

Google Analytics vs AWStats

Understanding web traffic statistics software

Willem L. Middelkoop

Aug. 20, 2018



This month I have been working on website statistics, tracking traffic using different technologies. Some of my customers use Google Analytics, others use AWStats, and some use both. Which is better is often debated, but few people really understand the differences. Time to shed some light on the magic of web statistics.

Comparing AWStats with Google Analytics

You might expect that both web analytics programs would report the same amount of website traffic, but this is not the case. I'll try to explain why reports from Analytics and AWStats differ (and why I think it doesn't really matter).



Google Analytics and AWStats - web traffic statistics software

Google Analytics

As part of the Google Marketing Platform, Google Analytics is one of the most widely used web traffic analytics programs available. Google launched the service in 2005 after acquiring multiple software companies specialised in analytics, most notably Urchin and Adaptive Path.

How Google Analytics works

Google Analytics works by placing a small piece of programming code on your website (or inside your app). This initiates the tracking and profiling used by Google to generate web statistics. The code makes your internet device connect to the servers from Google - which then count all these connections to generate usage statistics.

Google Analytics availability

Google offers Analytics without charge: when you use Google Analytics you agree to share your traffic data with Google. This allows Google to combine Analytics data from different websites to create accurate profiles of people surfing the web. While these profiles are great to deliver personalised ads, they are also reason for privacy concern. Google Analytics is used on websites, inside apps and games.

AWStats

Originally created by Frenchman, [Laurent Destailleur](#) in 2000, AWStats is a widely used server program that analyses web server log files. The name AWStats is an abbreviation for "Advanced Web Statistics".

How AWStats works

The web server log contains the complete record of website traffic. Inside the log file, individual connections between the web server and the internet device (phone, computer, tablet) are registered. After analysing this activity record, AWStats provides detailed insights in web traffic.

AWStats availability

Everybody can use AWStats as it is available as free software, including its source code. Earlier, I wrote about [helping people with free software](#), explaining the benefits of free software (it's not just money). AWStats must have access to the server log files, which

is often done by installing AWStats on the server itself. Because of this, AWStats has a limited scope. It is not able to track individual users across different websites. AWStats is used for web, mail, streaming and file servers.

Difference: unit of measurement

Probably the most important difference between the systems is the unit of measurement. Google Analytics mainly works with "Sessions" while AWStats uses (unique)visits:

- **Google Analytics session:** a series of page views from the same browser on a website with no longer than 30 minutes between clicks, except at midnight. See [explanation from Google](#).
- **AWStats (unique)visit:** a (unique) IP-address accessing a series of pages on a web server within one hour. If the series spans beyond the hour, the visit is counted as two. See [explanation in the AWStats docs](#).

Difference: where data is collected

In addition to the unit of measurement, the place where data is collected is different, too. AWStats uses the web server as place to measure. Quite literally the software tallies how often the server receives a request to deliver a specific page. The server responds to human visitors, but also to automated requests.

Bots

These automated requests are done by bots, computer programs that automatically surf the web. Often these bots have a specific task, like indexing a website for search engine listing. Most bots are identified by their originating IP-address or so-called UserAgent (a sort of identification text which states the origin of the bot; like "GoogleBot"). Most bots are not included in AWStats visits, but are specified in a separate listing.

Robots/Spiders visitors	Hits	File size	Last visit
Googlebot-Image	17,671	2.14 GB	20 Aug 2018 - 03:48
Bingbot	15,663+1489	4.19 GB	20 Aug 2018 - 04:27
Seznambot	11,926+3019	2.98 GB	20 Aug 2018 - 04:02
Ahrefsbot	13,724+289	3.36 GB	20 Aug 2018 - 04:30
empty user agent string	13,328+23	1.07 GB	20 Aug 2018 - 04:30
Crawl	13,146+135	1.06 GB	20 Aug 2018 - 04:35
Googlebot	12,891+249	576.62 MB	20 Aug 2018 - 04:35
Yandexbot	6,330+4257	300.78 MB	20 Aug 2018 - 19:22
Semrushbot	2,461+3359	282.19 MB	20 Aug 2018 - 02:57
DotBot	4,321	1.42 GB	20 Aug 2018 - 22:21
YandexImages	3,498	894.91 MB	20 Aug 2018 - 00:39
Facebookexternalhit	3,188	243.94 MB	24 May 2018 - 14:15
Mediapartners-Google	2,088+827	115.42 MB	20 Aug 2018 - 14:15
MJ12bot	2,414+89	262.81 MB	19 Aug 2018 - 14:15
Unknown robot identified by bot *	2,338+16	718.51 MB	19 Aug 2018 - 07:59
Google Web Preview	2,338+16	118.29 MB	17 Aug 2018 - 03:31
robot	1,700+48	101.83 MB	26 Jul 2018 - 06:17
BaiduSpider	1,238+64	158.68 MB	20 Aug 2018 - 06:17
CCBot	1,125+141	79.05 MB	19 Aug 2018 - 18:00
BLEXBot	1,118+1	51.80 MB	19 Aug 2018 - 06:49
HubSpot Webcrawler	1,095+7	45.05 MB	19 Aug 2018 - 18:38
MegaIndex.ru	960	48.86 MB	10 Aug 2018 - 18:00
BingPreview	932	124.54 MB	18 Aug 2018 - 04:46
Avicon	814+33	26.81 MB	19 Aug 2018 - 15:50
dataproducer.com	692+107	47.08 MB	18 Aug 2018 - 18:11
Firefox version 19 and lower - various robots	784	48.20 MB	18 Aug 2018 - 11:39
	447+254	21.02 MB	20 Aug 2018 - 00:58
	573	24.50 MB	20 Aug 2018 - 02:40
	538	29.96 MB	20 Aug 2018 - 02:40
	518+6	22.03 MB	19 Aug 2018 - 04:08
	0+483	40.88 MB	15 Aug 2018 - 00:29
	272+185	20.64 MB	16 Apr 2018 - 05:24
	204+82	12.35 MB	19 Aug 2018 - 06:22
	208+98	14.77 MB	19 Aug 2018 - 21:18
	167	4.43 MB	13 May 2018 - 01:29
		13.07 MB	20 Aug 2018 - 19:06
		4.42 MB	11 Mar 2018 - 10:50
			2018 - 03:49

Robots surfing the web: bot traffic listed by AWStats

Bad bots

Unfortunately not all bots behave nicely. Some of them are working for hackers to find potential security weaknesses. Others, like some search engine indexation robots, behave like humans to test if the website shows the same page to both humans and bots. Serving different content based on user agent is a bad practice used by stupid websites to manipulate search engine ranking.

Recognising bots

AWStats tries to recognise bot behaviour by taking into account the sequence and speed of page requests (like a normal, healthy human, should not be able to read more than two pages per second). But, this cannot prevent some bots to successfully disguise themselves as human, causing AWStats to count them along.

```
42-45.166.21 - [20/Aug/2018:20:40:36 +0000] "GET /en/upload HTTP/2.0" 200 46779 "https://willem.com/en/" "Mozilla/5.0 (X11; Linux x86_64; rv:52.0) Gecko/20100101 Firefox/52.0"
42-45.166.21 - [20/Aug/2018:20:40:47 +0000] "GET /blog/2018-03-07-ipad-pre-review-comparing-the-microsoft-surface-pre-to-ipad-pre/ HTTP/2.0" 200 43079 "-" "Mozilla/5.0 (compatible; Googlebot/2.1; http://www.google.com/bot.html)"
42-45.166.21 - [20/Aug/2018:20:41:13 +0000] "GET / HTTP/2.0" 200 308 "https://willem.com/en/upload/" "Mozilla/5.0 (X11; Linux x86_64; rv:52.0) Gecko/20100101 Firefox/52.0"
42-45.166.21 - [20/Aug/2018:20:41:36 +0000] "GET /blog HTTP/2.0" 200 51233 "https://willem.com/en/" "Mozilla/5.0 (X11; Linux x86_64; rv:52.0) Gecko/20100101 Firefox/52.0"
42-45.166.21 - [20/Aug/2018:20:41:37 +0000] "GET /blog/2018-06-02-creating-a-memoir-iphone-app/ HTTP/2.0" 200 46022 "https://willem.com/en/" "Mozilla/5.0 (X11; Linux x86_64; rv:52.0) Gecko/20100101 Firefox/52.0"
42-45.166.21 - [20/Aug/2018:20:41:42 +0000] "GET /blog/2018-06-02-removing-paint-from-a-bike-frame/ HTTP/2.0" 200 46926 "https://willem.com/en/" "Mozilla/5.0 (X11; Linux x86_64; rv:52.0) Gecko/20100101 Firefox/52.0"
42-45.166.21 - [20/Aug/2018:20:41:46 +0000] "GET /blog/2018-07-10-enjoy-a-festival-without-alcohol/ HTTP/2.0" 200 66605 "https://willem.com/en/" "Mozilla/5.0 (X11; Linux x86_64; rv:52.0) Gecko/20100101 Firefox/52.0"
42-45.166.21 - [20/Aug/2018:20:41:55 +0000] "GET /blog/2018-06-06-launch-by-gnome/ HTTP/2.0" 200 63308 "https://willem.com/blog/" "Mozilla/5.0 (X11; Linux x86_64; rv:52.0) Gecko/20100101 Firefox/52.0"
42-45.166.21 - [20/Aug/2018:20:41:56 +0000] "GET /blog/2018-01-23-something-the-smartwatch-will-never-have-patino/ HTTP/2.0" 200 60380 "https://willem.com/blog/" "Mozilla/5.0 (X11; Linux x86_64; rv:52.0) Gecko/20100101 Firefox/52.0"
42-45.166.21 - [20/Aug/2018:20:42:00 +0000] "GET /blog/2018-07-20-helping-people-with-free-software/ HTTP/2.0" 200 63919 "https://willem.com/blog/" "Mozilla/5.0 (X11; Linux x86_64; rv:52.0) Gecko/20100101 Firefox/52.0"
42-45.166.21 - [20/Aug/2018:20:42:08 +0000] "GET /blog/working HTTP/2.0" 200 47897 "https://willem.com/blog/2018-07-20-helping-people-with-free-software/" "Mozilla/5.0 (X11; Linux x86_64; rv:52.0) Gecko/20100101 Firefox/52.0"
42-45.166.21 - [20/Aug/2018:20:42:08 +0000] "GET /blog/2017-07-31-working-from-4g-turned-off-my-lan-and-went-4g-only/ HTTP/2.0" 200 44042 "https://willem.com/blog/working/" "Mozilla/5.0 (X11; Linux x86_64; rv:52.0) Gecko/20100101 Firefox/52.0"
42-45.166.21 - [20/Aug/2018:20:42:10 +0000] "GET /blog/2017-07-31-the-day-i-killed-my-lan/ HTTP/2.0" 200 61792 "https://willem.com/blog/2017-07-31-working-from-4g-turned-off-my-lan-and-went-4g-only/" "Mozilla/5.0 (X11; Linux x86_64; rv:52.0) Gecko/20100101 Firefox/52.0"
42-45.166.21 - [20/Aug/2018:20:42:10 +0000] "GET /blog/2017-07-31-working-from-4g-turned-off-my-lan-and-went-4g-only/" "Mozilla/5.0 (X11; Linux x86_64; rv:52.0) Gecko/20100101 Firefox/52.0"
```

Part of a web server log file, each line corresponds with a page request - do you see the bot?

JavaScript

Instead of running software on the server like AWStats, Google Analytics works by executing code (JavaScript) on the device of the visitor. Depending on how the Analytics JavaScript is integrated, the measurement is often delayed because it takes some time

to load JavaScript. Sometimes it only starts after the entire page (including images and fonts) has completely finished loading. People that leave the page before it is entirely loaded, or people that have JavaScript disabled, are therefore not measured. Because Google Analytics requires JavaScript, most bots are not measured because it is hard for bots to simulate a JavaScript enabled browser.

```

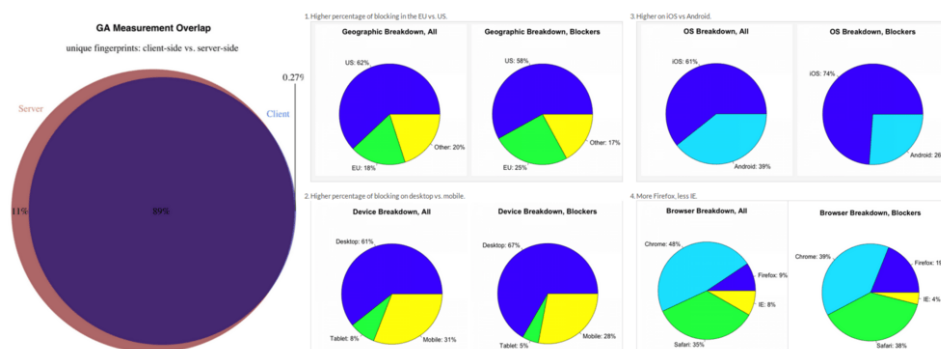
1 <script>
2 (function(i,s,o,g,r,a,m){i['GoogleAnalyticsObject']=r;i[r]=i[r]||function(){
3 (i[r].q=i[r].q||[]).push(arguments)},i[r].l=1*new Date();a=s.createElement(o),
4 m=s.getElementsByTagName(o)[0];a.async=1;a.src=g;m.parentNode.insertBefore(a,m)
5 })(window,document,'script','https://www.google-analytics.com/analytics.js','ga');
6
7 ga('create', 'UA-XXXXXXX-1', 'auto');
8 ga('send', 'pageview');
9
10 </script>

```

Typical piece of Google Analytics JavaScript code - you can clearly see it referring to Google's servers

Blocking Google Analytics

Because of the privacy concerns with regard to Google Analytics, some people actively block the JavaScript used by Google Analytics. There are browser plugins and ad blockers that do this automatically, without requiring any technical knowledge of the user. Another clever web developer used a purpose built server setup to find out how many people block Google Analytics. In his experiment, [Jason Packer](#) found that about 8-11% of the users are blocking Google Analytics.



Percentage of people blocking Google Analytics (2016), quantable.com - Jason Packer

Conclusion

Both the unit of measurement and the place where the measurement is performed cause the differences in statistics by Google Analytics and AWStats. Bots included, AWStats counts too much, while Google Analytics counts too little because of problems with JavaScript loading and blocking.

But in the end it doesn't really matter as you should use these statistic programs mainly to analyse relative performance. Consistent inaccuracies in absolute numbers do not really influence answers to questions like these:

- Is my website doing better than yesterday, last month or last year?

- What page, blog post or webshop product is most popular?
- Is average traffic increasing? Is my advertising campaign having any effect?

If you really want to know *who* is visiting your website, you should really consider asking the visitor to identify him or herself. That is way more accurate (and polite!) than relying on Google Analytics or AWStats to identify unique persons. That way you don't have to worry about people having multiple devices and accessing your website from different locations like work, home or on the go (train, bus).

Use statistics software to track relative performance, but consider getting to know your visitor by connecting to them instead of collecting data from them. Build relationships instead of data profiles. That's where the value is.