

Visual Analytics – Buzzword oder Innovation?

Jens Rauch, Informationsfabrik GmbH

Im Umfeld von Business Intelligence stößt man regelmäßig auf neue Begriffe, deren Sinnhaftigkeit oder genaue Bedeutung sich auf den ersten Blick nicht erschließt. „Visual Analytics“ gehört zu diesen Begriffen und scheint zunächst auf neue Techniken und Möglichkeiten zu verweisen, um Daten zu visualisieren, die schließlich innovative Ansätze der Daten-Analyse eröffnen. Wirft man jedoch einen kritischen Blick auf BI-Software, die mit Visual Analytics für sich wirbt, stellt sich schnell die Frage, was an Balken- und Linien-Diagrammen innovativ sein soll. Tatsächlich besitzt das Konzept eine solide theoretische Basis und liefert Methoden, die Analyse, Verständnis und Kommunikation von Daten für Menschen vereinfachen.

Aktuell sehen sich viele Unternehmen vor der Herausforderung, das unaufhaltsame Datenwachstum in den Griff zu bekommen. Cloud-Lösungen und neue Technologien für verteilte Systeme bieten Möglichkeiten, Big Data zu verarbeiten. Jedoch besteht nach wie vor viel Unsicherheit bei der Frage, wie große und unstrukturierte Datenmengen analytisch zu handhaben sind. Bildlich gesprochen geht es darum, die in den Daten versteckten „Schätze“ zu heben. Dies können wertvolle Einsichten über das Kaufverhalten von Kundengruppen oder Produktionsparameter für Fehlerquoten in der Fertigung sein. Visual Analytics ist neben Data Mining und Predictive Analytics als ein Lösungsansatz für dieses Problem vielerorts im Gespräch.

Ein bloßes Marketingwort?

Will man zum Stichwort „Visual Analytics“ mehr Informationen einholen, ergibt sich in der Mehrheit der Publikationen nur ein vages Bild: Einige Beiträge zum Thema sind bloße Sammlungen wohlmeinender Ratschläge, wie ein Diagramm am besten zu gestalten ist. Andere versuchen sich an einer Systematik, die Auskunft gibt, wann welcher Diagrammtyp einzusetzen ist. Wiederum andere postulieren eine Revolution der Daten-Analyse durch neue Visualisierungstools, bleiben jedoch wirklich überzeugende Beispiele schuldig. In allen drei Fällen ist nicht erkennbar, worin die innovative Kraft von Visual Analytics liegen soll: Was ist daran neu,

Daten zu visualisieren oder Diagramme leserfreundlich aufzubereiten? Warum fällt es so schwer, die neuen Möglichkeiten aktueller Visualisierungstools konkret zu benennen? Es scheint nicht einmal festzustehen, was Visual Analytics selbst eigentlich ist: Eine Methode? Eine Technologie? Ein Fachgebiet?

Schnell liegt der Verdacht nahe, dass es sich bei Visual Analytics um ein sogenanntes „Buzzword“ handelt. Damit bezeichnet man Begriffe mit unklarer Bedeutung, die zu Marketing-Zwecken verwendet werden, weil sie gut klingen und Innovationsgeist ausstrahlen. Als Beispiele könnte man „Web 2.0“, „agil“ oder auch „Big Data“ nennen.

Eine bemerkenswerte Eigenschaft von Buzzwords ist, dass niemand das Gegenteil von ihnen tun oder sein möchte. Wer würde sich rühmen, „Small Data“ zu bewirtschaften oder ein „starres“ Vorgehensmodell einzusetzen? Nicht selten werden Buzzwords eingesetzt, um einer alten Sache einen neuen Anstrich zu verleihen. Denn es lässt sich angesichts der unklaren Bedeutung ja kaum widerlegen, dass der Begriff auf das Bestehende zutrifft. Dergleichen geschieht in so manch einer Entwicklungsabteilung, die plötzlich „agil“ arbeitet und somit einen Ausweg gefunden hat, fehlende Dokumentation und unstrukturierte Arbeitsweise als innovatives Vorgehen positiv umzudeuten. Ist auch Visual Analytics nur alter Wein in neuen Schläuchen?

Die Verwendung des Begriffs ist entscheidend

Ob man es mit einem Buzzword zu tun hat, hängt tatsächlich weniger mit dem Begriff selbst zusammen, sondern damit, wie er verwendet wird. Beispielsweise ist der Begriff „agil“ in Zusammenhang mit Vorgehensmodellen im „Agile Manifesto“ sehr genau definiert worden und damit alles andere als unklar. Um nachvollziehen zu können, wie Buzzwords entstehen, muss man sich vergegenwärtigen, wie sich aus neuen Trends und Strömungen handfeste Innovationen herausbilden: Es gibt über einen längeren Zeitraum einen Trend von Praktiken oder Technologien, die einem bestimmten Muster folgen.

Beim Web 2.0 waren das zum Beispiel Wikipedia, Blogs und Google Mail. Schließlich findet jemand einen passenden Begriff dafür und es werden Aufsätze und Manifeste verfasst. So geschehen beim „Agile Manifesto“ und bei Aufsätzen zu „Web as a platform“. Für Visual Analytics gibt es das im Übrigen gleich in mehrfacher Form. Als umfängliche Einstiegsreferenz eignet sich der kürzlich erschienene Sammelband „Expanding the Frontiers of Visual Analytics and Visualization“ von Dill et al. (2012). Wenn begriffsprägenden Publikationen dieser Art genau den Zeitgeist treffen, werden sie viral; es wird über den Begriff gebloggt, getwittert, gesprochen und publiziert. Dabei passiert es leicht, dass der Bezug zum ursprünglichen Konzept verloren geht, und es resultiert ein Buzzword.

Letztlich werden Begriffe zu Buzzwords, wenn sie ohne existierende Definition oder genaue Kenntnis ihrer eigentlichen Bedeutung verwendet werden. Eine solche Verwendung zielt zumeist nur darauf ab, Eindruck zu machen, und ist einem der Sache angemessenen Diskurs nicht dienlich.

Die Leitidee hinter Visual Analytics

Um nun den Begriff „Visual Analytics“ wieder zu erden, klar einzugrenzen und auf seine innovativen Aspekte zu verweisen, ein Blick auf das dahinter liegende Konzept. Was leicht in der Diskussion um Daten-Analyse und Big Data übersehen wird, ist die Tatsache, dass der Mensch in allen Fällen die letzte Instanz darstellt, die nach geschehener Auswertung die Ergebnisse

interpretieren und entscheiden muss. Gilt es, zusehends komplexe und unübersichtliche Datenmengen zu handhaben, darf der Übergang von Information (Maschine) zu Wissen (Mensch) nicht außer Acht gelassen werden. Mit Visual Analytics wird diese Schnittstelle in Angriff genommen und optimiert.

Festzuhalten ist, dass Daten in graphischer Form (Visualisierungen) vom menschlichen Wahrnehmungsapparat erheblich besser verarbeitet werden als maschinelle Daten in numerischer, tabellarischer Repräsentation. Dieser Effekt beschränkt sich nicht nur auf Wertevergleiche (Balken A ist länger als Balken B), sondern erstreckt sich ebenso auf komplexe Muster und Zusammenhänge (Clusterbildung, nicht-lineare Abhängigkeiten). Im Hinblick darauf kann das menschliche

Auge zu modernen statistischen Algorithmen der Muster-Erkennung als ebenbürtig angesehen werden. Diese Erkenntnis ist zugegeben nicht neu, wurde aber in der Vergangenheit insbesondere in BI-Lösungen stark vernachlässigt. Der Grund ist, dass bislang vorrangig gefordert war, Werte exakt darzustellen.

Wird ein Wert graphisch dargestellt (zum Beispiel codiert als Abstand zur Achse oder als Fläche), dann kann er schneller verarbeitet werden, verliert aber gegenüber der numerischen Darstellung an Genauigkeit. Erst seitdem durch unbegrenztes Datenwachstum und Big-Data-Anwendungsfälle in den Vordergrund rücken, bei denen kaum noch einzelne Werte, sondern Wertzusammenhänge und -veränderungen interessieren, gewinnt auch die Visualisierung an Aufmerksamkeit.

Visual Analytics als konzeptionelles Framework

Visual Analytics lässt sich vor diesem Hintergrund als konzeptionelles Framework definieren, mit dem der Einsatz und Nutzen von Daten-Visualisierungen optimiert wird. Ein optimierter Einsatz bedeutet hierbei, dass Datenvisualisierungen durch Visual Analytics in viel mehr Bereichen und Anwendungsfällen als bislang eingesetzt werden. Ein optimierter Nutzen bedeutet, dass die inhaltliche Qualität, Lesbarkeit und Aussagekraft von Datenvisualisierungen erhöht wird.

Zunächst ist Visual Analytics also keine klar umgrenzte Disziplin oder Technologie, sondern eine Methodensammlung, die sich auf oberster Abstraktionsebene in drei Bausteine unterteilen lässt: Daten-Analyse, Information Design und Tools. Sie gliedern Visual Analytics in Fachgebiete, die jeweils für sich bearbeitet oder vorangetrieben werden können. Wenn sich die Frage nach dem innovativen Potenzial von Visual Analytics stellt, so ist auf die Leistungen in diesen Teilbereichen zu verweisen. Daher soll im Folgenden jeder Baustein kurz anschaulich im Hinblick auf neuere Entwicklungen beleuchtet werden, die zum Potenzial des Gesamtkonzepts beitragen.

Daten-Analyse und Visualisierungen gehen Hand in Hand

Das wesentliche Versprechen von Visual Analytics ist, dass der Einsatz von Visualisierungen die Daten-Analysen stark ver-

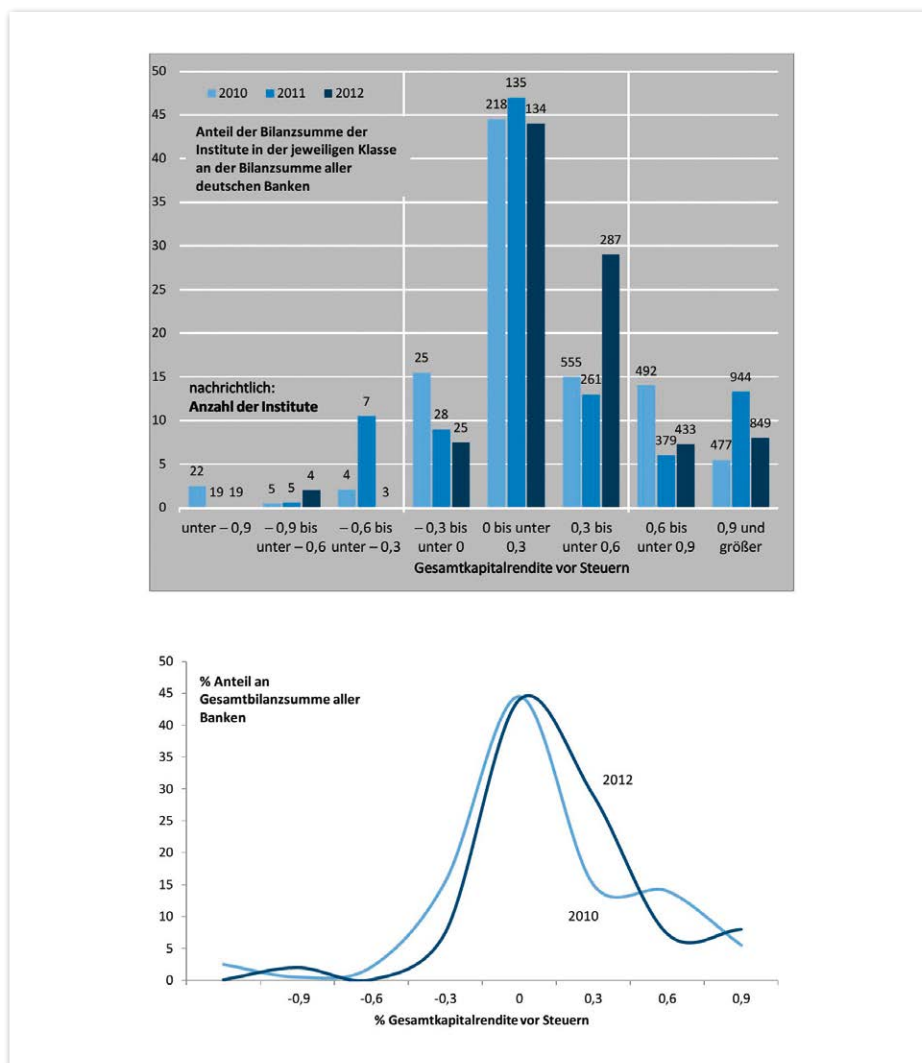


Abbildung 1: Wie visualisiert man, wie sich die Verteilung der Gesamtkapital-Rendite im deutschen Bankensystem verändert hat? Die untere Abbildung folgt den Prinzipien des Information Design

einfacht. Einige Anbieter von BI-Tools behaupten gern, dass es an der Zeit sei, die Daten von unnötig komplizierten Statistiken und Algorithmen zu befreien. Sobald die Daten visualisiert seien, lägen sie für jeden sichtbar da. Demnach könne bald jeder Mitarbeiter eines Unternehmens zum Data Scientist werden und sich selbst mit den benötigten Informationen versorgen. Bedauerlicherweise genügt ein leicht bedienbares Visualisierungs-Tool für diesen Anspruch nicht.

Selbst wenn Anwender einfachste Standard-Diagramme mit einem überschaubaren Datensatz erstellen, sind datenanalytische Grundkenntnisse erforderlich. Sie müssen zum Beispiel wissen, in welchen Fällen der Median dem arithmetischen Mittel vorzuziehen ist oder wie sich Varianz definiert. Vor allem müssen sie erkennen können, welche analytischen Fragestellungen sich überhaupt mit einem Datensatz sinnvoll beantworten lassen. Diese Aufgaben kann eine Software dem Anwender nicht abnehmen.

Man sollte auch nicht versucht sein zu glauben, dass Statistik oder Data Mining sich durch Daten-Visualisierungen ersetzen ließen. Zwar gibt es viele Fälle, in denen bereits einfache Visualisierungen erstaunliche Einsichten in Daten geben und zu denen man andernfalls nur durch mehrfache statistische Explorations-Tests gelangt. Allerdings müssen diese Zusammenhänge in den Daten auch so vorliegen, dass sie in der Visualisierung ablesbar sind. Sind sie das nicht, ist es unvermeidlich, auf Statistik zurückzugreifen.

Wohlgemerkt sollte jedoch jede statistische Kennzahl möglichst wieder visualisiert werden (etwa ein Regressionskoeffizient durch eine Gerade), da sie dann für Menschen besser interpretierbar ist. Daten-Analyse und Visualisierung müssen also Hand in Hand gehen. Damit umfasst der erste Baustein datenanalytische Grundkenntnisse und -fähigkeiten, Techniken der visuellen Analyse sowie statistische Verfahren und Algorithmen.

Information Design

Während der erste Baustein den Einsatz von Visualisierungen zur Analyse von Daten abdeckt, umfasst der zweite die Frage nach der geeigneten visuellen Darstellung. Dies hat maßgeblichen Einfluss darauf, wie gut lesbar und interpretierbar eine Visualisierung ist, und bestimmt damit nicht zuletzt den Erfolg einer Visualisierung als Kommunikationsinstrument. Das Information Design bemüht sich hierbei darum, die Reichhaltigkeit einer Visualisierung optimal auszuschöpfen. Das heißt, dass eine Visualisierung gemessen an den dargestellten Daten weder zu kompliziert noch zu einfach sein darf und dass die Form der Darstellung der zu beantwortenden Fragestellung entspricht.

Information Design gibt zur Visualisierung Richtlinien an die Hand, wie Daten in welchem Fall visuell codiert werden (Länge, Winkel, Form, Farbe etc.) und welche grafischen Hilfselemente ungeeignet oder überflüssig sind. Typen von Visualisierungen können als Design Patterns



Specialized
Oracle Database



Datenbanken mit iQ



(Entwurfsmuster) für konkrete Arten von Fragestellungen (wie Visualisierungen für Zeitreihen-Vergleiche) zusammengefasst werden. *Abbildung 1* zeigt ein Beispiel für gelungene Chart-Optimierung mithilfe von Information Design.

Usability vs. Customizability beim Entwurf

Ein verbreiteter und effizienter Einsatz von Visualisierungen ist natürlich nicht ohne entsprechende Tools denkbar. Interaktivität von Visualisierungen und eine verbesserte Usability während des Entwurfsprozesses mit BI-, Statistik- oder reinen Visualisierungs-Tools sind aktuelle Themen. Leitend ist hier vor allem der Gedanke, die technischen Hürden für Anwender zu senken, die Visualisierungen erstellen möchten. Dadurch wird das visuelle Sichten und Analysieren von Daten einem breiteren Publikum, dem der informationstechnische Hintergrund fehlt, zugänglich gemacht.

Für Anwender ist es wünschenswert, dass ihnen möglichst Tool-unterstützt die passende Visualisierung zu vorgegebenen Datenmodellen vorgeschlagen wird. Das Bedienen und das Browsen in Daten sollen intuitiv erfolgen, große Datenmengen möglichst live in Visualisierungen überführbar sein. Dass statistische Grund-

kenntnisse für eine belastbare Interpretation solcher Visualisierungen dennoch nicht fehlen dürfen, soll noch einmal hervorgehoben werden.

Aber auch ein zum Tool-unterstützten Entwurf von Standard-Visualisierungen gegenläufiger Trend ist zu verzeichnen: Ausgangspunkt bildet die von L. Wilkinson veröffentlichte „Grammar of Graphics“, eine Spezifikationsprache, mit der sich beliebige Visualisierungen entwerfen lassen. Dadurch müssen Autoren von Datenvisualisierungen nicht länger auf Standard-Diagramme zurückgreifen und können stattdessen nach dem Baukastenprinzip individuelle, auf die Daten zugeschnittene Visualisierungen erzeugen.

In den vergangenen Jahren wurde Wilkinsons Beschreibungssprache auch von Herstellern von BI- und Statistik-Software aufgegriffen. So gibt es etwa für die Open-Source-Statistik-Software R das Paket „ggplot2“ und für die IBM Cognos BI Suite die „RAVE“-Technologie. Beide implementieren Wilkinsons formale Visualisierungssprache und bieten damit in Bezug auf Visualisierungen eine echte technische Innovation.

Fazit

Der technologische Fortschritt der Data Sciences stellt nicht nur neue Anforderungen an Hard- und Software, sondern auch

an die Menschen, die mit diesen Technologien arbeiten. Statistische Grundkenntnisse sowie Techniken des Entwurfs, der Analyse und Kommunikation von Daten mit Visualisierungen werden in Zukunft immer wichtiger. Visual Analytics als konzeptionelles Framework vereint die dazu erforderlichen Teilbereiche von Datenanalyse, Information Design und Tool-Know-how. Entsprechend geschulten Anwendern werden so die Methoden und Kenntnisse an die Hand gegeben, den neuen Herausforderungen im Umgang mit Daten gerecht zu werden.



Jens Rauch
jrauch@informationsfabrik.com

Die DOAG hat zwei neue Communities

Die Delegiertenversammlung der DOAG hat die Gründung einer BI Community sowie einer Java Community beschlossen. 42 stimmberechtigte Delegierte kamen dafür am Rande der DOAG Konferenz + Ausstellung zu einer außerordentlichen Sitzung zusammen. Im Zuge dessen wurde auch das Budget 2015 der DOAG e.V. abgesegnet.

Die Themen „BI“ und „Java“ waren bisher in der Development Community angesiedelt. Deren Leiter Robert Szilinski erläutert die Motivation für die Neugründung: „Es gab in unserer Community zu viele Themen, die wir nicht alle optimal abdecken konnten. Da laut

Satzung die Communities nicht beliebig groß werden dürfen, haben wir dem DOAG-Vorstand die Gründung einer BI Community sowie einer Java Community vorgeschlagen.“

Robert Szilinski sieht aufgrund von Technologie-Treibern wie Big Data ein immenses Potenzial für die neue BI Community. Für die Java Community hingegen spricht neben der riesigen Zielgruppe auch die Sichtbarkeit an den Hochschulen, was Wachstums-Chancen für die DOAG bei den jungen Leuten verspricht.

Die Delegiertenversammlung diskutierte ausgiebig über die beiden Vorschläge und stimmte dann mit überwältigender

Mehrheit der Neugründung zu. Anschließend wurde einstimmig das Budget 2015 der DOAG e.V. verabschiedet. Auch der Vorschlag, Frank Stöcker, Rolf Scheuch und Martin Klier als Wahlvorbereitungsausschuss für die kommende Vorstandswahl einzusetzen, wurde einstimmig angenommen.

Dr. Dietmar Neugebauer, Vorstandsvorsitzender der DOAG, zieht ein positives Fazit: „Mit den beiden neuen Communities ist die DOAG für die Zukunft bestens aufgestellt. Ich gehe davon aus, dass sowohl die BI Community als auch die Java Community bald ihre Arbeit aufnehmen werden.“