic Frontier Foundation to test your browser for these things. Visitor

Analytics doesn't leverage all possible fingerprinting techniques, but it does use most of the typical ones.

5. Browsers are actively working to stop fingerprinting. One could correctly point out they're also trying to restrict cookie-based tracking, but the browser and web standards community recognizes the utility of cookies (especially server-side cookies) more than fingerprinting — meaning they're looking to *limit* cookies but *eliminate* fingerprinting. Google, for instance, explicitly bans fingerprinting in GA:

<https://support.google.com/analytics/answer/9682282>

I also find Visitor Analytics' service-level agreement (SLA) lacking, which reads like a service with little faith in their own architecture and uptime. In their terms at <https://www.visitor-analytics.io/en/support/legal-data-privacy-certificates/standard-integration/terms-of-use/> (section 21), the company defines average availability as 99.9%, which is pretty typical.

Then, they define availability as only *unscheduled* downtime. In other words, planned downtime doesn't count against the SLA. Although not unheard of, this means the defined uptime should be accordingly higher, but it's not. Concerning penalties, they give a minimal 2.5% service credit for downtime greater than or equal to 99% and less than 99.75%, so they could have up to 1% unscheduled downtime and still only owe customers a 2.5% credit. That's a pretty toothless SLA.

Compare this to Piwik PRO (<https://piwik.pro/cloud-sla/>), which gives a generous 25% credit at less than 99% and includes all downtime in their percentage.

GA4 360's SLA (https://marketingplatform.google.com/about/analytics_products/sla/) starts with a 5% credit at less than 99.9%. Google is stingy with their penalty payouts, but has high standards for availability. For individual high-availability services like BigQuery, Google sets an SLA of 99.99%.

If 99%, 99.9%, and 99.99% all seem basically the same, consider this table:

Availability	Downtime per Month
99%	7.3 hours
99.9%	44 minutes
99.99%	4.4 minutes

If my analytics service was out for seven hours a month, I'd be an unhappy customer, and a 2.5% credit would do little to appease me. In the analytics world, outages most frequently occur at times like the first of the month, when everyone is in the product looking for last month's stats.

My issue isn't that I want higher service credits or availability guarantees; it's that a low-expectation SLA doesn't send the right message about how the vendor operates.

The good news is that in February 2023 Visitor Analytics switched from fingerprinting to User-Agent + IP as their default tracking method. Unfortunately fingerprinting still exists in the other three tracking levels.

3

Data Privacy Mode

Choose a Data Privacy Mode

Four modes of protecting your website visitors' privacy. Pick the privacy mode that fits you best. The stricter your selected privacy mode, the less information will be displayed about your visitors. You can update this setting later.



Complete Protection
Data Privacy Level 4

Highest Visitors Anonymity & Reduces Data Insights

No Cookies & No Fingerprinting Used for Data Tracking

Mode Complies
European Union (GDPR Level)

Consent Needed for Data Tracking:
No Consent Needed



GDPR Privacy
Data Privacy Level 3

High Visitors Anonymity & Reduced Data Insights

No Cookies. Fingerprinting Used for Data Tracking.

Mode Complies
European Union (GDPR Level)

Consent Needed for Data Tracking:
Visitors Consent Potentially Needed



Basic Privacy
Data Privacy Level 2

Medium Visitors Anonymity & Full Data Insights, No IP

No Cookies Used. Fingerprinting Used for Data Tracking

Mode Complies
European Union

Consent Needed for Data Tracking:
Visitors Consent Potentially Needed



Default Privacy
Data Privacy Level 1

Medium Visitors Anonymity & Full Data Insights

No Cookies Used. Fingerprinting Used for Data Tracking

Mode Complies
Almost Everywhere Outside E.U.

Consent Needed for Data Tracking:
Visitor Consent Potentially Needed

Continue

The options during signup are pictured above. Confusingly, "Complete Protection" is the selected default, but there is also a different plan named "Default Privacy" (which was the previous default).

Visitor Analytics has the potential to be an excellent replacement for UA with small sites. It's easy to use and contains a well-considered selection of features. Moving away from fingerprinting by default is a big improvement, but there are still privacy and product maturity issues that need to be addressed before I could give the product a full-throated endorsement.

PRODUCT ANALYTICS

GOOGLE ANALYTICS 4

Mixpanel

Snowplow

Amplitude Analytics

Heap

PostHog

CHAPTER 24 - GOOGLE ANALYTICS 4

Type	Product Analytics
Default Tracking Method	Cookies
Popularity	#1, 47% (all versions), 17% (GA4 only)
Price	Free with minimal limitations Reportedly starts at \$4,167/mo. (much cheaper than UA 360's \$12,500/mo.)
Home page	https://marketingplatform.google.com

Pros: Price, Google integration, community, underlying tech

Cons: Privacy, product immaturity, lack of prebuilt reports

As previously mentioned in this guide, I believe GA4 is a pretty good tool. Considered as a new, standalone product, it has many great features and a favorable pricing model. The torrent of criticism it's received is understandable however, since as a forced upgrade it's not up to the standards set by Universal Analytics.

GA4 is also changing rapidly, with new features being released on a regular basis. This is nice, but it also makes migration more difficult since we're having to migrate to a moving target. Coming from UA, it's great to see new features, as UA has had few substantial updates over the past several years. If you haven't taken a look at GA4 in a while, it's worth another look.

From June to August, 2022, GA4 released the following major updates:

- Behavior modeling for consent mode
- EU-based traffic collection
- Google Ad Manager integration into GA4
- Addition of bounce rate
- Two new UTM parameters
- Access logging API

(full release notes here:

<https://support.google.com/analytics/answer/9164320>)

I'm sympathetic to the argument that many of these recently released features should have existed before GA4 came out of beta, or at least before announcing a forced upgrade. If you're in the process of migrating to GA4 and running into its shortcomings, I feel your pain.

As far as price, nobody gives you more for free than Google (though, as discussed, I see this as subsidized by users' data rather than free). Free GA4 currently has no enforced monthly hit limit. Since GA4's paid tier starts at 25 million, it seems probable that Google could implement a 25 million hit per month limit at some point. The other free plan limit most commonly encountered is the one million daily event limit in the BigQuery export.

Additionally, reduced data retention affects all accounts, with a max of 14 months for free (in Explore reports) and 50 months for paid.

UA offers unlimited data retention, making GA4's 14-month maximum (with a very limited two month default) a significant downgrade. As with UA, the GA4 data retention policy doesn't apply to standard reporting; all the aggregate reporting sticks around indefinitely, but you can't use "Exploration" reports (GA4 report builder) on data older than 14 months. So, if you frequently create custom reports with older data, this could be an issue.

In this data exploration example you can see the UI simply doesn't let us choose a date more than two months in the past because of the default data retention period.

The screenshot shows the Google Analytics 4 interface. On the left, there's a sidebar with 'Variables' and 'Tab Settings'. The 'Exploration Name' is 'Free form'. The 'Tab Settings' section shows 'TECHNIQUE' and 'Free form'. The main area displays a table with columns: 'Device category', 'City', 'Active users', and 'Totals'. The table data includes rows for 'Totals', '(not set)', 'Ashburn', 'Columbus', 'London', 'New York', 'Bengaluru', 'Leominster', 'Pune', 'Amsterdam', and 'Austin'. A date range selector is open, showing 'CUSTOM' with dates 'Aug 6, 2022' to 'Oct 5, 2022'. A calendar view for August 2022 is shown, with a note 'Nothing more than 2 months old' and an arrow pointing to the date '5' in the first row of the calendar grid. The 'Compare' toggle is turned off. The 'Nestled rows' dropdown is set to 'No'. The 'COLUMNS' section shows 'Device category'.

We can increase that lookback limit from the default of two months up to 14 very easily, however 14 is still quite low, and raising this limit doesn't restore already discarded data. The data explorer is one of the core features of GA4, so this is a significant restriction.

You can read Google's full list of limits in free vs. paid GA4 here: <https://support.google.com/analytics/answer/11202874>. Free versions of GA4 and UA are similar in that they've withheld few paid features compared to other "freemium" offerings.

One major advantage paid GA4 360 has over UA 360 is a much lower starting cost at \$50,000 vs. \$150,000 annually. The GA4 pricing structure is significantly different from UA's, although it's still opaque and subject to individual agreements with your GA reseller

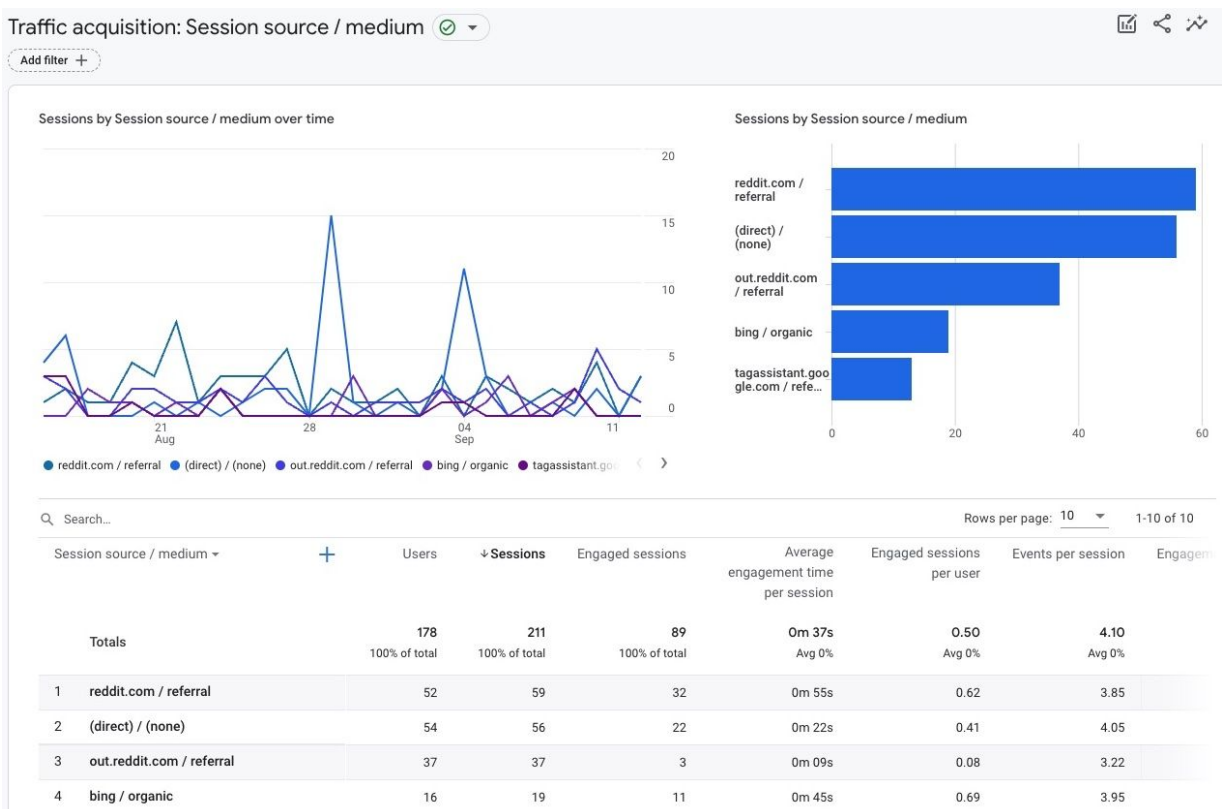
(<https://www.cardinalpath.com/blog/ua-360-vs-ga4-360-pricing-model> has more details on the pricing model).

The data models of UA and GA4 are also starkly different. Here's a helpful comparison of the two:

<https://support.google.com/analytics/answer/9964640>

I think the primary obstacle most people have when migrating from UA to GA4 is that GA4's overall approach is more in line with what we see in the product analytics category rather than traditional web analytics. Google has telegraphed this impending paradigm shift over the last few years, but it still caught much of the industry unawares.

UA has many pre-defined reports that cover most typical reporting requests. In GA4, the data you seek is likely still there (given a similar instrumentation), but either it takes more clicks to find, or it requires building a custom report. Let's look at a common source/medium report as an example:



This is available in UA under *Audience > All Traffic > Source/Medium*. In GA4, it's found under *Acquisition > Traffic acquisition*, then by selecting "source / medium" in the drop-down above the table. The path into the report is a little different, but the output is similar. Compared to UA, it's more complicated to drill down deeper from that point, as it involves customizing the report, creating a filter, or moving over to the data explorer. We can pick a secondary dimension, but we can't perform an advanced search, or use regex in the search box.

This style of reporting focuses on creating custom reports or going deep into exploratory data analysis, compared to ad-hoc data reporting. Simply put, UA encouraged you to poke around and try different dimensions, whereas GA4 assumes a more structured approach that encourages knowing what you want up front.

As stated at the beginning, we shouldn't consider Universal Analytics an available option for comparison, although it's difficult not to use UA as our baseline, especially in the context of GA4. Ultimately, our options going forward don't include UA, so we need to evaluate our other alternatives against GA4 and each other rather than UA.

We placed GA4 in the product analytics category, whereas UA is in the traditional category. GA4 fits less cleanly in the product analytics category than some of the other options like Heap, Mixpanel, or Amplitude. Evaluated solely as a product analytics tool, my experience has been that GA4 underperforms compared to other well-established options in the space.

However, GA4 can do more on the marketing analytics side than many of other options in this category, so, as usual, whether GA4 will work better for you depends on your use case. Part of the pain people are experiencing with GA4 is that it's a product analytics tool, but based on GA's history, many users were expecting another traditional web analytics/marketing analytics tool.

Before GA4, it was relatively common to run UA alongside a product analytics tool. The question now when replacing UA is, does it make sense to run both GA4 and another product analytics tool like Amplitude when there's so much more overlap? The answer to that should be led by reporting needs. It's highly preferable if everyone can work in the same tool, but that's not always practically possible.

If your marketing team wants to use GA4 while your product team wants to use Amplitude and neither teams' critical reporting needs can be serviced from just one tool, then running both tools is a reasonable solution. Working in different tools means those teams won't have the identical answers to the same questions, and it also means they aren't as likely to collaborate. This is why defining what is "critical" in your reporting is so important. If the marketing team can get what they need out of Amplitude or the product team can be satisfied by GA4, then a single tool will be a better outcome. However this sort of compromise is not always possible. Using multiple tools can work well, especially in larger organizations that have dedicated analytics teams.

With smaller organizations, GA4 can be problematic. The lack of prebuilt reports pushes more reporting into Looker Studio dashboards. In November 2022, Google implemented GA4 API query limits which make pulling significant amounts of data directly from GA4 to Looker Studio problematic (limits here: <https://developers.google.com/analytics/devguides/reporting/data/v1/quotas>). This move has encouraged many people to use Google BigQuery as an intermediary to avoid the API limits, further increasing the difficulty level for small teams. Small teams are finding it challenging to go from using UA alone to using GA4 plus BigQuery plus Looker Studio.

As we discussed in Chapter 3, Google has a big advantage when it comes to the synergistic benefits of running multiple tools from the Google Marketing Platform together. If you heavily rely upon Google Ads, Google Search Console, Google Tag Manager, Google Ad Manager, etc., then using GA4 alongside those may be the best option. While you shouldn't pick GA4 by default, you also shouldn't choose an alternative just because GA4 isn't Universal Analytics.

CHAPTER 25 - MIXPANEL

Type	Product Analytics
Default Tracking Method	Cookies
Popularity	#10, .5% of top 1M sites
Price	Free plan up to 20M events/mo \$28/mo. for 10K hits \$59/mo. for 100K hits \$216/mo. for 2M hits \$2,170/mo. for 25M hits
Home page	https://mixpanel.com

Pros: Low starting price, high volume free tier, straightforward product analytics

Cons: Add-ons prices as % of base plan, some add-on features require enterprise plan

Mixpanel launched in 2009 and is one of the originators of the product analytics methodology. In 2012, Mixpanel famously railed against more traditional web analytics stats such as pageviews and unique visitors as being "bullshit metrics" (<https://allthingsd.com/20121217/andreessen-and-mixpanel-call-for-an-end-to-bullshit-metrics/>). They advocated instead for a focus on user actions (i.e., events) and engagement. Mixpanel still operates with an emphasis on events rather than pageviews and highlights user retention, flow, and funnels.

I disagree with the idea that pageviews, unique visitors, or even bounce rate are inherently useless vanity metrics. It's how you use the metrics that determines their worth. If you measure bounce rate change over time to determine how well certain segments of traffic engage with what they see upon page landing, that can be highly valuable. If you report bounce rate to your investors without context compared to benchmark sites to show how great your site is, that is indeed bullshit. An analyst declaring a metric to be passé does not affect what the business actually needs to operate. 10 years later, pageviews are still the most vital measure of content success for many sites.

I found Mixpanel to be most similar to Amplitude, but with a reduced feature set. It also has a low starting price, so it could be a good option if you like the style of Amplitude but want something simpler and more affordable. Mixpanel's entry-level price is \$28 per month, far below Amplitude's lowest price of \$240 per month.

Mixpanel's base view of events is similar to Amplitude as well. One positive difference is that Mixpanel includes referrer and UTM parameters by default, whereas Amplitude requires manually enabling those fields. Coming from UA, Mixpanel's reporting of traffic sources still leaves something to be desired, because it treats referrals and UTMs separately. UA, meanwhile, has rules that set traffic source by a number of factors, outlined in this famous processing flow chart (famous in my mind at least):

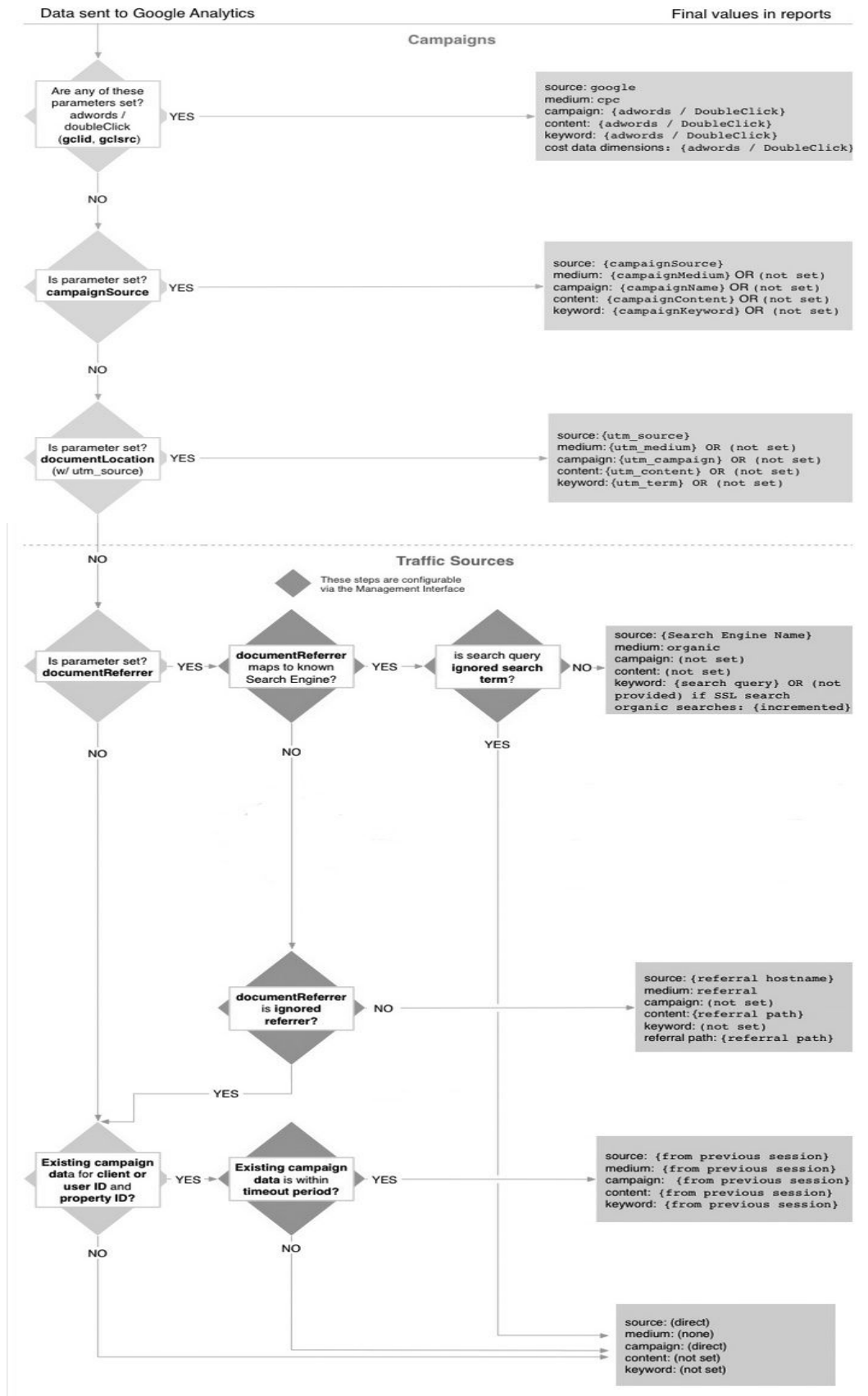


chart from <https://support.google.com/analytics/answer/6205762?hl=en#flowchart>

In Mixpanel (and many of the other product analytics options), we're forced to implement this flow chart manually if we want to recreate the same attribution logic. Additionally, UTM source is a separate field from referrer, and gclid on a URL has no meaning at all. So, if we hit Mixpanel with a URL like "https://socamper.com/?

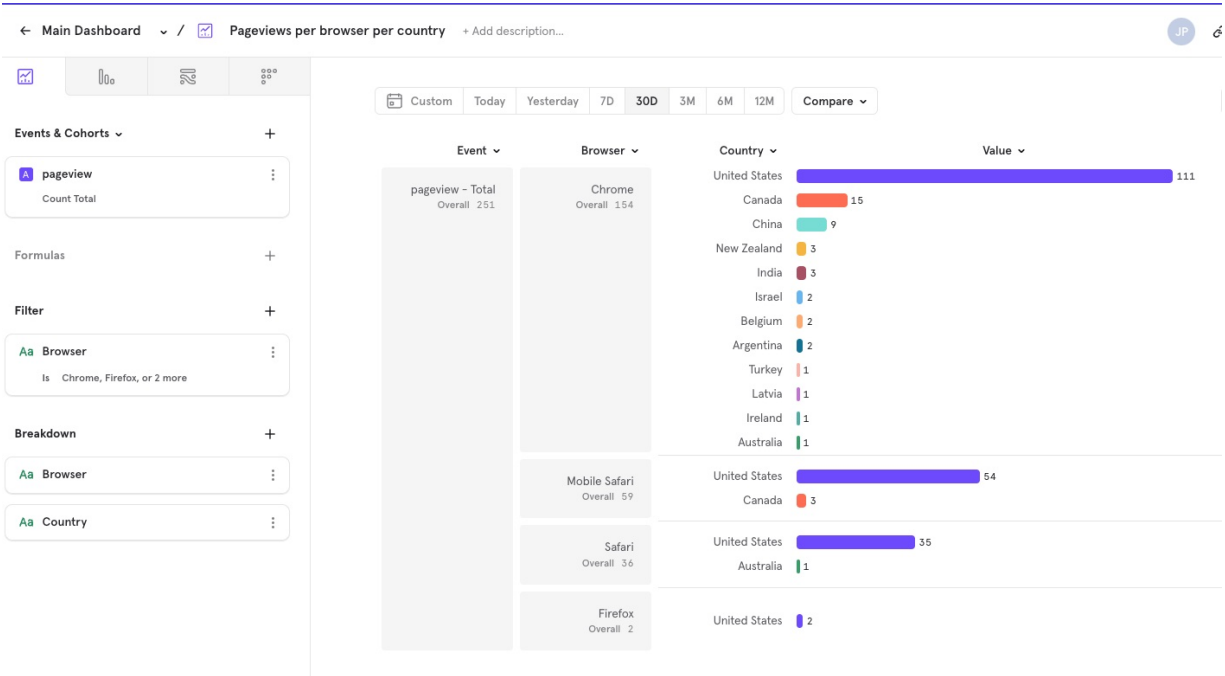
utm_source=foo&utm_medium=bar&gclid=123," it would look like the following:

The screenshot shows the Mixpanel Events interface. At the top, it says "Events" and "Updated 1 minute ago". Below this is a navigation bar with tabs for "Custom", "Today", "Yesterday", "7D", "30D", "3M", "6M", and "12M". To the right of the tabs are links for "Hide Filter", "Edit Columns · 6", "Export", and a search icon. Below the navigation bar is a "SELECT EVENT" section with a dropdown menu showing "A" and "All Events", and a "+ Add" button. Below this is a "FILTERS" section. The main content area shows "Showing 3 most recent results of 3 matches". Below this is a table with columns: "Event Name", "Time", "Distinct ID", "City", "Country", and "OS". The first row shows a "PV" event, "2 minutes ago", a distinct ID, "Columbus", "United States", and "Mac OS X". Below the table is a detailed view of the event properties. It is divided into three tabs: "All Properties", "Your Properties", and "Mixpanel Properties". The "All Properties" tab is selected. It shows various properties: "API Endpoint: api-js.mixpanel.com", "API Timestamp: 1664463352965", "Browser: Chrome", "Browser Version: 105", "City: Columbus", "Country: United States", "Current URL: https://socamper.com/?utm_sou...foo&utm_medium=bar&gclid...", "Device ID: 18389bfffcc05-0f651b4cf8330b-1a525635-232800-18389bfffcd62b", "Distinct ID: 18389bfffcc05-0f651b4cf8330b-1a525635-232800-18389bfffcd62b", "Initial Referrer: \$direct", "Initial Referring Domain: \$direct", "Insert ID: hra9244px8pkjkwe", "Library Version: 2.45.0", "Mixpanel Library: web", "OS: Mac OS X", "Region: Ohio", "Screen Height: 1200", "Screen Width: 1920", "Time: 7:58:47.664 am, Thu, Sep 29, 2022", "Time Processed (UTC): 2:56:19.338 pm, Thu, Sep 29, 2022", "UTM Medium: bar", and "UTM Source: foo".

With that URL, referrer is considered direct, and the UTMs are stored under their own fields. There's can also be a "search engine" field populated by the referrer and stored associated with the user. **There's no single "source" field we can use.**

While attribution isn't one of Mixpanel's strengths, the ability to report on events sliced however you'd like is a major benefit. Like others in this category, you have to build your own reports, but the interface to do so is

pretty intuitive. Here's an example drill down report with multiple dimensions:



Beyond event reporting, Mixpanel also has a number of add-on "apps" such as: A/B testing and auto-correlation detection (requires enterprise plan), statistical significance measurement (any paid plan), a JavaScript-based query language called "JQL" (all plans), and revenue analysis.

The JQL query language is adequate, but I'd prefer to use SQL. Here's an example of what the query builder in the product looks like:

←

+ New

Segmentation

```

1  /* =====
2    * Segmentation
3    *
4    * How many events did we see from different country/referrer
5    * groups?
6  */
7  function main() {
8    return Events({
9      from_date: "2022-08-25",
10     to_date: "2022-09-24"
11   })
12   .groupBy(
13     ["properties.mp_country_code", "properties.$referring_domain"],
14     mixpanel.reducer.count()
15   );
16 }

```

RUN QUERY

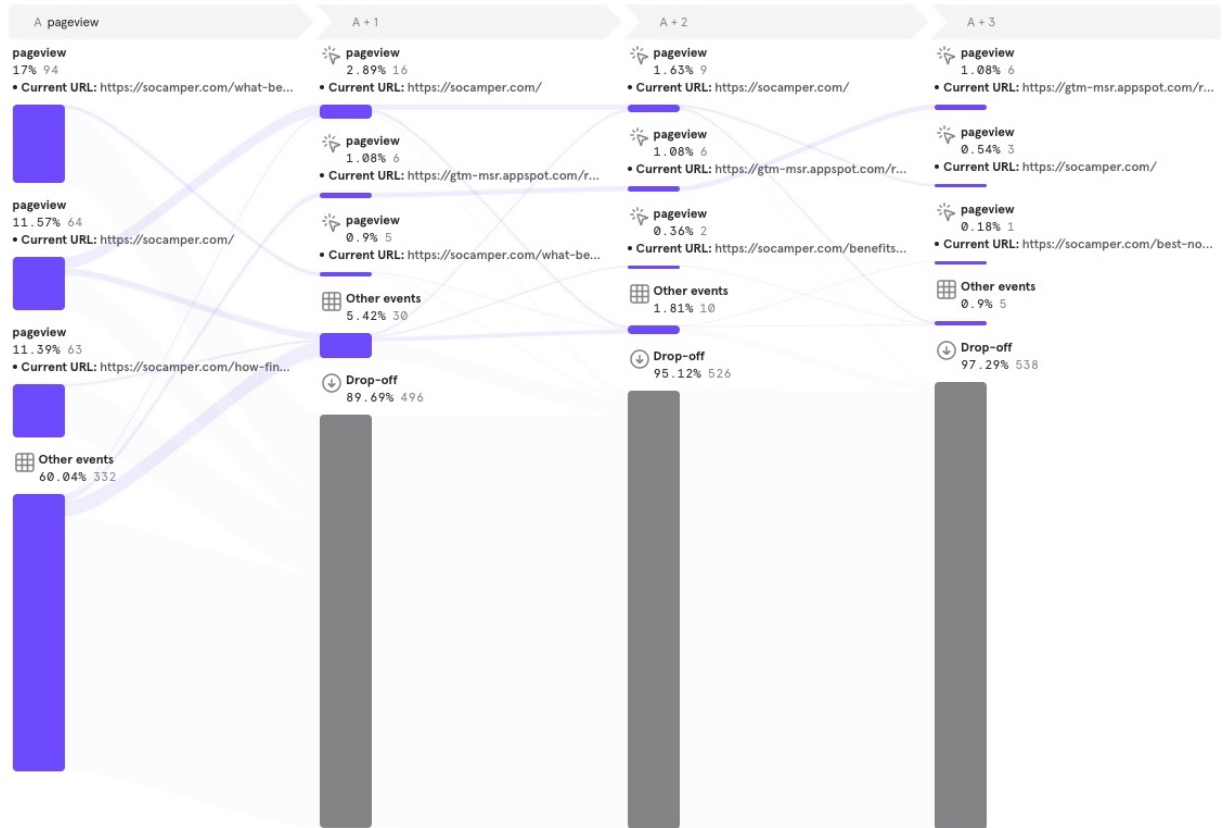
SAVE & RUN QUERY

TABLE JSON

key.0	key.1	value
AR	out.reddit.com	2
AT	null	1

While there are some benefits to doing this in JavaScript, such as allowing you to build more sophisticated functions into queries, the learning curve is much more significant than SQL. By my estimation, many more analysts know SQL than JavaScript. The JQL user interface does have a clear design, and comes with some examples and documentation so you don't start from a blank page needing to learn a completely new query language.

To further illustrate how a good UI can make a confusing report more usable, let's compare the flow reports in Mixpanel with UA. I've never found UA's version of flow reporting particularly valuable — it's difficult to use well, and easy to end up with nonsensical data. Here's what Mixpanel's flow report looks like:



This screenshot may look similar to what you get in UA, but interacting with this report in Mixpanel was much more user-friendly. This makes flow reporting a feature I would use in Mixpanel, vs. one I avoid in UA.

Mixpanel also contains a data governance dictionary called "lexicon," which keeps track of all the events being fired on your site and allows you to add descriptions and aliases. We didn't have a "data governance" chapter in part one since so few of the products we look at have such features, but I like Mixpanel's approach here. This is a great improvement on UA's complete free-for-all, and the way Mixpanel has implemented it could make it useful even if it is only used to document some of the events.

There's no way to enforce naming, but that fits with my personal preference for soft rather than hard naming convention requirements; I prefer an approach where a naming convention is agreed upon and documented but not enforced by the tool itself.

In January 2023, Mixpanel switched from a user-based pricing meter to a hit-based model. Interestingly that same month Amplitude also announced a switch in the opposite direction to a user-based model. Both products also lowered their entry-level price, which is interesting to note during this time of pricing increases. As discussed in Chapter 6 - Pricing, the fit of a particular pricing meter basis depends on your site's usage pattern. Sites with highly engaged users that log in will prefer Amplitude's user basis, whereas sites with less engaged and more anonymous users will prefer Mixpanel's. With this pricing change, Mixpanel upped their free tier limit to 20M events per month, the highest limit of any tool we are evaluating. However, I would expect that most organizations will want paid-only features such as automated data exports or cohorts well before they arrive at the volume limit.

According to our BuiltWith stats, Mixpanel was previously the most popular product analytics tool in our list of Google alternatives, but was passed in 2022 by Snowplow and Amplitude. Despite what appears to be declining popularity, there are still cases where Mixpanel could be an effective choice. I found Mixpanel's best fit to be for smaller sites that want product analytics but aren't ready for some of the more sophisticated options.

CHAPTER 26 - SNOWPLOW

Type	Product Analytics
Default Tracking Method	Cookies
Popularity	#6, 0.9% of top 1M sites
Price	?
Home page	https://snowplow.io

Pros: Open source, built for scale, very flexible

Cons: Learning curve, no built-in reporting, complicated install for self-hosting, opaque pricing

Are you the type of analyst who goes straight into Google BigQuery to search for answers? Do you get frustrated with cumbersome UIs for building reports and think, "This would be a whole lot easier if I could write some SQL to get an answer"? If so, then Snowplow may be the tool for you.

What does Snowplow's reporting UI look like? There isn't one, so it looks like whatever BI or viz tool you hook up to it. Some BI tools, like Looker, have more built-in support for the Snowplow data models than others; however we will only analyze the Snowplow tool itself in this guide. Snowplow Behavioral Data Platform (BDP) is their commercial cloud-hosted version, which is what I evaluated, not the open source version.

The Snowplow BDP trial data ends up in Postgres, so the reporting interface I worked with was a SQL prompt (don't expect a lot of pretty screenshots in this chapter):

```
snowplow=> select * from atomic.events where page_url like '%romfind%' and page_title != 'Page Not Found' and geo_country != 'DE' limit 10;
-[ RECORD 1 ]-----+-----
app_id          | try-snowplow
platform        | web
etl_tstamp      | 2022-09-22 22:45:29.954
collector_tstamp | 2022-09-22 22:45:29.945
dvce_created_tstamp | 2022-09-22 22:45:29.152
event           | page_view
event_id        | 1155c19d-8d69-4335-b726-f3e639db5e7a
txn_id          |
name_tracker    | try
v_tracker       | js-3.3.1
v_collector     | --stdout$
v_etl          | try-snowplow-pipeline-0.0.0-31074638c1f02a-common-3.1.4
user_id         |
user_ipaddress  |
user_fingerprint |
domain_userid   | b908f09f6d4714d82005f942b11074a183a441b9
domain_sessionid | 1
network_userid  | 2b2dc6422b0139a27b4cf4f6f339c224e8446fde
geo_country     | AR
geo_region      | C
geo_city        | Buenos Aires
geo_zipcode     | 1871
geo_latitude    |
geo_longitude   |
geo_region_name | Buenos Aires F.D.
ip_isp          |
ip_organization |
ip_domain       |
ip_netspeed     |
page_url        | https://romfind.com/roms/gbc/p.html
page_title      | GBC Roms: P
page_referrer    |
page_urscheme   | https
page_urlhost     | romfind.com
page_urlport     | 443
page_urlpath     | /roms/gbc/p.html
page_urlquery    |
page_urlfragment |
refr_urscheme    |
refr_urlhost     |
refr_urlport     |
refr_urlpath     |
refr_urlquery    |
```

If you already bring your data into a centralized data warehouse, what really is the point of a measurement platform reporting interface anyways? Embrace your inner data engineer and work with the raw data and its pipelines.

Snowplow's approach stands in stark contrast to UA; in fact, I'd say it's the furthest from UA of any of our alternatives. Snowplow is a general data collection and processing tool with a focus on user behavior data. Their product descriptor, "behavioral data platform", provides an excellent explanation of its function. Rather than focusing on product or marketing analytics specifically, it's more agnostic about the data it collects. As such, it has the longest, most involved setup process (even if you use their cloud-hosted version instead of self-hosting). It also can handle the greatest

number of use cases of any tool. So, Snowplow can undoubtedly replace UA, but not without effort.

What's the difference between a behavioral data platform vs. a web analytics platform like UA? Don't all of our alternatives record user behavior as data? Snowplow offers the flexibility to store behavioral data from multiple sources using whatever schemas you want. You could use Snowplow to track data from your website, your physical storefront, your IoT device, your mobile app, etc., with different schemas defined for each source, but ultimately the data flows into one place.

All of our product analytics tools provide the ability to report data from various platforms, but I believe Snowplow is the only one designed to **allow different custom data schemas per source, with a centrally managed schema repository.**

If your only source of data is a single website with modest levels of traffic, Snowplow would probably be overkill. Also, Snowplow was designed with developers as the primary target, not analysts. This is evident in their open source version, which is marketed as a "developer-first engine."

The tool's deep customizability and associated significant learning curve make Snowplow hard to evaluate. Having never set up and run a production Snowplow installation, my ability to give detailed feedback on the product is limited.

Snowplow is also oriented toward self-hosting more than any of our other alternatives. When Snowplow was first released in 2012, it could *only* be run self-hosted, and it wasn't until three years later that the hosted Snowplow Behavioral Data Platform (first called Snowplow Insights) launched.

If you do opt for the self-hosted version, it's important to know it has a relatively complex setup process and is designed for AWS or GCP. Compare this to Matomo's open source install, where you can set it up

quickly on just about any cloud environment and utilize component tools that may already be installed.

Unfortunately, I'm unable to provide pricing info on Snowplow BDP. As Snowplow is an enterprise-style data platform, you should expect enterprise-level pricing.

If you want to use Snowplow with a smaller level of traffic your best choice financially would either be to self-host or find a third-party managed host. As Snowplow is open source those are both potential options. For example: Snowcat Cloud, a small company separate from Snowplow, offers managed Snowplow starting at \$750 per month for 15 million hits (<https://www.snowcatcloud.com/#pricing>), and also has a developer plan for 50,000 hits. A third-party managed host such as Snowcat is likely to be somewhat less expensive than Snowplow BDP itself.

Snowplow's architecture is composed of a variety of enterprise-level building blocks. For example, with an AWS install you might use some or all of these tools: Terraform, Redshift, RDS Postgres, S3, or Kinesis. Snowplow is a data platform that can scale up to fit nearly any use case, so the underlying tech stack it requires is more complex. With great power comes great dependencies, one might say.

Since I've pointed out questionable marketing claims for other tools where I've seen them, I won't spare Snowplow. Its open source home page says,

"Snowplow is the leading open source Data Creation technology in the world today, and the third most adopted web tracker, behind Google Analytics and Facebook."

I'm going to have to call "citation needed" on that claim of being #3. Their site reports 1.7 million websites, which is consistent with the BuiltWith numbers (in fact, BuiltWith could be the source of their data since Snowplow themselves may not know how many self-hosted installs exist), but BuiltWith places them closer to #7 overall, after Google, Baidu, Meta, FullStory, Yandex, and NewRelic. Our ranking list is specific to web

analytics tools within the top one million sites, and it's #8 in popularity on that list.

When looking at Snowplow's install base, it's an unusual mix of higher market penetration among the top 10,000 sites (claiming 5% of the top 10,000 vs. only 1% in the top one million) and also a long tail outside the top one million.

If that seems strange, consider that many of those smaller sites obtain their analytics via a third-party CMS. Weebly, a popular hosted CMS, includes Snowplow-powered web analytics built into their system. Of the 1.7 million sites BuiltWith detects running Snowplow, 50% also run Weebly. This explains much of Snowplow's popularity among smaller sites that would not seem to have the technological wherewithal to install and manage the tool.

Snowplow is unlikely to be a good fit for most small sites, but it could be appropriate for small- to medium-traffic sites that have high levels of technical sophistication. Cloud hosted Snowplow BDP would probably still be on the expensive side for those sites. There is a 14-day free trial, which is what I used to evaluate the tool. Two weeks is not very long to learn the tool, and Snowplow was unable to extend my trial, so I had to create multiple trials to finish my evaluation — so be ready to start testing before you sign up for the trial.

For those sites that are a good fit for Snowplow, it could be a great cornerstone to build a data practice on. It's about as far from a drop-in replacement to UA as you can get, but for sites with a DIY approach to their analytics stack, it may be just the thing.

CHAPTER 27 - AMPLITUDE ANALYTICS

☆ FAVORITE

Type	Product Analytics
Default Tracking Method	Cookies
Popularity	#8, 0.6% of top 1M sites
Price	Free plan for up to 100K users/mo. Basic plan ("plus") starts at \$240/mo. for 25K users \$480/mo for 50K \$960/mo for 100K Standard plan ("growth") reportedly starts around \$900/mo.
Home page	https://amplitude.com

Pros: Sophisticated reporting tools, collaboration, scales well

Cons: Significant implementation work on instrumentation and reporting sides, free plan has a high hit limit but excludes some very useful tools

The product analytics category is by far the most challenging to review, because all of the tools in it are so flexible. Unlike the traditional category (where most of the reporting is pre-defined), or the simplified category

Amplitude is a good example of this: when you set it up, it's a blank canvas on which to develop your own reporting. The base tag will record pageviews, but no other events without explicit instrumentation. Unlike UA with more than 100 standard reports, Amplitude has no pre-loaded reports other than the User Lookup — which is similar (but superior) to UA's User Explorer, allowing for easy lookup of individual users by attribute.

User Look-Up

Search for your users to dive into their specific properties and event path.

Quantable

Quantable

463346292507

User Property values shown are the most recent values. [Learn More](#)

	User ID -	Amplitude ID 463346292507	First Seen Sep 14, 2022 (a day ago)	Last Seen Sep 14, 2022 (a day ago)
	Total Events 1	Total Sessions 1	Usage Time a few seconds	
	Language English	Country United States	Region Georgia	City Atlanta
	Device Type Apple iPhone	OS Mobile Safari 15	Version -	Platform Web
	Device ID(s) -pCaEE98Xz9ooCH9W3Mvao			

Event Stream

All Event(s) Select Event(s) from select Device ID(s)

Showing most recent 1 events.

September 14, 2022

a few seconds

Device ID
Family, OS

-pCaEE98Xz9ooCH9W3Mvao
Apple iPhone, Mobile Safari 15

9:39:42 pm • pageview

September 14, 2022
9:39:42.517 pm UTC

pageview

Event ID	1
Session ID	1663191582515

User Properties

User ID

-

Device ID

-pCaEE98Xz9ooCH9W3Mvao

Platform

Web

Device Type

Apple iPhone

Version

-

Library

amplitude-js-gtm/2.0.2

Country

United States

*** See all other 11 User properties

Search Users

Q device="Apple iPhone"

Tip: you can search any User Property us

USERS BASED ON YOUR SEARCH

463395643642
United States Web

463346292507
United States Web

462694353634
United States Web

461984151597
United States Web

461986270363
United States Web

461780228076
United States Web

462197738019
United States Web

461511191347
United States Web

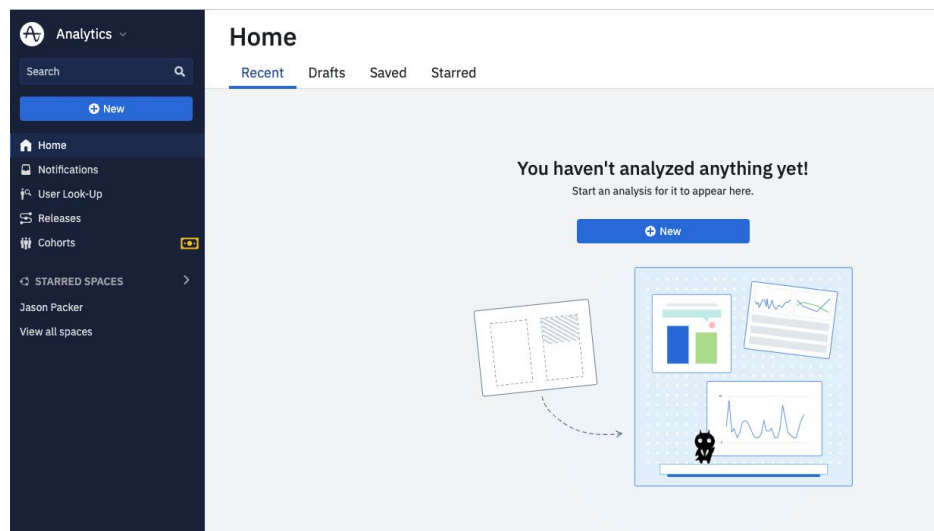
461240110642
United States Web

461565581535
United States Web

The tool provides numerous templates you can use as a starting point, but nothing pre-loaded. It's as if there were no built-in reports in UA and all that existed was Data Studio. **This minimalist starting point is by design** and solidly in the paradigm of product analytics in conjunction with explicit tagging.

Amplitude is a powerful tool, but out of the box, is very different from UA and requires a new way of thinking, especially if you're unfamiliar with GA4 or other product analytics tools. The switch from UA to Amplitude involves significant effort to learn the platform and methodology — more than moving from UA to GA4.

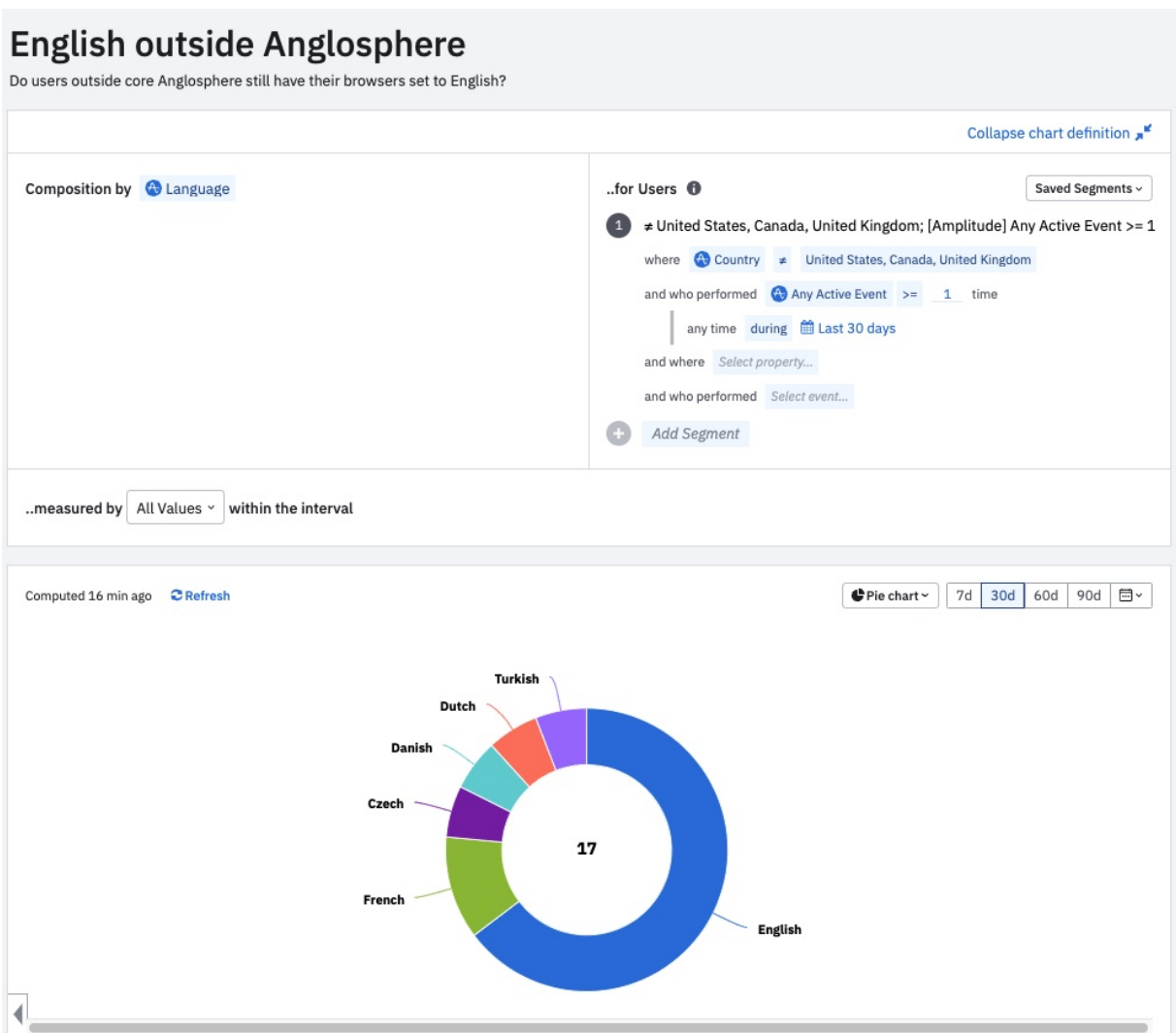
This is what you start with:



Your Amplitude installation could have recorded millions of event hits, but if nobody on your team sets up reports, this home page would remain empty. Starting from that blank page can be hard, and I've seen multiple product analytics installations of this type fail because stakeholders had unrealistic expectations about the level of work required.

Amplitude focuses on answering questions and sharing those questions and answers with a team. Collaboration is built into the platform on a fundamental level, which I like, especially if you work within a team of analysts.

Let's run through a simple example. My test site is an English-only content site, but it draws traffic from many countries. Do the users in countries where English isn't the predominant language still have their browsers set to English when they view my site? If not, then perhaps I should create content in their localized language? Amplitude excels at this kind of hypothesis testing. Here's what that report would look like:



So, my site visitors outside the core English-speaking countries that generated traffic still largely have their browsers set to English, meaning the answer to my question is that I don't have to create any new content, phew.

We could answer the same question with UA, but it'd be harder to build the report and much more difficult to collaborate on it.

Now let's say we want to segment these users to Google Organic only. Annoyingly, we can't do this by default, and there's more setup required on both the client and reporting sides. On the client side in GTM, we need to add "includeReferrer: true" as an option to the JavaScript sensor. On the reporting side in Amplitude Data, we have to add "referrer" as a user property.

Things like **UTMs and referrer are not tracked by default** in Amplitude. This may be a shock to users coming from more of the marketing analytics world, where source/medium is one of the most important things to track and report on. It's not hard to set up, but it is not part of the default setup here. Recently Amplitude has added a "Channels" feature that can be used in a similar way to GA custom channel groupings, but that feature is only available to Growth and Enterprise plans, not the Plus plan that I have listed prices for. This emphasizes how Amplitude focuses on what users do once they're on your site instead of how they arrived there.

If this seems like more of a complement to GA than a competitor, Amplitude might agree with you. As stated on a blog article from their site (<https://amplitude.com/blog/amplitude-vs-google-analytics>).

You can absolutely use both Google Analytics and Amplitude. Some marketing teams find GA really helpful when it comes to tracking ads and channels. This is possible even with the free tier of Google Analytics. If this is the case for your team, keep the free tier of GA and head on over to Amplitude for everything past your marketing site.

I've worked with clients who run both UA and Amplitude together, and that can be a robust solution. With GA4, this combination loses some of its strength because there's more overlap in the products. We discussed this same dual tool usage some in the GA4 chapter. This being a guide to "GA

alternatives" though, can you replace UA with Amplitude? Yes, though it depends on what you do with UA and how much customization you're willing to build.

Amplitude has a generous free tier similar to Universal Analytics, giving you up to 100K monthly tracked users (MTUs) for free. Unlike UA, there are significant features and integrations left out of Amplitude's free plan. Some of Amplitude's paid-only features include: behavioral cohorts, anomaly detection and forecasting (one of the strengths of the product), emailed reports, dashboard templates, and automatic syncing to third-party data destinations. While users receive plenty of useful functionality for free, much is held back for paying customers. Additionally, since you start without pre-built instrumentation or reporting, having paid customer support is nice to have.

Paid plans for Amplitude start at a lower price than you might expect from a top-tier product (if you are used to Adobe and GA360). The "Plus" level isn't on the general pricing page but becomes available once you have a free account. This starts at \$240 per month for 25 thousand MTUs and has some of the paid features we listed above (like cohorts), but not all. The standard plan above that level, called "Growth," reportedly starts at around \$900 a month. As we covered in the pricing chapter, it would be nice if Amplitude had more transparent pricing. The generous free tier in conjunction with the limited pricing information is a bad combination for users, as it can encourage sites to put significant effort into a platform that may ultimately prove to be unaffordable.

As mentioned in the Mixpanel chapter, Amplitude switched in January 2023 from a hit-based pricing meter to counting monthly tracked users. This makes Amplitude unique as the only product we are evaluating priced on user counts. If you are using Amplitude for product analytics on a site with significant engagement per user this could be a great change for you. It all depends on what your monthly events per active user is. PostHog has released some data on this and found that most of their customers were in the range of 50-100 tracked events per user per month (<https://posthog.com/blog/calculating-events-from-users>). Amplitude seems to have made a similar calculation, changing their free tier from 10M hits to

100K users (100 event hits per user) and their ~\$500/mo. "Plus" plan from 5M hits to 50K users (also 100 hits per user). This means if you have more than 100 event hits per MTU this was a favorable change to you, and if you have fewer it was unfavorable. In either case, there is no option to pick and choose your pricing model, so if the new model is unfavorable to you then you may want to look to other products.

Amplitude is an excellent product analytics tool that is improving rapidly. I first evaluated it for a client over two years ago, and it's greatly expanded its capabilities since then. It has even added a number of features since the first version of this book was completed in October 2022. Implementing and learning the product involves significant effort, but the benefits are worthwhile for sites that want product analytics. For sites looking for marketing analytics and an out-of-the-box implementation, Amplitude is a poor fit.

CHAPTER 28 - HEAP

Type	Product Analytics
Default Tracking Method	Cookies
Popularity	#12, 0.3% of top 1M sites
Price	Free (limited features) to 10k/sessions \$375/mo. for 17k sessions
Home page	https://heap.io

Pros: ML-based insights with "Illuminate" tool, pricing plan amenable to auto-tracking, governance tools

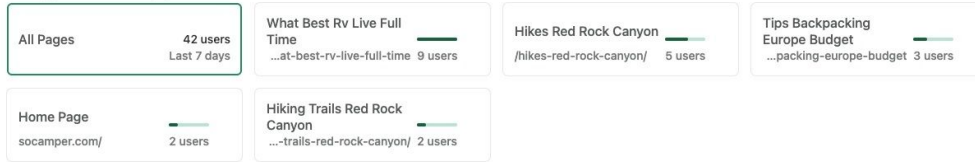
Cons: Feature limits on basic plan

One of Heap's defining characteristics is its automatic event tracking. Heap launched in 2013 and was one of the originators of the "implicit" tracking concept that auto-captures user actions on a site (<https://stackshare.io/heap/how-heap-built-an-analytics-platform-that-auto-tracks-every-user-event>). Heap doesn't automatically capture *all* interactions (for example, it does not auto-capture scroll depth), and if you are working with a single-page app then you will need to do some coding — but it does track most things without any setup work.

Explore auto-captured events

Did you know? Heap automatically captures **everything**. Find, analyze, and save the events you care about most.

Filter by your most visited pages

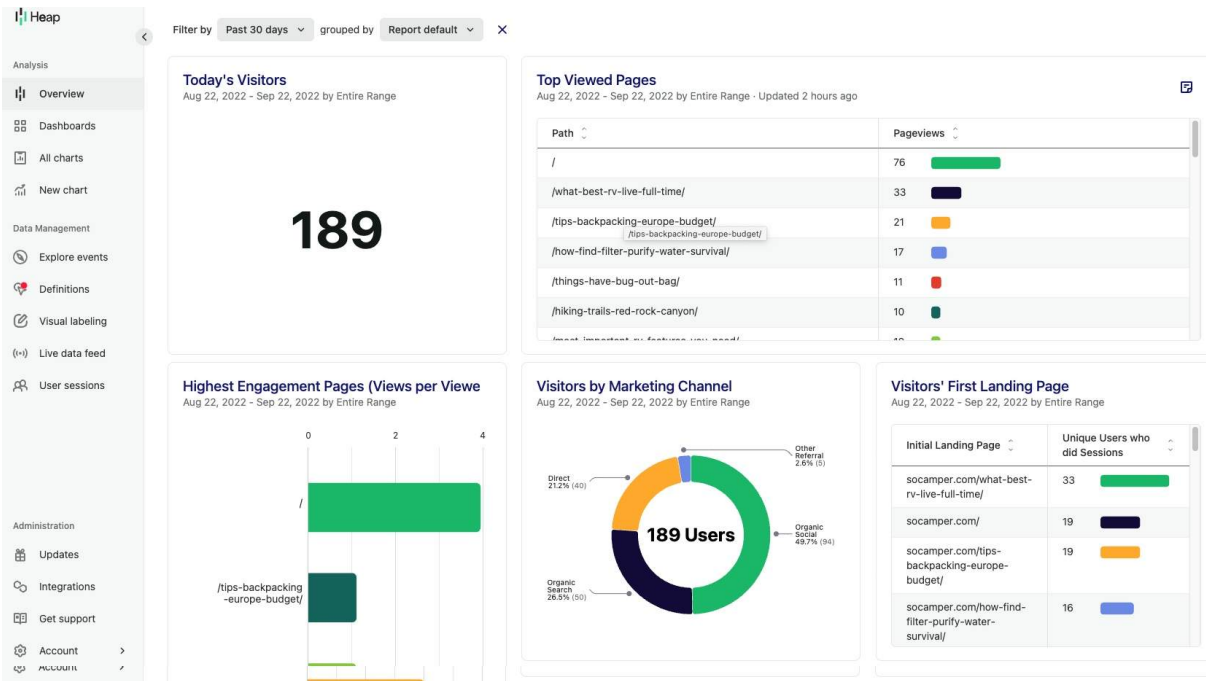


Search events by button text, form label, etc.

User action	# of users (7 days)	Generate chart	Define event
Click - Right Sidebar	2 (5% of all users)	  	Defined 
Click - Posthog Toolbar (Home Page)	1 (2%)	  	Define >
Click - Link to /category/camping-gear/ (Home Page)	1 (2%)	  	Define >
Click - Ppms Cm Close Popup Id (Benefits Walking Barefoot Earth Page)	1 (2%)	  	Define >
Click - Right Sidebar (Hikes Red Rock Canyon Page)	1 (2%)	  	Defined 
Click - Right Sidebar (Tips Backpacking Europe Budget Page)	1 (2%)	  	Define >
Click - Content (What Best Rv Live Full Time Page)	1 (2%)	  	Define >
Click - Get Started (Home Page)	1 (2%)	  	Define >
Click - Class C Motorhomes (What Best Rv Live Full Time Page)	1 (2%)	  	Define >
Click - Some disadvantages to Class A's: (What Best Rv Live Full Time Page)	1 (2%)	  	Define >
Click - Site (Home Page)	1 (2%)	  	Define >
Click - Share on Facebook (Hiking Trails Red Rock Canyon Page)	1 (2%)	  	Define >

Heap tracks all clicks, but the ones that are then highlighted as "defined events" receive first-class treatment in the reporting interface. This definition can take place after the fact or through a visual overlay tool. This approach lets you enjoy the best of both worlds with event-tracking. The auto-event capture collects everything, while you can favorite the events you're most interested in reporting.

Heap has a friendly user-interface and some good pre-built dashboards. I found it to be one of the more intuitive products in the product analytics category.



Heap also has a machine learning-based feature called "Illuminate" that claims to give automated actionable insights based on this data. I wasn't able to test this feature, since it's not part of the free trial and had no pricing listed. While I listed this as an upside for the product, I'm wary of the usefulness of automated insights. The description of this feature is similar to GA's automated insights, which I have never found useful and typically return insights like this actual example from GA4:

Users for the campaign medium "referral" dropped We forecasted that users for the campaign medium "referral" would be between 11 and 48. Your actual users of 10 were less than this range.

I suspect Heap's auto-tracking allows for a better dataset to work with, but this kind of automated anomaly detection is at best a reminder for a human to take a deeper look.

Another of Heap's strengths is its data governance capabilities. It has a data dictionary area where all defined events and other assets live:

Definitions		Search events	Filters
Get started			
Events	Events (8) ↑		Status Last modified
Segments	Personal (3)		
Properties	Click - 16. Go on picnics. (Tips Backpacking Europe Budget Page) Click element with text "16. Go on picnics." on page "socamper.com/tips-backpacking-europe-budget"	Unpublis...	Jul 14, 2022 by you
Recent	Click - Content Click element with id "content" on any page	Unpublis...	Aug 20, 2022 by you
Suggestions	Click - Site (Subscribe Page) Click element with aria "Site" on page "socamper.com/subscribe/"	Unpublis...	Aug 20, 2022 by you
Pending verification			
Inactive	Generated from Explore Events (2)		
Unused	Click - Right Sidebar Click element with id "right-sidebar" on any page	Unverified	Sep 21, 2022 by you
	Click - Right Sidebar (Hikes Red Rock Canyon Page) Click element with id "right-sidebar" on page "socamper.com/hikes-red-rock-canyon/"	Unverified	Sep 22, 2022 by you
Archived	Top Viewed Pages (3)		
	View - Home Page View page socamper.com	Unverified	Apr 17, 2022
	View - Tips Backpacking Europe Budget View page socamper.com/tips-backpacking-europe-budget	Unverified	Apr 9, 2022
	View - What Best Rv Live Full Time View page socamper.com/what-best-rv-live-full-time	Unverified	Apr 14, 2022

It can also enforce a naming convention on assets. I wouldn't personally use that option, but it may be helpful for certain installations. The tool offers other more advanced data governance options like event repair, which identifies and allows you to redefine events, that are only available in higher-level plans.

I'd like these kinds of helpful features to be available at all paid levels. Products like Heap that have done a lot of picking and choosing which features are available at different pricing levels can be frustrating to deal with. It's inevitable that there will be a certain feature that you want from a higher tier, but your volume and overall usage pattern doesn't otherwise justify that tier. I prefer pricing models where the main factors determining tiers are volume and support level, rather than picking and choosing individual features that the vendors think are more "enterprise."

Heap is the only tool in this guide with a sessions-based pricing model. As discussed in the Pricing chapter, your best fit model depends on your site's usage pattern. With auto-tracking like Heap provides, there's solid logic in having the pricing model structured around sessions rather than events — otherwise, your bill would go up as you improve your user

engagement. PostHog bases billing on total event volume, so your usage patterns will determine which product is better for you financially.

Neither Heap nor PostHog publish product comparisons with the other, even though they're very similar.

Heap compares itself to:

GA, Mixpanel, Pendo, Amplitude, FullStory, Contentsquare, and Quantum Metric (<https://heap.io/compare>).

PostHog compares itself to:

GA, Mixpanel, Matomo, and Amplitude (<https://posthog.com/blog/categories/comparisons>).

This is also true of Fathom and Plausible: similar tools that have feature comparisons with other products but not each other. It can be hard to compare products that are so similar, but skipping the obvious comparison is a disservice to potential buyers (and the vendors almost certainly have detailed comparisons available internally).

These comparisons are also an interesting example of tool functionality convergence within the field. In Heap's case, it chose to compare itself to FullStory, which has its roots in session recording. Heap has only recently added session recording, and FullStory has added more product analytics functionality.

If you're looking for a product analytics tool that offers implicit tracking, Heap is an excellent option. Although I gave PostHog my pick for favorite rather than Heap because of its larger breadth of features and self-hosted option, either could be a good option.

CHAPTER 29 - POSTHOG

☆ FAVORITE

Type	Product Analytics
Default Tracking Method	Cookies
Popularity	#16, 0.1% of top 1M sites
Price	Free up to 1M hits/mo. \$1,100/mo. for 5M hits/mo. \$3,400/mo. for 25M hits/mo.
Home page	https://posthog.com

Pros: Open source, rich feature set, transparent pricing

Cons: Newer product, volume-based pricing can become expensive at scale (especially for high-engagement sites)

The first time I saw the "PostHog" name was as the creators of the isgoogleanalyticsillegal.com website. The site failed to impress me and came across as a marketing gimmick that sowed more fear, uncertainty, and doubt into an already confusing situation, so I initially ignored PostHog.

As we discussed in chapter 4, use of Google Analytics in the EU is legally fraught. If you do business in the EU, do you need to stop using GA? I've read a lot on the subject and find it challenging to give an answer beyond "maybe?" Yet, PostHog's fear-inducing marketing site is #1 in Google for the search "is google analytics illegal." While the text it uses is

reasonable — "If your country is bound by the GDPR, continuing to use Google Analytics may now be illegal." — the scrolling red banner and giant red "ILLEGAL" sends a starkly different message.

The screenshot shows a dark-themed website banner. On the left, a vertical red bar contains the text "IS GOOGLE ANALYTICS ILLEGAL?". The main heading reads "Is Google Analytics **ILLEGAL** in your country?". Below this, text explains that due to recent GDPR rulings, Google Analytics users are under fire for transmitting personal data outside the EU. It advises that if a country is bound by the GDPR, continuing to use Google Analytics may now be illegal. The safest solution is to use an analytics provider that keeps data on your own infrastructure, with a link to learn about data privacy settings. It also mentions that more rulings are expected throughout 2022 and provides a link to submit a PR. A button labeled "Browse Google Analytics alternatives" is at the bottom left. On the right, a map of Europe shows several countries highlighted in red, including France, Spain, Italy, Austria, and the Netherlands. A legend at the bottom right of the map says "Select a country in red to see rulings concerning Google Analytics legality".

The best synopsis I've seen on GA's legality comes from Iubenda (<https://www.iubenda.com/en/help/72017-is-google-analytics-illegal-in-the-eu-and-uk>), but their headline:

"Do you need to stop using Google Analytics altogether? There's no direct answer yet as this is a developing case."

makes for poor marketing copy. The legality of GA and actions required is a complicated question. I view PostHog's site as SEO bait rather than constructively adding to the conversation. Ironically, PostHog had the same issues as UA until very recently. PostHog launched an EU hosting option on Oct 11th, 2022 — but previous to that the only cloud housing option was Virginia, USA at AWS. Meaning from when the site was created in January through October they had the same legal compliance problem (data transfers to the U.S.) as UA.

Clearly PostHog didn't leave the best first impression on me. However, in the process of researching popular products for this guide, I came across

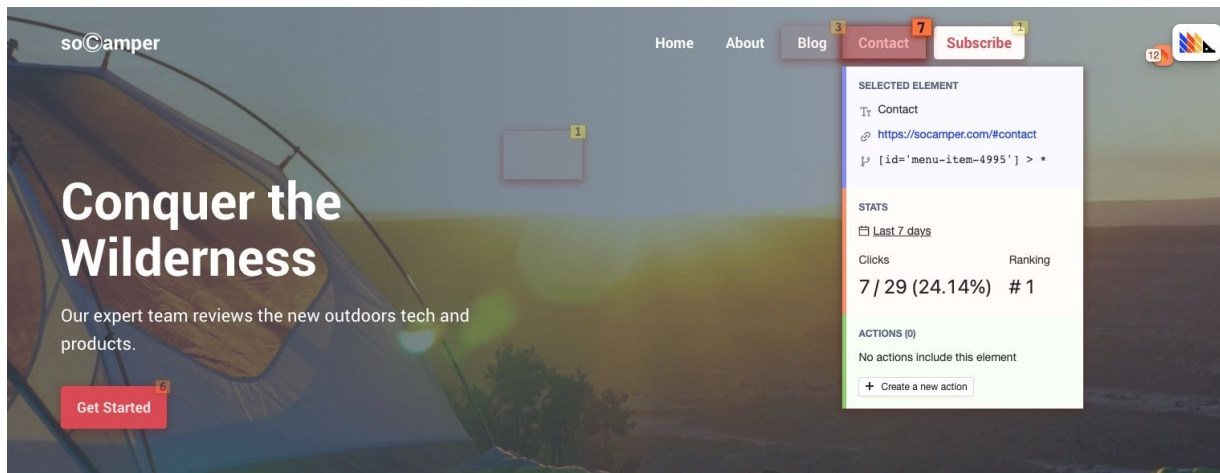
it again and, after looking into the product, was very impressed.

PostHog is the fastest-growing product of any of the alternatives we evaluate, registering 141% year-over-year growth. It's new and thus starting from a lower position, but those numbers are nonetheless very impressive.

In addition to product analytics, PostHog also provides session recording, heatmaps, A/B testing, and data pipelining. PostHog bills itself as an "open source product OS." I don't know what that means, but it definitely does have a rich feature set beyond just product analytics.

PostHog has a transparent pricing plan and a strong free-tier offering for up to one million hits a month without limiting critical features to paid only. This makes it a great option for many small to medium sites. Even if all you want out of the product is session recordings, PostHog could be worth it. For example, session recording tool HotJar is \$80 a month for 500 daily sessions — while PostHog can give you a near-identical base feature set for free (up to 15,000 recordings per month).

There is also a heat-mapping tool, which is less sophisticated than what you'd find in a tool like HotJar, but can still be useful. It works similarly to the heat-mapping Chrome extension for UA (called Page Analytics, no longer updated by Google). This feature is available via the PostHog toolbar, a site overlay that includes click stats and the ability to convert those clicks to tracked events and does not require a browser extension.



This is similar to Heap's "visual labeling" tool, although PostHog treats it as a multi-function toolbar (in Heap, it's focused on event labeling). Overall Heap and PostHog are very similar in their core analytics functionality.

Some of PostHog's paid-only features include: A/B testing, group analytics, and funnel correlation analysis. Overall, PostHog gives away a considerable amount of functionality up to high levels of traffic, which is likely a primary driver of their impressive growth curve.

PostHog also boasts a wide array of data connectors and apps.

Like all of the options in our product analytics category, PostHog takes a "build what you need" approach to reporting as opposed to UA's "pre-built report for every occasion" philosophy. PostHog has a few predefined reports (one dashboard, six "insights" reports), but treats these as sample templates to modify rather than something designed for actual use. Here's the default dashboard:

My App Dashboard

Description (optional)

+ Add tag

...

J

Share

Add insight

UTC

Custom

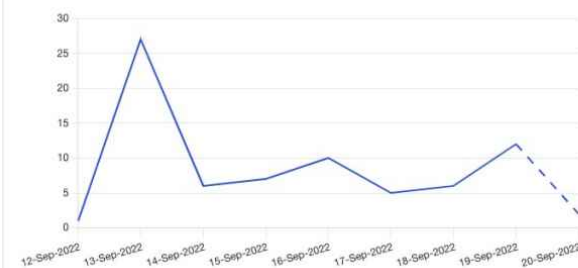
+ Add filter

Last updated 5 minutes ago

TRENDS - LAST 7 DAYS

Daily active users (DAUs)

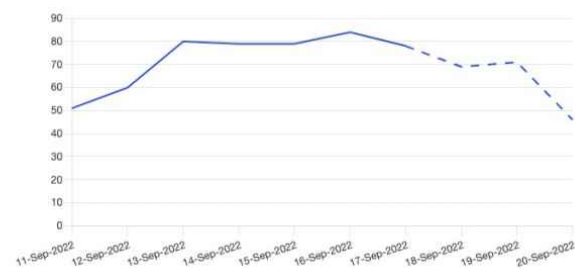
Shows the number of unique users that use your app every day.
Last modified 2 months ago



TRENDS - LAST 7 DAYS

Weekly active users (WAUs)

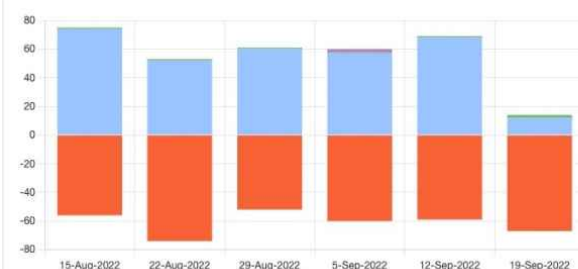
Shows the number of unique users that use your app every week.
Last modified 2 months ago



LIFECYCLE - LAST 30 DAYS

Growth accounting

How many of your users are new, returning, resurrecting, or dormant each week.
Last modified 2 months ago



RETENTION - LAST 7 DAYS

Retention

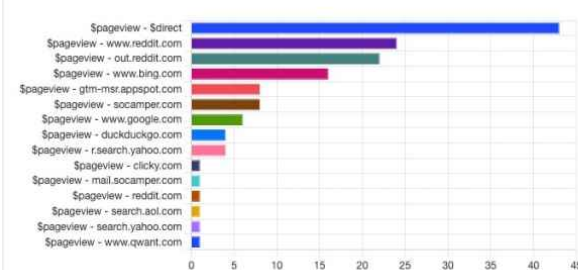
Weekly retention of your users.
Last modified 2 months ago

Cohort	Size	Week 0	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8
Jul 10	0	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Jul 17	30	100.0%	10.0%	6.7%	3.3%	3.3%	3.3%	3.3%	3.3%	3.3%
Jul 24	68	100.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Jul 31	43	100.0%	2.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Aug 7	47	100.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Aug 14	80	100.0%	0.0%	0.0%	2.5%	0.0%	0.0%	0.0%	0.0%	0.0%

TRENDS - LAST 14 DAYS

Referring domain (last 14 days)

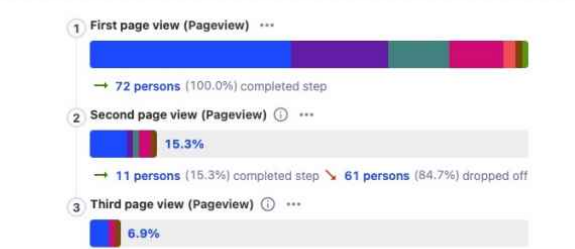
Shows the most common referring domains for your users over the past 14 days.
Last modified 2 months ago



FUNNEL - LAST 7 DAYS

Pageview funnel, by browser

This example funnel shows how many of your users have completed 3 page views, broken down by browser.
Last modified 2 months ago



It's nice to see some reasonable default reports, but this alone isn't enough, so you'll need to build more reporting. I prefer this approach where there's a starting point vs. others like Amplitude where you begin with a blank page.

You won't have to do much event instrumentation since PostHog takes an implicit approach to tracking, auto-capturing all clicks and form submissions and making them available for reporting. PostHog also offers what it calls "actions," which are events you've selected to highlight in reporting. This is very similar to the approach Heap pioneered, except the company's nomenclature of "actions" and "events" is reversed.

Think of actions as a list of favorite events pulled from the full list of auto-captured events.

Live events

Event history limited to the last twelve months.

Autocapture

+ Add filter

Automatically load new events

Configure columns

Export

EVENT	PERSON	URL / SCREEN	SOURCE	TIME	
↕ clicked svg	1 183589b71ac1e0-0f535...6500-183589b71ad727	socamper.com/benefits-walking-barefoot-earth/	web	a day ago	...
September 17, 2022					
↕ clicked link with text "Get Started"	1 1834a0afe2019f-03a36...8f40-1834a0afe21161	socamper.com/	web	4 days ago	...
September 13, 2022					
↕ clicked link with text "Learn More"	1 18339f1760b4d4-0235d...c680-18339f1760ca0f	socamper.com/	web	7 days ago	...
↕ clicked link with text "Learn More"	1 18339f1760b4d4-0235d...c680-18339f1760ca0f	socamper.com/	web	7 days ago	...
↕ clicked image	1 18336a2f21c1ed-0c5d5...800-18336a2f21d3afd	socamper.com/#contact	web	8 days ago	...
↕ clicked link with text "Contact"	1 18336a2f21c1ed-0c5d5...800-18336a2f21d3afd	socamper.com/	web	8 days ago	...
↕ submitted form	1 18336a2b5cf1271-0b03...800-18336a2b5d012d8	socamper.com/#contact	web	8 days ago	...
↕ clicked button with text "Submit"	1 18336a2b5cf1271-0b03...800-18336a2b5d012d8	socamper.com/#contact	web	8 days ago	...
↕ clicked link with text "Contact"	1 18336a2b5cf1271-0b03...800-18336a2b5d012d8	socamper.com/blog/	web	8 days ago	...
↕ clicked link with text "Blog"	1 18336a2b5cf1271-0b03...800-18336a2b5d012d8	socamper.com/	web		Create action from event
September 12, 2022					

Action selection is similar to creating an event and trigger in GTM:

Match Group #1

Autocapture
Custom event
Page view

Select element on site
See documentation.

Link target equals
If your element is a link, the location that the link opens (href tag)

AND

Text equals
Text content inside your element

AND

HTML selector matches ⓘ
CSS selector or an HTML attribute that ideally uniquely identifies your element. Example: `[data-attr="signup"]`

AND

Page URL
contains
matches regex
matches exactly
Elements will match only when triggered from the URL (particularly useful if you have non-unique elements in different pages).

As mentioned in part one, my theoretical ideal implementation would be based on dataLayer event pushes instead of CSS selectors, but PostHog's method is easy to implement and doesn't require coding on the application side.

Another downside of this automatic capture of events is that it can be expensive at scale. Since PostHog's price is metered by event volume, if you have auto-capture on it'll drive up your volume much higher than an explicit tracking scenario where you only fire the events you choose. You can turn off auto-capture to lower event volume, but that defeats one of the benefits of the product.

Their pricing does include significant discounts at scale, but, as we talked about in the Pricing chapter, sometimes event volume-based just isn't the most advantageous model.

Events	Cost Per Million Events
--------	-------------------------

Events	Cost Per Million Events
< 1 million	\$0
1-2 million	\$450
2-10 million	\$225
10-100 million	\$75
100 million-1 billion	\$25

This means a site with 10 million events a month would cost \$2,250 monthly, and 25 million would be \$3,400. Step functions like this can also be less advantageous if you sit near the top of a range, meaning PostHog is a better deal for a site with 15 million events vs. 10 million.

As an example of how the landscape around a fast growing new product can change, let's look at the status of PostHog and ad blockers. PostHog is not currently blocked by the EasyPrivacy list (used by uBlock, Adblock Plus, and many others), but it has been in the past. PostHog was blocked via this list from June through October 2022, when it was removed. Even though it's not blocked currently, it's my expectation that it will be re-added to the list at some point. These lists tend towards very inclusive definitions of what a tracker is which most likely do include PostHog. This is a mixed blessing — it's nice avoid the blockers, but problematic to have that segment of traffic disappear and then re-appear.

PostHog is one of my favorites in this category because of the rich feature set and ease of use, though the product is still less mature than some of the alternatives. If you're willing to roll with the obstacles that might result from this, then PostHog could be a great fit. If you're looking for something more established, then Heap or Amplitude would be a better option for a cloud-hosted version, or Snowplow if you're looking for open source.

CSS: <css/styleSheet.css>

Chapter 30 - Conclusion

THERE'S A VERY OLD adage in tech: "Nobody ever got fired for buying IBM." This dates to the early days of computing, when IBM held a seemingly unassailable stranglehold on the industry. In the 1960s and 1970s, IBM dominated IT much more thoroughly than any of today's giants like Facebook, Apple, Microsoft, Google, etc. In 1975, IBM had over 288,000 employees — more than any of the aforementioned giants have now — in an era when the overall size of industry was much smaller (Amazon has many more employees for obvious reasons). IBM was all of FAAMG rolled into one monolith, at least in regards to the tech sides of those companies.

One meaning behind that IBM quote is that the market dominator is always the safe pick. However it can also be read that the opposite may be true. It warns that, although following the crowd and conventional wisdom can keep you out of trouble with your bosses, it's also a narrow view that puts self-interest ahead of deep thought and investigation.

IBM is still a huge and successful company, but it now has to compete on more equal terms with everyone else, and the same should be true of Google Analytics. GA could still be the best choice for you, but I hope I've given you useful insight into how to evaluate alternative products beyond simply picking the default choice for no other reason beyond convention or "best practice."

The tech industry also has a tendency to believe that paradigmatic change will happen faster than it actually does. Small changes do occur rapidly, but the underlying fundamentals stay relatively consistent and shift slowly.

We're in the midst of that sort of deep change in analytics now, as discussed in part one. However, here's a brief example of why that change is still in its early stages.

Product analytics tools have positioned themselves as replacements for GA. The reality is that most of the vendors in that category still run GA on their homepages and, in some cases, inside their web apps.

Tool	Category	Runs GA on their homepage?
Google Analytics	Traditional	✓
Cloudflare Web Analytics Simplified		✗
Matomo	Traditional	✗
Mixpanel	Product	✓
Snowplow	Product	✓
Statcounter	Simplified	✓
Amplitude	Product	✓
Chartbeat	Simplified	✓

Tool	Category	Runs GA on their homepage?
Heap	Product	✓
Clicky	Traditional	⊗
Fathom Analytics	Simplified	⊗
Plausible Analytics	Simplified	⊗
Visitor Analytics	Simplified	⊗
PostHog	Product	⊗
Piwik PRO	Traditional	⊗

In the above table, the red "X"s are a positive sign, while the green checks should be read as cautionary symbols indicating who doesn't eat their own dog food.

Within the traditional and simplified categories, only Statcounter runs GA. In the product analytics category, PostHog is the only one that **doesn't** run GA. Some don't even run their own tool on their homepage at all. I think this speaks volumes about how marketing teams remain uncomfortable and dissatisfied using product analytics tools alone. It's hard to claim you're a complete GA alternative when you still run GA on your own site.

This is one of the reasons why I like looking at actual implementations and numbers derived from them, such as the BuiltWith numbers we've referenced throughout this guide. Popularity as a single measurement point is noisy and fallible, but very useful to see the state of the field and the direction it's going.

When looking at vendors and tools, I want to pick one that:

- Is currently used widely on sites with real traffic
- Will continue to operate and be supported and improved
- Is growing and has a promising looking future

Let's look one final time at popularity, this time including year-over-year trends:

Tool	Deployments on Top 1M Sites	Yearly Trend	YoY Change
Google Analytics	464,897		-5%
Cloudflare Web Analytics	20,068		18%
Matomo	20,033		-1%
Snowplow	8,907		10%
Amplitude	6,193		16%
Statcounter	6,060		-8%
Mixpanel	5,395		-66%
Chartbeat	3,417		-7%
Heap	3,358		16%
Clicky	2,925		-7%
Fathom Analytics	1,667		30%
Plausible Analytics	1,358		92%
PostHog	1,279		498%
Visitor Analytics	1,207		1%
Piwik PRO	111		13%

All of the options we've evaluated are credible solutions, but I do question the products that have been unable to grow in this highly advantageous environment. That 5% loss for GA means there are plenty of potential customers out there looking.

While there are benefits to picking a tool whose popularity is increasing, it's only one factor to consider. PostHog's incredible growth numbers don't mean we should all jump on the bandwagon. Similarly, Clicky's lack of growth shouldn't eliminate it as a possibility.

As analysts, we should also keep in mind possible data quality issues with BuiltWith. Mixpanel's huge drop is concentrated around one time frame and could be the result of a measurement error, or the loss of a single large-scale

partner. Mixpanel is under competitive pressure from Amplitude, Heap, and PostHog — all relatively similar products with impressive growth numbers — but as we see with GA4 migration does not happen as quickly as that graph indicates.

Moreover, this is only a snapshot of a single point in time in a rapidly changing industry. These trending numbers are just another data point to help us make a better-informed decision.

Evaluating and discussing tools can be fun, and the choice of tool is important. Ultimately, how you implement and use a tool is the more important part. Having a clear tracking plan, a robust reporting methodology, and a strong data culture will take you much further than the marginal benefits of tool A vs. tool B.

I hope this guide has been helpful in your analytics tool selection process, and I wish you the best of luck in your UA migration. Thanks for reading.

THANKS AND ACKNOWLEDGMENTS

Andrew Packer for reading and commenting incisively on whatever version of my writing I put in front of him. Quantable wouldn't be Quantable without him, and not just because he came up with the name.

My parents.

Ahmad Ahmad for the idea that someone might actually pay money for a book I wrote.

Lily Liaw for being my writing buddy.

Tim Wilson for all the great suggestions and help.

Kathryn Bradley for being a beta reader and for the marketing assistance.

Mac Liaw for the encouragement and nth-tier tech support.

Most importantly, Andrea for always believing in me.

APPENDIX

This appendix contains the full comparison spreadsheet I've created. As spreadsheets aren't very usable in printed form, you may prefer the full Google Sheet available here:

<https://gaalternatives.guide/sheet>

All usage numbers from BuiltWith.com as of November, 2022

Per-month prices as of January, 2023 — subject to change

Prices in italics are estimates

Company size is rough estimate based upon LinkedIn users

Tool	Deployments on Top 1M Sites	Yearly Trend	YoY Change	Category
Google Analytics	464,897		-5%	traditional
Cloudflare Web Analytics	20,068		18%	simplified
Matomo	20,033		-1%	traditional
Snowplow	8,907		10%	product
Amplitude	6,193		16%	product
Statcounter	6,060		-8%	simplified
Mixpanel	5,395		-66%	product
Chartbeat	3,417		-7%	simplified
Heap	3,358		16%	product
Clicky	2,925		-7%	traditional
Fathom Analytics	1,667		30%	simplified
Plausible Analytics	1,358		92%	simplified
PostHog	1,279		498%	product
Visitor Analytics	1,207		1%	simplified
Piwik PRO	111		13%	traditional
GA4				product

Tool	Default Tracking Method	Mobile SDK	Self Hosted Option	Open Source	Level of automatic event tracking
Google Analytics	cookies	✓	✗	✗	low
Cloudflare Web Analytics	referrer	✗	✗	✗	n/a (no events)
Matomo	cookies	✓	✓	✓	low
Snowplow	cookies	✓	✓	✓	low
Amplitude	cookies	✓	✗	✗	very low
Statcounter	cookies	✗	✗	✗	low
Mixpanel	cookies	✓	✗	✗	low
Chartbeat	cookies	✗	✗	✗	n/a
Heap	cookies	✓	✗	✗	high
Clicky	UA + IP	✗	✗	✗	low
Fathom Analytics	UA + IP	✗	✗	✗	low
Plausible Analytics	UA + IP	✗	✓	✓	low
PostHog	cookies	✓	✓	✓	high
Visitor Analytics	UA+IP	✗	✗	✗	high (if enabled)
Piwik PRO	cookies	✓	✓	✗	low
GA4	cookies	✓	✗	✗	medium

Tool	Free Tier	Small Site Cost (100k hits)	Medium Site Cost (2M hits)	Large Site Cost (25M hits)	pricing model
Google Analytics	10M hits	\$0	\$0	\$12,500	hits
Cloudflare Web Analytics	always free	\$0	\$0	\$0	n/a
Matomo	none	\$39	\$369	\$4,390	hits
Snowplow	none	?	?	?	hits
Amplitude	100K users	\$0	\$240	?	users
Statcounter	500 pvs	\$9	\$49	\$249	pvs
Mixpanel	20M hits	\$59	\$216	\$2,170	hits
Chartbeat	none	n/a	\$580	?	pvs
Heap	10k sessions	\$375	?	?	sessions
Clicky	90k pvs	\$10	\$20	\$188	pvs
Fathom Analytics	none	\$14	\$74	?	pvs
Plausible Analytics	none	\$19	\$89	?	pvs
PostHog	1M hits	\$0	\$1,100	\$3,400	hits
Visitor Analytics	400 visits	\$13	\$56	?	visitors
Piwik PRO	500k hits	\$0	\$545	\$3,500	hits
GA4	25M hits?	\$0	\$0	\$4,167	hits

Tool	Company HQ	Company Size (LinkedIn)	Runs GA on their homepage?	UA data importer
Google Analytics	United States	307,000	✓	n/a
Cloudflare Web Analytics	United States	3,300	✗	✗
Matomo	New Zealand	24	✗	✓
Snowplow	United Kingdom	175	✓	✗
Amplitude	United States	780	✓	✗
Statcounter	Ireland	6	✓	✗
Mixpanel	United States	440	✓	✗
Chartbeat	United States	75	✓	✗
Heap	United States	411	✓	✗
Clicky	United States	?	✗	✗
Fathom Analytics	Canada	2	✗	⚠️
Plausible Analytics	Estonia	5	✗	✓
PostHog	United States	34	✗	✗
Visitor Analytics	Germany	33	✗	⚠️
Piwik PRO	Poland	100	✗	✗
GA4			✓	✗



Your gateway to knowledge and culture. Accessible for everyone.



z-library.se

singlelogin.re

go-to-zlibrary.se

single-login.ru



[Official Telegram channel](#)



[Z-Access](#)



<https://wikipedia.org/wiki/Z-Library>