

# Kijken naar

# grafieken



Centraal Bureau  
voor de Statistiek



**Kijken naar**

**grafieken**

## Colofon

### *Uitgever*

Centraal Bureau voor de Statistiek  
Henri Faasdreef 312, 2492 JP Den Haag  
[www.cbs.nl](http://www.cbs.nl)

Prepress: Centraal Bureau voor de Statistiek, Grafimedia  
Ontwerp: Edenspiekermann

### *Inlichtingen*

Tel. 088 570 70 70, fax 070 337 59 94  
Via contactformulier: [www.cbs.nl/infoservice](http://www.cbs.nl/infoservice)

© Centraal Bureau voor de Statistiek, Den Haag/Heerlen, 2014.  
Verveelvoudigen is toegestaan, mits het CBS als bron wordt vermeld.

# Contents

<b>1. Waarom grafieken?</b>	<b>4</b>
<b>2. De bouwstenen van een grafiek</b>	<b>10</b>
2.1 Variabelen	11
2.2 De gegevens	12
2.3 Titel	13
2.4 Namen van de staven	13
2.5 Schaalverdeling	14
2.6 Roosterlijnen	15
2.7 Kleuren	15
2.8 Legenda	16
<b>3. Staafdiagrammen</b>	<b>19</b>
3.1 Horizontaal of verticaal	20
3.2 Doorknippen van de staven?	22
3.3 3D-grafieken?	24
3.4 Staafdiagrammen voor twee variabelen	25
<b>4. Cirkeldiagrammen</b>	<b>30</b>
<b>5. Tijdreeksen</b>	<b>34</b>
5.1 Tijdreeks voor een kwantitatieve variabele	35
5.2 Tijdreeksen voor een kwalitatieve variabele	38
<b>6. Foute grafieken</b>	<b>41</b>
6.1 Te veel chartjunk	42
6.2 Gerommel met de assen	45
6.3 Pictogrammen in plaats van staven	48
Gebruikte bronnen	51
Gebruikte illustraties	52
Medewerkers	53

1.

# Waarom grafieken?

**Het Centraal Bureau voor de Statistiek (CBS) verzamelt heel veel informatie over de toestand van Nederland. Met die gegevens probeert het CBS te beschrijven hoe het met de mensen gaat en hoe het met de economie gaat. Ook kijkt het CBS naar de ontwikkelingen in de tijd: gaat het beter of gaat het slechter?**

Het CBS verzamelt veel van die gegevens met enquêtes. Mensen of bedrijven krijgen vragenlijsten die ze moeten invullen en terugsturen. Het CBS krijgt zo heel veel gegevens binnen. De vraag is nu hoe je de informatie in al die gegevens op een zo duidelijke en bondige manier kunt samenvatten dat iedereen het begrijpt. Hoe kun je de boodschap die in de gegevens verstopt zit, zichtbaar maken? Daarvoor zijn drie manieren.

De eerste manier is het schrijven van een verhaaltje over de gegevens. Hier is een voorbeeld dat gaat over cyberpesten:

*In 2012 gaf 3 procent van de Nederlandse bevolking van 15 jaar of ouder aan weleens gepest te zijn via internet. Dit zijn naar schatting ruim 400 duizend personen. Jongeren worden veel meer met cyberpesten geconfronteerd dan ouderen. Van de 15- tot 18-jarigen had 10 procent in 2012 te maken met pesten via internet, tegen minder dan 1 procent van de 65-plussers.*

Deze tekst bevat de belangrijkste uitkomsten van het onderzoek naar cyberpesten. Veel details staan erin. Toch is het lastig om het globale beeld eruit te halen.

De tabel is een andere manier om de informatie samen te vatten. Het is een numeriek overzicht waarin vooral cijfers staan en haast geen tekst. Tabel 1.1 bevat de uitkomsten van het onderzoek naar cyberpesten:

Deze kleine tabel is nog wel redelijk overzichtelijk. Zo is eenvoudig te zien in welke leeftijdscategorie cyberpesten het meest voorkomt (15 tot 18 jaar) en in welke groep het minst (65 jaar en ouder). Als tabellen groter en ingewikkelder worden, dan wordt de interpretatie van een tabel een stuk lastiger.

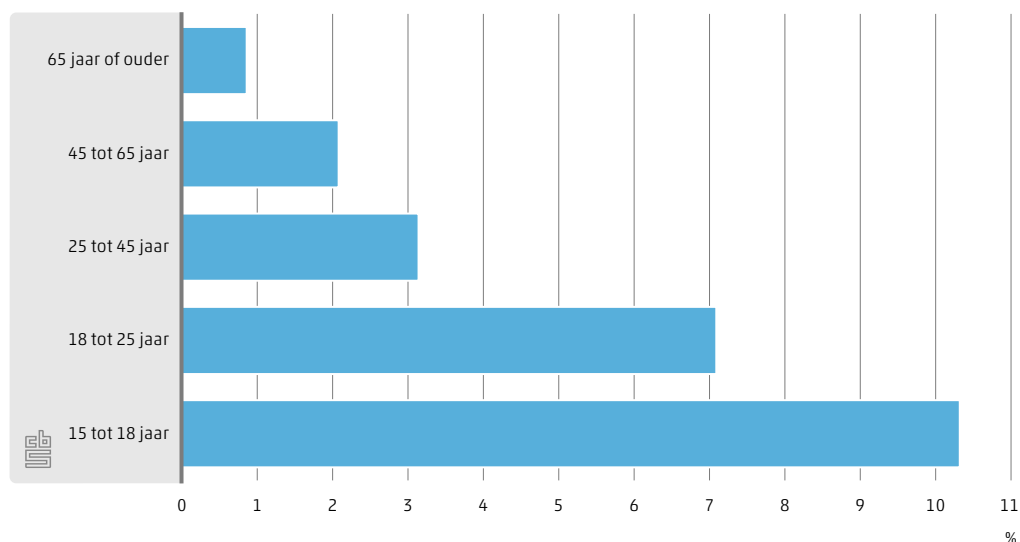
### **1.1 Personen die te maken hebben gehad met cyberpesten, naar leeftijd, 2012**

Leeftijd	Percentage
15 tot 18 jaar	10,3
18 tot 25 jaar	7,1
25 tot 45 jaar	3,1
45 tot 65 jaar	2,1
65 jaar en ouder	0,9

Een derde manier om de uitkomsten van een onderzoek weer te geven, is met een grafiek. In een grafiek is vaak eenvoudiger te zien wat er aan de hand is. Dat maakt een grafiek geschikt voor een groot en breed publiek. Maar pas op: er zijn ook gevaren. Je kunt gegevens op allerlei manieren grafisch weergeven. Niet elke manier is goed. Het is vrij eenvoudig mogelijk om mensen op het verkeerde been te zetten. Daardoor kunnen ze de verkeerde conclusies uit de gegevens trekken. Het is daarom belangrijk om zorgvuldig te zijn als je zelf een grafiek maakt. En je moet oppassen als je een grafiek van iemand anders onder ogen krijgt.

Er zijn veel verschillende soorten grafieken in gebruik. Het hangt van het soort gegevens af welk type grafiek het best is om de 'boodschap' op een goede manier over te brengen. We beperken ons tot staafdiagrammen, cirkeldiagrammen en lijndiagrammen. Er zijn, bijvoorbeeld, ook nog histogrammen en thematische kaarten, maar die komen hier niet aan bod. Het hangt van het soort gegevens af welk type grafiek het beste is. Voor de gegevens over cyberpesten komt het *staafdiagram* in aanmerking. Het resultaat staat in figuur 1.2.

## 1.2 Slachtofferschap cyberpesten naar leeftijd, 2012



### Vraag 1.1

Probeer uit de grafiek in figuur 1.2 af te lezen wat er aan de hand is met cyberpesten. Beantwoord daarvoor de volgende vragen:

a. Welke trend kun je zien als je kijkt naar cyberpesten en leeftijd?

.....

.....

b. Ongeveer hoeveel procent van de 15- tot 18-jarigen is in 2012 het slachtoffer geweest van cyberpesten?

.....

.....

c. Hoe zou je kunnen verklaren dat cyberpesten minder vaak voorkomt bij ouderen?

.....

.....

Een grafiek kan vooral zijn kracht tonen als hij veel gegevens in beeld brengt. Een grafiek kan onthullend zijn. Een grafiek kan het inzicht in de gegevens vergroten door het spontaan onthullen van allerlei aspecten die in tekst of tabellen verborgen zouden zijn gebleven.

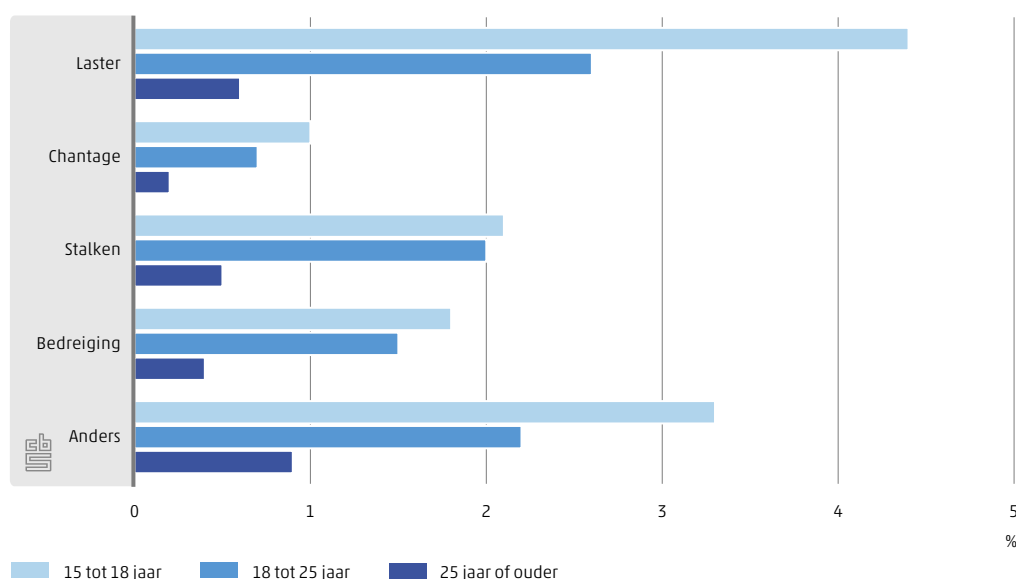
We maken het voorbeeld wat ingewikkelder door verschillende soorten cyberpesten te onderscheiden: laster (kwetsende teksten plaatsen in sociale media), chantage, stalken, bedreiging met geweld en andere vormen van cyberpesten. De leeftijd beperken we tot drie categorieën: 15 tot 18 jaar, 18 tot 25 jaar en 25 jaar of ouder. Tabel 1.3 bevat de gegevens in tabelvorm.

### 1.3 Personen die te maken hebben gehad met cyberpesten, naar soort en leeftijd, 2012

	15 tot 18 jaar	18 tot 25 jaar	25 jaar of ouder
	%		
Laster	4,4	2,6	0,6
Chantage	1,0	0,7	0,2
Stalken	2,1	2,0	0,5
Bedreiging	1,8	1,5	0,4
Anders	3,3	2,2	0,9

Het wordt nu toch wel wat lastiger om een duidelijk beeld te krijgen van de gegevens in de tabel. Daarvoor is het beter om een grafiek te maken. Een type grafiek dat hiervoor in aanmerking komt, is het *samengesteld staafdiagram*. Figuur 1.4 bevat het samengesteld staafdiagram voor de gegevens over cyberpesten. We gebruiken de term 'samengesteld' om aan te geven dat we tegelijk twee aspecten van de gegevens in beeld brengen: de vorm van cyberpesten en de leeftijdscategorie. Je zou ook kunnen zeggen dat figuur 1.4 een staafdiagram bevat voor elke vorm cyberpesten.

### 1.4 Personen die te maken hebben gehad met cyberpesten, naar soort en leeftijd, 2012





### Vraag 1.2

Aan de grafiek in figuur 1.4 is van alles te zien. Probeer de volgende vragen te beantwoorden:

a. Welke vorm van cyberpesten komt het meest voor bij de verschillende leeftijdsgroepen?

.....

b. Welke vorm van cyberpesten komt het minst voor bij 18- tot 25-jarigen?

.....

c. Beschrijf in je eigen woorden hoe het zit met stalken en leeftijd.

.....

### Vraag 1.3

Het CBS voert elk jaar het onderzoek 'ICT-gebruik van huishoudens en personen' uit. Daarin worden aan een steekproef van ongeveer 4500 Nederlanders tussen 12 en 75 jaar vragen gesteld over het gebruik van computers en internet. In een publicatie uit 2013 staat onderstaande grafiek over het gebruik van verschillende apparaten voor mobiel internet in 2012.

Probeer de grafiek te 'lezen'. Geef vier conclusies die je kunt trekken uit deze grafiek.

.....

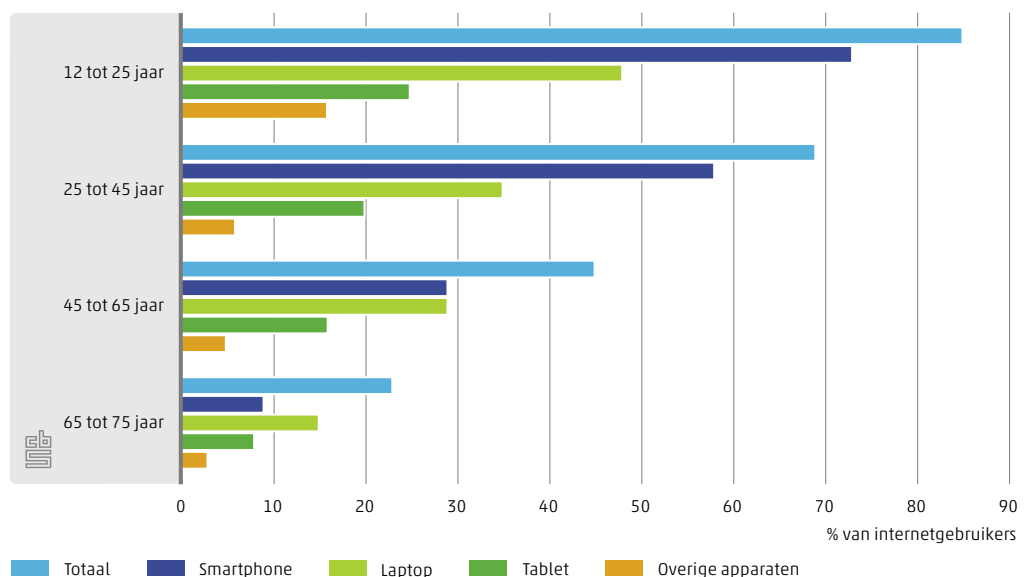
.....

.....

.....

.....

### 1.5 Het gebruik van mobiele apparatuur voor internet door internetgebruikers, naar leeftijd in 2012



In de grafiek in figuur 1.4 zijn de omvang van de verschillende groepen weergegeven in de vorm van percentages. We hadden het ook in de vorm van aantallen personen kunnen doen. Je kunt een percentage vaak wat makkelijker interpreteren. Een percentage is een manier om aan te geven hoe groot het aandeel in het totaal is. Zie je ergens bijvoorbeeld 50 procent staan, dan weet je meteen dat het om de helft van de groep gaat. Met percentages kun je ook vaak wat makkelijker groepen met elkaar vergelijken. Je kunt bijvoorbeeld in de tabel snel zien dat in de leeftijdsgroep van 15 tot 18 jaar veel meer jongeren (4,4%) te maken hebben met laster dan in de leeftijdsgroep van 18 tot 25 jaar (2,6%).

De voorbeelden in dit hoofdstuk zijn gemaakt met gegevens van het CBS. Het CBS is echter niet de enige die gegevens verzamelt. Er zijn andere overheidsorganisaties die dit ook doen. Een voorbeeld is het Sociaal en Cultureel Planbureau (SCP), dat sociaal-wetenschappelijk onderzoek doet. Een ander voorbeeld is het Koninklijk Meteorologisch Instituut (KNMI) dat heel veel gegevens over het weer verzamelt, zoals temperatuur, windkracht, windrichting, zonneschijn en regenval. Verder zijn er nog de universiteiten waar veel onderzoek gedaan wordt. Een voorbeeld zijn de experimenten met nieuwe behandelingen en nieuwe medicijnen bij de medische faculteiten. En dan zijn er ook nog de commerciële marktonderzoekbureaus. Die voeren heel vaak opiniepeilingen uit. Denk hierbij aan bedrijven als Ipsos, TNS Nipo en Intomart GfK.

De uitkomsten van die peilingen worden vaak in de vorm van grafieken weergegeven. Een goede grafiek is een krachtig hulpmiddel om snel inzicht te krijgen in de uitkomsten van een onderzoek. Voorwaarde is dan wel dat je een goede grafiek maakt. Dat is een grafiek die de 'boodschap' goed overbrengt en je niet op het verkeerde been zet. In de praktijk kom je helaas nogal eens foute grafieken tegen. Deze lesmodule leert je hoe je naar grafieken moet kijken. Ook leggen we uit hoe een goede grafiek er uit ziet en wat er mis kan gaan met een slechte grafiek.

**2.**

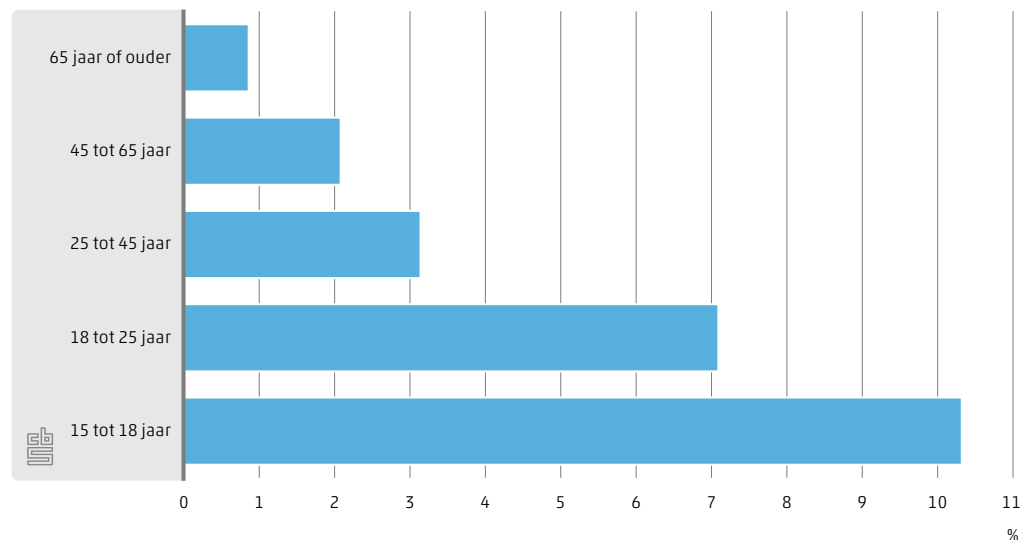
**De bouwstenen**

**van een grafiek**

Een grafiek bestaat uit een aantal bouwstenen. Die bouwstenen zijn nodig voor het maken van een goede grafiek. Een goede grafiek is een grafiek die makkelijk is te 'lezen'. Het moet eenvoudig zijn om juiste conclusies te trekken. En een grafiek mag je niet op het verkeerde been zetten.

In dit hoofdstuk bespreken we de belangrijkste bouwstenen van een grafiek. Daarbij gebruiken we de grafiek in figuur 2.1 als voorbeeld. De grafiek beschrijft hoeveel mensen in 2012 het slachtoffer waren van cyberpesten.

## 2.1 Slachtofferschap cyberpesten naar leeftijd, 2012



## 2.1 Variabelen

In een grafiek wordt een plaatje gemaakt van gegevens die afkomstig zijn uit een onderzoek. Die gegevens zijn de waarden van variabelen. Een *variabele* is een kenmerk dat wordt gemeten. De waarde van een variabele kan voor iedereen anders zijn. Vandaar dat we het een 'variabele' noemen, afkomstig van *variëren*.

Een voorbeeld van een variabele is de lengte van een leerling (in centimeters). Elke leerling heeft een andere lengte, al kan het natuurlijk ook wel eens voorkomen dat twee leerlingen dezelfde lengte hebben. Een ander voorbeeld van een variabele is de provincie waarin een persoon woont. Er zijn dan 12 mogelijke waarden: Groningen, Friesland, Drenthe, Overijssel, Flevoland, Gelderland, Utrecht, Noord-Holland, Zuid-Holland, Noord-Brabant en Limburg.

Voor het maken van grafieken is het handig om twee soorten variabelen te onderscheiden: kwalitatieve variabelen en kwantitatieve variabelen:

Een *kwantitatieve variabele* heeft getallen als waarden. Die getallen betekenen ook echt iets. De lengte van een leerling is een kwantitatieve variabele. Als de lengte van iemand 220 centimeter is, dan is hij langer dan iemand van 110 centimeter. Andere voorbeelden van kwantitatieve variabelen zijn het gewicht van een persoon, het maandelijkse inkomen van iemand of het aantal uren dat een persoon dagelijks bezig is op internet.

Soms heeft een kwantitatieve variabele ook een *natuurlijk nulpunt*. Een waarde van 0 betekent dan dat er niets meer is. Denk maar eens aan een lengte van 0 centimeter, of een gewicht van 0 gram. Dan kun je ook over verhoudingen praten. Iemand van 220 centimeter is dan twee keer zo lang als iemand van 110 centimeter. En iemand van 120 kilo weegt twee keer zo zwaar als iemand van 60 kilo. Temperatuur (in graden Celsius) is wel een kwantitatieve variabele, maar deze variabele heeft geen natuurlijk nulpunt. Een temperatuur van 20 graden is niet twee keer zo warm als een temperatuur van 10 graden.

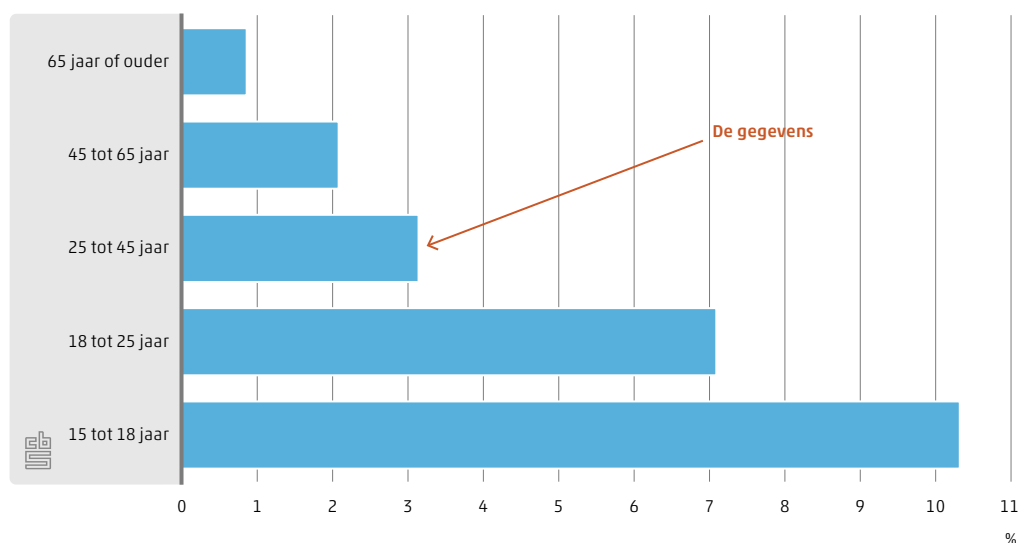
Een *kwalitatieve variabele* verdeelt de mensen in groepjes. De waarden van deze variabele zijn geen getallen maar etiketjes die je op de mensen plakt. De provincie waarin iemand woont is een kwalitatieve variabele. Deze variabele verdeelt de mensen dus in 12 groepen. Andere voorbeelden van kwalitatieve variabelen zijn de haarkleur van iemand (bruin, blond, rood, grijs, wit), het favoriete vak op school (lichamelijke opvoeding, wiskunde, Nederlands, economie,...) en de manier waarop een leerling naar school is gekomen (lopend, fiets, scooter, openbaar vervoer, auto, ...).

We maken dit onderscheid in variabelen omdat er vele soorten grafieken zijn. Sommige grafieken moet je gebruiken voor kwalitatieve variabelen en andere voor kwantitatieve variabelen. Voor een kwalitatieve variabele kun je bijvoorbeeld een staafdiagram gebruiken. Daarin geef je met staven aan hoe groot de groepjes zijn. In figuur 2.1 zie je daar een voorbeeld van. In hoofdstuk 3 leggen we uit hoe je een staafdiagram moet maken.

## 2.2 De gegevens

De grafiek vat de gegevens op een grafische manier samen. De gegevens moeten dus op een of andere manier duidelijk te zien zijn. Dat kan zijn in de vorm van staven, punten of lijnen. In figuur 2.2 gebeurt het in de vorm van staven.

### 2.2 Slachtofferschap cyberpesten naar leeftijd, 2012

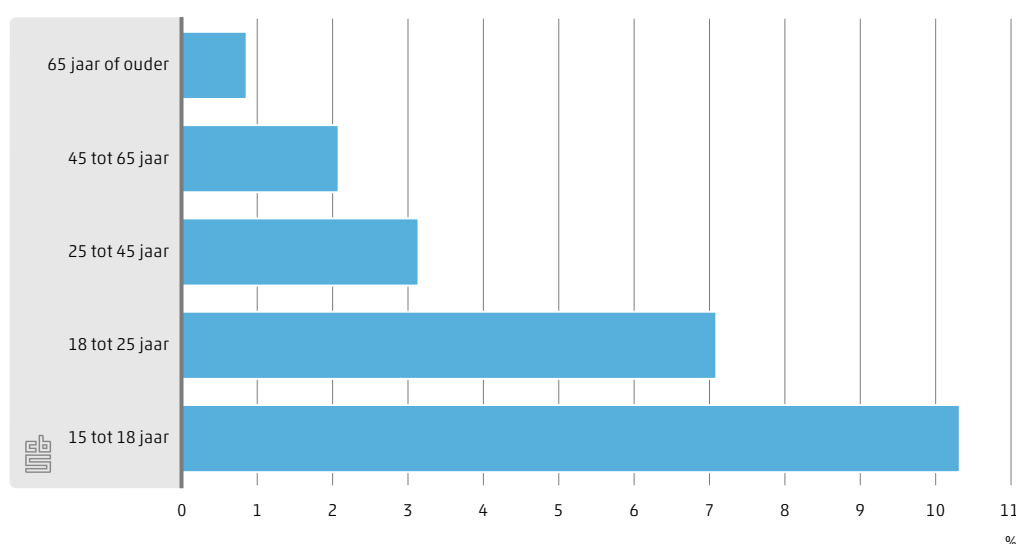


Die gegevens moeten duidelijk te zien zijn. Het mag niet zo zijn dat je de grafiek niet duidelijk kunt interpreteren omdat de gegevens verstopt zitten in allerlei overbodige tierlantijnen. Je moet 'chartjunk' vermijden (letterlijk: 'grafiekrommel'). Hoofdstuk 6 (Foute grafieken) bevat een aantal voorbeelden van grafieken die de verkeerde boodschap overbrengen.

## 2.3 Titel

Elke grafiek heeft een titel. Die staat vaak boven de grafiek. Dat is ook het geval in figuur 2.3. De titel legt uit wat er in de grafiek te zien is. De titel moet voldoende informatie geven. Het mag niet zo zijn dat je eerst uitgebreid in de tekst moet gaan zoeken naar uitleg. Als dat voor de begrijpelijkheid nodig is, dan mag de titel langer dan één regel zijn.

**2.3 Slachtofferschap cyberpesten naar leeftijd, 2012** ← De titel

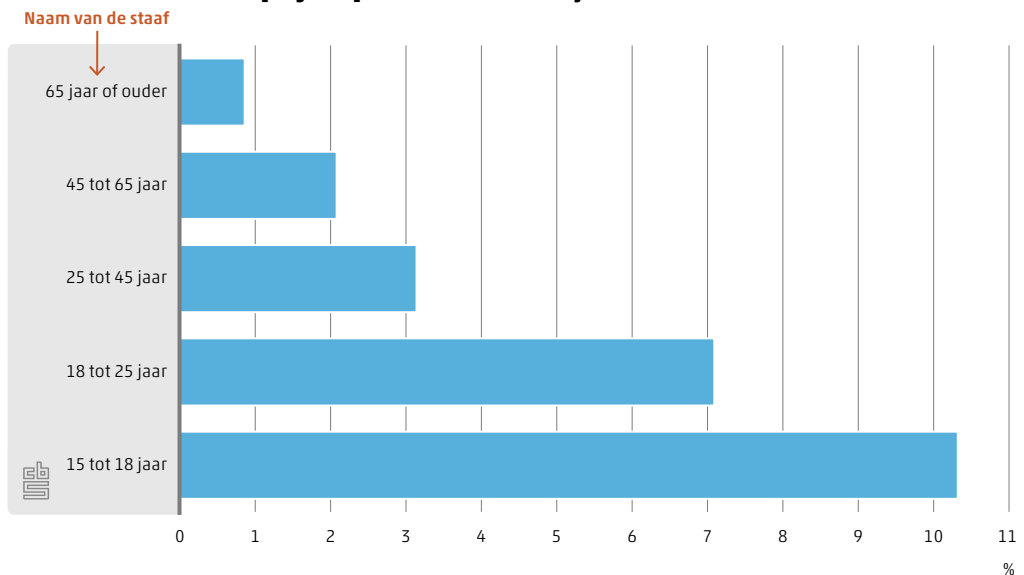


## 2.4 Namen van de staven

Het moet in de grafiek duidelijk zijn op welke groepen de verschillende elementen (staven, lijnen, punten) betrekking hebben. Figuur 2.4 bevat vijf staven. Elke staaf hoort bij een groep. Elke groep moet duidelijk omschreven worden.

Er moet in de grafiek voldoende ruimte zijn voor teksten. Dat zou, bijvoorbeeld, een stuk lastiger zijn in figuur 2.4 als de staven verticaal (in plaats van horizontaal) zouden zijn getekend. We komen daar nog op terug in hoofdstuk 3 (staafdiagrammen).

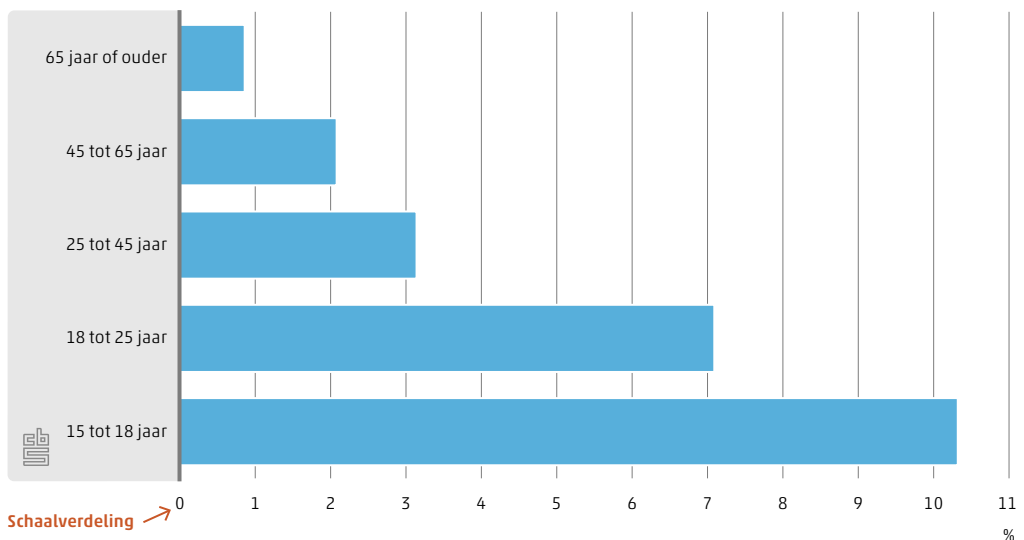
## 2.4 Slachtofferschap cyberpesten naar leeftijd, 2012



## 2.5 Schaalverdeling

Om goed te kunnen inschatten welke waarden horen bij de lengte van de staven, moeten grafieken als het staafdiagram in figuur 2.5 een goede schaalverdeling bevatten. De schaalverdeling loopt hier van 0 tot 11. Kleine maatstreepjes (tick marks) onder de horizontale lijn helpen bij het aflezen van de juiste waarden.

### 2.5 Slachtofferschap cyberpesten naar leeftijd, 2012



In figuur 2.5 staan de waarden 0 tot en met 11 bij de schaalverdeling. Je zou kunnen overwegen om minder schaalwaarden (bijvoorbeeld alleen de even getallen 0, 2, 4, ...) of alleen de vijfvouden (0, 5, 10) bij de schaalverdeling te zetten. Dan wordt het aflezen van de lengtes van de staven weer lastiger. En heel veel waarden is ook weer niet goed. Het is zaak om een goed bruikbare middenweg te vinden.

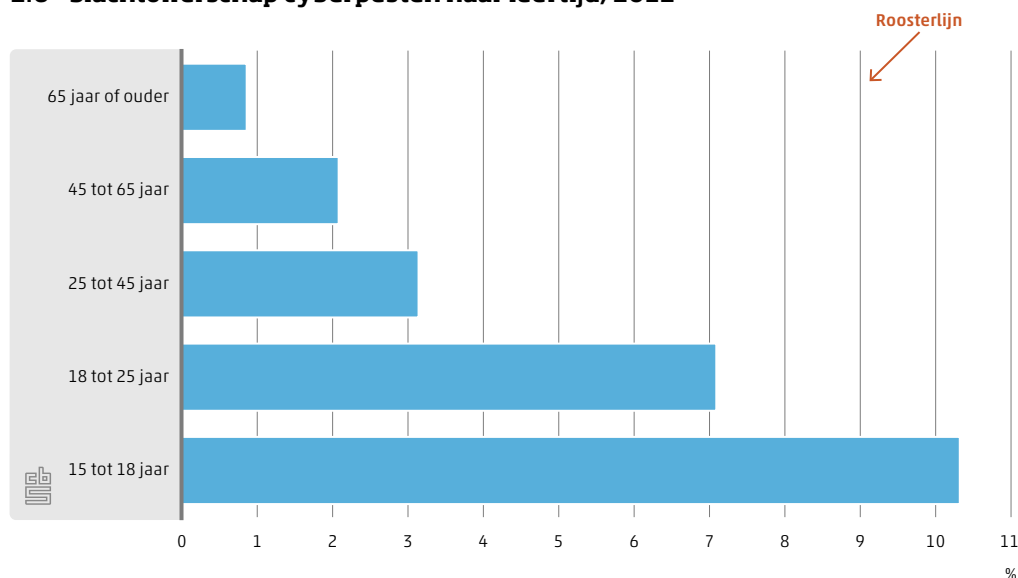
De staven in figuur 2.5 staan voor percentages. Percentages kunnen de waarde 0 hebben. Er is hier daarom sprake van een 'natuurlijk nulpunt'. Als een schaalverdeling zo'n natuurlijk nulpunt heeft, dan moet de horizontale as daar ook mee beginnen. Als je de as met een andere waarde laat beginnen, dan kan dit tot verkeerde interpretatie leiden. Hoofdstuk 6 (Foute grafieken) bevat daarvan een voorbeeld.

Het moet duidelijk zijn wat die getallen precies voorstellen. Een korte tekst of symbool moet uitleggen wat de *meeteenheid* is. In figuur 2.5 gaat het om een percentage. De waarden zullen dus liggen tussen de 0 en de 100.

## 2.6 Roosterlijnen

Om het aflezen van de lengtes van de staven nog makkelijker te maken, kun je roosterlijnen in de grafiek tekenen. In figuur 2.6 zijn dat de verticale grijze lijnen. Het ligt voor de hand om de roosterlijnen te laten beginnen bij de maatstreepjes op de as met de schaalverdeling.

### 2.6 Slachtofferschap cyberpesten naar leeftijd, 2012



De roosterlijnen zijn in een bescheiden kleur getekend. Ze moeten niet te veel op de voorgrond treden. Ze mogen het beeld van de echte gegevens niet vertroebelen. Het mag geen chart junk worden.

## 2.7 Kleuren

Je kunt op allerlei manieren gebruik maken van kleur in een grafiek. Als kleur alleen maar is bedoeld om de grafiek wat op te vrolijken en niet helpt bij het interpreteren van de



grafiek, dan is dit chart junk. Dat is niet goed. Kleur kan echter ook een functioneel gebruik hebben. Kleuren worden dan ingezet om te helpen bij het lezen en begrijpen van de grafiek.

Er gelden regels voor het gebruik van kleuren:

- Gebruik alleen verschillende kleuren als het echt nodig is.
- Bij kwantitatieve variabelen zijn de waarden getallen die echt iets betekenen. Een grote waarde is meer dan een kleine waarde. Gebruik hiervoor verschillende tinten van één kleur. Die tinten variëren van licht naar donker, of omgekeerd, van donker naar licht. Een voorbeeld zijn de verschillende tinten blauw hieronder:



- Bij een kwalitatieve variabele worden de mensen in groepen ingedeeld. Er is geen natuurlijke volgorde. Daarom moet je voor verschillende groepen verschillende kleuren gebruiken. Die kleuren moeten allemaal ongeveer even fel zijn. Hier is een voorbeeld:

### 2.8 Kleuren voor een kwalitatieve variabele



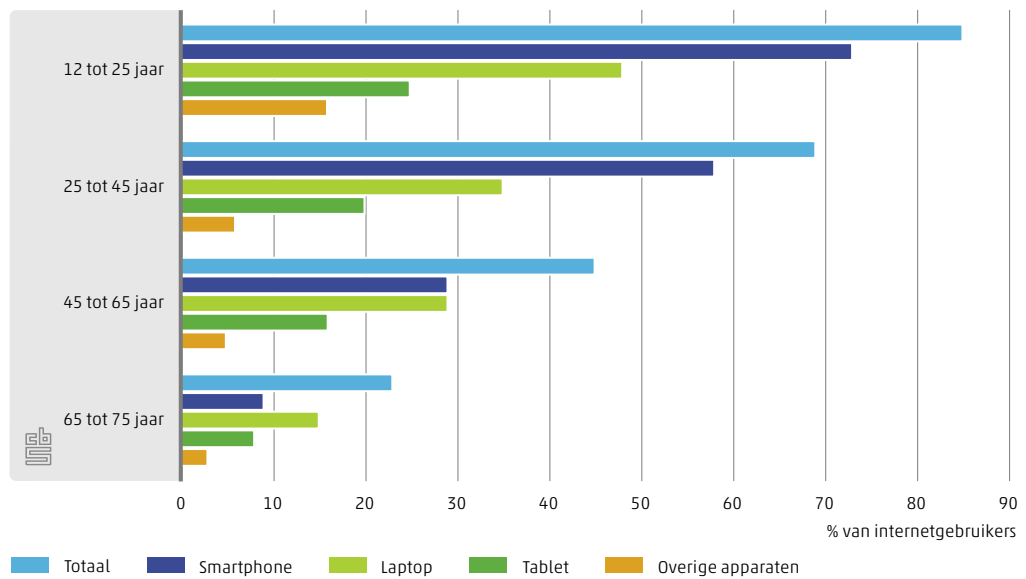
In figuur 1.4 is leeftijd een kwantitatieve variabele. Daarom zijn verschillende tinten blauw gebruikt om de staven te kleuren. De tint is donkerder naarmate iemand ouder is. In figuur 1.5 is het type internetapparaat een kwalitatieve variabele. Daarom zijn daar verschillende kleuren gebruikt om de staven te kleuren.

## 2.8 Legenda

Als je in een staafdiagram maar één variabele weergeeft, dan moet je de namen van de verschillende groepen bij de assen zetten. Dat zie je bijvoorbeeld in figuur 2.1 voor de leeftijdscategorieën. Wil je in één grafiek echter meer variabelen laten zien, dan is dat vaak lastig. Er is vaak geen ruimte om teksten bij al die staven te zetten. De oplossing is dan om gebruik te maken van een legenda. Je geeft de verschillende groepen aan met verschillende kleuren en naast de grafiek geef je aan wat elke kleur betekent.

Figuur 2.9 bevat een voorbeeld van een grafiek met een legenda. Hij staat onder de grafiek. In die legenda kun je bijvoorbeeld zien dat donkerblauw is gebruikt voor een smartphone en donkergroen voor een tablet.

## 2.9 Het gebruik van mobiele apparatuur voor internet door internetgebruikers, naar leeftijd in 2012



Het is duidelijk dat er geen plaats is bij de staven voor de teksten Totaal, Smartphone, enz. Bovendien zou je die teksten dan moeten herhalen voor elke leeftijdsgroep. De volgorde van de groepen in de legenda is dezelfde als de volgorde in de grafiek. Dat maakt het makkelijker om te zien welke staaf hoort bij welk apparaat.

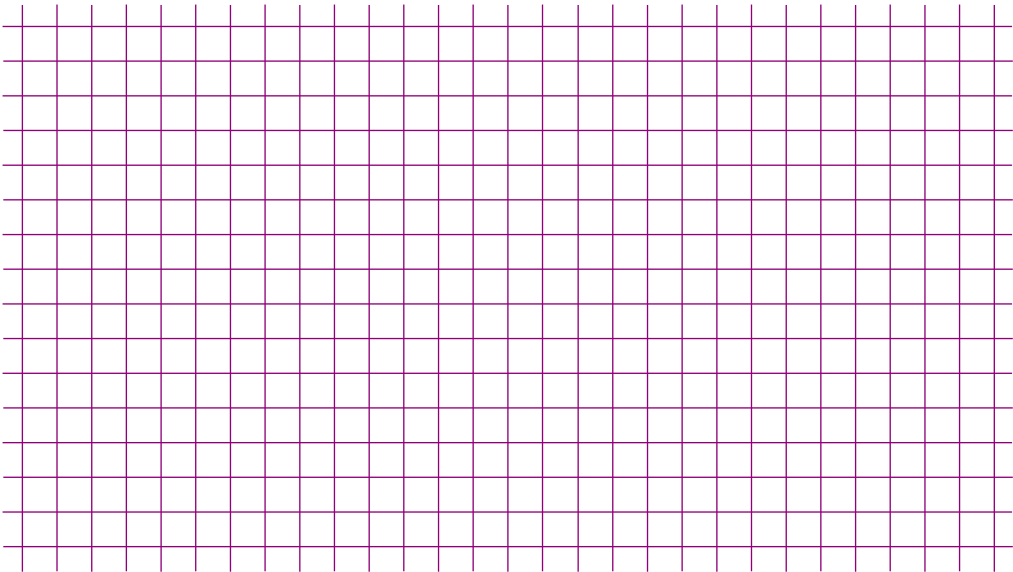
### Vraag 2.1

Het CBS voert elk jaar het Permanent Onderzoek LeefSituatie (POLS) uit. Daarin worden gegevens verzameld over de leefsituatie van de Nederlandse bevolking. Een van die onderwerpen die aan bod komt, is geluidshinder. Hebben de Nederlanders last van lawaai in de woonomgeving? Tabel 2.10 bevat de uitkomsten voor het jaar 2011. Bij de eerste groep (Verkeer/industrie) gaat het om mensen die tegelijk last hebben van lawaai van verkeer en industrie. Bij de andere groepen gaat het om mensen die maar van één soort lawaai last hebben.

## 2.10 Personen die in 2011 last hadden van lawaai, uitgesplitst naar bron van het lawaai

Bron	Percentage
Verkeer/industrie	40
Vliegverkeer	14
Railverkeer	7
Wegverkeer	30
Industrie	3
Buren	20

Maak een grafiek met een staafdiagram van de gegevens in tabel 2.10. Houd je daarbij zoveel mogelijk aan de regels die hiervoor in dit hoofdstuk staan. Gebruik het ruitjespapier.



3.

# Staafdiagrammen

**Voor het weergeven van een kwalitatieve variabele in een de vorm van een grafiek kun je twee typen grafieken gebruiken: het staafdiagram en het cirkeldiagram. In dit hoofdstuk gaat het over het staafdiagram. In het volgende hoofdstuk komt het cirkeldiagram aan de beurt.**

Nog even ter herinnering: een kwalitatieve variabele verdeelt alles in groepen. Het enige wat je met zo'n variabele kunt doen is tellen hoeveel mensen er in de groepen zitten. Eventueel kun je die aantallen omzetten in percentages. Je kunt dan bijvoorbeeld vaststellen welke groep relatief het grootst of het kleinst is.

---

### Vraag 3.1

Geef van elke van de onderstaande variabelen aan of het een kwalitatieve of kwantitatieve variabele is:

a. Het gewicht van een persoon (in kilogram)

.....

b. De kleur van iemands ogen (blauw, bruin, groen, ...)

.....

c. De temperatuur (in graden Celsius) om 12.00 uur in de woonplaats van een persoon

.....

d. Het land waar iemand is geboren

.....

e. Het geslacht van iemand (man, vrouw)

.....

f. Het aantal minuten dat een leerling nodig heeft om van zijn huis naar school te komen

.....

---

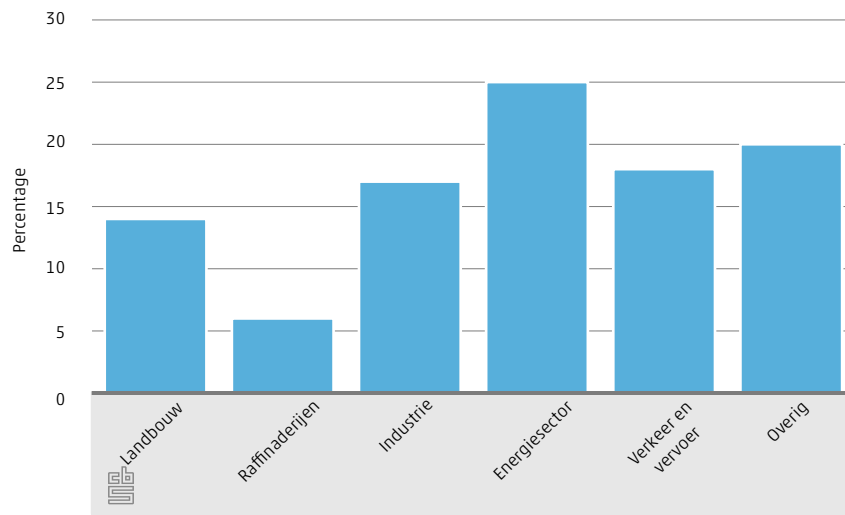
Bij het staafdiagram tekenen we de omvang van de groepen als staven. De lengtes van de staven moeten overeenkomen met de aantallen (of percentages) in de groepen. In de vorige hoofdstukken zijn we al een paar voorbeelden van staafdiagrammen tegengekomen. Hier geven we nog een voorbeeld over broeikasgassen. Die gassen hebben invloed op het klimaat op aarde. Als er teveel broeikasgassen (zoals kooldioxide en lachgas) in de lucht komen, stijgt de temperatuur op aarde. Hierdoor stijgt de zeespiegel en kunnen planten of dieren verdwijnen. Ons land heeft met andere landen afspraken gemaakt om de uitstoot van broeikasgassen te verminderen. Het CBS houdt bij hoeveel broeikasgassen er in Nederland in de lucht terecht komen en waar die gassen precies vandaan komen. Dat zie je in de vorm van een staafdiagram in figuur 3.1.

## 3.1 Horizontaal of verticaal

Je kunt een staafdiagram op twee manieren tekenen: met verticale staven zoals in figuur 3.1 of met horizontale staven zoals in figuur 3.2. De grafiek met verticale staven heeft twee nadelen:

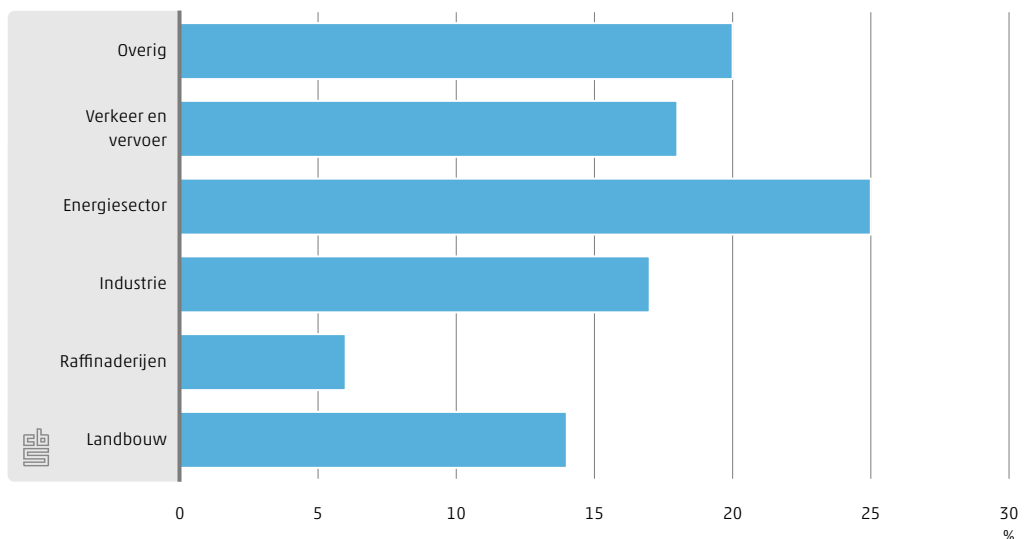
- Die grafiek lijkt erg op een andere type grafiek. Dat is het histogram. Je gebruikt een histogram voor een kwantitatieve variabele. Om verwarring te voorkomen, kun je het staafdiagram maar beter met horizontale staven tekenen.
- In een staafdiagram met verticale staven is er maar weinig ruimte voor teksten. Dat kun je oplossen door die teksten schuin neer te zetten, zoals de namen van de staven in figuur 3.1. De naam van de variabele bij de verticale as is zelfs helemaal dwars neergezet. Dit maakt de grafiek nogal moeilijk leesbaar.

### 3.1 Een staafdiagram met verticale staven over de bronnen van broeikasgassen



De voorkeur gaat dus uit naar een staafdiagram met horizontale staven, zoals in figuur 3.2. Daarin kun je voldoende ruimte maken voor goed leesbare teksten.

### 3.2 Een staafdiagram met horizontale staven over de bronnen van broeikasgassen



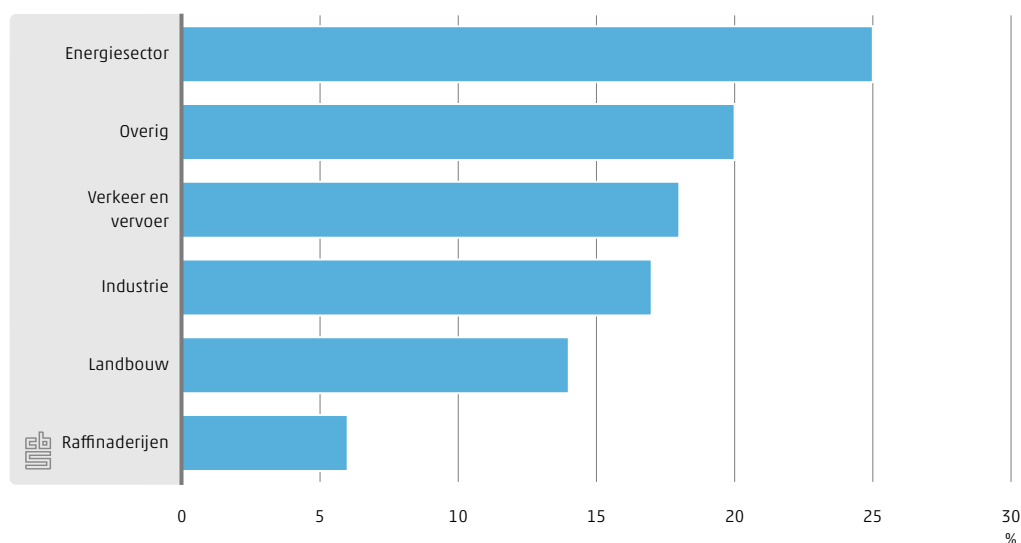
De staven in het staafdiagram hebben allemaal dezelfde kleur. Omdat de staven goed van elkaar zijn gescheiden, kunnen ze allemaal dezelfde kleur krijgen. Het heeft geen zin om ze allemaal een andere kleur te geven. Dit voegt niets toe aan de grafiek. Het leidt alleen maar af.

Het voordeel van die ene kleur zoals het blauw in figuur 3.2 is dat de gegevens zelf goed in het oog springen.

Hoe groot moet de ruimte tussen de staven zijn? Te veel ruimte is niet goed en te weinig ruimte is niet goed. Er is een vuistregel die zegt dat de ruimte tussen de staven ongeveer de helft moet zijn van de breedte van de staven. Dat is het geval in figuur 3.2.

In figuur 3.2 zijn de staven in dezelfde volgorde getekend als waarin de bijbehorende groepen in de tabel stonden. Als die groepen geen natuurlijke, vanzelfsprekende volgorde hebben, kun je overwegen de staven van klein naar groot te tekenen, of omgekeerd van groot naar klein. Je krijgt dan een grafiek zoals in figuur 3.3.

### 3.3 Een staafdiagram met staven van groot naar klein over de bronnen van broeikasgassen



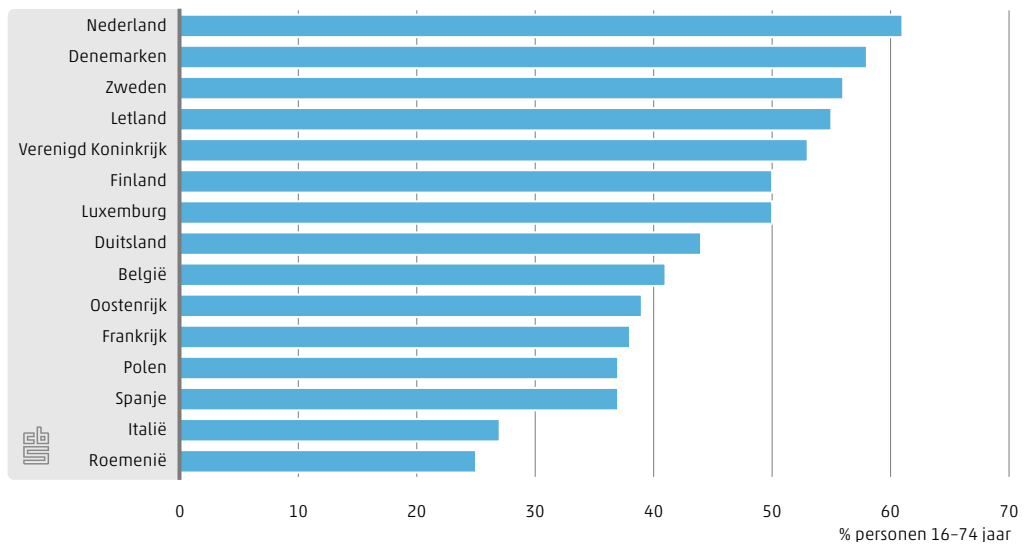
In zo'n geordende grafiek kun je makkelijk aflezen welke groep de grootste en welke de kleinste is. Ook is makkelijker te zien welke groep groter of kleiner is dan een andere groep.

## 3.2 Doorknippen van de staven?

De Nederlanders zijn erg actief met de sociale media op het internet. Bij sociale media gaat het niet alleen om sociale netwerken zoals Twitter, Facebook en LinkedIn, maar ook om het meedoen met discussieforums, het uitwisselen van tekstberichten en het lezen en bijhouden van weblogs. Hoe zit het met het gebruik van sociale media in de rest van Europa? De meest recente gegevens hierover dateren van 2011. Het CBS heeft er een grafiek van gemaakt en die staat in figuur 3.4.

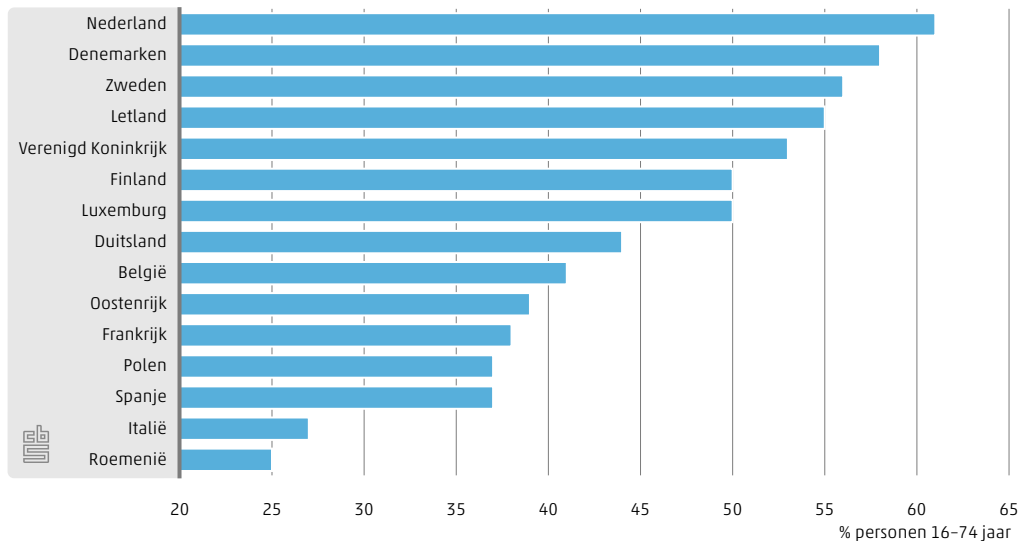
Al die lange staven nemen nogal wat ruimte in beslag. Het linkerdeel van de grafiek is ook niet zo interessant. Het gaat er vooral om waar de staven eindigen aan de rechterkant, en niet waar ze beginnen aan de linkerkant. Daarom komen makers van grafieken wel eens op het idee om een deel van die staven weg te laten. Dat kan op verschillende manieren. In grafiek 3.5 is dit gedaan door de horizontale as te laten beginnen bij 20 in plaats van bij 0.

### 3.4 Het gebruik van sociale media in Europa, 2011



Mensen zien makkelijk over het hoofd dat de as niet bij 0 begint. Daardoor kunnen ze een verkeerde indruk krijgen. Zo lijkt het er bij eerste oogopslag op dat in Duitsland maar half zoveel mensen sociale media gebruiken als in Nederland. De staaf van Duitsland is immers ongeveer half zo lang als die van Nederland. Maar als je naar de cijfers kijkt, dan klopt dat niet. In Nederland is het percentage 61 procent en in Duitsland is het 44 procent. Daar zit helemaal geen factor 2 tussen. Kortom, het is niet goed om de beginwaarde van de as te veranderen.

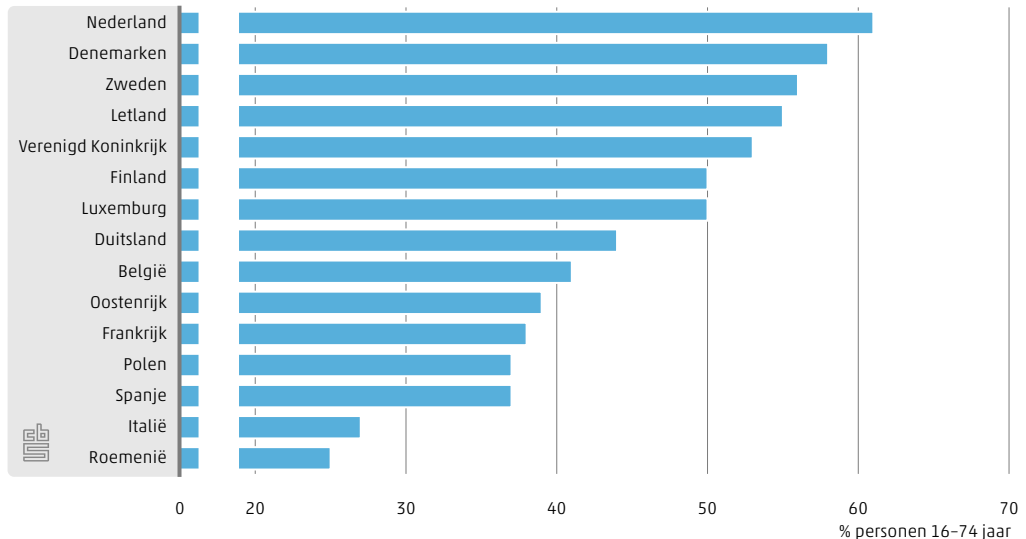
### 3.5 Het gebruik van sociale media in Europa, 2011



Figuur 3.6 laat nog een andere manier zien om de staven korter te maken. De as begint wel bij 0, maar er is een stuk tussenuit gehaald. Dat kun je zien aan de schaalwaarden onder de as: het stukje tussen 0 en 20 is een stuk korter als het stuk tussen 20 en 30. Ook is een witte verticale band door de staven getekend om aan te geven dat de staven doorgeknipt zijn.

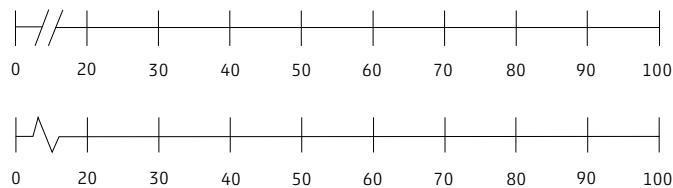


### 3.6 Het gebruik van sociale media in Europa, 2011



De grafiek in figuur 3.6 heeft hetzelfde probleem als de grafiek in figuur 3.5: je wordt al snel op het verkeerde been gezet als je alleen naar de lengte van de staven kijkt en niet naar de schaalwaarden bij de as. Als het even kan moet het doorknippen van de staven worden vermeden. En als het toch moet, geef het doorknippen dan ook duidelijk aan bij de horizontale as. Figuur 3.7 bevat twee gangbare manieren om dit te doen.

### 3.7 Het aangeven van een knip in de horizontale as



## 3.3 3D-grafieken?

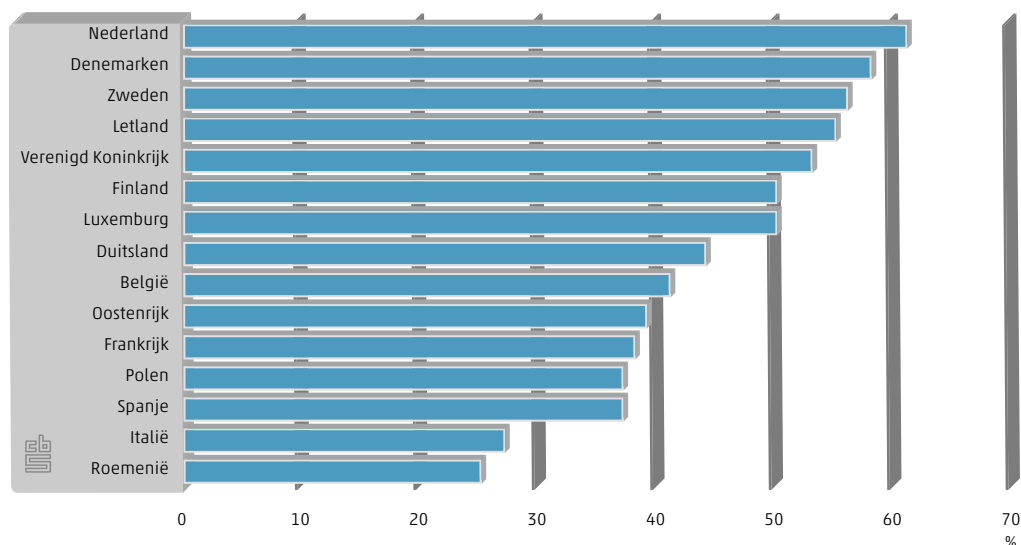
Grafieken voor publicaties worden soms gemaakt door vormgevers die niet altijd veel verstand van statistiek hebben. Ze vinden grafieken vaak saai en proberen ze daarom op allerlei manieren op te vrolijken. Dat kan bijvoorbeeld met chart junk, maar ook door er een driedimensionaal perspectief aan te geven. Dat leidt misschien wel tot leukere plaatjes, maar lang niet altijd tot beter interpreteerbare grafieken.

Een voorbeeld van zo'n driedimensionaal perspectief is te vinden in figuur 3.8. Daarin staan dezelfde gegevens als in figuur 3.4. Alleen zijn de staven nu vervangen door balkjes. Het is lastig om de lengte van de balken precies af te lezen. Moet je kijken naar de voorkant van de balk of de achterkant?

Met allerlei computerprogramma's is het vrij eenvoudig om grafieken als in figuur 3.8 te maken. Een voorbeeld is het spreadsheet-programma Excel van Microsoft. Dat is op heel

veel computers aanwezig. Daardoor zie je vaak grafieken in publicaties die er voor het oog leuk uitzien, maar die niet de goede 'boodschap' overbrengen. Het is maar beter om geen driedimensionaal perspectief te gebruiken bij grafieken.

### 3.8 Het gebruik van sociale media in Europa, 2011



## 3.4 Staafdiagrammen voor twee variabelen

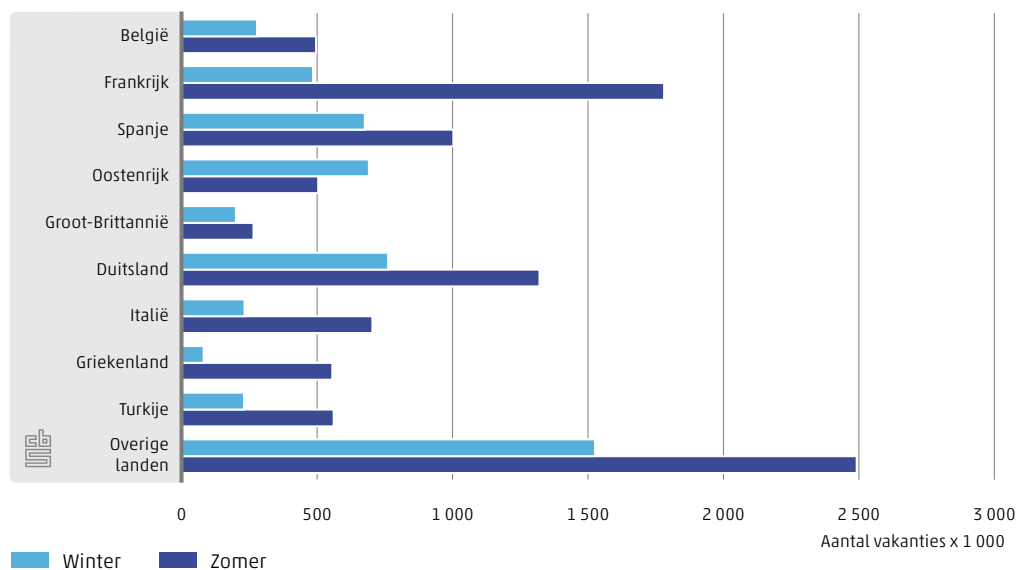
Bij de staafdiagrammen die we tot nog toe hebben bekeken in dit hoofdstuk ging het steeds om het in beeld brengen van één kwalitatieve variabele: bron van de broeikasgassen of Europees land. Soms komt het ook voor dat je twee kwalitatieve variabelen tegelijk in beeld wilt brengen. Dat levert meer informatie op. Het kan je bijvoorbeeld vertellen of er misschien een verband bestaat tussen die twee variabelen.

Het CBS voert elk jaar het Continu Vakantie Onderzoek (CVO) uit. Doel van dit onderzoek is het verzamelen van informatie over korte en lange binnen- en buitenlandse vakanties. In het CVO wordt gevraagd hoe lang mensen op vakantie gaan, waar ze naar toe gaan, hoeveel geld ze uitgeven, waar ze slapen (tent, hotel, enz.) en hoe ze op vakantie gaan (auto, vliegtuig, enz.).

De uitkomsten van het onderzoek laten dus bijvoorbeeld zien hoe vaak mensen op vakantie gaan, waar ze naar toe gaan en of ze in de zomer of in de winter op vakantie gaan. Je kunt je dan afvragen of er verband bestaat tussen het land van bestemming en tijd van het jaar (zomer of winter). Gaan mensen vooral in de zomer naar bepaalde landen en in de winter naar andere landen? Zoiets kun je zichtbaar maken met een staafdiagram, maar dan geen gewoon staafdiagram. Het is een wat ingewikkelder staafdiagram, we laten hier drie vormen zien: het samengesteld staafdiagram, het stapeldiagram, en het stapeldiagram met percentages.

Figuur 3.9 bevat een voorbeeld van een *samengesteld staafdiagram*. Voor elk groep van de ene variabele (land van bestemming) is een staafdiagram getekend van de andere variabele (tijd van het jaar). Er is dus een reeks van staafdiagrammen getekend. Om de gegevens voor de verschillende landen goed van elkaar te kunnen onderscheiden, zit er witruimte tussen de staafdiagrammen. Binnen een land zijn de staven tegen elkaar aan getekend. Om die staven goed van elkaar te kunnen onderscheiden, moeten ze een verschillende kleur krijgen.

### 3.9 Staafdiagram van lange vakanties in het buitenland, per land van bestemming, naar tijd van het jaar, in 2012



In grafiek 3.9 gaat het alleen om lange vakanties. Dat zijn vakanties met minstens vier overnachtingen. Let wel op bij het aflezen van waarden op de horizontale as. Die waarden lopen van 0 tot 3 000, maar onder de as staat ook 'Aantal vakanties x 1 000'. Die 'x 1 000' betekent dat je alle waarden met 1 000 moet vermenigvuldigen. Dus het aantal zomervakanties in Oostenrijk is niet 500, maar 500 000.

#### Vraag 3.2

Probeer de informatie in de grafiek in figuur 3.9 te interpreteren. Geef antwoord op de volgende vragen:

a. Wat zijn de drie belangrijkste vakantiebestemmingen in de zomer?

.....

b. Wat zijn de vier belangrijkste vakantiebestemmingen in de winter?

.....

c. Welke land is in de winter populairder dan in de zomer?

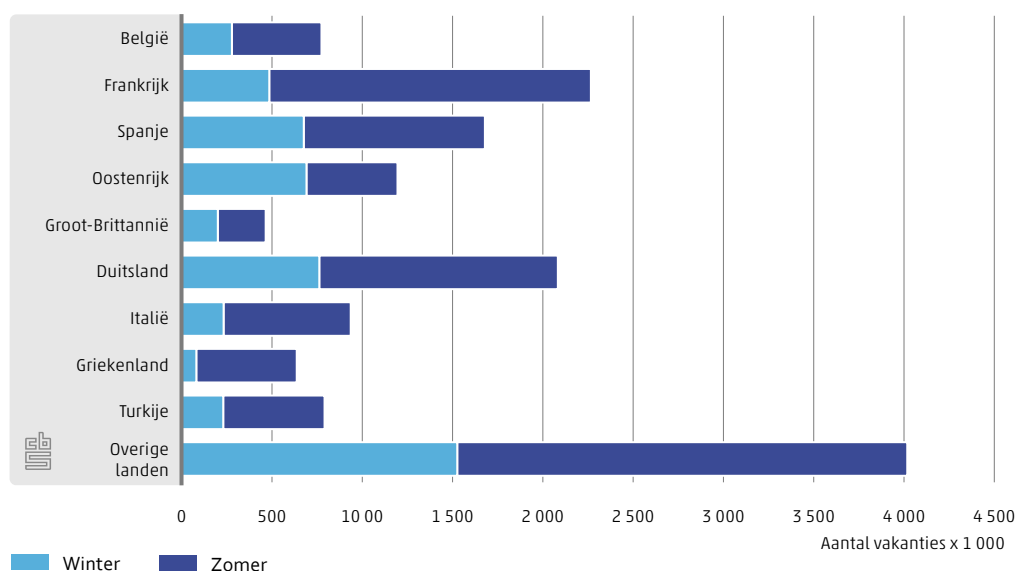
.....

d. Hoeveel zomervakanties zijn in België doorgebracht?

.....

Uit grafiek 3.9 kun je allerlei interessante feiten halen, maar je kunt niet alle informatie er even makkelijk uit halen. Zo kun je niet het totaal aantal vakanties per land (winter- en zomervakanties samen) aflezen. Daarvoor zou je eigenlijk de winterstaaf en de zomerstaaf achter elkaar moeten tekenen in plaats van onder elkaar. Dat kan, maar dan krijg je een andere type grafiek: het stapeldiagram. Figuur 3.10 bevat een voorbeeld van zo'n stapeldiagram.

### 3.10 Stapeldiagram van lange vakanties in het buitenland, per land van bestemming, naar tijd van het jaar, in 2012

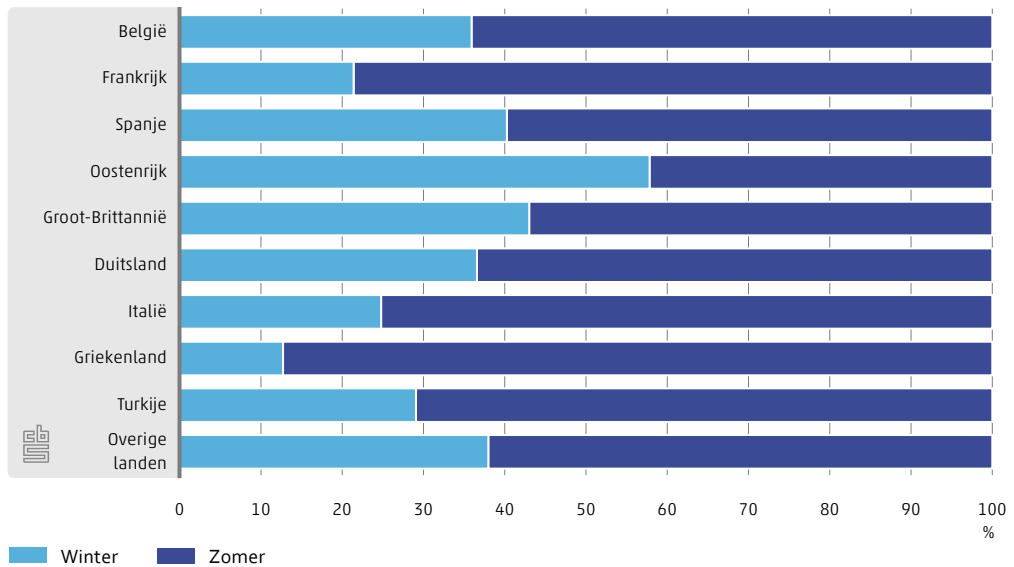


Afgezien van de groep 'Overige landen' is Frankrijk het populairste vakantieland. Daar worden de meeste vakanties doorgebracht. Frankrijk wordt op de voet gevolgd door Duitsland. Spanje staat op de derde plaats. In deze grafiek is het weer lastiger om de zomervakanties met elkaar te vergelijken. Waren er nu meer zomervakanties in Griekenland dan in Turkije? Zo'n vergelijking is moeilijk omdat de staven niet op hetzelfde punt beginnen. Het samengesteld staafdiagram is daar weer beter voor.

Er is nog een derde manier om twee kwalitatieve variabelen te bekijken en dat is het stapeldiagram met percentages. Die is vooral bedoeld om de verhoudingen binnen de groepen met elkaar te vergelijken. Zo zou het kunnen dat je wilt weten in welke landen het aandeel wintervakanties groot is en in welk land juist klein. Dit soort vragen kun je het beste beantwoorden met een grafiek als in figuur 3.11.

De staven in figuur 3.11 zijn allemaal even lang. Elke staaf stelt 100% van de vakanties naar dat land voor. De staven worden gesplitst in een aandeel wintervakanties en een aandeel zomervakanties.

### 3.11 Stapeldiagram met percentages van lange vakanties in het buitenland, per land van bestemming, naar tijd van het jaar, in 2012



#### Vraag 3.3

Probeer de informatie in de grafiek in figuur 3.11 te interpreteren. Geef antwoord op de volgende vragen:

- a. In welk land zijn relatief de meeste zomervakanties doorgebracht?

.....

- b. In welk land zijn relatief de minste zomervakanties doorgebracht?

.....

- c. Welke zaken kun je niet aflezen uit de grafiek?

.....

#### Vraag 3.4

Van alle huishoudens in Nederland heeft 28 procent geen auto. De helft van de huishoudens (50 %) heeft één auto. En de resterende 22 procent van de huishoudens heeft 2 of meer auto's. In de grafieken hieronder is deze informatie in beeld gebracht. Geef bij elke grafiek aan of het een goede grafiek is. Als het geen goede grafiek is, leg dan uit waarom niet.

.....

.....

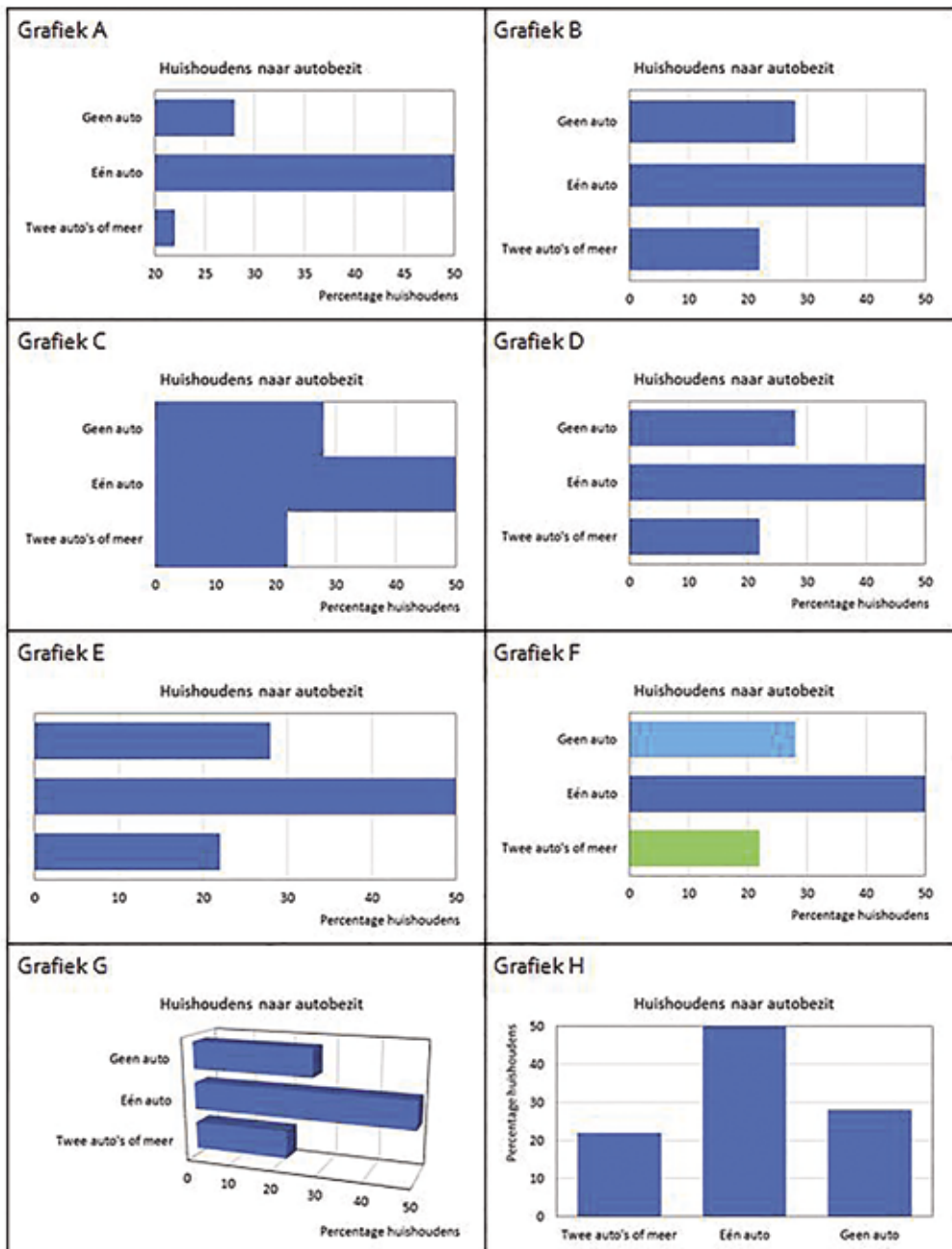
.....

.....

.....

.....

.....



4.

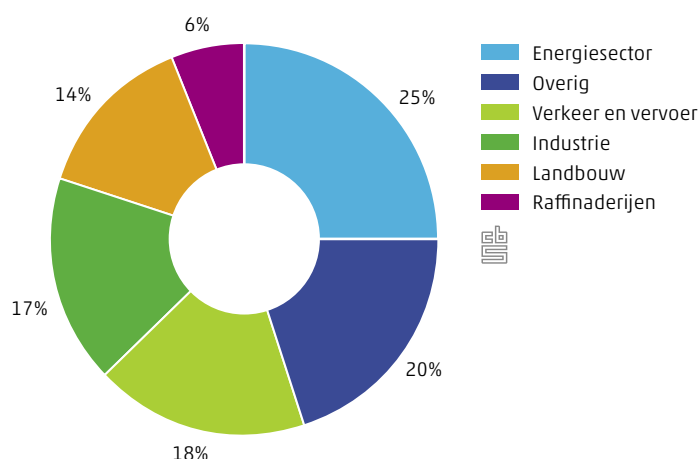
# Cirkel- diagrammen

**Je kunt een kwalitatieve variabele (een variabele die alles in groepen verdeelt) weergeven met een staafdiagram. Dat hebben we in de vorige hoofdstukken gezien. Het kan ook nog op een andere manier. Dat is met een *cirkeldiagram*. Een populaire naam voor het cirkeldiagram is het *taartdiagram*. Als je een cirkeldiagram ziet, dan is meteen duidelijk waarom: een cirkeldiagram lijkt op een taart.**

Cirkeldiagrammen zie je vooral veel in de media (radio, tv, internet). De voorkeur gaat vaak niet uit naar cirkeldiagrammen omdat ze de 'boodschap' in de gegevens goed overbrengen, maar omdat ze er leuker uitzien. En dat is dus geen goede reden.

Bij een cirkeldiagram verdelen we een cirkel (de taart) in evenveel parten (punten) als de variabele groepen (categorieën) heeft. De oppervlaktes van deze parten weerspiegelen de grootte van de groep: is een groep twee keer zo groot, dan is ook de taartpunt twee keer zo groot. Figuur 4.1 bevat een voorbeeld van een cirkeldiagram. De figuur geeft dezelfde gegevens weer als het staafdiagram in figuur 3.3. Het gaat om broeikasgassen in Nederland. Het CBS houdt bij hoeveel broeikasgassen er in Nederland in de lucht terecht komen en waar die gassen precies vandaan komen.

#### 4.1 Een cirkeldiagram van de bronnen van broeikasgassen in Nederland



Een taart als in figuur 4.1 ziet er wat minder saai uit dan een staafdiagram, zeker als hij vrolijk gekleurd is. Het is wel zo dat je cirkeldiagrammen minder makkelijk kunt interpreteren. Het is eenvoudiger om lengtes van staven met elkaar te vergelijken dan de sectoren in de cirkel. Dat pleit voor het gebruik van staafdiagrammen in plaats van cirkeldiagrammen. Om de interpretatie van het cirkeldiagram te vereenvoudigen, zie je vaak dat de percentages (of aantallen) bij de parten zijn gezet. Dat is ook gebeurd in figuur 4.1.

Er is een legenda naast het cirkeldiagram gezet om de verschillende groepen te kunnen terugvinden in de grafiek. Je had ook kunnen proberen om de namen van de groepen in of bij de sectoren te zetten, maar daar is meestal onvoldoende plaats voor.

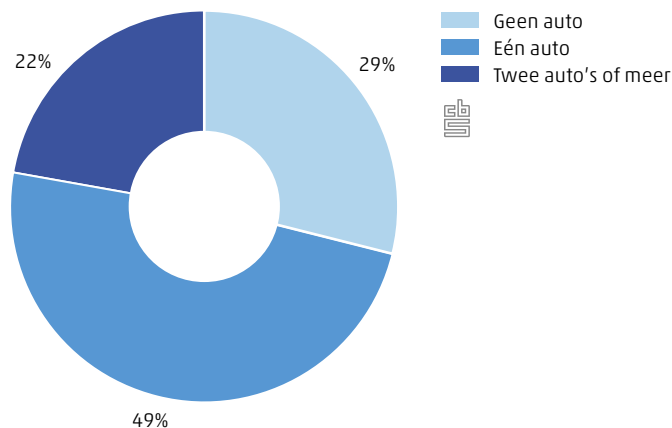
De verschillende groepen van de kwalitatieve variabele hebben meestal geen natuurlijke ordening. Daarom moet je verschillende kleuren gebruiken om de verschillende groepen in het cirkeldiagram goed van elkaar te onderscheiden. Die kleuren moet je zo kiezen dat



de ze allemaal ongeveer even fel zijn. De ene kleur mag niet feller zijn (en daardoor meer opvallen) dan de andere kleur.

Hebben de groepen van de kwalitatieve variabele wel een natuurlijke ordening (denk aan een variabele als opleidingsniveau), dan is het beter om verschillende tinten van één kleur te gebruiken. Figuur 4.2 toont daarvan een voorbeeld. Hier is de kwalitatieve variabele het aantal auto's per huishouden. Naarmate er meer auto's zijn in een huishouden, krijgt de bijbehorende sector in de grafiek een donkerder kleur blauw.

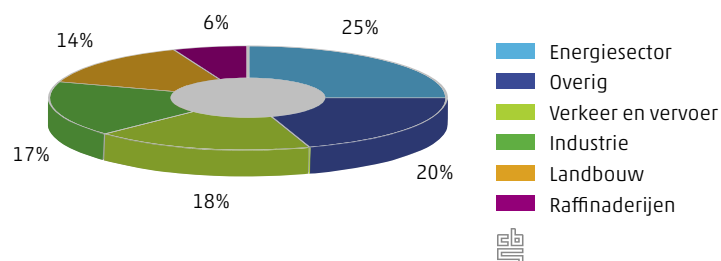
#### 4.2 Een cirkeldiagram van het autobezit per huishouden



Soms kunnen vormgevers de verleiding niet weerstaan om cirkeldiagrammen in een driedimensionaal perspectief te tekenen. Bij de staafdiagrammen is al uitgelegd dat dit de grafiek onduidelijker maakt. Hetzelfde geldt voor cirkeldiagrammen. Ook daar komt dit driedimensionale perspectief de interpreteerbaarheid van de grafiek niet ten goede.

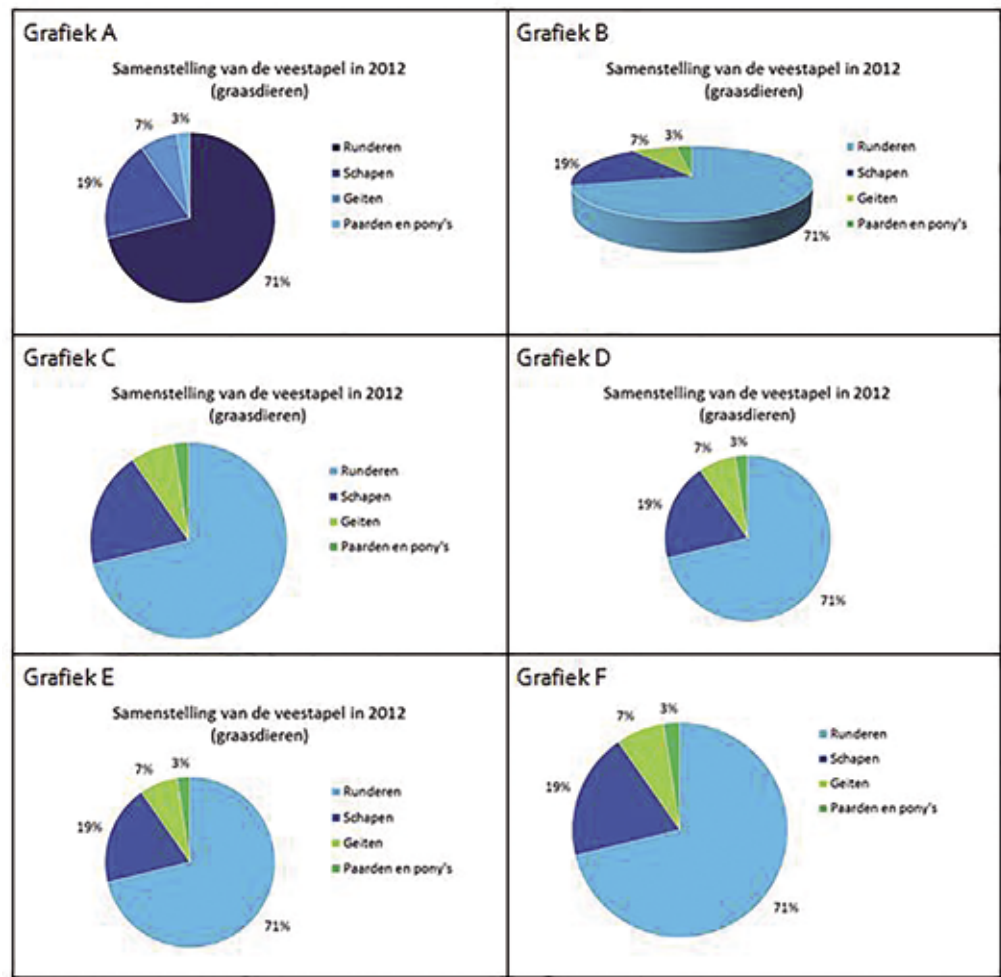
In figuur 4.3 zie je een voorbeeld van een cirkeldiagram met driedimensionaal perspectief. Door dit perspectief zijn de vormen van de verschillende sectoren verwrongen. Daarom is het lastig ze met elkaar te vergelijken. Verder is het zo dat de sectoren aan de voorkant belangrijker lijken omdat daar de dikte van de sectoren ook te zien is. Gezien de problemen met het aflezen van cirkeldiagrammen is het maar beter om dit type grafiek niet te gebruiken. Het gebruik van driedimensionaal perspectief is al helemaal uit den boze.

#### 4.3 Een cirkeldiagram met driedimensionaal perspectief van de bronnen van broeikasgassen in Nederland



Vraag 4.1

De veestapel bestaat in Nederland uit vele miljoenen dieren. Die veestapel bestaat uit graasdieren en hokdieren. De graasdieren staan buiten in de wei. Maar hoe is die veestapel van graasdieren samengesteld? Wat is het aandeel runderen, schapen, geiten, paarden en pony's? Dat is weergegeven in onderstaande cirkeldiagrammen. Geef bij elke grafiek aan of het een goede grafiek is of niet. Als het geen goede grafiek is, leg dan uit waarom niet.



.....

.....

.....

.....

.....

.....

.....

.....

**5.**

# Tijdreeksen

Onderzoek van instituten als het CBS is er heel vaak op gericht om veranderingen in de tijd in kaart te brengen. Hoe gaat het met de werkloosheid in Nederland? Neemt die toe of neemt die af? Hoe gaat het met de prijzen van huizen? Worden huizen goedkoper of duurder? Is er sprake van global warming? Neemt de temperatuur op aarde toe of af? Bij al deze voorbeelden wil je dus eigenlijk weten hoe de waarde van een variabele verandert in de loop van de tijd. Om daarachter te komen moet je zo'n variabele op regelmatige momenten meten. Sommige variabelen meet je misschien wel elke dag (temperatuur). Bij andere variabelen kan het minder vaak zijn, bijvoorbeeld eens per maand (werkloosheid) of eens per jaar (toegang tot internet). De gegevens die we zo krijgen, noemen we *tijdreeksen*.

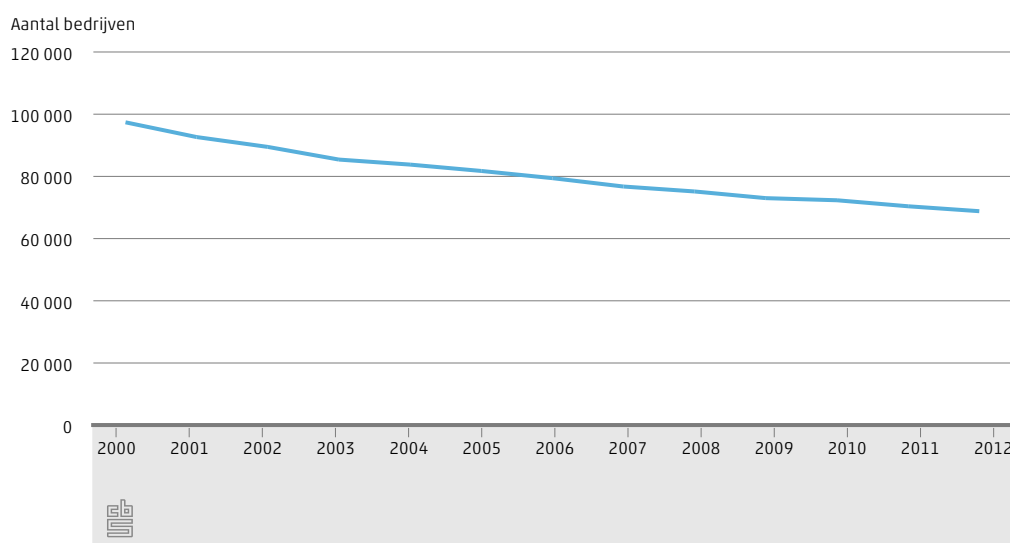
We kunnen twee typen grafieken gebruiken voor het weergeven van tijdreeksen: het lijndiagram en het staafdiagram. Beide typen zullen we bespreken. Lijndiagrammen en staafdiagrammen kunnen goed veranderingen in de tijd van een variabele in beeld brengen. Die variabele kan zowel een kwalitatieve als een kwantitatieve variabele zijn.

Bij tijdreeksen geeft een van de assen de tijd weer. Het is gebruikelijk om daarvoor de horizontale as te nemen. Die as heeft meestal geen nulpunt. Er is geen natuurlijk nulpunt voor de tijd. Je zou natuurlijk het begin van onze jaartelling als nulpunt kunnen nemen, maar dat is maar een willekeurig moment in de tijd. Andere volken en culturen (Joden, Chinezen, Maya's) hebben andere kalenders en dus andere nulpunten. Laat dus het nulpunt maar weg en beperk je tot de periode in de tijd waarover je gegevens hebt.

## 5.1 Tijdreeks voor een kwantitatieve variabele

De meest voor de hand liggende manier om een tijdreeks weer te geven is met een lijndiagram. De grafiek bevat voor elk tijdstip de waarde van de variabele. En die punten

### 5.1 Een tijdreeks in de vorm van een lijndiagram over land- en tuinbouwbedrijven



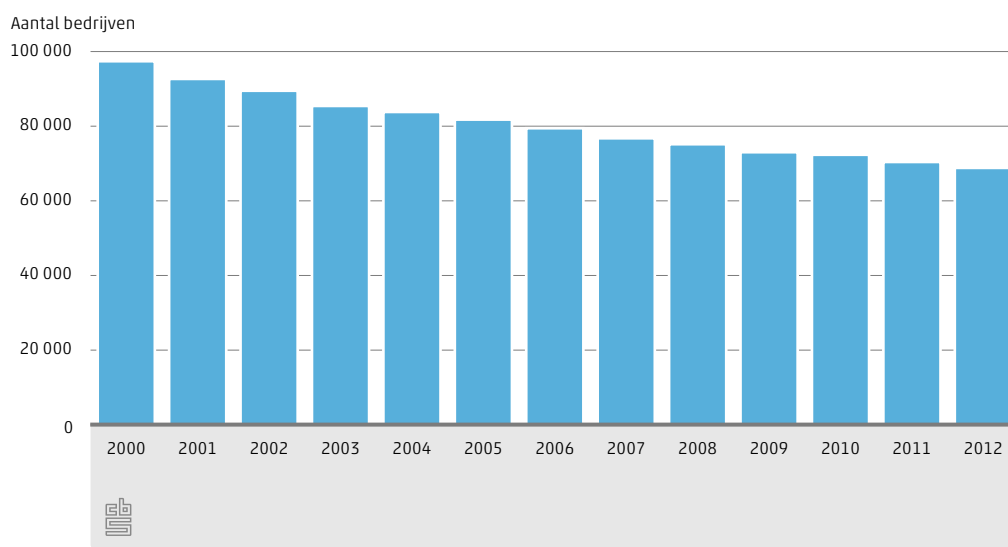
zijn met elkaar verbonden door een curve (kromme). Figuur 5.1 bevat een voorbeeld. Het gaat hier om het in beeld brengen van het aantal landbouwbedrijven in Nederland door de tijd heen. In de grafiek is duidelijk te zien dat het aantal landbouwbedrijven afneemt. In het jaar 2000 waren er nog bijna 100 000 bedrijven. In 2012 is dat aantal afgenomen tot ongeveer 70 000.

De aantallen per jaar, de punten, zijn in de grafiek niet zichtbaar. Je kunt ze eventueel wel laten zien door ze met een speciaal symbool (cirkeltje, blokje) weer te geven, maar dit leidt wel de aandacht af van het patroon dat in de lijn zit.

Als je de punten door lijnstukken verbindt, dan geef je daarmee wel aan dat de ontwikkeling tussen twee opeenvolgende tijdstippen geleidelijk verloopt. Dat hoeft niet altijd het geval te zijn.

Als de variabele steeds is gemeten op aparte tijdstippen en daartussen niet, en je wilt dat benadrukken, dan kan je beter het lijndiagram vervangen door een staafdiagram, zoals in figuur 5.2. Dit staafdiagram is beter om de stand van zaken op een bepaald moment af te lezen, terwijl het lijndiagram zich vooral leent om juist de ontwikkelingen in de tijd te bestuderen.

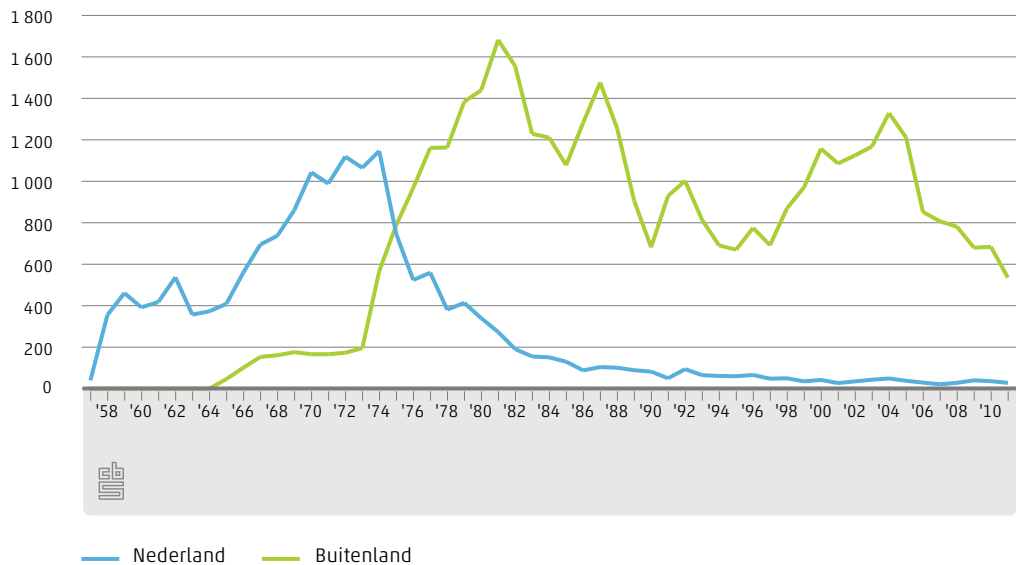
## 5.2 Een tijdreeks in de vorm van een staafdiagram over land- en tuinbouwbedrijven



Het is vrij simpel om in een lijndiagram verschillende variabelen tegelijk tegen de tijd af te zetten. Zeker als het gaat om een beperkt aantal variabelen, blijft interpretatie van de grafiek eenvoudig. In figuur 5.3 staat het aantal adoptiekinderen door de tijd heen weergegeven. De grafiek bevat twee variabelen: het aantal adoptiekinderen afkomstig uit Nederland zelf en het aantal adoptiekinderen dat uit het buitenland afkomstig is.

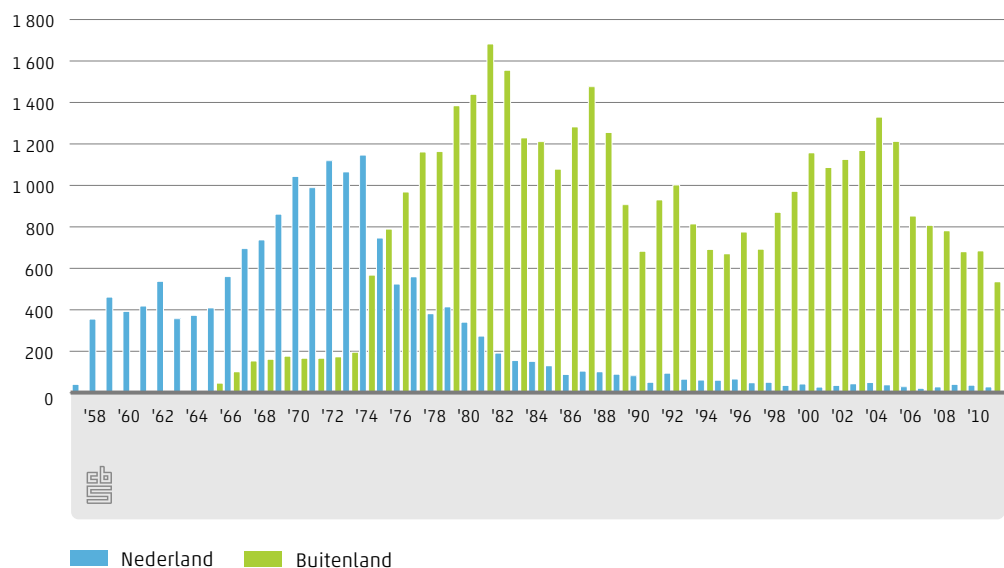
Duidelijk is te zien hoe omstreeks 1974 het aantal adoptiekinderen uit Nederland sterk afneemt en het aantal adoptiekinderen uit het buitenland sterk groeit. Uit de grafiek valt niet goed af te lezen hoe het verloop van het totaal aantal adoptiekinderen in de tijd is.

### 5.3 Een lijndiagram met twee variabelen over het aantal adoptiekinderen naar herkomst



Figuur 5.4 laat zien wat er gebeurt als de grafiek in figuur 5.3 wordt vervangen door een staafdiagram. Het grote aantal staven in combinatie met meer dan één variabele vermindert een eenvoudige interpretatie van de grafiek. Het lijkt zelfs alsof er nog een derde variabele in de grafiek staat (tussen 1967 en 1987), maar dat is een optisch effect.

### 5.4 Een staafdiagram met twee variabelen over het aantal adoptiekinderen naar herkomst



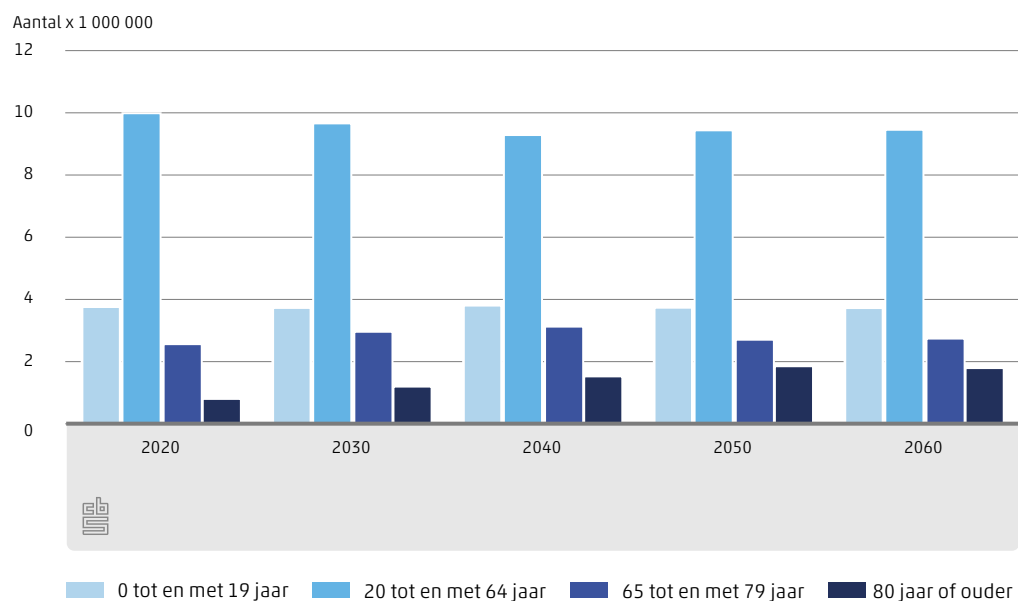
Het weergeven van twee variabelen in een tijddiagram, zoals in figuur 5.3, kan goed werken. Het is echter niet zo dat je willekeurig veel variabelen in de grafiek kunt opnemen. Op een gegeven moment wordt de grafiek onleesbaar.

## 5.2 Tijdreeksen voor een kwalitatieve variabele

Ook bij een kwalitatieve variabele heb je de keuze om de tijdreeks te tekenen met een staafdiagram of een lijndiagram. Als je kiest voor een staafdiagram, dan zijn er nog twee varianten mogelijk: een samengesteld staafdiagram en een stapeldiagram.

Figuur 5.5 bevat een bevolkingsprognose van het CBS. Die prognose probeert te voorspellen hoeveel mensen er in de toekomst in Nederland wonen. De prognose geeft ook aan of de verdeling over de verschillende leeftijdsgroepen verandert.

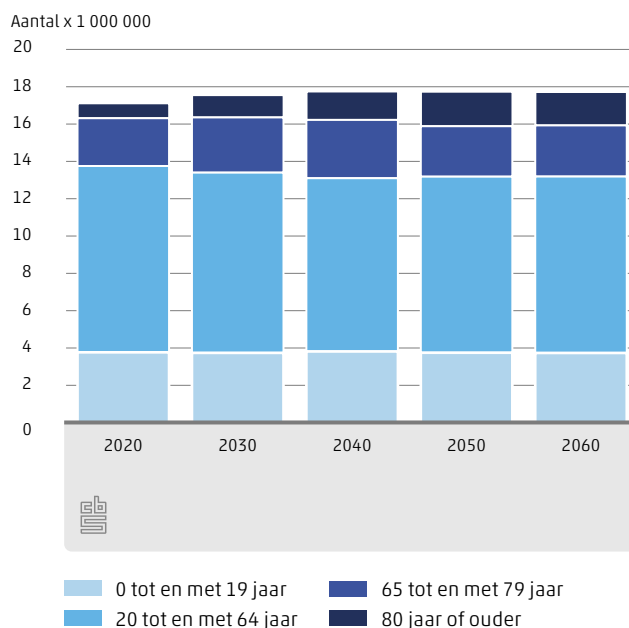
### 5.5 Een tijdreeks als samengesteld staafdiagram over de bevolkingsprognose naar leeftijd



De kwalitatieve variabele leeftijd is verdeeld in vier groepen: 0–19 jaar, 20–64 jaar, 65–79 jaar en 80 jaar en ouder (80+). Figuur 5.5 toont de bevolkingsprognose in de vorm van een samengesteld staafdiagram. Hierbij kun je de verandering in de tijd van de verschillende leeftijdsgroepen goed volgen. Zo is duidelijk een toename te zien van het aantal personen van 80 jaar en ouder. En het aantal mensen tussen de 20 en 64 jaar lijkt tot 2040 af te nemen en daarna weer licht toe te nemen.

Uit figuur 5.5 valt niet af te lezen wat er gebeurt met de bevolking als geheel. Daarvoor is het beter om een stapeldiagram te maken. Figuur 5.6 bevat de bevolkingsprognose als een stapeldiagram. Nu is te zien dat de bevolking eerst nog licht toeneemt en daarna afneemt. Het is echter lastiger om het verloop van de omvang van de verschillende leeftijdsgroepen in de tijd te volgen.

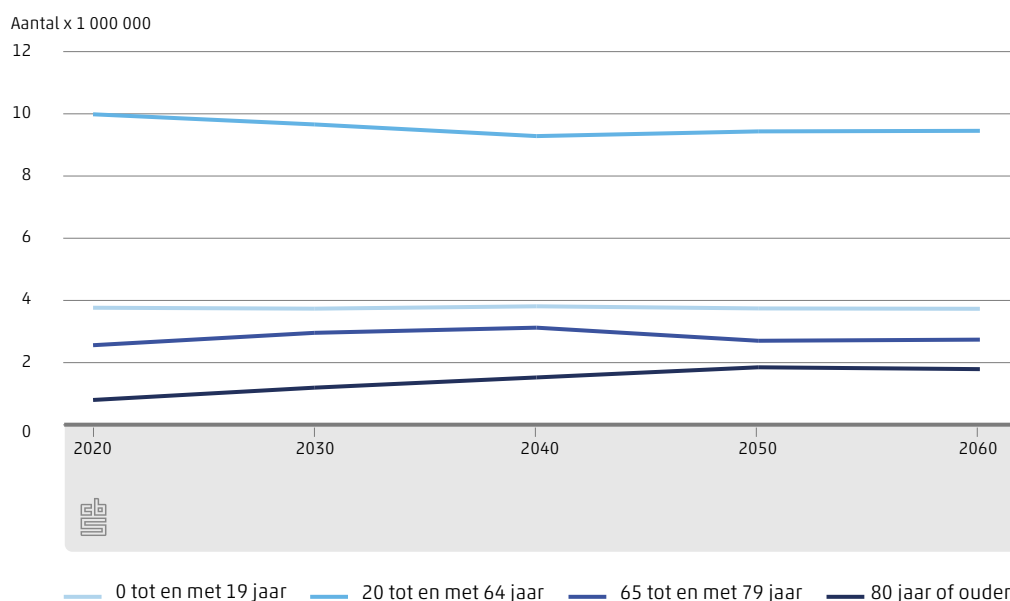
## 5.6 Een tijdreeks als stapeldiagram over de bevolkingsprognose naar leeftijd



Merk op dat in figuren 5.5 en 5.6 de groepen van de kwalitatieve variabele een natuurlijke ordening hebben. Daarom zijn verschillende tinten van één kleur gebruikt om de categorieën aan te duiden (en niet verschillende kleuren).

In figuur 5.7 is de ontwikkeling in de tijd van een kwalitatieve variabele weergegeven in de vorm van een lijndiagram. De ontwikkeling in de tijd is voor elke groep apart goed te volgen. Moeilijker is het om vast te stellen of de relatieve omvang van een leeftijdsgroep in de tijd verandert. Dat lukt in figuur 5.7 nog wel voor de groep 80+, maar voor de andere leeftijdsgroepen is het schatten van de ruimte tussen de twee begrenzendende lijnen lastig.

## 5.7 Een tijdreeks als lijndiagram over de bevolkingsprognose naar leeftijd





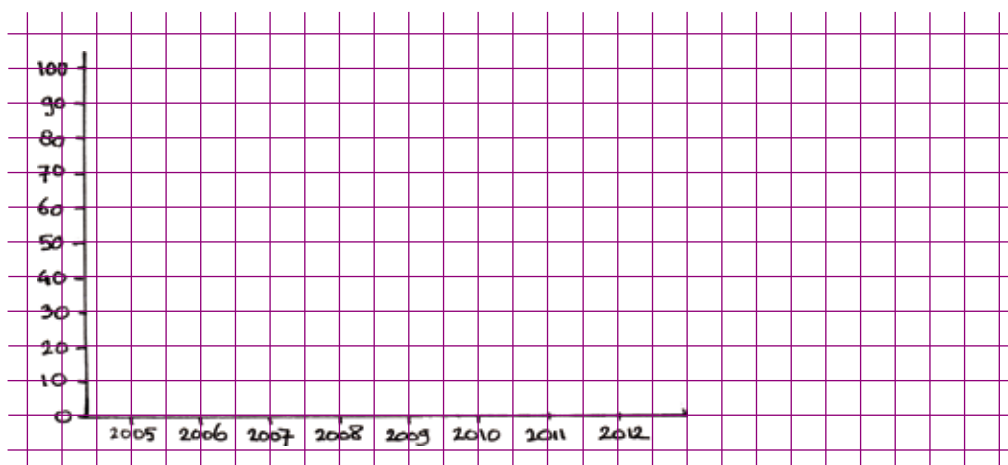
### Vraag 5.1

Nederland loopt voorop in de wereld als je kijkt naar het gebruik van het internet. We hebben het hoogste percentage huishoudens met toegang tot het internet. Alleen Noorse, Zweedse en Deense huishoudens komen bij ons in de buurt.

## 5.8 Percentage huishoudens in Nederland met toegang tot internet

Jaar	Percentage
2005	78
2006	80
2007	83
2008	86
2009	90
2010	91
2011	94
2012	94

In tabel 5.8 kun je zien hoe het percentage huishoudens met internet zich in de loop van de tijd heeft ontwikkeld. Maak een grafiek van deze tijdreeks op het ruitjespapier hieronder. Er is al een beginnetje gemaakt. Je mag zelf kiezen of je een lijndiagram wilt maken of een staafdiagram. Als de grafiek klaar is, probeer hem dan te interpreteren. Geef twee conclusies die je uit de grafiek kunt trekken.



.....

.....

.....

.....

6.

Foute

grafieken

Een grafiek is bedoeld om de informatie te laten zien die verborgen zit in een hoeveelheid gegevens. De grafiek kan daarvoor een krachtig middel zijn, maar dan moet hij wel goed zijn gemaakt. We hebben al eerder gezien dat het eenvoudig is om mensen op het verkeerde been te zetten. Daardoor lopen ze het risico een verkeerde conclusie te trekken uit de gegevens.

In de media kun je heel wat voorbeelden tegenkomen van foute grafieken. Meestal zijn die grafieken niet bewust verkeerd gemaakt. Vaak is het een vormgever die in zijn enthousiasme een leuk, vrolijk plaatje wil maken en die niet beseft dat hij daarmee de 'boodschap' in de informatie kan verminken.

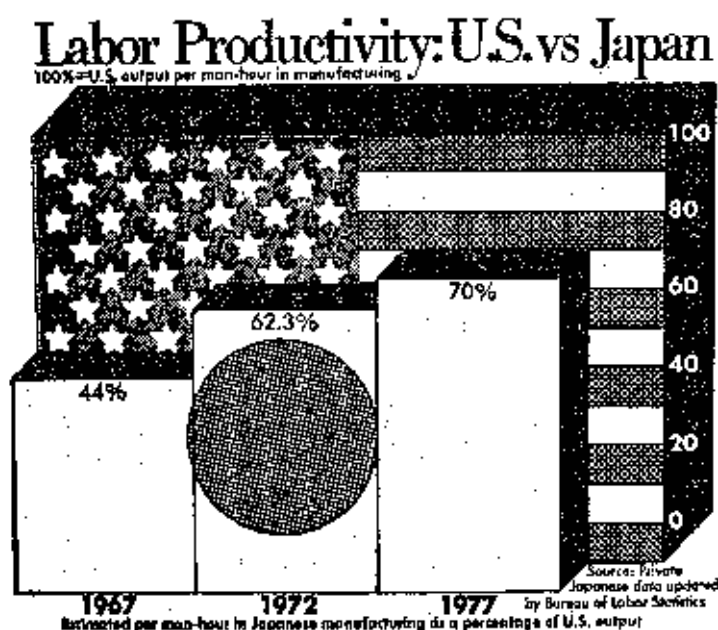
In dit hoofdstuk laten we een paar voorbeelden zien van foute grafieken. Waar mogelijk proberen we een betere versie van deze grafieken te maken.

## 6.1 Te veel chartjunk

De term chartjunk is in 1983 bedacht door de Amerikaanse statisticus Edward Tufte. Daarmee bedoelt hij alle zaken die zijn toegevoegd aan een grafiek, maar die niet helpen om de grafiek te begrijpen. Ze kunnen je zelfs afleiden. Chartjunk moet je zoveel mogelijk vermijden. Helaas zie je toch nog regelmatig grafieken met een hoop chartjunk.

Figuur 6.1 bevat een voorbeeld van een grafiek met veel chartjunk. Daarin wordt de arbeidsproductiviteit (hoeveelheid productie per gewerkt uur) in Japan vergeleken met die in Amerika. Een percentage kleiner dan 100% betekent dat de Amerikanen harder werken dan de Japanners.

### 6.1 Een grafiek met veel chartjunk

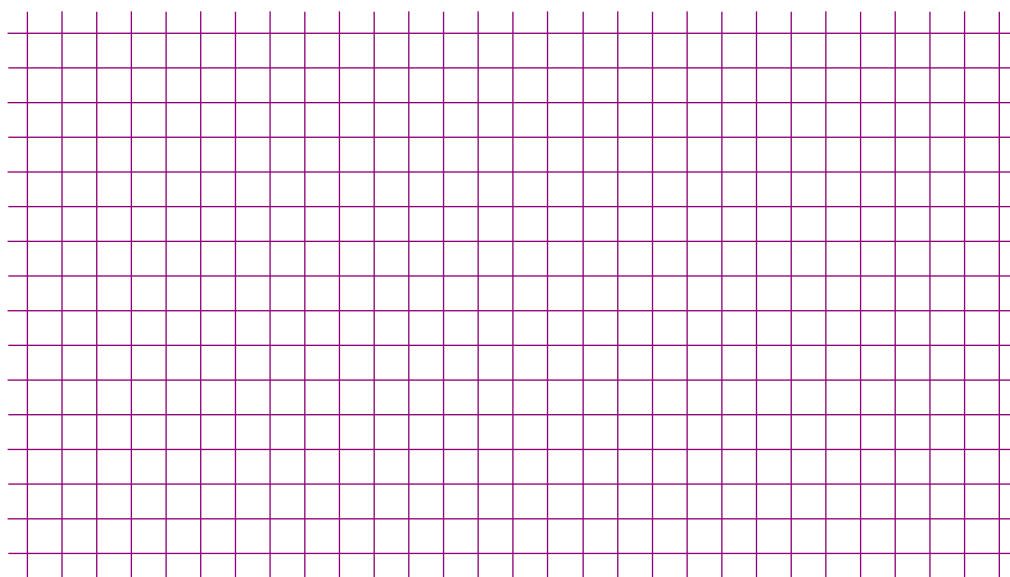


Door alle toeters en bellen is het niet meteen duidelijk wat er precies in de grafiek staat. En het driedimensionale karakter helpt ook al niet. Het blijkt om een tijdreeks te gaan met maar drie metingen: de arbeidsproductiviteit van Japan ten opzichte van die van Amerika in de jaren 1967, 1972 en 1977.

---

### Opgave 6.1

Maak een betere grafiek door alle chartjunk weg te halen en de regels toe te passen uit de vorige hoofdstukken. Kies voor een tijdreeks in de vorm van een staafdiagram.



---

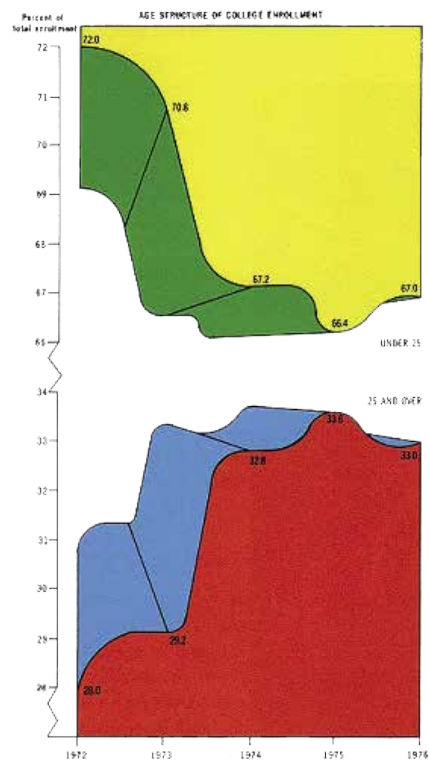
De boodschap in de grafiek die je hebt gemaakt in opgave 6.1 is veel duidelijker: Het verschil in arbeidsproductiviteit tussen Amerika en Japan neemt geleidelijk af.

Figuur 6.2 bevat een wel heel bizarre grafiek. Hij probeert in beeld te brengen hoe de leeftijdsopbouw van inschrijvers bij Amerikaanse colleges is veranderd in de loop der tijd. Het gaat om het aandeel van de twee leeftijdsgroepen: studenten jonger dan 25 jaar en studenten van 25 jaar of ouder.

Dit is een wel heel slechte grafiek. De verticale as is maar liefst op twee plaatsen onderbroken. Verder zijn de waarden voor opeenvolgende jaren met elkaar verbonden door mooie ronde lijnen, waardoor je ten onrechte de suggestie krijgt van een glad verloop in de tijd. Tenslotte is een driedimensionaal perspectief toegevoegd. Dit perspectief en de inkleuring bemoeilijkt de juiste interpretatie van de grafiek behoorlijk. De grafiek suggereert een sterke toename voor het rode vlak tussen 1973 en 1974, en tegelijk een grote afname voor het gele vlak ertussen. Maar is dat wel zo?

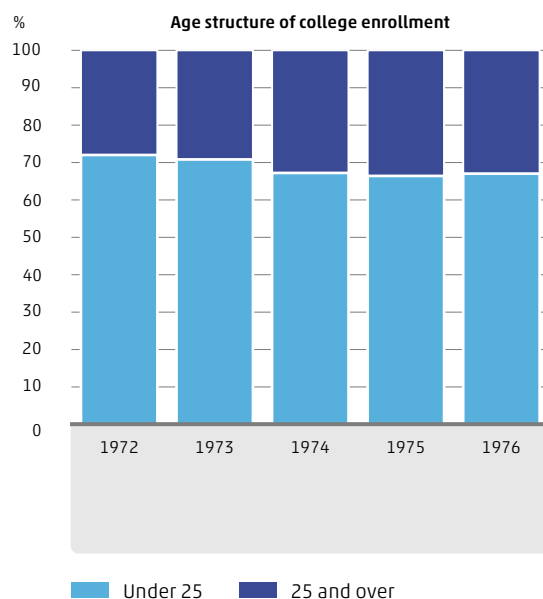
In feite bevat de grafiek in figuur 6.2 slechts tien waarden, namelijk de percentages inschrijvers onder de 25 jaar en daarboven in de vijf opeenvolgende jaren 1972 tot en met 1976.

## 6.2 Nog een grafiek met veel chartjunk



Een simpel en helder alternatief voor deze grafische mislukking staat in figuur 6.3. Het stapeldiagram bevat exact dezelfde informatie, maar dan veel duidelijker.

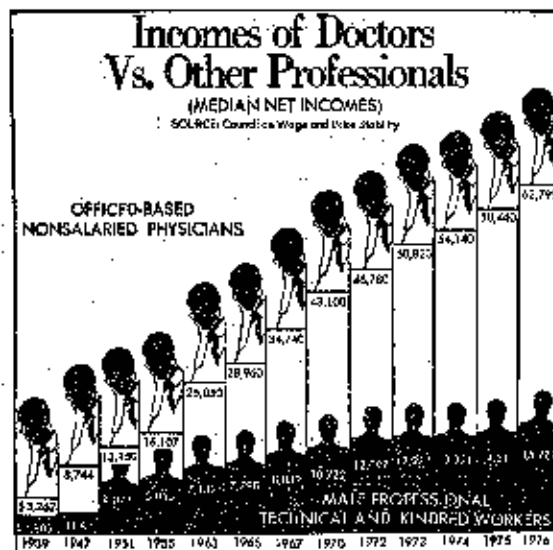
## 6.3 De verbeterde versie van de grafiek in figuur 6.2



## 6.2 Gerommel met de assen

Als je op een horizontale of verticale as de waarden van een kwantitatieve variabele uitzet, dan moet je dat zorgvuldig doen. Dat betekent bijvoorbeeld dat je het nulpunt moet vermelden als er een natuurlijk nulpunt is. Het is ook niet verstandig om de as door te knippen. Een andere veel gemaakte fout is dat de waarden op de as niet gelijkmatig oplopen: soms stijgen de waarden snel en soms langzaam. Kijk maar eens naar de grafiek in figuur 6.4.

### 6.4 Een grafiek met een ongelijkmatige horizontale as



De grafiek bevat tijdreeksen van de inkomens van artsen en van werkenden in technische en aanverwante beroepen. De grafiek suggereert dat de inkomens van artsen gelijkmatig stijgen door de jaren heen, met een licht afname van de stijging in de laatste paar jaar. Er lijkt dus geen sprake te zijn van spectaculaire inkomensstijgingen. Maar is dat wel zo? Als je de horizontale tijd-as wat nader bekijkt, zie je dat daar iets vreemds aan de hand is. Tussen de opeenvolgende staven van het staafdiagram zit niet steeds evenveel jaar. Zo zit er tussen de eerste twee staven 8 jaar en tussen de laatste twee staven maar één jaar. De tijd neemt dus niet gelijkmatig toe, maar snel in het begin en langzaam aan het eind. Anders gezegd: de as is aan het eind steeds meer uit elkaar getrokken.

### Opgave 6.2

Maak een verbeterde versie van de grafiek in figuur 6.4. Gebruik daarbij de gegevens in tabel 6.5. Om het werk wat te vereenvoudigen zijn de inkomens afgerond op veelvouden van 1 000.

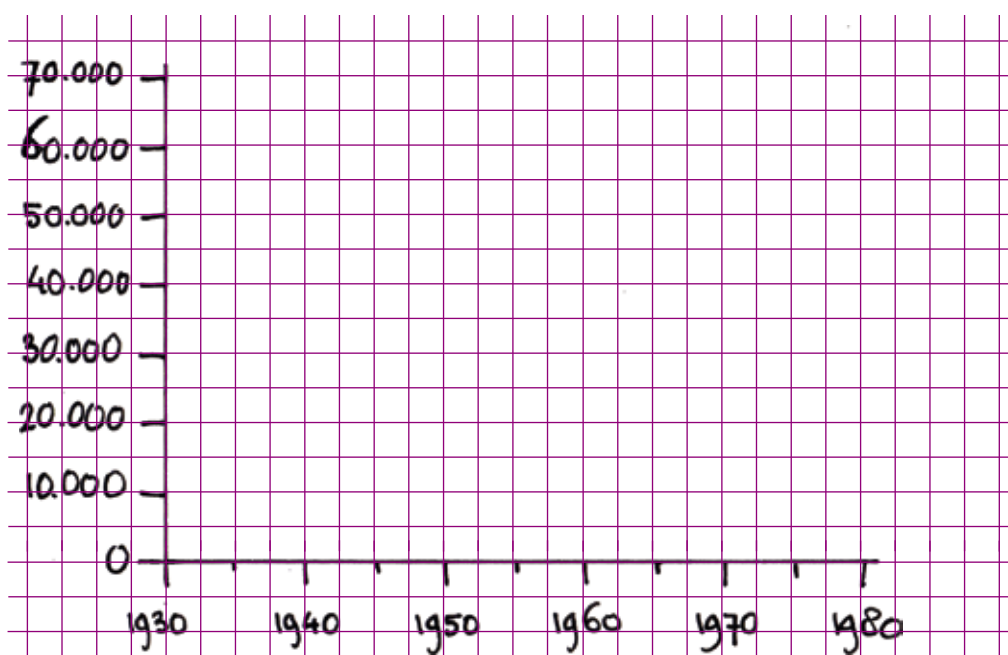
## 6.5 Inkomen van doktoren en andere professionals

Year	Income of doctors	Income of other professionals
1939	3 000	2 000
1947	9 000	3 000
1951	13 000	4 000
1955	16 000	5 000
1963	25 000	7 000
1965	29 000	8 000
1967	35 000	9 000
1970	43 000	11 000
1972	46 000	12 000
1973	50 000	13 000
1974	54 000	13 000
1975	58 000	14 000
1976	63 000	15 000

De nieuwe grafiek moet een horizontale tijd-as krijgen. Die moet beginnen bij het jaar 1930 en eindigen bij het jaar 1980. Om de 10 jaar moet een klein maatstreepje (tick mark) komen, met daarbij het jaartal.

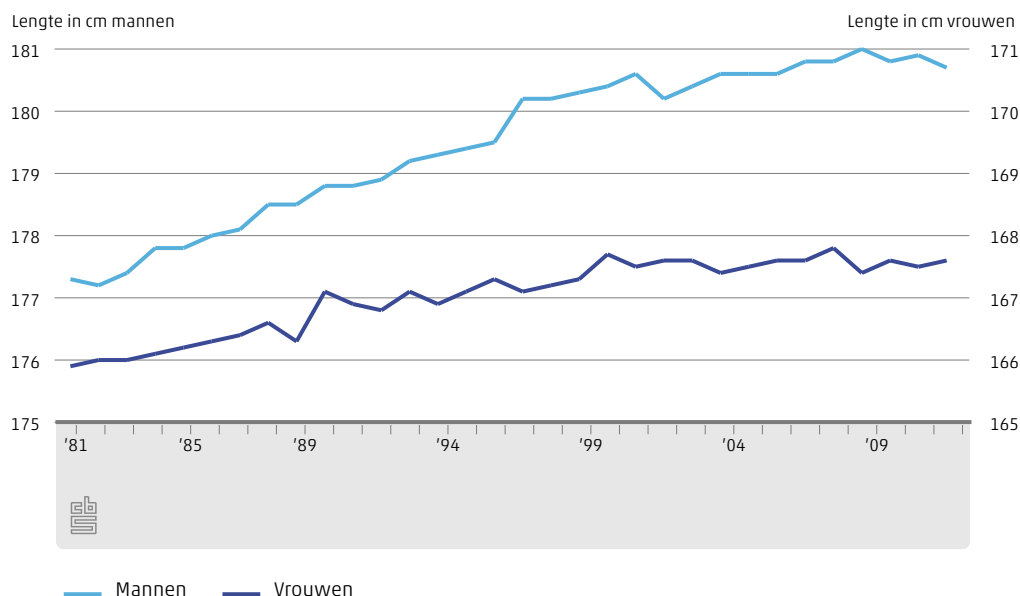
Op de verticale as moet je het inkomen afzetten. Het inkomen begint bij 0 en eindigt bij 70 000 (dollar). Zet om de 10 000 dollar een maatstreepje met daarbij het bedrag.

Vergeet niet alle andere benodigde bouwstenen in de grafiek op te nemen. Op het ruitjespapier hieronder is al een eerste begin gemaakt. Maak de grafiek verder af.



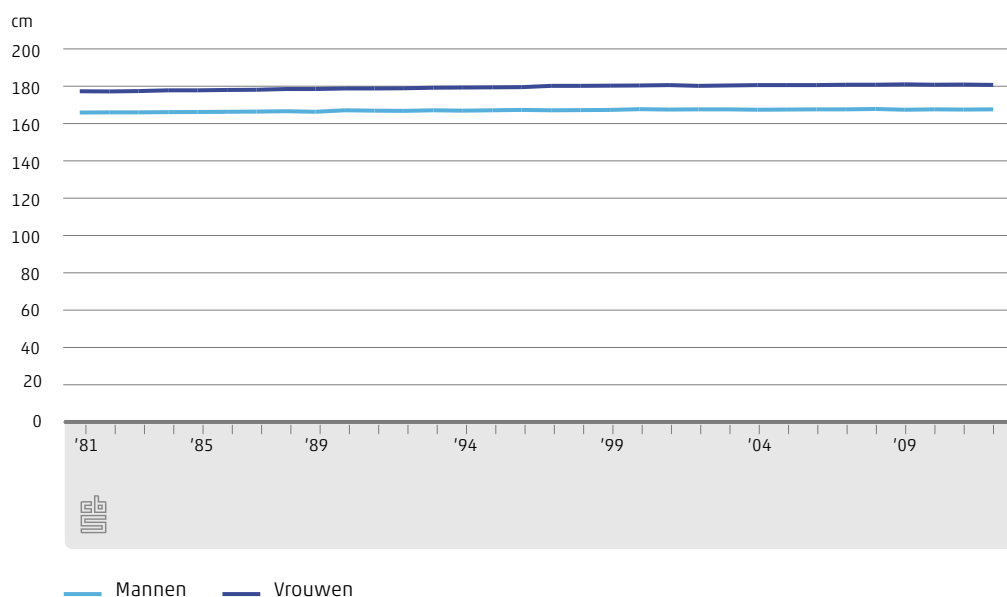
Figuur 6.6 bevat een grafiek met daarin een tijdreeks van de gemiddelde lengte van mannen en vrouwen in Nederland. Op het eerste gezicht ziet de grafiek er nogal spectaculair uit. De grafiek roept de indruk op van een sterke toename van de gemiddelde lengte. Dat geldt vooral voor mannen en iets minder voor vrouwen. Verder wekt de grafiek de suggestie dat vrouwen gemiddeld maar half zo lang zijn als mannen.

## 6.6 Een grafiek zonder nulpunt en met twee verticale assen over de gemiddelde lengte van volwassenen



Er zijn twee problemen met de grafiek in figuur 6.6. In de eerste plaats begint de verticale as niet bij 0. In de tweede plaats heeft de grafiek twee verschillende verticale assen: een as voor de lengte van mannen (aan de linkerkant) en een as voor de lengte van de vrouwen (aan de rechterkant). Beide assen komen niet met elkaar overeen.

## 6.7 De verbeterde versie van de grafiek in figuur 6.6 over de gemiddelde lengte van volwassenen





In de grafiek in figuur 6.7 zijn beide defecten gerepareerd. In die grafiek is te zien dat de lengte van mannen en vrouwen nog steeds wel toeneemt, maar veel langzamer. Er zijn ook nog steeds verschillen tussen mannen en vrouwen, maar die zijn veel minder groot. Verder lijkt het verschil tussen mannen en vrouwen langzaam toe te nemen.

## 6.3 Pictogrammen in plaats van staven

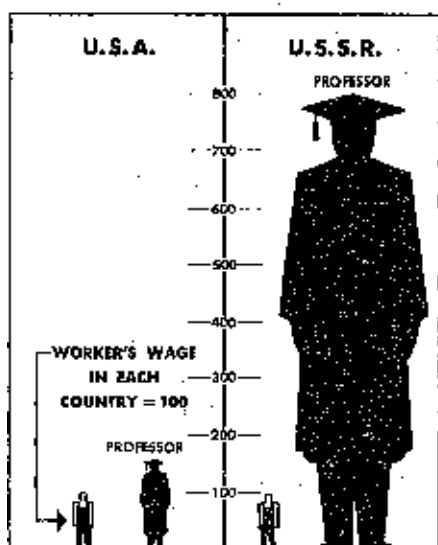
Een staafdiagram mag dan een duidelijk beeld geven van een kwalitatieve variabele, het is ook wel een beetje een saaie grafiek. Daarom proberen vormgevers soms de grafiek wat leuker en aantrekkelijker te maken. Dat doen ze door de staven te vervangen door pictogrammen.

Figuur 6.8 geeft een voorbeeld waarin de staven zijn vervangen door poppetjes. De grafiek wil laten zien hoe in Amerika en de vroegere Sovjet Unie de inkomens van professoren en arbeiders van elkaar verschilden. De staven voor de professoren zijn vervangen door plaatjes van professoren en die van arbeiders door plaatjes van arbeiders. De hoogte van de poppetjes geeft aan hoeveel de mensen in die groep gemiddeld verdienen.

Uit de grafiek kun je aflezen dat een Sovjet professor acht keer zoveel verdient als een Sovjet fabrieksarbeider. In de grafiek wordt echter over het hoofd gezien dat als een plaatje acht keer zo hoog wordt, hij ook acht keer zo breed wordt. In feite is het poppetje van de Sovjet professor dus  $8 \times 8 = 64$  keer zo groot als het poppetje van de Sovjet fabrieksarbeider. Als je alleen naar de omvang van de poppetjes kijkt, dan trek je dus de conclusie dat een professor in de Sovjet Unie 64 keer zoveel verdient als een fabrieksarbeider in de Sovjet Unie.

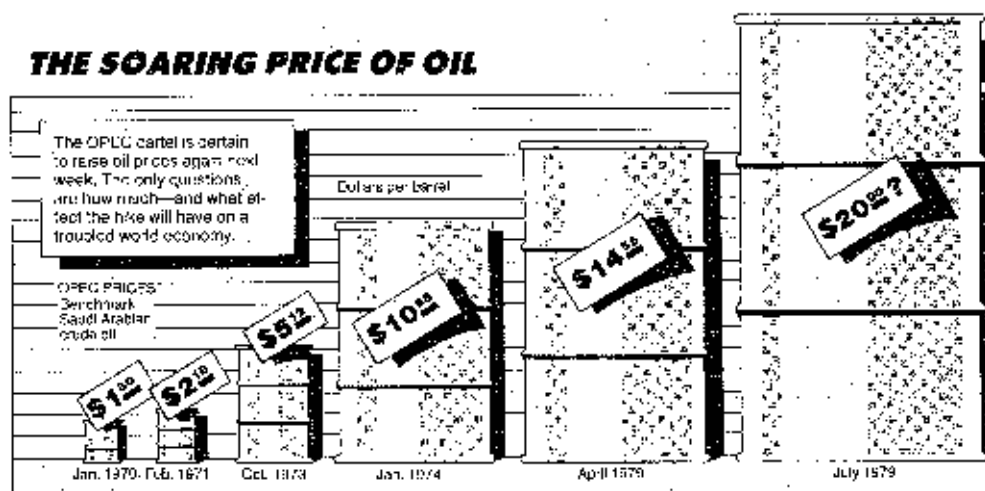
De grafiek in figuur 6.8 heeft nog een ander gebrek. Doordat de figuren voor de fabrieksarbeiders in de VS en de USSR even groot zijn gemaakt, wordt ten onrechte de indruk gewekt dat de Russische hoogleraar veel meer verdient dan de Amerikaanse hoogleraar.

### 6.8 Een staafdiagram met pictogrammen



Figuur 6.9 geeft nog een ander voorbeeld van het gebruik van pictogrammen in plaats van staven. Die grafiek probeert de stijging van de olieprijs in de jaren 1970 tot en met 1979 in beeld te brengen. De staven in het staafdiagram hebben de vorm van olievaten gekregen. De hoogte van de vaten weerspiegelt de olieprijs op dat moment. De ontwerper heeft echter over het hoofd gezien dat de argeloze gebruiker van de grafiek zal kijken naar de oppervlakte van de grafiek. En de verhoudingen tussen de oppervlaktes komen niet overeen met de verhoudingen tussen de olieprijsen. De oppervlaktes nemen veel sneller toe. Daardoor krijg je onbewust de indruk van een veel snellere stijging van de olieprijsen dan in werkelijkheid het geval was.

## 6.9 Nog een staafdiagram met pictogrammen



Er valt nog wel wat meer aan te merken op de grafiek in figuur 6.9. Zo is er duidelijk geen sprake van gelijkmatig oplopende waarden op de horizontale as. Tussen twee opeenvolgende olievaten verstrijkt niet evenveel tijd. Dat kan ook een verkeerd beeld geven van de ontwikkeling in de tijd.

## Opgave 6.3

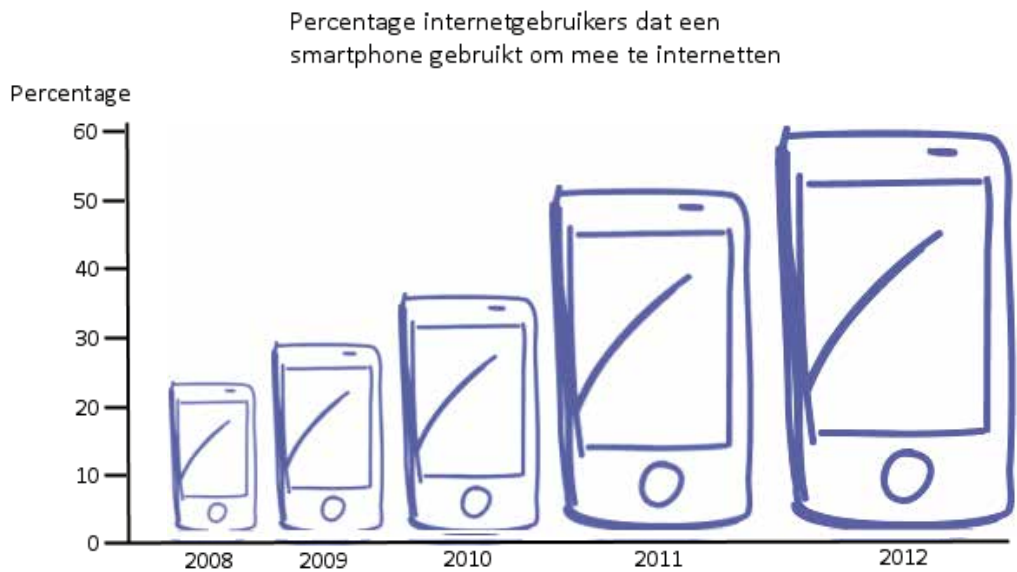
Je kunt op allerlei manieren toegang krijgen tot het internet. Dat kan met allerlei verschillende apparaten, zoals desktop-computers, laptops, tablets en smartphones. Het CBS houdt de ontwikkelingen in deze statistieken bij. In tabel 6.10 kun je zien hoeveel procent van de internetgebruikers een smartphone gebruikt voor internet.

## 6.10 Het aandeel internetgebruikers dat een smartphone gebruikt om mee te internetten

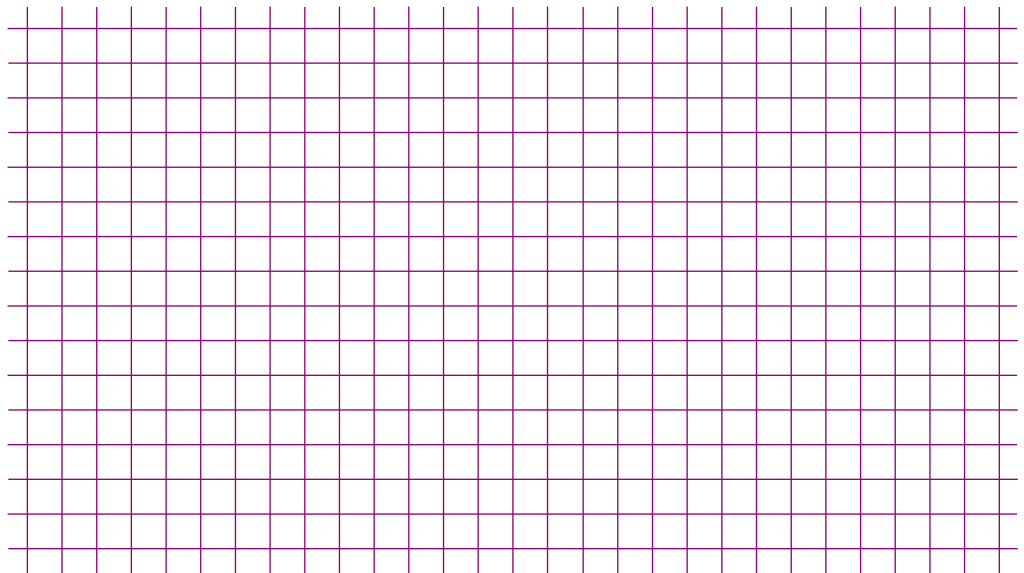
Jaar	Percentage
2008	22
2009	28
2010	35
2011	50
2012	59

Figuur 6.11 bevat de tijdreeks in de vorm van een foute grafiek. De staven van het staafdiagram zijn vervangen door een symbool (van een smartphone).

### 6.11 Een foute grafiek



Leg uit wat er mis is met de grafiek. Maak vervolgens een goede grafiek, door een tijdreeks op basis van staafdiagrammen te tekenen.



# Gebruikte bronnen

CBS (2013), *ICT, Kennis en Economie 2013*. Centraal Bureau voor de Statistiek, Den Haag.

Schmid, C.F. (1983), *Statistical Graphics, Design Principles and Practices*. John Wiley & Sons, New York.

Tufte, E. (1983), *The Visual Display of Quantitative Information*. Graphics Press, Cheshire, Connecticut.

Wainer, H. (1997), *Visual revelations: graphical tales of Fate and Deception from Napoleon Bonaparte to Ross Perot*. Copernicus / Springer Verlag, New York.

Wallgren, A., Wallgrens, B., Persson, R., Jorner, U. and Haaland, J-A. (1996), *Graphing Statistics and Data, Creating Better Charts*. Sage Publications, London.

# Gebruikte illustraties

Figuur 6.1, Labor productivity: Washington Post, 1978.

Figuur 6.3, College enrolment: Calvin Schmid, Statistical Graphics, Design Principles and Practices, 1983.

Figuur 6.5, Incomes of doctors: Washington Post, 1979.

Figuur 6.7: Gemiddelde lengte van volwassenen: CBS, Webmagazine, 2008.

Figuur 6.9: Incomes of professors and worker, Calvin Schmid, Statistical Graphics, Design Principles and Practices, 1983.

Figuur 6.10: The soaring price of oil, College enrolment: Calvin Schmid, Statistical Graphics, Design Principles and Practices, 1983.

# Medewerkers

Jelke Bethlehem (methodoloog CBS)