

1. Kijken naar data

Boekje 1 havo wiskunde A, domein E: Statistiek



Verantwoording



© 2015, SLO (nationaal expertisecentrum leerplanontwikkeling), Enschede

Dit lesmateriaal is ontwikkeld in het kader van de nieuwe examenprogramma's zoals voorgesteld door de commissie Toekomst Wiskunde Onderwijs (cTWO) en herzien door SLO.

Mits de bron wordt vermeld, is het toegestaan zonder voorafgaande toestemming van de uitgever deze uitgave geheel of gedeeltelijk te kopiëren en/of verspreiden en om afgeleid materiaal te maken dat op deze uitgave is gebaseerd.

Auteurs: Erik van Barneveld, Wouter Boer, Carel van de Giessen, Peter Kop, Heleen van der Ree, Henk Reuling, Frits Spijkers, Tanja Stroosma, Anneke Verschut

Met medewerking van: Nico Alink, Martine de Klein (eindredactie)

Informatie: SLO
Afdeling: tweede fase
Postbus 2041, 7500 CA Enschede
Telefoon (053) 4840 661
Internet: www.slo.nl



Overzicht lesmateriaal in het domein Statistiek

1. Kijken naar data

- § 1.1 Wat is statistiek?
- § 1.2 Data
- § 1.3 Diagrammen
- § 1.4 Interpretaties
- § 1.5 Overzicht

2. Data en datasets verwerken

- § 2.0 Begrippenlijst
- § 2.1 Data presenteren
- § 2.2 Verbanden tussen datarepresentaties
- § 2.3 Frequentieverdelingen typeren
- § 2.4 Twee groepen vergelijken
- § 2.5 Samenhang tussen twee variabelen

3. Data verwerven

- § 3.0 Pas op voor valkuilen
- § 3.1 Onderzoeks- en enquêtevragen
- § 3.2 Steekproeven en fouten
- § 3.3 Standaardafwijking
- § 3.4 Steekproeffout: variatie bij steekproeven
- § 3.5 Normale verdeling
- § 3.6 Toevallige steekproeffouten in getallen
- § 3.7 Terugblik op boekje 3

4. Statistische uitspraken doen

- § 4.1 Voorkennis
- § 4.2 Doel van deze module
- § 4.3 Populatieproportie
- § 4.4 Populatiegemiddelde
- § 4.5 Verschil tussen twee groepen
- § 4.6 Samenhang tussen twee kwantitatieve variabelen
- § 4.7 Gemengde opgaven
- § 4.8 Terugblik
- § 4.9 Lessenserie: Statistiek op een groot gegevensbestand
- § 4.10 Diagnostische computertoets



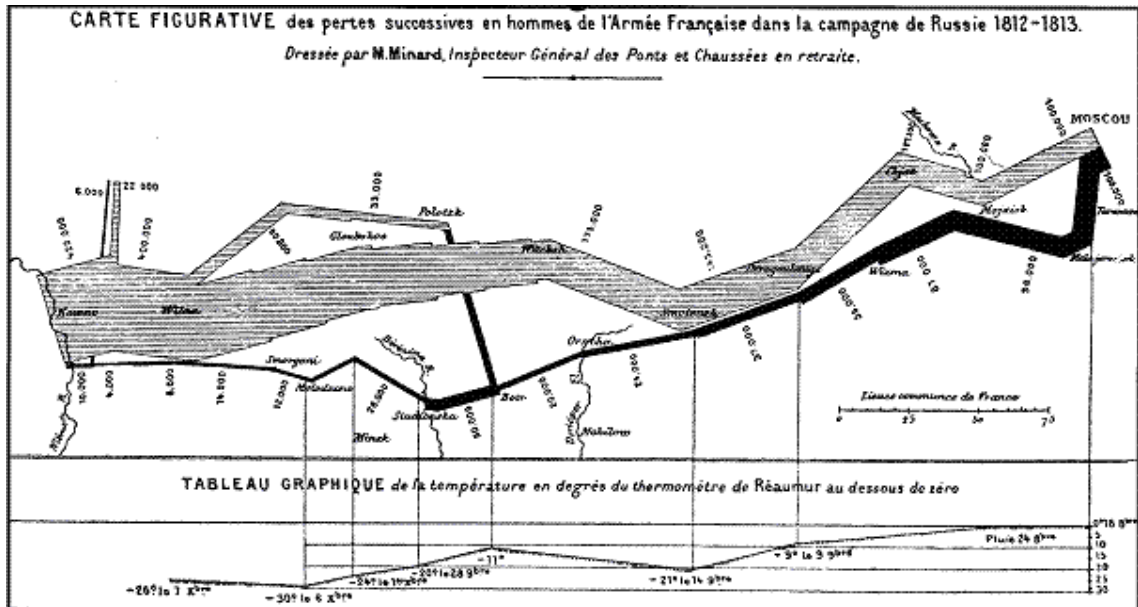
Inhoud

Overzicht lesmateriaal in het domein Statistiek.....	3
§ 1.1 Wat is statistiek?	5
§ 1.2 Data.....	14
§ 1.3 Diagrammen.....	20
§ 1.4 Interpretaties	35
§ 1.5 Overzicht	41



§ 1.1 Wat is statistiek?

Verkennen



In 1812 trok het Franse leger onder aanvoering van keizer Napoleon ten strijde tegen Rusland. Deze campagne begon aan de rivier de Niemen en leidde naar Moskou. Het leger stuitte op flinke weerstand en werd bij Moskou verslagen. De terugtocht was vreselijk, mede door de extreme koude.

Dit is de figuur die de Franse ingenieur Charles Joseph Minard (1781-1870) tekende om het verloop van deze Russische veldtocht weer te geven.

▣ Opgave 1

Je ziet dat het Franse leger met 422.000 man uit Polen vertrok richting Moskou. Het aantal mensen waaruit het leger bestond is een statistische variabele.

- Hoeveel mensen in dit leger kwamen aan in Moskou?
- Hoeveel mensen kwamen weer bij het startpunt terug?
- Hoeveel procent van dit Franse leger is tijdens de Russische veldtocht overleden?
- De slag aan de Berezina vond plaats van 26 tot en met 29 november 1812.
Hoe komt het dat Napoleons leger vlak daarvoor in omvang meer dan verdubbelde?
Welk deel van dit nieuwe leger bleef er na deze slag nog over?
- Op de terugweg zijn ook af en toe de dagtemperaturen gemeten.
Welke laagste dagtemperatuur is er gemeten? En op welke datum?
- De temperatuur is gegeven in graden Réaumur (°Re).
Nu is 0 graden Celsius (°C) hetzelfde als 0 °Re en 100 °C hetzelfde als 80 °Re.
Hoeveel graden was de laagste dagtemperatuur in graden Celsius?

▣ Opgave 2

Bekijk de figuur van Minard nog eens goed. Je kunt er in aflezen uit hoeveel personen Napoleons leger op bepaalde momenten bestond, hoe de gevolgde route er uitzag en met welke temperaturen het leger op de terugweg te maken kreeg. Bekijk nu vooral die terugweg.

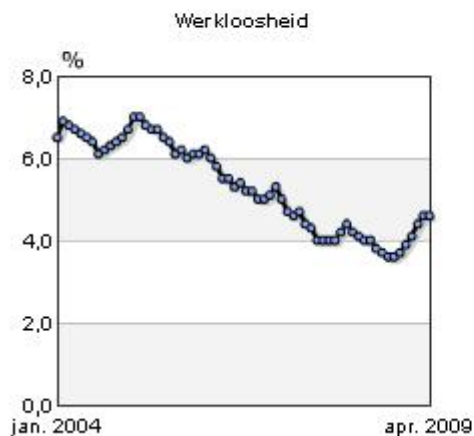
- Op welke twee manieren geeft dit diagram de grootte van het leger weer?
- Welke andere twee variabelen worden er in getallen uitgedrukt?
- Welke variabele wordt er niet in een getal uitgedrukt? Hoe wordt die variabele zichtbaar gemaakt?
- Maak voor de terugtocht een tabel waarin je bij elke datum die je in de figuur kunt aflezen het aantal sterfgevallen onder de soldaten uitzet tegen de temperatuur. Hoe zou je deze gegevens op een andere manier dan in Minards diagram grafisch kunnen voorstellen?
- Welk doel heeft Minard waarschijnlijk gehad met zijn beelddiagram? Licht je antwoord toe.

▣ Opgave 3

Op de website van het Centraal Bureau voor de Statistiek (CBS) vind je een schat aan informatie. Bekijk de startpagina van de website van het CBS. Op de startpagina stond op een zeker moment deze informatie:

Kerncijfers

Prijzen Arbeid Groei Vertrouwen



Bevolkingsteller

Nederland telt

16.521.632

inwoners op

maandag 6 juli 2009

10:13:38 GMT +02:00

- Op welke dag en op welk tijdstip is deze schermafdruck gemaakt?
- Hoeveel mensen telde Nederland op dat moment? En hoeveel nu?
- Het CBS spreekt in haar toelichting over een schatting van het aantal inwoners. Licht dit toe.
- Wat lees je af onder *Kerncijfers > Arbeid*?
- Het percentage werklozen is begin 2009 aan het stijgen. Kun je op grond van alleen deze gegevens zeggen dat ook het aantal werklozen in die periode stijgt? Waarom is dit wel erg waarschijnlijk?
- Beschrijf wat je afleest onder *Prijzen*, onder *Groei* en onder *Vertrouwen*.
- Bij *Groei* wordt een staafdiagram gebruikt en bij de andere variabelen niet. Waarom zou men dat hebben gedaan?
- Waarom heeft het CBS juist deze kerncijfers op de startpagina staan?

Uitleg

Het woord **statistiek** ontstond in de tijd van Napoleon. Toen werd het bestuur van de verschillende Europese staten die in de achttiende eeuw opkwamen zo ingewikkeld dat een nauwkeurige kennis van staatszaken belangrijk werd.

Tegenwoordig moet je dat wat ruimer opvatten: ook bedrijven, onderzoeksinstituten en particulieren buiten de overheid verzamelen gegevens en doen dus aan statistiek. Die gegevens worden in de vorm van tabellen, diagrammen, kaarten of een combinatie daarvan overzichtelijk weergegeven. Op de website van het CBS vind je daar mooie voorbeelden van.

In de figuur van Minard gaat het om de soldaten van het Franse leger dat onder Napoleon naar Moskou trok. Deze soldaten vormen de **elementen** van een statistisch onderzoek naar hun wel en wee. Napoleons leger is de onderzochte **populatie**. Een van de onderwerpen van het onderzoek is het verband tussen het aantal soldaten en de vordering van de tocht. Dit aantal soldaten is een van de **statistische variabelen** die Minard onderzocht. De gegevens omtrent het aantal soldaten op een bepaald punt tijdens de tocht vormen de door Minard verzamelde **data**. Data zijn dus statistische gegevens.

In Minards diagram zijn ook data over andere variabelen te vinden.

Bij de weergave van de data heeft de onderzoeker meestal wel een bepaald doel. Minard wil kennelijk de verschrikkingen van deze veldtocht van Napoleon in beeld brengen. Hij heeft daarom de vorm van zijn diagram zo gekozen dat duidelijk zichtbaar wordt hoeveel kleiner het leger telkens werd tijdens de veldtocht.

📄 Opgave 4

Bekijk nog eens de gegevens uit Minards diagram over de veldtocht van 1812.

- Welke variabelen zijn in het diagram terug te vinden?
- Het grootste deel van Minards diagram heeft de vorm van een kaart.
Welk voordeel heeft die keuze? En welk nadeel?
- Welke variabele wordt vooral als nadere toelichting op de grote sterfte onderzocht?
- Is het diagram dat bij deze variabele past een echt lijndiagram?

📄 Opgave 5

Open het bij deze opgave horende bestand met data van de Nederlandse bevolking zoals die in juli 2009 te vinden waren op de website van het CBS.

- Waarom staat achter het jaartal 2009 een sterretje? (Zoek eventueel op het tabblad *Beschrijving*.)
- Over welke populatie gaat deze tabel?
- Wat zijn de elementen van de populatie als het gaat over *Bevolking*?
En wat zijn de elementen als het gaat over *Huishoudens*?
- Over welke variabelen gaat het bij *Bevolking naar geslacht*?
- Over welke variabelen gaat het bij *Particuliere huishoudens*?
- Wat stelt de *Bevolkingsdichtheid* voor? Wat betekent de toename van de bevolkingsdichtheid?
- Hoe komt het dat het totale aantal Nederlanders gedeeld door de bevolkingsdichtheid niet voor elk jaar hetzelfde getal is? (Geef minstens twee verklaringen.)

Theorie

Statistiek is de wetenschap, de methodiek en de techniek van het verzamelen, bewerken, interpreteren en presenteren van gegevens. Een statisticus is iemand die statistisch onderzoek verricht.

- Je begint een statistisch onderzoek met een **vraag** die alleen met statistische gegevens kan worden beantwoord.
- Dan stel je vast over welke **populatie** (doelgroep) het onderzoek gaat. De leden van de populatie zijn de **elementen** van het onderzoek.
- Vervolgens stel je nauwkeurig vast op welke **statistische variabele** het onderzoek betrekking heeft.
- Daarna verzamel je de bij die variabele passende **data** (de statistische gegevens).
- Vervolgens **orden** je de verzamelde gegevens om meer overzicht te krijgen.
- Tenslotte probeer je een **conclusie** te trekken.

De verzamelde data geef je weer met behulp van tabellen en diagrammen. Welke vorm je kiest hangt af van wat je wilt zeggen: tabellen zijn nauwkeuriger, diagrammen vaak overzichtelijker.

Opgave 6

Voorbeeld

Stel je voor dat je de vraag voorgelegd krijgt hoe het aantal Nederlanders in de loop van de jaren verandert. Dat is een statistische vraag waarvoor het CBS tabellen bijhoudt

Ze ziet hier een klein deel van de tabel **BEVOLKING; KERNCIJFERS NAAR DIVERSE KENMERKEN**.

Bevolking; kerncijfers naar diverse kenmerken			CBS juli 2009						
Onderwerpen	Onderwerpen	Perioden	1950	1960	1970	1980	1990	2000	2008
Bevolking naar geslacht	Totale bevolking	aantal	10026773	11417254	12957621	14091014	14892574	15863950	16405399
	Mannen	aantal	4998251	5686152	6465081	6994280	7358482	7846317	8112073
	Vrouwen	aantal	5028522	5731102	6492540	7096734	7534092	8017633	8293326
Bevolkingsgroei	Totale bevolkingsgroei	aantal	173507	138754	161809	117572	117871	123125	81188
	Totale bevolkingsgroei, relatief	per 1 000 inw	17,3	12,2	12,5	8,3	7,9	7,8	4,9
	Levendgeborenen	aantal	229718	239128	238912	181294	197965	206619	184669
	Overledenen	aantal	75929	87825	109619	114279	128824	140527	134996
	Geboorteoverschot	aantal	153789	151303	129293	67015	69141	66092	49673
	Geboorteoverschot, relatief	per 1 000 inw	15,2	13,2	9,9	4,7	4,6	4,2	3
Bevolkingsdichtheid	Bevolkingsdichtheid	aantal	309	352	384	415	439	468	486

Beschrijf hoe dit deel van de tabel is opgebouwd: welke variabelen worden er gebruikt en hoe hangen die variabelen samen?

Uitwerking

Dit deel van de tabel gaat over de Nederlandse bevolking (de populatie). Er komen nogal wat variabelen voor in de tabel: *totale bevolking*, *aantal mannen*, *aantal vrouwen*, *totale bevolkingsgroei*, *aantal levendgeborenen*, *aantal overledenen*, *geboorteoverschot* en *bevolkingsdichtheid*. Deze variabelen variëren met de *tijd*. Het CBS heeft de bijbehorende data verzameld.

De variabelen *aantal mannen*, *aantal vrouwen* en *totale bevolking* hangen als volgt samen:

$$\text{aantal mannen} + \text{aantal vrouwen} = \text{totale bevolking}$$

En zo kun je ook een samenhang beschrijven tussen de *totale bevolkingsgroei, relatief*, de *totale bevolkingsgroei* en de *totale bevolking*.

Bekijk het voorbeeld nog eens.

- Hoe komt het CBS aan dergelijke gegevens?
- Licht toe hoe je de waarden van *totale bevolkingsgroei, relatief* kunt berekenen uit de *totale bevolking* en de *totale bevolkingsgroei*.
- Laat met een berekening zien dat de waarde 4,9 voor 2008 klopt.
- Welke uitspraak zou je op grond van deze tabel kunnen doen over de variabele *totale bevolkingsgroei, relatief*?

Opgave 7

In het voorbeeld komt ook de variabele *bevolkingsdichtheid* voor.

- Hoe berekent het CBS de bevolkingsdichtheid?
- Klopt het dat tussen 1950 en 2000 de oppervlakte van Nederland groter is geworden? Reken na hoeveel vierkante kilometer ongeveer.

Opgave 8

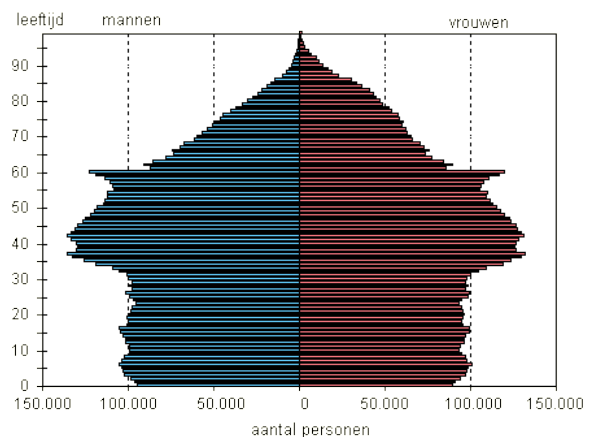
Bekijk nog eens de variabele *bevolkingsgroei*.

- Waaraan zie je dat deze variabele niet alleen kan worden bepaald uit *totale bevolking* en *geboorteoverschot*?
- Waar moet je om de bevolkingsgroei te berekenen ook nog rekening mee houden?
- In welke periode was het geboorteoverschot groter dan de bevolkingsgroei? Wat betekent dit?

Opgave 9

Open het bestand met data van de Nederlandse bevolking. Bekijk ook dit leeftijdsdiagram uit 2008 voor de Nederlandse bevolking.

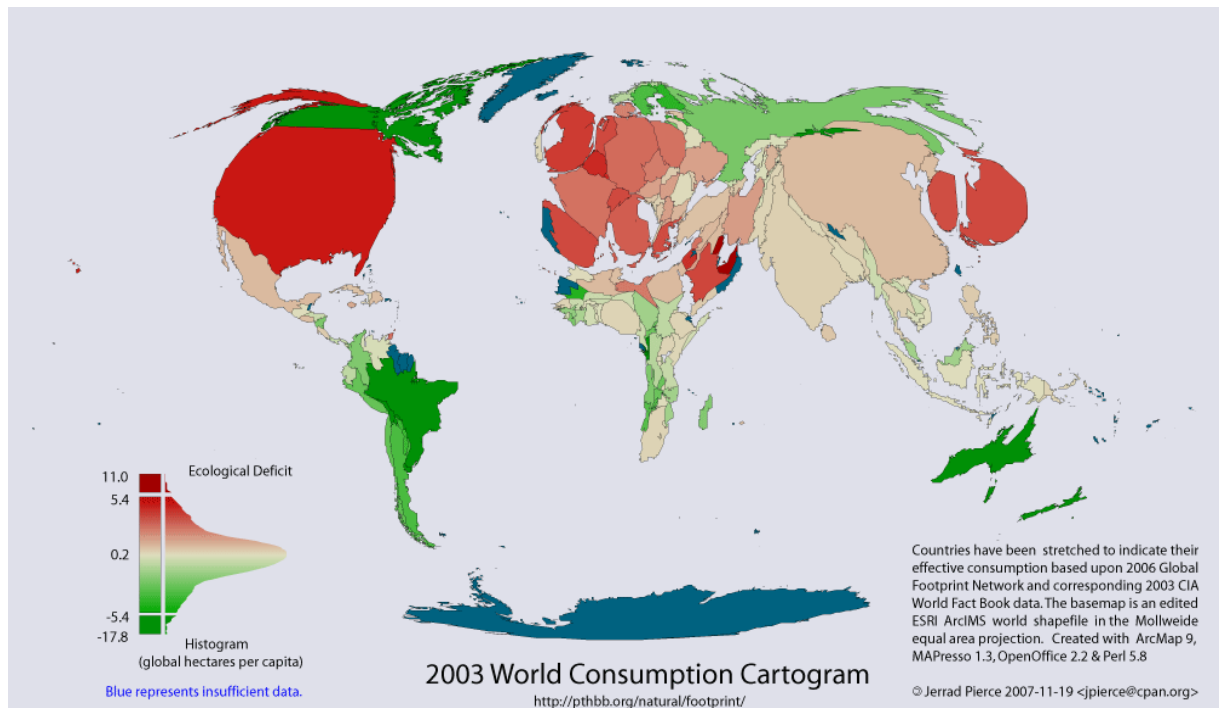
- Wat wordt verstaan onder *demografische druk*?
- Laat zien hoe de demografische druk wordt berekend uit de rest van deze tabel.
- De *grijze druk* neemt toe, de *groene druk* neemt af. Wat betekent dit?
- Hoe kun je dit verschijnsel uit het leeftijdsdiagram afleiden?



Verwerken

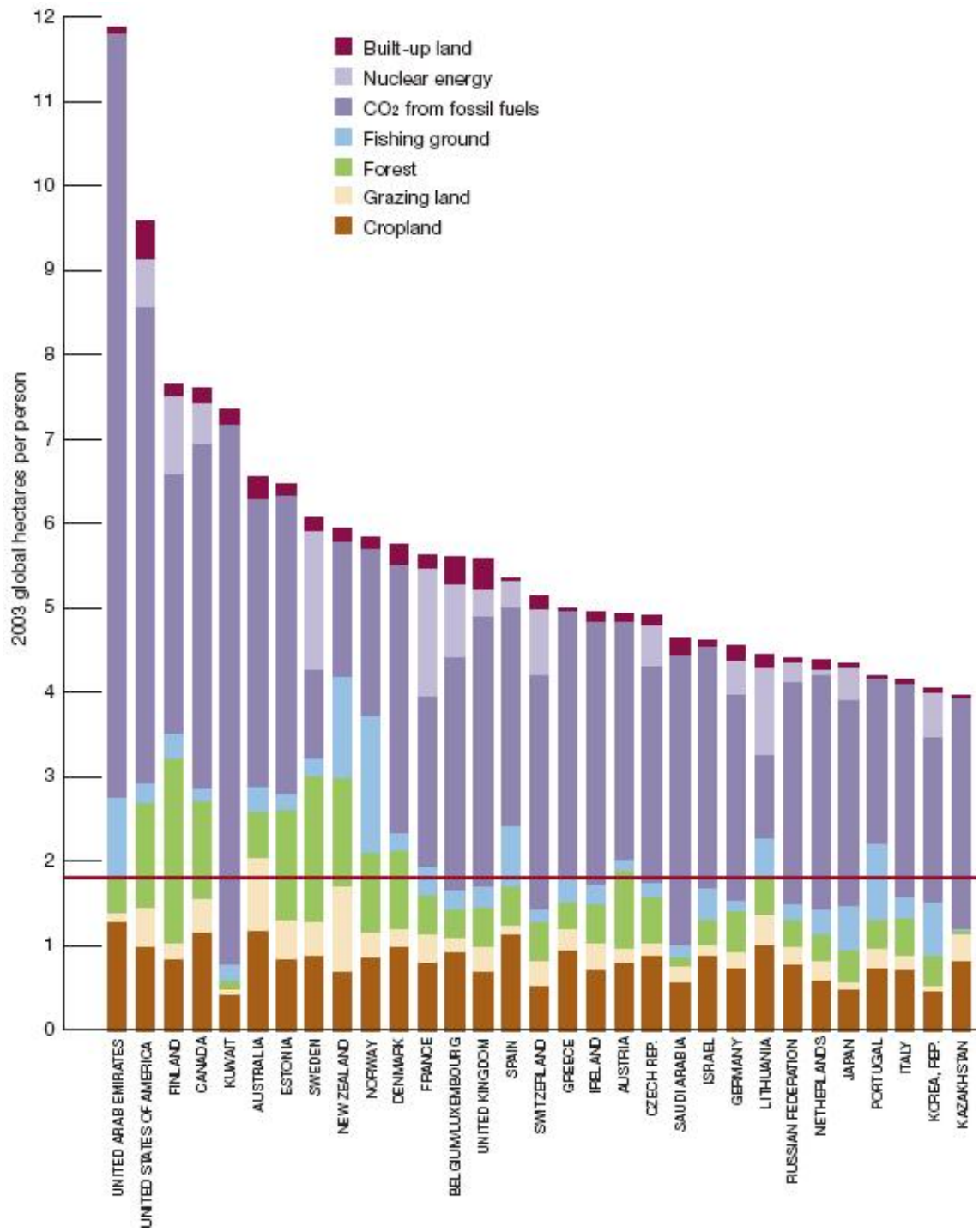
▣ Opgave 10

De ecologische voetafdruk (ook 'mondiale voetafdruk' of kortweg 'voetafdruk') is een getal dat weergeeft welk deel van het aardoppervlak een bepaalde bevolkingsgroep in een jaar gebruikt om van te leven. Dit getal wordt uitgedrukt in zogenaamde mondiale hectares.



Deze vervormde wereldkaart geeft de ecologische voetafdruk weer van de gemiddelde inwoner van een bepaald land.

- Wat betekent het dat Nederland aan de donkere kant van het rood zit?
- Waarom zijn bijvoorbeeld Groenland en Portugal blauw gekleurd?
- De Verenigde Staten is nogal opgeblazen getekend en Rusland is nogal ingekrompen. Waarom?
- Wat valt je op aan Italië?
- Milieuorganisatie De Kleine Aarde heeft berekend dat een gemiddelde ecologische voetafdruk van 1,8 voor de hele wereldbevolking haalbaar is. Maak op grond daarvan eens een schatting van de totale biologisch productieve oppervlakte op aarde. (Zoek eerst op hoeveel mensen de aarde op dit moment telt.)
- Het diagram op de volgende pagina laat meer details zien. Welke ecologische voetafdruk heeft de gemiddelde Nederlander?
- Litouwen heeft een vergelijkbare ecologische voetafdruk als Nederland, terwijl de welvaart er lager is. Geef daarvoor een verklaring.
- De Verenigde Arabische Emiraten hebben een voetafdruk van ongeveer 11,9. De gemiddelde voetafdruk per inwoner van dit land en Nederland samen is niet ongeveer $(11,9 + 4,4) / 2$. Waarom niet? En hoe moet je die dan wel berekenen?



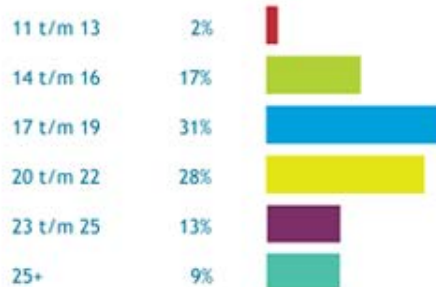
Opgave 11

Dit overzicht van de bijbanen van 44.000 actieve deelnemers van de site WWW.BIJBANEN.NL staat in juni 2008 op de website.

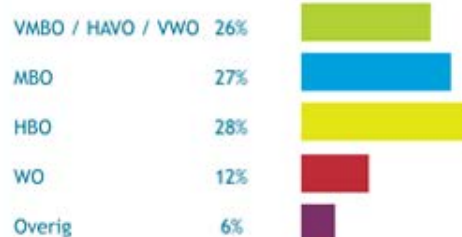
Geslacht *



Leeftijd *



Opleidingsniveau *



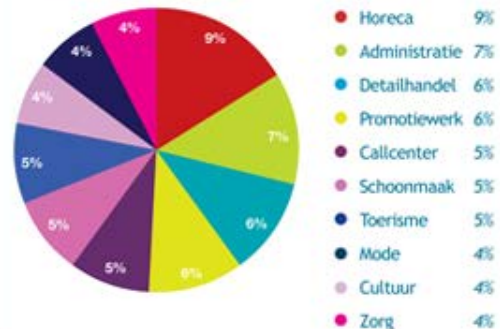
Kandidaten per regio *

De top 10 van de in totaal 60 regio's



Populairste branches *

De top 10 meest populaire branches



* Bovengenoemde cijfers hebben betrekking op de 44.000 actieve profielen uit de database van BijBanen.nl in juni 2008.

- Hoeveel deelnemers van de leeftijdscategorie 14 tot en met 16 jaar zijn er ongeveer?
- Hoeveel deelnemers zitten er ongeveer op vmbo/havo/vwo?
- In welke branche hebben de meeste deelnemers een bijbaan? Hoeveel ongeveer?
- In het cirkeldiagram lijkt het dat bijna een kwart van de deelnemers in de regio Amsterdam zit. Klopt dat ook?
- Waarom zijn de twee cirkeldiagrammen misleidend? Wat is er fout?
- Kun je nu zeggen dat 9 procent van alle mensen in Nederland met een bijbaan in de horeca werkt?

Opgave 12

De tabel hieronder stamt uit een onderzoek van het CBS naar vakanties van Nederlanders. Een lange vakantie is een vakantie met minstens vier overnachtingen.

- Waaraan kun je zien dat de grenzen van de regio's niet zuiver langs de provinciegrenzen lopen?
- Over welke populatie gaat deze tabel? Welke variabelen zijn er?
- Hoeveel personen uit regio Noord hielden in 2008 een lange vakantie?
- Uit welke twee variabelen bereken je het *Gemiddeld aantal lange vakanties per participant*? Hoe gaat die berekening?
- Kun je op grond van deze tabel concluderen dat mensen uit stedelijke gebieden meer behoefte aan lange vakanties hebben? Motiveer je antwoord.

Tabel 1 Participatie aan lange vakanties en aantal lange vakanties naar bestemming van enkele categorieën van de bevolking, 2008 (slot)

	Omvang van de populatie	Participatie aan lange vakanties	Gemiddeld aantal lange vakanties per participant	Totaal aantal lange vakanties	w.v.	
					in Nederland	in het buitenland
	x 1 000	%		x 1 000	%	
Woonprovincie						
Groningen	550	70	1,82	700	48	52
Friesland	620	71	2,10	930	45	55
Drenthe	470	68	1,88	600	41	59
Overijssel	1 080	74	1,92	1 530	38	62
Flevoland	360	83	1,99	590	42	58
Gelderland	1 900	74	1,96	2 770	34	66
Utrecht	1 140	82	2,14	2 000	32	68
Noord-Holland	2 440	77	2,09	3 890	31	69
Zuid-Holland	3 220	78	2,14	5 390	40	60
Zeeland	360	67	1,95	470	47	53
Noord-Brabant	2 310	75	1,95	3 370	33	67
Limburg	1 060	66	1,97	1 380	31	69
Totaal	15 510	75	2,03	23 630	36	64
Regionale indeling vakantiespreiding						
Regio Noord	5 520	74	2,01	8 270	37	63
Regio Midden	5 650	78	2,11	9 300	38	62
Regio Zuid	4 340	72	1,93	6 050	32	68
Totaal	15 510	75	2,03	23 630	36	64
Stedelijkheid gemeenten						
Zeer sterk stedelijk	2 610	75	2,12	4 140	32	68
Sterk stedelijk	4 380	77	2,06	6 990	37	63
Matig stedelijk	3 140	76	2,04	4 880	38	62
Weinig stedelijk	3 420	74	1,96	4 970	36	64
Niet stedelijk	1 960	71	1,90	2 650	38	62
Totaal	15 510	75	2,03	23 630	36	64

Bron: Continu Vakantie Onderzoek.

§ 1.2 Data

Verkennen

Opgave 13

Stel je voor dat je een statistisch onderzoekje houdt onder de leerlingen in jouw eigen jaargroep.

Je vraagt naar:

- Geslacht (mannelijk of vrouwelijk).
- Lichaamslengte.
- Lichaamsgewicht.
- Omvang van het gezin.
- Afstand tot de school.
- Bloedgroep.
- Kleur van de ogen.
- Zakgeld per maand in euro's.
- Voorkeur voor klassieke muziek.

Sommige van deze variabelen geven een hoeveelheid (een grootte) aan, andere een kenmerk.

- Welke van deze variabelen geven een kenmerk aan?
- Welke van deze variabelen kun je in een getal uitdrukken?
- Bij welke van deze variabelen is het zinvol om ze op volgorde te zetten?

Opgave 14

Hieronder staan de gegevens over de gezinsomvang van leerlingen uit twee havo 4-klassen:

H4A: 3, 5, 2, 6, 5, 5, 2, 3, 4, 6, 4, 3, 5, 2, 5, 4, 4, 3, 4, 4, 5, 3, 7, 4, 4, 5, 2, 4

H4B: 4, 2, 5, 3, 4, 2, 4, 3, 5, 2, 3, 4, 6, 3, 3, 4, 3, 5, 4, 2, 4, 2, 5, 3

- Ga na dat de frequentietabel hieronder bij de data past en leg uit waarom deze frequentietabel handiger is:

Gezinsomvang (aantal personen)	Frequenties H4A	Frequenties H4B
2	4	
3	5	
4	9	
5	7	
6	2	
7	1	

- Hoeveel leerlingen zitten er in H4A?
- Hoeveel personen zitten er in totaal in de gezinnen waaruit de leerlingen van H4A kwamen?
- Hoeveel bedraagt de modus van de gezinsomvang in H4A?
- Hoeveel bedraagt de gemiddelde gezinsomvang in H4A?
- Vul de tabel verder in voor H4B.

In de frequentietabel staan de absolute frequenties. Door ze allemaal te delen door het totale aantal leerlingen krijg je relatieve frequenties die je ook kunt geven als percentages. Je kunt dan gemakkelijker beide tabellen vergelijken.

- g. Maak een nieuwe tabel met relatieve frequenties.
- h. Beschrijf de verschillen tussen H4A en H4B.

Uitleg

Kijk je naar de voorgaande twee opgaven, dan kun je vaststellen:

- Statistische variabelen als geslacht, kleur ogen, godsdienst, bloedgroep, naam, e.d. geven alleen een kenmerk van de populatie weer, maar hoeven niet noodzakelijkerwijs in een getal te worden uitgedrukt; dit zijn *kwantitatieve variabelen*.
- Statistische variabelen als lengte, hoogte van het inkomen, omvang van het gezin, e.d. moeten wel in een getal worden uitgedrukt; dit zijn *kwantitatieve variabelen*.

Het is vaak handig om je gegevens te ordenen in *frequentietabellen*. Daarin hebben niet alleen de data een logische volgorde, maar is ook meteen te zien hoe vaak een bepaalde waarde van de statistische variabele voorkomt, zijn *absolute frequentie*.

Bij de gegevens over de gezinsomvang van leerlingen uit twee havo 4-klassen heb je twee datasets bij eenzelfde kwantitatieve variabele.

Om beide datasets goed te kunnen vergelijken is het handiger om met *relatieve frequenties* te werken, d.w.z. absolute frequenties gedeeld door het totaal, eventueel in procenten.

Opgave 15

Hieronder wordt telkens een variabele beschreven die betrekking heeft op de werknemers van een bepaald bedrijf.

Geef aan of het een kwalitatieve of een kwantitatieve variabele betreft. Geef ook aan of het maken van een frequentietabel zinvol is en waarom.

- a. Schoenmaat.
- b. Lengte.
- c. Inkomen in euro's.
- d. Burgerlijke staat.
- e. Aantal dienstjaren bij het bedrijf.
- f. Aantal personen waaruit het huishouden van de betreffende werknemer bestaat.

Theorie

Er bestaan verschillende soorten variabelen en dus verschillende soorten data. Een belangrijk onderscheidend kenmerk is:

- Een **kwantitatieve variabele** beschrijft van elk element van de populatie een bepaald kenmerk dat niet noodzakelijk in een getal wordt uitgedrukt (bijvoorbeeld geslacht, kleur ogen, godsdienst, bloedgroep, naam, e.d.).
- Een **kwantitatieve variabele** kan wel in een getal worden uitgedrukt (zoals lengte, hoogte van het inkomen, omvang van het gezin, e.d.).

Bij veel statistische variabelen kun je een **frequentietabel** maken.

Door tellen hoe vaak een bepaalde waarde van die variabele voorkomt krijg je de **absolute frequentie** van die waarde. Deel je die absolute frequentie door het totale aantal waarden dan krijg je de **relatieve frequentie** van die waarde, eventueel in procenten.

Relatieve frequenties maken het vergelijken van twee datasets gemakkelijker.

Voorbeeld

In 1947 hielden de wiskundigen Freudenthal en Sittig een statistisch onderzoek ten behoeve van een nieuw maatsysteem voor vrouwenkleding in opdracht van het warenhuis De Bijenkorf. Onder andere maten zij nauwkeurig de mouwlengte van 5001 vrouwen in centimeters.

Hier zie je een frequentietabel met hun data.

- Welke absolute en welke relatieve frequentie horen er bij een mouwlengte van 60 centimeter?
- Hoeveel procent van deze vrouwen had een mouwlengte vanaf 64 centimeter?

Uitwerking

- Er zijn 653 vrouwen met een mouwlengte van 60 centimeter, de absolute frequentie is dus 653.
De relatieve frequentie is nu $\frac{653}{5001} \approx 0,131$.
Dat is ongeveer 13 procent.
- Een mouwlengte vanaf 64 centimeter betekent dat je met de mouwlengtes 64, 65, ..., 71 te maken hebt.
Het gaat daarbij om $159 + 106 + 52 + 18 + 15 + 3 + 0 + 1 = 354$ vrouwen.
En dat is: $\frac{354}{5001} \approx 0,071$. Dat is ongeveer 7 procent.

mouw- lengte	frequentie
49	3
50	11
51	22
52	53
53	89
54	163
55	250
56	405
57	519
58	660
59	578
60	653
61	560
62	421
63	260
64	159
65	106
66	52
67	18
68	15
69	3
70	0
71	1
	5001

Opgave 16

Bekijk de tabel uit het voorbeeld nog eens.

- Is *mouwlengte* een kwalitatieve of een kwantitatieve statistische variabele?
- Maak zelf een tabel met relatieve frequenties voor de mouwlengtes.
- Welk nut heeft zo'n tabel voor De Bijenkorf?

		Statistisch onderzoek 1947																								
		De Bijenkorf - Amsterdam																								
		5001 vrouwen met leeftijd > 18 jr.																								
		kniehoogte in cm																								
		34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54				
mouwlengte in cm	71																			1			1			
	70																									
	69															1										
	68										2	2	1			3	3	2	2							15
	67											3	1	2	4	2	2	3			1					18
	66											4	5	5	4	6	10	7	7	3				1		52
	65								1		2	7	3	15	13	26	16	11	6	3	1	1	1	1		106
	64					1		2	2	5	5	13	22	26	31	22	18	4	4	4						159
	63						3	1	5	5	20	36	44	48	40	29	16	6	4	2	1					260
	62						2	6	15	21	61	61	84	62	59	27	14	6	1	2						421
	61					3	2	6	29	43	65	99	111	91	61	33	9	7	1							560
	60					3	5	17	40	68	112	117	121	84	49	31	2	2	2							653
	59			1	1	2	11	20	60	110	108	88	66	72	22	12	3	1			1					578
	58				3	8	21	50	72	110	141	103	90	34	22	5	1									660
	57				1	2	9	28	60	80	92	87	72	53	20	12	2	1								519
	56		1	2	4	13	30	50	79	82	60	40	32	8	3			1								405
55			3	6	11	27	38	46	46	39	14	16	1	3											250	
54			4	7	11	17	32	37	23	14	8	4	1	1	3	1									163	
53			3	6	6	16	12	17	15	8	6														89	
52		1	2	3	5	9	13	10	5	4	1														53	
51		1	2		4	4	5	2	2	1	1														22	
50	1					3		1	3		3														11	
49			1	2																					3	
		1	3	19	34	76	178	313	495	632	738	675	665	466	343	195	88	45	18	12	2	3		5001		

Opgave 17

Je ziet hierboven de kruistabel die Freudenthal en Sittig maakten van de mouwlengte en de kniehoogte in centimeters van 5001 vrouwen.

- Hoeveel procent van de onderzochte vrouwen heeft mouwlengte 60 centimeter en kniehoogte 50 centimeter?
- Welke combinatie van mouwlengte en kniehoogte komt het meeste voor?
- Hoeveel procent van de vrouwen met een mouwlengte van 60 centimeter heeft een kniehoogte van 50 centimeter?
- Hoeveel procent van deze vrouwen heeft een mouwlengte van meer dan 68 centimeter en een kniehoogte van meer dan 50 centimeter?
- Hoeveel bedroeg de gemiddelde mouwlengte van de 5001 vrouwen?
En de gemiddelde kniehoogte?

Opgave 18

Waarom hebben Freudenthal en Sittig een kruistabel zoals die in opgave 17 gemaakt?

Welke andere kruistabellen zou je ook nog kunnen maken bij een onderzoek dat als doel heeft een nieuw maatsysteem voor vrouwenkleding te ontwerpen?

Opgave 19

Stel je voor dat je het onderzoek van Freudenthal en Sittig dit jaar opnieuw zou uitvoeren.

Je kunt nu (met de moderne hulpmiddelen) veel meer vrouwen in De Bijenkorf laten opmeten.

Als je de resultaten wilt vergelijken met die van Freudenthal en Sittig is het verstandig om met relatieve frequenties te werken. Laat met een rekenvoorbeeld zien waarom dat zo is.

Verwerken

▣ Opgave 20

De tabel hieronder laat het aantal geslaagden zien op havo en vwo gedurende drie schooljaren (bron: CBS juli 2009).

Voortgezet onderwijs; geslaagden naar onderwijssoort									
Onderwerpen	Aantal examenkandidaten			Aantal geslaagden			Percentage geslaagden		
Perioden	2005*06	2006*07	2007*08 ^a	2005*06	2006*07	2007*08 ^a	2005*06	2006*07	2007*08 ^a
Onderwijssoort	aantal	aantal	aantal	aantal	aantal	aantal	%	%	%
Vwo	31633	32961	35150	29287	30331	32398	93	92	92
Vwo natuurprofiel totaal	14967	15940	17233	13750	14570	15723	92	91	91
Vwo natuur en techniek	3765	3510	3970	3476	3213	3662	92	92	92
Vwo natuur en gezondheid	9641	10268	11259	8771	9275	10097	91	90	90
Vwo natuur ongedeeld	1561	2162	2004	1503	2082	1964	96	96	98
Vwo maatschappijprofiel totaal	16657	17007	17908	15528	15747	16666	93	93	93
Vwo economie en maatschappij	10087	10142	10801	9413	9392	10112	93	93	94
Vwo cultuur en maatschappij	6359	6623	6858	5911	6126	6309	93	92	92
Vwo maatschappij ongedeeld	211	242	249	204	229	245	97	95	98
Havo	42983	44560	46313	38105	39755	41371	89	89	89
Havo natuurprofiel totaal	12008	12641	13634	10743	11181	12306	89	88	90
Havo natuur en techniek	3699	3618	3871	3319	3203	3492	90	89	90
Havo natuur en gezondheid	7360	7853	8704	6546	6906	7815	89	88	90
Havo natuur ongedeeld	949	1170	1059	878	1072	999	93	92	94
Havo maatschappijprofiel totaal	30974	31911	32679	27361	28571	29065	88	90	89
Havo economie en maatschappij	15705	16203	16959	13607	14321	14993	87	88	88
Havo cultuur en maatschappij	14751	15182	15264	13294	13778	13649	90	91	89
Havo maatschappij ongedeeld	518	526	456	460	472	423	89	90	93

- Om welke populatie gaat het bij deze data?
- Welke variabelen worden er onderzocht? Welke soort variabelen betreft het?
- Waar vind je absolute frequenties en waar relatieve frequenties?
- Op havo is in 2007/2008 het geslaagdenpercentage 89. Toon dit aan door een berekening.
- Hoeveel procent van de examenkandidaten havo heeft in 2007/2008 een N-profiel gekozen?
En hoeveel M-profielers zijn er dat jaar op havo?
- Kun je het geslaagdenpercentage havo in 2007/2008 berekenen vanuit de percentages geslaagden van de M-profielen en de N-profielen afzonderlijk?
- Het aantal examenkandidaten met een N-profiel op havo neemt absoluut gezien toe.
Is dat relatief bekeken ook zo?
- Kun je verklaren waarom de examenkandidaten in alle ongedeelde N-profielen of M-profielen een hoger geslaagdenpercentage hebben dan het totaal per profiel?

Opgave 21

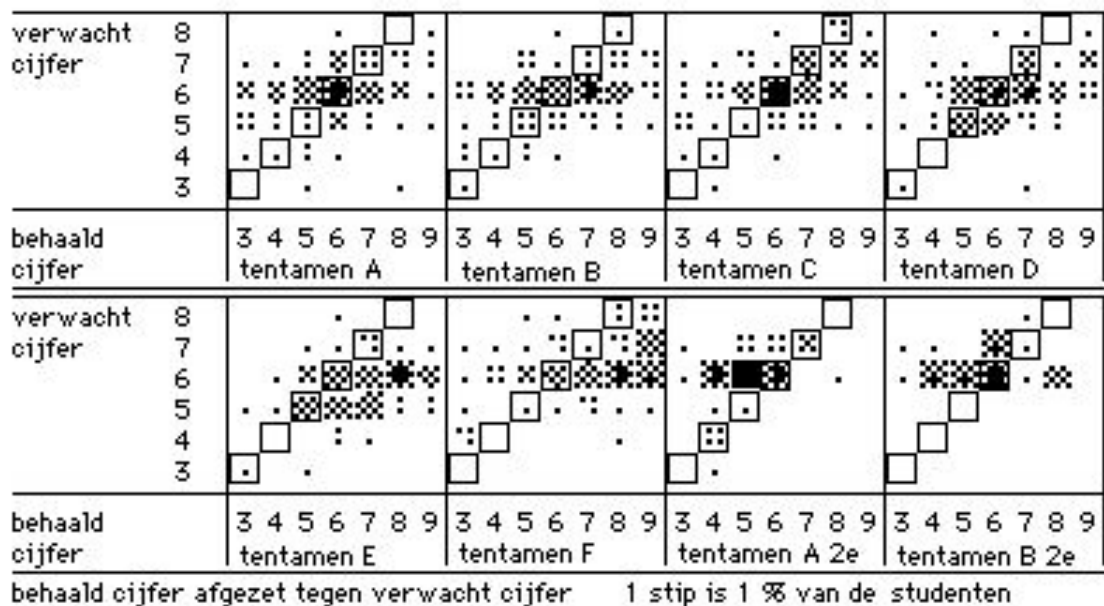
Voor een biologiepracticum moet het aantal slakken op een stuk grond worden geteld. Het stuk grond wordt daartoe in stukken van een vierkante meter verdeeld. Iedere leerling telt het aantal slakken op vier van die stukken. Hieronder zie je de resultaten:

Aantal slakken per m ²	1	3	4	5	6	7	8	9
Frequentie	16	14	7	4	2	3	1	1

- Om welke populatie gaat het hier en om welke variabele? Wat voor soort variabele betreft het?
- Hoeveel vierkante meter is de oppervlakte van dat stuk grond?
- Hoeveel leerlingen hebben er geteld?
- Hoeveel slakken zijn er in totaal geteld?
- Hoeveel slakken zijn er gemiddeld per vierkante meter gevonden?

Opgave 22

Hier zie je kruistabellen van zes tentamens en twee hertentamens.



- Welk tentamen werd duidelijk slechter gemaakt dan de meeste studenten hadden verwacht?
- Hoeveel studenten behaalden bij tentamen A een beter cijfer dan verwacht als er 240 studenten deelnamen?
- Kun je zeggen dat de twee hertentamens moeilijker waren dan de bijbehorende eerste versies?

§ 1.3 Diagrammen

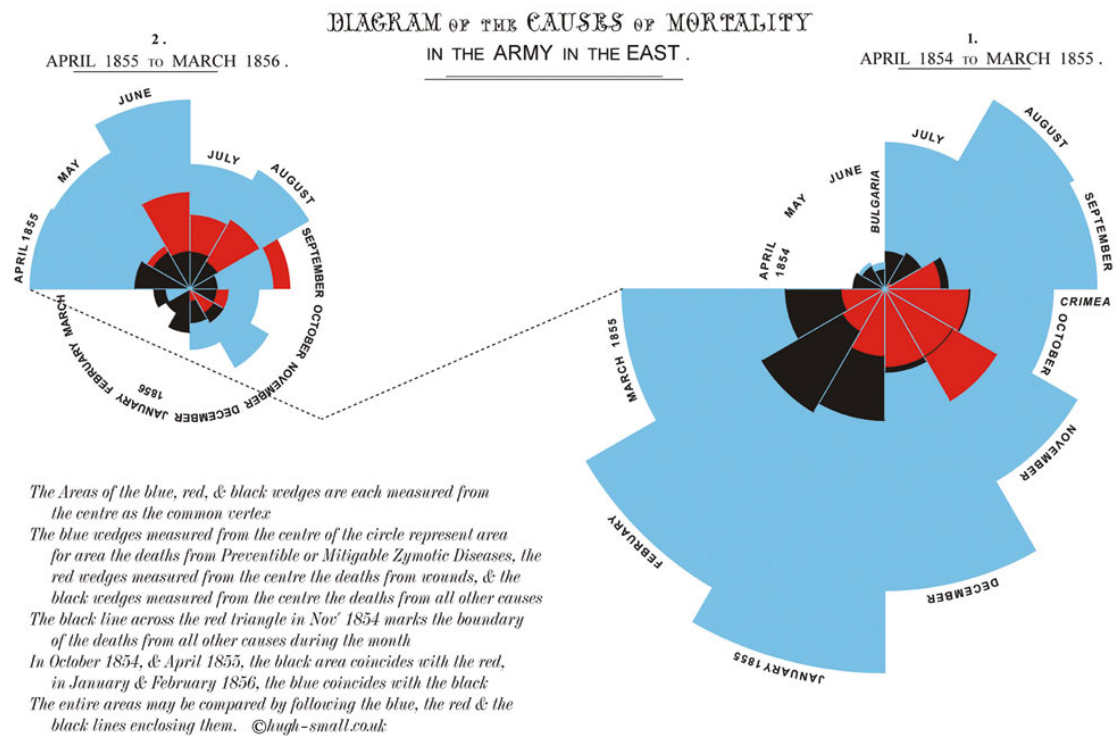
Verkennen

▣ Opgave 23

Verpleegster Florence Nightingale was zo geschokt door het gebrek aan hygiëne onder het Britse leger dat vocht tijdens de Krimoorlog, dat ze bij terugkeer deze diagrammen maakte.

Rood stelt 'dood door verwonding' voor, blauw is 'dood door ziektes als gevolg van de slechte omstandigheden' en zwart is 'dood door andere oorzaken'.

Elk diagram is in twaalf sectoren verdeeld en bij elke sector is *vanaf het middelpunt* een deel rood gekleurd, een deel zwart en een deel blauw. Die drie delen overlappen elkaar! De oppervlakte van elke sector stelt het aantal militairen voor.

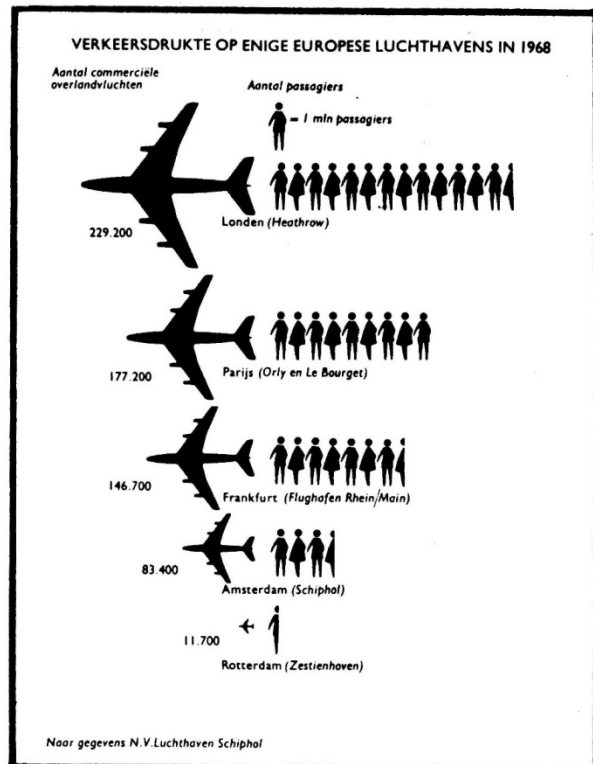


- Vooral in januari 1855 was de sterfte onder de soldaten hoog. Waaraan stierf verreweg het grootste deel van de militairen in die maand?
- Wat betekent het in deze figuur als in een bepaalde sector het rode gebied het zwarte gebied overlapt? In welke maanden is dat zo?
- In welke maand is het aantal doden door verwondingen relatief gezien het hoogst?
- In welke maanden zijn er helemaal geen doden door verwondingen?
- Waarom zou Florence Nightingale voor dit type diagram hebben gekozen? Welke andere mogelijkheden zijn er?

Opgave 24

In 1968 gaf de toenmalige NV Luchthaven Schiphol dit overzicht uit van de verkeersdrukte op sommige Europese luchthavens.

- Waarom wordt dit type diagram wel 'beelddiagram' genoemd?
- Hoeveel passagiers werden er in 1968 op Heathrow vervoerd?
- Bereken hoeveel passagiers er gemiddeld per vlucht op elk van deze luchthavens werden vervoerd.
- Het diagram bestaat uit twee delen, waarbij elk deel op een eigen manier de aantallen weergeeft. Leg dat uit.



Uitleg

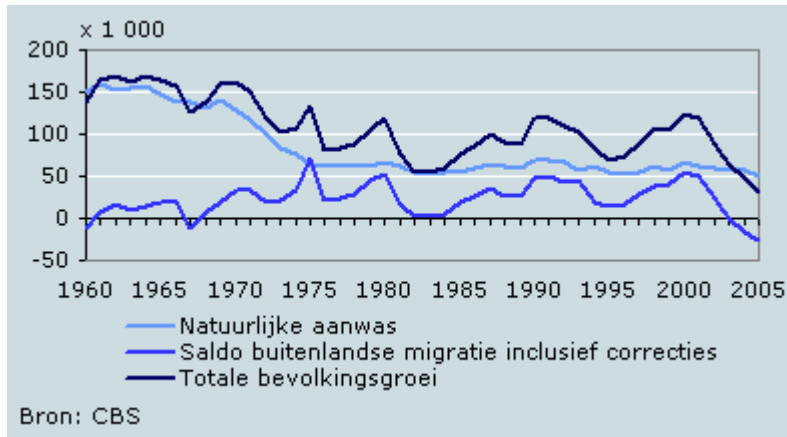
Als je data overzichtelijk wilt weergeven of een bepaalde tendens zichtbaar wilt maken, dan zijn tabellen vaak niet erg handig. Je vindt er wel alle gegevens in, maar je mist overzicht, je kunt slecht in één oogopslag zien wat er aan de hand is. Wil je (zoals Florence Nightingale) een boodschap overbrengen, dan zeggen diagrammen meestal meer.

Het lijkt wel of er oneindig veel verschillende soorten diagrammen bestaan. Je zult er in deze paragraaf een flink aantal voorbij zien komen. Het **beelddiagram**, **staafdiagram**, **lijndiagram** en **cirkeldiagram** ken je al. Daarnaast bestaat er een veelheid aan bijzondere diagrammen voor heel specifieke situaties. In deze paragraaf kom je ook daarvan een aantal tegen.

Opgave 25

Als je van het diagram van Florence Nightingale een staafdiagram wilt maken, kun je dan de staven even hoog nemen als de stralen van de cirkelsectoren?

Opgave 26



Deze diagrammen geven een beeld van de groei van de Nederlandse bevolking.

- Wat voor soort diagrammen zijn het?
- Beschrijf wat je hieruit kunt aflezen.
- Hoe leid je het diagram van de totale bevolkingsgroei af uit de andere twee?
- Wat betekent het afnemen van de totale bevolkingsgroei voor het aantal inwoners van Nederland?

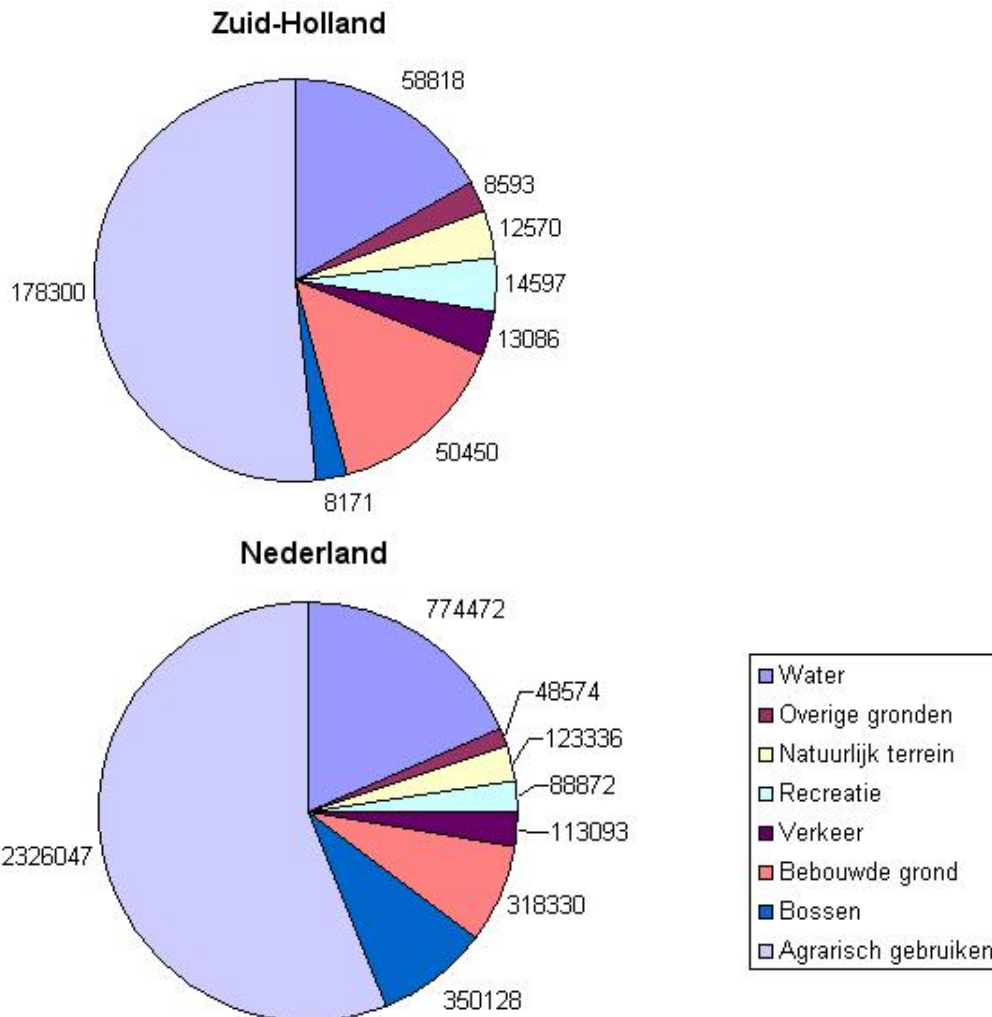
Opgave 27

Bij opgave 10 tref je twee soorten diagrammen aan. Een daarvan is een *gestapeld staafdiagram*. Leg uit wat dat is en wat er in dit geval is gestapeld.



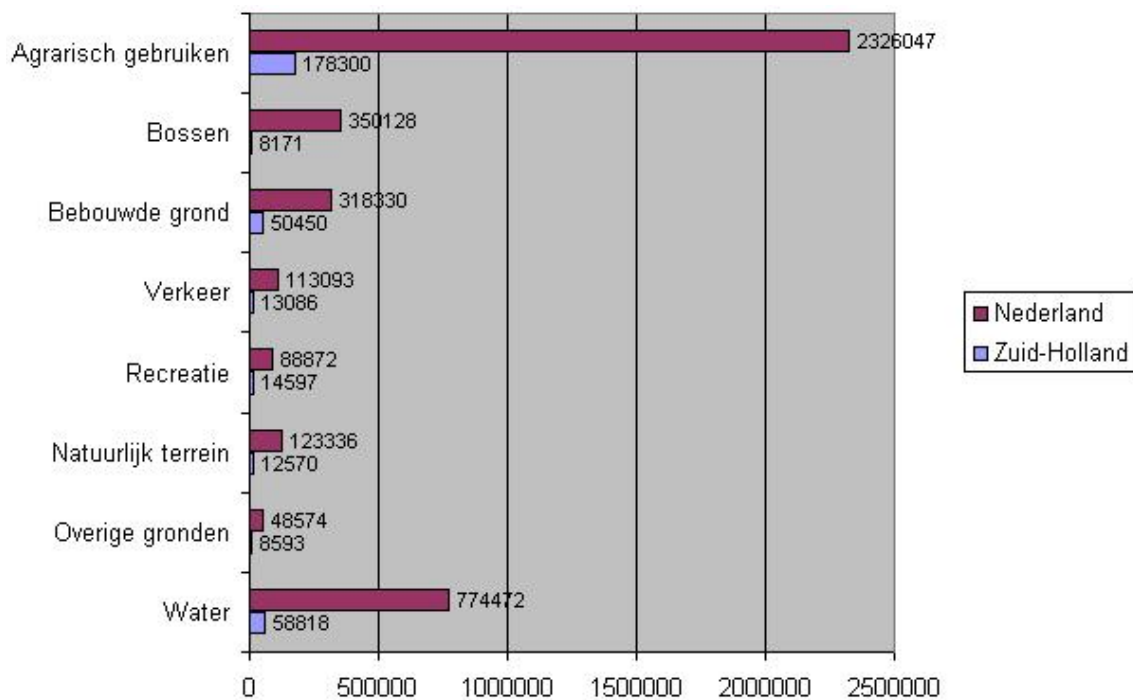
Opgave 28

Deze twee cirkeldiagrammen vergelijken het bodemgebruik van Zuid-Holland met dat van heel Nederland.



- De getallen in beide diagrammen verschillen nogal, toch kun je in één oogopslag zien dat in Zuid-Holland relatief gezien veel minder bossen zijn dan in Nederland als geheel. Hoe komt dat?
- Laat door berekening zien dat de *sectorhoeken* van de categorie *bossen* correct zijn.
- Bij een cirkeldiagram gaat het eigenlijk om de oppervlakte van een cirkelsector.
Wat is er dus mis in de bovenstaande figuur?

Dezelfde gegevens kun je ook in een staafdiagram weergeven. Je ziet het op de volgende bladzijde. Wat geeft deze figuur beter weer?



Theorie

In een tabel heb je vaak alle data nauwkeurig voorhanden, maar een **diagram** geeft vaak beter overzicht en duidelijker een bepaald verloop weer.

Er zijn veel soorten diagrammen. In deze paragraaf kom je de belangrijkste tegen:

- Een **beelddiagram** (*pictogram*) geeft de waarden of de frequenties weer door de oppervlakte van een figuur of het aantal figuren.
- Een **staafdiagram** (*bar chart*) geeft de waarden of de frequenties weer door de oppervlaktes van rechthoekige staven.
- Een **lijndiagram** (*line chart*) geeft de waarden of frequenties weer door punten in een assenstelsel die door rechte lijnstukken met elkaar worden verbonden.
- Een **cirkeldiagram** (*pie chart*) geeft de waarden of frequenties weer door de oppervlakte van een cirkelsector.
- Een **kaartdiagram** (*cartogram*) presenteert de gegevens in de vorm van een kaart.
- Een **puntenwolk** of **spreidingsdiagram** (*scatterplot*) zet twee variabelen tegen elkaar uit en duidt de gegevens aan door punten in een assenstelsel (past bij een kruistabel).

Voorbeeld

Hier zie je nog eens de tabel van de mouwlengte van 5001 vrouwen die Freudenthal en Sittig maakten bij hun onderzoek.

Maak er een staafdiagram van.

Uitwerking

Op de horizontale as komt een schaalverdeling die logisch oploopt met stappen van 1 van 49 tot en met 71. Bij elke waarde voor de voetslengte hoort een interval van getallen die op die waarde worden afgerond: bijvoorbeeld bij 54 horen waarden vanaf 53,5 tot 54,5. Dit is het interval $[53,5; 54,5)$.

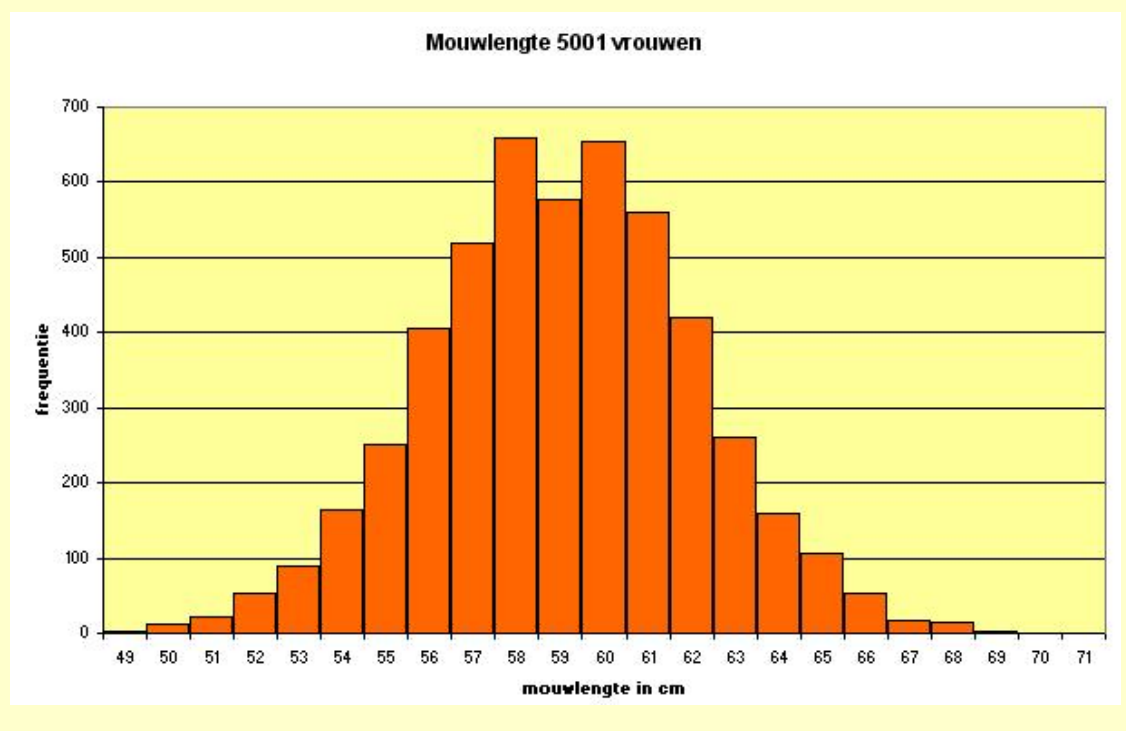
De staaf die bij de waarde 54 hoort heeft een hoogte van 163 eenheden.

Het getal 54 staat onder het midden van deze staaf.

En zo gaat dat ook voor alle andere voetslengtes.

Je krijgt dan dit staafdiagram:

mouw- lengte	frequentie
49	3
50	11
51	22
52	53
53	89
54	163
55	250
56	405
57	519
58	660
59	578
60	653
61	560
62	421
63	260
64	159
65	106
66	52
67	18
68	15
69	3
70	0
71	1
	5001



Opgave 29

Bekijk het voorbeeld.

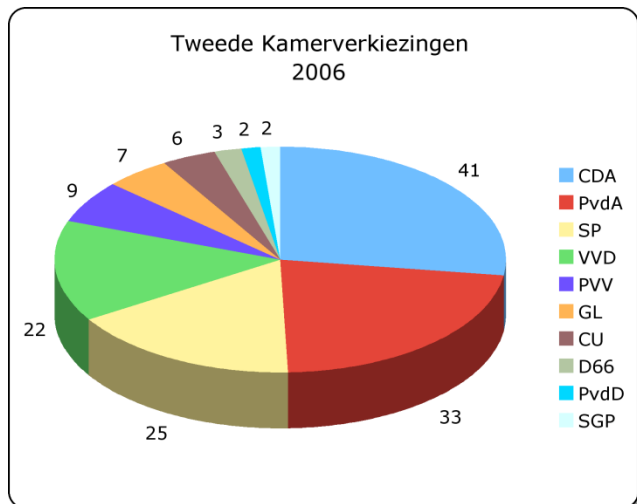
- Teken een lijndiagram bij de data over de mouwlengtes van de 5001 vrouwen die Freudenthal en Sittig hebben onderzocht.
- Is een cirkeldiagram hier een geschiktere manier om de gegevens te presenteren dan een staafdiagram? Verklaar je antwoord.

Opgave 30

Dit cirkeldiagram geeft de uitslag van de Tweede Kamerverkiezingen in 2006

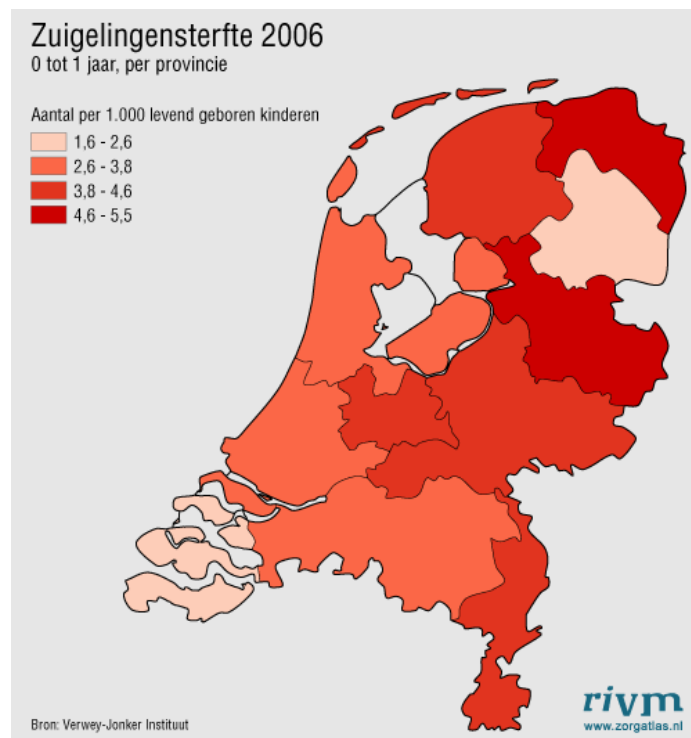
(bron: Wikipedia).

- Stellen de getallen percentages voor of aantallen Kamerzetels?
- Het cirkeldiagram is in 3D weergegeven en dus wat vertekend. Maak zelf het correcte cirkeldiagram door de sectorhoeken te berekenen.
- Maak een staafdiagram van deze data. Is het belangrijk in welke volgorde de partijen in jouw staafdiagram staan?



Opgave 31

Dit kaartdiagram geeft informatie over de zuigelingensterfte in Nederland.

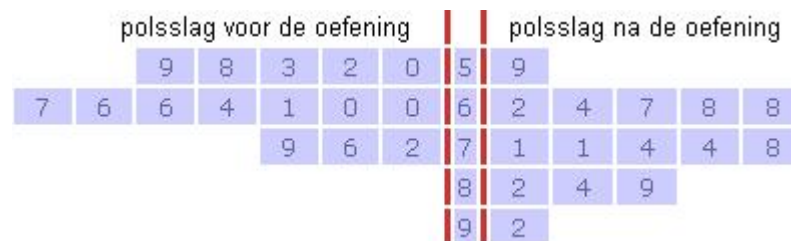


- Waarom is er gekozen voor deze presentatievorm ?
- Kun je hierbij een staafdiagram of een lijndiagram maken? Beschrijf de nadelen daarvan.



Opgave 32

Als je in de sportzaal een tijdje een bepaalde oefening hebt gedaan, gaat vaak je polsslag wat omhoog. In dit tweezijdige steelbladdiagram vind je bijbehorende data. Van elke sporter werd één keer voor en één keer na de oefening de polsslag gemeten.

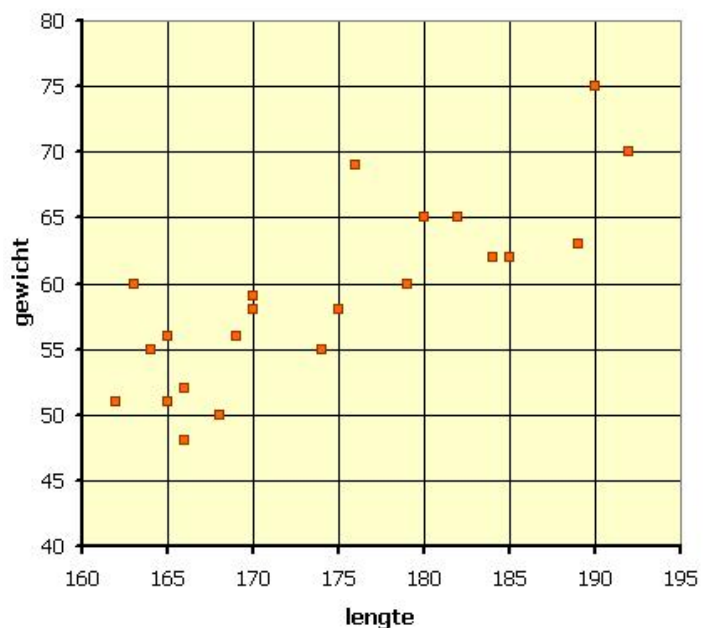


De laagst gemeten polsslag voor de oefening is 50, na de oefening is dat 59.

- Hoeveel sporters doen er mee aan het onderzoekje?
- Hoeveel bedraagt de hoogste polsslag voor de oefening? En na de oefening?
- Hoeveel bedraagt de gemiddelde polsslag voor en na de oefening?
- Gaat de polsslag door de oefening inderdaad omhoog?
- Er wordt wel gezegd dat een **steelbladdiagram** een combinatie is van een staafdiagram en een tabel. Verklaar dat.

Opgave 33

Hier zie je een puntenwolk van lengte en gewicht van de havoscholieren uit één klas (*lengte* in cm en *gewicht* in kg).

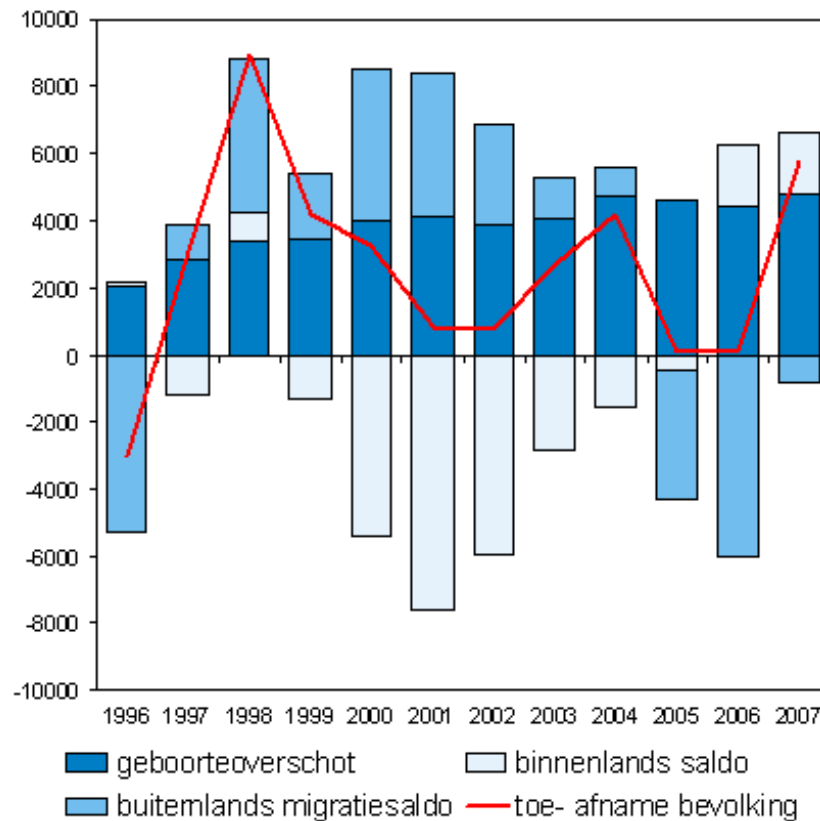


- Om hoeveel scholieren gaat het hier?
- Hoe lang is de zwaarste leerling? Is hij ook het langst?
- Hoeveel weegt de kleinste leerling? Is hij ook het lichtst?
- Toevallig is er bij het meten van deze gegevens iemand ziek. Hij is 186 centimeter lang. Hoe zwaar zal hij waarschijnlijk zijn?

Verwerken

Opgave 34

Dit diagram geeft informatie over de bevolking van Amsterdam.



bron: O+S

- Welke soorten diagrammen tref je aan in de figuur?
- Wat betekenen de variabelen *geboorteoverschot*, *buitenlands migratiesaldo* en *binnenlands saldo*?
- Hoeveel ongeveer bedraagt de bevolkingstoename van Amsterdam in 2004? Hoeveel zijn in dat jaar het geboorteoverschot, het buitenlands migratiesaldo en het binnenlands saldo?
- Het migratiesaldo zit de ene keer boven en de andere keer onder de nullijn. Geef daar een verklaring voor.
- Aan het lijndiagram zie je dat in 2007 de Amsterdamse bevolking met ongeveer 6000 personen is toegenomen. Laat zien hoe je dit kunt berekenen vanuit het staafdiagram.

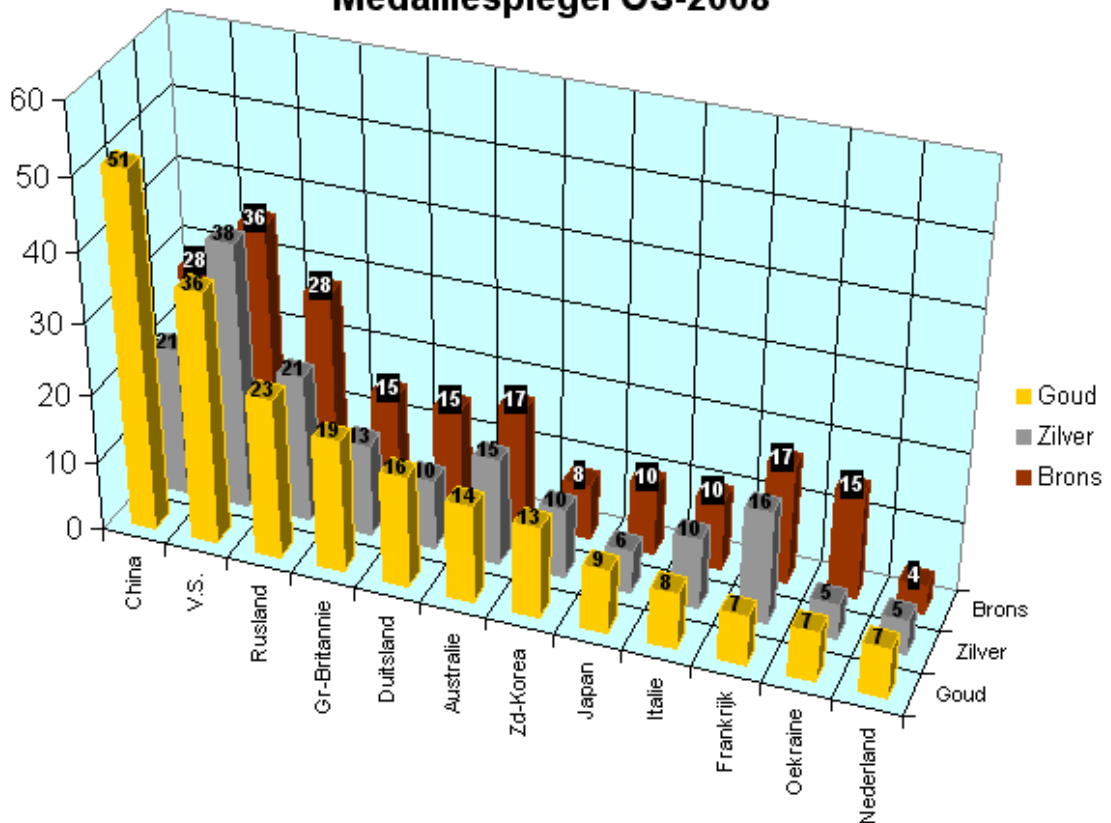
Opgave 35

Op de volgende pagina zie je de medaillespiegel van de Olympische Spelen van 2008 in Peking, alleen de beste twaalf landen.

- Wat geeft elke staaf in dit diagram weer?
- Waarom is er een 3D-diagram gebruikt? Wat staat er op elk van de drie assen weergegeven?
- Welk land heeft de meeste gouden medailles gewonnen?
- Welk land won de meeste zilveren medailles?
- Welk land heeft in totaal de meeste medailles gewonnen?
- Deze gegevens kun je ook in een gestapeld staafdiagram weergeven. Wat is daarvan het voordeel? En wat is het nadeel?
- Maak zo'n gestapeld staafdiagram.
- Maak ook een staafdiagram waarbij je voor elk land drie staafjes naast elkaar zet voor goud/zilver/brons.
- Kun je nog een andere wijze van presentatie bedenken die alle gewenste informatie bevat en toch een duidelijk overzicht geeft?



Medaillespiegel OS-2008



Opgave 36

Bekijk nog eens de kruistabel voor *mouw lengte* en *knieshoogte* van Freudenthal en Sittig.

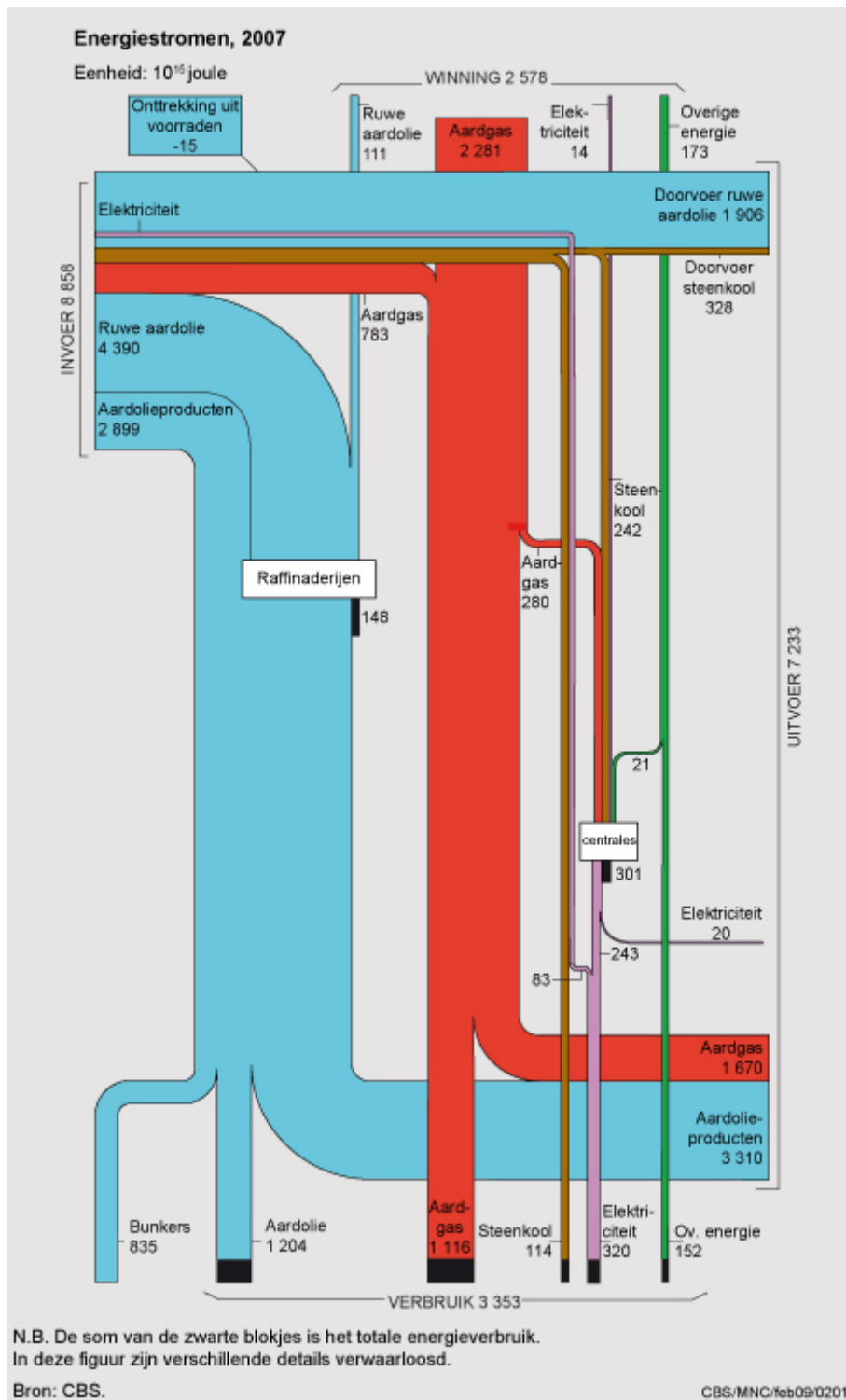
- Teken een staafgram van de variabele *knieshoogte* (in cm).
- Teken ook een lijndiagram van de *knieshoogte*. Gebruik nu relatieve frequenties.
- Kun je een manier bedenken om een puntenwolk te tekenen van de gegevens uit onderstaande kruistabel?

		Statistisch onderzoek 1947																						
		De Bijenkorf - Amsterdam																						
		5001 vrouwen met leeftijd > 18 jr.																						
		knieshoogte in cm																						
		34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54		
mouw lengte in cm	71																					1	1	
	70																							3
	69														1								1	3
	68										2	2	1		3	3	2	2						15
	67											3	1	2	4	2	2	3						18
	66											4	5	5	4	6	10	7	7	3			1	52
	65							1		2	7	3	15	13	26	16	11	6	3	1	1	1		106
	64				1		2	2	5	5	13	22	26	31	22	18	4	4	4					159
	63					3	1	5	5	20	36	44	48	40	29	16	6	4	2	1				260
	62						2	6	15	21	61	61	84	62	59	27	14	6	1	2				421
	61					3	2	6	29	43	65	93	111	91	61	33	9	7	1					560
	60					3	5	17	40	68	112	117	121	84	49	31	2	2	2					653
	59			1	1	2	11	20	60	110	108	88	66	72	22	12	3	1			1			578
	58				3	8	21	50	72	110	141	103	90	34	22	5	1							660
	57			1	2	9	28	60	80	92	87	72	53	20	12	2	1							519
	56		1	2	4	13	30	50	79	82	60	40	32	8	3			1						405
	55			3	6	11	27	38	46	46	39	14	16	1	3									250
	54			4	7	11	17	32	37	23	14	8	4	1	1	3	1							163
	53			3	6	6	16	12	17	15	8	6												89
	52		1	2	3	5	9	13	10	5	4	1												53
51		1	2		4	4	5	2	2	1	1												22	
50	1					3		1	3		3												11	
49			1	2																			3	
		1	3	19	34	76	178	313	495	632	738	675	665	466	343	195	88	45	18	12	2	3	5001	



▣ Opgave 37

Dit stroomdiagram stelt de energiebalans van Nederland voor. Je vindt er de hoeveelheid energie die Nederland zelf opwekt en invoert en ook de energie die we met z'n allen verbruiken dan wel doorvoeren/uitvoeren naar het buitenland. De gebruikte eenheid is 10^{15} joule.



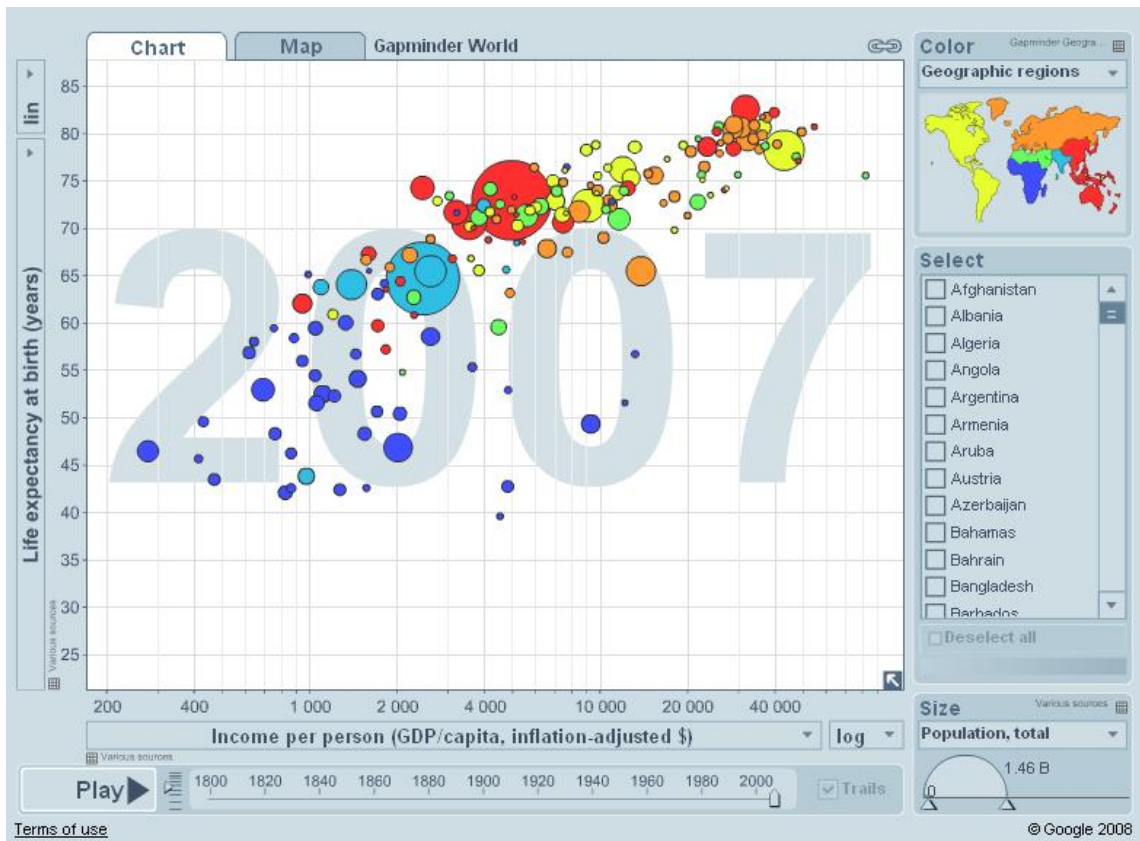
- a. Wat betekent het getal 2281 bij de aardgaswinning?
- b. Hoeveel joule aan energie is er in 2007 verbruikt door onze energiecentrales om elektriciteit op te wekken?
- c. Deze energiecentrales halen hun energie behalve uit aardgas en steenkool ook uit andere energiebronnen. Waaruit blijkt dat? En om welke energiebronnen zou dat gaan?
- d. Hoeveel joule aan energie is er in Nederland in 2007 verbruikt?
- e. Hoeveel joule aan energie is er in Nederland in 2007 ingevoerd?
- f. Hoeveel joule aan energie is er als elektriciteit ingevoerd?
- g. Waarom was het vinden van aardgas in de Nederlandse bodem de afgelopen jaren zo belangrijk voor onze economie? Motiveer je antwoord met gegevens uit het stroomdiagram.
- h. Nederland kent ook opgeslagen energievoorraden. Je ziet dat in het schema op twee plaatsen. Waar?



▣ Opgave 38

Mooie diagrammen met gegevens over landen vind je via WWW.GAPMINDER.ORG.

Klik op 'Gapminder World' en je vindt zo'n diagram:



- Wat voor soort diagram betreft het hier?
- Welke twee variabelen zijn er tegen elkaar uitgezet?
- Welke betekenis heeft de grootte van elke 'punt'?
- Speel deze animatie af en beschrijf wat je ziet gebeuren.
- Welke twee landen stellen de grootste cirkels voor?
- Wat is in 2007 het armste land ter wereld? Hoe zit het met de levensverwachting in dat land? Hoe groot is de bevolking? Waar ligt dit land?
- Heeft in 2007 het rijkste land ook de hoogste levensverwachting? Om welk land gaat het? Kun je dit verklaren?
- Noem een land dat sinds 1900 (voor die tijd zijn de gegevens onbetrouwbaar) spectaculair is gestegen in levensverwachting. Is het gemiddelde inkomen ook zo spectaculair gegroeid?
- Je kunt ook naar het tabblad *Map*. Wat voor diagram krijg je dan? Is er op dit tabblad extra informatie beschikbaar of is de informatie alleen anders geordend?
- Klik op *Family size & length of life*. Speel de animatie af. Beschrijf wat je ziet gebeuren en vertel ook iets over de verschillende categorieën landen. Probeer in ieder geval het merkwaardige bewegen van het 'punt' dat China voorstelt te verklaren.

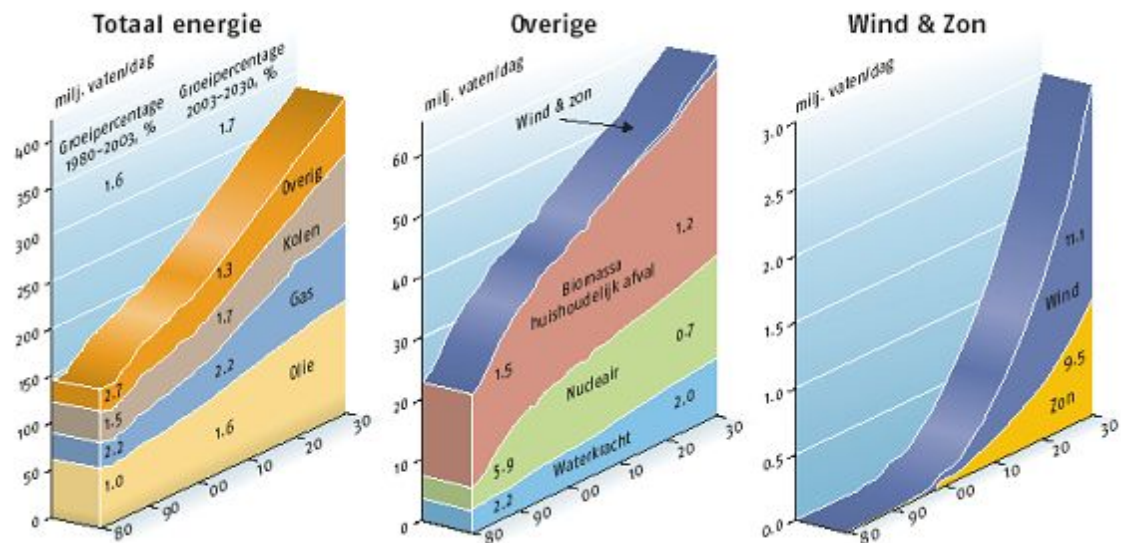


§ 1.4 Interpretaties

Verkennen

Opgave 39

De volgende diagrammen komen uit **EXXON MOBILE'S ENERGY OUTLOOK 2004**.



Het bijbehorende onderschrift luidt:

'In het eerste van de drie diagrammen is goed te zien dat de wereld voor zijn energievoorziening zeker de eerste 25 jaar grotendeels van fossiele brandstoffen afhankelijk blijft. Olie blijft de belangrijkste energiedrager, maar de productie van gas en steenkool groeit ook flink.

Het tweede diagram is een detaillering van het oranje vlak in het eerste (overige energiedragers). De groei van kernenergie stagneert enigszins doordat het maatschappelijk omstreden blijft.

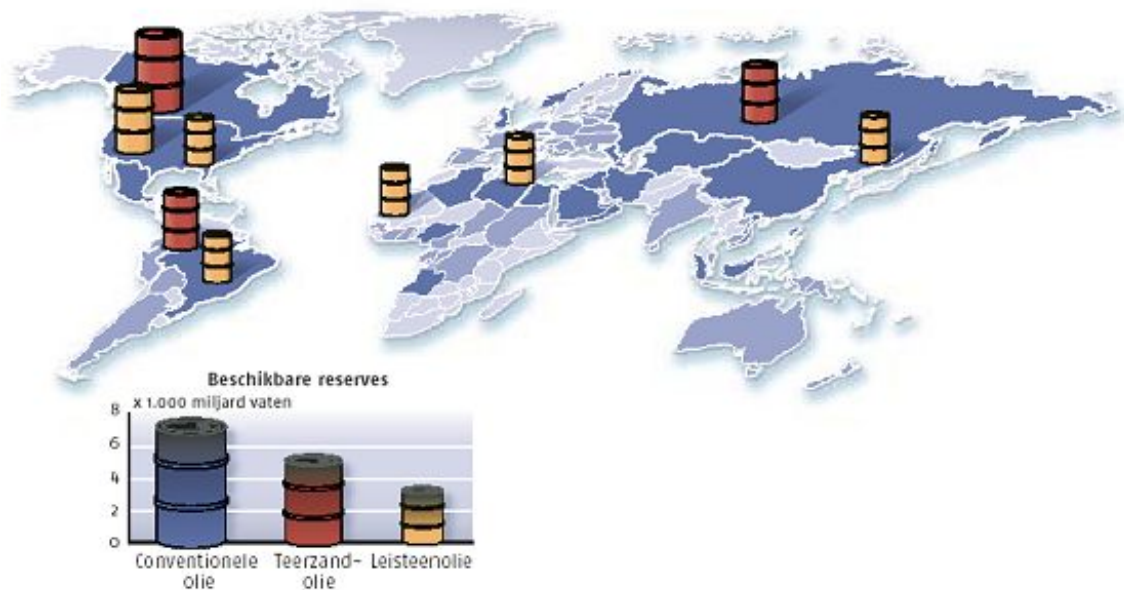
Voor de sector windmolens en zonnepanelen (derde diagram) is niet meer dan een bijrol weggelegd, ondanks de indrukwekkende groeicijfers.'

- Hoeveel procent van het energieverbruik in 2000 is gebaseerd op fossiele brandstoffen?
- Hoeveel procent van het energieverbruik is dat in 2030 volgens dit scenario?
- Ondersteunen deze diagrammen de uitspraak dat de eerste 25 jaar de energievoorziening grotendeels van fossiele brandstoffen afhankelijk blijft?
- In de derde grafiek lijkt het alsof er een enorme toename van het gebruik van wind- en zonne-energie zal plaatsvinden. Waarom is er toch volgens Exxon maar een bijrol voor die sector weggelegd?
- Waarom zie je dat de ontwikkeling van kernenergie volgens Exxon zal gaan stagneren?

Exxon is in zijn vooruitblik 2004 niet pessimistisch over het vinden van olie.

Men verwacht tot het einde van deze eeuw voldoende olie te kunnen winnen, alleen zal het dan niet alleen conventionele aardolie zijn.

Exxon's **OUTLOOK 2004** toont een diagram met de volgende beschikbare oliereserves en belangrijkste vindplaatsen.



Bijschrift bij deze figuur:

'Er is in ieder geval genoeg aardolie tot het einde van deze eeuw. De kleurschakering geeft aan hoe waarschijnlijk het is dat de getoonde reserves ook daadwerkelijk gewonnen kunnen worden: hoe donkerder, hoe waarschijnlijker. Op de wereldkaart zijn de vindplaatsen per land weergegeven: conventionele oliebronnen in blauw, teerzandolie in rood en leisteenolie in geel.'

- Waar denkt men bij Exxon kan er veel teerzandolie worden gewonnen?
- Op hoeveel miljard vaten schat men de voorraad teerzandolie?
- Wat bepaalt de beschikbare hoeveelheid reserves: de hoogte van een vat of de inhoud van een vat? Wat is er dus niet correct in deze figuur?

Theorie

Statistieken – en vooral diagrammen die statistieken presenteren – brengen meestal een bepaald verloop van een variabele tot uitdrukking. Diagrammen geven vaak meningen weer, waarbij grafische trucs een bepaald verloop soms overdrijven. Daarom is het erg belangrijk om kritisch te zijn op statistisch materiaal.

Let goed op:

- Keuze van de eenheden.
- Keuze van een nulpunt op een schaalverdeling.
- Manier van vergroten of verkleinen van figuren om een grotere of kleinere hoeveelheid aan te geven.
- Het feit dat het bij staafdiagrammen gaat om de oppervlakte van de staaf, zeker als je staven van verschillende breedte gebruikt.

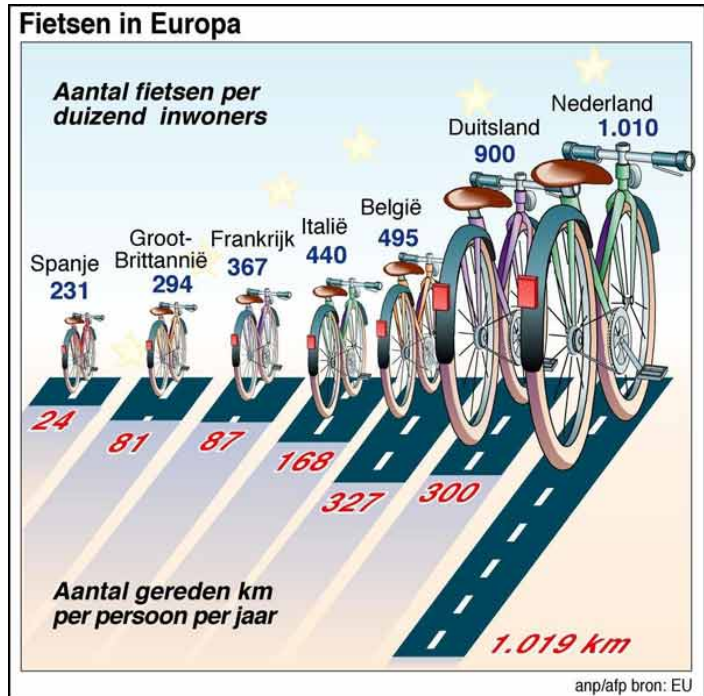
In de opgaven in deze paragraaf tref je een aantal situaties aan waarin dit soort zaken voor interpretatieproblemen kunnen zorgen.

Verwerken

Opgave 40

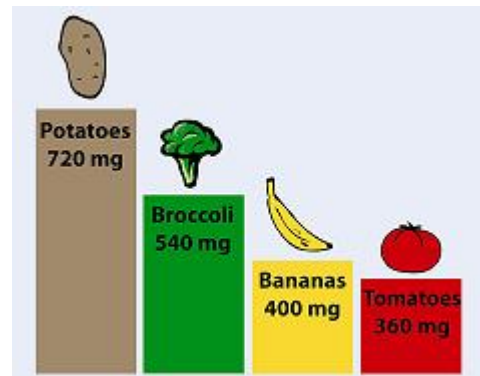
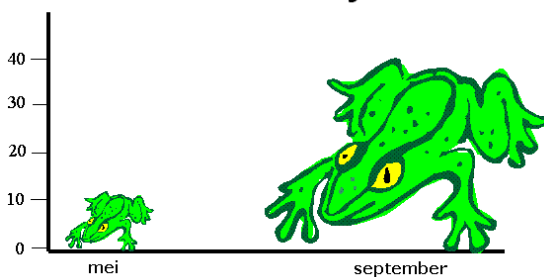
Wij Nederlanders doen nauwelijks iets anders dan fietsen. Kijk maar...

- Welk land heeft vermoedelijk de meeste fietsen?
- Welke informatie ontbreekt om dit met zekerheid te kunnen zeggen?
- Hoeveel kilometer fietsen we ongeveer per persoon in Nederland per jaar?
- Het onderste deel van de figuur is eigenlijk een staafdiagram. De lengte van de staaf past dan bij de grootte van de waarde die hij aangeeft. Klopt dat hier wel? Licht je antwoord toe met een voorbeeld.
- Waarom is zo bekeken het bovenste deel van de figuur ook een staafdiagram? Kloppen dan de hoogtes van de 'staven'?



Opgave 41

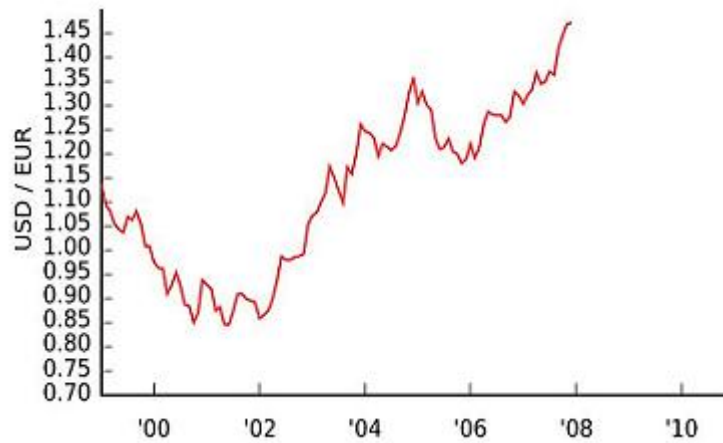
Aantal kikkers in de vijver



- De linker figuur laat zien hoe het aantal kikkers in de vijver toeneemt. Hoe heeft de maker van deze figuur gebruik gemaakt van overdrijving?
- In de rechter figuur leggen aardappelhandelaars uit dat er in aardappelen 720 milligram kalium per portie zit. Hoeveel is dat bij broccoli, bij bananen en bij tomaten? Waarom is het rechter staafdiagram misleidend?

Opgave 42

Deze figuur geeft de koers weer van de Amerikaanse dollar (USD) in de loop van de jaren. Er zijn behoorlijke schommelingen, maar de figuur overdrijft dat toch wel.

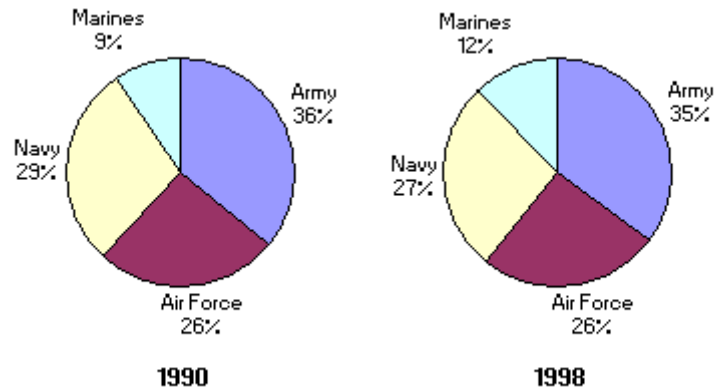


- Wat wordt bedoeld met USD/EUR?
- Hoeveel moest een Europeaan voor een dollar betalen toen de dollar op zijn dieptepunt was?
- Leg uit hoe de overdrijving in deze figuur ontstaat. (Er zijn twee oorzaken.)

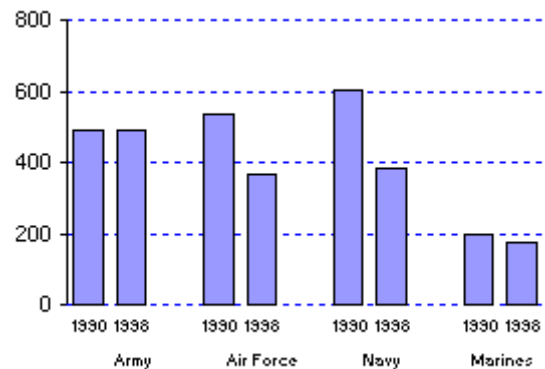


Opgave 43

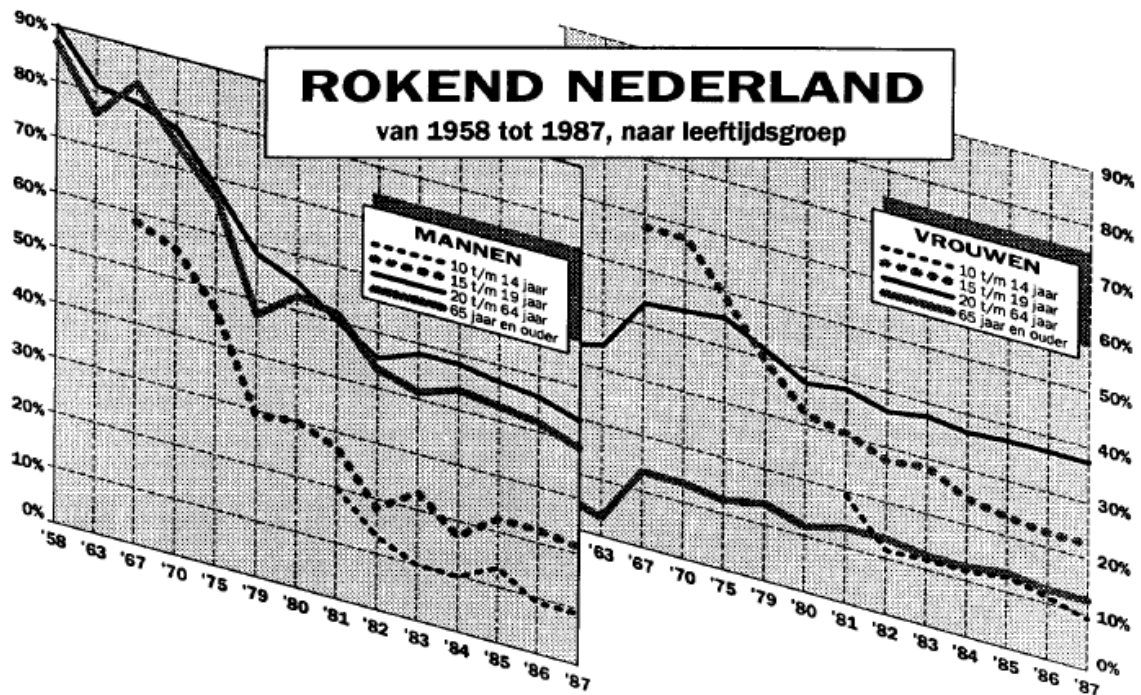
Deze twee cirkeldiagrammen geven de uitgaven weer aan het militaire apparaat in de Verenigde Staten (VS) in twee verschillende jaren.



- Kun je op grond van deze figuur vaststellen dat de *marines* (mariniers) in 1998 een groter budget hadden dan in 1990?
- Welke informatie ontbreekt in deze figuur?
- In dit staafdiagram zie je de werkelijke uitgaven in miljoenen dollars. Schat met hoeveel het totale budget voor het militaire apparaat in de VS in deze jaren is afgenomen.
- Wat betekent dit voor het cirkeldiagram voor 1998?



Opgave 44



Deze lijndiagrammen komen uit een krantenartikel uit 1988.

Volgens de linker grafiek rookte in 1958 nog 90 procent van de mannen in de leeftijdsgroep van 20 tot 65 jaar. In 1987 was dit percentage gedaald tot 43 procent.

Deze sterke daling wordt door de tekenaar op een misleidende wijze benadrukt.

- Wat veroorzaakt deze misleiding?
- Bekijk het diagram van de mannen van 15 tot 20 jaar. De grafiek ziet er ook voor de jaren 1982 tot 1987 dalend uit. Daalt het percentage rokers van die categorie ook werkelijk?
- Bij welke van deze acht diagrammen is er vrijwel nooit van daling sprake?

In het krantenartikel staat:

'Een overzicht van de rookgewoonten in Nederland in 1987 gaf, net als in de jaren daarvoor, opnieuw een daling te zien van het aantal rokers in ons land. Hoewel de betrekkelijk snelle daling in de jaren zeventig en het begin van de jaren tachtig is afgenomen, heeft die tendens zich de afgelopen drie jaar gestabiliseerd op een daling van 1 procent per jaar. Kon in 1958 worden becijferd dat 60 procent van de Nederlandse mannen en vrouwen in de leeftijdsgroep van 15 tot 65 jaar rookte, volgens cijfers van de Stichting Volksgezondheid en Roken was dat in 1987 afgenomen tot 37 procent.'

Een lezer van dit artikel denkt dat die 37 procent niet kan kloppen. Hij redeneert zo:

- De laatste drie jaar is er een daling van 1 procent.
 - Volgens de tekst en de figuur was de daling in de periode daarvoor nog sterker.
 - In 1958 is het percentage rokers 60.
 - In de 29 jaar van de periode 1958-1987 is daar zeker 29×1 procent = 29 procent van afgegaan, dus in 1987 moet het percentage minder dan 31 procent zijn.
- Leg uit waarom het percentage van 37 procent wel correct kan zijn als je de 1 procent daling per jaar goed interpreteert.

§ 1.5 Overzicht

Je hebt alle theorie van het onderwerp **KIJKEN NAAR DATA** doorgewerkt. Het is tijd om een overzicht over het geheel te krijgen.

Begrippen

- § 1.1 Statistiek – populatie – elementen – variabele – data.
- § 1.2 Meetbaar – kwalitatief – kwantitatief.
- § 1.3 Beelddiagram – staafdiagram – lijndiagram – cirkeldiagram – kaartdiagram – puntenwolk.
- § 1.4 Interpretatieproblemen.

Vaardigheden

- § 1.1 Populatie en objecten van een statistisch onderzoek herkennen – te onderzoeken variabele herkennen – soorten data herkennen.
- § 1.2 Verschillende soorten variabelen en hun data herkennen.
- § 1.3 Verschillende soorten diagrammen herkennen en op grond daarvan de wijze van aflezen eruit herkennen – een aantal basisdiagrammen zelf kunnen tekenen.
- § 1.4 Kritisch kijken naar statistisch materiaal.

Samenvatten

Opgave 45

Maak een samenvatting van dit hoofdstuk door bij elk van de genoemde *begrippen* een omschrijving of een voorbeeld(opgave) te geven en bij elk van de genoemde *vaardigheden* een voorbeeld te geven.



Terugblikken

Opgave 46

Hieronder zie je de kruistabel voor lengte en gewicht die de wiskundigen Freudenthal en Sittig maakten van 5001 vrouwen in het kader van hun onderzoek naar een beter systeem voor kledingmaten van vrouwen in opdracht van warenhuis De Bijenkorf uit 1947.

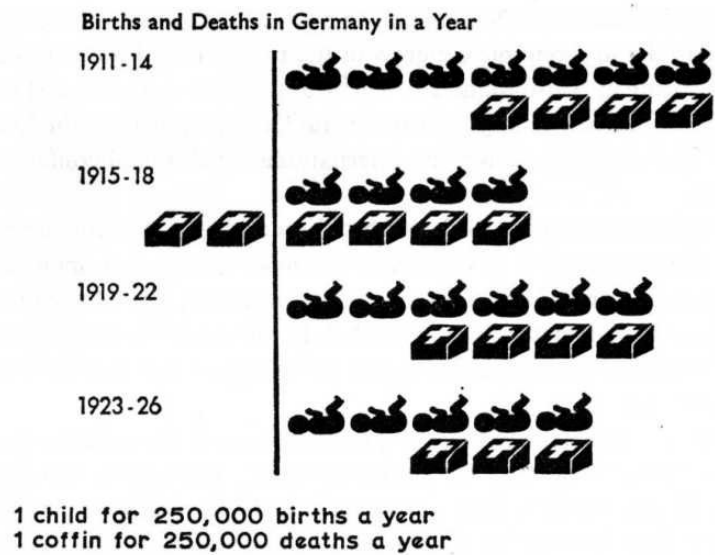
		Statistisch onderzoek 1947																											
		De Bijenkorf - Amsterdam																											
		5001 vrouwen met leeftijd > 18 jr.																											
		gewicht in kg																											
		42	45	48	51	54	57	60	63	66	69	72	75	78	81	84	87	90	93	96	99	102	105	108	111	114			
lengte in cm	186																										1	1	
	183									1				1	2	2													6
	180									1	1	5	4	2	3	2													21
	177				1	2	2	4	13	19	15	10	9	4	4	4	1	1				4							94
	174				1	1	6	15	17	22	22	20	13	12	5	3	7	4	2	2	1	1	1						156
	171				2	7	25	41	50	68	34	32	22	25	18	8	9	8	9	1						1	1	1	362
	168			1	13	24	39	59	74	73	55	43	35	22	40	20	20	20	5	8	4	3	2	1	1	1	1	563	
	165			1	3	20	52	109	108	92	85	85	64	55	48	44	28	19	10	8	6	6	2						846
	162			2	10	42	92	103	127	100	95	90	59	56	47	28	38	13	11	9	4	3	3	1		1			924
	159		1	3	22	58	79	80	83	96	89	82	59	53	37	32	23	12	10	6	4	3	2						834
	156		1	7	16	35	45	62	57	70	58	51	59	38	28	25	9	10	6	6		1	2						586
	153		1	2	18	24	22	39	42	44	33	32	20	21	9	10	7	9	2	2	1	2			1	1			342
	150			7	11	15	12	19	19	20	22	19	8	13	4	2	1	3	1										176
	147		3	5	3	4	10	8	6	9	4	4	6	2		1	1												67
	144				2	1	1		2	3	1	1	1	1	1														14
	141		1				3	1						1	1														8
	138									1																			1
		7	27	86	216	350	493	564	591	560	495	386	322	242	213	142	103	76	48	27	24	13	7	3	3	3	5001		

Bekijk de tabel goed.

- Over welke populatie gaat deze tabel?
- Iemands lengte is een kwantitatieve variabele die is afgerond. Welk interval is afgerond op 150 centimeter?
- Hoe heeft de afronding van de gewichten plaatsgevonden? Geef ook hiervan een voorbeeld.
- Teken een staafdiagram van de lengtes van de 5001 vrouwen. Gebruik daarbij relatieve frequenties. Teken ook een lijndiagram van deze lengtes.
- Hoeveel procent van de vrouwen is kleiner dan 151,5 centimeter?
- Kun je ook schatten hoeveel procent van de vrouwen kleiner is dan 153 centimeter?
- Bereken hoeveel procent van de vrouwen groter is dan 172,5 centimeter, maar lichter dan 64,5 kilogram.
- Schat hoeveel procent van de vrouwen groter is dan 175 centimeter, maar lichter dan 80 kilogram. Licht toe hoe je te werk gaat.

Opgave 47

Het beelddiagram hieronder geeft informatie over de geboorte en de sterfte in het Duitsland in de jaren 1911-1926.

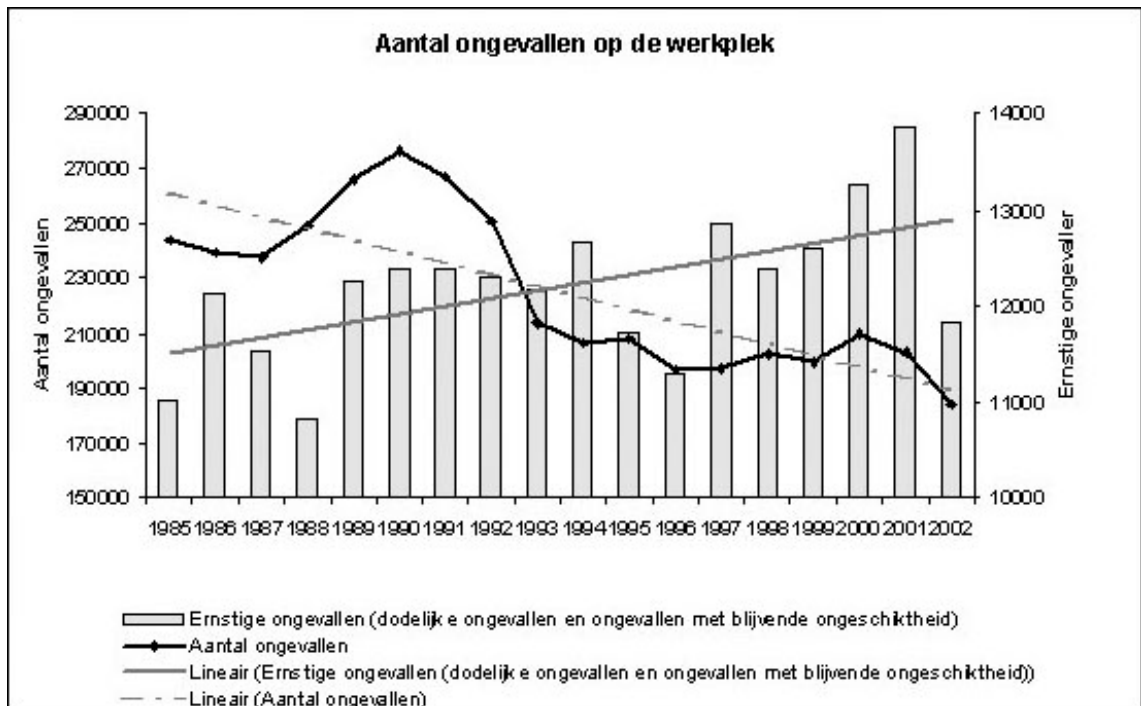


- Waaraan herken je de periode van de Eerste Wereldoorlog?
- Hoeveel bedroeg het geboorteoverschot in de jaren voor de Eerste Wereldoorlog?
- Hoeveel bedroeg het geboorteoverschot in de jaren na de Eerste Wereldoorlog?
Heb je een verklaring voor het verschil?



Opgave 48

Je ziet hieronder een overzicht van het aantal ongevallen op de werkplek in Vlaanderen in de periode 1985-2002.



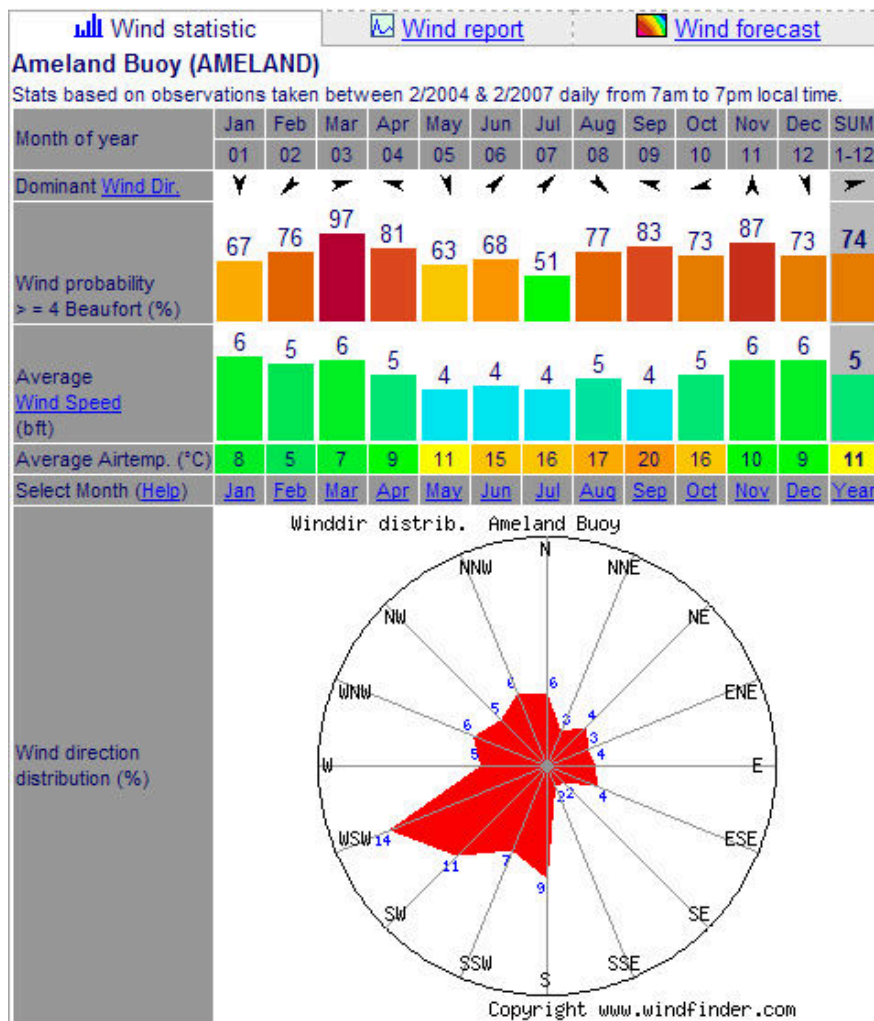
- Over welke populatie gaat dit statistisch onderzoek? En over welke twee statistische variabelen?
- Wat valt je op aan het staafdiagram?
- Welke schaalverdeling hoort bij *Aantal ongevallen*?
- Hoeveel procent van het aantal ongevallen op de werkplek in 2000 waren ernstig? Geef de berekening.
- Hoeveel bedroeg het aantal ongevallen in de periode 1985-2002 op zijn hoogst?
- De twee getekende lijnen zijn bedoeld om een bepaalde trend weer te geven. Beschrijf de trend voor wat betreft het totale aantal ongevallen op de werkplek. Doe hetzelfde voor het aantal ernstige ongevallen.
- Wat betekent dit voor het aantal ernstige ongevallen als percentage van het totale aantal ongevallen?



Opgave 49

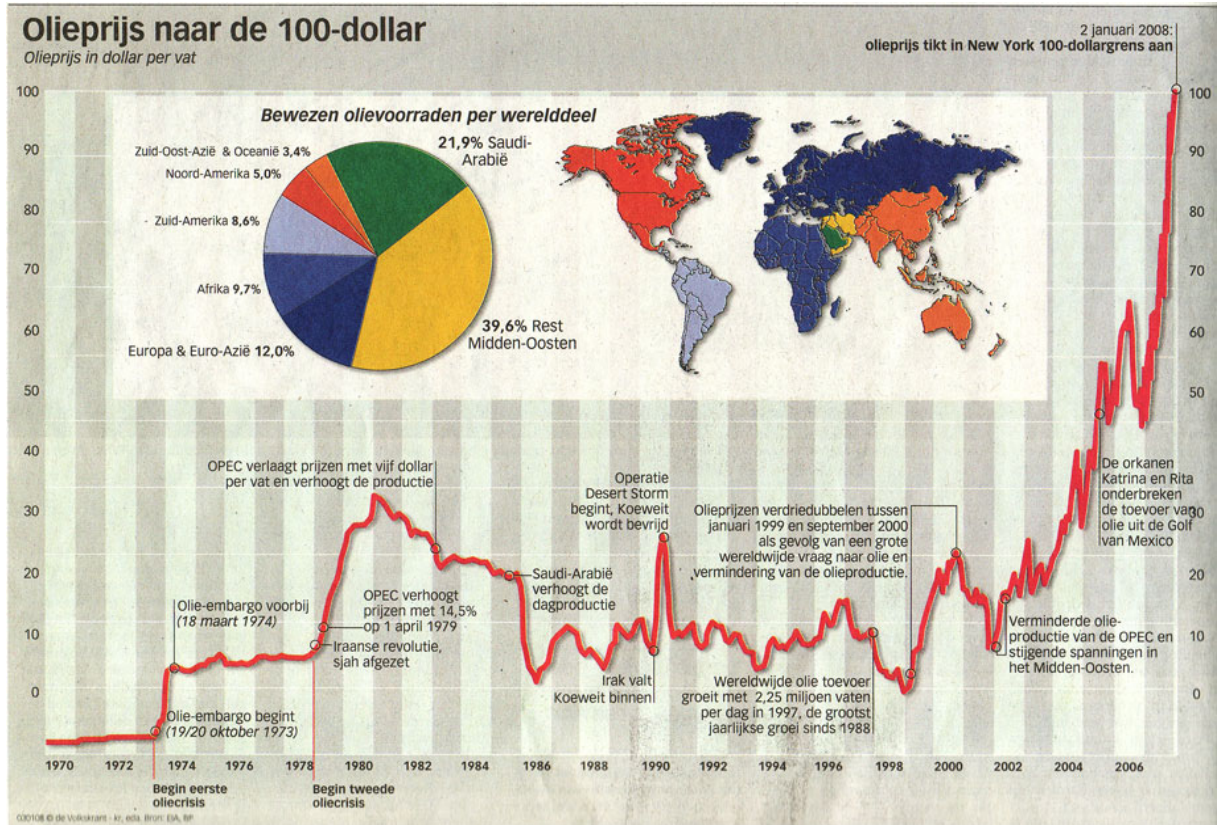
Hieronder zie je een overzicht van de windgegevens op Ameland in de loop van het jaar. De windkracht wordt gemeten in Beaufort (Bft); 4 Bft komt overeen met matige wind.

- Over welke statistische variabelen gaat dit overzicht?
- Uit welke windrichting komt de wind op Ameland meestal?
- Met betrekking tot de wind op Ameland in juli kun je drie gegevens uit deze figuur aflezen. Welke drie?
- In welke maand waait het op Ameland het vaakst harder dan 4 Bft? Hoe hard waait het dan gemiddeld?
- De onderste figuur is geen cirkeldiagram, maar eigenlijk meer een lijndiagram. Licht dat toe.
- Hoe zou dit diagram er uit zien wanneer het als een normaal lijndiagram zou zijn getekend? Welk nadeel heeft dat?



Opgave 50

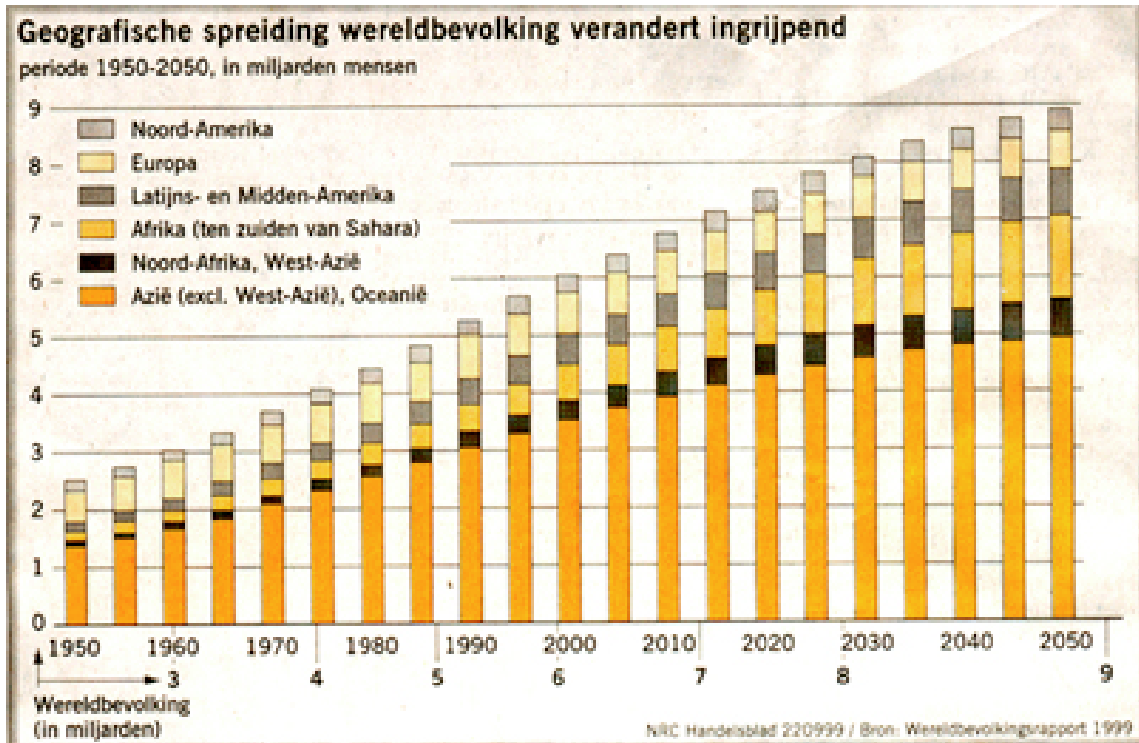
In DE VOLKSKRANT van 2 januari 2008 staat onderstaande infographic.



- Welke soorten diagrammen tref je aan in deze infographic?
- Waarom is het kaartje eigenlijk geen cartogram maar een legenda?
- Hoeveel procent van de totale bewezen olievoorraad zit in het Midden-Oosten? Waarom wordt er gesproken over de 'bewezen' olievoorraden?
- Let eens op de verticale schaal van het lijndiagram. Wat is daar fout gegaan?
- Wat was er aan de hand tijdens de eerste en de tweede oliecrisis? Welke gevolgen had dit?
- Wat probeert de krant met dit diagram weer te geven?
Verzin een andere kop voor het bijbehorende artikel dan *Olieprijs naar de 100 dollar*.

▣ Opgave 51

Onderstaande infographic over de geografische spreiding van de wereldbevolking staat op 22 september 1999 in **NRC/HANDELSBLAD**.



- Om wat voor soort diagram gaat het hier?
- Licht de kop van deze infographic toe.
- Hoeveel procent groeit de bevolking van Noord-Amerika volgens deze voorspelling?
Hoe zit het met Europa?
- Van welk werelddeel wordt relatief gezien de grootste groei verwacht?
- In 1950 woonde ongeveer de halve wereldbevolking in Azië.
Hoe zit dat volgens deze voorspelling in 2050?



Onderzoeken

Om echt de *ins* en *outs* van het bedrijven van statistiek te leren moet je zelf een keer een statistisch onderzoek uitvoeren. Dat ga je ook daadwerkelijk doen als je wat meer kennis over statistiek hebt opgebouwd. Nu bereid je jezelf daarop voor.

Opdracht

Beantwoord de volgende vragen:

- Welk onderwerp zou je kiezen voor een statistisch onderzoek? Met andere woorden: welke vraag zou je met behulp van statistisch onderzoek willen beantwoorden?
- Welke populatie zou je kiezen om dit onderzoek op uit te voeren?
- Hoe zou je het onderzoek willen inrichten?

Houd hierbij rekening met je mogelijkheden om ook echt onderzoek te doen, te meten, een enquête af te nemen, etc.

Hieronder krijg je een paar voorbeelden aangereikt.

De eerste mogelijkheid zul je in de komende hoofdstukken af en toe tegenkomen.

Voorbeeld 1

Hoofdvraag

Hebben havo-leerlingen in het N&G-profiel die wiskunde B hebben gekozen betere cijfers dan havo-leerlingen in het N&G-profiel die wiskunde A hebben gekozen?

Populatie

Alle havo-leerlingen met een N&G-profiel.

Onderzoeksmethode

Het gemiddelde eindcijfer voor het havo-examen wordt vergeleken met de gekozen wiskundesoot. Ook worden de cijfers voor natuurkunde, scheikunde en biologie vergeleken met de gekozen wiskundesoot.

Voorbeeld 2

Hoofdvraag

Je school gaat alle leerlingen voorzien van sportkleding en koopt daarom shirts en sportbroeken. Daarvoor moet je een bestaand maatsysteem voor kleding (van een bepaalde fabrikant?) gebruiken om vast te stellen hoeveel shirts en broeken je van een bepaalde maat moet inkopen. Hoeveel koop je van elke maat in?

Populatie

Alle leerlingen op jouw school.

Onderzoeksmethode

Je meet een goed gekozen steekproef van de leerlingen van je school op. Belangrijk is om vooraf vast te stellen welke maten van belang zijn voor de kledingmaat. Vervolgens stel je vast hoeveel procent van de gemeten personen een bepaalde maat broek of shirt heeft. Daarna bereken je hoeveel leerlingen dit jaarlijks op jouw school betreft.

Voorbeeld 3

Hoofdvraag

Bij de lessen lichamelijke opvoeding (of bewegingsonderwijs) doe je regelmatig de Coopertest en/of de *shuttle run test* (piepjestest). Doen sporters deze tests structureel beter dan niet-sporters?

Populatie

Alle leerlingen in het vierde leerjaar.

Onderzoeksmethode

Eerst stel je vast wat je onder een 'sporter' verstaat en vervolgens ga je de resultaten van Coopertest en/of *shuttle run test* vergelijken met het aantal uren dat iemand aan zijn sport besteedt.

