

Zonder open data geen datajournalistiek

Sjoerd Mouissie, Nederlands Dagblad

3 juni 2021

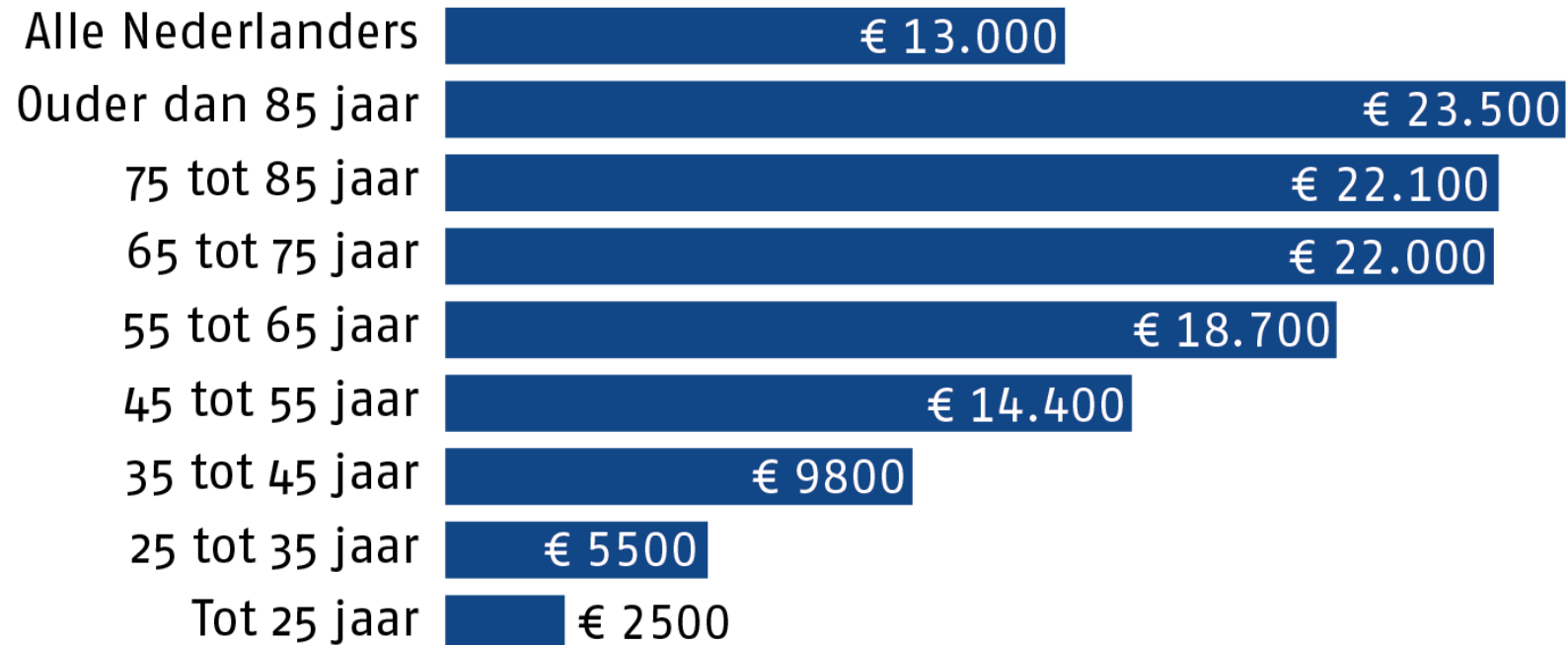
Twitter / LinkedIn: @SjoerdMouissie

Kleine graphics bij geplande nieuwsberichten

Bijvoorbeeld: hoe zit het met landbouwgrond,
aantallen koeien, ontwikkeling bevolkingscijfers

HOVEEL SPAARGELD HEBBEN WE?

Een doorsnee Nederlands huishouden had in 2017 zo'n 13.000 euro spaargeld. In deze grafiek staan de doorsnee spaarbedragen van huishoudens per leeftijd van de hoofdkostwinner.



bron data: CBS

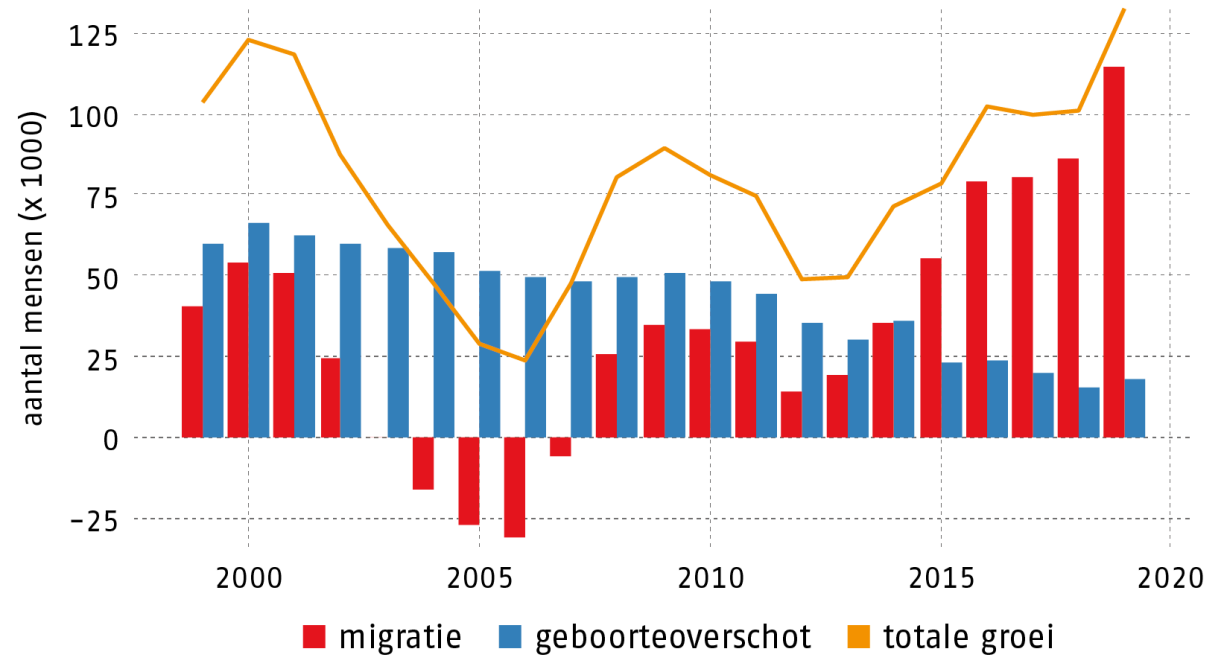
LANDBOUWGROND IN NEDERLAND

legenda

- landbouwgrond
- kassen
- natuurgebieden
- bebouwing, wegen

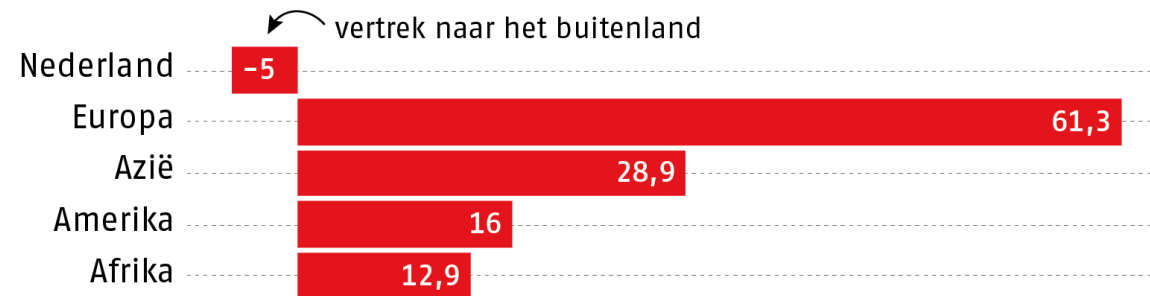


FORSE BEVOLKINGSGROEI IN 2019

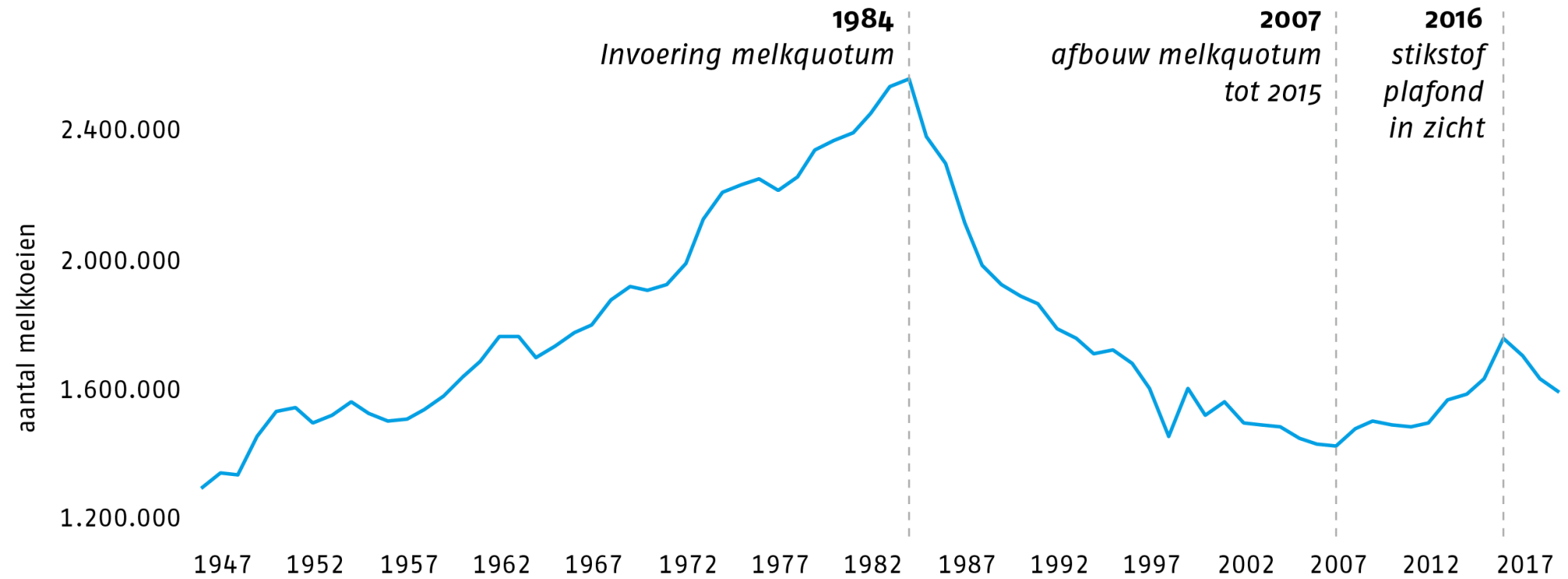


Herkomst van migranten

(saldo van binnenkomende en vertrekkende migranten, aantallen x 1000)

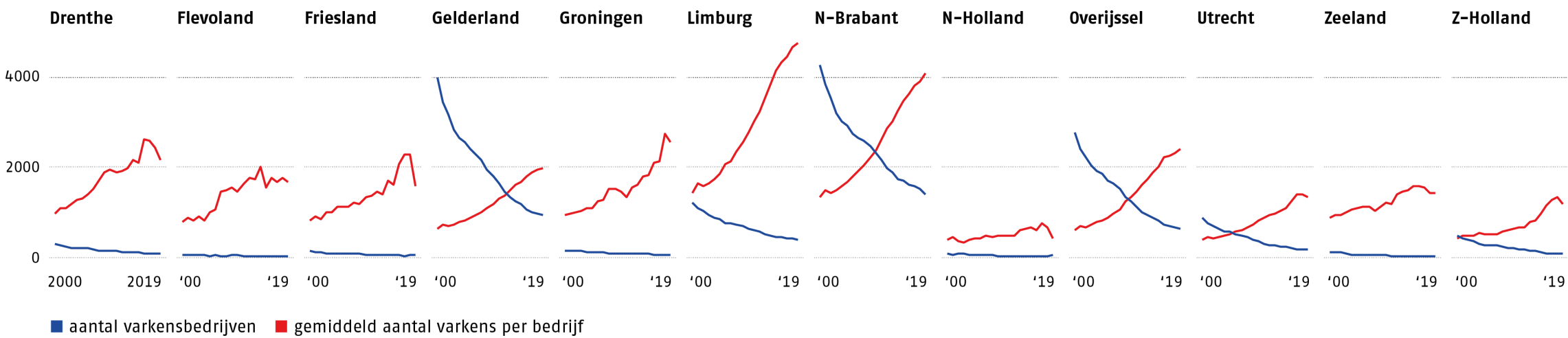


AANTAL MELKKOEIEN SINDS 1946



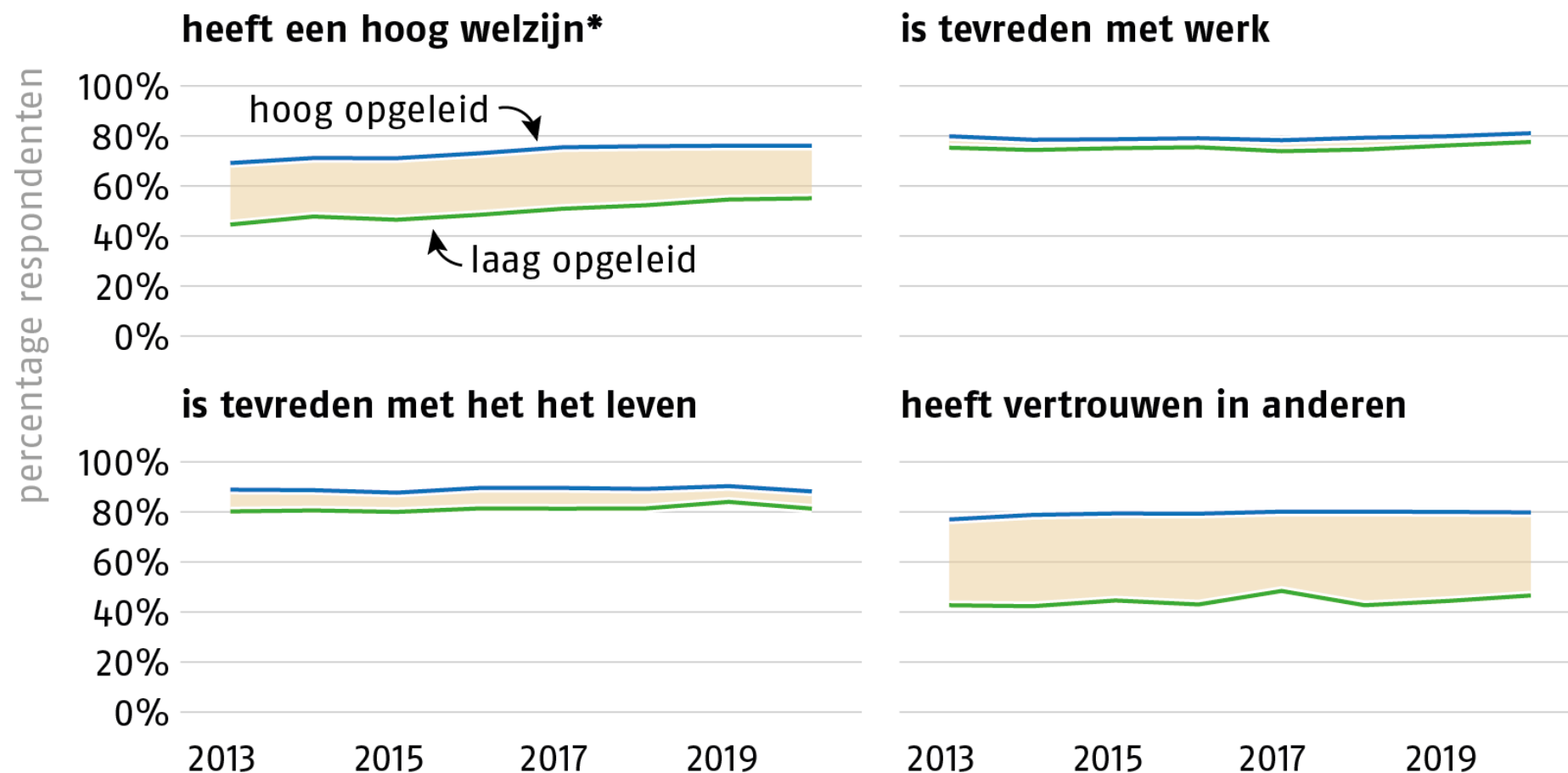
© 2019/2020 Sjoerd Mouissie, Nederlands Dagblad. Bron data: CBS

DE VARKENSHOUDERIJ: HIER IS DE AGRARISCHE SCHAALVERGROTING HET MEEST ZICHTBAAR



OPLEIDING MAAKT VEEL UIT IN VERDELING WELVAART

CBS-onderzoek toont verschil in tevredenheid, welzijn en vertrouwen



* goede score op onder andere gezondheid, financiën, woonomgeving

© 20210519 Sjoerd Mouissie, Nederlands Dagblad. Bron data: CBS

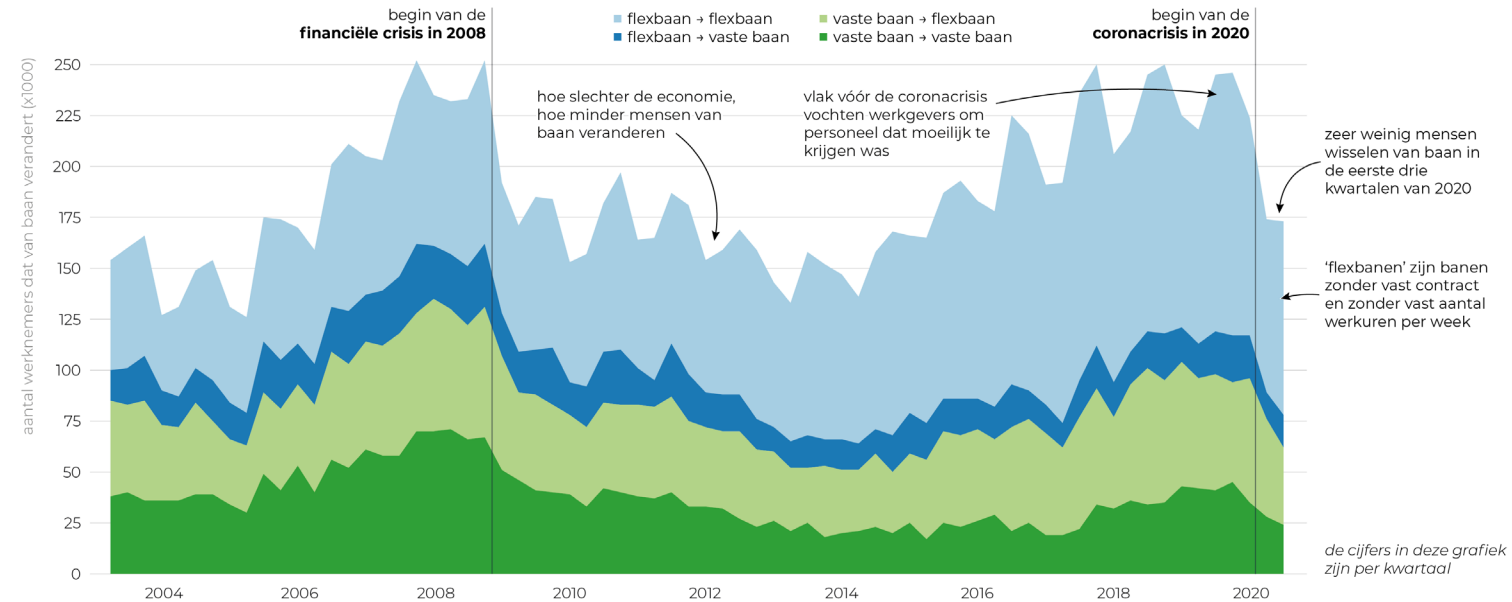
Grotere graphics met CBS-gegevens

Bijvoorbeeld: krantenbijlage over onderwijs en
werkgelegenheid; series graphics over
interessante data van het afgelopen jaar

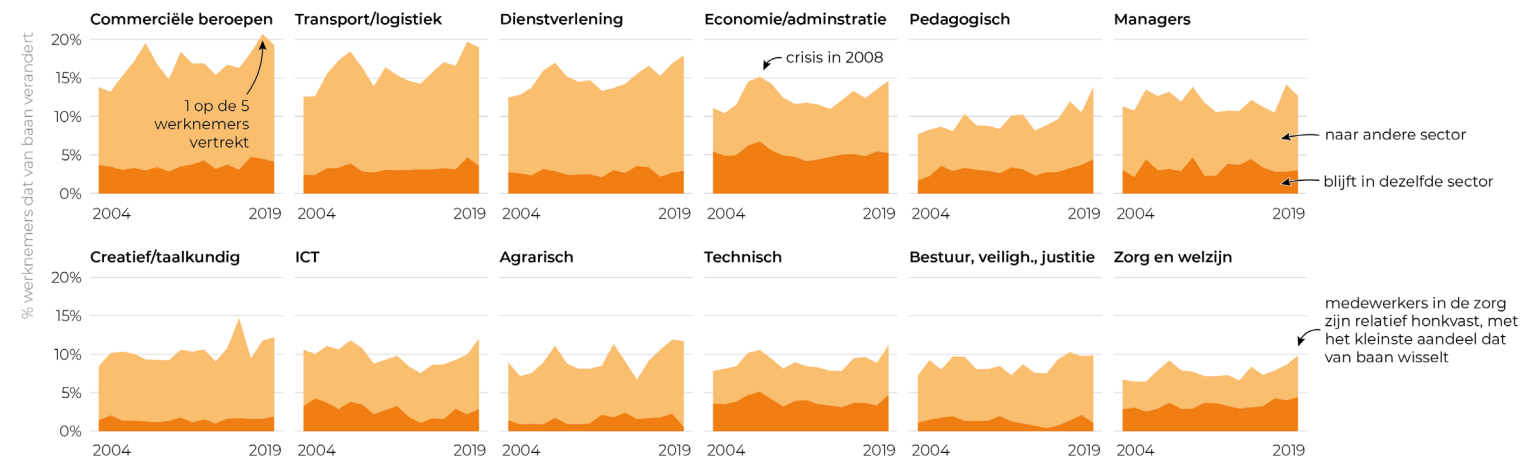
IN SOMMIGE SECTOREN VERTREKT ÉÉN OP DE VIJF NAAR EEN ANDERE BAAS

Aan de statistieken over hoe veel mensen van baan veranderen, is precies te zien hoe het met de economie gaat. Hoe slechter de omstandigheden, hoe minder werknemers risico nemen. Maar als mensen van baan veranderen, kiezen ze meestal voor werk in een andere sector.

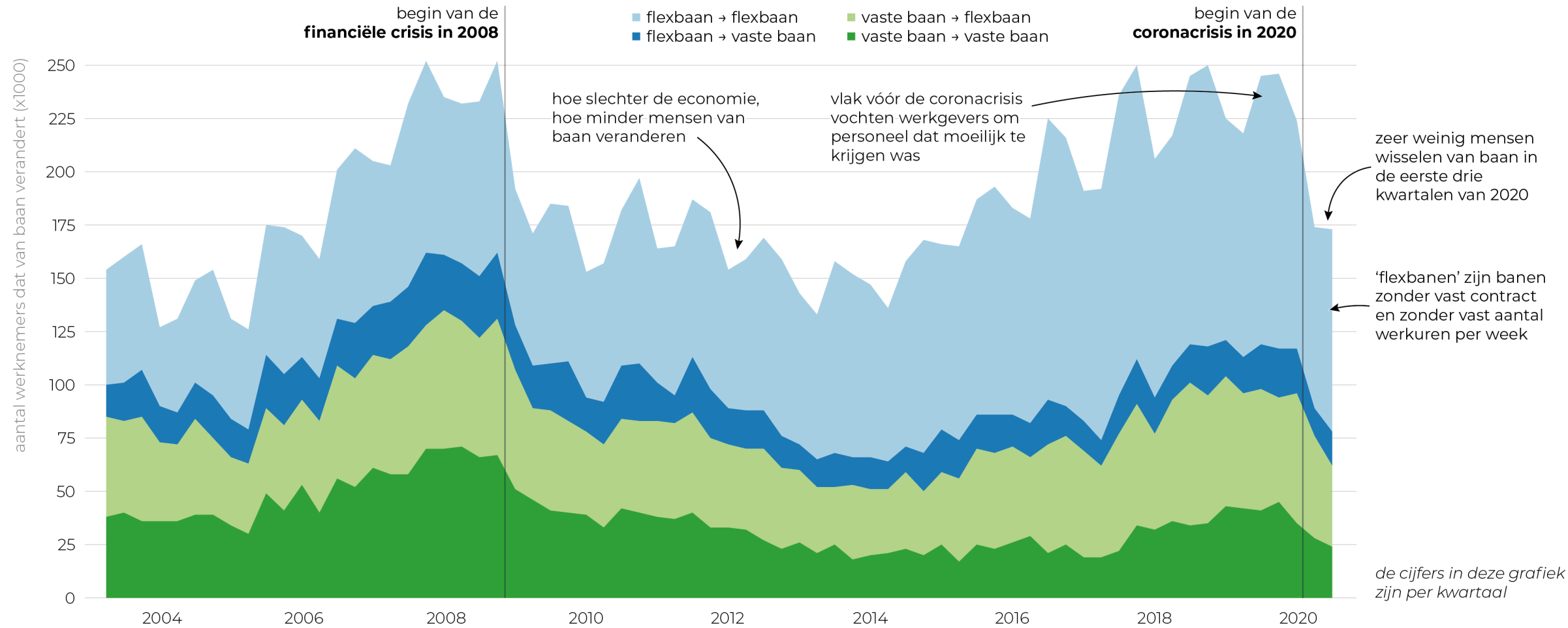
Hoe veel werknemers veranderen van baan, en welk contract krijgen ze dan?



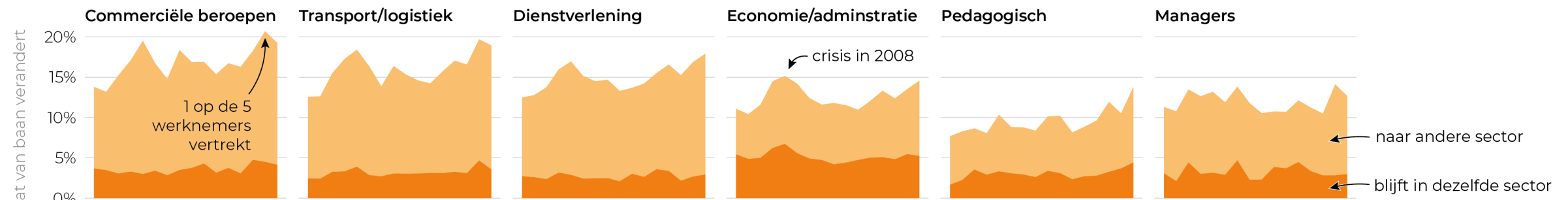
Werknemers die vertrekken naar een nieuwe baan in **dezelfde** of een **andere** sector



Hoe veel werknemers veranderen van baan, en welk contract krijgen ze dan?



Werknemers die vertrekken naar een nieuwe baan in dezelfde of een andere sector

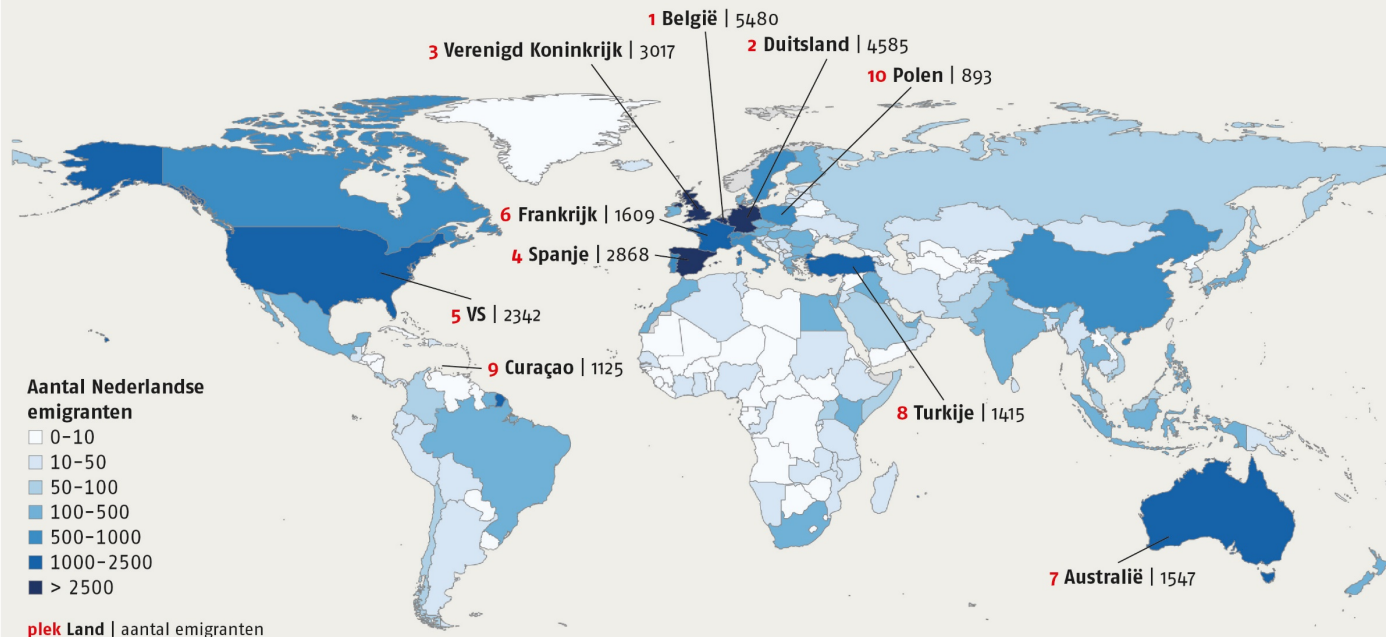


Hierheen verhuisden duizenden Nederlanders

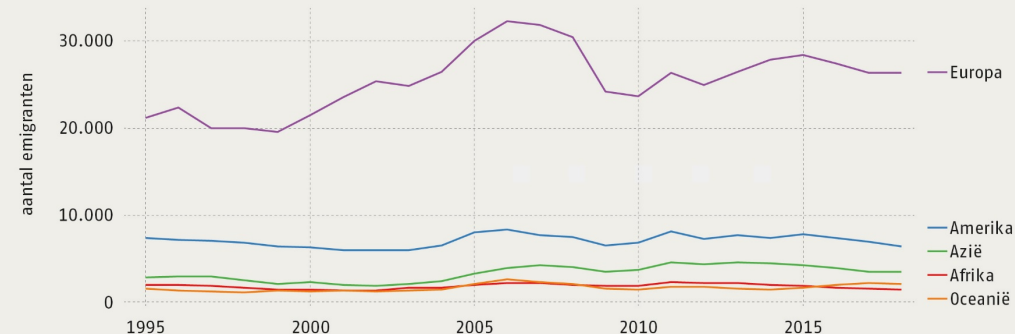
Data zijn overal. Elke dag verschijnen nieuwe gegevens. In deze serie belichten we met infographics opvallende bevindingen van afgelopen jaar. Vandaag: emigratie van Nederlanders.

Het CBS presenteerde vorige week nieuwe cijfers: er wonen nu zo'n 17,4 miljoen mensen in Nederland. De bevolking groeit snel, mede door migratie. Tienduizenden mensen uit alle windstreken vestigen zich in Nederland voor werk, gezin of studie. Uit de cijfers blijkt ook dat ongeveer 39.000 Nederlanders naar andere landen emigreerden. Waar verhuizen al die landgenoten naartoe? Precieze cijfers daarover zijn lastig te vinden. De Verenigde Naties publiceerden in de zomer een rapport over hoeveel buitenlanders er in diverse landen wonen. Die informatie is echter summier en incompleet – er zijn immers nogal wat landen die geen goede administratie bijhouden. Eurostat, het Europese statistiekbureau, houdt alleen gegevens bij over de Europese lidstaten. Het beste beeld komt uit een tabel van het CBS met cijfers over 2018, die vorig jaar werd gepubliceerd. Daarin staat hoeveel mensen met de Nederlandse nationaliteit bij de gemeente gemeld hebben dat ze naar een bepaald land gaan emigreren. Dat levert gegevens op over de verhuizingen naar bijna 240 landen. Daaruit blijkt dat Europese landen verreweg de meest populaire bestemming zijn voor emigranten. Daarna komen de Verenigde Staten.

Top tien van landen waar Nederlanders in 2018 naartoe geëmigreerd zijn



Ontwikkeling van het aantal emigranten per continent



In de migratiecijfers van deze infographic zijn geen 'administratieve correcties' verwerkt. Deze correcties worden bijvoorbeeld gedaan als een gemeente iemand uitschrijft, omdat hij of zij onvindbaar is. Hierdoor kunnen de werkelijke migratiecijfers iets verschillen van de aantallen die hier staan.

Nertsen: cijfers over een verdwenen sector

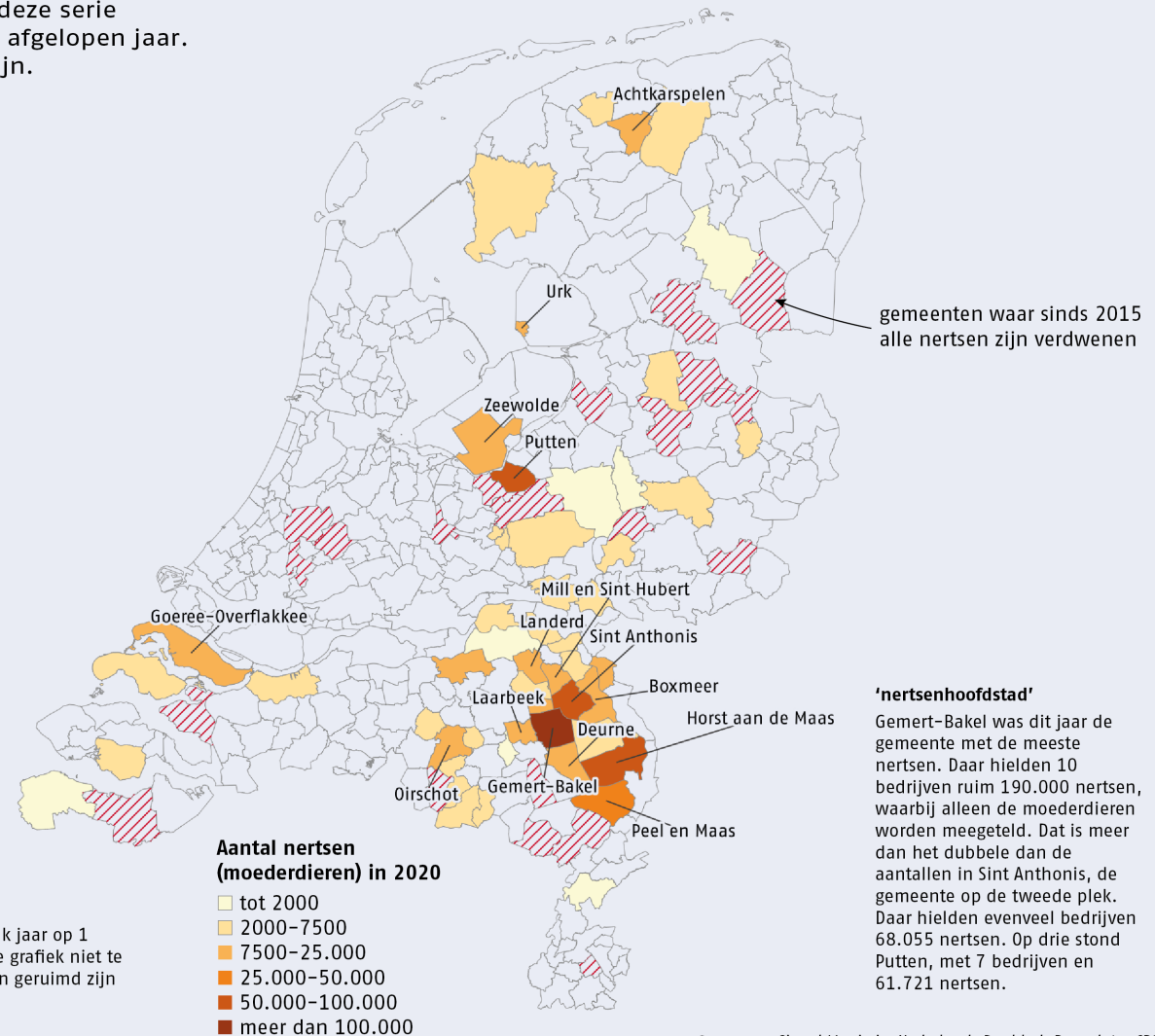
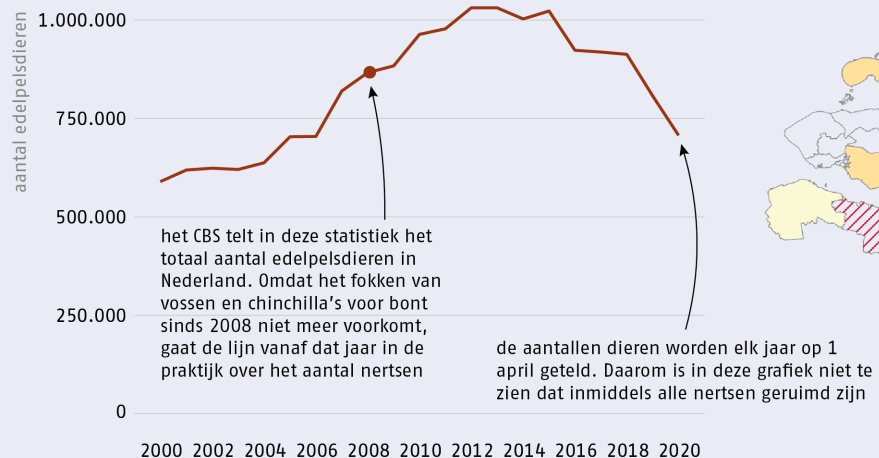
Data zijn overal. Elke dag verschijnen nieuwe gegevens. In deze serie belichten we met infographics opvallende bevindingen van afgelopen jaar. Vandaag: nertsbedrijven, die dit jaar allemaal geruimd zijn.

alle nertsen geruimd

Eerder dit jaar besloot het kabinet alle nertsbedrijven te ruimen, omdat op diverse bedrijven de dieren besmet bleken met het coronavirus. Volgens de afspraken zouden alle bedrijven in 2024 moeten stoppen. Die termijn werd vervroegd naar maart 2021, maar later kwam het kabinet daar op terug. De bedrijven mochten het 'seizoen' afmaken, waarbij de dieren worden gedood voor hun vacht. Begin december meldde minister Carola Schouten (Landbouw) dat ook de laatste overgebleven nertsen waren geruimd. Fokkers krijgen een overbruggingsvergoeding tot 2024, om zich op de toekomst te oriënteren.

Op de kaart en in de grafiek over nertsen worden alleen de moederdieren meegeteld. In de praktijk is het totaal aantal dieren dus hoger, gezien een moeder vier à vijf jongen krijgt.

aantallen edelpelsdieren in Nederland, 2000-2020



Corona op de Biblebelt: kerk of grote gezinnen?

Data zijn overal. Elke dag verschijnen nieuwe gegevens. In deze serie belichten we met infographics opvallende bevindingen van afgelopen jaar. Vandaag: coronabesmettingen op de 'Biblebelt'.

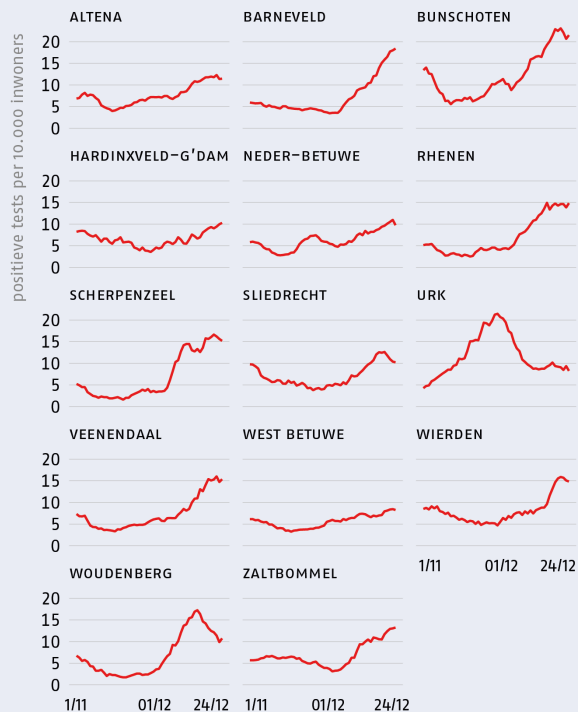
overlappende statistieken

Het RIVM meldde deze week dat 'christelijke gemeenten' opvallend hoge besmettingscijfers hebben. Een gemeente is volgens het RIVM 'christelijk' als de ChristenUnie en SGP in 2019 meer dan 15 procent van de stemmen haalden. Kerkdiensten werden niet direct als oorzaak aangewezen, waarschijnlijk speelt de hechte samenleving een belangrijke rol.

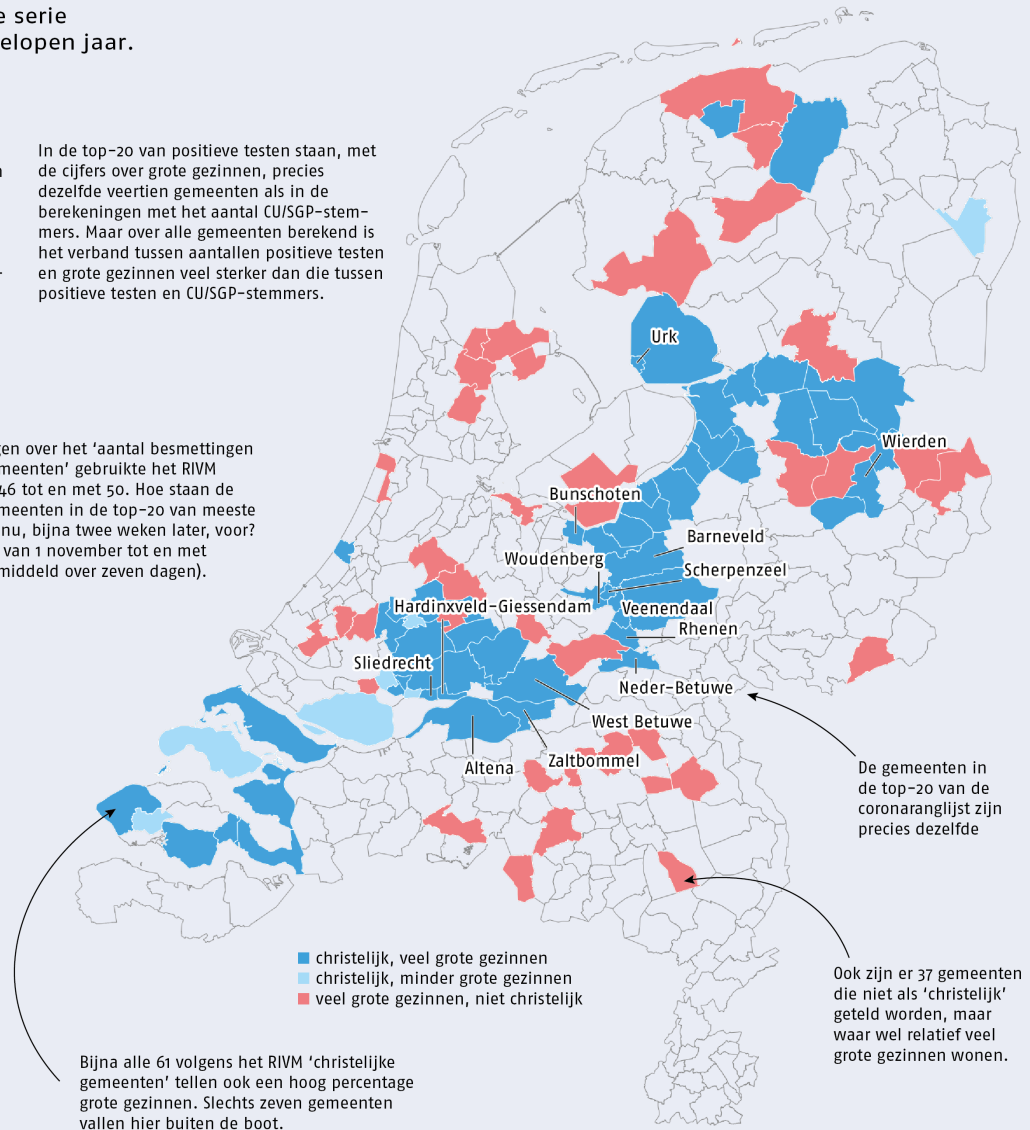
Er is echter een andere statistiek die dat aantoont: die van het aantal grote gezinnen (van vijf of meer personen). Ook hier geldt: gemeenten met relatief veel grote gezinnen (6,5 procent of meer), scoren hoog op de coronaranglijst. Dat komt waarschijnlijk omdat grote gezinnen meer contactmomenten hebben dan kleine huishoudens.

In de top-20 van positieve testen staan, met de cijfers over grote gezinnen, precies dezelfde veertien gemeenten als in de berekeningen met het aantal CU/SGP-stemmers. Maar over alle gemeenten berekend is het verband tussen aantallen positieve testen en grote gezinnen veel sterker dan die tussen positieve testen en CU/SGP-stemmers.

actuele cijfers 'christelijke gemeenten'



In de berekeningen over het 'aantal besmettingen in christelijke gemeenten' gebruikte het RIVM cijfers van week 46 tot en met 50. Hoe staan de 14 christelijke gemeenten in de top-20 van meeste besmettingen er nu, bijna twee weken later, voor? Dit zijn de cijfers van 1 november tot en met 24 december (gemiddeld over zeven dagen).



Bijna alle 61 volgens het RIVM 'christelijke gemeenten' tellen ook een hoog percentage grote gezinnen. Slechts zeven gemeenten vallen hier buiten de boot.

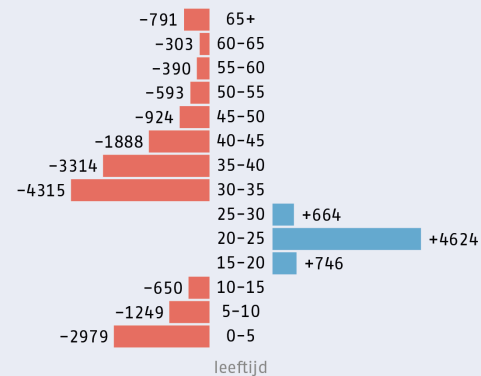
De gemeenten in de top-20 van de coronaranglijst zijn precies dezelfde

Ook zijn er 37 gemeenten die niet als 'christelijk' geteld worden, maar waar wel relatief veel grote gezinnen wonen.

Corona of niet, we blijven gretig verhuizen

Data zijn overal. Elke dag verschijnen nieuwe gegevens. In deze serie belichten we met infographics opvallende bevindingen van afgelopen jaar. Vandaag: verhuizingen, die ondanks alles onverminderd doorgaan.

Aantal mensen dat van/naar Amsterdam verhuist, per leeftijd

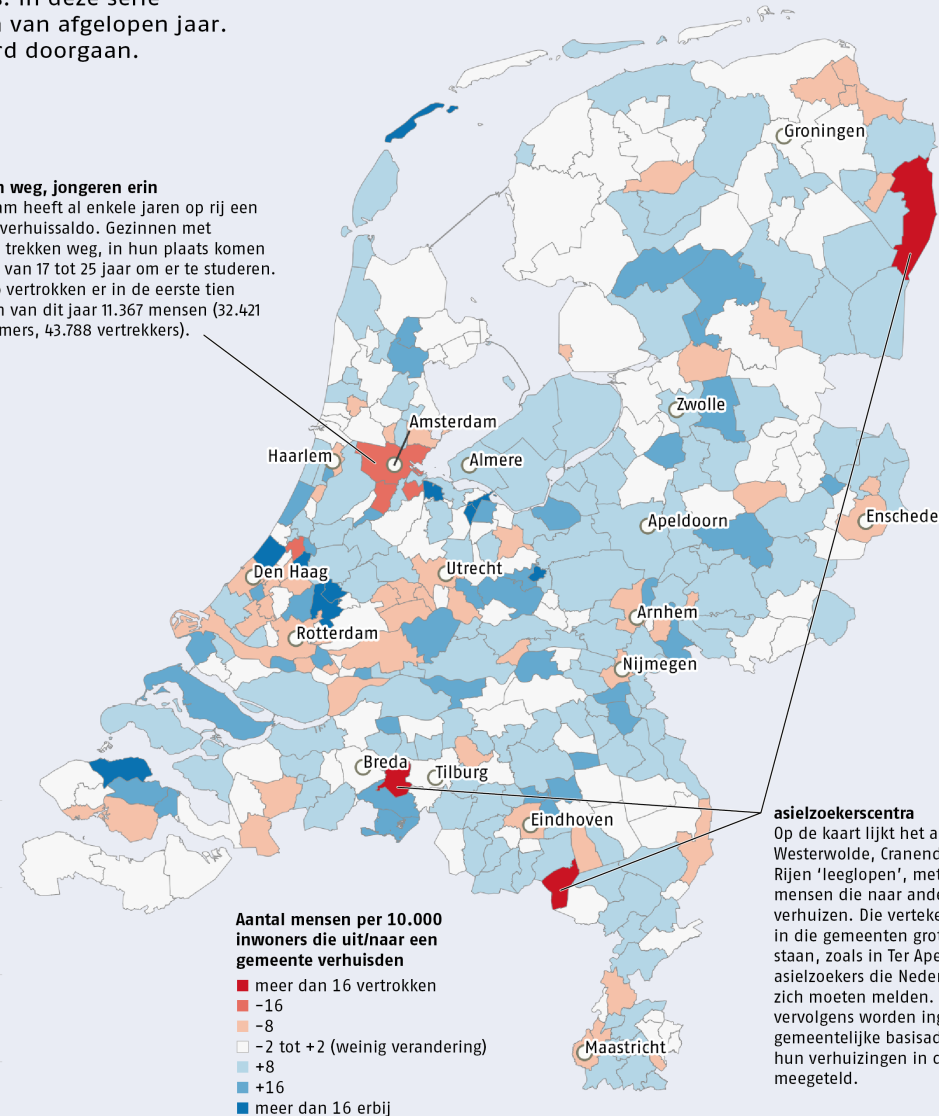
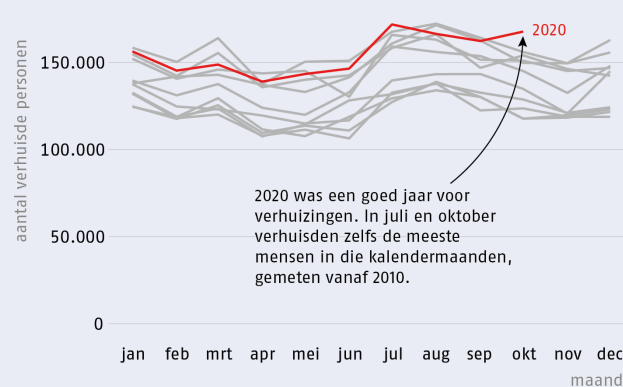


De grafieken en de kaart gaan over de periode januari t/m oktober

gezinnen weg, jongeren erin

Amsterdam heeft al enkele jaren op rij een negatief verhuissaldo. Gezinnen met kinderen trekken weg, in hun plaats komen jongeren van 17 tot 25 jaar om er te studeren. Per saldo vertrokken er in de eerste tien maanden van dit jaar 11.367 mensen (32.421 nieuwkomers, 43.788 vertrekkers).

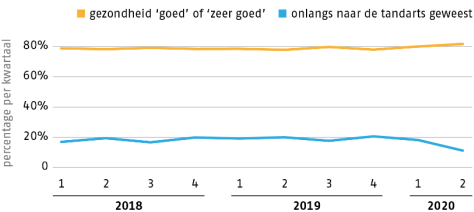
Aantal personen dat binnen of tussen gemeenten verhuisde, 2010-2020



Hoe gaat met ons in coronatijd?

CBS komt met eigen dashboard over gevolgen crisis

Het Centraal Bureau voor de Statistiek (CBS) presenteert vandaag een eigen dashboard: 'Welvaart in coronatijd'. Dat is geïnspireerd op de jaarlijkse Monitor Brede Welvaart, die door alle ontwikkelingen echter achterhaald voelde. Op deze pagina staat een greep uit de cijfers van het CBS.

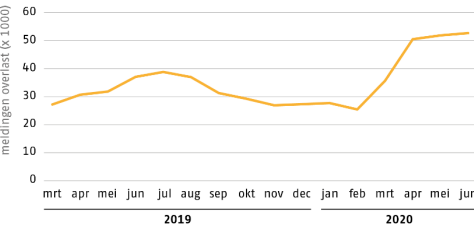
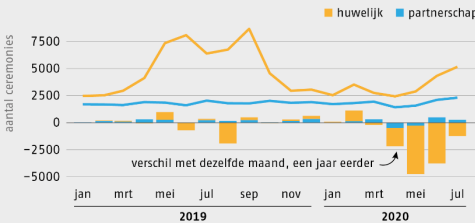


gezondheid

Door COVID-19 overleden er dit jaar bijna negenduizend mensen meer dan verwacht. Tegelijk voelden we ons in het tweede kwartaal van 2020 duidelijk gezonder dan in eerdere jaren – 82 procent van de mensen noemt de eigen gezondheid 'goed of zeer goed'. Het bezoek aan huisartsen, medisch specialisten en tandartsen daalde.

samenleving

Door de coronacrisis mochten er bij huwelijken en andere gelegenheden maar dertig bezoekers zijn. Het aantal huwelijken en geregistreerde partnerschappen daalde vrijwel direct. In april en mei trouwden bijna de helft minder stellen dan in dezelfde maanden van 2019.

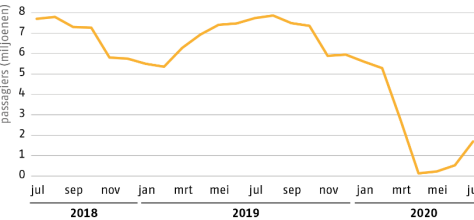
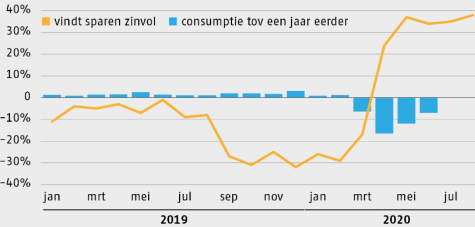


veiligheid

Tijdens de coronacrisis registreerde de politie fors meer meldingen van overlast. Vooral in april en mei was dit het geval – het aantal meldingen was 60 procent hoger dan vorig jaar. Daarbij ging het vooral om geluids- overlast. Dat het aantal meldingen in juni toeneemt, is normaal voor het (zomer)seizoen.

materiële welvaart

De consumptie door huishoudens, een belangrijke graadmeter voor materiële welvaart, daalde op een manier die nooit eerder gezien werd. In plaats daarvan sloegen mensen aan het sparen. Waar zij sparen eerder niet zo zinvol vonden, kanteelde dit beeld vanaf maart volledig. Nu vindt een ruime meerderheid van de deelnemers naar onderzoeken over consumentenvertrouwen sparen wél zinvol.



mobilititeit

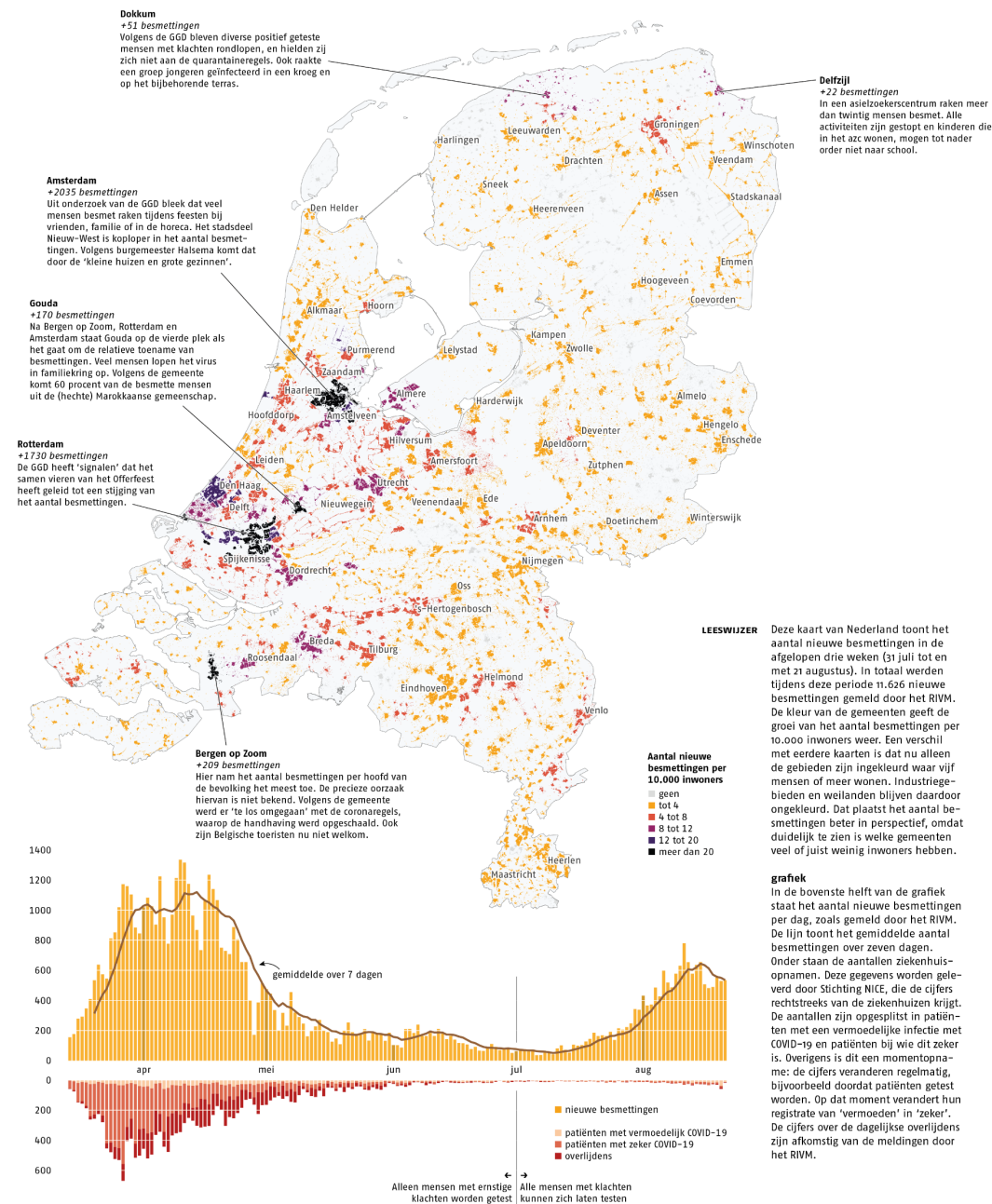
Het aantal passagiers op Nederlandse vliegvelden werd direct na het uitbreken van de coronacrisis, in maart, al gehalveerd. In april en mei werd er nauwelijks gevlogen – het aantal passagiers daalde tot bijna nul. Het vliegverkeer lag die periode dan ook nagenoeg stil.

CBS-data als 'tussenlaag' in project

Bijvoorbeeld: data over bevolkingsdichtheid
in kaart over coronabesmettingen

Ruim elfduizend coronabesmettingen in drie weken

Vooral in het westen van Nederland stijgen de aantallen soms flink.



+2035 besmettingen

Uit onderzoek van de GGD bleek dat veel mensen besmet raken tijdens feesten bij vrienden, familie of in de horeca. Het stadsdeel Nieuw-West is koploper in het aantal besmettingen. Volgens burgemeester Halsema komt dat door de 'kleine huizen en grote gezinnen'.

Gouda

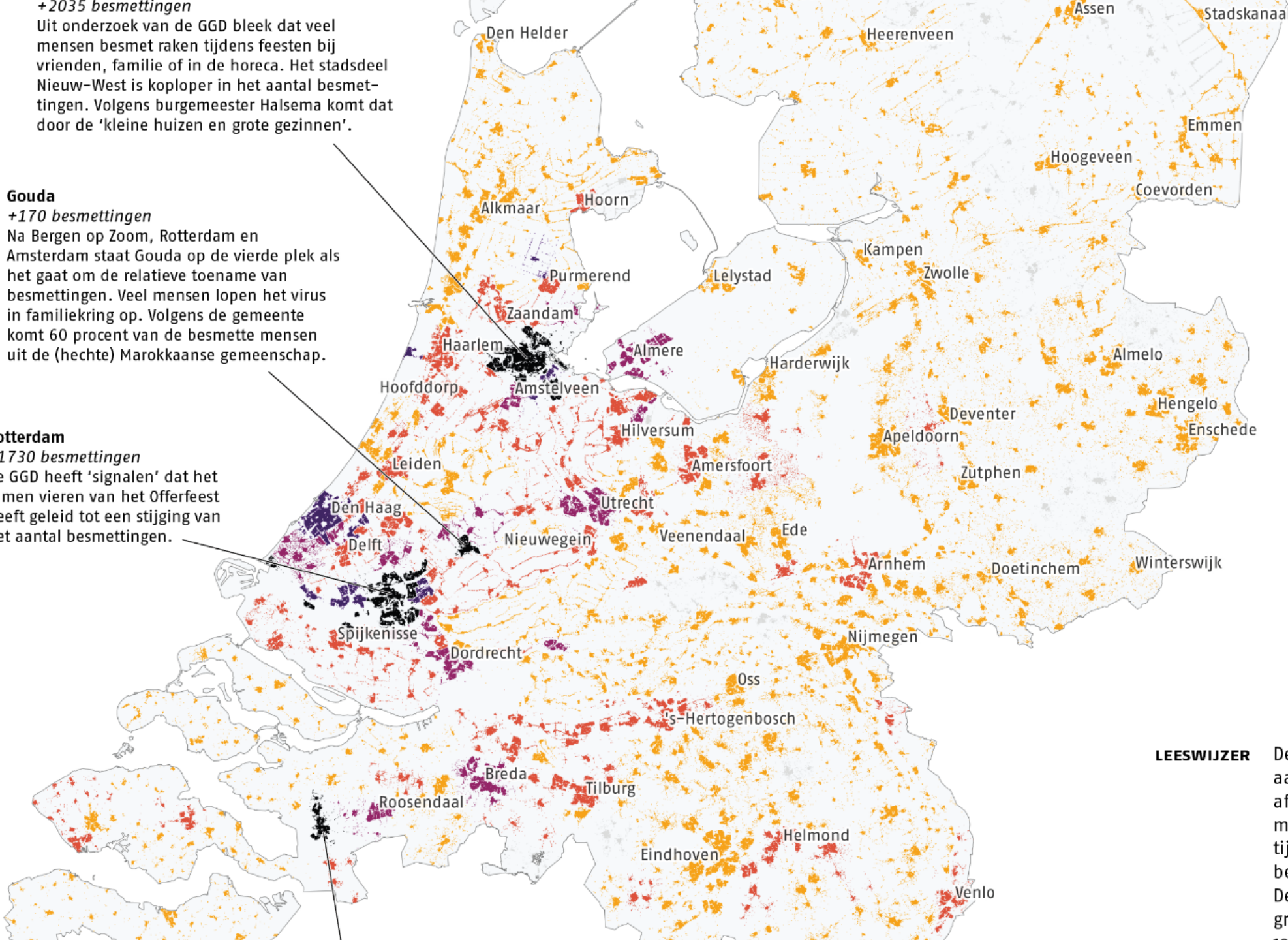
+170 besmettingen

Na Bergen op Zoom, Rotterdam en Amsterdam staat Gouda op de vierde plek als het gaat om de relatieve toename van besmettingen. Veel mensen lopen het virus in familiekring op. Volgens de gemeente komt 60 procent van de besmette mensen uit de (hechte) Marokkaanse gemeenschap.

Rotterdam

+1730 besmettingen

De GGD heeft 'signalen' dat het samen vieren van het Offerfeest heeft geleid tot een stijging van het aantal besmettingen.



LEESWIJZER

Deze kaart van Nederland toont het aantal nieuwe besmettingen in de afgelopen drie weken (31 juli tot en met 21 augustus). In totaal werden tijdens deze periode 11.626 nieuwe besmettingen gemeld door het RIVM. De kleur van de gemeenten geeft de groei van het aantal besmettingen per 10.000 inwoners weer. Een verschil

Drie weken positieve coronatesten in 3D

RIVM noemt cijfers 'zorgwekkend'; regionale verschillen zijn groot

Waar het noorden van Nederland relatief weinig merkt van het coronavirus, worden er veel mensen positief getest in Twente, Utrecht, Brabant en Zuid-Holland. Vooral de hoge aantallen in de gemeenten Urk en Bunschoten springen eruit, al gaat het er in de afgelopen dagen beter. Alle genoemde aantallen zijn het aantal gemelde positieve tests per 10.000 inwoners.

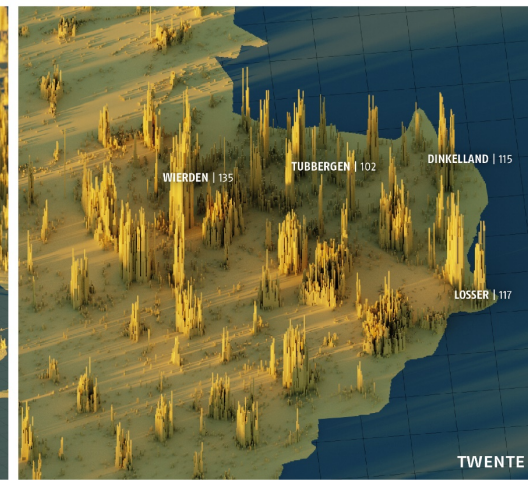
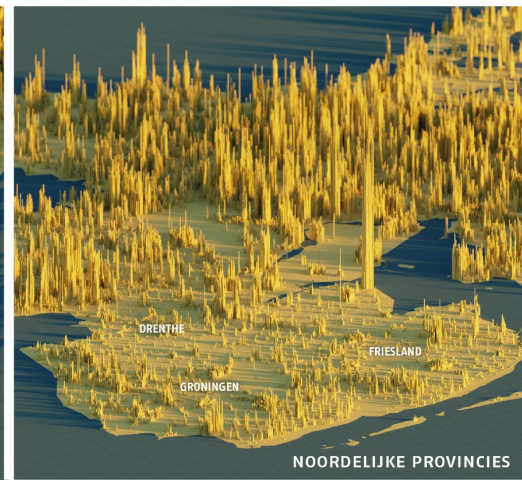
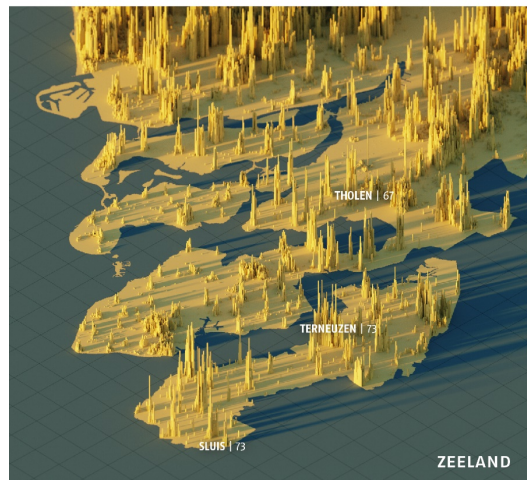
WERKWIJZE

De basis voor deze kaart is een bestand van het Centraal Bureau voor de Statistiek. Daarin is Nederland verdeeld in bijna 400.000 vakjes van honderd bij honderd meter. Van elk vakje is bekend hoe veel mensen er wonen. Ook is per gemeente bekend hoe veel positieve coronatesten er gemeld zijn. Die twee getallen zijn vervolgens gebruikt om de hoogte van de 'balkjes' te bepalen. In gebieden met weinig inwoners zijn ze minder hoog dan waar veel mensen wonen – vandaar dat de hoogte van de balkes in één gemeente varieert. De maximale hoogte van de balkjes in een plaats wordt bepaald door het relatieve aantal positieve tests.

ILLUSTRATIE

Door de toegepaste 3D-techniek is dit meer een illustratie dan een kaart waarop aantallen en gemeenten nauwkeurig te vergelijken zijn. Het doel is inzicht te geven in hoe regio's van elkaar verschillen.

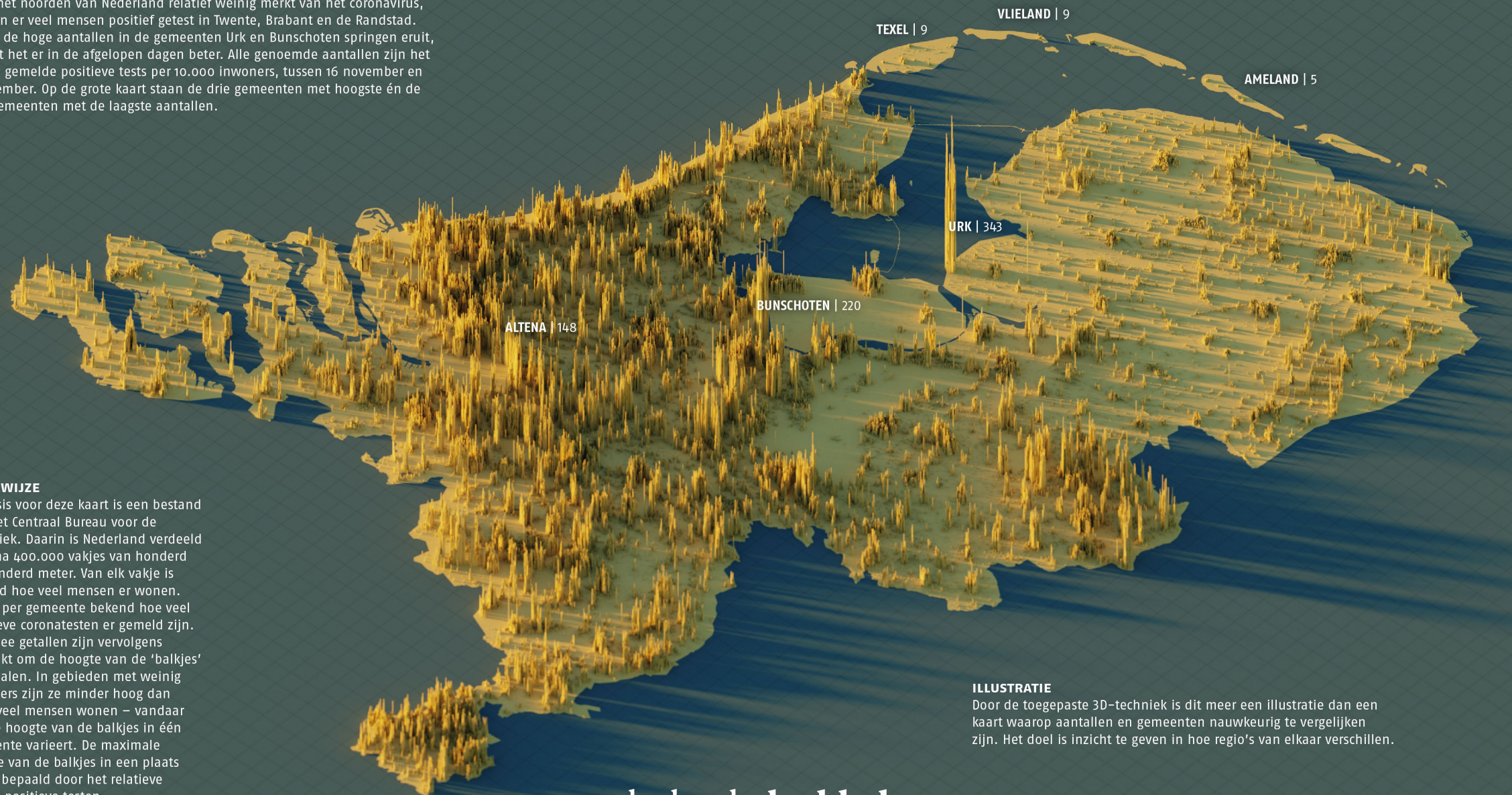
© 2020/208 Sjoerd Mouissie, Nederlands Dagblad. Bron data: RIVM, CBS, Kadaster



Drie weken positieve coronatesten in 3D

RIVM noemt cijfers 'zorgwekkend'; regionale verschillen zijn groot

Waar het noorden van Nederland relatief weinig merkt van het coronavirus, worden er veel mensen positief getest in Twente, Brabant en de Randstad. Vooral de hoge aantallen in de gemeenten Urk en Bunschoten springen eruit, al gaat het er in de afgelopen dagen beter. Alle genoemde aantallen zijn het aantal gemelde positieve tests per 10.000 inwoners, tussen 16 november en 8 december. Op de grote kaart staan de drie gemeenten met hoogste én de drie gemeenten met de laagste aantallen.

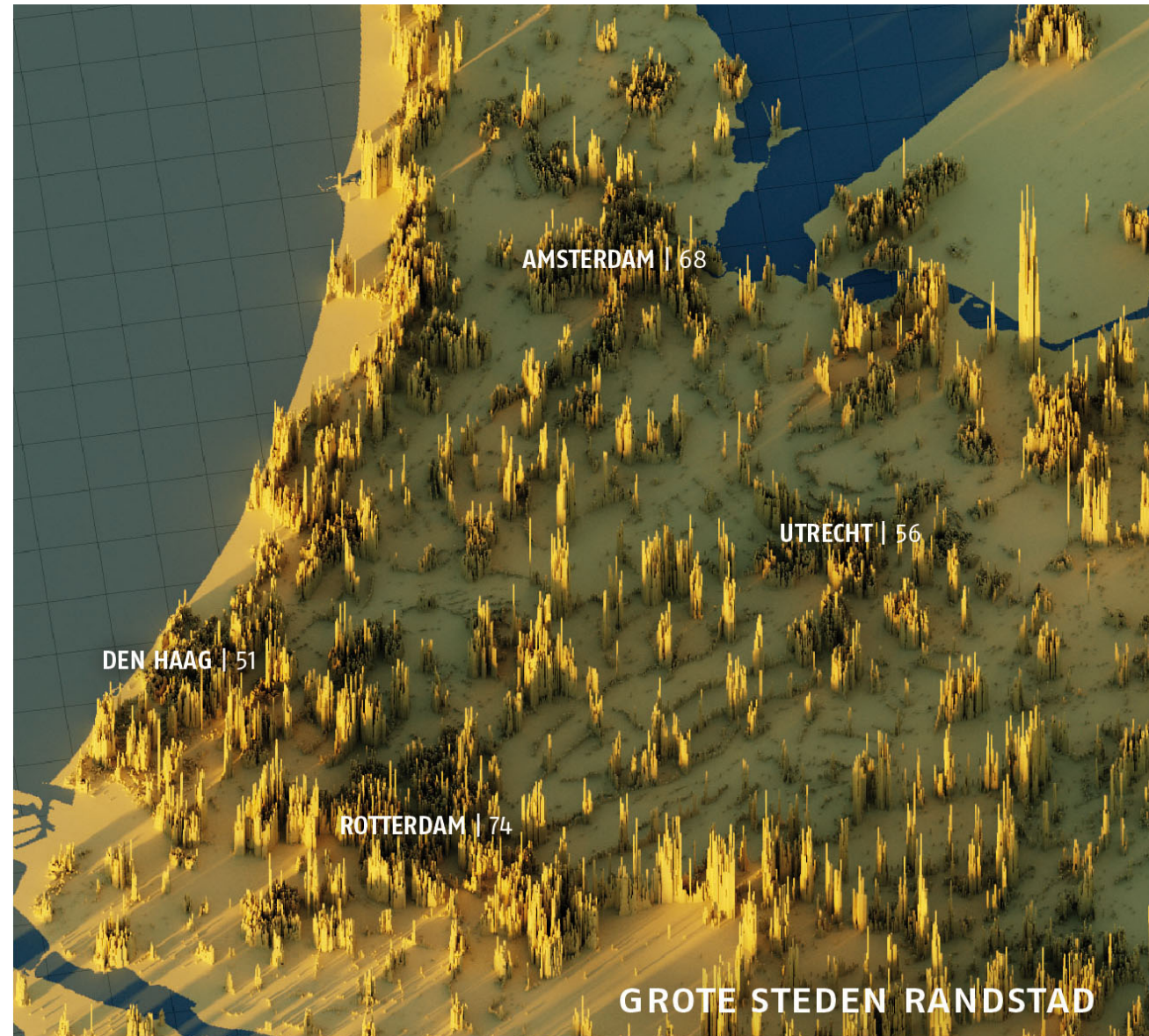


WERKWIJZE

De basis voor deze kaart is een bestand van het Centraal Bureau voor de Statistiek. Daarin is Nederland verdeeld in bijna 400.000 vakjes van honderd bij honderd meter. Van elk vakje is bekend hoe veel mensen er wonen. Ook is per gemeente bekend hoe veel positieve coronatesten er gemeld zijn. Die twee getallen zijn vervolgens gebruikt om de hoogte van de 'balkjes' te bepalen. In gebieden met weinig inwoners zijn ze minder hoog dan waar veel mensen wonen – vandaar dat de hoogte van de balkjes in één gemeente varieert. De maximale hoogte van de balkjes in een plaats wordt bepaald door het relatieve aantal positieve testen.

ILLUSTRATIE

Door de toegepaste 3D-techniek is dit meer een illustratie dan een kaart waarop aantallen en gemeenten nauwkeurig te vergelijken zijn. Het doel is inzicht te geven in hoe regio's van elkaar verschillen.



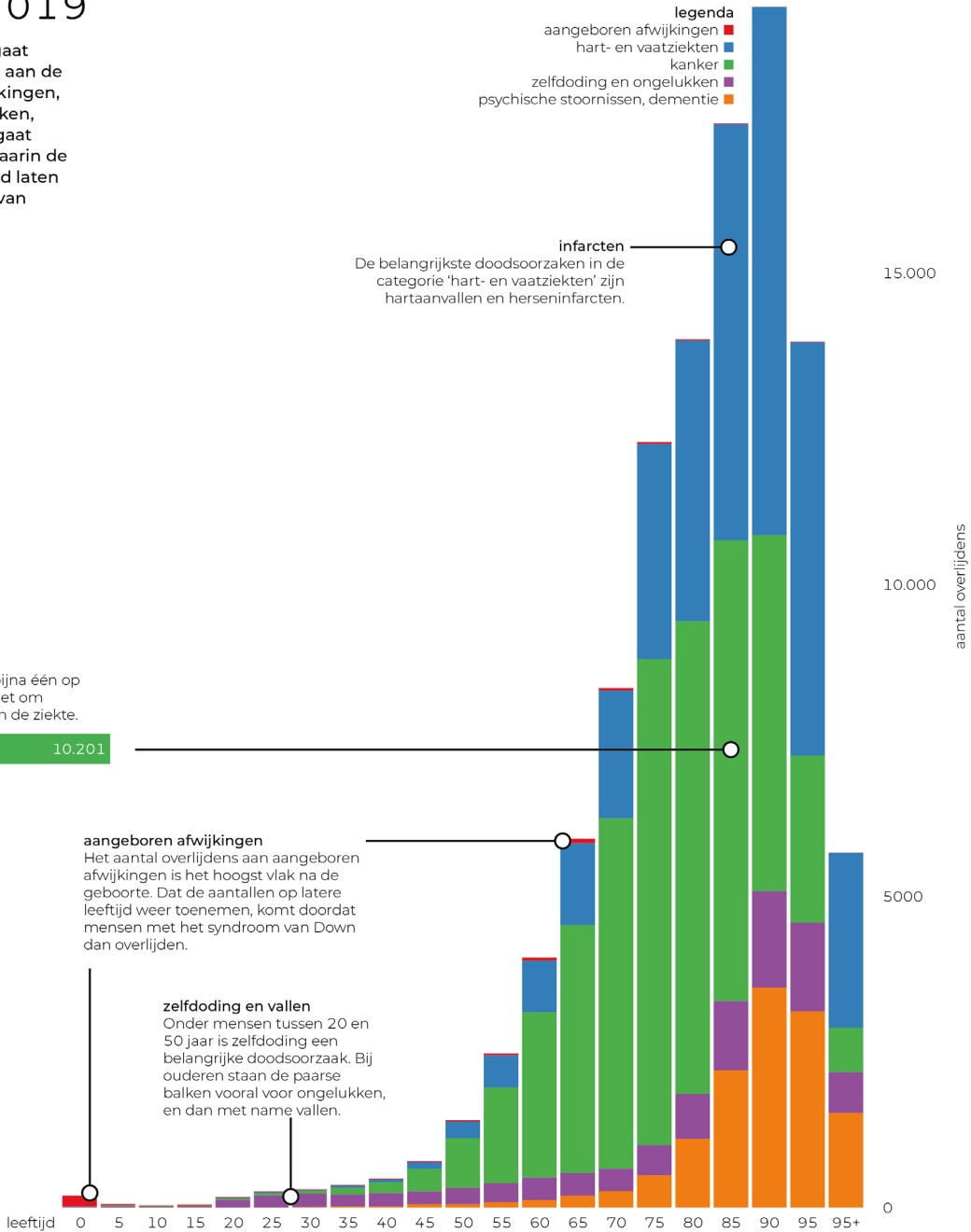
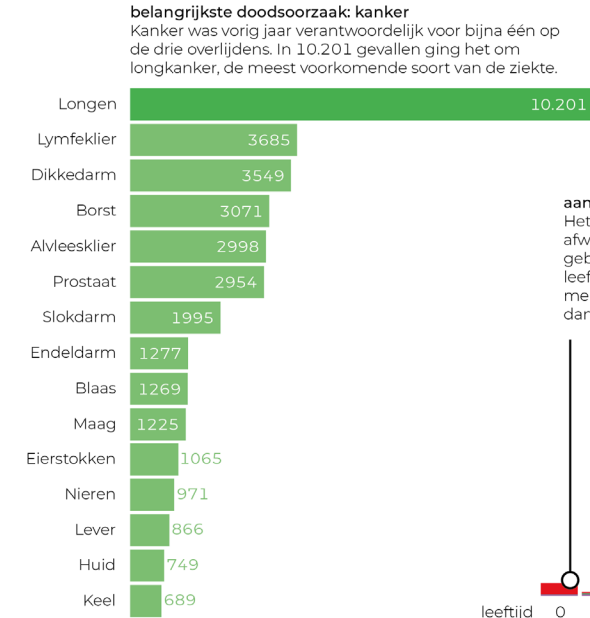
‘Sterven in Nederland’

Serie artikelen en datavisualisaties met CBS-gegevens

Nederlands Dagblad, september 2020

HIERAAN STIERVEN NEDERLANDERS IN 2019

Vorig jaar overleden 151.885 mensen. Deze grafiek gaat over 70 procent van hen – de mensen die overleden aan de vijf belangrijkste doodsoorzaken: aangeboren afwijkingen, hart- en vaatziekten, kanker, zelfdoding en ongelukken, psychische stoornissen en dementie. Dit overzicht gaat nadrukkelijk over 2019 – voorlopig het laatste jaar waarin de statistieken over doodsoorzaken een 'normaal' beeld laten zien. De cijfers van 2020 zullen er, door de uitbraak van COVID-19, ongetwijfeld anders uitzien.

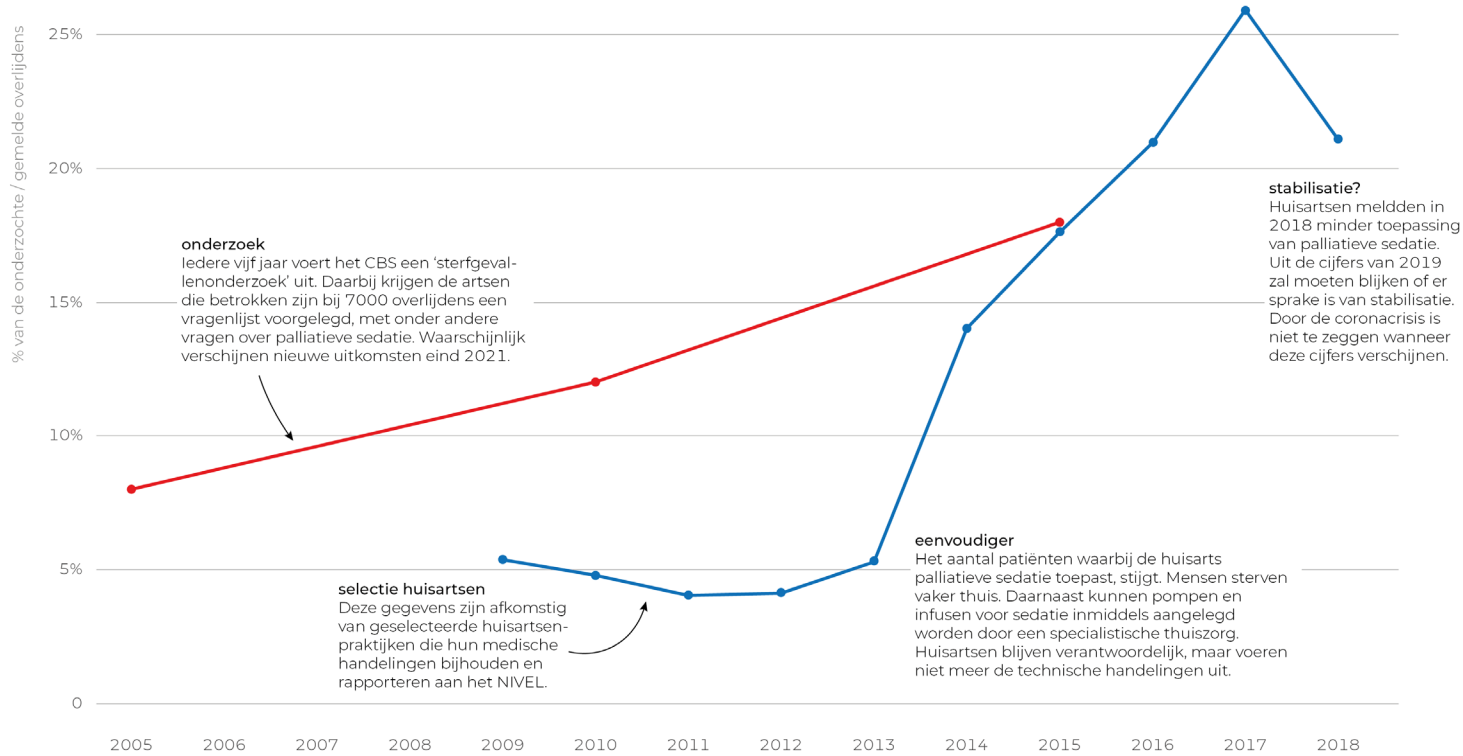


PALLIATIEVE SEDATIE STEEDS VAKER TOEGEPAST

Stervende patiënten krijgen steeds vaker palliatieve sedatie (pijnbestrijding in de laatste levensfase), blijkt uit gegevens van het Nederlands Instituut voor Onderzoek van de Gezondheidszorg (NIVEL) en het CBS. Hoe vaak sedatie precies wordt toegepast, blijft echter onduidelijk. Er bestaat bijvoorbeeld geen centrale registratie van welke middelen

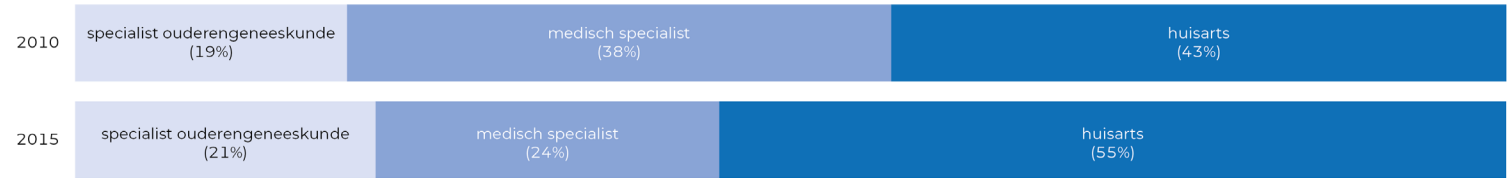
worden toegediend in aanloop naar het overlijden. De weinige beschikbare cijfers komen uit jaarlijkse steekproeven onder huisartsenpraktijken en enquêtes onder artsen, die iedere vijf jaar worden gehouden. Desondanks is de stijgende trend onmiskenbaar. Hier staat een overzicht van de gegevens en de slagen om de arm.

Aantal overlijdens waarbij palliatieve sedatie is toegepast



Welke arts palliatieve sedatie toepast

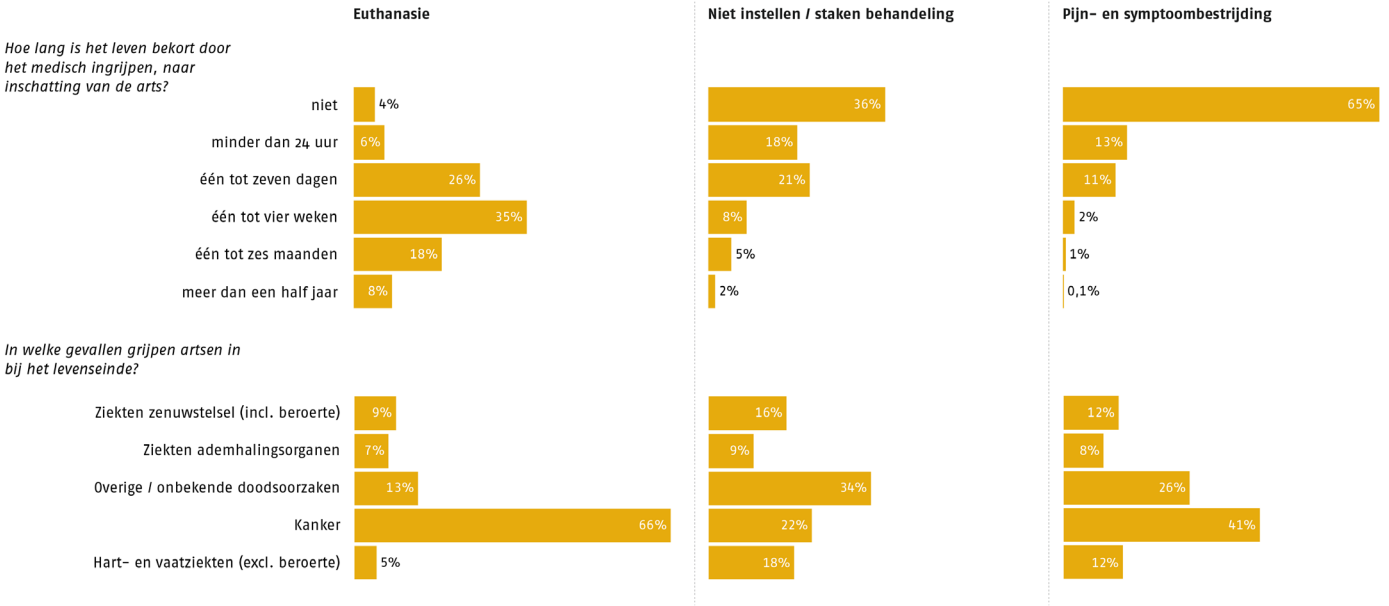
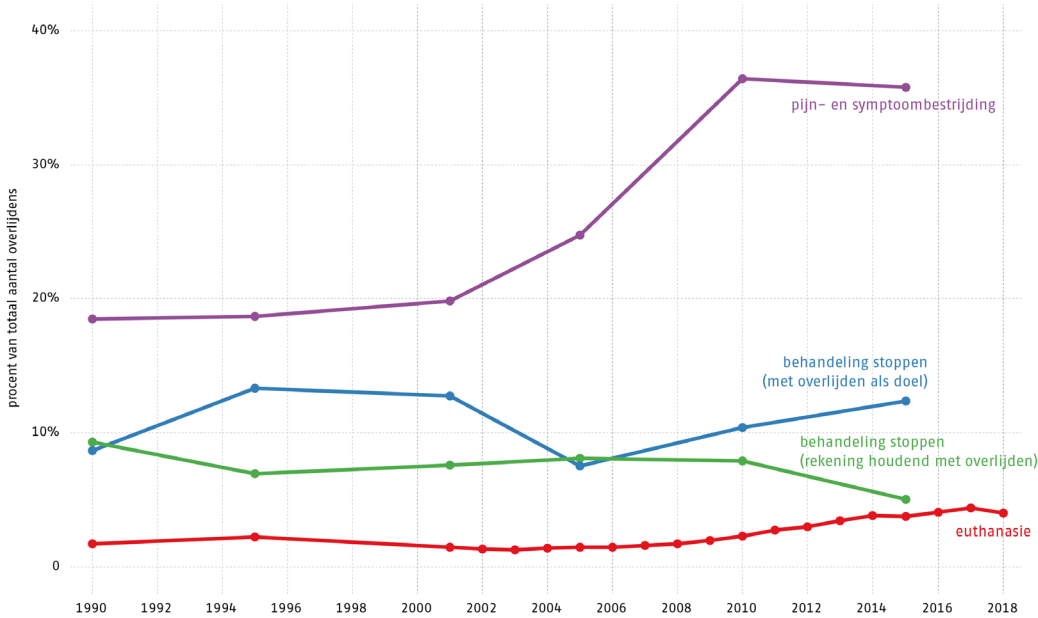
Dit zijn de meest recente gegevens uit het sterfgevallenonderzoek van het CBS (2015). Het aandeel van de huisartsen steeg van 43 procent in 2010 naar 55 procent in 2015.



DE INVLOED VAN ARTSEN OP HET LEVENSEINDE

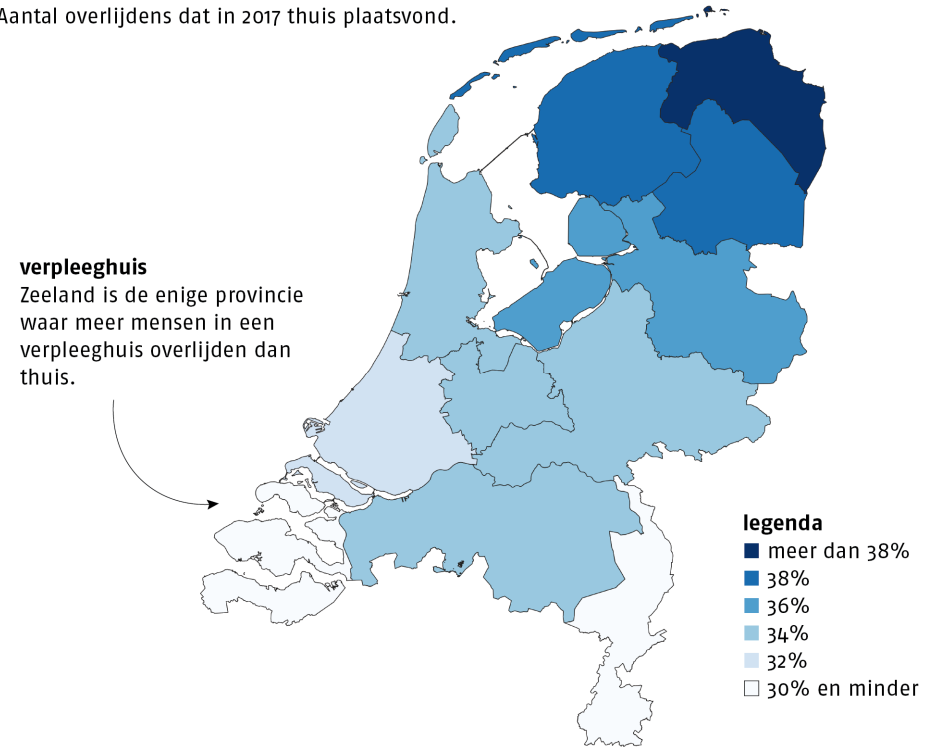
Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat. Ut wisi enim ad minim veniam, quis nostrud exerci tation ullamcorper suscipit lobortis nisl ut aliquip ex ea commodo consequat. Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis at vero eros et accumsan et iusto odio dignissim qui blandit praesent luptatum zzril delenit augue duis dolore te feugait nulla facilisi.

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat. Ut wisi enim ad minim veniam, quis nostrud exerci tation ullamcorper suscipit lobortis nisl ut aliquip



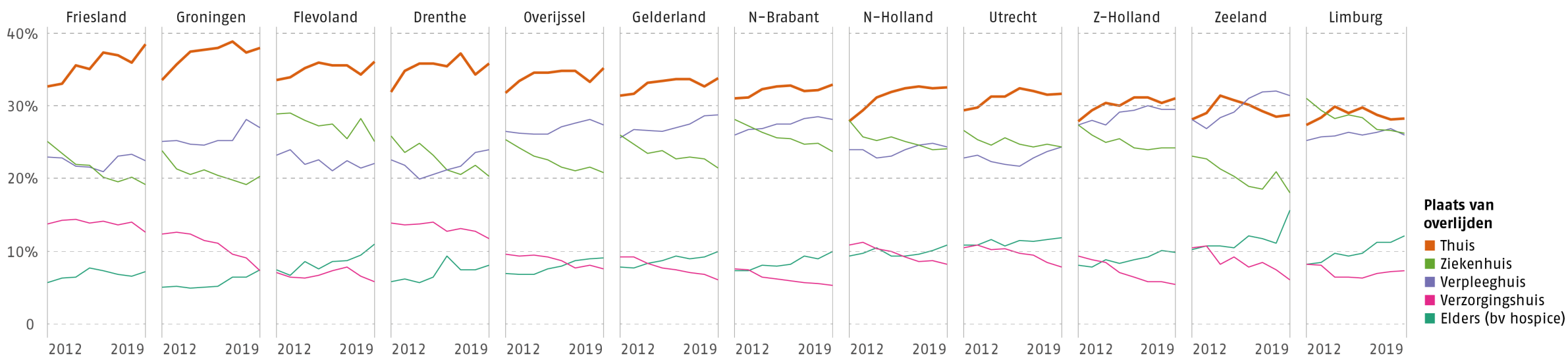
HOE NOORDELIJKER, HOE MEER MENSEN THUIS STERVEN

Aantal overlijdens dat in 2017 thuis plaatsvond.



© 20191031 Sjoerd Mouissie, Nederlands Dagblad. Bron data: CBS

IN FRIESLAND STERVEN DE MEESTE MENSEN THUIS, IN LIMBURG DE MINSTE



© 20201023 Sjoerd Mouissie, Nederlands Dagblad. Bron data: CBS

Mijn workflow

Open data in R

Eeh, R?

- Gratis, open source statistiekprogramma
- Werkt op basis van een script
- Te gebruiken met het programma Rstudio (eveneens gratis en open source)
- CBS heeft een uitstekende koppeling gemaakt tussen R en Statline
- Maar: R is minder gebruik(ers)vriendelijk dan Excel of SPSS, en de leercurve kan steil zijn

CBS-data in R: cbsodataR

```
CA0lonen <- cbs_get_data('82838NED')
```

CBS-data in R: cbsodataR

```
CA0lonen <- cbs_get_data('82838NED', CaoSectoren = "T001020")
```

R: maak een grafiek met x en y

```
ggplot(CAOlonen.permaand,  
       aes(x = Perioden_Date, y = ContractueleLoonkostenPerUur_13)) +  
  geom_line(color = "skyblue4", size = 1.5) +  
  scale_x_date(date_labels = "%b %y", date_breaks = "1 month") +  
  scale_y_continuous(limits = c(0, NA)) +  
  labs(x = "", y = "stijging loonkosten per uur",  
       title = "Ontwikkeling CAO-loonkosten per uur",  
       caption = "bron data: CBS") +  
  theme_ND_krant()
```

R: het moet een lijngrafiek zijn

```
ggplot(CAOlonen.permaand,  
       aes(x = Perioden Date, y = ContractueleLoonkostenPerUur_13)) +  
  geom_line(color = "skyblue4", size = 1.5) +  
  scale_x_date(date_labels = "%b %y", date_breaks = "1 month") +  
  scale_y_continuous(limits = c(0, NA)) +  
  labs(x = "", y = "stijging loonkosten per uur",  
       title = "Ontwikkeling CAO-loonkosten per uur",  
       caption = "bron data: CBS") +  
  theme_ND_krant()
```

R: op de x-as staan data, per maand

```
ggplot(CAOlonen.permaand,  
       aes(x = Perioden_Date, y = ContractueleLoonkostenPerUur_13)) +  
  geom_line(color = "skyblue4", size = 1.5) +  
  scale_x_date(date_labels = "%b %y", date_breaks = "1 month") +  
  scale_y_continuous(limits = c(0, NA)) +  
  labs(x = "", y = "stijging loonkosten per uur",  
       title = "Ontwikkeling CAO-loonkosten per uur",  
       caption = "bron data: CBS") +  
  theme_ND_krant()
```

R: de y-as moet op nul beginnen

```
ggplot(CAOlonen.permaand,  
       aes(x = Perioden_Date, y = ContractueleLoonkostenPerUur_13)) +  
  geom_line(color = "skyblue4", size = 1.5) +  
  scale_x_date(date_labels = "%b %y", date_breaks = "1 month") +  
  scale_y_continuous(limits = c(0, NA)) +  
  labs(x = "", y = "stijging loonkosten per uur",  
       title = "Ontwikkeling CAO-loonkosten per uur",  
       caption = "bron data: CBS") +  
  theme_ND_krant()
```

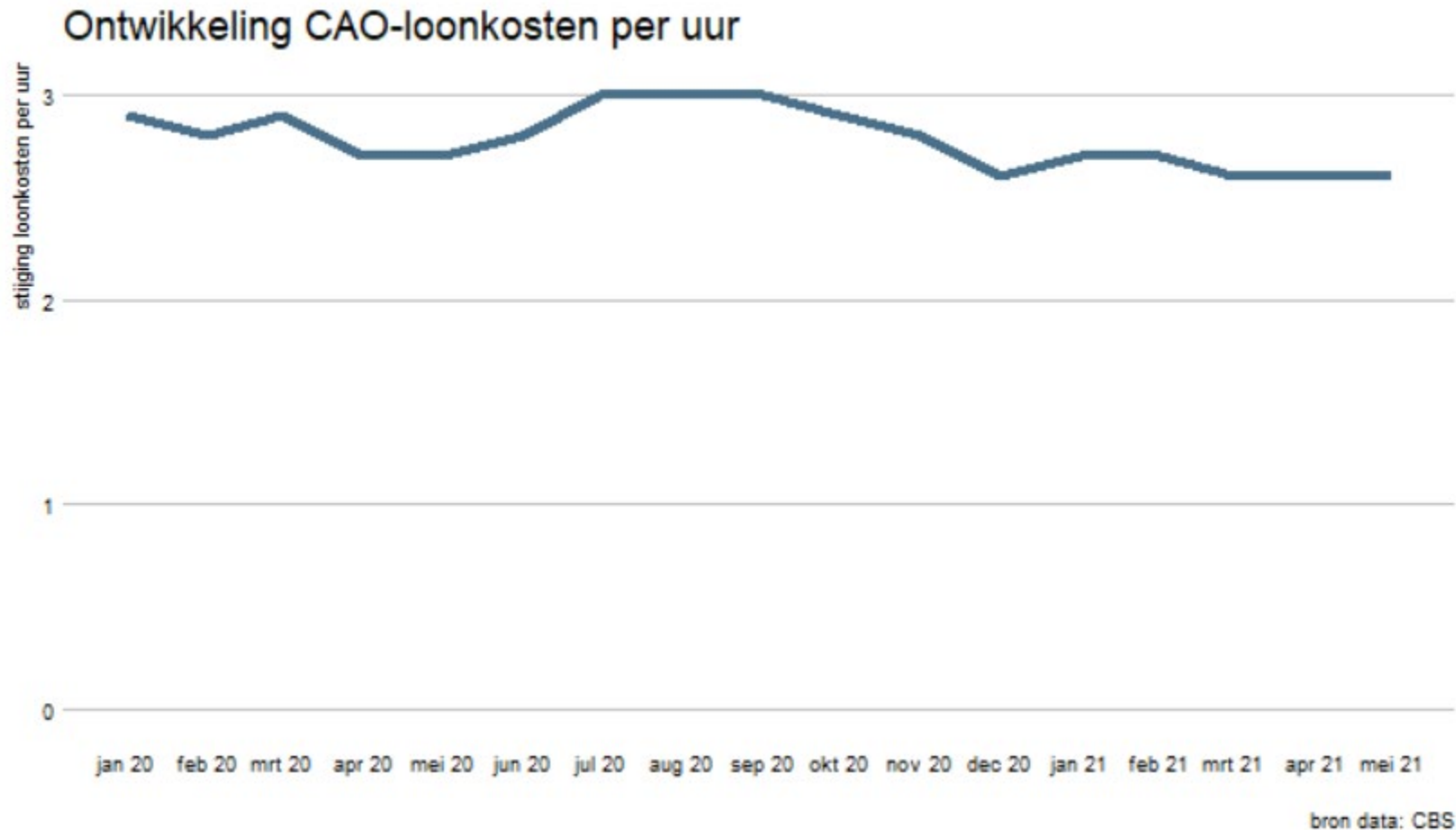
R: titel grafiek, labels bij assen, bron

```
ggplot(CAOlonen.permaand,  
       aes(x = Perioden_Date, y = ContractueleLoonkostenPerUur_13)) +  
  geom_line(color = "skyblue4", size = 1.5) +  
  scale_x_date(date_labels = "%b %y", date_breaks = "1 month") +  
  scale_y_continuous(limits = c(0, NA)) +  
  labs(x = "", y = "stijging loonkosten per uur",  
       title = "Ontwikkeling CAO-loonkosten per uur",  
       caption = "bron data: CBS") +  
  theme_ND_krant()
```

R: ND-format voor Adobe Illustrator

```
ggplot(CAOlonen.permaand,  
       aes(x = Perioden_Date, y = ContractueleLoonkostenPerUur_13)) +  
  geom_line(color = "skyblue4", size = 1.5) +  
  scale_x_date(date_labels = "%b %y", date_breaks = "1 month") +  
  scale_y_continuous(limits = c(0, NA)) +  
  labs(x = "", y = "stijging loonkosten per uur",  
       title = "Ontwikkeling CAO-loonkosten per uur",  
       caption = "bron data: CBS") +  
  theme_ND_krant()
```

R: resultaat voor Adobe-bewerking



Waarom al die moeite?

- Het grote voordeel van R: herhaalbaarheid
- Als er nieuwe cijfers verschijnen, is het vaak voldoende een script opnieuw uit te voeren
- Geen downloads, versies van bestanden, opnieuw opschonen en grafieken opnieuw handmatig maken
- Voorbeeld: project 'Sterven in Nederland' lag een jaar op de plank totdat er gepubliceerd kon worden. In de tussentijd verschenen er nieuwe cijfers en kon de grafiek moeiteloos bijgewerkt worden.