

1.1 Data ordenen

Inleiding

Bij statistiek draait het allemaal om onderzoek. Het doel ervan is uitspraken doen; deze uitspraken moeten onderbouwd worden met cijfers, diagrammen en berekeningen. Je kunt bij een onderzoek nooit alle Nederlanders of alle jongeren onderzoeken. Je moet bij een onderzoek een goede selectie uit alle Nederlanders of uit alle jongeren bekijken. Hoe doe je dat? Het woord statistiek is afkomstig uit het Latijn: 'statisticum collegium'. Dit betekent: les over staatszaken. Je zou kunnen zeggen de analyse van staatsgegevens.



Figuur 1

Je leert in dit onderwerp

- statistische variabelen herkennen en ordenen;
- beoordelen of er conclusies getrokken kunnen/mogen worden.

Voorkennis

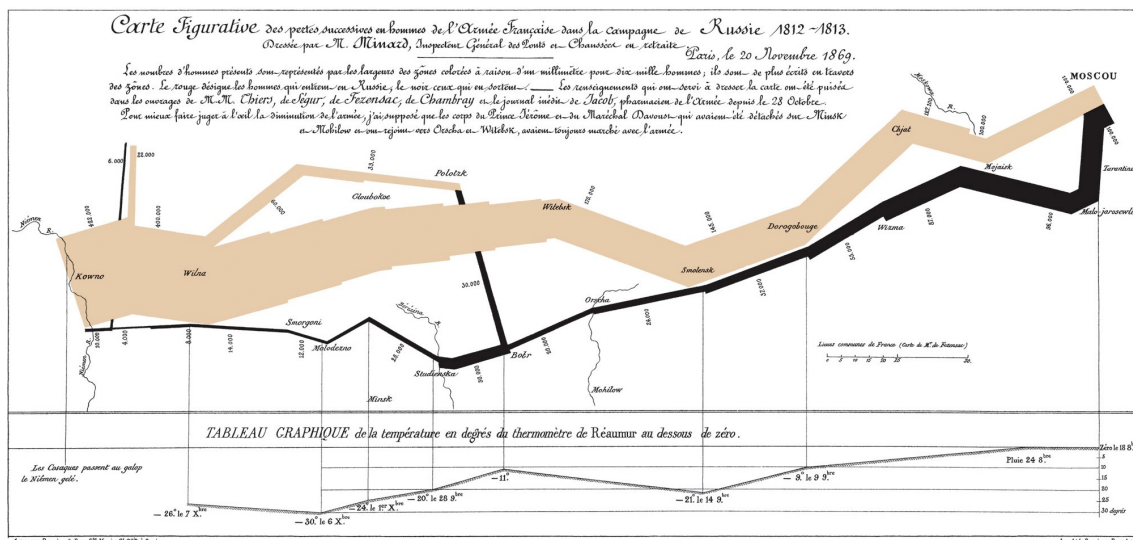
- rekenen met getallen, ook met procenten;
- tabellen en diagrammen interpreteren.

Verkennen

Opgave V1

In 1812 trok het Franse leger onder aanvoering van de Franse keizer Napoleon ten strijde tegen Rusland. Deze campagne begon aan de rivier de Niemen en leidde naar Moskou. Dit 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.

De kaart kun je ook bekijken op wikipedia.



Figuur 2

Je ziet dat het Franse leger met 422000 man uit Polen vertrok richting Moskou. Het aantal mensen waaruit het leger bestond is een statistische variabele.

- a Hoeveel mensen van dit leger kwamen aan in Moskou?

- b Hoeveel mensen kwamen weer bij het startpunt terug?
- c Hoeveel procent van dit Franse leger is tijdens de Russische veldtocht overleden?
- d Op de terugweg zijn ook af en toe de dagtemperaturen gemeten. Welke laagste dagtemperatuur is er gemeten?
- e De temperatuur is aangegeven in graden Réaumur ($^{\circ}\text{Re}$). Nu is $0\text{ }^{\circ}\text{C}$ hetzelfde als 0°Re en $100\text{ }^{\circ}\text{C}$ hetzelfde als 80°Re . Hoeveel graden was de laagste dagtemperatuur in $^{\circ}\text{C}$?

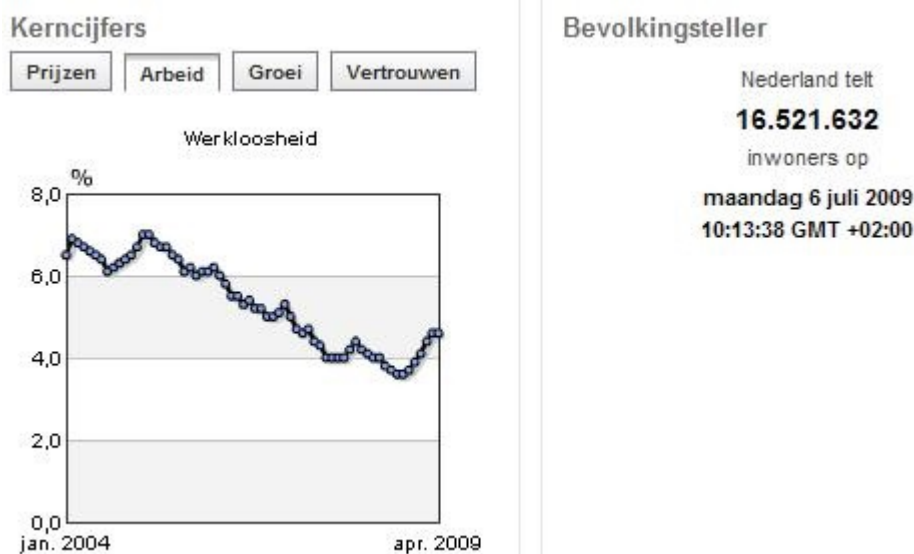
Opgave V2

Bekijk de figuur van Minard in de vorige opgave. Je kunt erin 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.

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

Opgave V3

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



Figuur 3

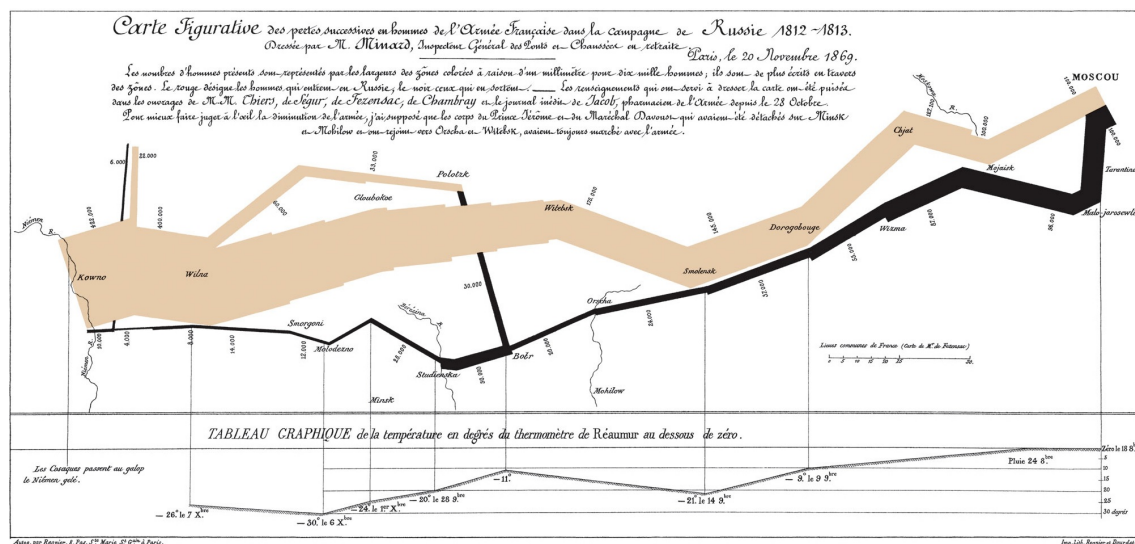
- a Hoeveel mensen telde Nederland op dat moment? En hoeveel nu?
- b Het CBS spreekt in haar toelichting over een schatting van het aantal inwoners. Licht dit toe.
- c Wat lees je af onder Kerncijfers Arbeid?
- d Het percentage werklozen is begin 2009 aan het stijgen. Kun je op grond van alleen deze gegevens zeggen dat ook het aantal werklozen in die periode stijgt? Waarom is dit wel erg waarschijnlijk?

Uitleg

Het woord statistiek ontstond in de tijd van Napoleon, toen het bestuur van de verschillende Europese staten (in de achttiende eeuw) zo ingewikkeld werd dat een nauwkeurige kennis van staatszaken belangrijk werd.

Tegenwoordig moet je dat wat ruimer opvatten: ook bedrijven, onderzoeksinstituten en particulieren buiten de overheid verzamelen en ordenen 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 (Centraal Bureau voor de Statistiek)** vind je hier veel 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 was de onderzochte populatie. Een van de onderwerpen van het onderzoek was het aantal soldaten afhankelijk van de vordering van de tocht. Dit aantal soldaten was 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.



Figuur 4

Bij het ordenen en weergeven van de data heeft de onderzoeker meestal een bepaald doel. Minard wilde 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 1

Bekijk in de **Uitleg** 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?

Opgave 2

Open het **bestand met data van de Nederlandse bevolking** zoals dat in juli 2009 te vinden was 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?

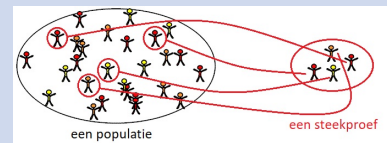
- c Wat zijn de elementen van de populatie als het gaat over 'bevolking'? En wat zijn de elementen als het gaat over 'huishoudens'?
- d Over welke variabelen gaat het bij 'bevolking naar geslacht'?
- e Over welke variabelen gaat het bij 'particuliere huishoudens'?
- f Wat stelt de 'bevolkingsdichtheid' voor? Wat betekent de toename van de bevolkingsdichtheid?
- g Hoe komt het dat het totale aantal Nederlanders gedeeld door de bevolkingsdichtheid niet voor elk jaar hetzelfde getal is? Geef minstens twee verklaringen.

Theorie en voorbeelden

Om te onthouden

De wetenschap, de methodiek en de techniek van het verzamelen, bewerken, interpreteren en presenteren van gegevens heet **beschrijvende statistiek**. Een statistisch onderzoek kent diverse fasen:

- Je begint een statistisch onderzoek met een **onderzoeksvraag**.
- Je stelt vast over welke **populatie** (doelgroep) het onderzoek gaat. Bij een grote doelgroep neem je een deel van de groep, dit heet een **steekproef**. Deze steekproef moet **aselect** (volstrekt willekeurig samengesteld) en **representatief** (kenmerkend) zijn voor de populatie.
- Je stelt vast op welke **statistische variabele** het onderzoek betrekking heeft.
- Je verzamelt de bij die variabele passende **data** (de statistische gegevens). Deze data moet je meestal eerst **ordenen**. De verzamelde data geef je weer in de vorm van tabellen en diagrammen: tabellen zijn nauwkeuriger, diagrammen vaak overzichtelijker.
- Je probeert een (eerlijke) **conclusie** te trekken.



Figuur 5

Voorbeeld 1

Statistiek is het verzamelen, bewerken, interpreteren en presenteren van gegevens. Dat is waar onder andere het CBS zich mee bezig houdt.

Bekijk de tabel 'Bevolking; kerncijfers'.

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	184069
	Overledenen	aantal	75929	87825	109619	114279	128824	140527	134996
	Geborteoverschot	aantal	153789	151303	129293	67015	69141	66092	49673
	Geborteoverschot, 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

Tabel 1

- Welke mogelijke onderzoeksvraag kun bij deze tabel opstellen?
- Stel vast over welke populatie (doelgroep) het onderzoek gaat.
- Stel vast op welke statistische variabelen het onderzoek betrekking heeft.
- De data zijn al verzameld en geordend in een tabel door het CBS. Noem een voordeel van het ordenen van data in een tabel en een voordeel van het ordenen van data in een diagram.
- Waarover kunnen er aan de hand van deze gegevens conclusies getrokken worden?
- Welke conclusie kan er getrokken worden over de bevolkingsdichtheid?

Antwoord

- Een mogelijke onderzoeksvraag is:
Hoe is het aantal Nederlanders in de loop van de jaren veranderd?
- De populatie is:
De Nederlandse bevolking.

- De variabelen zijn:
totale bevolking, aantal mannen, aantal vrouwen, totale bevolkingsgroei, aantal levendgeborenen, aantal overledenen, geboorteoverschot en bevolkingsdichtheid
Deze variabelen variëren met de tijd.
- Tabellen zijn nauwkeuriger, diagrammen vaak overzichtelijker.
- Er kunnen conclusies getrokken worden over:
totale bevolking, aantal mannen, aantal vrouwen, totale bevolkingsgroei, aantal levendgeborenen, aantal overledenen, geboorteoverschot en bevolkingsdichtheid.
- Een juiste conclusie over de bevolkingsdichtheid is:
De bevolkingsdichtheid is in de loop van de tijd toegenomen, dat betekent dat er steeds meer Nederlanders op één vierkante kilometer wonen.

Opgave 3

Bekijk **Voorbeeld 1**. Het CBS verzamelt dergelijke gegevens door het bevragen van de gemeenten in Nederland.

- a Licht toe hoe je de waarden van *totale bevolkingsgroei, relatief* kunt berekenen uit de *totale bevolking* en de *totale bevolkingsgroei*.
- b Laat met een berekening zien dat de waarde *totale bevolkingsgroei, relatief* voor 2008 klopt.
- c Welke uitspraak zou je op grond van deze tabel kunnen doen over de variabele *totale bevolkingsgroei, relatief*?

Opgave 4

In **Voorbeeld 1** komt ook de variabele *bevolkingsdichtheid* voor.

- a Hoe berekent het CBS de bevolkingsdichtheid?
- b Klopt het dat tussen 1950 en 2000 de oppervlakte van Nederland groter is geworden? Reken na hoeveel km² ongeveer.

Opgave 5

Bekijk nog eens de variabele *bevolkingsgroei*.

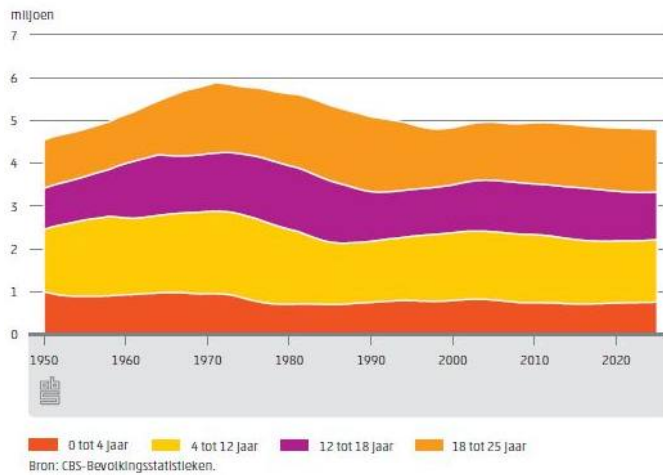
- a Waaraan zie je dat deze variabele niet alleen kan worden bepaald uit *totale bevolking* en *geboorteoverschot*?
- b Waar moet je om de bevolkingsgroei te berekenen ook nog rekening mee houden?
- c In welke periode was het geboorteoverschot groter dan de bevolkingsgroei?
- d Wat betekent dit?

Verwerken

Opgave 6

Het CBS doet in Nederland veel onderzoek en rapporteert hierover. Deze diagrammen komen uit de Jongerenmonitor 2013.

1.1.2 Aantal jongeren (0 tot 25 jaar) naar leeftijdsgroep, 1950-2025

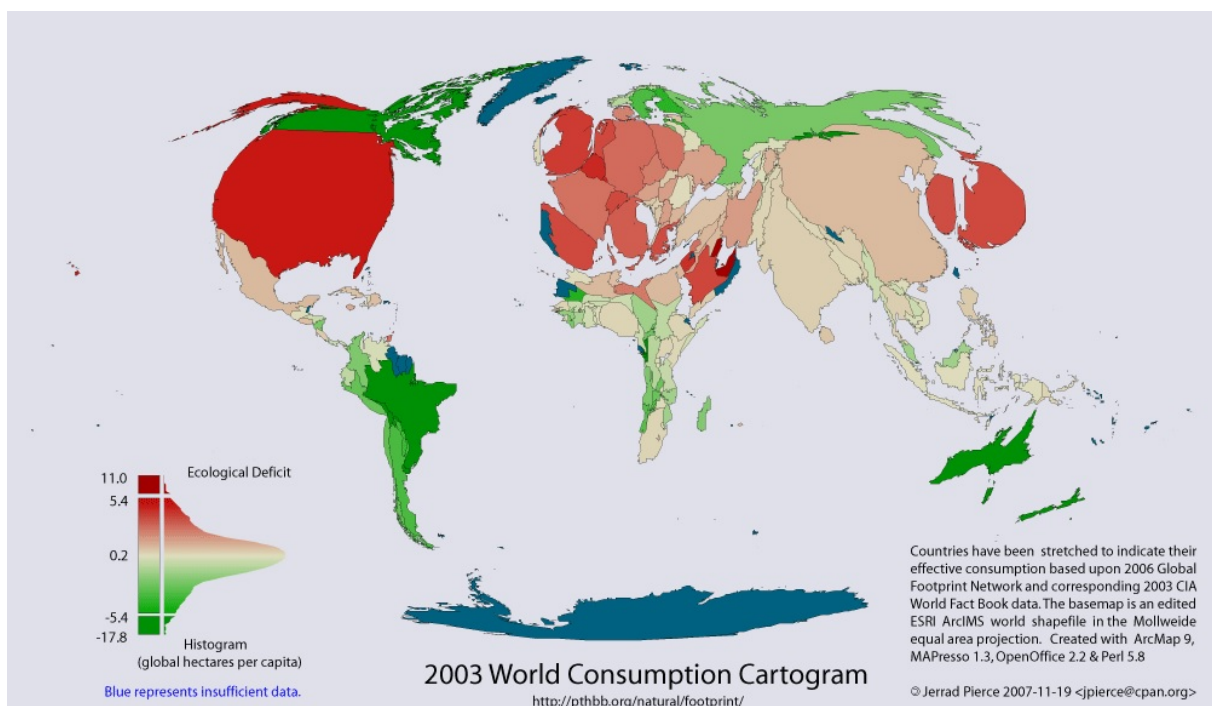


Figuur 6

- Wat is de populatie bij dit onderzoek van het CBS geweest?
- Wat is de definitie van jongeren in dit onderzoek geweest, denk je?
- Welke deelgroepen zijn onderscheiden?
- Welke variabele is in dit diagram weergegeven?

Opgave 7

De ecologische voetafdruk is een getal dat weergeeft welk deel van het aardoppervlak door een bepaalde bevolkingsgroep gebruikt wordt om van te leven. Dit getal wordt uitgedrukt in mondiale hectares.



Figuur 7

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

- a Wat betekent het dat Nederland aan de donkere kant van het rood zit?
- b Waarom zijn bijvoorbeeld Groenland en Portugal blauw gekleurd?
- c Griekenland heeft een vergelijkbare ecologische voetafdruk als Nederland terwijl de welvaart er lager is. Geef daarvoor een verklaring.
- d Welke data heb je nodig om deze wereldkaart te kunnen maken?

Opgave 8

In de tabel zie je het aantal kilometer fietspad op 23 januari 2013.

Provincie	lengte fietspad (km)	fietspad/opp (m/km ²)	fietskm per dag / km fietspad
Groningen	1600	700	882
Friesland	2000	599	848
Drenthe	2100	791	522
Overijssel	2800	837	1193
Flevoland	1500	1041	487
Noord-Holland	4200	1563	1601
Zuid-Holland	4800	1696	1583
Utrecht	1800	1306	1809
Noord-Brabant	5000	1019	1145
Gelderland	5000	997	982
Limburg	2000	918	1064
Zeeland	1400	764	660
Nederland	34600	1009	1157

Tabel 2

- a Welke data heb je nodig om de tweede kolom in deze tabel te kunnen maken?
- b Welke data heb je nodig om de derde kolom in deze tabel te kunnen maken?
- c Waarom zou de derde kolom gemaakt zijn?
- d Welke extra data zijn er nodig geweest om de vierde kolom te kunnen maken?
- e Wat is de betekenis van de data in de vierde kolom?
- f Drenthe profileert zich graag als fietsprovincie. Welke argumenten kan men daarvoor aandragen op basis van gegevens uit deze tabel?

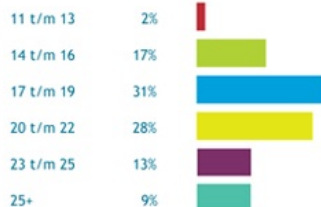
Opgave 9

Aan 44000 jongeren is een aantal vragen gesteld. De resultaten staan in de diagrammen.

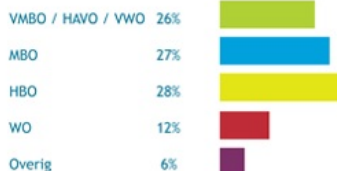
Geslacht *



Leeftijd *

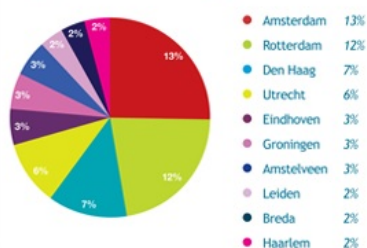


Opleidingsniveau *



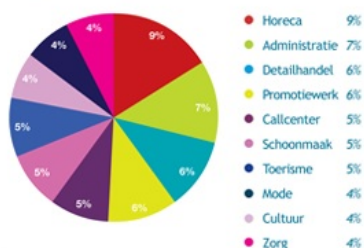
Kandidaten per regio *

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.

Figuur 8

- Welke data heb je nodig van de deelnemers om dit overzicht te kunnen maken?
- Voor wie zou dit overzicht gemaakt zijn?
- Maak een beschrijving van de deelnemers die zich ingeschreven hebben.
- Kun je nu zeggen dat 9% van alle mensen een bijbaan in de horeca heeft?

Opgave 10

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

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 Neder- land	in het buiten- land
	<i>x 1 000</i>	%		<i>x 1 000</i>		
Woonprovincie						
Groningen	550	70	1,82	700	48	52
Friesland	620	71	2,10	930	45	55
Drenthe	470	68	1,88	600	41	39
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.

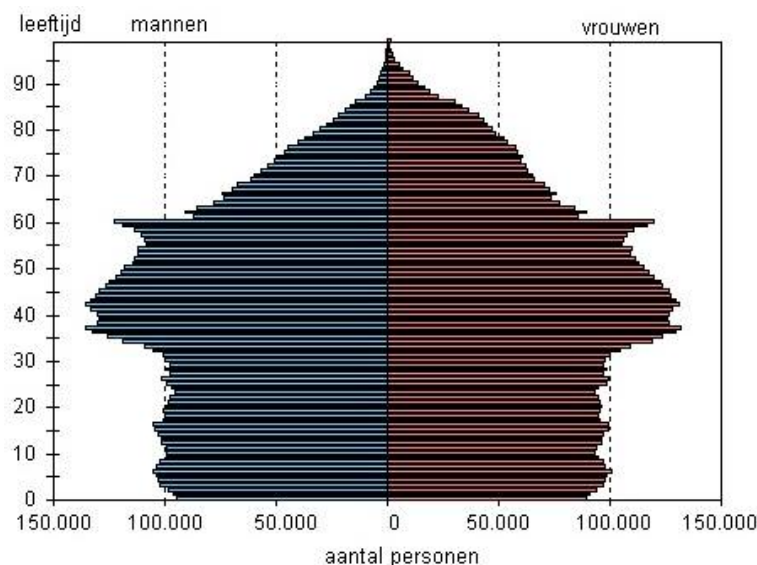
Tabel 3

- Welke data heb je nodig om dit overzicht te kunnen maken?
- Voor wie is deze informatie belangrijk?
- Hoe komen de data in de kolom 'gemiddeld aantal lange vakanties per participant' tot stand? Hoe kun je controleren of de getallen in de tabel kloppen?
- Een reisbureau concludeert dat mensen uit stedelijke gebieden meer behoefte aan lange vakanties hebben. Wat vind je van die conclusie?

Toepassen

Opgave 11: Demografische druk

Open het [bestand met data van de Nederlandse bevolking](#). Bekijk het tabblad 'Beschrijving'. Bekijk ook het leeftijdsdiagram uit 2008 voor de Nederlandse bevolking.



Figuur 9

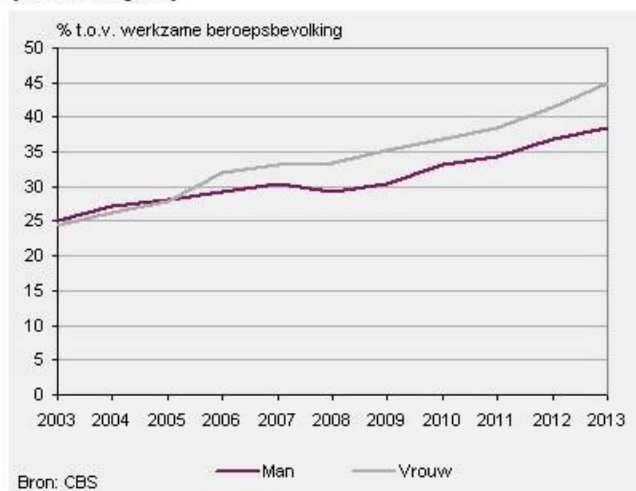
- Wat wordt verstaan onder 'demografische druk'?
- Laat zien hoe de demografische druk wordt berekend met de rest van de tabel.
- De 'grijze druk' neemt toe, de 'groene druk' neemt af. Wat betekent dit?
- Hoe kun je dit verschijnsel uit het leeftijdsdiagram afleiden?

Testen

Opgave 12

Bekijk de grafieken uit de Jongerenmonitor 2013. Ze gaan over het percentage flexwerkers onder jongeren ten opzichte van de werkzame bevolking.

Ontwikkeling flexibele arbeid niet-onderwijsvolgende jongeren (15 tot 27 jaar)

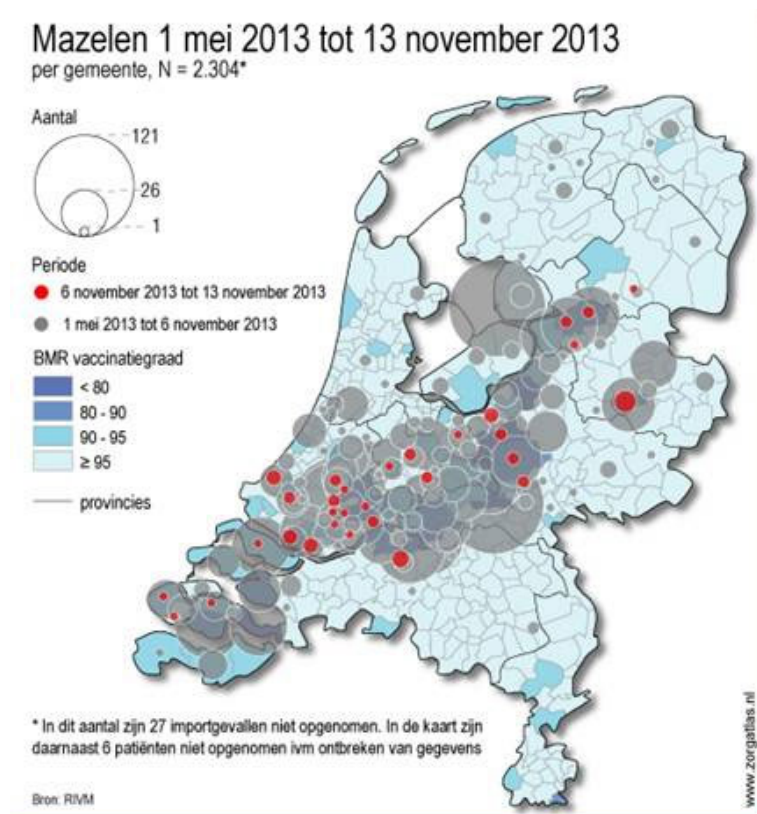


Figuur 10

- Welke doelgroepen worden hier vergeleken?
- Welke uitspraken kun je doen aan de hand van dit diagram?

Opgave 13

De landkaart van Nederland geeft het beeld weer van de mazelenepidemie in de periode mei tot en met november 2013.



Figuur 11

- Welke data heb je nodig om deze kaart te kunnen maken?
- Hoe zijn de data geordend?
- Welke betekenis hebben de rode puntjes op de kaart?
- Voor wie is deze kaart vooral bedoeld?



© 2024

Deze paragraaf is een onderdeel van het Math4All wiskundemateriaal.

Math4All stelt het op prijs als onvolkomenheden in het materiaal worden gemeld en ideeën voor verbeteringen in de content of dienstverlening kenbaar worden gemaakt. Klik op  in de marge bij de betreffende opgave. Uw mailprogramma wordt dan geopend waarbij het emailadres en onderwerp al zijn ingevuld. U hoeft alleen uw opmerkingen nog maar in te voeren.

Email: f.spijkers@math4all.nl

Met de Math4All Foliostraat kunnen complete readers worden samengesteld en toetsen worden gegenereerd. Docenten kunnen bij a.f.otten@math4all.nl een gratis inlog voor de maatwerkdienst aanvragen.
