

DeltaMaster clicks!

02/2012

Liebe Datenanalysten,

seit über 5 Jahren ist das Blog von Bella, unserem Bürohund, ein Quell der Inspiration für alle, die nach Beispielen für den richtigen Umgang mit Daten und für gute Visualisierung suchen. Weit über 100 Artikel sind bereits erschienen. Meist sind es Darstellungen aus den Medien, an denen Bella ihre Beobachtungen macht, aus denen sie ihre Schlussfolgerungen zieht und Regeln ableitet. Diese Beispiele sind echt – und haben es deshalb immer wieder in sich. Eine gefällige Darstellung ist schnell erzeugt, wenn man sich auch die Daten dafür zurechtgelegt hat. Echte Daten sind weniger zahm. Ausreißer, Lücken, Sprünge, Niveauunterschiede machen uns die Arbeit schwer. In diesen *DeltaMaster clicks!* nehmen wir Sie mit in unsere Blog-Redaktion und zeigen, womit wir Bella helfen, wenn sie einen neuen Fall entdeckt hat. Natürlich arbeiten wir dabei mit *DeltaMaster*, bereiten Daten auf, analysieren sie, probieren Varianten aus und tun, was nötig ist, damit wir andere verständlich informieren können. Und immer wieder freuen wir uns, dass die Methoden und Funktionen von *DeltaMaster* so verzahnt sind, dass wir uns frei in den Daten umschauen können. Dieses Mal gehen wir damit ein häufiges Problem an: die Darstellung mehrerer Zeitreihen. Zugleich ist unser Themenrundgang über den Umgang mit Zeitreihen ein Beispiel, wie Sie sich auch in unbekanntem Terrain auf die zahlreichen Werkzeuge von *DeltaMaster* verlassen können. Kommen Sie mit!

Herzliche Grüße
Ihr Team von Bissantz & Company

DeltaMaster für den Handel EuroCIS ab 28. Februar 2012, Düsseldorf

Gemeinsam mit unserem Partner Ritter aus Mexiko stellen wir im Rahmen der EuroShop aus – das ist die wohl wichtigste Messe für den Handel.
www.eurocis.com

Bissantz Campus Nächster Seminarblock: 5. bis 14. März 2012

Unser Schulungsangebot für *DeltaMaster* und Microsoft SQL Server/Analysis Services.
www.bissantz.de/campus
www.bissantz.de/olap-seminar

DeltaMaster-Matinee 21. März 2012, Zürich

Wir freuen uns auf Andreas Palmer, den Leiter Information Systems der AMAG. Er wird präsentieren, wie sich *DeltaMaster* von einer Lösung für einzelne Fachabteilungen zu einem Instrument für die strategische Unternehmenssteuerung entwickelt hat.
www.bissantz.de/matinee

DeltaMaster@Work 29. März 2012, Nürnberg

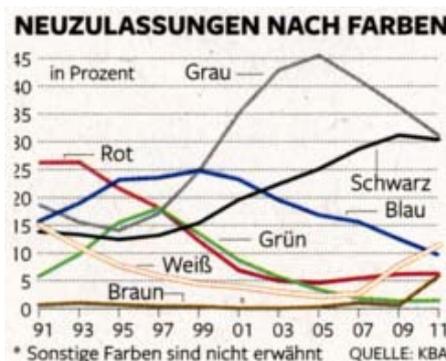
Der Februar-Termin ist bereits ausgebucht.
www.bissantz.de/dm@w

Archiv

Aktuelle und alle früheren *DeltaMaster clicks!*:
www.bissantz.de/clicks

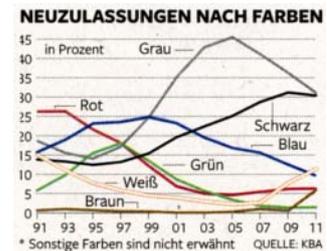
Kacheln statt Spaghetti

Um dieses Beispiel geht es: eine Statistik der Pkw-Neuzulassungen nach Farben, gefunden in der Welt vom 27. November 2011, Seite 99.
www.bella-beraet.de/ueberlagerte-zeitreihen



Kniff des Monats Zeichnen der Zeit – Varianten von Zeitreihendarstellungen

Schauen wir uns das Liniendiagramm aus der Welt vom 27. November 2011 genauer an. Es zeigt, wie sich die Beliebtheit von Autofarben im Laufe der letzten 20 Jahre verändert hat. Sieben Farben sind dargestellt, jeweils als prozentualer Anteil an allen Neuzulassungen im betreffenden Jahr. Wie wirkt die Grafik auf Sie?



„Ganz schön spaghetti“, schreibt Bella in ihrem Blog. Das Durcheinander der Linien ist schwer zu lesen und schwer zu verstehen, für die Zahlenwerte ist kein Platz, die Beschriftungen geistern darin herum. Solche Datenkonstellationen finden wir auch im Controlling immer wieder: Mehrere zeitliche Entwicklungen sind zu berichten und zu analysieren, seien es Marktforschungsdaten (wie hier), Umsatzzahlen aus verschiedenen Regionen oder der EBIT der Konzerngesellschaften.

Wir wollen das Beispiel aus Bellas Blog aufgreifen und diskutieren, wie man zeitlichen Entwicklungen mit den Standardwerkzeugen von *DeltaMaster* zu Leibe rückt.

Vorarbeiten: Daten beschaffen und würfeln

Die Daten für die Grafik auf der Titelseite bekommt man beim Kraftfahrt-Bundesamt (www.kba.de). Wir haben sie in eine Excel-Datei kopiert und daraus mit dem *DeltaMaster TableWizard* und dem *DeltaMaster CubeWizard* eine lokale Cube-Datei erzeugt (siehe *DeltaMaster clicks!* 04/2010).

Das Analysemodell besteht aus zwei Dimensionen (den Jahren und den Farben) und einem Analysewert (Anteil). Diese Modellbestandteile erzeugen wir über die entsprechenden Funktionen im Kontextmenü der Spaltenköpfe. Als Controller lassen wir bei der Würfelerzeugung eine Hilfsdimension „Periodenansicht“ anlegen – mit Sicherheit werden uns irgendwann die kurzfristigen Abweichungen interessieren, etwa wenn wir frische Daten bekommen haben, und dann benötigen wir Zeitanalyseelemente (siehe *DeltaMaster clicks!* 08/2007). Auf die Daten in der Excel-Tabelle könnten wir mit *DeltaMaster* auch direkt zugreifen, ohne vorher einen Würfel zu erzeugen. Das

Dimension	Ausgewählte Elemente
Jahr	Alle Jahre
Farbe	Alle Farben

'Neuzulassungen nach Farben'		
Jahr	Farbe	Anteil der Farben in Prozent
1991	Weiß	13,8
1992	Weiß	
1993	Weiß	
1994	Weiß	
1995	Weiß	
1996	Weiß	
1997	Weiß	
1998	Weiß	
1999	Weiß	
2000	Weiß	
2001	Weiß	
2002	Weiß	

OLAP-Datenbank erzeugen

Datenquelle: Provider=Microsoft.Jet.OLEDB.4.0;Data Source=S:\Marketi...

Typ: Cube-Datei Microsoft Analysis Services 2005/2008

Cube-Datei: C:\DeltaMaster\Neuzulassungen nach Farben.cub [Auswählen](#)

Bestehende Rechte erhalten Server-Zeitdimension erzeugen

MDX-Skripte erhalten Periodenansichtsdimension erzeugen

Fakttabellen zusammenfassen

ginge ein wenig schneller, aber wir müssten dann auf berechnete Elemente verzichten. Wenn Sie unsere Beispiele nachvollziehen möchten, kommen Sie hier aber ohne berechnete Elemente aus.

Nach diesem Schritt liegen die Daten multidimensional vor und wir können mit allen Funktionen von *DeltaMaster* damit arbeiten.

Erste Sichtkontrolle: alle Daten in der Pivottabelle

Auf der Basis des soeben erstellten Würfels legen wir eine neue Analysesitzung an und stellen die Daten zunächst als Pivottabelle dar. Das reicht für eine Sichtkontrolle: Alle 140 Datenwerte sind „drin“.

Anteil	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Blau	15,8	17,3	20,5	23,3	23,8	22,9	24,5	24,9	24,5	24,1	22,5	20,2	18,7	16,9	16,4	16,3	15,1	13,0	13,6	10,5
Braun	0,6	1,0	1,1	1,0	0,4	0,3	0,5	0,6	0,2	0,1	0,1	0,1	0,1	0,1	0,4	0,9	1,1	1,3	1,8	3,6
Grau	17,7	16,3	14,5	13,6	12,1	15,3	17,6	21,5	26,6	32,9	38,4	41,6	44,9	46,4	45,3	42,2	39,5	37,3	32,5	32,6
Grün	6,0	7,9	10,6	14,1	17,6	18,6	16,7	14,8	12,1	9,5	7,7	6,3	5,0	3,7	2,6	2,0	1,6	1,3	1,6	1,4
Rot	28,4	27,5	25,9	22,2	21,8	19,2	16,8	13,6	10,5	7,4	5,8	5,1	4,6	4,2	4,5	5,1	5,2	5,9	9,1	11,5
Schwarz	13,0	13,6	12,7	12,3	11,9	12,9	13,7	14,2	16,2	18,9	19,8	21,7	22,6	24,2	25,5	27,8	30,6	31,3	27,6	30,4
Weiß	13,8	12,2	10,4	8,2	7,6	6,1	5,3	4,7	4,4	4,2	3,5	3,1	2,4	2,3	2,1	1,6	2,9	6,3	9,8	6,8

Um vergleichen zu können, müssen wir jedoch weiter gehen – Entwicklungen, Größenunterschiede, Abstände, Verhältnisse, Muster sehen wir so noch nicht.

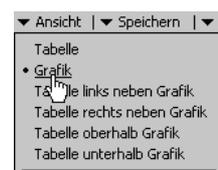
Grafische Tabelle

Über das Kontextmenü der Datenzellen blenden wir *Balken* ein und wir *transponieren* die Achsen (Menü *Ändern*). Jetzt stehen die Jahre in den Zeilen und die Farben auf den Spalten. Das Verlaufsmuster einer jeden Farbe ist gut zu erkennen und wir müssen auf keinen Einzelwert verzichten. Schade, dass die Zeit nun von oben nach unten verläuft statt von links nach rechts, wie es sein sollte. Mit Säulen könnten wir auch arbeiten, dann wird die Tabelle weniger kompakt. Probieren Sie es aus.

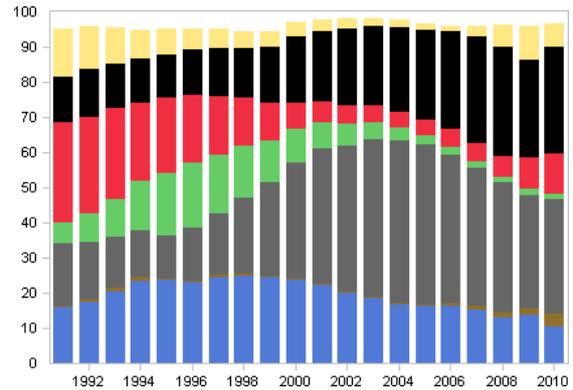
Anteil	Grau	Schwarz	Blau	Rot	Grün	Weiß	Braun
1991	17,7	13,0	15,8	28,4	6,0	13,8	0,6
1992	16,3	13,6	17,3	27,5	7,9	12,2	1,0
1993	14,5	12,7	20,5	25,9	10,6	10,4	1,1
1994	13,6	12,3	23,3	22,2	14,1	8,2	1,0
1995	12,1	11,9	23,8	21,8	17,6	7,6	0,4
1996	15,3	12,9	22,9	19,2	18,6	6,1	0,3
1997	17,6	13,7	24,5	16,8	16,7	5,3	0,5
1998	21,5	14,2	24,9	13,6	14,8	4,7	0,6
1999	26,6	16,2	24,5	10,5	12,1	4,4	0,2
2000	32,9	18,9	24,1	7,4	9,5	4,2	0,1
2001	38,4	19,8	22,5	5,8	7,7	3,5	0,1
2002	41,6	21,7	20,2	5,1	6,3	3,1	0,1
2003	44,9	22,6	18,7	4,6	5,0	2,4	0,1
2004	46,4	24,2	16,9	4,2	3,7	2,3	0,1
2005	45,3	25,5	16,4	4,5	2,6	2,1	0,4
2006	42,2	27,8	16,3	5,1	2,0	1,6	0,9
2007	39,5	30,6	15,1	5,2	1,6	2,9	1,1
2008	37,3	31,3	13,0	5,9	1,3	6,3	1,3
2009	32,5	27,6	13,6	9,1	1,6	9,8	1,8
2010	32,6	30,4	10,5	11,5	1,4	6,8	3,6

Pivotgrafik

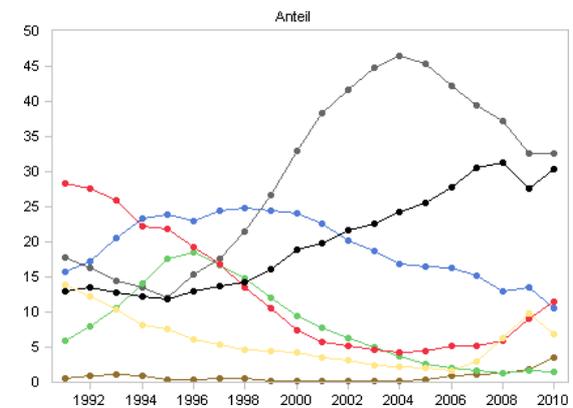
Die Pivottabelle ist mit einem Mausklick in eine *Grafik* überführt (Menü *Ansicht* im Fenster *Bericht*, *Analyse* oder *Cockpit*). Das Umschalten zwischen Tabelle und Grafik ist in allen Stufen möglich, vom *Reader* bis zum *Miner*. So können Sie als Grafik gespeicherte Berichte immer auch als Tabelle anzeigen, zum Beispiel um die exakten Werte abzulesen, die in der Grafik womöglich nicht enthalten sind. Zum Bearbeiten der Grafik ist der Modus *Pivotizer*, *Analyzer* oder *Miner* erforderlich.



Das Kraftfahrt-Bundesamt zeichnet die Daten in Pressemitteilungen bisweilen als Stapeldiagramm. Das haben wir in *DeltaMaster* nachgestellt. Aus den Spaghetti in der Zeitung wird dann Lasagne. Die einzelnen Segmente lassen sich nicht gut vergleichen, weil sie auf unterschiedlicher Höhe anfangen. Delikat ist zudem: Wie wir eine Entwicklung wahrnehmen, hängt davon ab, wo sie im Stapel liegt. Bella hat in ihrem Blog gezeigt, dass dieselbe Reihe einmal wie ein Anstieg und einmal wie ein Rückgang wirkt, je nachdem, auf welche andere Reihe man sie stapelt (www.bella-beraet.de/linien-kumuliert-man-nicht).



Wenn wir den Diagrammtyp ändern und ein Liniendiagramm zeichnen, wird unser Nachbau wie in der Zeitung – alles spaghetti. Es sind zu viele Linien, als dass wir sie mit dem Auge gut verfolgen könnten. Dass sich die Linien immer wieder kreuzen, macht die Angelegenheit nicht einfacher. Immerhin, die Interaktivität am Bildschirm kann die Lektüre etwas erleichtern: Wenn Sie mit der Maus auf eine Linie zeigen, werden die übrigen Linien abgeblendet und die selektierte wird betont. Anstelle aller Linien sieht man dann praktisch nur noch eine – ein Gedanke, den wir weiter unten wieder aufgreifen werden. Die *Zahlenwerte* fehlen hier. Über das Kontextmenü könnte man sie einblenden. Dann wird aber aus Spaghetti Buchstabensuppe. Probieren Sie es aus: Machen Sie die Grafik größer. Alles wird wieder lesbar und lässt sich intensiv studieren. Und wir merken dabei: Die Skalierung vermag uns nur die großen Farben überzeugend zu zeigen. Skalierungsvarianten gibt es auch in der Pivotgrafik; prädestiniert dafür ist aber das Modul *Zeitreihenanalyse*.



Zeitreihenanalyse

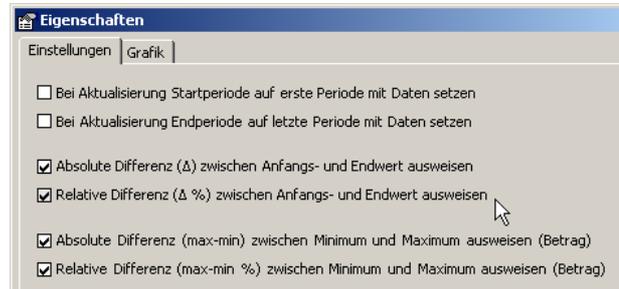
Die *Zeitreihenanalyse* im Modus *Miner* erzeugt eine ähnliche Darstellung wie in der vorherigen Abbildung. (Auch im *Analyzer* lassen sich *Zeitreihenanalysen* erstellen, jedoch nur für eine Datenreihe.) Anders als in der *Pivotgrafik* steht bei der *Zeitreihenanalyse* nicht die Grafik im Vordergrund, sondern es geht vor allem um Funktionen, die man aus betriebswirtschaftlicher Sicht braucht, um *Zeitreihendaten* zu analysieren.

Beispielsweise kann man über das Kontextmenü *Trendgeraden* einblenden; diese wiederum lassen sich trennen, etwa wenn sich zu einem bestimmten Zeitpunkt die äußeren Umstände so geändert haben, dass „vorher“ und „nachher“ nicht vergleichbar sind. *Kumulierte Werte* können auch ohne eine spezielle Dimension im Analysemodell dargestellt werden. Bei *absoluten Werten* werden negative Vorzeichen umgekehrt, was den betragsmäßigen Vergleich von Wertreihen erleichtert, die mit unterschiedlichen Vorzeichen erfasst sind. Die *Schrittweite* ist hilfreich, wenn beispielsweise Daten auf Tagesebene vorliegen, aber der Wochentag kein Kriterium im Datenmodell ist: Mit einer Schrittweite von 7 stehen dann trotzdem etwa nur die

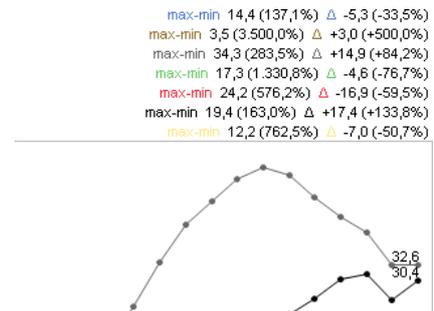
Normieren nach '1991'	
Normierung aufheben	
Kumulierte Werte	
Absolute Werte	
Schrittweite	▶
Nullwerte	▶
Reihen auf 2. Y-Achse	▶
Logarithmische Skala	
Trendgeraden	

Montage nebeneinander. Diese und andere Funktionen sind dem Analyseverfahren *Zeitreihenanalyse* zugeordnet und nicht mit Grafikoptionen in Pivotgrafiken zu verwechseln.

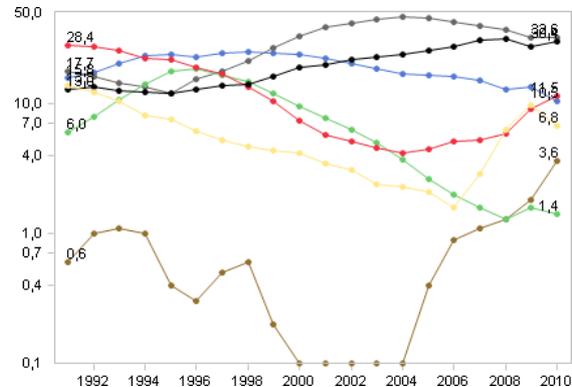
Für unsere Aufgabe erweisen sich die *Eigenschaften* (Kontextmenü oder Menü *Ich möchte*) als hilfreich. Hier lässt sich einstellen, dass die *absolute und/oder relative Differenz zwischen Anfangs- und Endwert* sowie *zwischen Minimum und Maximum* ausgewiesen werden soll.



Diese Angaben erscheinen oberhalb des Diagramms. Sie geben wertvolle Hinweise auf die Spannweite der Werte. Die wiederum ist ein wichtiger Indikator, ob bei der Skalierung Gefahr droht. In unserem Beispiel ist das so: Der Marktanteil der Farbe Braun hat von 1991 bis 2010 um 500 Prozent zugenommen! Das ist bei weitem die größte Steigerung. Aus dem Diagramm kann man das aber beim besten Willen nicht ablesen: Weil sich Braun insgesamt auf einem niedrigen Niveau bewegt, sind auch die enormen prozentualen Veränderungen nicht sichtbar. Reihen mit kleinen Werten werden von den Reihen mit großen Werten regelrecht „flach gedrückt“. Das ist ein ganz typischer Effekt, wenn man versucht, Zeitreihen unterschiedlicher Größenordnung in einem Diagramm darzustellen.

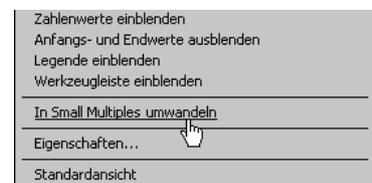


Abhilfe schafft die *logarithmische Skala* (Kontextmenü). Sie macht die relativen Unterschiede vergleichbar: Der gleiche Höhenunterschied steht für den gleichen Wertunterschied in Prozent. Mehr über die logarithmische Skalierung und ihre Vorzüge entnehmen Sie den *DeltaMaster clicks! 07/2010*. Die Lesbarkeit des Diagramms leidet aber immer noch daran, dass es so viele Linien sind.



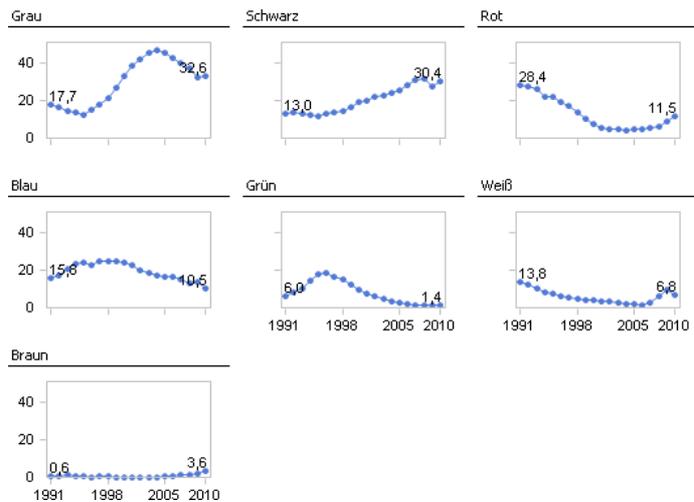
Small Multiples

Deshalb sollten wir die Spaghetti loswerden. Im Modus *Miner-Expert* gibt es dafür einen Menübefehl: Umfasst die Analyse mehr als eine Reihe, bietet *DeltaMaster* an, die Darstellung in *Small Multiples* umzuwandeln (Menü *Ich möchte*).

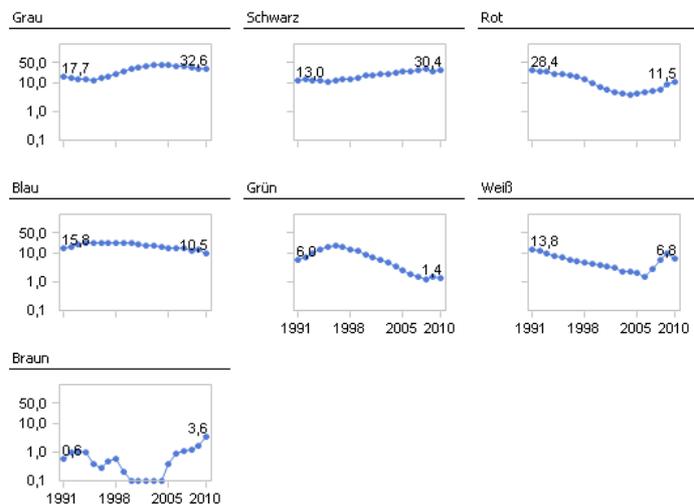


Als *Small Multiple* wird für jede Farbe eine eigene, kleine Zeitreihendarstellung erzeugt. Diese fasst *DeltaMaster* zu einem Bericht zusammen. Hier haben wir uns für die globale Skalierung entschieden. Die einzelnen Verläufe sind klar zu erkennen.

Allerdings haben wir noch das Problem mit der Skalierung: Die starke Zunahme des Marktanteils von Braun geht unter.



Das ändert sich, wenn wir die *Small Multiples* aus einer *Zeitreihenanalyse* erzeugen, die logarithmisch skaliert ist. In der Abbildung rechts gilt ebenfalls eine globale Skalierung, das heißt, alle Kacheln verwenden die gleiche Achseneinteilung. Diese ist aber logarithmisch – und schon erkennt man, wie viel Bewegung es bei Braun gab und wie wenig bei Blau. Ähnliche Ergebnisse erzielt man mit der vergleichbaren Skalierung, einer Option in *Small Multiples*. Sie bewirkt, dass die prozentuale Spannweite in allen Kacheln übereinstimmt, und stellt darüber die Vergleichbarkeit her.

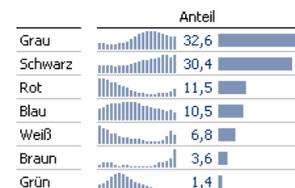


Haben wir mit dieser Darstellung erreicht, was wir wollten? Die Verlaufsmuster kann man jetzt gut erkennen und miteinander vergleichen. Wie aber steht es um die absolute Bedeutung? Das starke Wachstum von Braun kann ein wenig darüber hinwegtäuschen, dass der Anteil dieser Farbe mit nicht einmal 4 Prozent immer noch sehr niedrig ist.

Sparklines

Für den nächsten Schritt wollen wir daher Größenvergleich und Zeitvergleich trennen. Zu diesem Zweck kommen wir auf die Grafischen Tabellen zurück. Eingangs hatten wir für jedes Jahr den Marktanteil mit Balken visualisiert. Das tun wir jetzt nur noch für das letzte Jahr, für 2010.

Dessen Werte werden in Form von Balken visualisiert. Die zeitliche Entwicklung haben wir in Sparklines abgebildet, die zellweise skaliert sind. Dadurch wird die Zeilenhöhe in jeder Zeile voll ausgenutzt und die Sparkline differenziert maximal. Der Bericht ist sehr kompakt und ansprechend. Wegen der Deutungspriorität sind die Sparkline-Säulen linear skaliert (siehe *DeltaMaster clicks! 12/2011*). Der Mustererkennung tut das keinen Abbruch. Wir sehen:



Der Anteil von Braun ist mit 3,6 Prozent gering. Bemerkenswert ist das dennoch, weil Braun lange überhaupt keine Rolle spielte und erst in den letzten Jahren die Werte steigen.

Minima und Maxima als Zahlenwerte geben weitere Orientierung. Für Grün ist das dringend nötig: Hier ist der aktuelle Marktanteil von 1,4 Prozent so weit vom Maximum entfernt, dass die rein visuelle Deutung an ihre Grenzen stößt. Minimum und Maximum haben wir als benutzerdefinierte Analysewerte in *DeltaMaster* angelegt.

	Anteil	Min	Max
Grau	32,6	12,1	46,4
Schwarz	30,4	11,9	31,3
Rot	11,5	4,2	28,4
Blau	10,5	10,5	24,9
Weiß	6,8	1,6	13,8
Braun	3,6	0,1	3,6
Grün	1,4	1,3	18,6

Fazit

Wir sehen mal wieder: Die Analyse von Zeitreihen ist ein tückisches Geschäft. In Form von Liniendiagrammen kommen Zeitreihen daher, als könnte man leicht ablesen, wie sie steigen, stagnieren oder fallen. Dem ist nicht so. So weit Liniendiagramme verbreitet sind, so oft werden sie vorschnell interpretiert. Gut, dass Sie mit *DeltaMaster* alles Rüstzeug zur Hand haben, um sie zu nutzen, wo sie taugen, und ihnen zu trotzen, wo sie täuschen.