

Bonushoofdstuk

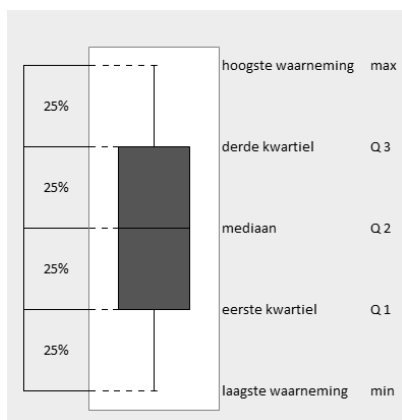
Een verticale boxplot bouwen

In mijn boek 'Excel voor professionals' leest u in hoofdstuk 15, 'Speciale grafieken maken', hoe u een horizontale boxplot maakt, zowel voor één groep als voor drie groepen gegevens. U kunt in Excel ook boxplots maken die rechtop staan. U vindt de uitleg voor deze verticale boxplots in dit bonushoofdstuk. U leest hoe u een verticale boxplot maakt, compleet met markeringen voor de uitschieters. Aan het eind geef ik aan dat u een eenvoudige boxplot ook kunt maken met de optie Omhoog/omlaag-balken (als u weinig eisen stelt).

■ Een boxplot bouwen

De boxplot is geen standaard ingebouwd grafiektype in Excel. Om die toch in Excel te maken, is een uitdaging die we graag aangaan. Voor de verticale boxplot gebruiken we de zogenoemde Gestapelde kolom. Dat houdt in dat we een aantal waarden op elkaar stapelen, die we niet zonder meer uit de tabel kunnen overnemen. We moeten de afstanden ertussen berekenen en die vervolgens als blokjes op elkaar zetten.

Bij een boxplot staat in het midden een rechthoek, de doos (een *box* in het Engels). Deze doos geeft de beide middelste kwartielen weer en loopt van Q1 tot en met Q3. De afstand tussen deze beide kwartielen wordt berekend met $Q3 - Q1$ en wordt de interkwartielafstand (IKA) genoemd; binnen deze grenzen liggen 50 procent van alle waarden.



■ **Afbeelding 1**

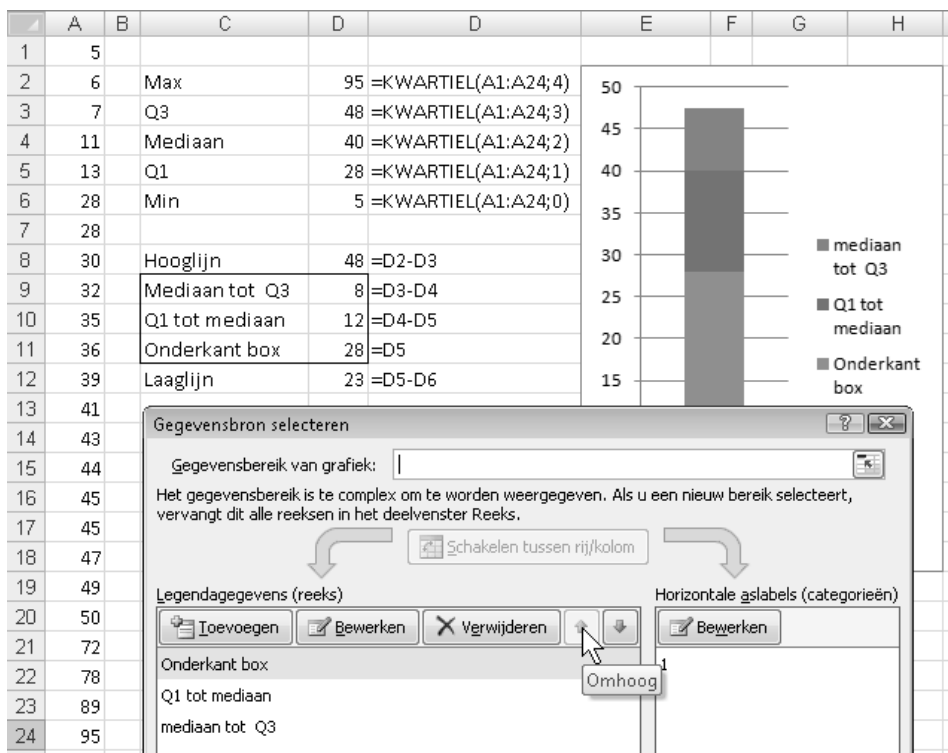
Een boxplot verdeelt de gegevens in vier gebieden. Zo ziet de boxplot eruit als de waarden precies gelijkmatig verdeeld zijn.

Vanaf de doos lopen er twee lijnen naar de uiteinden. Er loopt vanaf de onderkant van de box (Q1) een lijn naar de minimumwaarde, en van de bovenkant van de doos (Q3) loopt een lijn naar het maximum. Deze lijnen worden *whiskers* genoemd (snorharen, als van een kat). Eventuele uitschieters in de verzameling worden boven en onder deze lijnen als cirkels weergegeven.

Waarden voor de boxplot berekenen

Een boxplot (of doosdiagram) is een grafiek die van een serie getallen de vijfgetalensamenvatting weergeeft. Om een boxplot te kunnen maken, hebben we nodig:

- de hoogste waarde (maximum);
- het derde kwartiel (Q3); deze vormt de bovenkant van de rechthoek;
- de mediaan, deze wordt gemarkeerd door een lijn in de rechthoek;
- het eerste kwartiel (Q1); deze vormt de onderkant van de rechthoek;
- de laagste waarde (minimum).



■ **Afbeelding 2**

Voor een verticale boxplot neemt u de formules over en begint u met de grafiek Gestapelde kolom.

In dit voorbeeld gaan we uit van de bedragen die 24 klanten in een supermarkt hebben uitgegeven. Typ in de cellen A1 tot en met A24 willekeurige, hele bedragen tussen 5 en 95. Neem voor de vijfgetallensamenvatting de aanduidingen en de formules uit de afbeelding op de vorige pagina over in uw werkblad.

De rijen 2 tot en met 6 zijn voor de vijfgetallensamenvatting.

- Wij berekenen deze vijf waarden voor het gemak steeds met de functie KWARTIEL (in hoofdstuk 8 in het boek leest u dat de mediaan gelijk is aan het tweede kwartiel, het minimum gelijk is aan het nulde kwartiel en het maximum aan het vierde kwartiel). Gebruikt u hiervoor liever andere functies, dan neemt u de formules uit de volgende tabel over.

Rij	Formules in kolom D (alternatief)
2	=MAX(A1:A24)
3	=PERCENTIEL(A1:A24;0,25)
4	=MEDIAN(A1:A24)
5	=PERCENTIEL(A1:A24;0,75)
6	=MIN(A1:A24)

Gebruikt u deze werkmap ook met oudere versies van Excel, neem dan de functie KWARTIEL over (zonder toevoeging). Als de boxplot erg nauwkeurig moet zijn en u werkt uitsluitend in Excel 2010, dan kunt u beter de functie KWARTIEL.EXC gebruiken. Deze geeft het kwartiel gebaseerd op percentielwaarden van 0 tot en met 1. De functie KWARTIEL.INC is gelijk aan KWARTIEL, dus dat zou geen verschil moeten maken.

De verticale boxplot tekenen

Voor de verticale boxplot gebruiken we de Gestapelde kolom. Dat houdt in dat we een aantal waarden als een blokkentoren op elkaar stapelen, maar daarvoor kunnen we de waarden uit rij 2 tot en met 6 niet zonder meer gebruiken. We moeten de afstanden daartussen berekenen en die vervolgens op elkaar zetten. Hiervoor zijn de formules in rij 8 tot en met 12.

Selecteer de cellen C9 tot en met D11 (Mediaan tot Q3, Q1 tot mediaan en Onderkant box). Klik op **Invoegen**, klik op **Kolom** en kies onder **2D-kolom** de **Gestapelde kolom** (let op: kies niet de 100% gestapelde kolom). Er verschijnt een grafiek met drie kolommen naast elkaar. De tabs Hulpmiddelen voor grafieken staan ook boven in beeld. Klik hierin op de tab **Ontwerpen** en kies **Rijen/kolommen omdraaien**.

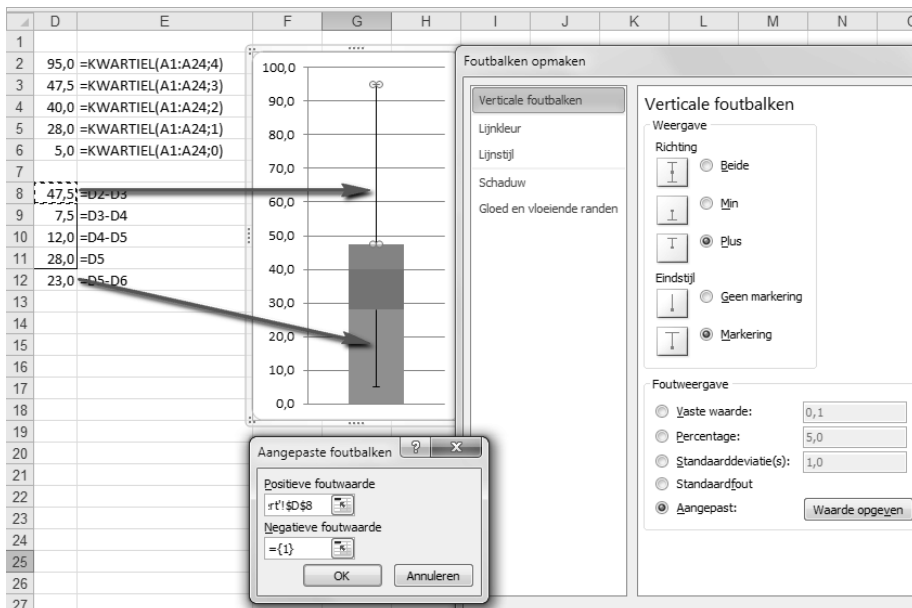
U hebt nu één kolom die uit drie delen bestaat. Alleen de volgorde klopt nog niet. Klik daarom op de grafiek en klik in de tab **Ontwerpen** op **Gegevens selecteren** (of rechtsklik op de grafiek en kies deze optie in het snelmenu). In het dialoogvenster dat verschijnt, ziet u de drie reeksen terug. Verander met behulp van de pijltjes

midden in dit venster de volgorde, zodat u hierin van boven naar beneden Onderkant box, Q1 tot mediaan en Mediaan tot Q3 ziet. In dit dialoogvenster staan ze precies andersom dan in de grafiek zelf (zie de legenda), maar zo komt het goed.

- Maak deze grafiek maar meteen wat smaller door de rechterraand naar binnen te slepen.
- Haal de categorieas onder aan de grafiek weg.
- Verwijder de legenda.

De whiskers aanbrengen

Het middelste en het bovenste vlak vormen straks de box. Hier vandaan trekken we de lijnen. De ene *whisker* moet van de onderkant van het middelste blok tot het minimum lopen. Klik hiervoor op het onderste deel van de kolom (de reeks Onderkant box), klik op de tab **Indeling**, klik op **Foutbalken** en dan op **Meer opties voor foutbalken**. In het venster dat opengaat, klikt u op de optie **Min**, op **Aangepast** en op **Waarde opgeven**. Er verschijnt een klein venster. Klik in het vak onder **Negatieve foutwaarde**, haal $\{1\}$ weg en klik op cel D12 (de waarde van de Laaglijn). Dit zorgt voor een lijn in de grafiek van de onderkant van het middelste vlak tot het minimum. De lengte van deze Laaglijn is het verschil tussen Q1 en de minimumwaarde (D5 minus D6). In ons voorbeeld is dat $28-5$ is 23. Vanaf Q1 (28) gaat deze lijn 23 omlaag en zo staat het dwarsstreepje precies op de minimumwaarde 5.



■ Afbeelding 3

Met zogenoemde foutbalken brengt u lijnen aan van de box naar de uiteinden.

Klik voor een lijn vanaf de box naar boven op het bovenste vlak (de reeks Mediaan tot Q3), klik weer op **Foutbalken** en op **Meer opties voor foutbalken**. Klik in het venster dat opengaat op de optie **Plus**, op **Aangepast** en op **Waarde opgeven**. Klik in het kleine venster in het vak onder Positieve foutwaarde, maak dat leeg en klik op cel D8 (de waarde van de Hooglijn). Zo krijgt u een lijn van de bovenkant van het vlak tot het maximum.

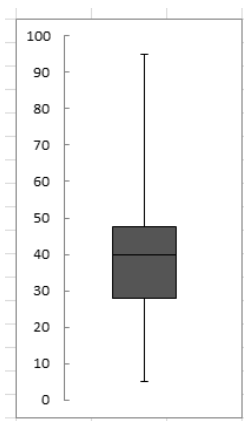
De vlakken aanpassen

Het onderste vlak van de kolom draagt de box, maar dit vlak hoeven we niet te zien. Om dat onzichtbaar te maken, klikt u op het onderste vlak, klikt u in de tab **Opmaak** op **Opvulling van vorm** en op **Geen opvulling**.

De beide vlakken die overblijven, vormen samen de box. Excel heeft deze automatisch twee verschillende kleuren gegeven (want het zijn twee verschillende waarden). Het is gebruikelijk om de beide delen van de box dezelfde kleur te geven. Klik hiervoor op het ene vlak, klik in de tab **Opmaak** op **Opvullen van vorm** en kies uw kleur. Herhaal dit voor het andere vlak (of klik erop en druk op Ctrl+Y).

Om de mediaan te kunnen zien, hoeven we alleen maar de rand om de beide vlakken van de box aan te brengen. Klik hiervoor op het ene vlak, klik in de tab **Opmaak** op **Omtrek van vorm** en op **Automatisch**. Herhaal dit voor het andere vlak.

- Wilt u op de linker as de cijfers van 0 tot en met 100 in stappen van 10 zien, rechtsklik dan op die as en kies **Waardeas opmaken**. In het venster dat verschijnt, kiest u onder Opties voor as bij Minimum, Maximum en Primaire eenheid de optie **Vast** en vult u respectievelijk 0, 100 en 10 in.
- Hoeft u de horizontale rasterlijnen niet te zien, klik dan op een van die lijnen en druk op de Delete-toets.



■ **Afbeelding 4**

En uw verticale boxplot is klaar. Zoals u ziet, ligt de helft van de uitgaven in de supermarkt tussen 28 en 48 euro. Aan de plaats van de mediaan in de box ziet u meteen of de verzameling scheef verdeeld is.

Uitschieters bepalen

Wat u misschien wel ziet aan de getallen zelf maar niet aan de boxplot, is dat er eigenlijk vier uitschieters zijn (ook wel uitbijters genoemd, in het Engels *outliers*). We kunnen de boxplot uitbreiden om de uitschieters ook weer te geven. De uitschieters worden opgespoord met de regel: 1,5 x interkwartielafstand (IKA, in het Engels IQR, *Inter Quartile Range*). Eerst wordt de interkwartielafstand genomen. Dat is de afstand tussen de beide kwartielen aan weerszijden van de mediaan, dus het verschil tussen Q3 en Q1 (en niet tussen Q3 en Q2, want Q2 is de mediaan). Deze afstand wordt met 1,5 vermenigvuldigd en vervolgens van Q1 afgetrokken om de ondergrens te bepalen en bij Q3 opgeteld om de bovengrens te bepalen. Alle waarden die buiten dit bereik vallen, zijn dan uitschieters.

Als de uitschieters worden weergegeven, loopt de onderste lijn niet meer van Q1 naar de minimumwaarde, maar van Q1 naar de nieuwe ondergrens (die ligt op de onderkant van de box minus 1,5 x de interkwartielafstand). En de bovenste lijn loopt nu niet meer van Q3 naar het maximum, maar van Q3 naar de nieuwe bovengrens (die ligt op de bovenkant van de box plus 1,5 x IKA).

Interkwartielafstand berekenen

We moeten de interkwartielafstand berekenen en de nieuwe grenzen. We nemen als voorbeeld weer de bestedingen in de supermarkt, die in A1 tot en met A24 staan. De formules uit de volgende afbeelding verschaffen de informatie voor de boxplot. Hebt u de verticale boxplot al gemaakt met de aanwijzingen uit de paragraaf *De verticale boxplot tekenen*, dan moet u alleen de formules op rij 2 en 6 veranderen en de formules vanaf rij 14 toevoegen.

	A	B	C	D	D	E
1	5				Formules in kolom D	<i>Betekenis:</i>
2	6	Max		77	=MIN(D16;MAX(A1:A24))	
3	7	Q3		48	=KWARTIEL(A1:A24;3)	↙
4	11	Mediaan		40	=KWARTIEL(A1:A24;2)	<i>interkwartielafstand</i>
5	13	Q1		28	=KWARTIEL(A1:A24;1)	↘
6	28	Min		5	=MAX(D17;MIN(A1:A24))	
7	28					
8	30	Hooglijn		29	=D2-D3	<i>Max-Q3</i>
9	32	Mediaan tot Q3		8	=D3-D4	<i>Q3-mediaan</i>
10	35	Q1 tot mediaan		12	=D4-D5	<i>mediaan-Q1</i>
11	36	Onderkant box		28	=D5	<i>Q1</i>
12	39	Laaglijn		23	=D5-D6	<i>Q1-min</i>
13	41					
14	43	1,5 * IKA		29	=(D3-D5)*1,5	<i>1,5 * interkwartielafstand</i>
15	44					
16	45	bovengrens uitschieters		77	=D3+D14	<i>Q3 + 1,5 * IKA, uitschieters boven Max</i>
17	45	ondergrens uitschieters		-1	=D5-D14	<i>Q1 - 1,5 * IKA, uitschieters onder Min</i>
18	47					

Afbeelding 5

Om de uitschieters weer te geven in de grafiek, verleggen we eerst de grenzen van de boxplot. Neemt u de formules maar over.

De formule in D14 berekent de interkwartielafstand, dat is het verschil tussen Q3 en Q1, en die wordt met 1,5 vermenigvuldigd, met de formule:

`=(D3-D5)*1,5`

In D16 wordt deze afstand bij Q3 opgeteld om de bovengrens te bepalen met

`=D3+D14`

De waarden daarboven zijn de uitschieters omhoog.

In D17 wordt deze afstand van Q1 afgetrokken om de ondergrens te bepalen met

`=D5-D14`

De waarden daaronder zijn de negatieve uitschieters.

De formule in D2 is veranderd. Deze regelt nu hoever de hooglijn loopt. Die gaat niet meer gewoon tot de maximumwaarde, maar tot het punt waarboven de waarden als uitschieters gelden. Maar als er geen uitschieter is hoger dan de maximumwaarde, dan gaat de lijn tot dat maximum. Deze formule zegt eigenlijk: als D16 kleiner is dan het maximum, neem dan D16; neem anders het maximum. Dat had gekund met

`=ALS(D16<MAX(A1:A24);D16;MAX(A1:A24))`

maar die hebben we ingekort tot

`=MIN(D16;MAX(A1:A24))`

Ook de formule in D6 is aangepast en geeft niet zonder meer de minimumwaarde. Deze is het spiegelbeeld van D2 en regelt hoever de laaglijn loopt. Die gaat niet meer tot de laagste waarde, maar de waarde waaronder de uitschieters omlaag beginnen. Maar als er geen uitschieter is lager dan de minimumwaarde, dan gaat de lijn tot dat minimum. Hier staat voluit: als D21 groter is dan het minimum, neem dan D21; neem anders het minimum. Hiervoor hebben we

`=ALS(D21>MIN(A1:A24);D21;MIN(A1:A24))`

ingekort tot

`=MAX(D21;MIN(A1:A24))`

Hebt u de boxplot gemaakt met de aanwijzingen in de vorige paragrafen, dan hoeft u de celverwijzingen van de grafiek niet te veranderen. De lengte van de foutbalken (de Min en de Max van de whiskers) haalt de grafiek nog steeds uit dezelfde cellen (namelijk D8 en D12), alleen worden die nu anders berekend.

Met de formules uit de tabel zal het dwarsstreepje van de bovenlijn iets verschoven zijn. Het streepje stond eerst bij het maximum van 95, maar staat nu op het punt waarboven de uitschieters beginnen, bij 77 in dit voorbeeld. Het streepje van de onderste lijn in dit voorbeeld is niet verschoven. Want pas onder -1 zou er sprake zijn van uitschieters; de laagste waarde is 5, dus dat was en blijft de ondergrens.

Uitschieters aan de enkele boxplot toevoegen

Om de uitschieters in de grafiek weer te geven, berekenen we die in een apart tabeltje. We zullen ze straks op een speciale manier in de grafiek plaatsen. Zouden we die nu meteen automatisch berekenen, dan zijn ze straks lastig terug te vinden om te bewerken. Daarom doen we het andersom: we typen eerst zelf een aantal waarden en pas als die goed in de grafiek staan, plaatsen we in die cellen de formules die de waarden automatisch berekenen.

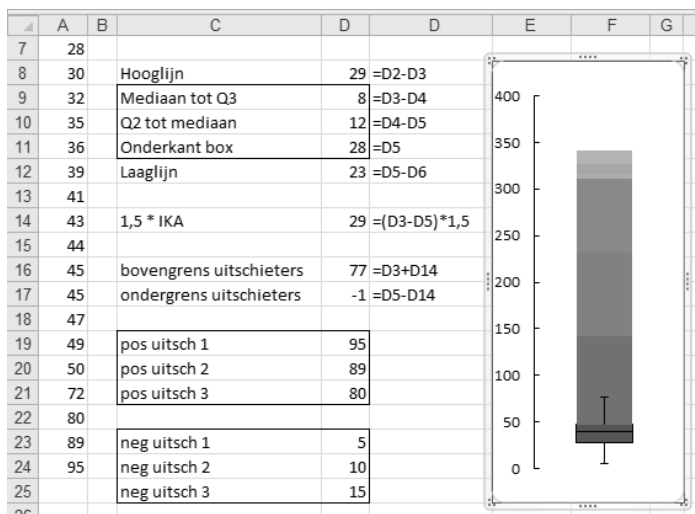
Typ daarom in D19, D20 en D21 drie verschillende getallen, die groter zijn dan het getal dat bij 'bovengrens uitschieters' staat. Typ ook in D23, D24 en D25 drie verschillende getallen, die kleiner zijn dan Q1 (maar neem nul niet).



Bekende uitschieters

Hebt u al een bestaande dataset, waarin ook de uitschieters bekend zijn, dan hoeft u die natuurlijk niet meer apart te berekenen. U typt dan meteen de waarden uit uw dataset in dit werkblad in kolom D.

Rechtsklik op de grafiek en klik op **Gegevens selecteren**; het venster Gegevensbron selecteren verschijnt. Klik op **Toevoegen**, laat in het venstertje dat opengaat, het vak Reeksnaam leeg, klik in het vak Reekswaarden, verwijder het $=\{1\}$ dat er staat, klik op cel D19 en klik één keer op **OK**. Klik weer op **Toevoegen**, klik in het vak Reekswaarden, maak dat leeg, klik op cel D20 en klik één keer op **OK**. Herhaal dit voor alle zes cellen. U hebt nu de reeksen 6 tot en met 11 erbij gemaakt. Sluit het venster Gegevensbron selecteren. Er is voor iedere reeks een vlak op de toren gestapeld; daar gaan we meteen iets aan doen.



Afbeelding 6

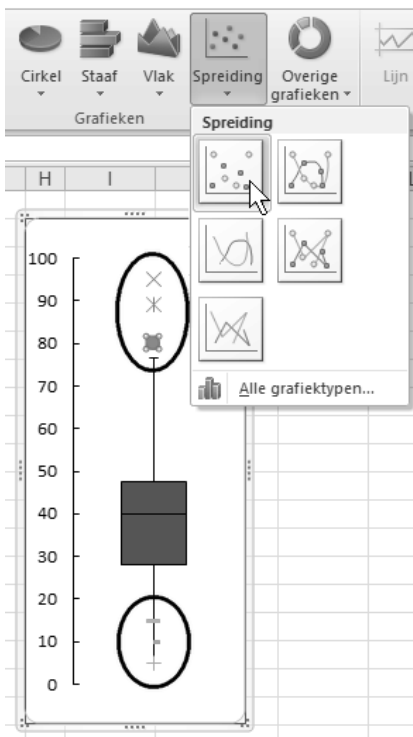
We zetten de uitschieters in een apart tabeltje bij elkaar. Ze komen voorlopig boven op de blokkentoren.

Om de uitschieters boven en onder de hoog- en laaglijn te krijgen, kiezen we voor hun waarden het type Spreidingsgrafiek; dat doen we in de bestaande grafiek van de box. Klik op een van de reeksen die u zojuist boven op de kolom hebt gekregen.

- Als u de muisaanwijzer op een van die vlakjes houdt, verschijnt er Reeks met een nummer. Klik daarop.
- Is een vlak zo klein dat u er niet gemakkelijk op kunt klikken, vergroot dan het beeld door de Ctrl-toets ingedrukt te houden en aan het scrollwielje van de muis te draaien. Lukt het nog niet, klik dan op de grafiek; de tabs Hulpmidde-len voor grafieken verschijnen. Klik op de tab **Opmaak**; helemaal links in beeld staat een keuzelijst. Selecteer om te beginnen **Reeks6** in de lijst.

Klik op de tab **Invoegen**, klik op de knop **Spreiding** en kies het eerste type: **Spreiding met alleen markeringen**.

Herhaal deze stappen voor alle zes vlakjes van de uitschieters. Hierna worden Reeks 6 tot en met 11 weergegeven met kleine markeringen. Dit zijn de uitschieters en die komen nu boven- en onderaan terecht.



■ **Afbeelding 7**

Om de uitschieters te tonen, geven we Reeks 6 tot en met 11 weer als Spreidingsgrafiek.

Uitschieters als cirkels weergeven

Het is in de statistiek de gewoonte om uitschieters met een cirkel weer te geven. Dat kan in Excel ook.

- Rechtsklik op een markering en kies **Gegevensreeks opmaken**; dit opent een venster. Kies in het tabblad **Markeringsopties** voor **Standaard** en kies als Type het rondje, met als grootte 7. Klik op **Markeringsopvulling** en kies **Geen opvulling**. Klik op **Kleur van markeringslijn**, kies **Ononderbroken streep** en kies via **Kleur** zwart.
- Klik nog niet op **Sluiten**, want u kunt gewoon op de volgende markering klikken terwijl dit venster openstaat en die dan met dit venster opmaken.

Selecteer de volgende markering en herhaal deze stappen voor alle zes reeksen.



■ **Afbeelding 8**

Via de opmaak geeft u de uitschieters in de boxplot weer als witte cirkels.

Staan de uitschieters goed in de grafiek, dan is het tijd om hun waarde automatisch te berekenen. Neemt u de volgende formules over in uw werkblad. Deze vervangen de waarden die u daar zopas zelf had getypt.

Rij	Kolom C	Formules in kolom D
19	pos uitsch 1	=ALS(GROOTSTE(A1:A24;1)>D16;GROOTSTE(A1:A24;1);NB())
20	pos uitsch 2	=ALS(GROOTSTE(A1:A24;2)>D16;GROOTSTE(A1:A24;2);NB())
21	pos uitsch 3	=ALS(GROOTSTE(A1:A24;3)>D16;GROOTSTE(A1:A24;3);NB())
22		
23	neg uitsch 1	=ALS(KLEINSTE(A1:A24;1)<D17;KLEINSTE(A1:A24;1);NB())
24	neg uitsch 2	=ALS(KLEINSTE(A1:A24;2)<D17;KLEINSTE(A1:A24;2);NB())
25	neg uitsch 3	=ALS(KLEINSTE(A1:A24;3)<D17;KLEINSTE(A1:A24;3);NB())

De formule in D19 zegt: als de hoogste waarde van A1 tot en met A24 groter is dan de bovengrens voor de uitschieters (die in D16 staat), geef dan de hoogste waarde; geef anders NB(), dat staat voor 'niet beschikbaar'. Zouden we aan het eind van de formule een nul of lege tekst opgeven, dan zou dat in de grafiek een markering op de nullijn opleveren. Maar door deze NB() geeft de grafiek niets weer (zie ook hoofdstuk 14, de paragraaf *Lege cellen niet weergeven met NB()*).

De formules in D20 en D21 kijken of er eventueel een tweede en derde uitschieter is.

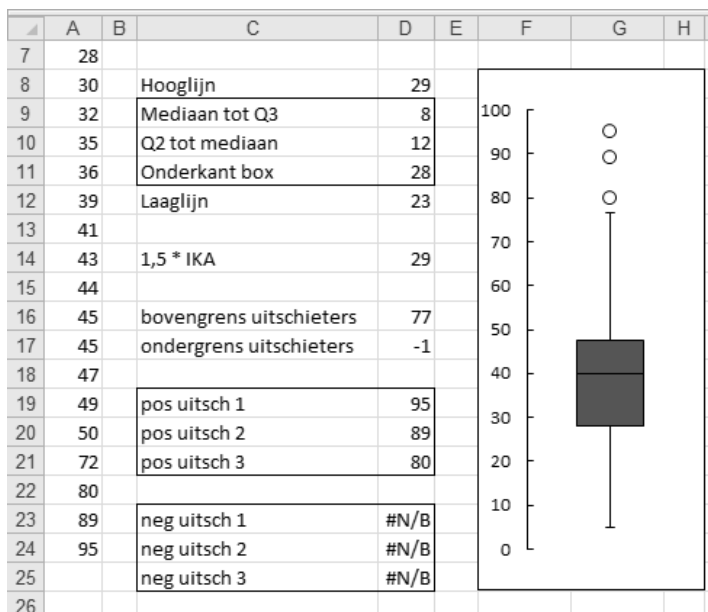
De formules in de onderste helft van de tabel kijken of de laagste drie waarden van A1 tot en met A24 kleiner zijn dan de ondergrens voor de uitschieters (die grens staat in D17).

Als er in de gegevens waarden voorkomen die buiten deze grenzen vallen, worden ze weergegeven als uitschieters. Met de getallen in het voorbeeld ziet u mooi de cirkels voor de bovenste uitschieters 95, 89 en 80 (die steken boven de grens van 77 uit). We hebben geen waarden kleiner dan de ondergrens van -1, dus daar zien we geen cirkels; in het staatje met formules staat daarvoor drie keer #NB.

In dit voorbeeld zijn wij berekend op drie positieve en drie negatieve uitschieters. Komen er bij u meer uitschieters voor, dan:

- maakt u in uw werkblad ruimte voor langere lijstjes;
- typt u eerst zelf getallen groter dan de bovengrens dan wel de ondergrens van de uitschieters;
- voegt u die cellen toe aan de grafiek;
- geeft u de uitschieters weer als een spreidingsgrafiek;
- maakt u ze op als cirkels;
- kopieert u voor de vierde positieve uitschieter de formule en past u deze zo aan dat er twee keer GROOTSTE(A1:A24;4 staat, voor de vijfde komt er na de puntkomma een 5 enzovoort;

- moet voor de vierde negatieve uitschieter in het deel `KLEINSTE(A1:A24;3` van de formule na de puntkomma een `4` staan (twee keer), voor de vijfde komt na de puntkomma een `5` enzovoort.



■ **Afbeelding 9**

De uitschieters worden automatisch berekend en weergegeven als witte cirkels. Deze uitgebreide boxplot geeft door de uitschieters een beter beeld van de werkelijkheid.

■ Groepen vergelijken met een boxplot

Boxplots zijn niet alleen handig om de verdeling van een reeks gegevens uit te beelden, maar ook om meer reeksen te vergelijken. Zo kunt u bijvoorbeeld de besteedde bedragen in drie supermarkten naast elkaar in een grafiek zetten, als drie verticale boxplots. In het voorbeeld staan de uitgaven van Winkel A, B en C in de cellen A1 tot en met C100. Hiervoor zetten we weer de vijfgetallensamenvatting onder elkaar en berekenen we enkele afstanden daartussen die we in een Gestapelde kolom weergeven.

Neem hiervoor de aanduidingen over uit kolom E van de afbeelding. Plaats de formules in kolom F (voor Winkel A) en kopieer deze naar kolom G en H voor de beide andere supermarkten.

De grafiek maakt u als volgt. Selecteer van de onderste tabel de middelste drie items, in dit voorbeeld zijn dat E9 tot en met H11. Klik op **Invoegen**, klik op **Kolom** en kies onder **2D-kolom** de **Gestapelde kolom**. Er verschijnt een grafiek met drie blokkentorens naast elkaar.

	A	B	C	D	E	F	F	G
1	Winkel A	Winkel B	Winkel C			Winkel A	Formules in kolom F	
2	1	22	8	Max	14,00	=KWARTIEEL(A1:A100;4)		
3	2	25	13	Q3	10,75	=KWARTIEEL(A1:A100;3)		
4	3	34	14	Mediaan	7,50	=KWARTIEEL(A1:A100;2)		
5	4	35	16	Q1	4,25	=KWARTIEEL(A1:A100;1)		
6	5	41	23	Min	1,00	=KWARTIEEL(A1:A100;0)		
7	6	41	26					Uitleg:
8	7	46	28	Hooglijn	3,25	=F2-F3		Max-Q3
9	8	46	33	Mediaan tot Q3	3,25	=F3-F4		Q3-Mediaan
10	9	47	39	Q1 tot mediaan	3,25	=F4-F5		Mediaan-Q1
11	10	49	61	Onderkant box	4,25	=F5		Q1
12	11	54	72	Laaglijn	3,25	=F5-F6		Q1-Min
13	12	54	86					
14	13	59	100					
15	14	60	115					

■ **Afbeelding 10**

Voor de vijfgetallensamenvatting van drie groepen en enkele afstanden daartussen, hebt u deze formules nodig.

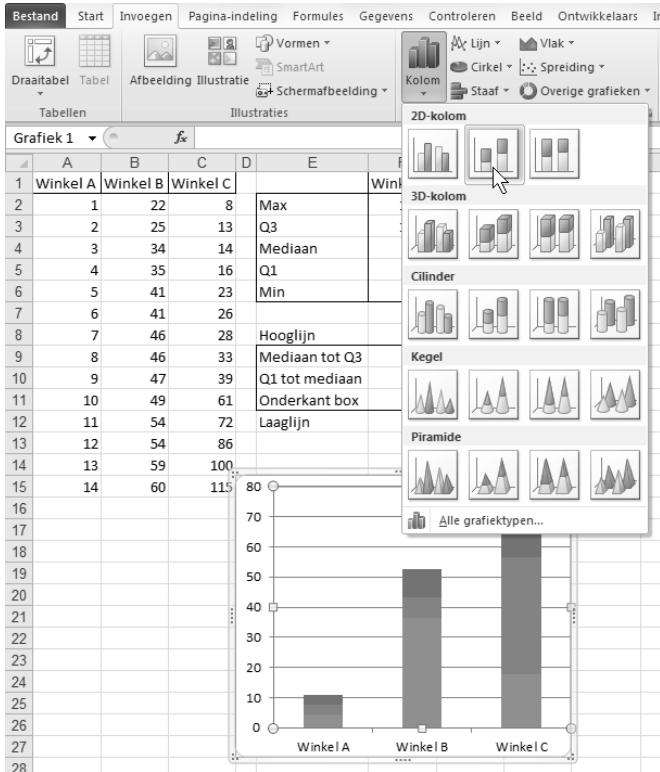
De volgorde (hoog-laag) van de vlakken moeten we even omkeren. Klik hiervoor op de grafiek en klik in de tab **Ontwerpen** op **Gegevens selecteren** (of kies dat na een rechtermuisklik); in het dialoogvenster dat verschijnt, ziet u de drie reeksen terug. Verander met behulp van de pijltjes de volgorde zo, dat u van boven naar beneden Onderkant box, Mediaan tot Q3 en Q1 tot mediaan in dit dialoogvenster ziet. In dit venster staan ze precies andersom dan in de grafiek zelf (zie de legenda), maar zo komt het goed.

Het is wel zo helder om de namen van de winkels in de grafiek te zien. Nu u dit dialoogvenster toch in beeld hebt, kan dat hiermee. Klik hiervoor onder Horizontale aslabels op **Bewerken**; er verschijnt een venster waarin u het Aslabelbereik opgeeft. Sleep hiervoor over de reeksnamen in F1 tot en met H1. Onder iedere staaf verschijnen de aanduidingen Winkel A enzovoort.

- Maak de grafiek breder en minder hoog door aan de randen te slepen.
- Wilt u de staven breder hebben, rechtsklik dan op een van de boxen, kies **Gegevensreeks opmaken** en kies onder Opties met het schuifje onder Breedte tussenruimte bijvoorbeeld **30%**. Vindt u het niet makkelijk om daarmee te schuiven, typ dan gewoon 30 in het vakje eronder.
- En verwijder de legenda.

Bonushoofdstuk

Een verticale boxplot bouwen



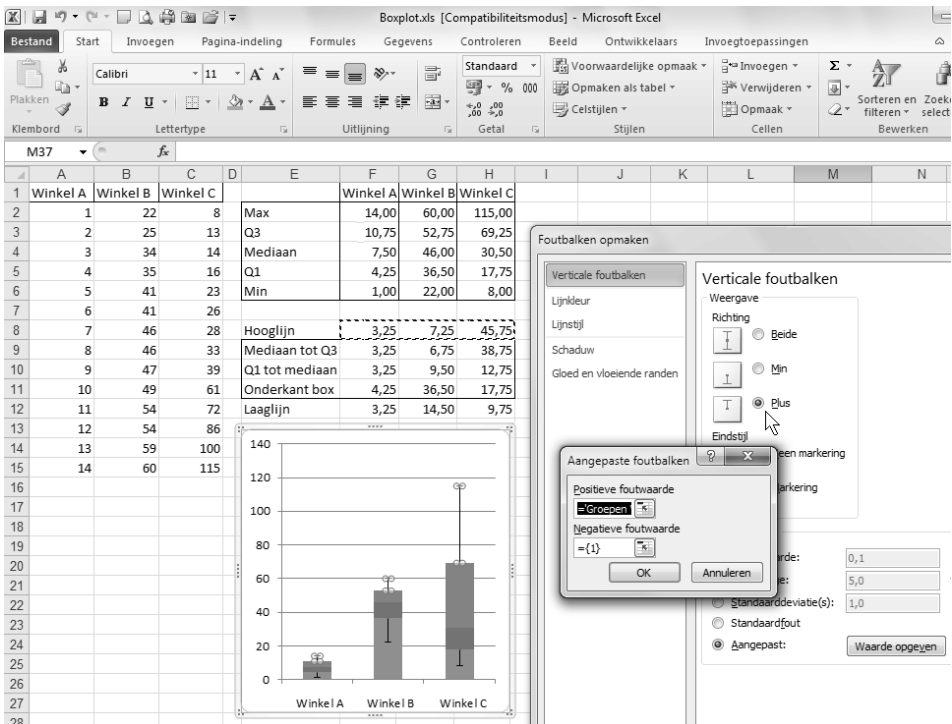
■ Afbeelding 11

Om groepen met elkaar te vergelijken in een verticale boxplot, nemen we de Gestapelde kolom.

De hoog- en laaglijnen trekken

De lijnen van de box tot het minimum maken we voor alle kolommen tegelijk. Klik op het onderste deel van een van de kolommen (de reeks Onderkant box), klik op de tab **Indeling**, op **Foutbalken** en dan op **Meer opties voor foutbalken**. In het venster dat opengaat, klikt u op de optie **Min**, op **Aangepast** en op **Waarde opgeven**; er verschijnt een klein venster. Klik in het vak onder Negatieve foutwaarde, haal ={1} weg en sleep over de cellen F12 tot en met H12 (de waarden van de Laaglijn). U krijgt hiermee in alle drie kolommen een lijn van de onderkant van het middelste blok tot de minimumwaarden.

Voor de lijnen vanaf de box omhoog klikt u op een van de bovenste vlakken (de reeks Mediaan tot Q3), dan weer op **Foutbalken** en op **Meer opties voor foutbalken**. Klik in het venster dat opengaat, op de optie **Plus**, op **Aangepast** en op **Waarde opgeven**. Klik in het kleine venster in het vak onder Positieve foutwaarde, maak dat leeg en sleep over de cellen F8 tot en met H8 (de waarden van de Hooglijn). Zo krijgt elke kolom een lijn van de bovenkant tot zijn maximum.



■ **Afbeelding 12**

U maakt de foutbalken van de box naar de uiteinden voor alle groepen tegelijk.

Alleen de boxen tonen

Elke box wordt gedragen door het onderste vlak van de kolom, maar die hoeven we niet te zien. Klik hiervoor op een van de onderste vlakken, klik in de tab **Opmaak** op **Opvulling van vorm** en op **Geen opvulling**.

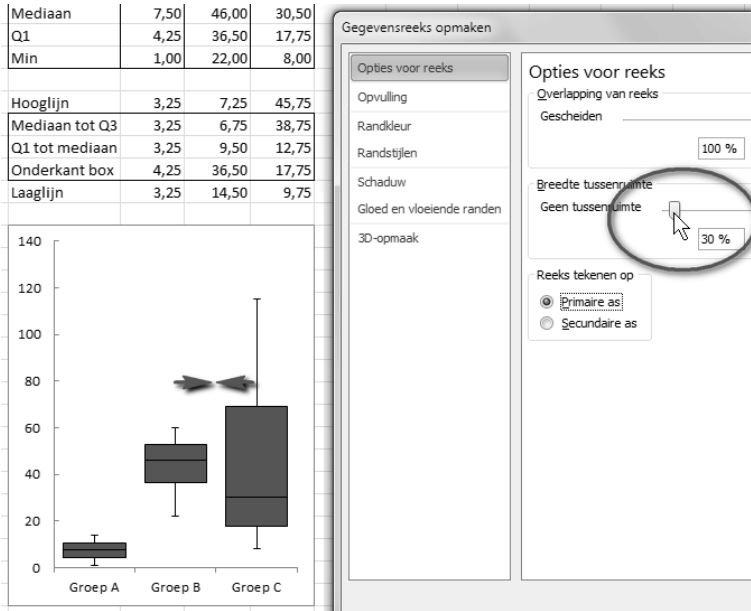
Om hierna de mediaan te kunnen zien, moeten we nog even de rand om de vlakken van de box aanbrengen. Klik hiervoor op een van de bovenste vlakken, klik in de tab **Opmaak** op **Omtrek van vorm** en op **Automatisch**. Klik op een van de andere vlakken en herhaal dit.

- U kunt de horizontale rasterlijnen verwijderen en de beide helften van de box dezelfde kleur geven. Volg hiervoor de aanwijzingen uit de tweede helft van de paragraaf *De whiskers aanbrengen*.
- Wilt u de boxen iets verbreden? Rechtsklik dan op een van de boxen, kies **Gegevensreeks opmaken** en kies onder Opties met het schuifje onder Breedte tussenruimte bijvoorbeeld **30%**. Vindt u het niet makkelijk om daarmee te schuiven, typ dan 30 in het vakje eronder.

U kunt de verzamelingen met gegevens nu in een oogopslag met elkaar vergelijken.

Bonushoofdstuk

Een verticale boxplot bouwen



■ Afbeelding 13

Even afwerken en uw verticale boxplots zijn klaar.

Uitschieters bij de groepen aanbrengen

Evenals in een boxplot met één reeks, kunt u in deze boxplot met meer reeksen de uitschieters laten zien. Voor uitleg van de uitschieters verwijzen we naar de paragrafen *Uitschieters bepalen* en *Interkwartielafstand en grenzen berekenen* eerder in dit hoofdstuk. U neemt in kolom E de aanduidingen uit de afbeelding over. Neem ook de formules in kolom F over en kopieer die naar kolom G en H.

- Of ga verder met het voorbeeld van de enkele verticale boxplot waarin de getallen in A1 tot en met A24 staan en waarin u de uitschieters hebt weergegeven. Voeg daarin links van kolom B twee extra kolommen in. Die zullen dan overeenkomen met de formules in onderstaande afbeelding. Kopieer de formules van kolom F opzij, naar kolom G en H.

Uitschieters in de grafiek plaatsen

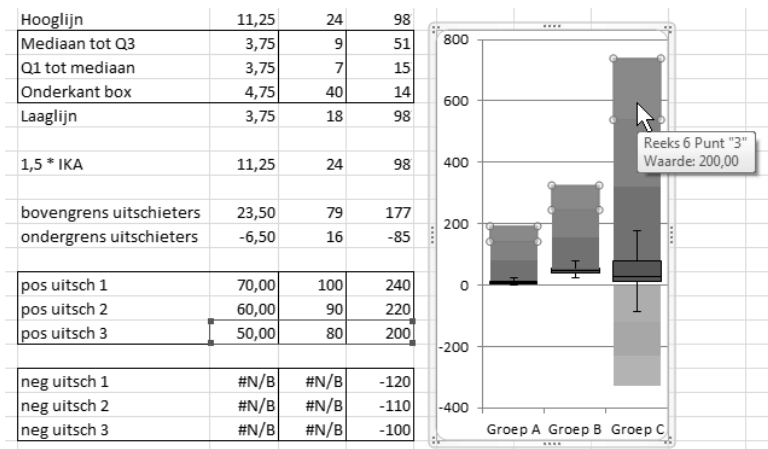
We hebben zes groepen van steeds drie uitschieters: positieve uitschieters van Winkel A, B en C en negatieve uitschieters van A, B en C. Deze plaatsen we als volgt in de grafiek. Rechtsklik op de grafiek en klik op **Gegevens selecteren**; het venster **Gegevensbron selecteren** gaat open. Klik op **Toevoegen**, laat in het vensterje dat opengaat het vak **Reeksnaam** leeg, klik in het vak **Reekswaarden**, verwijder het **=**{1} dat er staat, sleep over de cellen F19 tot en met H19 (de cellen naast positieve uitschieters) en klik één keer op **OK**. Klik weer op **Toevoegen**, klik in het vak **Reeks-**

	A	B	C	D	E	F	F
1	Winkel A	Winkel B	Winkel C			Winkel A	Formules in kolom F
2	1	22	8	Max	25,00	=MIN(F16;MAX(A1:A100))	
3	2	25	13	Q3	11,50	=KWARTIEL(A1:A100;3)	
4	3	34	14	Mediaan	7,00	=KWARTIEL(A1:A100;2)	
5	4	35	16	Q1	2,50	=KWARTIEL(A1:A100;1)	
6	5	41	23	Min	-11,00	=MAX(F17;MIN(A1:A100))	
7	6	41	26				
8	7	46	28	Hooglijn	13,50	=F2-F3	
9	8	46	33	Mediaan tot Q3	4,50	=F3-F4	
10	9	47	39	Q1 tot mediaan	4,50	=F4-F5	
11	10	49	61	Onderkant box	2,50	=F5	
12	11	54	72	Laaglijn	13,50	=F5-F6	
13	12	54	86				
14	13	59	100	1,5 * IKA	13,50	=(F3-F5)*1,5	
15	50	80	200				
16	60	90	220	bovengrens uitschieters	25,00	=F3+F14	
17	70	100	240	ondergrens uitschieters	-11,00	=F5-F14	
18	-100	-100	-100				
19	-110	-110	-110	pos uitsch 1	70,00	=ALS(GROOTSTE(A1:A100;1)>F16;GROOTSTE(A1:A100;1);NB())	
20	-120	-120	-120	pos uitsch 2	60,00	=ALS(GROOTSTE(A1:A100;2)>F16;GROOTSTE(A1:A100;2);NB())	
21				pos uitsch 3	50,00	=ALS(GROOTSTE(A1:A100;3)>F16;GROOTSTE(A1:A100;3);NB())	
22							
23				neg uitsch 1	-120,00	=ALS(KLEINSTE(A1:A100;1)<F17;KLEINSTE(A1:A100;1);NB())	
24				neg uitsch 2	-110,00	=ALS(KLEINSTE(A1:A100;2)<F17;KLEINSTE(A1:A100;2);NB())	
25				neg uitsch 3	-100,00	=ALS(KLEINSTE(A1:A100;3)<F17;KLEINSTE(A1:A100;3);NB())	

Afbeelding 14

Met deze formules berekenen we de uitschieters van drie groepen.

waarden, verwijder daaruit = {1}, sleep over de cellen F20 tot en met H20. Mocht daarin #N/B staan, dan is dat geen probleem, het gaat erom dat we nu de celverwijzing instellen. Klik één keer op OK. Herhaal dit voor alle zes rijtjes van cellen die naast elkaar liggen. U hebt nu Reeks 6 tot en met 11 erbij gemaakt. Sluit het venster Gegevensbron selecteren. Er is voor iedere reeks een vlak op de toren gestapeld, of eronder gezet.



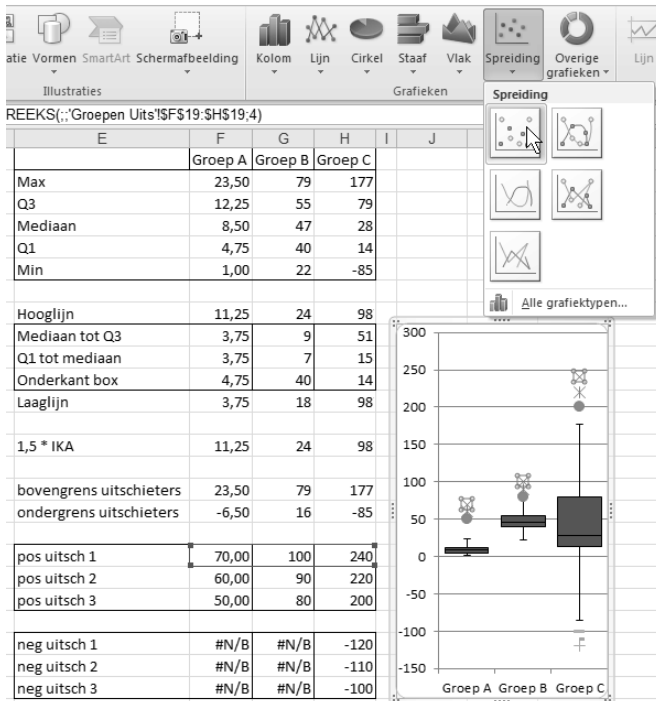
Afbeelding 15

U voegt de uitschieters van de reeksen toe aan de grafiek. Die worden voorlopig boven op en onder de kolom gestapeld.

Bonushoofdstuk
Een verticale boxplot bouwen

Vervolgens willen we de uitschieters als cirkels boven en onder de hoog- en laaglijn hebben. Hiervoor moeten we hun waarden in een spreidingsgrafiek plaatsen.

- De uitschieters worden automatisch berekend. Kunt u ze lastig terugvinden om ze te bewerken of zijn er met de huidige cijfers nog geen uitschieters (u ziet dan #N/B staan), dan kunt u ook eerst zelf een aantal waarden typen. Typ in dat geval in F19 tot en met H21 verschillende getallen, die groter zijn dan het getal dat bij bovengrens uitschieters staat. En typ in F23 tot en met H25 getallen die kleiner zijn dan het getal bij ondergrens uitschieters. Voer de volgende stappen uit en plaats dan naderhand de formules in de cellen.
- Of typ in de lijst met gegevens zelf getallen die er duidelijk uitschieten. Bent u straks klaar, vergeet dan vooral niet om ze weer weg te halen.
- Klik op een van de vlakken die zojuist onder en boven op de kolom zijn gestapeld. Klik op de tab **Invoegen**, klik op de knop **Spreiding** en kies het eerste type, **Spreiding met alleen markeringen**. Herhaal deze stappen voor alle zes uitschieters; u hoeft slechts de vlakken van één kolom te behandelen, de andere kolommen komen vanzelf mee. Hierna worden Reeks 6 tot en met 11 weer gegeven met kleine markeringen: dit zijn de uitschieters en die komen nu boven en onder hun box terecht.



■ **Afbeelding 16**

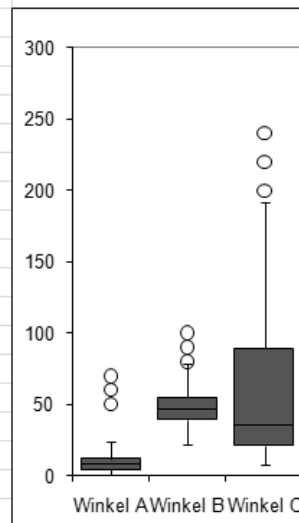
Om de uitschieters te tonen, geven we Reeks 6 tot en met 11 weer als Spreidingsgrafiek.

- Is een vlak zo klein dat u er niet gemakkelijk op kunt klikken, vergroot dan het beeld door met ingedrukte Ctrl-toets aan het scrollwiel van de muis te draaien. Lukt het nog niet, klik dan op de grafiek; de tabs Hulpmiddelen voor grafieken verschijnen. Klik op de tab **Opmaak**; helemaal links in beeld staat een keuzelijst. Kies om te beginnen Reeks6 in die keuzelijst.

Ten slotte geeft u de uitschieters als cirkels weer met behulp van de aanwijzingen in de paragraaf *Uitschieters als cirkels weergeven*.

- Staan de uitschieters eenmaal goed in de grafiek, dan kan hun waarde automatisch worden berekend. Hebt u vóór deze operatie zelf waarden getypt in de cellen F19 tot en met H25, zet dan de formules in die cellen.
- Of verwijder de uitschieters die u zelf in de lijst met gegevens hebt getypt.

Hooglijn	11,25	24	102
Mediaan tot Q3	3,75	9	54
Q1 tot mediaan	3,75	7	15
Onderkant box	4,75	40	21
Laaglijn	3,75	18	13
1,5 * IKA	11,25	24	102
bovengrens uitschieters	23,50	79	192
ondergrens uitschieters	-6,50	16	-81
pos uitsch 1	70,00	100	240
pos uitsch 2	60,00	90	220
pos uitsch 3	50,00	80	200
neg uitsch 1	#N/B	#N/B	#N/B
neg uitsch 2	#N/B	#N/B	#N/B
neg uitsch 3	#N/B	#N/B	#N/B



■ **Afbeelding 17**

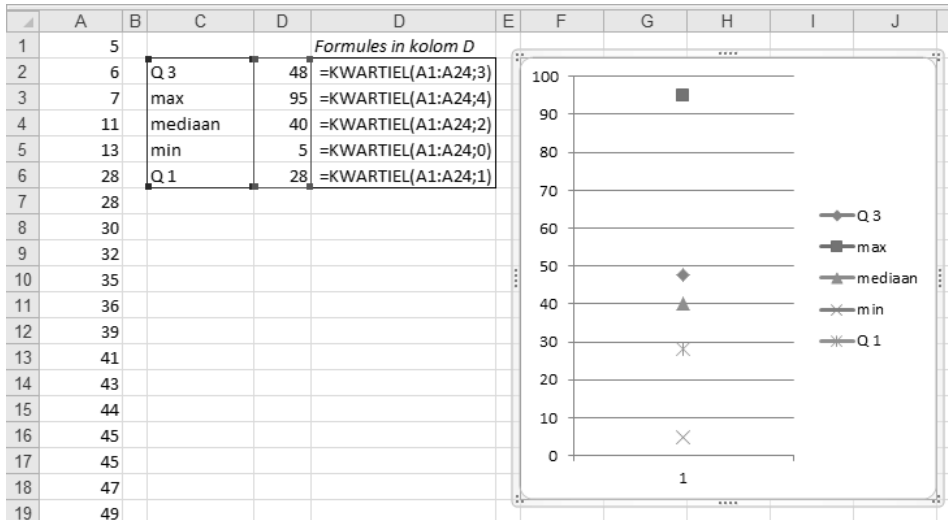
De uitschieters van iedere groep worden automatisch berekend en weergegeven als witte cirkels.

■ **Waarom geen boxplot tekenen met balken?**

Als u op de hoogte bent van de optie Omhoog/omlaag-balken voor grafieken, vraagt u zich misschien af waarom u daarmee geen boxplot zou maken. Dat hebben we voor u uitgezocht. Het grootste nadeel hiervan is dat de lijn van de mediaan niet over de volle breedte van de box loopt.

Wilt u het proberen, neem dan de formules over uit de afbeelding. Let u even op de volgorde van de kwartielen. Het is essentieel dat u exact dezelfde volgorde aanhoudt, want dan berekent Excel automatisch de lengte van de lijnen en de balken.

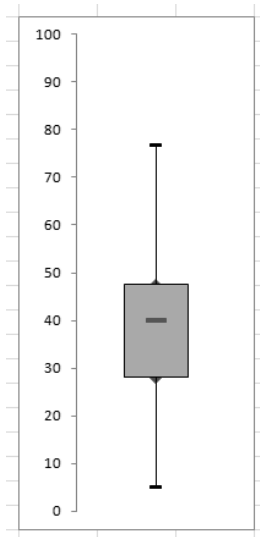
Bonushoofdstuk
Een verticale boxplot bouwen



Afbeelding 18

Voor een boxplot met Omhoog/omlaag-balken plaatst u de vijfgetallensamenvatting in deze volgorde in uw werkblad. U maakt daarmee eerst een grafiek met vijf markeringen.

- Selecteer de tabel van C2 tot en met D6, klik op **Invoegen**, klik in de groep Grafieken op **Lijn** en kies de optie **Lijn met gegevensmarkeringen**; er verschijnt een grafiek. Klik in de tabs Hulpmiddelen voor grafieken op de tab **Ontwerpen** en kies **Rijen/kolommen omdraaien**. Hierdoor verandert de lijn in markeringen boven elkaar.



Afbeelding 19

U kunt een boxplot ook maken met lijnen en balken, al loopt de mediaan dan niet over de volle breedte.

- Klik op de grafiek, op de tab **Indeling**, op **Lijnen** en kies **Hoog/laag-lijnen**. Excel trekt een lijn van de maximumwaarde naar het minimum.
- Klik vervolgens in de tab **Indeling** op **Omhoog/omlaag-balken** en kies **Omhoog/omlaag-balken**. U krijgt een rechthoek in het midden. Excel laat deze automatisch van Q1 tot Q3 lopen (de onderste en bovenste waarden in de tabel). Pas eventueel de kleur van de rechthoek aan.

U kunt de markeringen boven en onder veranderen in dwarsstreepjes (met **Opmaken** en **Markeringsopties**). Maar de markering van de mediaan blijft te kort. Als u die opmaakt, kunt u met **Markeringsopties** en **Standaard** een streepje kiezen en dat bijvoorbeeld 15 punten groot maken, maar het zal niet zo lang worden dat het de randen van de box raakt. Daarom is dit voor ons gevoel geen echte boxplot. Ook de uitschieters ontbreken. Maar hier staat tegenover dat Excel de lengte van de lijnen en de balken automatisch berekent. Dus wilt u even snel een boxplot maken, dan kan het op deze manier.

