



**TQU VERLAG**

**Der  
Violin Plot**

Probieren und Studieren

# Der Violin-Plot

Daten attraktiv präsentieren und intelligent interpretieren

[Dr. Konrad Reuter](#)

Mit statistischen Ergebnissen überzeugen heißt nicht zuletzt, diese Daten in geeigneter grafischer Form darzustellen. Sowohl statistische Software als auch Bürosoftware bieten hierfür vielfältige Möglichkeiten.

Aber die Entwicklung bleibt auch hier nicht stehen. Insbesondere für den Vergleich von Datenreihen sind neue Grafiken entwickelt worden. Unter dem Namen "Violin-Plot" ist im deutschsprachigen Raum eine kleine Grafik bekannt. Diese Grafik kombiniert statistisch relativ einfache Box-Plots mit einer Schätzung der Dichtekurve empirischer Verteilungen. Die Dichtekurve wird an der Merkmalsachse gespiegelt dargestellt, so dass dem Diagramm eine gewisse Ähnlichkeit mit dem Resonanzkörper einer Violine zugesprochen werden kann.

Die vorliegende Datei ermöglicht es dem Nutzer bereits mit der Tabellenkalkulation diese Grafik zum Vergleich von Datenreihen zu verwenden.

Die Datei ist ebenfalls geeignet, einen Zugang zur Schätzung von Dichtekurven zu gewinnen (Kernel-Schätzer).

Eine praktikable Lösung mit EXCEL basiert auf der Verwendung von dynamischen Bereichsnamen und der Matrixfunktionalität von Formeln.

Ansprechpartner: Dr. Konrad Reuter

Telefon: 0171/6006604

TQU Verlag, Magirus-Deutz-Straße 18, 89077 Ulm Deutschland, Telefon 0731/14660200, [verlag@tqu-group.com](mailto:verlag@tqu-group.com), [www.tqu-verlag.com](http://www.tqu-verlag.com)

## QUALITY APPS Applikationen für das Qualitätsmanagement

### Lizenzvereinbarung

Dieses Produkt "Violin Plot" wurde vom Autor Dr. Konrad Reuter mit großem Aufwand und großer Sorgfalt hergestellt.

Dieses Werk ist urheberrechtlich geschützt (©). Die dadurch begründeten Rechte, insbesondere die der Weitergabe, der Übersetzung, des Kopierens, der Entnahme von Teilen oder der Speicherung bleiben vorbehalten.

Bei Fehlern, die zu einer wesentlichen Beeinträchtigung der Nutzung dieses Softwareproduktes führen, leisten wir kostenlos Ersatz. Beschreibungen und Funktionen verstehen sich als Beschreibung von Nutzungen, Möglichkeiten und nicht als rechtsverbindliche Zus

Sie erklären sich damit einverstanden, dieses Produkt nur für Ihre eigene Arbeit und für die Information innerhalb Ihres Unternehmens zu verwenden. Sollten Sie es in anderer Form, insbesondere in Schulungs- und Informationsmaßnahmen bei anderen Unternehme

[www.tqu-verlag.de](http://www.tqu-verlag.de)

Alle Ergebnisse basieren auf den vom Autor eingesetzten Formeln und müssen vom Anwender sorgfältig geprüft werden.

Die berechneten Ergebnisse sind als Hinweise und Anregungen zu verstehen.

Wir wünschen viel Spaß und Erfolg mit dieser Applikation

TQU Verlag, Magirus-Deutz-Straße 18, 89077 Ulm Deutschland, Telefon 0731/14660200, verlag@tqu-group.com, www.tqu-verlag.com

## 1. Statistik

### 1.1 Lösungsidee

Violin Plots sind Methoden zur grafischen Darstellung von Daten. Die Methode ist in der deutschen Literatur kaum zu finden [1] [2].

Eine Reihe von Software, unter anderem "R", enthält entsprechende Module bzw. es ist auf solche Algorithmen verwiesen.

Ein "Violin Plot" kombiniert eine Box Plot mit symmetrisch zugeordneten Dichtekurven.



Bild1: Beispielplot mit einer Normalverteilung

Die Boxplot wird klassisch mit Median und Whiskers dargestellt.

Die Dichtekurve wird zwischen den Maxima berechnet und gespiegelt um die Box positioniert, so dass mit einiger Vorstellungskraft das nebenstehende Bild einer Violinkontur passt.

### 1.2 Boxplot

Box-Whisker beruhen auf Quantilen und sind nicht von einer Verteilung abhängig.

Die Box erstreckt sich vom unteren Quartil bis zum oberen Quartil und umfasst somit 50% der Daten. Der Abstand zwischen den Quartilen wird als Interquartilsabstand IQR bezeichnet.

Die Box wird vom Median geteilt. Ein Kreuzsymbol kann für den arithmetischen Mittelwert verwendet werden.

Der Abstand der Whiskers (Antennen) kann zwischen 1 und 1,5 IQR gewählt werden.

Die Darstellung von "Ausreißern" in Box Plots ist in [4] beschrieben.

### 1.3 Dichtekurven

Von empirischen Daten liegt die Wahrscheinlichkeitsdichte in der Regel nicht vor. Anzahl der Treppenstufen bzw. die Klassenbreite sind wählbar in bestimmten sinnvollen Grenzen.

Histogramme zeigen die den gewählten Klassen zugeordneten mittleren Wahrscheinlichkeiten. Insofern ergibt sich immer eine Treppenkurve.

Die Anwendung der Formatierungsoption "Linie glätten" durch EXCEL auf ein Histogramm bzw. Polygonzug führt nicht zu einem mathematisch sinnvollem Ergebnis!

Eine mathematisch begründete geglättete Darstellung der Dichtefunktion wird durch einen sogenannten Kernel (Dichte) Schätzer erhalten.

Dichteschätzer verwenden folgende Funktion :

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n W\left(\frac{x-x_i}{h}\right)$$

mit  $n$  als Stichprobenumfang und  $h$  als Bandbreite.

Für die Funktion  $W$  sind unterschiedliche Kernelschätzer anwendbar. Hierzu gibt es eine Reihe von Spezialliteratur, auf die hier nicht eingegangen wird.

Als brauchbar hat sich folgenden Schätzer erwiesen [3]:

$$W(u) = \begin{cases} 1 + \cos(2\pi u) & \text{falls } |u| < 1/2 \\ 0 & \text{sonst} \end{cases}$$

Die Wahl von  $h$  richtet sich nach dem Stichprobenumfang  $n$  und der gewünschten Glättung.

Die Größe von  $h$  wird als % Anteil der Spannweite  $R$  angegeben.

Größere Werte  $h$  bewirken ein bessere Glättung, führen aber auch zu einer Verzerrung (Bias).

Näheres dazu auch in der Spezialliteratur.

Die Testdatensätze geben Beispiele zur Wahl der Bandbreite  $h$ .

Variable	Verteilung	n	h
Dat_1	empirisch1	47	60%
Dat_2	Normalverteilung	500	50%
Dat_3	Weibullverteilung	1000	40%
Dat_4	empirisch2	1933	30%

## 2. Anwendung

Wie auch bei Box Plots lebt die Grafik vom Vergleich von Stichproben [2].

Im Blatt Merkmal sind zugehörigen Angaben einzutragen.

Im Blatt Daten sind die Daten einzutragen oder aus Anwendungen zu übernehmen.

Im Blatt Testdaten sind als Beispiel vier Datensätze vorbereitet.

Sie können einen Datensatz auswählen und sehen sofort im Blatt "Plot" die Veränderung.

Die Anpassung einer geeigneten Bandbreite müssen Sie mit dem Schieberegler selbst vornehmen.

[Merkmale](#)

[Daten](#)

[Testdaten](#)

Der Datenumfang ist nur durch EXCEL begrenzt. Der Datenbereich ist als dynamischer Bereichsnamen festgelegt.

Im Blatt Kennwerte sind übliche Ergebnisse der deskriptiven Statistik aufgeführt.

Die Auswahl von  $h$  erfolgt über die Bildlaufleiste in der Grafik, so dass das Ergebnis direkt beobachtet werden kann.

[Kennwerte](#)

## 3. Die Auswahl des Faktors für den Interquartilsabstand $k$ erfolgt im Blatt Kennwerte.

Die y-Achse der Grafik ist ggf. dem Datenbereich manuell anzupassen.

### EXCEL

Die EXCEL-Lösung stützt sich auf folgende Prinzipien:

Funktionale Arbeit über verschiedene Blätter

Optische Hervorhebung von Zellen in Abhängigkeit von ihrer Funktion.

Kommentierung wesentlicher Zellen

Vergaben von Namen für Variable.

Die Liste muss nach Änderungen mit F3 manuell aktualisiert werden.

Einsatz sog. dynamischer Namen, der Umfang sich den Inhalten anpasst.

Textliche Kommentierungen in rot kursiv.

Auslesen von Tabelleninhalten

Bezug auf benannte Zellbereiche (ist ein sog. volatile Funktion in EXCEL).

Feststellen von Zellinhalten

Reagieren auf Bedingungen /Verzweigungen

Erzwingung einer geeigneten Zahlenformatierung im Protokoll mit vorgegebener Stellenzahl.

Verknüpfung von Zellinhalten über "&"

Zellen werden in Berechnungsblättern und Protokollen nicht über "Zellen verbinden" formatiert!

Als Lösung dient die Formatierung schmaler Spalten und die Formatierung

benachbarter Zellen mit "Über Auswahl zentrieren".

Die Nachteile verbundener Zellen sind damit vermieden..

Ausblenden von Zellinhalten, die nicht zutreffend sind.

Ausblenden von Blättern, die nicht ständig gebraucht werden.

Blattschutz gegen versehentliches Überschreiben, bitte nutzerintern anpassen.

Matrixfunktionen werden aus "normalen" Funktionen insofern erzeugt, als beim Eingeben die Funktion mit "strg", "shift" und "enter" gleichzeitig (beidhändig arbeiten) abgeschlossen wird.

Die geschweiften Klammern werden bei dieser Prozedur von EXCEL automatisch eingefügt, sind also nicht mit der Tastatur erzeugbar!

Matrixfunktionen arbeiten gewissermaßen als "Schleife" und ersparen so häufig eine sonst notwendige VBA Prozedur!

*Achtung, die Berechnung dauert ein wenig länger!*

## 4. In den Grafiken werden z.B. Grenzl意思en nur mit einem Datenpunkt erzeugt.

Die Linie wird dann mit der Formatierung als "Fehlerindikator" erzeugt.

Dieser Fehlerindikator ist sehr flexibel einsetzbar und erweitert die Möglichkeiten grafischer Gestaltung bedeutend.

[Merkmale](#)

[Eingabedaten](#)

[Kennwerte](#)

[Bezeichnungen](#)

[Namen](#)

[BEREICH.VERSCHIEBEN](#)

[Bemerkungen](#)

[VERGLEICH\(;;0\)](#)

[INDEX\(;;\)](#)

[INDIREKT\(\)](#)

[ISTLEER\(\)](#)

[ISTZAHL\(\)](#)

[WENN\(;;\)](#)

[FEST\(;;\)](#)

[="text1"&BEZUG](#)

[Berechnung](#)

[1]

[2]

[3] Quellen

[4] [wikipedia.org/wiki/Violin\\_plot](http://wikipedia.org/wiki/Violin_plot)

[eike-klima-energie.eu/climategate-anzeige/ein-modell-eine-wahl/](http://eike-klima-energie.eu/climategate-anzeige/ein-modell-eine-wahl/)

STATGRAPHICS XVI statpoint inc.

Quality APP 693TQU Ausreißerbehandlung

## Der Violin Plot

<i>Bezeichnung</i>	Testdaten
<i>Teile-Nr.</i>	0815
<i>Merkmal</i>	Durchmesser
<i>Maßeinheit</i>	mm
<i>Messmittel</i>	Messzylinder
<i>Auflösung</i>	0,001

QUALITY APPS im TQU VERLAG

www.tqu-verlag.de

TQU Verlag, Magirus-Deutz-Straße 18, 89077 Ulm Deutschland, Telefon 0731/14660200, verlag@tqu-group.com, www.tqu-verlag.com

## Der Violin Plot

Stichprobenumfang	$n$	47
Mittelwert		12,821915
Standardabweichung	$s$	10,445045
Schiefte		1,307534
Größtwert	$max$	47,78
Kleinstwert	$min$	0,74
Perzentil 1%	$P_1$	47,04
Perzentil 5%	$P_5$	10,65
Median	50%	9,44
Perzentil 95%	$P_{95}$	10,65
Perzentil 99%	$P_{99}$	47,04
Interquartilsabstandsfaktor	$k$	1
Interquartilsabstand	$IQR$	15,575
ob. Ausreißergrenze		36,24
unt. Ausreißergrenze		-10,485
Kernelschätzung		
Bandbreite in % von $R$	$h$	60%
mit Bildlaufleiste in Plot auswählen		

QUALITY APPS in TQU-VERLAG

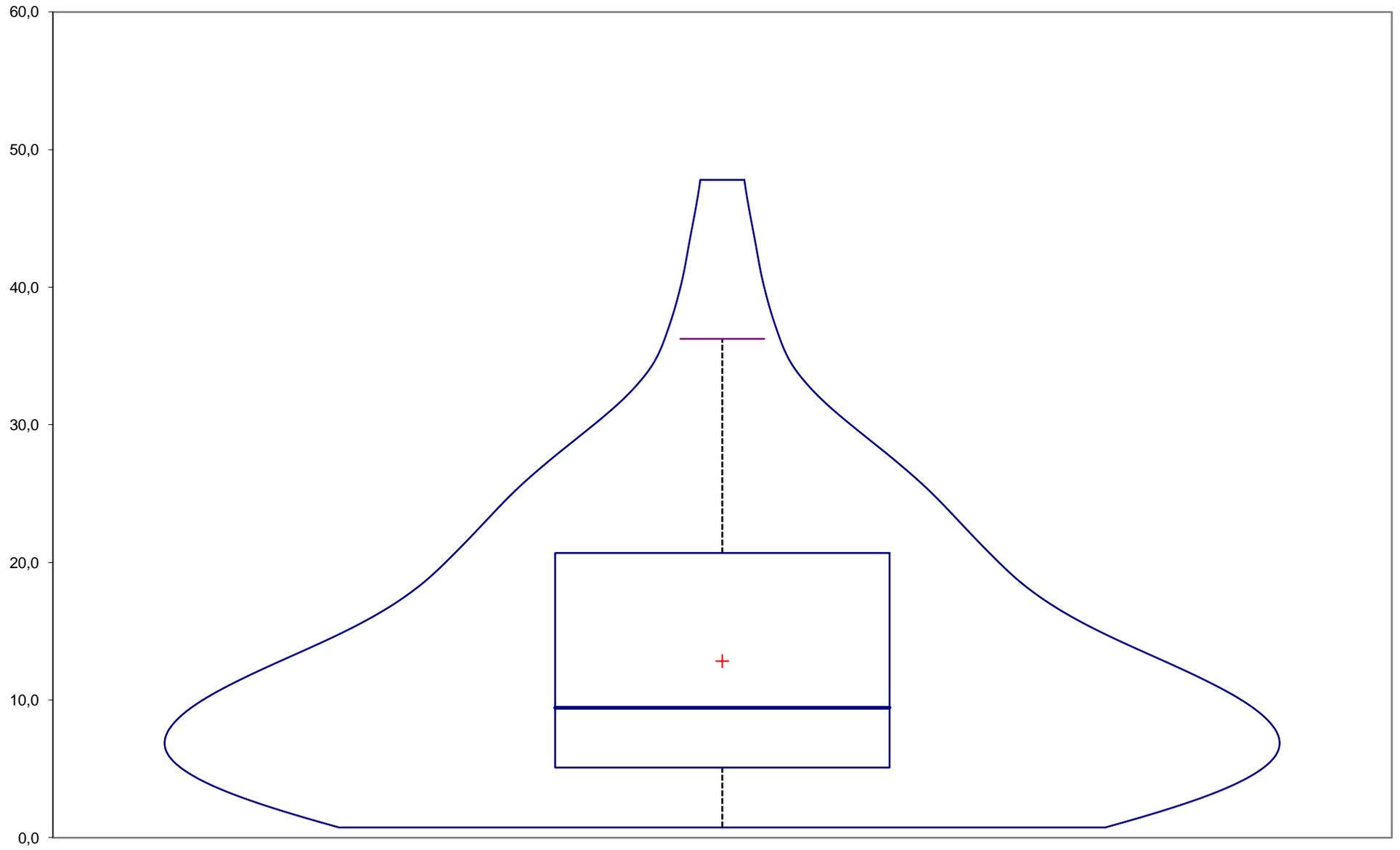
www.tqu-verlag.de

auswählen

WAHR

FALSCH

60



Der Violin Plot

Box	
x	y
-0,0125	20,665
0,0125	20,665
0,0125	5,090
-0,0125	5,090
-0,0125	20,665

Whiskers		
min	Q1	Q3
0	0,74	4,35

Interquartile	
Q1	Q3
-0,0031	36,24
0,0031	36,24
0,0031	#NV
C.V.	1,17
max y	0,04164

Median	
Q1	Q3
0,0125	9,44
0,0125	9,44
Mittelwert	
0	12,822

Maximumlinie		Zeile
-0,0016	47,78	101
0,0016	47,78	
Minimumlinie		
-0,0286	0,74	
0,0286	0,74	

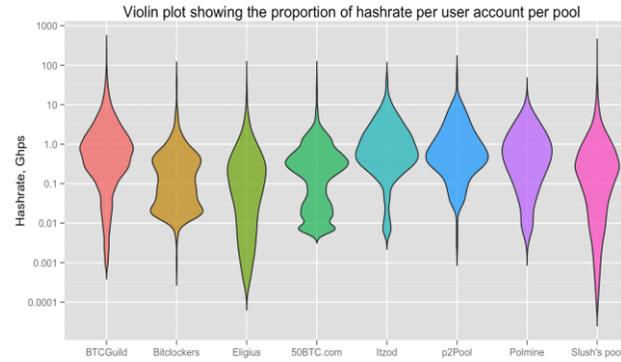
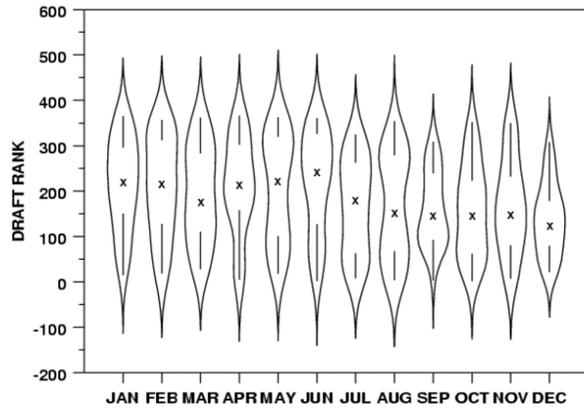
Dat_1	Dat_2	Dat_3	Dat_4
8,25	-0,81	917,2	18,7041
2,82	-0,164	489,7	5,8839
4,16	0,395	299,6	12,5175
18,66	-0,76	426,7	15,165
12,72	-0,51	663,8	11,1806
8,75	1,626	684,6	2,0766
2,29	-1,281	651,7	-1,1967
7,22	0,517	310	10,0656
9,76	-0,86	412,8	7,1581
7,72	0,115	761,4	6,9582
27,38	-1,011	14,9	0,5226
5,14	-0,532	160,1	-0,3462
47,78	-0,593	741	6,0062
35,92	-0,613	1,9	53,9
21,36	-0,567	230,8	7,5637
4,97	-0,358	565,3	3,5816
0,74	0,024	419,1	3,1105
5,96	-0,712	532,6	11,9821
5,04	-0,582	264,9	3,4034
14,98	-1,742	619,5	10,3959
6,09	-1,467	548,1	2,767
11,05	0,19	390,9	8,1101
5,31	-0,612	334,6	2,2043
13,36	0,938	153,6	2,8305
22,27	1,009	189,9	12,123
28,31	-0,873	735,5	12,9816
21,95	-0,227	383,3	7,5046
20,87	-0,086	95,3	8,372
4,85	2,141	119,3	14,8389
3,1	-0,752	364,9	2,8989

Variable	Verteilung	n	h	Testdaten übernehmen
Dat_1	empirisch1	47	60%	
Dat_2	Normalverteilung	500	50%	
Dat_3	Weibullverteilung	1000	40%	
Dat_4	empirisch2	1933	30%	
1	Verteilung	n	h	
	empirisch1	47	60%	

TQU Verlag, Magirus-Deutz-Straße 18, 89077 Ulm Deutschland, Telefon 0731/14660200, verlag@tqu-group.com, www.tqu-verlag.com

QUALITY APPS im TQU VERLAG  
www.tquverlag.de

# Der Violin Plot: Beispiele

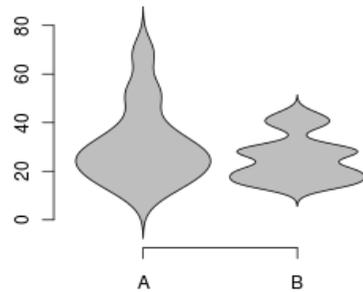


Quelle: <http://organofcorti.blogspot.de/2013/04/121-pre-asic-network-hashrate.html>

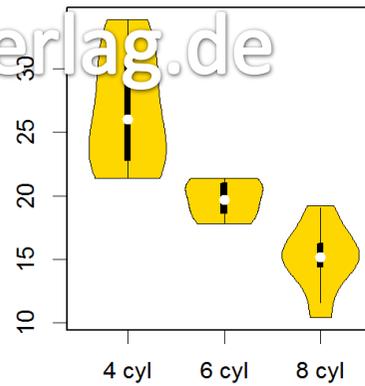
Quelle: [http://en.wikipedia.org/wiki/Violin\\_plot](http://en.wikipedia.org/wiki/Violin_plot)

QUALITY APPs im TOU VERLAG  
Violin Plots of Miles Per Gallon

www.tqu-verlag.de



Quelle: <http://users.monash.edu.au/~murray/AIMS-R-users/ws/ws11.html>



Quelle: <http://f.dataguru.cn/thread-107352-1-1.html>