

# Kapitel 1

## Ziele der Webanalyse

*»Für einen, der nicht weiß, welchen Hafen er ansteuern will, gibt es keinen günstigen Wind.« (Seneca)*

### 1.1 Was ist Webanalyse?

Die Digital Analytics Association (DAA) definiert Webanalyse wie folgt:

*Web analytics is the measurement, collection, analysis and reporting of web data for purposes of understanding and optimizing web usage. (2008)*

Diese Definition ist allerdings schon ein wenig in die Jahre gekommen, denn mittlerweile werden Webanalysten schon häufiger Digitalanalysten genannt, da sie nicht nur Daten aus dem Web analysieren, sondern auch aus Apps und anderen Quellen. Und so hat auch die Web Analytics Association das »Web« aus dem Namen gestrichen und durch »Digital« ersetzt. Eine aktualisierte Definition könnte somit lauten:

*Digitale Analyse beinhaltet das Messen, Sammeln, Analysieren und Berichten von digitalen Daten zum Zweck des Verstehens und der Optimierung der Nutzung von digitalen Produkten.*

In diesem Buch wird zwar weiterhin der Begriff *Webanalyse* verwendet (da dieser zurzeit noch populärer ist), allerdings sind viele der Prinzipien auch auf das erweiterte Feld der Digitalanalyse anwendbar.

### 1.2 Ohne Ziele keine Analyse

Das Seneca-Zitat bringt einen wesentlichen Aspekt der Webanalyse auf den Punkt: Webanalyse wird nicht für sich selbst betrieben, sondern um zu messen, ob man auf dem richtigen Weg zum Erreichen eines Ziels ist. Ist das Ziel unbekannt, kann man so viel messen, wie man will – das Ergebnis spielt keine Rolle. Um bei dem Seneca-Zitat zu bleiben: Natürlich kann man messen, wie viele Knoten ein Boot in welche Richtung fährt, aber was macht man mit diesen Daten, wenn man ohnehin nicht weiß, wohin man will?

Eine Website ist natürlich kein Boot. Aber eine Website kann im übertragenen Sinne als eine Art Vehikel verstanden werden, mit dem Ziele verfolgt werden. Sind diese Ziele nicht vorhanden oder schwammig formuliert («Awareness steigern» oder »wichtiger Baustein der digitalen Strategie«), fehlt ein Referenzpunkt, der für die Einordnung der eigenen Daten notwendig ist. Um im Bild der Seefahrt zu bleiben: Wenn die gegenwärtige Position gemessen 54°03'59.7"N 8°05'24.2"E ist und das Boot mit 10 Knoten den Kurs 090 hält, kann man mit diesen Werten wenig anfangen, wenn wir nicht wissen, wo wir eigentlich hin wollen oder nur eine grobe Richtung angegeben ist (Osten). Beispiel: Der Anteil der wiederkehrenden Nutzer liegt bei 17,5 %. Ist das gut oder schlecht? Darauf kann keine Antwort gegeben werden, wenn die Ziele nicht klar sind. Webanalyse ist kein Ersatz für eine mangelhafte Zielsetzung. Ein Webanalyst ist dafür verantwortlich, dass sinnvolle Ziele definiert werden, idealerweise zu Beginn eines Projekts.<sup>1</sup>

Das klingt zunächst trivial, in der Regel ist es aber kein einfacher Vorgang. Typische (und leider auch authentische) Zielformulierungen aus der Praxis, die *auf keinen Fall* als Vorbild dienen sollten, sind:

- ▶ Unsere Website soll die Firma repräsentieren.
- ▶ Ich will Aufträge über die Website bekommen.
- ▶ Ich will Geld mit Werbung auf meiner Website verdienen.
- ▶ Die Website dient dem Abverkauf unserer Produkte.
- ▶ Wir wollen die Awareness für das Produkt steigern.
- ▶ Ich will mehr Besucher haben als mein Konkurrent.

Keines dieser Ziele ist so konkret, dass es für eine Analyse geeignet wäre. Beispiel »Awareness für ein Produkt«: Wenn eine Person zusätzlich von dem Produkt weiß (zum Beispiel der Webanalyst, der gerade über das Produkt aufgeklärt wird), ist die Awareness bereits um eine Person gesteigert und das Ziel erfüllt, oder? Wenn eine Person nicht reicht, wie viele müssten es denn sein, damit das Ziel erfüllt ist? Spätestens zu diesem Zeitpunkt wird klar, dass vage Ziele nicht zum Vorteil gereichen. Gleichzeitig ist der Ansatz der klaren Zieldefinition zum Teil ungewohnt und kann für Teilnehmer auch bedrohlich wirken oder nervig sein.<sup>2</sup>

Ein Ziel vorab zu definieren, ist nicht nur für Webanalyse-Daten wichtig, sondern zum Beispiel auch für die Analyse von Kampagnen-Daten. Gibt es kein Ziel für eine

<sup>1</sup> Der Autor hat mehrere Workshops in Firmen geleitet, bei denen sich die Teilnehmer aus derselben Firma nicht einig werden konnten, was das Ziel ihrer Firmenwebsite ist. Zum Teil waren das emotionale Diskussionen, die zum ersten Mal geführt wurden.

<sup>2</sup> So wurde einmal in einem Workshop von einem Kunden formuliert, dass man lieber keine Ziele haben möchte, da dann ja die eigene Leistung messbar sei.

Kampagne, kann eine Auswertung dieser Kampagne auch nicht belegen, dass die Kampagne erfolgreich war. Zwar gibt es Erfahrungswerte für manche Kampagnentypen (zum Beispiel Öffnungsraten bei E-Mail-Newslettern), aber diese lassen sich nur bedingt von einer Marke oder Kampagne auf eine andere übertragen.<sup>3</sup>

### 1.3 Ziele nach dem S.M.A.R.T.-Prinzip

Ziele sollten stets nach dem S.M.A.R.T.-Prinzip definiert werden, ein Akronym für:

- ▶ S: Spezifisch
- ▶ M: Messbar
- ▶ A: Attraktiv
- ▶ R: Realistisch
- ▶ T: Terminiert

Beispiele:

- ▶ Mit einer Content-Website sollen innerhalb eines Jahres 500.000 EUR Umsatz bei maximal 10.000 EUR Kosten im Monat erreicht werden.
- ▶ Mit einer NGO-Website sollen Spenden generiert werden, um die abnehmende Anzahl von Offline-Spenden zu kompensieren, idealerweise sogar mehr als zu kompensieren. Dies muss innerhalb von 24 Monaten bei einem Budget von 20.000 EUR und gleichbleibender oder steigender Spendenhöhe geschehen.
- ▶ Mit einer Gaming-Seite sollen innerhalb von 6 Monaten 10.000 EUR pro Monat durch den Verkauf von virtuellen Gegenständen bei maximalen Kosten von insgesamt 2.000 EUR erzielt werden.

Auch wenn Ziele für die Analyse elementar sind, sind sie nicht in Stein gemeißelt. So kann es passieren, dass bereits vor der Zielerreichung bemerkt wird, dass ein Ziel beziehungsweise ein Unterziel nicht unbedingt das bringt, was man sich wünscht. So könnte ein Ziel sein, dass die Anzahl der Bestellungen steigt. Wenn sich aber gleichzeitig die Anzahl der Rücksendungen disproportional erhöht, könnte sich damit herausstellen, dass eine höhere Anzahl von Bestellungen für sich allein gesehen nicht erstrebenswert ist.

<sup>3</sup> Mehr als einmal hat der Autor in über 20 Jahren in diesem Bereich erlebt, dass in solchen Fällen einfach irgendwelche Daten präsentiert wurden, denn »für irgendwas würden sie schon gut sein«. Schließlich haben die Erfinder des Webanalyse-Systems diese Daten »ab Werk« bereitgestellt und sich sicherlich etwas dabei gedacht.

## 1.4 Die Gewinnerzielungsabsicht

Alle oben genannten Ziele haben mit Geld zu tun. Fast alle Unternehmen verfolgen mit ihrer Website ein gemeinsames Ziel, nämlich in irgendeiner Weise dazu beizutragen, dass das Unternehmen einen finanziellen Gewinn erzielt.<sup>4</sup> Die Gewinnerzielungsabsicht ist Voraussetzung für einen Gewerbebetrieb. Selbst wenn es um eine reine Informationsseite geht, hat auch diese ein Ziel, das zumindest indirekt auf eine Gewinnerzielung hinausläuft (siehe dazu Abschnitt 1.5, »Business Value Generation: der Antrieb«).

Zurück in das Bild der Schifffahrt: Wenn die Website ein Schiff ist, hat der Webanalyst zu messen, ob und wie das Schiff den Gewinnhafen erreicht – und zwar nicht irgendeinen Gewinnhafen, sondern den Hafen mit dem geplanten Gewinn.

Das bedeutet nicht, dass der Gewinn einer Firma allein von der Website abhängt. Aber bei allem, was die Website leisten soll, muss auf das Verfolgen der Gewinnerzielungsabsicht geachtet werden.

## 1.5 Business Value Generation: der Antrieb

Da Webanalyse Kosten erzeugt, sowohl das Gehalt des Webanalysten als auch gegebenenfalls für das Webanalyse-Tool, müssen die Erkenntnisse der Webanalyse einen Wert schaffen, der die Kosten dafür übersteigt. Webanalyse muss zur Wertschöpfung beitragen, sodass bei jeder Aufgabe die Gretchenfrage gestellt werden sollte, inwieweit diese Aufgabe der Zielerreichung dient. Auf Neudeutsch wird dies manchmal als *Business Value Generation* bezeichnet. Gerade bei Webanalysten, die viele Aufgaben zu bewältigen haben, hilft diese Gretchenfrage, den Arbeitstag zu priorisieren und sich gegen »Mach-mal-eben-Anfragen« zu wehren. Gleichzeitig sollte der Beitrag zur Wertschöpfung der Antrieb für den Webanalysten sein. Er agiert wie ein Unternehmensberater, der dem Unternehmen hilft, mehr Gewinn zu erwirtschaften.

Ziele sollten stets irgendwie zum obersten Ziel, dem Gewinn, beitragen. Da die meisten Firmen limitierte Ressourcen haben, sollte transparent sein, was eine Initiative kostet und welchen Betrag sie einbringt, um priorisieren zu können. Auch hier kann der Webanalyst helfen, datengetriebene Entscheidungen zu ermöglichen.

Für Start-ups kann es schwierig sein, derart genaue Ziele zu definieren; der Businessplan sollte aber eine Indikation liefern, welche Werte erreicht werden müssen, um Profitabilität und Skalierung zu erreichen.

<sup>4</sup> NGOs (die aber eventuell auch Geld aus Spenden sammeln), Hobbyprojekte usw. werden hier ausgeklammert.

## 1.6 Was, wenn es nicht direkt um Geld geht?

Nicht alle Webseitenbereiche zielen direkt auf das Ziel des Gewinns ab, und bei manchen Websites fällt es zunächst schwer, einen monetären Zusammenhang zu sehen. Aber auch hier lassen sich Modelle finden:

- ▶ Ein Jobbereich soll dafür sorgen, dass Bewerber alle notwendigen Informationen finden, sodass die Personalakquise beschleunigt wird (und somit Headhunter-Kosten oder Auftragsausfälle durch zu geringe Ressourcen verhindert werden).
- ▶ Die Unternehmensdarstellung soll dafür sorgen, dass ein Nutzer Vertrauen in das Unternehmen bekommt und dazu gestimmt wird, sich bei diesem Unternehmen zu bewerben.<sup>5</sup>
- ▶ Ein Softwareentwickler betreibt eine Seite mit Open-Source-Software, die er geschrieben hat. Die Popularität der Software steigert seinen Marktwert.<sup>6</sup>

Für diejenigen, die tatsächlich kein Geld mit einer Website verdienen, können andere Ziele in Betracht kommen.

Default Channel Grouping	Acquisition			Behaviour			Conversions	
	Users ? ↓	New Users ?	Sessions ?	Bounce Rate ?	Pages/Session ?	Avg. Session Duration ?	YARPP gesehen (Goal 5 Conversion Rate) ?	YARPP gesehen (Goal 5 Completions) ?
	1,180 % of Total: 100.00% (1,180)	1,123 % of Total: 100.45% (1,118)	1,406 % of Total: 100.00% (1,406)	13.51% Avg for View: 13.51% (0.00%)	1.23 Avg for View: 1.23 (0.00%)	00:05:28 Avg for View: 00:05:28 (0.00%)	52.77% Avg for View: 52.77% (0.00%)	742 % of Total: 100.00% (742)
1. Organic Search	1,006 (85.04%)	969 (86.29%)	1,184 (84.21%)	10.14%	1.19	00:05:39	56.50%	669 (90.16%)
2. Direct	141 (11.92%)	128 (11.40%)	177 (12.59%)	35.03%	1.12	00:03:37	31.07%	55 (7.41%)
3. Referral	30 (2.54%)	25 (2.23%)	36 (2.56%)	19.44%	2.89	00:08:02	44.44%	16 (2.16%)
4. (Other)	3 (0.25%)	1 (0.09%)	3 (0.21%)	0.00%	2.00	00:07:31	0.00%	0 (0.00%)
5. Social	3 (0.25%)	0 (0.00%)	6 (0.43%)	16.67%	1.67	00:07:09	33.33%	2 (0.27%)

Abbildung 1.1 Weiches Ziel – die Sichtbarkeit eines Elements am Ende eines Artikels

In Abbildung 1.1 ist eine andere Art von Ziel zu sehen, die auf der Website des Autors verfolgt wird, nämlich die Sichtbarkeit eines Elements am Ende eines Artikels. Natürlich ist das zunächst auch kein Ziel (dazu später mehr in Kapitel 2, »Die Dreifaltigkeit der Datenanalyse«), aber die Idee dahinter ist, dass die Website nur dann ihre Intention erreicht, wenn die Inhalte auch tatsächlich gelesen werden, was hier durch das Erreichen des Endes eines Artikels gemessen wird.

<sup>5</sup> Hier müsste der Effekt genauer nachgewiesen werden.

<sup>6</sup> Allerdings existieren hier auch genügend Beispiele dazu, dass die Software auch ohne solche Hintergedanken geschrieben und zur Verfügung gestellt wird.

## 1.7 Findet ein gutes Tool nicht selbst heraus, was wichtig ist?

Der große Traum in der Data-Science-Welt ist, dass eine künstliche Intelligenz automatisiert Daten auswerten und Handlungsempfehlungen aussprechen kann. Abbildung 1.2 zeigt ein Beispiel aus Google Analytics, in dem automatisch Analysen durchgeführt wurden.

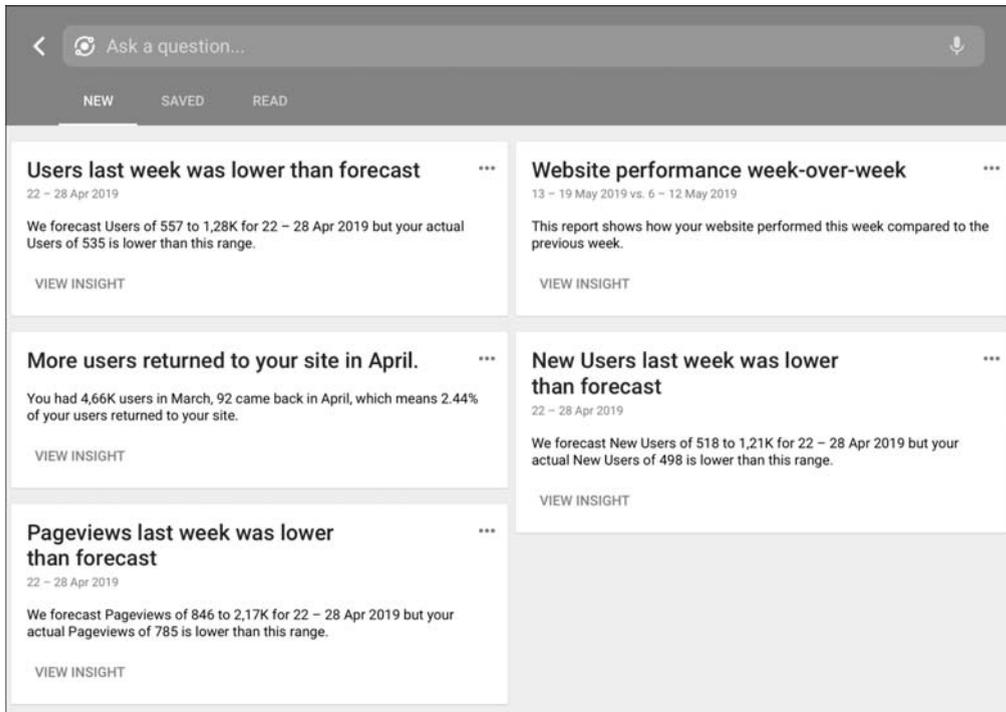


Abbildung 1.2 Automatische Analysen in Google Analytics

Die Relevanz solcher Empfehlungen ist allerdings mitunter fraglich, denn solange das Ziel nicht klar ist, kann auch keine Aussage getroffen werden, ob sich eine Veränderung der Werte tatsächlich auf die Zielerreichung auswirkt. Zwar können auch Ziele in den Webanalyse-Tools definiert werden, aber diese sind nicht als wirkliche Ziele definiert (x EUR Umsatz bis Ende des Jahres), sondern zum Beispiel lediglich die Variable *Umsatz*.

## 1.8 Vanity Metrics

Neben den gewinnorientierten Zielen hinaus existieren Eitelkeitsziele oder Eitelkeitsmetriken (Vanity Metrics), auf die hier nicht weiter eingegangen wird, da sie nicht messbar sind oder etwas messen, das toll aussieht, aber keinen tieferen Sinn

bietet. Ein Beispiel hierzu sind Ziele, wie »die geilste Seite im Wettbewerb zu haben«.<sup>7</sup> Dies ist weder messbar, noch dient es der Gewinnerzielung. Eine weitere Eitelkeitsmetrik ist zum Beispiel die Anzahl der Nutzer oder Seitenaufrufe, da sie nichts über den Wert der Nutzer oder der Seitenaufrufe aussagt.

### Wichtig zu wissen

- ▶ Klar definierte Ziele, zum Beispiel nach dem S.M.A.R.T.-Prinzip, sind die Grundlage der Webanalyse. Ohne Ziel ist jede Zahl irrelevant.
- ▶ In den meisten Fällen haben die Ziele mit dem Umsatz beziehungsweise Gewinn eines Unternehmens zu tun. Jedes Unterziel und jede Messung muss auf dieses Hauptziel ausgerichtet sein

<sup>7</sup> Auch das wurde dem Autor bei einem Website-Projekt schon als Ziel genannt.

# Kapitel 5

## Hits, Seitenaufrufe und Sitzungen

*Hits sind die kleinsten Bausteine in der Webanalyse, von denen der Seitenaufruf der populärste ist. Eine Sitzung ist der Rahmen von der ersten bis zur letzten Interaktion eines Nutzers innerhalb einer definierten Zeit.*

### 5.1 Hits

Ein *Hit*, selten Treffer genannt, ist die kleinste Einheit in der Webanalyse. Manche Autoren halten Hits für wenig sinnvoll (siehe zum Beispiel Hassler 2017), und für Hits allein gesehen stimmt das auch. Allerdings ist das Konzept des Hits elementar, da alle Messungen darauf beruhen, auch wenn Hits nicht in einem Report auftauchen sollten (aus gutem Grund). Ohne das Verständnis der Anatomie eines Hits ist eine Webanalyse schwierig; daher wird hier ausführlicher darauf eingegangen.

Wie in Abschnitt 3.4, »Server- und Client-basiertes Tracking«, beschrieben, kann ein Hit dadurch entstehen, dass eine Ressource von einem Server abgerufen oder ein Ereignis ausgelöst wird (siehe Kapitel 8, »Ereignisse und Datenschicht«), oder er entsteht auf dem Tracking-Server, weil eine andere Messung stattgefunden hat und das Ergebnis dieser Messung an den Server geschickt wird. Hits im Client-basierten Tracking werden auch als *Server Call* bezeichnet. Die Anzahl der Hits bestimmt nicht nur die Kosten, sondern auch, welches Tool beziehungsweise ob die kostenlose oder kostenpflichtige Variante für die eigene Website geeignet ist. Hits spielen in der Rohdatenanalyse eine größere Rolle; in der Arbeit mit aggregierten Daten sind sie meistens nicht sichtbar.

Die einfachste Form eines Hits ist die Messung eines Seitenaufrufs (siehe Abschnitt 5.2, »Seitenaufrufe«). Darüber hinaus lassen sich Hits beim Client-basierten Tracking konfigurieren, das heißt, dass der Webanalyst bestimmt, was getrackt werden soll und was nicht. Genau hierzu ist es immens wichtig, dass man sich im Vorfeld Gedanken darüber macht, welche Daten wichtig sind, damit diese dann in der Analyse auch zur Verfügung stehen.

Da die Anzahl der Hits kostenrelevant ist oder auch nicht unbegrenzt Hits abgefeuert werden dürfen (je nach System) wird zum Teil versucht, mehrere Informationen in einen Hit zu packen. So kann zum Beispiel die Information, auf welchen Link ein Nut-

zer geklickt hat, um in der Seite zu navigieren, erst dann an den Server übertragen werden, wenn die darauffolgende Seite erreicht und ein Seitenaufruf erfolgt ist. Anstatt also einen Hit mit dem Klick auf den Link und dann mit dem Aufruf der Seite zu verursachen, entsteht lediglich ein Hit mit beiden Informationen. Dies erfordert zwar einen größeren Programmieraufwand, aber kann insbesondere bei größeren Websites viel Geld einsparen.

Sent beacon:			VM105:1
v=1&_v=j73d&aip=1&a=225244659&t=event&ni=0&_s=1&dl=https%3A%2F%2Ftom.alby.de%2Fneue-versus-wiederkehrende-benutzer-ein-nutzloser-kpi%2F&ul=de-de&de=UTF-8&dt=Neue%20versus%20wiederkehrende%20Benutzer%3A%20Ein%20nutzloser%20KPI%20-%20data%20%3E%20opinion&sd=24-bit&sr=1440x900&vp=1440x900&je=0&ec=Page&ea=Scroll&el=Scroll%20Depth&ev=25&_u=yDCAAEADQ~&jid=1915686431&gjid=257301039&cid=1752958513.1556997616&tid=UA-276221-6&_gid=897430895.1556997616&_r=1&gtm=2wg430M3MC9X&cd3=1557072494786&cd4=1557072494787.v8o1q09&cd1=1752958513.1556997616&cd2=event&z=1721288846			
<unknown>	(&gtm)	2wg430M3MC9X	VM105:1
_j1	(&jid)	1915686431	VM105:1
_j2	(&gjid)	257301039	VM105:1
adSenseId	(&a)	225244659	VM105:1
anonymizeIp	(&aip)	1	VM105:1
apiVersion	(&v)	1	VM105:1
clientId	(&cid)	1752958513.1556997616	VM105:1
dimension1	(&cd1)	1752958513.1556997616	VM105:1
dimension2	(&cd2)	event	VM105:1
dimension3	(&cd3)	1557072494786	VM105:1
dimension4	(&cd4)	1557072494787.v8o1q09	VM105:1
encoding	(&de)	UTF-8	VM105:1
eventAction	(&ea)	Scroll	VM105:1
eventCategory	(&ec)	Page	VM105:1
eventLabel	(&el)	Scroll Depth	VM105:1
eventValue	(&ev)	25	VM105:1
hitType	(&t)	event	VM105:1
javaEnabled	(&je)	0	VM105:1
language	(&ul)	de-de	VM105:1
location	(&dl)	https://tom.alby.de/neue-versus-wiederkehrende-benutzer-ein-nutzloser-kpi/	VM105:1
nonInteraction	(&ni)	0	VM105:1
screenColors	(&sd)	24-bit	VM105:1
screenResolution	(&sr)	1440x900	VM105:1
title	(&dt)	Neue versus wiederkehrende Benutzer: Ein nutzloser KPI - data > opinion	VM105:1
trackingId	(&tid)	UA-276221-6	VM105:1
viewportSize	(&vp)	1440x900	VM105:1

Abbildung 5.1 Daten, die von Google Analytics in einem Hit mitgesendet werden

Mit jedem Hit werden alle notwendigen Informationen mitgesendet, also Uhrzeit, User Agent sowie andere Variablen, die den jeweiligen Umfang haben (siehe Abbildung 5.1 und Kapitel 7, »Dimensionen und Messwerte«).

## 5.2 Seitenaufrufe

Ein Hit kann, wie im vorangehenden Abschnitt 5.1, »Hits«, beschrieben, dadurch entstehen, dass ein Nutzer eine Seite aufruft. Wie in Abschnitt 3.4, »Server- und Client-basiertes Tracking«, ausgeführt, wird das Tracking-Skript beim Aufrufen der Seite zu einem definierten Zeitpunkt gestartet und sendet dann Informationen an den Tracking-Server, in diesem Fall, dass ein Seitenaufruf stattgefunden hat. Hits können nicht nur durch Seitenaufrufe oder durch nutzerinitiierte Ereignisse entstehen, sondern zum Beispiel auch durch definierte Zeitintervalle (siehe Kapitel 8, »Ereignisse und Datenschicht«).

Standardmäßig werden in den meisten Tools ab Werk nur Seitenaufrufe gemessen. Die Anzahl der Seitenaufrufe ist nur selten ein sinnvoller KPI, ebenso wenig wie die Hits, denn ein Aufruf sagt nichts darüber aus, ob der Inhalt der Seite auch angesehen wurde. Dennoch bilden Seitenaufrufe die nächste Grundeinheit oberhalb der Hits, denn Ereignisse werden häufig pro Seitenaufruf bewertet.

Dabei ist zu beachten, dass der Seitenaufruf zu verschiedenen Zeiten »gefeuert« werden kann, im Google Tag Manager zum Beispiel, sobald dieser Tag Manager geladen ist, wenn das DOM geladen ist<sup>1</sup> oder wenn die Seite komplett geladen und für den Nutzer sichtbar ist. Je nachdem, wie schnell eine Seite lädt, führt dies zu unterschiedlichen Zeitstempeln. Es kann aber auch erklären, warum manche Nutzer anscheinend gar nichts auf der Seite tun, da sie vielleicht einen Seitenaufruf erzeugt, die Seite aber, bevor sie überhaupt im Browser gerendert war, geschlossen hatten.

## 5.3 Sitzung

Eine Session beginnt, wenn ein Nutzer auf eine Website kommt und damit mindestens eine Seite aufruft. Wurde früher von einem »Besuch« gesprochen, wird heute meistens der Begriff *Session* verwendet, da auch Apps getrackt werden, wo es keine Besuche, aber sehr wohl Sessions gibt.

Abbildung 5.2 zeigt eine Sitzung eines Nutzers mit all den dazugehörigen Hits an. Jeder der folgenden Punkte ist ein Hit:

- ▶ Der Nutzer beginnt um 21:16 Uhr mit dem Seitenaufruf der Homepage.
- ▶ Der Nutzer scrollt auf 25 % der Seite herunter.
- ▶ Der Nutzer startet ein Video.
- ▶ Der Nutzer hat um 21:17 Uhr 25 % des Videos erreicht.
- ▶ Der Nutzer pausiert das Video nach insgesamt 27 Sekunden.

<sup>1</sup> Die Struktur der Seite ist im Browser komplett interpretiert, aber sie wird noch nicht dargestellt.

- ▶ Der Nutzer startet das nächste Video.
- ▶ Der Nutzer hat um 21:18 Uhr 25 % des Videos erreicht.
- ▶ Der Nutzer hat um 21:20 Uhr 50 % des Videos erreicht.
- ▶ So geht es weiter, bis der Nutzer um 21:35 Uhr die Website verlässt.

Time	Event
9:35 PM	Event Leave on Page
9:35 PM	Event Scroll on Page
9:21 PM	Event Pause on Video
9:20 PM	Event Reached 50% on Video
	Event category: Video
	Event action: Reached 50%
	Event label: Opera: King Arthur/Purcell- Cold Genius Song/Air du Fro
9:18 PM	Event Reached 25% on Video
9:17 PM	Event Start playing on Video
9:17 PM	Event Pause on Video
9:17 PM	Event Reached 25% on Video
9:16 PM	Event Start playing on Video
9:16 PM	Event Scroll on Page
9:16 PM	Viewed data > opinion /

Abbildung 5.2 Eine Sitzung mit einem Seitenaufruf und mehreren Ereignissen in Google Analytics

In diesem Beispiel wurden Ereignisse konfiguriert, um verschiedene Messungen der Nutzerinteraktionen durchzuführen; für jede dieser Messungen entsteht ein Hit. Der Nutzer hat sich nur eine Seite in dieser einen Sitzung angesehen, aber mit der Seite interagiert und somit mehrere Hits ausgelöst. In einer Standardinstallation wäre nur der Seitenaufruf gemessen worden; durch leichte Anpassungen sind sämtliche Interaktionen sichtbar. Anstatt eines Hits für den Seitenaufruf wurden 11 Hits ausgelöst.

## 5.4 Zeitliche Definition einer Sitzung

In den meisten Tools wird eine Sitzung oder Session so definiert, dass sie automatisch nach 30 Minuten Inaktivität des Benutzers als beendet gilt. Dieser Wert ist veränderbar. Der Timeout für die Inaktivität hat nichts mit der Sitzungsdauer selbst zu tun, sondern lediglich damit, ab wann ein erneutes Aufrufen einer Seite in die gleiche Sitzung fällt oder eine neue Sitzung auslöst. Ruft ein Nutzer auf einer Website innerhalb von 20 Minuten mehrere Seiten auf und macht dann eine Pause von 31 Minuten, beginnt bei einem neuen Seitenaufruf eine neue Sitzung, da der Nutzer 30 Minuten inaktiv war.

Das Konzept einer Sitzung soll anhand einer Nachrichtenseite verdeutlicht werden. Ein Benutzer besucht die Nachrichtenseite mehrmals täglich, zum Beispiel immer dann, wenn ihm langweilig ist oder er sich von einer kniffligen Arbeit ablenken will (zum Beispiel ein Buch über Webanalyse zu schreiben). Die Betreiber der Nachrichtenseite interessieren sich unter anderem für Seitenaufrufe, da diese Werbeeinblendungen und damit eine Finanzierung der Website ermöglichen. Eventuell existiert bei den Betreibern der Website ein KPI, wie Seiten pro Besuch mit Werbeeinblendungen, denn je mehr Seiten pro Besuch sich ein Benutzer ansieht, desto mehr Werbeeinblendungen werden erzeugt. Dafür werden zum Beispiel bei jedem Artikel weitere Artikel angezeigt, die für den Besucher interessant sein könnten.

Wird die Sitzungsdauer kurz angesetzt (10 Minuten), bedeutet eine kurze Pause von 11 Minuten beim Lesen, dass beim Aufrufen einer weiteren Seite eine neue Sitzung beginnt. Die Anzahl der Seiten pro Besuch ist damit reduziert. Wird die Sitzungsdauer zum Beispiel auf 60 Minuten festgelegt, kann ein prokrastinierender Autor durch zwei Besuche der Nachrichtenseite innerhalb von 45 Minuten nur einen Besuch erzeugen, obwohl er aus seiner Sicht zweimal dort war. In diesem Fall würde er mehr Seiten pro Besuch erzeugen.

## 5.5 Was ist die ideale Sitzungsdauer?

30 Minuten für eine Session sind ein Standard, der aber nicht immer sinnvoll ist. Beinhaltet eine Website Tutorials, kann ein Tutorial zum Beispiel erfordern, dass zwischen zwei Schritten eine Pause gemacht wird (um zum Beispiel etwas trocknen oder abkühlen zu lassen).

Dauert diese Pause länger als 30 Minuten, beginnt eine neue Session, obwohl es eigentlich noch dieselbe Session ist. Es ist also von vornherein notwendig, sich zu überlegen, wie Nutzer auf der Website agieren könnten und ob eine Sitzung durch längere Pausen unterbrochen werden könnte. Gleichzeitig muss diese Änderung unbedingt dokumentiert werden, da die Werte ansonsten viel Verwirrung stiften können.

Letztendlich sind die 30 Minuten Standard-Session-Timeout ein arbiträrer Wert. Ebenso könnten 60, 120, 240 oder einfach 42 Minuten gewählt werden.

## 5.6 Die Rolle des Nutzers

Zu guter Letzt soll noch einmal auf den Nutzer eingegangen werden. Wenn Hits die kleinste Einheit sind, dann sind Nutzer die größte Einheit, denn:

- ▶ Nutzer haben eine oder mehrere Sitzungen.
- ▶ In den Sitzungen werden eine oder mehrere Seiten aufgerufen.
- ▶ Auf den Seiten finden keine oder mehrere Interaktionen statt.
- ▶ Dies alles führt zu Hits.

Aber nicht jede Analyse stellt den Nutzer in den Vordergrund, sondern hin und wieder die folgenden Inhalte:

- ▶ Wie häufig wurde ein Video bis zum Schluss gesehen?
- ▶ Wie häufig wurde welcher Inhalt bis zum Ende durchgelesen?

Der Seitenaufruf als solcher muss also gar nicht die wichtigste Währung nach dem Nutzer sein. Insbesondere mit den zum Teil populären One Page Websites stellt sich die Frage, welche Rolle ein Seitenaufruf hier spielt. Gleichzeitig zeigen die Beispiele, dass aus dem Baustein Hit neue Schnitte aus den Daten möglich sind.

### Wichtig zu wissen

- ▶ Hits sind die kleinste Einheit in der Webanalyse; die meisten Webanalyse-Systeme berechnen Gebühren nach aufgerufenen Hits beziehungsweise schränken die Anzahl ausgewerteter Hits ein.
- ▶ Seitenaufrufe sind die populärsten Beispiele eines Hits, sagen aber in der Regel wenig über die Interaktion eines Nutzers mit einer Website aus.
- ▶ Durch zusätzliche Messungen können die Interaktionen eines Nutzers getrackt werden, was aber die Anzahl der Hits mit jedem Messpunkt nach oben treibt.
- ▶ Die Hierarchie der Daten besteht aus Hit, Seitenaufruf, Sitzung und Nutzer, wobei der Seitenaufruf eine geringere Rolle spielen kann. Auch können die Daten anders geschnitten werden, wenn es nicht um den Nutzer, sondern um die Inhalte geht.
- ▶ Der Ablauf einer Sitzung ist in den meisten Systemen auf 30 Minuten gesetzt; diese Zahl sollte aber für die jeweiligen eigenen Ziele überprüft und angepasst werden.

## Kapitel 25

# Die Kunst, das richtige Diagramm zu wählen

*Diagramme können die Informationsaufnahme für den Empfänger stark vereinfachen; allerdings ist die Auswahl des richtigen Diagrammtyps ein richtiges Handwerk.*

### 25.1 Vorsicht, Diagramm!

Viele Programme wie Excel, Tableau und die meisten Analytics-Tools bieten eine Vielzahl von Visualisierungsmöglichkeiten. Google bietet mit dem Data Studio ein eigenes Tool nur für das Erstellen von Berichten (Kertzel und Mylluks 2018). Gleichzeitig fehlt häufig das Wissen darüber, welches Diagramm am besten geeignet ist. Stattdessen besteht die Gefahr, dass einfach das genommen wird, was Excel und Co. zur Verfügung stellen. Eine gute Datenvisualisierung erfordert Gehirnschmalz, da sie dem Konsumenten von Informationen einen deutlichen Mehrwert bei der Vermittlung von Sachverhalten erbringen soll.

Auch hier ist es sinnvoll, nicht den Vorgaben eines Tools zu vertrauen. Ein Beispiel aus Google Analytics soll das verdeutlichen. Abbildung 25.1 zeigt ein Tortendiagramm, das viele Jahre zum Standard-Dashboard gehörte, das ein Nutzer nach dem Login sah. Es zeigt das Verhältnis von neuen und wiederkehrenden Nutzern innerhalb eines Zeitraums, meistens 4 Wochen.

Hier kommen gleich mehrere Fehler zusammen. So würde man bei einem Tortendiagramm davon ausgehen, dass Nutzer, die in einem Tortenstück sind, nicht gleichzeitig in dem anderen Tortenstück sein können. Allerdings kann ein Nutzer sowohl das erste Mal in einem Zeitraum auf eine Website kommen als auch diese im gleichen Zeitraum ein zweites Mal besuchen. Tatsächlich tauchen in Google Analytics solche Nutzer in beiden Tortenstücken auf, und das entspricht nicht der gewohnten Aussage eines Tortendiagramms.

Hinzu kommt, dass es das Ziel vieler Websites ist, die Anzahl neuer und wiederkehrender Nutzer zu erhöhen. Wird eine Marketing-Kampagne durchgeführt, ist in der Regel von einer Steigerung der Anzahl neuer Nutzer auszugehen (es sei denn, beste-

hende Nutzer sollen noch einmal angesprochen werden), und der Anteil der wiederkehrenden Nutzer erscheint kleiner, selbst wenn die absolute Zahl dieser Nutzer unverändert bleibt. Was nützt also das Verhältnis zueinander, wenn beide Zahlen sich unabhängig voneinander bewegen können? Schlimmer noch: Bei beiden Werten wird eine positive Veränderung erwartet, und Veränderung bedeutet, dass ein Tortendiagramm nicht sinnvoll ist. Denn Tortendiagramme werden für statische Verhältnisse verwendet, nicht für Entwicklungen, wie sie hier vorhanden sind.

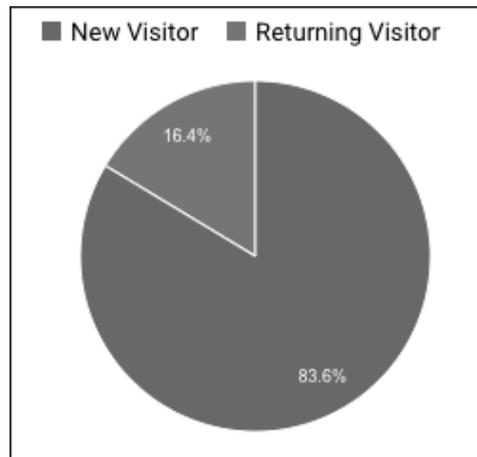


Abbildung 25.1 Tortendiagramm aus Google Analytics in einem Standard-Dashboard

Zu guter Letzt stellt sich die Frage, was mit einem solchen Tortendiagramm erreicht werden soll (siehe Kapitel 24, »Wie handlungsrelevante Berichte entstehen«). Welche Aktion soll abgeleitet werden?

Tatsächlich ergeben die beiden Zahlen unabhängig voneinander mehr Sinn, denn wenn weniger neue Nutzer auf die Seite kommen, ist es ein Problem des Marketings (sofern neue Nutzer das Ziel sind und nicht nur die Steigerung des CLV der bestehenden Nutzer). Wenn weniger Nutzer wiederkehren, sind es die Inhalte, die optimiert werden müssen, es sei denn, die Nutzer konnten auf der Website etwas final abschließen, sodass sie nicht mehr wiederkommen müssen. Wenn die Zahlen unterschiedliche Maßnahmen erfordern, sollten sie auch getrennt berichtet werden, zumal sie im Verhältnis zueinander nicht sinnvoll interpretiert werden können.

In den folgenden Beispielen werden verschiedene Diagrammtypen und ihre geeignete Nutzung erläutert. Diese Beispiele sind lediglich eine Auswahl von Diagrammen ohne Anspruch auf Vollständigkeit. Für eine weiterführende Beschäftigung mit diesem Thema werden Few 2013, Healy 2019 und Wilke 2019 empfohlen.

## 25.2 Welches Diagramm für was?

Diagramme können in vier unterschiedliche Typen unterteilt werden:

- ▶ **Vergleiche**
  - Welche Gruppe hat mehr oder weniger Items?
  - Wie hat sich eine Gruppe über die Zeit verändert?
- ▶ **Zusammensetzungen**
  - Aus welchen Komponenten besteht eine Menge, und wie hoch sind die jeweiligen Anteile?
- ▶ **Beziehungen**
  - Was haben verschiedene Werte miteinander zu tun?
- ▶ **Verteilungen**
  - Haben wir eine Normalverteilung? Oder eine andere Verteilung?

Darüber hinaus wird zwischen statischen und sich verändernden beziehungsweise einer oder mehreren Variablen unterschieden. Manchmal sind die Grenzen nicht ganz eindeutig.

## 25.3 Liniendiagramm

Liniendiagramme sind vor allem für Vergleiche über einen Zeitraum spannend, um eine Entwicklung eines oder mehrerer Werte zu beobachten.

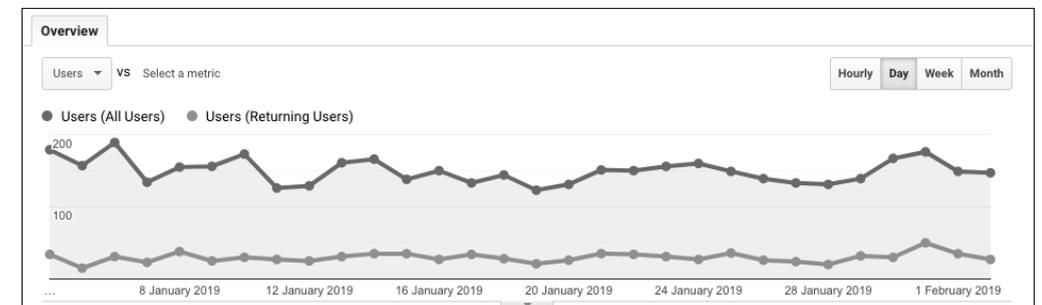


Abbildung 25.2 Liniendiagramm

In dem Diagramm in Abbildung 25.2 werden die Daten aus dem Tortendiagramm aus Abbildung 25.1 auf ein Liniendiagramm übertragen. Deutlich wird in dieser Visualisierung, dass die Ratio von neuen zu wiederkehrenden Nutzern nicht immer parallel läuft, aber auch, dass mehr neue Nutzer nicht nachhaltig für zusätzliche wiederkehrende Nutzer sorgen.

## 25.4 Säulendiagramm/Balkendiagramm

Das Säulendiagramm ist eines der einfachsten Diagramme und wird vor allem dazu verwendet, einen Vergleich zwischen Elementen zu ermöglichen, wenn nur eine Variable pro Element existiert.

Ein Säulendiagramm sollte nicht mit einem Histogramm verwechselt werden (siehe Abschnitt 25.5). In dem Beispiel in Abbildung 25.3 werden Sessions und Nutzer über Geräteklassen miteinander verglichen. Aus diesem Chart kann interpretiert werden, dass Desktop-Nutzer mehr Sessions haben als Tablet-Nutzer. Das Säulendiagramm ist hier wahrscheinlich die effektivste Darstellung dieses Unterschieds, da er so am schnellsten wahrgenommen werden kann.

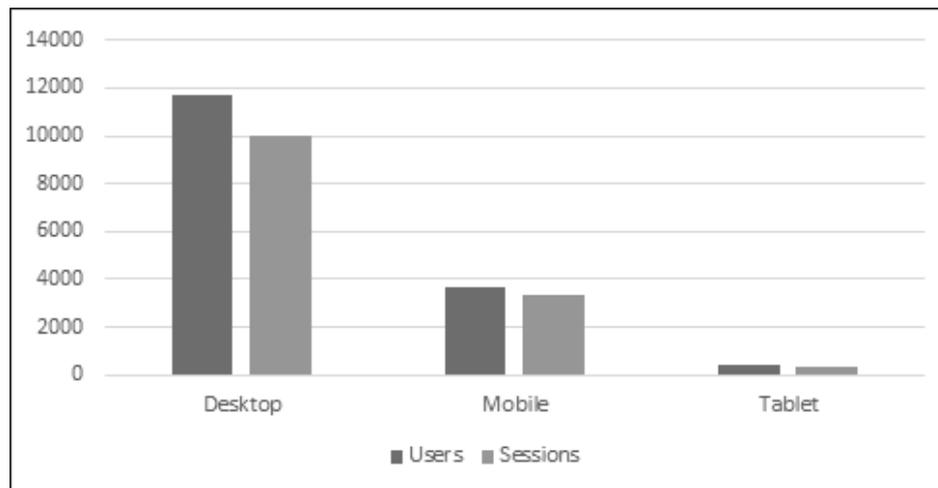


Abbildung 25.3 Säulendiagramm

Es existieren drei populäre Variationen des Säulendiagramms: das Balkendiagramm, das gruppierte Säulendiagramm und das gestapelte Säulendiagramm. Das Balkendiagramm wird vor allem dann verwendet, wenn viele Elemente angezeigt werden sollen, da die Säulen in die Waagerechte gelegt werden. Gestapelte Säulendiagramme werden verwendet, wenn die Zusammensetzung einer Gruppe ersichtlich sein muss. Gruppierte Säulendiagramme zeigen mehrere Kategorien in einer Gruppe nebeneinander, und es werden mehrere Gruppen miteinander verglichen. Abbildung 25.3 zeigt ein gruppiertes Säulendiagramm.

## 25.5 Histogramm

Ein Histogramm (siehe Abbildung 25.4) zeigt eine Häufigkeitsverteilung von Merkmalsausprägungen an. Dieses Histogramm wurde bereits in Abschnitt 13.3, »The mean Mean«, erwähnt, als es darum ging, die Verteilung von Seiten pro Besuch zu visualisieren.

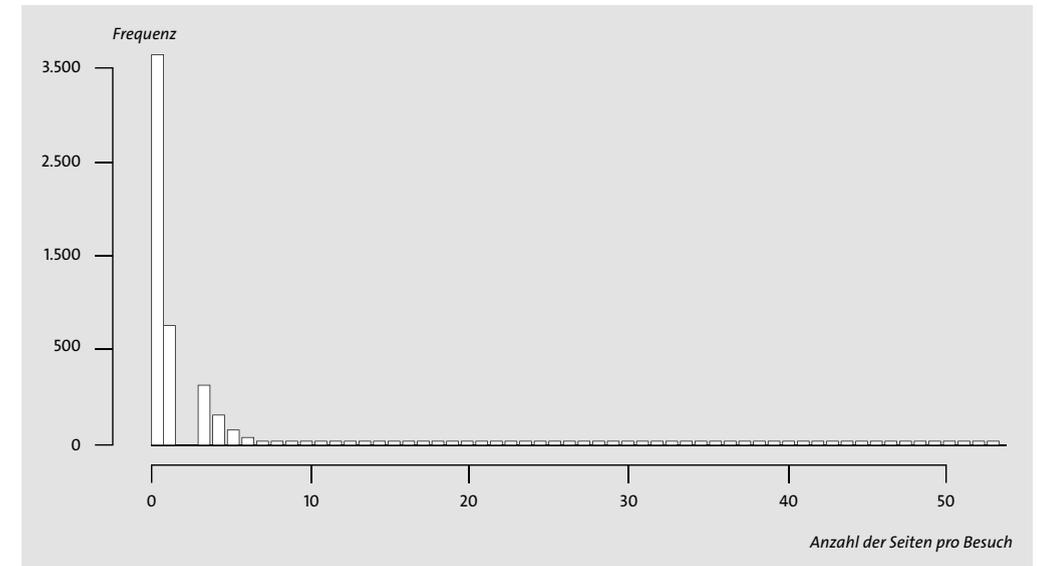


Abbildung 25.4 Histogramm

Im Prinzip ist ein Histogramm nichts anderes als die Visualisierung einer Häufigkeitstabelle. In einer solchen Tabelle wird für jedes Auftauchen eines Werts (wie oft wurde nur eine Seite angesehen? wie oft wurden zwei Seiten angesehen? usw.) sozusagen ein Strich gemacht und zum Schluss dann die Striche zusammengezählt.

Ein Histogramm ist dann sinnvoll, wenn die Verteilung einer Variablen mit wenigen Datenpunkten visualisiert werden soll. Mitunter müssen Ausreißer aus der Darstellung exkludiert werden, da ansonsten die Verteilung der häufigsten Fälle nicht mehr sichtbar ist.

## 25.6 Scatterplot

Ein Scatterplot ist dann sinnvoll, wenn zwei oder mehr Merkmale und ihr Bezug zueinander visualisiert werden sollen. In dem Beispiel in Abbildung 25.5 ist eine Scatterplot-Matrix zu sehen, in der mehrere Variablen wie in einer Kreuztabelle in Bezug zueinander gesetzt werden. Sehr schön ist hier zum Beispiel die Beziehung von Nutzern und Sessions zu sehen.

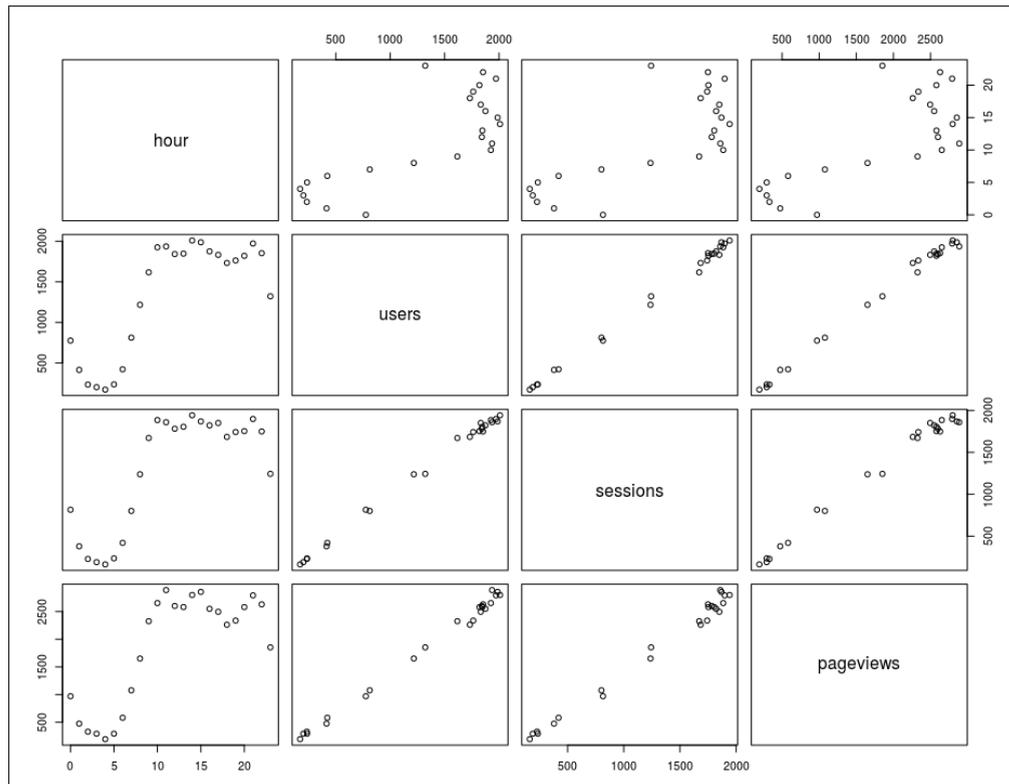


Abbildung 25.5 Scatterplot-Matrix

Ein Scatterplot ist ideal für eine erste Visualisierung von Daten, zum Beispiel um Korrelationen zu erkennen.

## 25.7 Bubble Chart

Ein Bubble Chart ist eine Art von Scatterplot und kann eine Verbindung von Merkmalen visualisieren; nur sind hier drei Variablen vorhanden (nähme man nun noch eine Farbtiefe hinzu, wäre man bei vier Variablen). Das Beispiel in Abbildung 25.6 zeigt die variablen Seitenaufrufe (Y-Achse), Sessions (X-Achse) und User (Größe der Bubble). Es ist zu erwarten, dass mehr Nutzer mehr Sessions generieren und damit auch mehr Page Views. Interessant ist in diesem Beispiel, dass die Nutzer aus Berlin leicht unter der Erwartung liegen, die Nutzer aus Hamburg und München hingegen weit darüber.

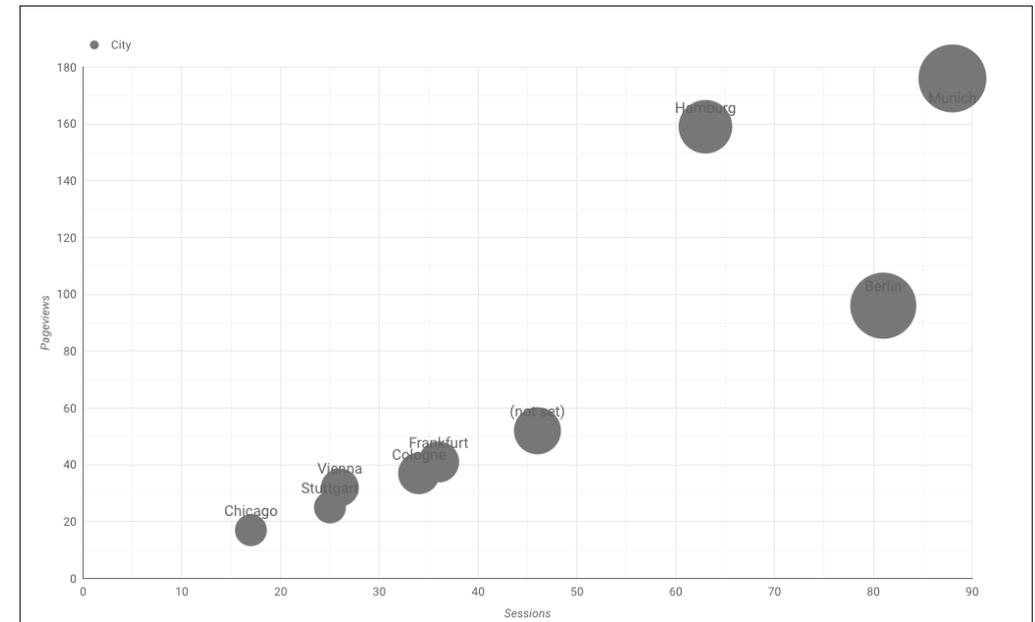


Abbildung 25.6 Bubble Chart

Auch wenn Bubble Charts oft gut aussehen, wird die Informationsvermittlung an den Leser damit nicht unbedingt beschleunigt. Der Betrachter muss mehrere Dimensionen gleichzeitig interpretieren, sodass er nicht selten zurück auf die Legende schauen muss, was zum Beispiel Größe und Farbtintensität in dem Diagramm aussagen. Die Verwendung eines Bubble Charts ist also nur mit Vorsicht zu empfehlen. Allerdings kann in einer interaktiven Grafik die Verwendung eines Bubble Charts aufschlussreich sein, wenn sich mehrere Werte gleichzeitig ändern und der Benutzer dies spielerisch erkunden kann.

## 25.8 Boxplot-Diagramm

Ein Boxplot-Diagramm, wie es in Abbildung 25.7 dargestellt ist, ist eher für Experten geeignet, bietet dafür aber eine Vielzahl von Informationen in einer Grafik. Die rechteckige Box in jedem Jahr zeigt an, wo sich 50 % der Daten befinden, der dicke Querstrich in der Mitte der Box bildet den Median und die Breite der Box die Wurzel aus der Sample-Menge ab. Die gestrichelten Linien zeigen die Streuung an.

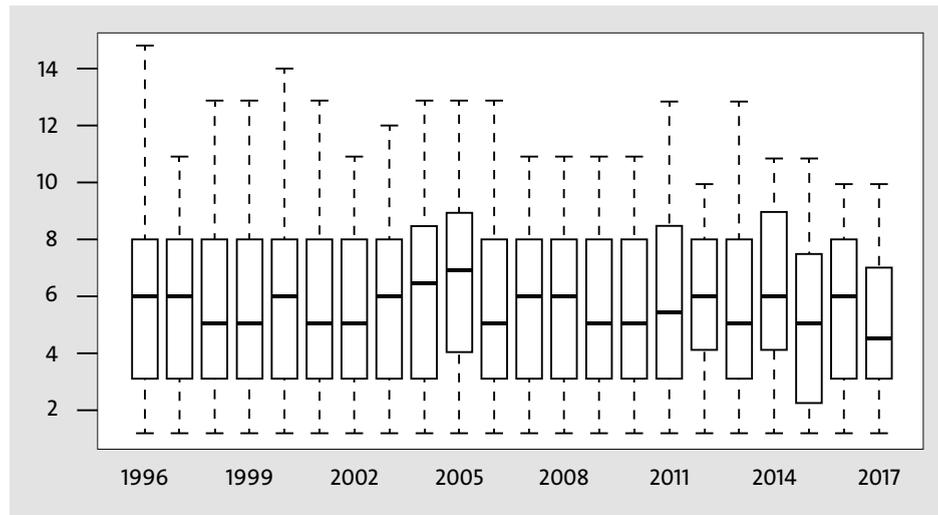


Abbildung 25.7 Boxplot-Diagramm

In dem abgebildeten Beispiel wird die Verteilung der Platzierungen von Domains in den Suchergebnissen von Google, aufgesplittet nach Alter, angezeigt. Eine Suchergebnisseite kann mehr als 10 Ergebnisse haben, deswegen zeigt die Y-Achse 14 an; ein Wert in der Grafik geht sogar bis 15. Unter Suchmaschinen-Optimierern wurde lange das Gerücht gepflegt, dass ältere Domains einen Ranking-Vorteil bei Google hätten. Tatsächlich ist in den Daten zu sehen, dass ältere Domains scheinbar keinen Vorteil haben, es allerdings weniger Domains in den Jahren 2014 bis 2017 auf die erste Suchergebnisseite geschafft haben (die Erhebung fand im Frühjahr 2017 statt, sodass die 2017er-Zahl mit Vorsicht zu genießen ist).

#### Wichtig zu wissen

- ▶ Die Auswahl eines Diagramms erfordert Sorgfalt, um zum einen den Daten, zum anderen aber auch der Intention eines Berichts gerecht zu werden.
- ▶ Auch den Diagrammen der Webanalyse-Tools ist grundsätzlich erst einmal zu misstrauen, da nicht davon ausgegangen werden kann, dass die Diagramme sinnvoll ausgewählt wurden.