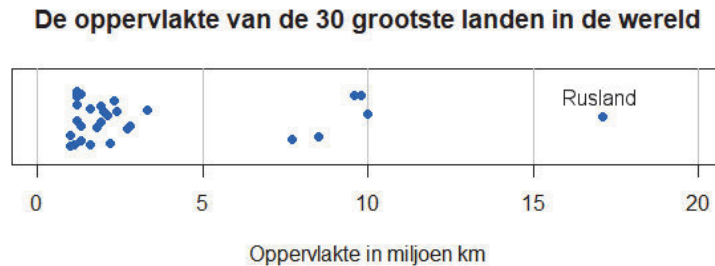


## Antwoorden opgaven

---

### Antwoord opgave 2.11.1.

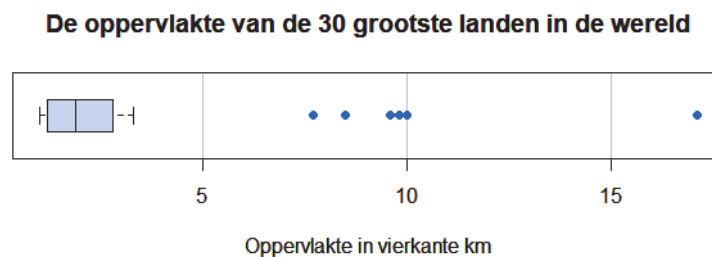
De grafiek hieronder toont het spreidingsdiagram:



Er is een grote groep van 24 kleine landen. Dat zijn landen met een oppervlakte van hooguit 3 miljoen vierkante km. Verder is er een groep van vijf middelgrote landen, met een oppervlakte van tussen de 7,7 en 10,0 miljoen vierkante km. Ten slotte is er nog een grote uitschieter. Dat is Rusland met een oppervlakte van 17,1 miljoen vierkante km. Voor de duidelijkheid is deze uitschieter gelabeld.

### Antwoord opgave 2.11.2.

- De grafiek hieronder toont het toont het spreidingsdiagram:
- $K1 = 1,2$ ,  $K2$  (mediaan) = 1,6 en  $K3 = 2.15$ .
- Lengte van de doos = 0,95.
- Anderhalf keer de lengte van de doos = 1,425.
- $K1 - 1,425 = -0,225$ . Naburige waarde links = 1,0.
- $K3 + 1,425 = 3,575$ . Naburige waarde rechts = 3,3.

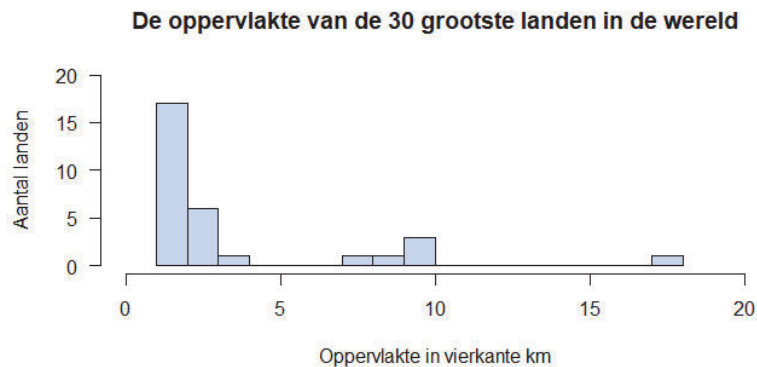


Uit de grafiek blijkt dat er een grote groep van kleine landen is met een oppervlakte kleiner dan 3 miljoen vierkante km. De verdeling van de oppervlaktes van die metingen is (bijna) symmetrisch om de mediaan, en die is hier gelijk aan 1,6. Dat is het midden tussen de twee waarden van Mongolië (1,6 miljoen vierkante km) en Iran (ook 1,6 miljoen vierkante km).

De boxplot vindt dat de 5 middelgrote landen (met waarden net onder de 10) uitschieters zijn. En uiteraard is Rusland met 17,1 ook een uitschieter. Rusland is gewoon heel groot (qua oppervlakte).

### Antwoord opgave 2.11.3.

De grafiek met het histogram staat hierna. De breedte van de klasse is hier gelijk genomen aan 1 (miljoen vierkante kilometer).



Ook in deze grafiek is weer duidelijk te zien dat er drie groepen landen: een grote groep van 24 kleine landen, een groep van 5 middelgrote landen, en één grote uitschieter. De verdeling is zeker niet symmetrisch. Hij is duidelijk scheef naar rechts (een lange rechter staart). De verdeling is verre van normaal.

### Antwoord opgave 2.11.4.

- Gemiddelde van de oppervlaktes = 3,5.
- Mediaan van de oppervlaktes = 1.6.

Er is een duidelijk verschil tussen beide waarden. De verdeling is dus zeker niet symmetrisch (en dus ook niet normaal). Het gemiddelde is groter dan de mediaan. Dat is een aanwijzing dat er uitschieters zijn in de rechter staart van de verdeling.

### Antwoord opgave 3.7.1.

Er zijn twee belangrijke zaken mis met deze grafiek. In de eerste plaats heeft de grafiek een driedimensionaal perspectief. Daardoor is het lastig om de verschillende taartpunten met het oog te vergelijken.

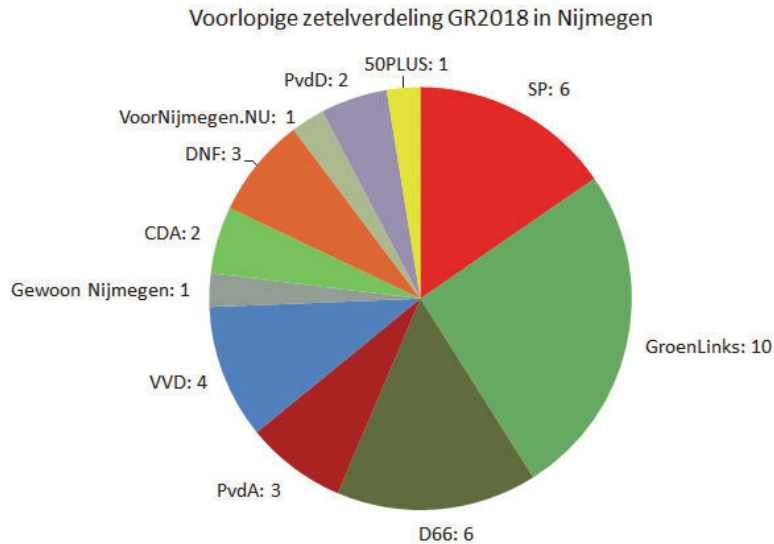
De taartpunten aan de voorkant (D66 en GroenLinks) lijken nog groter dan ze al zijn. Dat komt omdat ook de zijkanten van de taartpunten zichtbaar zijn. Je krijgt niet de indruk dat de taartpunt van D66 kleiner is dan die van GroenLinks. Bij de taartpunten aan de achterkant zijn de zijkanten niet zichtbaar. Daardoor lijken ze kleiner dan ze zijn.

Het beeld dat de grafiek oproept, is verwrongen. Kijk maar eens naar de punten van D66 en de SP. Beide partijen hebben zes zetels, maar de taartpunt van D66 lijkt veel groter dan die van de SP.

Een tweede probleem is dat om onduidelijke reden de taartpunten niet tegen elkaar aan zitten, maar van elkaar zijn gescheiden. Een dergelijk cirkeldiagram wordt ook wel een *geëxplodeerd cirkeldiagram* genoemd. Wat dragen deze gleuven bij aan de interpretatie? Het maakt de interpretatie lastiger.

### Antwoord opgave 3.7.2.

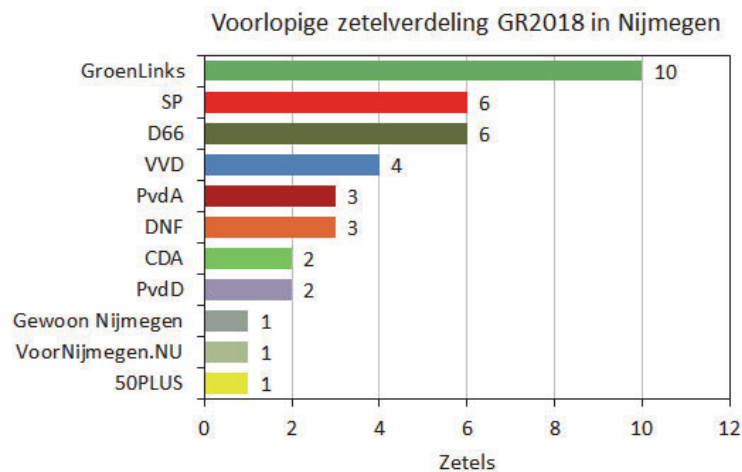
Het totaal aantal zetels is 39 en dat komt overeen met 360 graden. Het aantal graden voor de SP (6 zetels) is dan  $360 \times 6 / 39 = 55,4$ . Voor GroenLinks is het  $360 \times 10 / 39 = 92,3$ . Enz. Teken de sectoren een voor een in een lege cirkel. Dit moet dan leiden tot een 'plat' cirkeldiagram, zoals hierna:



Dit cirkeldiagram is aanzienlijk beter leesbaar. Toch blijft het lastig al die sectoren met elkaar te vergelijken. Sommige deskundigen raden aan geen cirkeldiagrammen te gebruiken als er meer dan zes categorieën zijn.

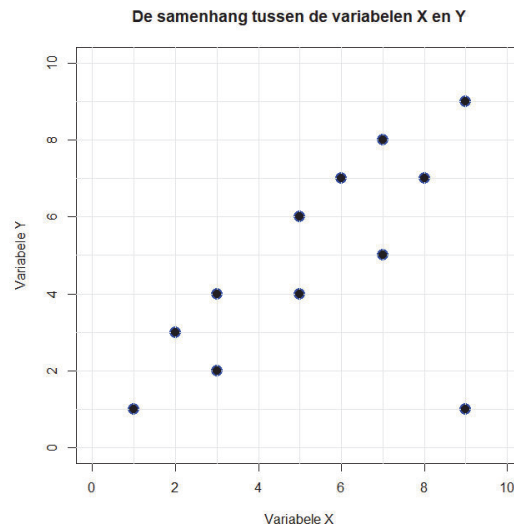
### Antwoord opgave 3.7.3.

Je zou de voorkeur kunnen geven aan horizontale staven. Dan heb je meer ruimte om de teksten bij de staven te zetten. Je kunt overwegen om verschillende kleuren voor de staven te gebruiken. Elke partij heeft immers zijn eigen kleur. Om te voorkomen dat sommige kleuren feller zijn dan andere kleuren, en daardoor meer aandacht krijgen, is het misschien beter om al die kleuren hetzelfde te houden. Hieronder staat het resulterende staafdiagram. Een staafdiagram is misschien wat minder frivool dan een cirkeldiagram, maar het is wel functioneler (beter interpreteerbaar).



### Antwoord opgave 4.5.1.

a. De grafiek hieronder is een voorbeeld van hoe de puntenwolk eruit zou kunnen zien:



b. Afgezien van de uitschieter rechtsonder liggen de punten redelijk keurig op (of dichtbij) een rechte lijn. Er is dus sprake van een lineair verband. Dat ene punt voor object 12 (met waarden  $X = 9$  en  $Y = 1$ ) verstoort het patroon.

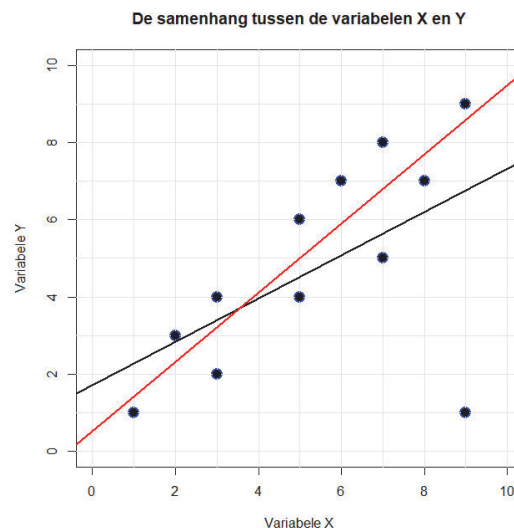
### Antwoord opgave 4.5.2.

a. De formule van de regressielijn is  $A + B \times X = 0,5326 + 0,8954 \times X$ .

b. De voorspelde waarde van  $Y$  voor  $X = 0$  is gelijk aan  $0,5326 + 0,8954 \times 0 = 0,5326$ . Het punt linksonder heeft dus coördinaten  $X = 0$  en  $Y = 0,5326$ .

c. De voorspelde waarde van  $Y$  voor  $X = 9$  is gelijk aan  $0,5326 + 0,8954 \times 9 = 8,5912$ . Het punt rechtsboven heeft dus coördinaten  $X = 9$  en  $Y = 8,5912$ .

d. Deze (rode) regressielijn in de grafiek hieronder is een goede samenvatting van de puntenwolk, als je het twaalfde object buiten de analyse laat, moet je je wel afvragen wat er aan de hand is met de uitschieter. Is er misschien een foutje gemaakt bij het verwerken van de gegevens? Dan moet je die gegevens corrigeren en de analyse opnieuw doen. Gaat het om een correcte waarneming, dan moet je concluderen dat er geen simpele structuur in de data zit.



### Antwoord opgave 5.7.1.

De tabel met randtotalen staat hieronder. Ook de celpercentages zijn toegevoegd

	Progressief	Conservatief	Weet niet	Totaal
Man	440 22%	220 11%	240 12%	900
Vrouw	320 16%	480 24%	300 15%	1100
Totaal	760	700	540	2000

De groep conservatieve vrouwen is met 24% de grootste groep. De groep conservatieve mannen is met 11% de kleinste groep. Het tabeltotaal is 2000. Als controle kun je zien dat het tabeltotaal overeen komt met het aantal respondenten in de steekproef. Er zitten ongeveer evenveel mannen en vrouwen in de datamatrix.

### Antwoord opgave 5.7.2.

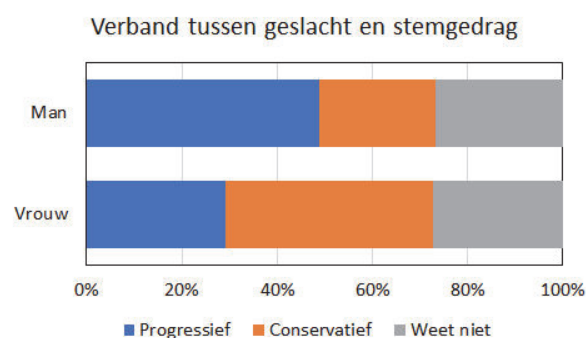
De tabel met rijpercentages staat hieronder. Die rijpercentages krijg je door de celaantallen te percenteren op het bijbehorende rijtotalen/

	Progressief	Conservatief	Weet niet	Totaal
Man	440 48,9%	220 24,4%	240 26,7%	900 100%
Vrouw	320 29,1%	480 43,6%	300 27,3%	1100 100%

Bij de mannen is de groep progressieven oververtegenwoordigd, en dus ondervertegenwoordigd bij de vrouwen. Bij de vrouwen is het net andersom. De groep 'weet niet' is bij mannen en vrouwen ongeveer even groot.

### Antwoord opgave 5.7.3.

Voor het maken van het stapeldiagram kun je de gegevens van opgave 5.2 gebruiken. Hieronder de grafiek zoals die is gemaakt met het spreadsheet-programma Excel.



Bij de mannen zijn de progressieven oververtegenwoordigd. Bij de vrouwen is het net andersom: de conservatieven zijn oververtegenwoordigd. Het percentage 'weet niet' is bij mannen en vrouwen ongeveer even groot. Die groep is helaas best wel groot.

#### Antwoord opgave 5.7.4.

In de situatie van een  $2 \times 3$  tabel is Cramér's  $V$  gelijk aan:

$$V = \sqrt{\frac{G/n}{\text{Min}(r,k)-1}} = \sqrt{\frac{103,2176/2000}{2-1}} = \sqrt{103,2176/2000} = 0,2276.$$

De waarde van Cramér's  $V$  is kleiner dan 0,3. We kunnen concluderen dat er sprake is van slechts een zwak verband tussen beide variabelen.

#### Antwoord opgave 6.6.1.

Het aantal calorieën per hotdog is bij kip-hotdogs wat lager dan bij rund-hotdogs en varken-hotdogs. Er is vrijwel geen verschil tussen varken-hotdogs en rund-hotdogs.

Bij alle drie soorten hotdogs is er sprake van een verdeling in twee groepen hotdogs. Er is een groep met wat minder calorieën en een groep met wat meer calorieën.

Bij de varken-hotdogs en de rund-hotdogs is er sprake van uitschieters.

Deze grafiek roept nieuwe vragen op: waarom zijn er aparte groepjes en wat is er aan de hand bij de uitschieters?

#### Antwoord opgave 6.6.2.

Het aantal calorieën per hotdog is bij kip-hotdogs behoorlijk wat lager dan bij rund-hotdogs en varken-hotdogs. Er is vrijwel geen verschil tussen varken-hotdogs en rund-hotdogs. De mediaan van het aantal calorieën is bij beide groepen hetzelfde.

Bij de varken-hotdogs en de rund-hotdogs is er kennelijk geen sprake van uitschieters.

De verdeling van de aantallen calorieën per groep is niet symmetrisch. De mediaan ligt immers niet in het midden van de doos.

De linker staarten van de verdeling zijn bij rund-hotdogs en varken-hotdogs groter dan de rechter staarten.

Uit het oogpunt van het aantal calorieën is het dus beter om kip-hotdogs te eten.

#### Antwoord opgave 6.6.3.

Er valt niet zo veel af te leiden uit deze tabel. De belangrijkste conclusie is misschien wel dat het gemiddelde aantal calorieën voor kip-hotdogs lager is dan bij rund-hotdogs en varken-hotdogs.

Bedenk hierbij wel dat er uitschieters zijn, al zie je ze hier niet. Die kunnen grootheden als het gemiddelde ernstig beïnvloeden. Misschien moet je de analyse nog een keer doen, maar dan zonder de twee uitschieters.