

Erstellung von Boxplots, Grafiken zur Darstellung der groben Kennwerte der Grundgesamtheit

Z.B. der Anzahl und Kaufpreis-Unterschiede einzelner Stadtbezirke oder Ortsteile in einer Stadt einschließlich der Spannen und Mittelwerte

Um von einer Mietdatenbank oder einer Kaufpreisauswertung ein BoxPlot erstellen zu können, muss zunächst eine statistische Auswertung vorgenommen werden und aus den berechneten Kennwerten eine Matrix mit nachstehendem Aufbau in einen eigenen Tabellen-Bereich geschrieben werden. Ich erledige dies als Nebenaufgabe der programmierten statistischen Auswertung gleich mit.

Es empfiehlt sich diese Kennwerte wie vorgestellt anzulegen, und zwar die ausgewählten Kennwerte in der Excel-Tabelle, in der Sie Ihre grundlegende Datenauswertung vorgenommen haben. (Werte aus der Spalte, die die relativen Kaufpreise je m², bzw. die relativen Mieten je m² enthält.)

Es gibt mehrere Darstellungsvarianten, bei der auch unterschiedliche Kennwerte verarbeitet werden können.

In dieser Anleitung wird die dritte von Excel bereitstellte Variante verwendet.

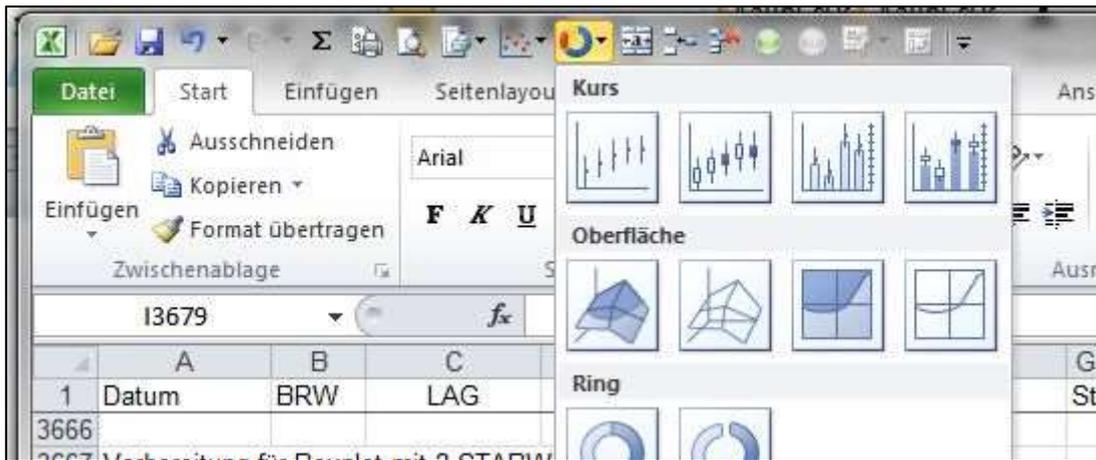
In diese Matrix kopieren Sie (Aufbau, Inhalt und Reihenfolge der Spalten ist entscheidend!)

- in Spalte A die die Stadtbezirksnamen,
- in Spalte B die Anzahl der Datensätze je Stadtbezirk,
- in Spalte C den unteren Schwellenwert (hier $\mu - 1 \sigma$) (Andere Kriterien sind denkbar)
- in Spalte D den oberen Schwellenwert (hier $\mu + 1 \sigma$) (Andere Kriterien sind denkbar)
- in Spalte E den Median, hier könnten auch die arithmetischen Mittelwerte stehen

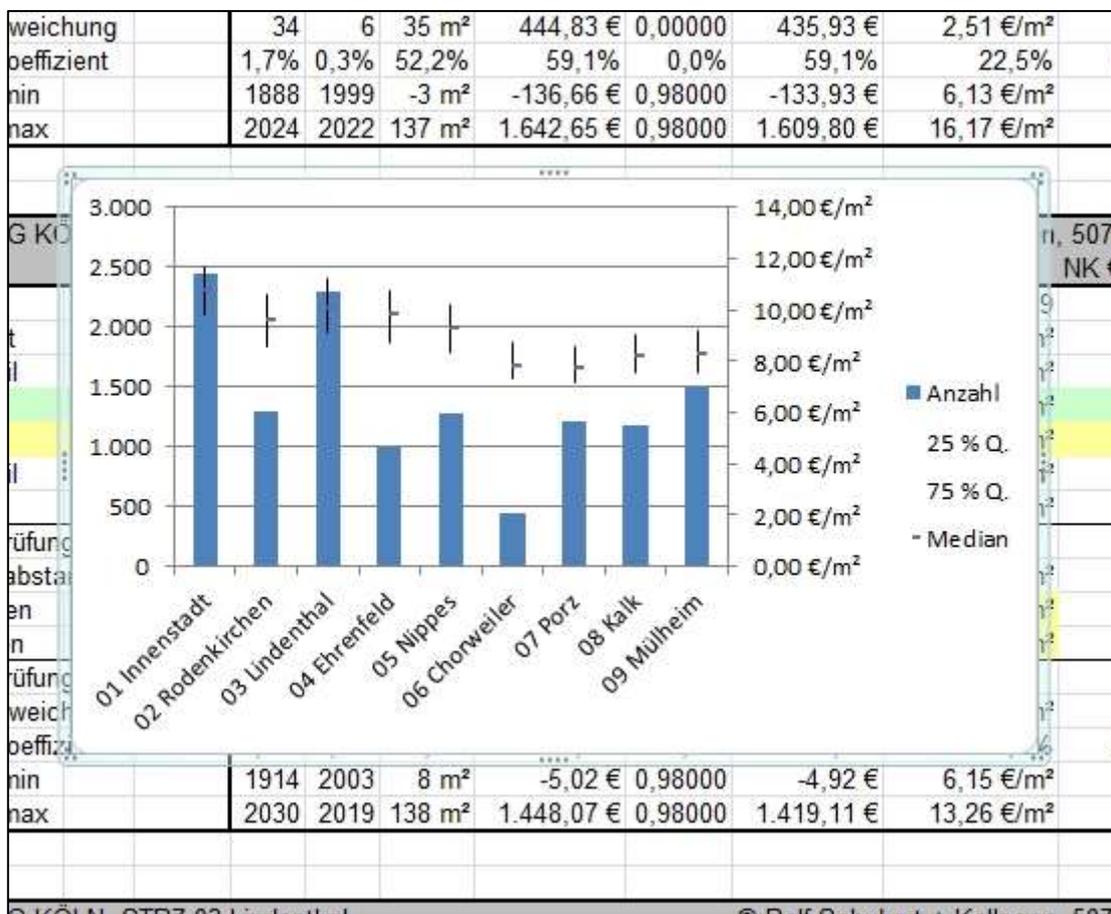
und formatieren die relativen Kaufpreise entsprechend.

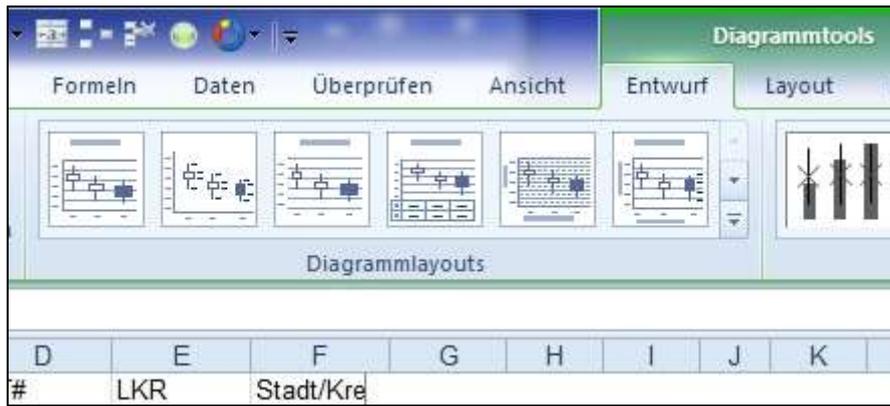
	A	B	C	D	E	F	G
1	Datum	BRW	LAG	STB	OT#	LKR	Stz
3666							
3667	Vorbereitung für Boxplot mit 2 STABW						
3668	STBZ	Anzahl	$\mu - 1 \text{ Sigma}$	$\mu + 1 \text{ Sigma}$	Median		
3669	Innenstadt	757	1.342,86 €/m ²	2.079,92 €/m ²	1.701,00 €/m ²		
3670	Rodenkirchen	337	1.208,81 €/m ²	1.957,56 €/m ²	1.572,29 €/m ²		
3671	Lindenthal	709	1.233,76 €/m ²	2.114,48 €/m ²	1.681,70 €/m ²		
3672	Ehrenfeld	211	1.187,24 €/m ²	1.864,19 €/m ²	1.543,59 €/m ²		
3673	Nippes	333	1.210,49 €/m ²	1.745,66 €/m ²	1.471,15 €/m ²		
3674	Chorweiler	127	1.161,73 €/m ²	1.676,59 €/m ²	1.451,25 €/m ²		
3675	Porz	366	1.014,94 €/m ²	1.597,55 €/m ²	1.279,41 €/m ²		
3676	Kalk	252	927,66 €/m ²	1.438,35 €/m ²	1.177,30 €/m ²		
3677	Mülheim	212	1.141,37 €/m ²	1.776,51 €/m ²	1.453,90 €/m ²		
3678	Erstellen Sie aus den Zellen A3668 bis E3677 eine Boxplot-Grafik mit nachstehendem Diagramm-Titel						
3679	Bereinigte, relative KP €/m ² für ETW in Köln nach Stadtbezirken im Jahr 2005						

Anschließend wird der im Beispiel markierte Bereich (Hier Zellen A3668:E3677) kopiert und danach im Menü <Einfügen> – <Diagramme> das Diagramm-Menü geöffnet. Ich empfehle Excel vorher so vorzubereiten, dass es eine Schnellformat-Leiste gibt, in die das Kreis-Icon für Diagramme aufgenommen ist.

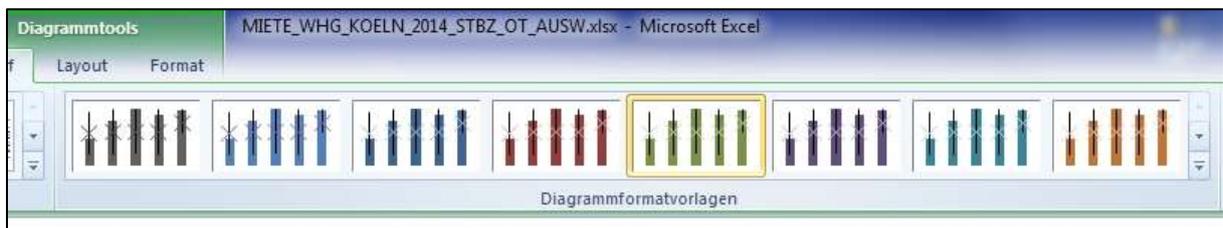


In der Auswahl wählt man den 3. Diagrammtyp aus. Excel erstellt eine Rohfassung





Im Diagrammtool „Entwurf“ wählt man die 4. Diagramm-Layout-Vorlage aus, worin unten im Diagrammbereich die Daten des Plots in Rechtecken eingetragen werden.



Im Diagrammtool Entwurf wählen Sie die Farbe aus, die die Anzahl der Datensätze darstellen soll.

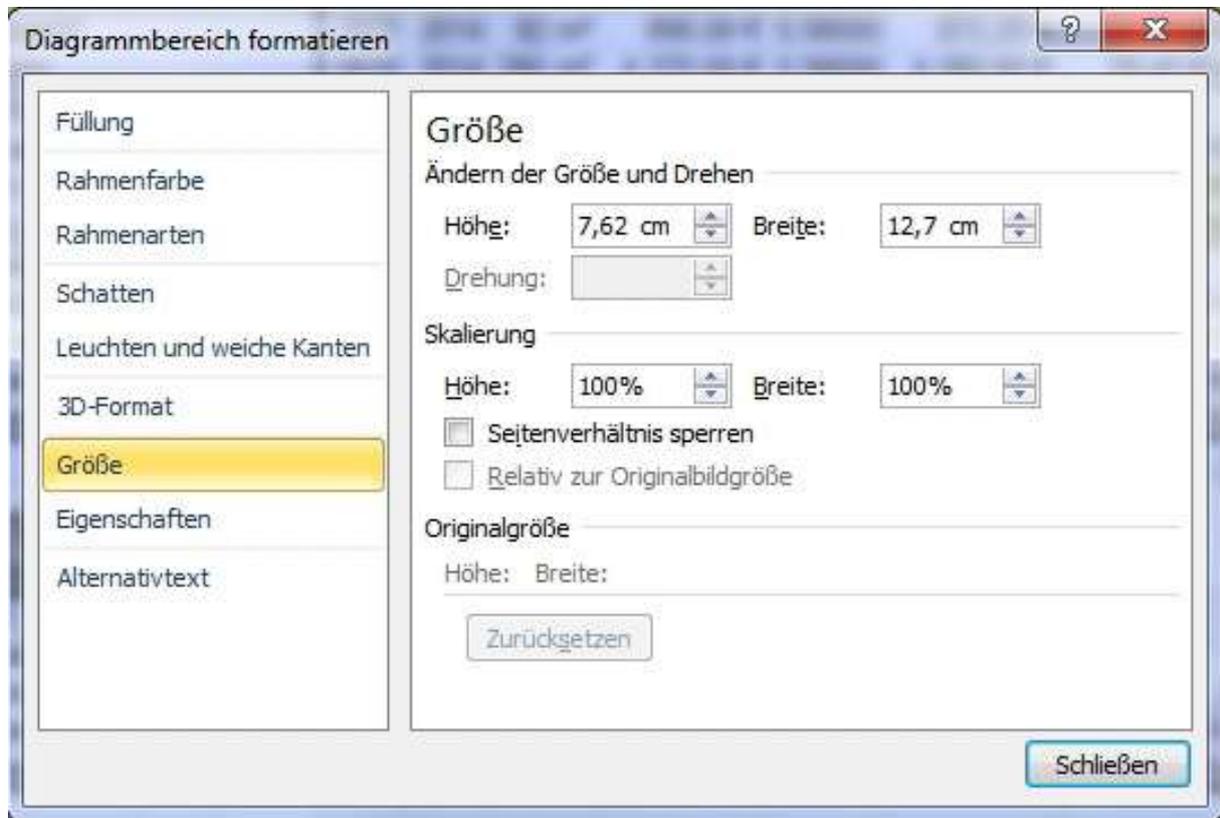
	01	02	03	04	05	06	07	08	09
Anzahl	2.444	1.294	2.297	1.003	1.281	447	1.210	1.175	1.489
25 % Q.	9,80	8,58	9,13	8,75	8,33	7,33	7,14	7,56	7,59
75 % Q.	11,71	10,57	11,19	10,75	10,21	8,72	8,55	9,01	9,17
Median	10,78	9,58	10,09	9,80	9,28	7,79	7,75	8,20	8,30

Danach klickt man mit der rechten Maus-Taste auf den Diagramm-Entwurf und wählt <Diagrammbereich formatieren ...> aus.

Zunächst wird die Größe festgelegt, das hat den Vorteil, dass bei entsprechender Größe die STBZ-Namen in einer Zeile neben einander stehen können.

Sinnvoll ist eine Diagrammgröße im Verhältnis Höhe zu Breite die Din A 4 Quer entspricht.

Es hat sich z.B. bewährt die Größe in der Höhe auf 20 cm und in der Breite auf 28 cm festzulegen.



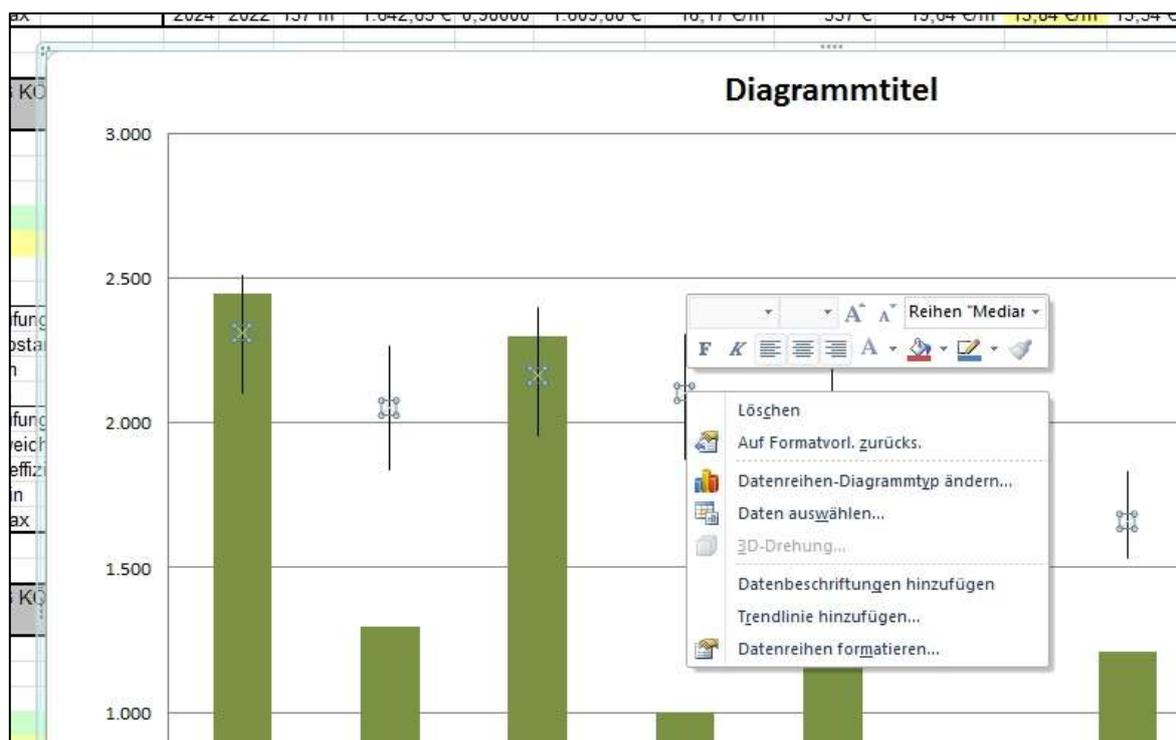
Nun klickt man die (hier grünen oder blauen) Balken mit der rechten Maustaste an, und wählt den Auswahlpunkt Füllung – hier hat sich „keine Füllung“ bewährt, dafür die Rahmenfarbe dunkelgrün ...



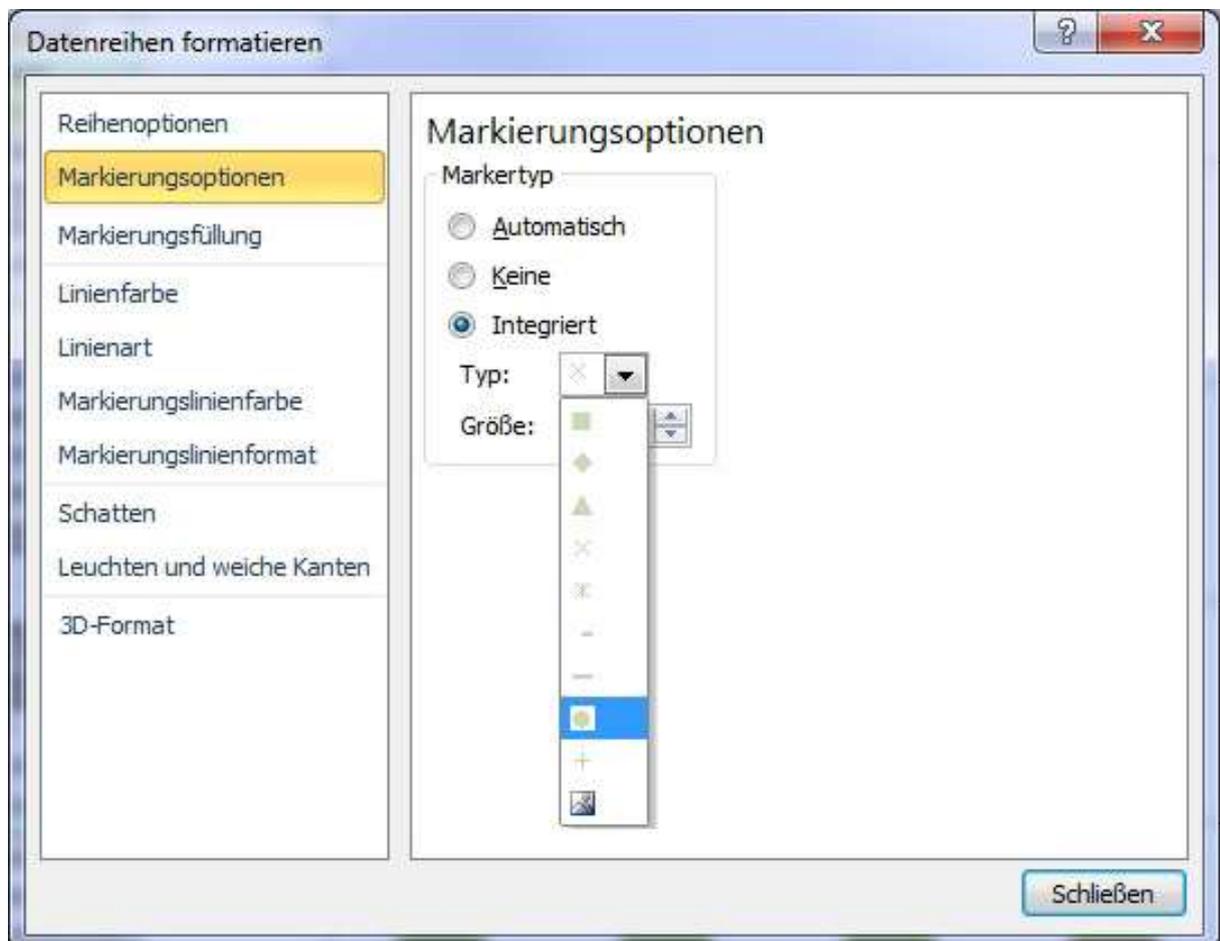
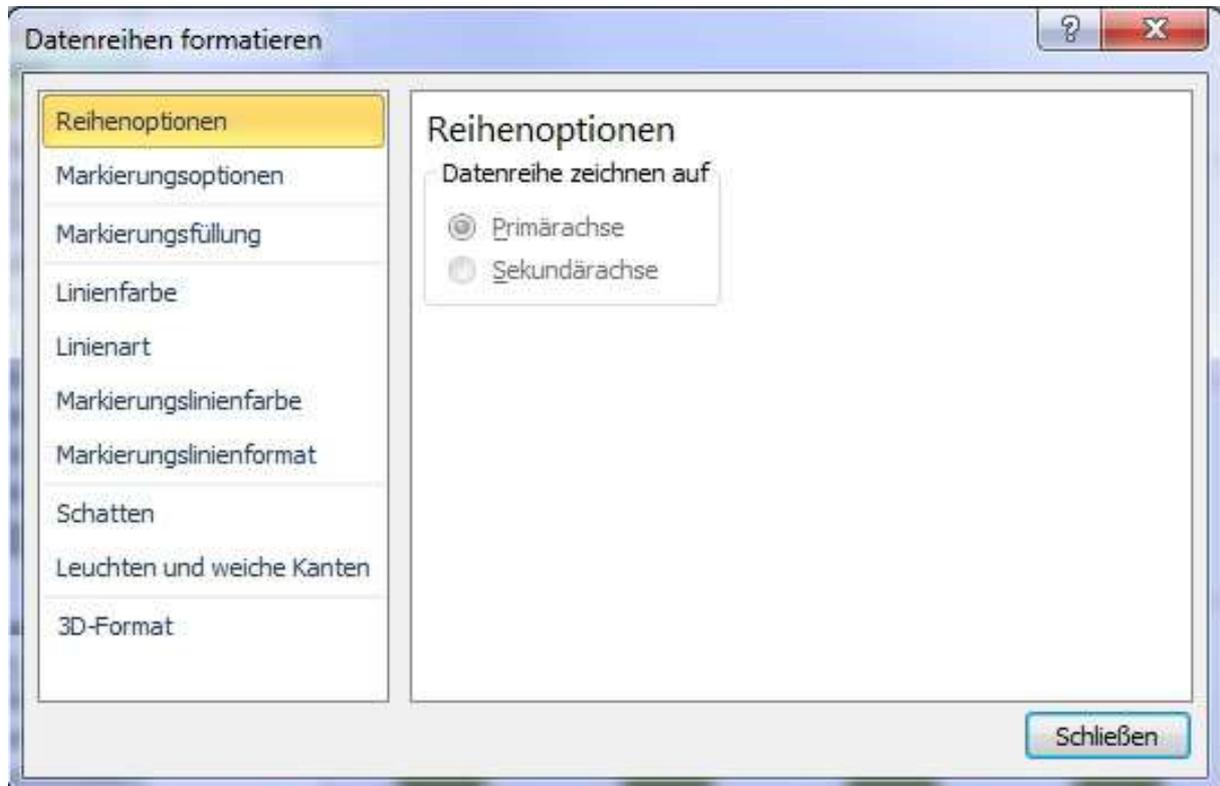
und die Rahmenart 1,5 bis 2 Pixel. Dadurch wirkt der Rahmen, der die linke Skala mit der Anzahl je Stadtbezirk darstellt, nicht so dominant, denn es kommt nicht so sehr auf die Anzahl der Datensätze an, sondern auf die Verteilung der Kauf- oder Mietpreise.



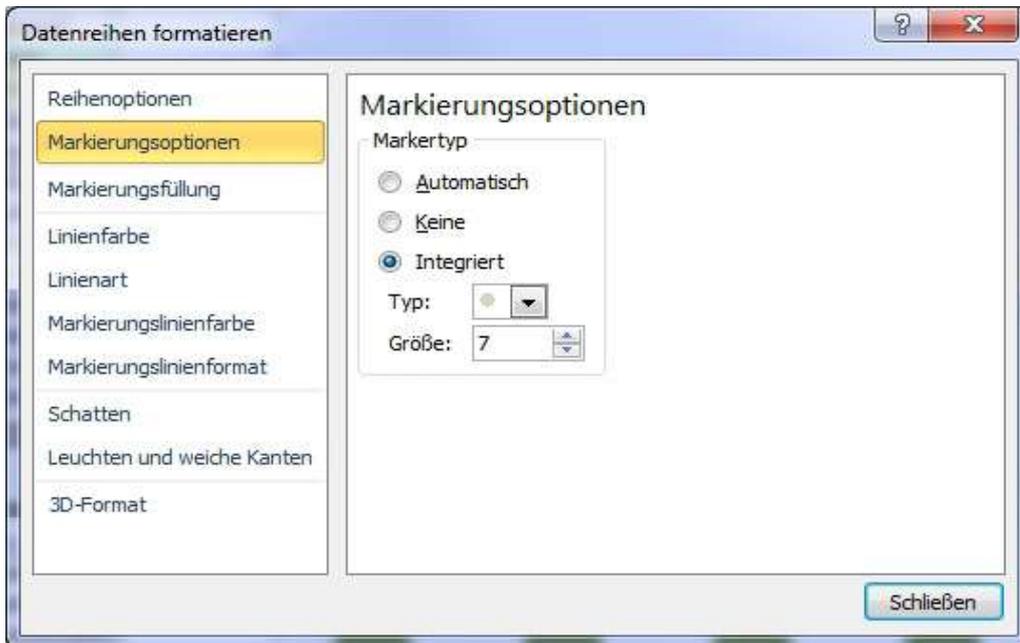
Jetzt klickt man mit der rechten Maustaste in die Mitte des Diagramms und kann im Kontextmenü den Diagrammbereich formatieren. Man klickt in einer der Spannweiten-Linien das schwach erkennbare „x“ für den Median an:



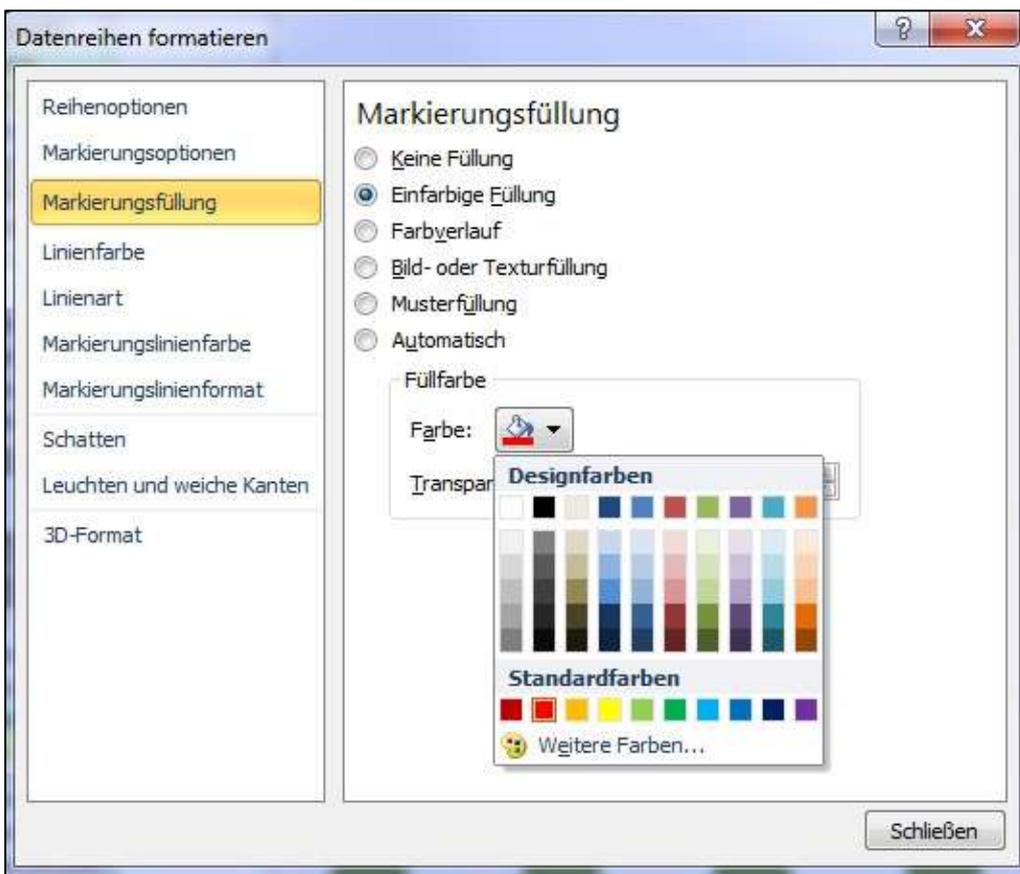
Durch Anklicken des Menüpunkts <Datenreihen formatieren> kommt man zum Menü für Markierungsoptionen:



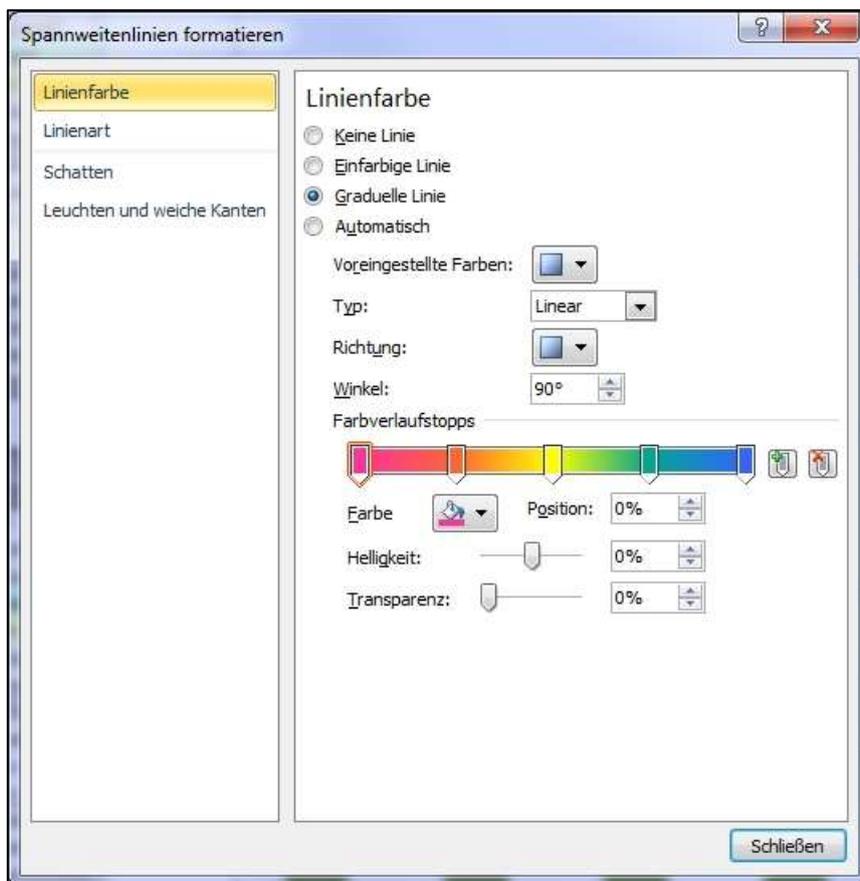
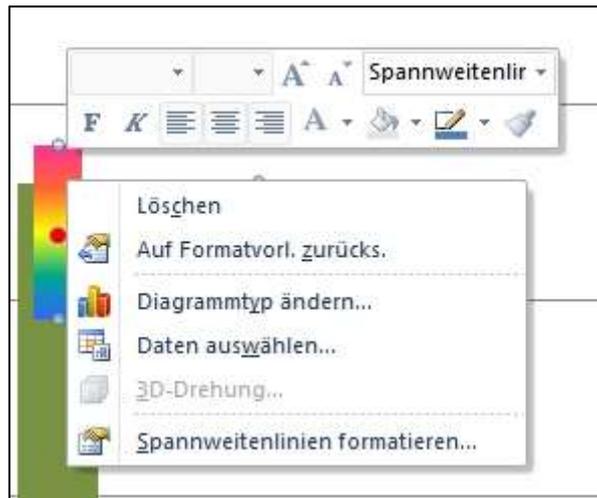
Hier wählt man die Option <Integriert> aus und für den Median als Typ den gefüllten Kreis und stellt die Größe auf 15 Pixel ein



Dann wählt man den Menüpunkt <Markierungsfüllung> aus, z.B. die Standard-Farbe feuerrot.



Dann klickt man auf eine der Spannweiten-Linien (der Wiskers), und öffnet dadurch das Menü <Spannweitenlinien formatieren>



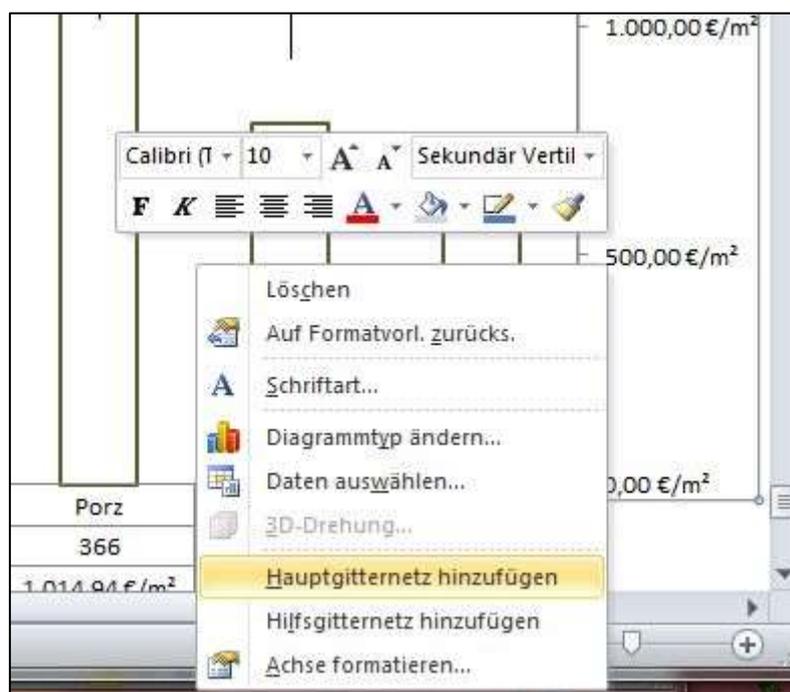
Hier kann man die Linienfarbe auswählen; bevorzugt die Option „graduierte Linie“, weil es so möglich wird, die Wertigkeit des unteren Schwellenwerts (das ist der rechte Bereich) blau, mit der Bedeutung „kalt, niedrig“ und den oberen Schwellenwert (der im linken Bereich liegt) weinrot, mit der Bedeutung „warm, heiß, hoch“ einzufärben. Für die mittleren Bereiche wählt man gelb, das passt gut zu dem roten Punkt des Medians.

Schließlich stellt man in der Linienart die Breite der Box ein. Hier z.B. 20 Pt.

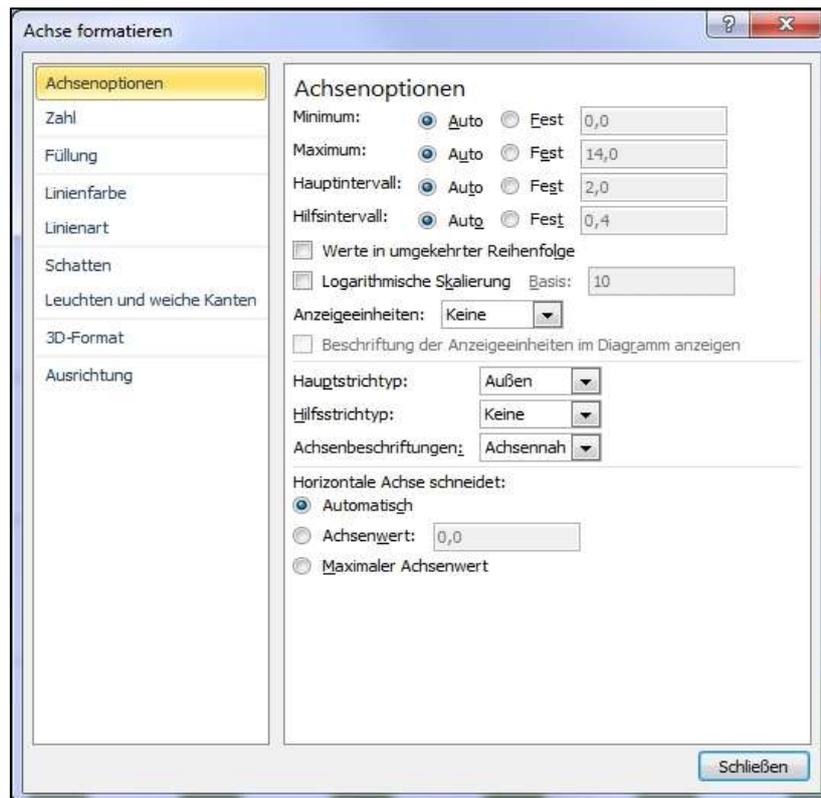
Nun klickt man im Diagramm auf die linke Skala, für die Anzahl, und klickt im Kontextmenü auf <Hauptgitternetz formatieren>.



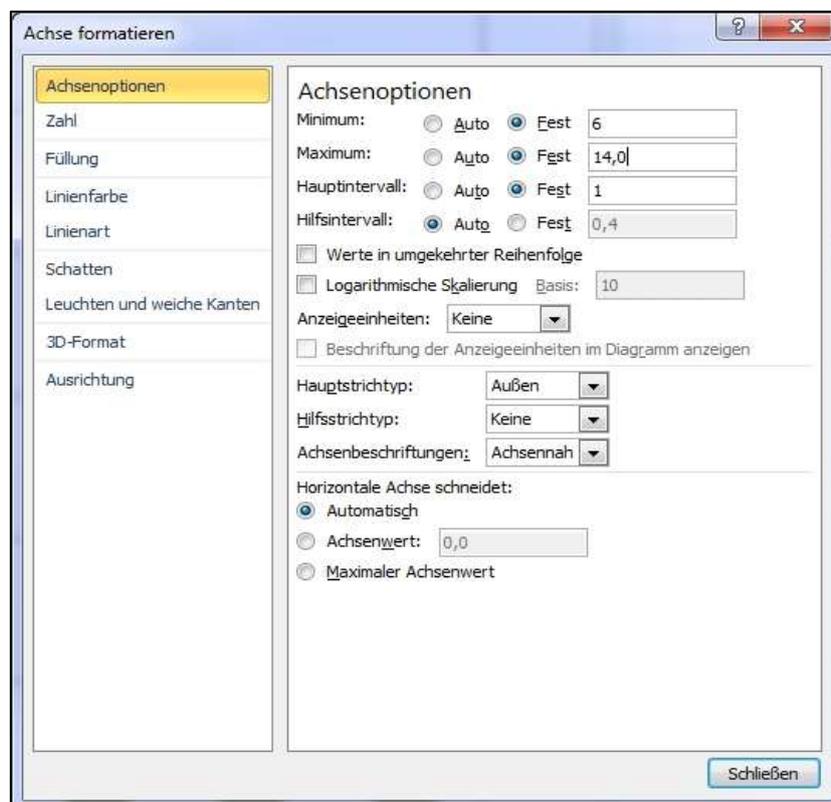
Hier wählt man die Option <keine Linie>, denn der Fokus der Darstellung liegt nicht auf der Anzahl, sondern auf den Kauf- bzw. Mietwerten. Die rechte Skala soll Hauptgitterlinien bekommen! Dazu klickt man mit der rechten Maustaste auf die rechte Skala und klickt im Kontextmenü <Hauptgitternetz hinzufügen> an.



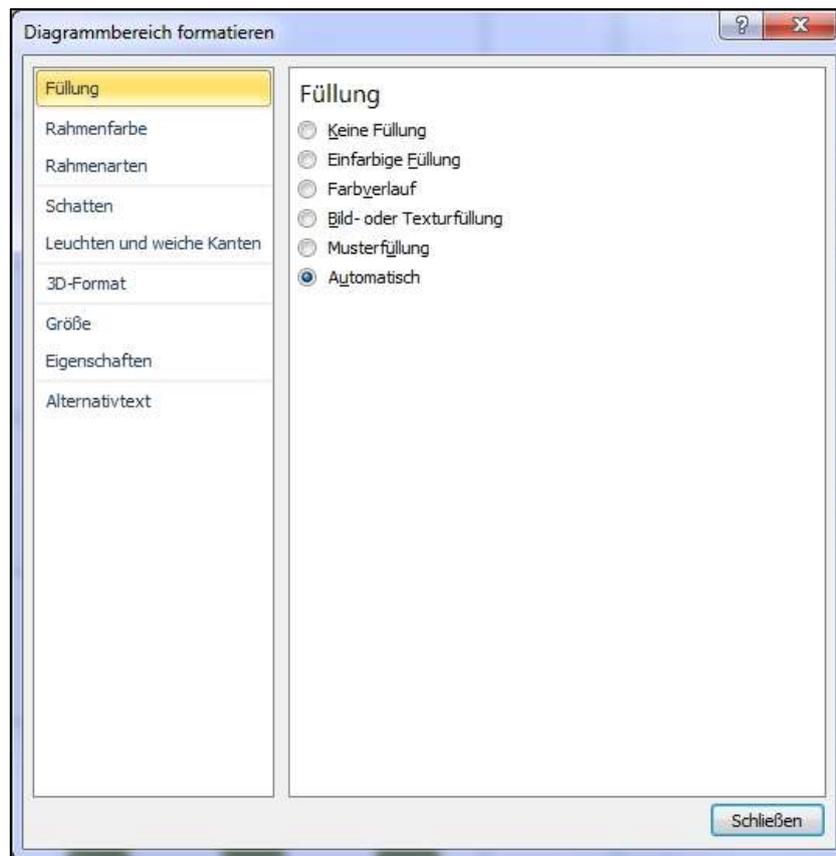
Nun klickt man im Diagramm auf die rechte Skala und wählt <Achse formatieren> aus. Insbesondere, wenn die Spannweiten hauptsächlich am oberen Rand des BoxPlots liegen, macht es Sinn, den Anzeige-Bereich erst (z.B. bei den Mieten) ab 4,00 €/m² oder 6,00 €/m² beginnen zu lassen. Wenn man mehrere ähnliche Auswertungen für unterschiedliche Städte etc. vergleichend in einer Dokumentation verarbeiten möchte, dann sollte man allerdings nicht von Diagramm zu Diagramm springen, sondern überall den gleichen Skalenbereich verwenden.



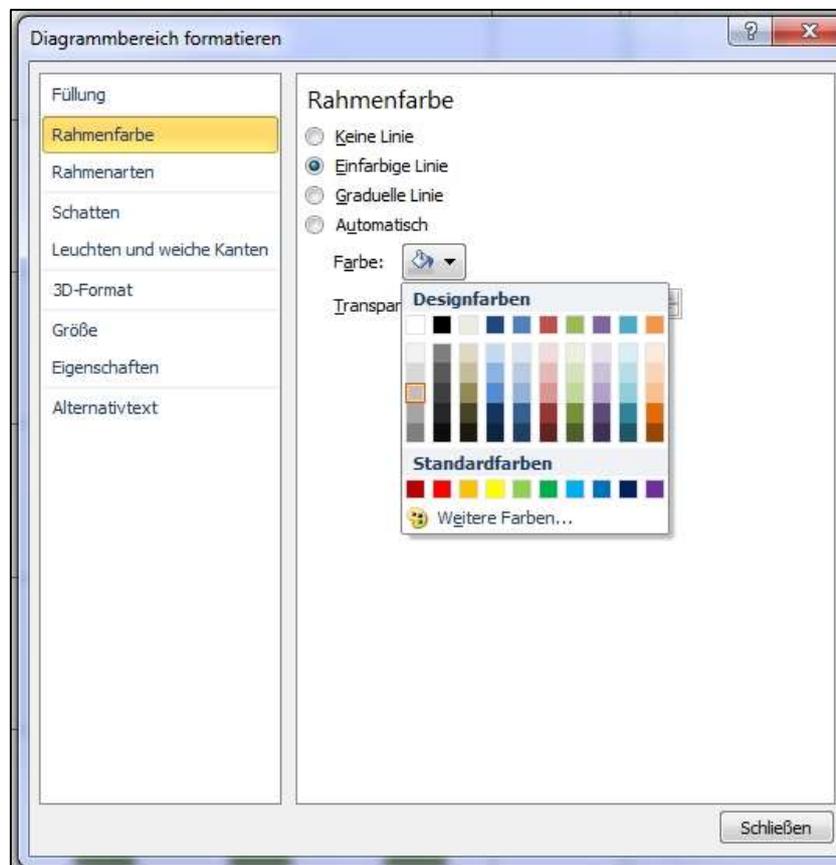
Um den sichtbaren Bereich möglichst groß darstellen zu können, setzt man das Minimum und das Maximum der Skala fest. Das Intervall stellt man auf z.B. 1, das bedeutet hier 1,00 €/m². Wenn man die Matrix vor Erstellung des Diagramms formatiert, wird diese Formatierung ins Diagramm automatisch übernommen.

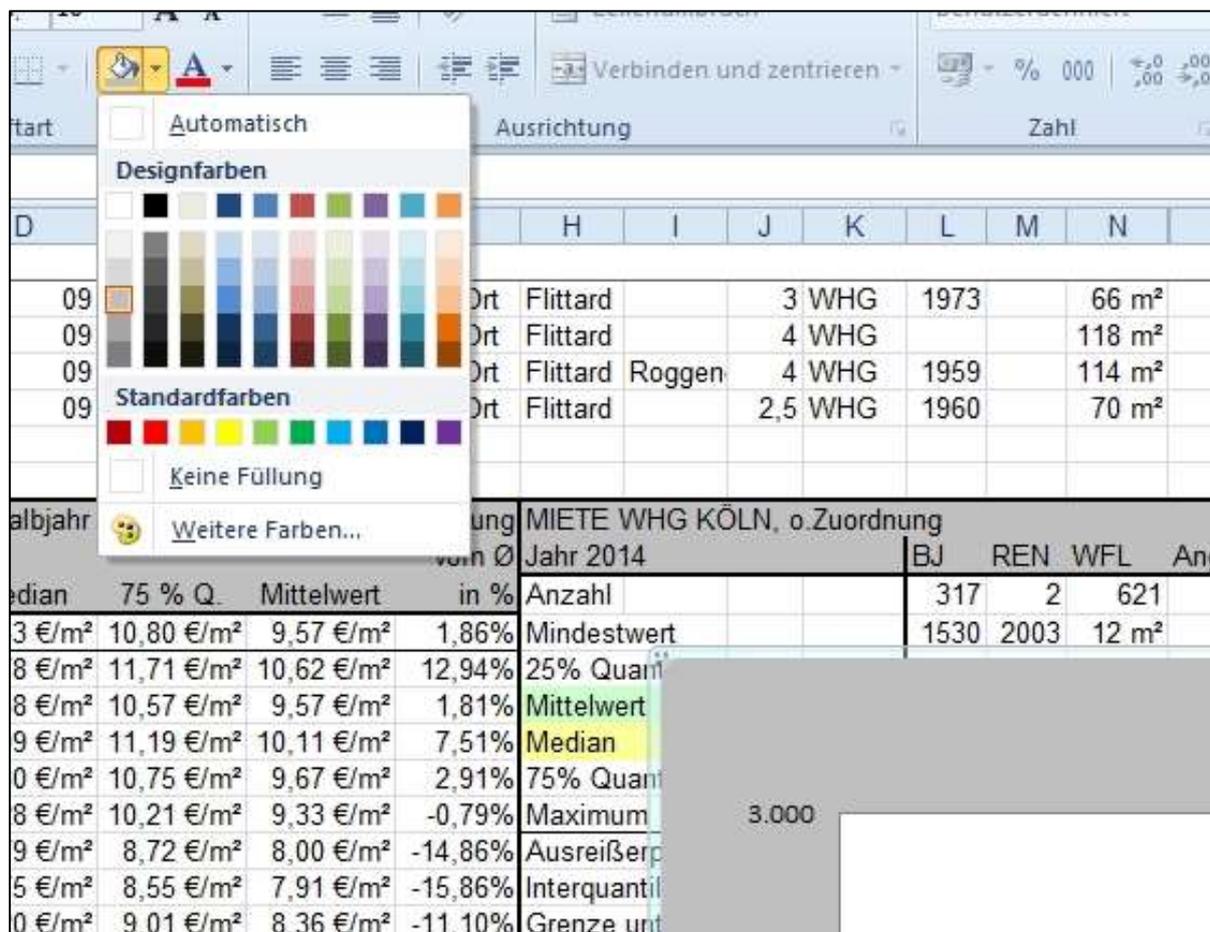
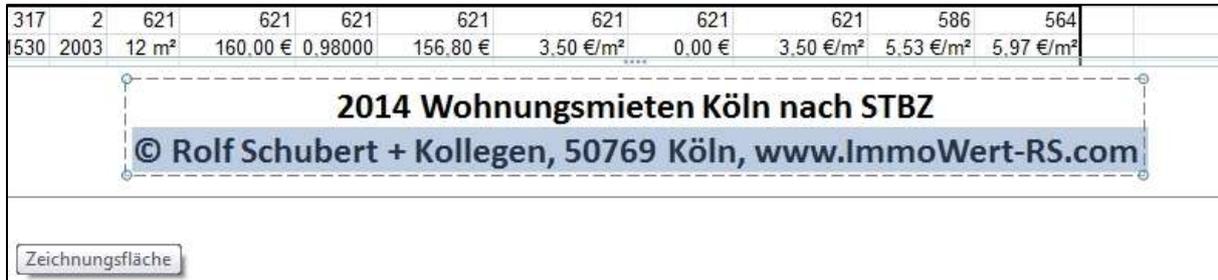


Nun klickt man auf den freien Rahmen des Diagramms um den Diagramm-Rahmen zu formatieren.

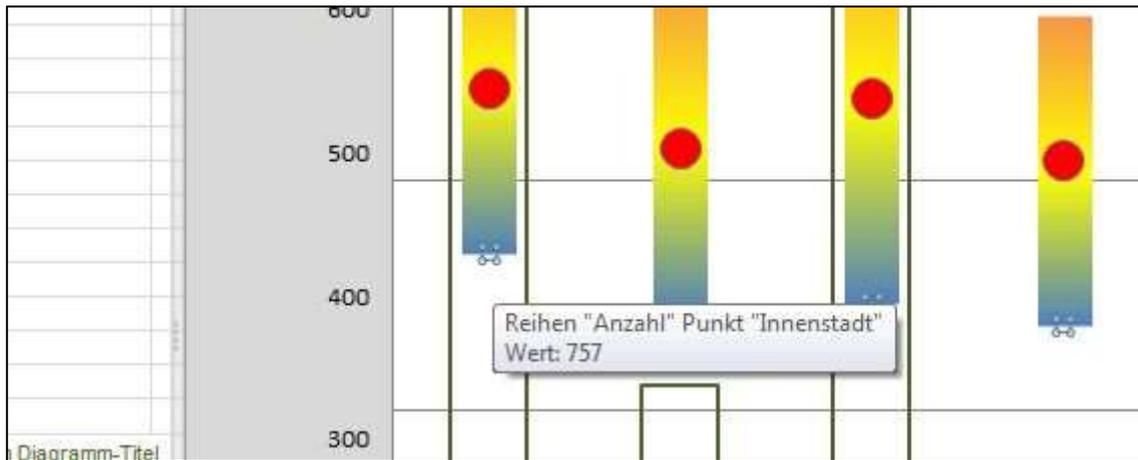


Man wählt ein helles Grau (25 %) aus.

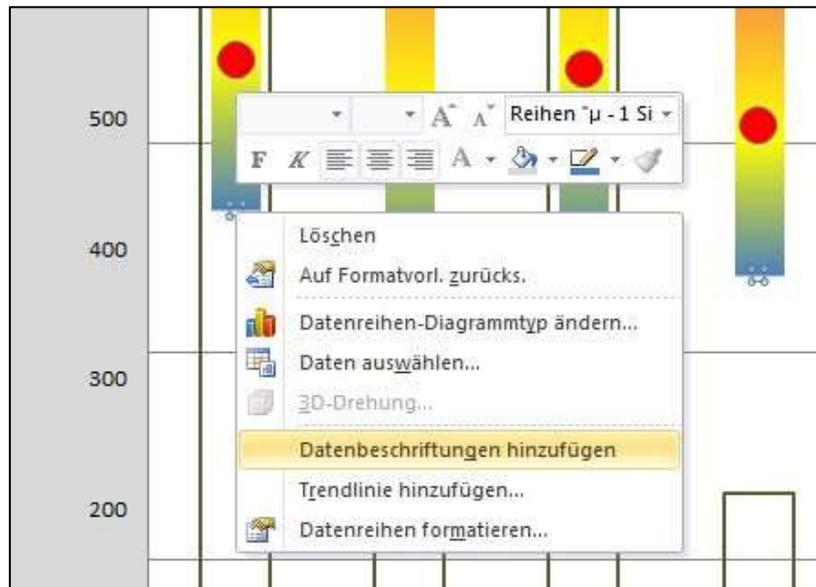




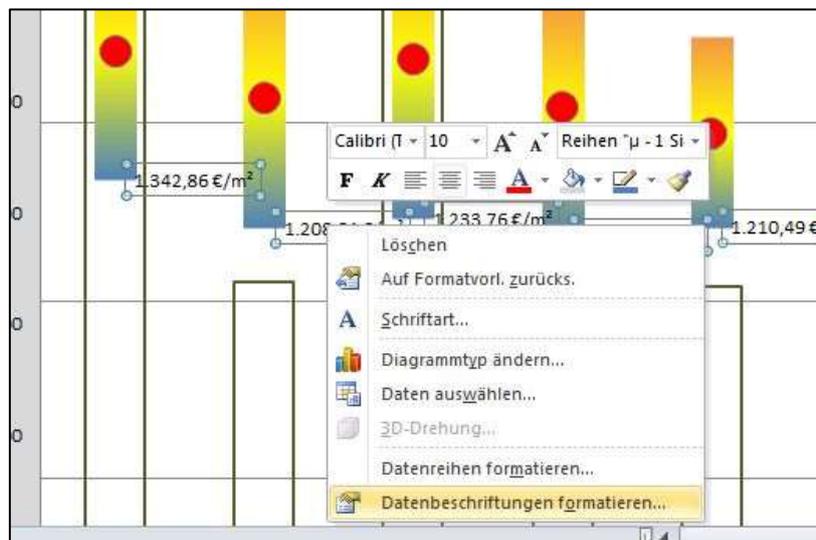
Man kann am unteren und dem oberen Ende der Wiskers noch eine Datenbeschriftung anbringen. Dazu klickt man mit der Maus auf den unteren Grenzwert der Box, bis vier kleine Punkte erscheinen.



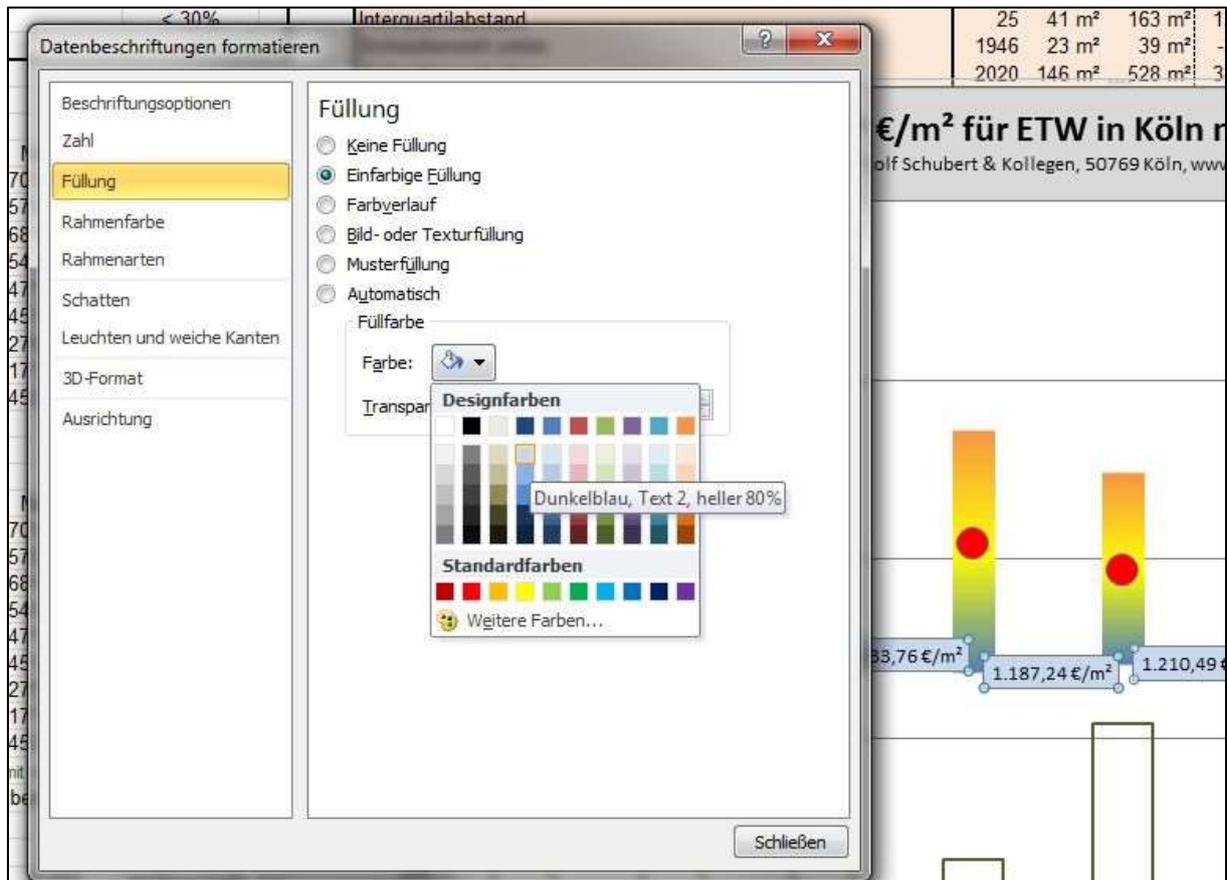
Nun klickt man mit der rechten Maustaste in das kleine Viereck und wählt im Kontextmenü <Datenbeschriftungen hinzufügen>:



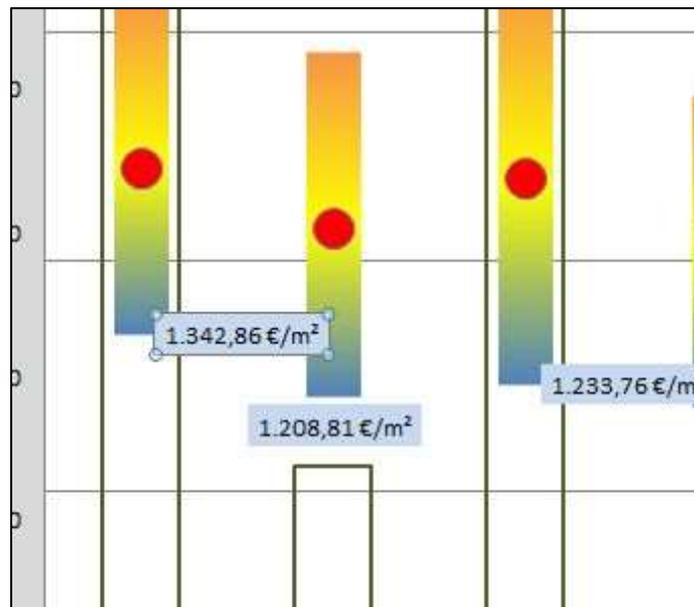
Unmittelbar danach setzt neben diesen Punkt Excel den unteren Schwellenwert des Stadtbezirks.



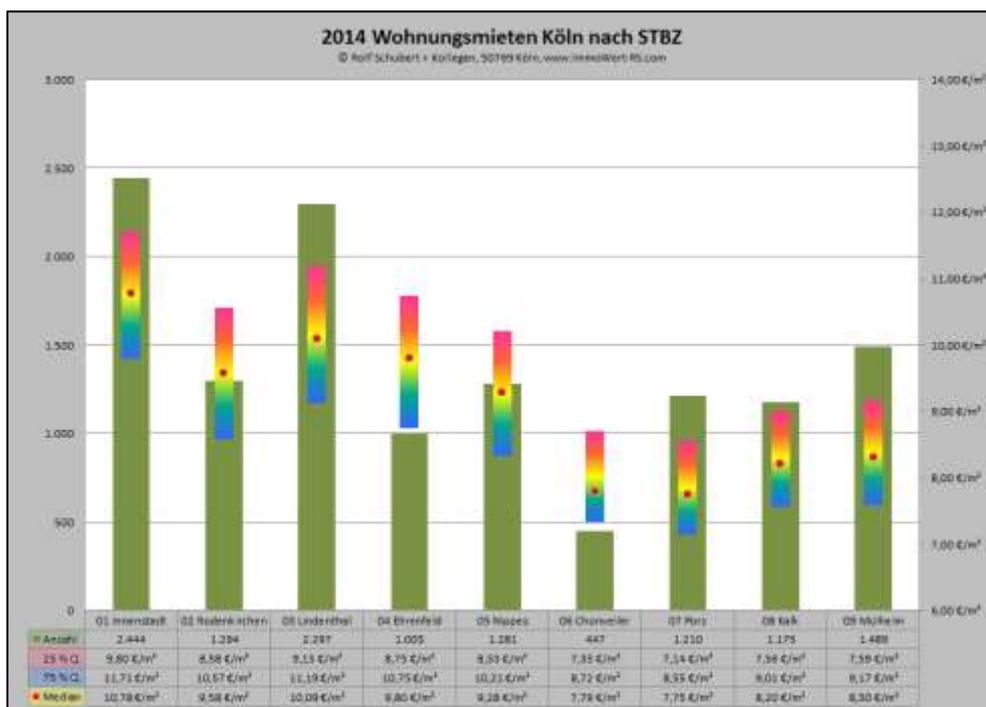
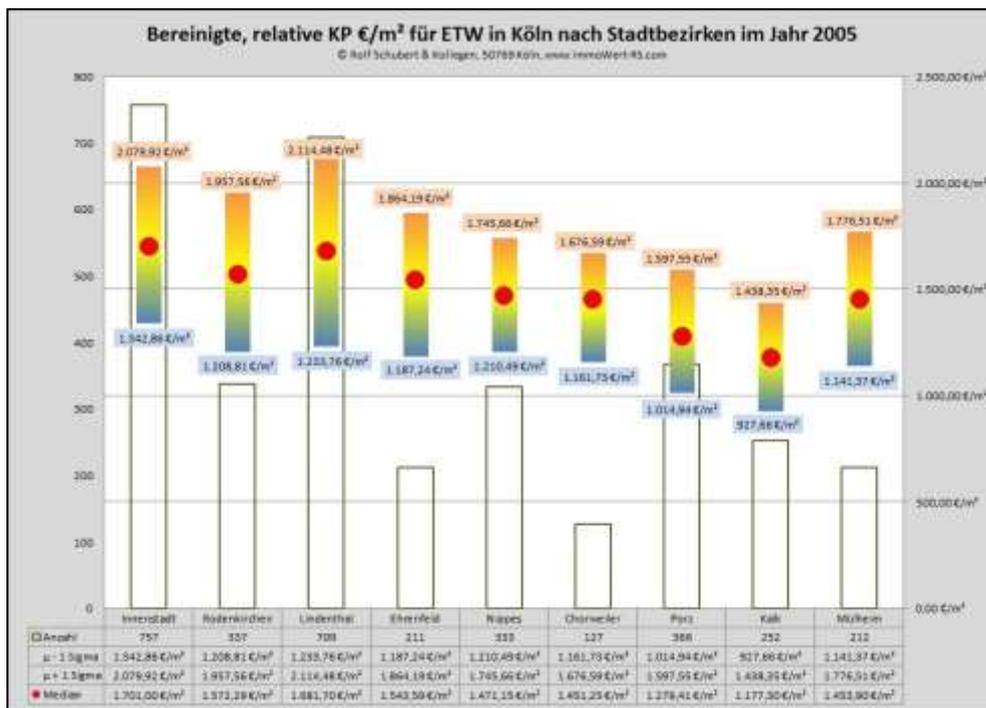
Diese Datenbeschriftungen werden automatisch mit der linken oberen Ecke auf die Mitte des Wiskers gesetzt und stehen damit immer rechts neben dem Wisker. Man kann die unteren bzw. oberen Datenbeschriftungen formatieren, indem man auf eine der Datenbeschriftungen klickt, dadurch werden alle mit einem Kasten markiert. Hier können Sie eine Hintergrundfarbe auswählen, in meinem Beispiel unten mit einem aufgehellten dunkelblau, oben mit einem aufgehellten rot.



Diese Datenbeschriftungen kann man nun jeweils einzeln doppelt anklicken und verschieben, so dass sie mittig unter der Box stehen:



Das macht man mit jeder Beschriftung. So sieht das fertige Boxplot aus:



Mit diesen Diagrammen kann man für ein Stadtgebiet in einer Grafik auch bei sehr vielen Datensätzen sowohl die Anzahl als auch den Median und die Spanne zwischen dem 25 % und dem 75 % Quantil oder die Spanne zwischen ± 1 Sigma, ggf. auch ± 2 Sigma grafisch darstellen und kann auf einen Blick alle Parameter gleichzeitig wiedergeben.

Man kann auch den Median mit einer Trendkurve unterlegen. Das macht aber nur Sinn, wenn zeitliche Aspekte in den Plot einfließen, d.h. wenn man eine zeitliche Entwicklung für einen STBZ darstellen möchte und nicht die absoluten oder relativen Mieten aller STBZ nebeneinander.