

# Data visualiseren

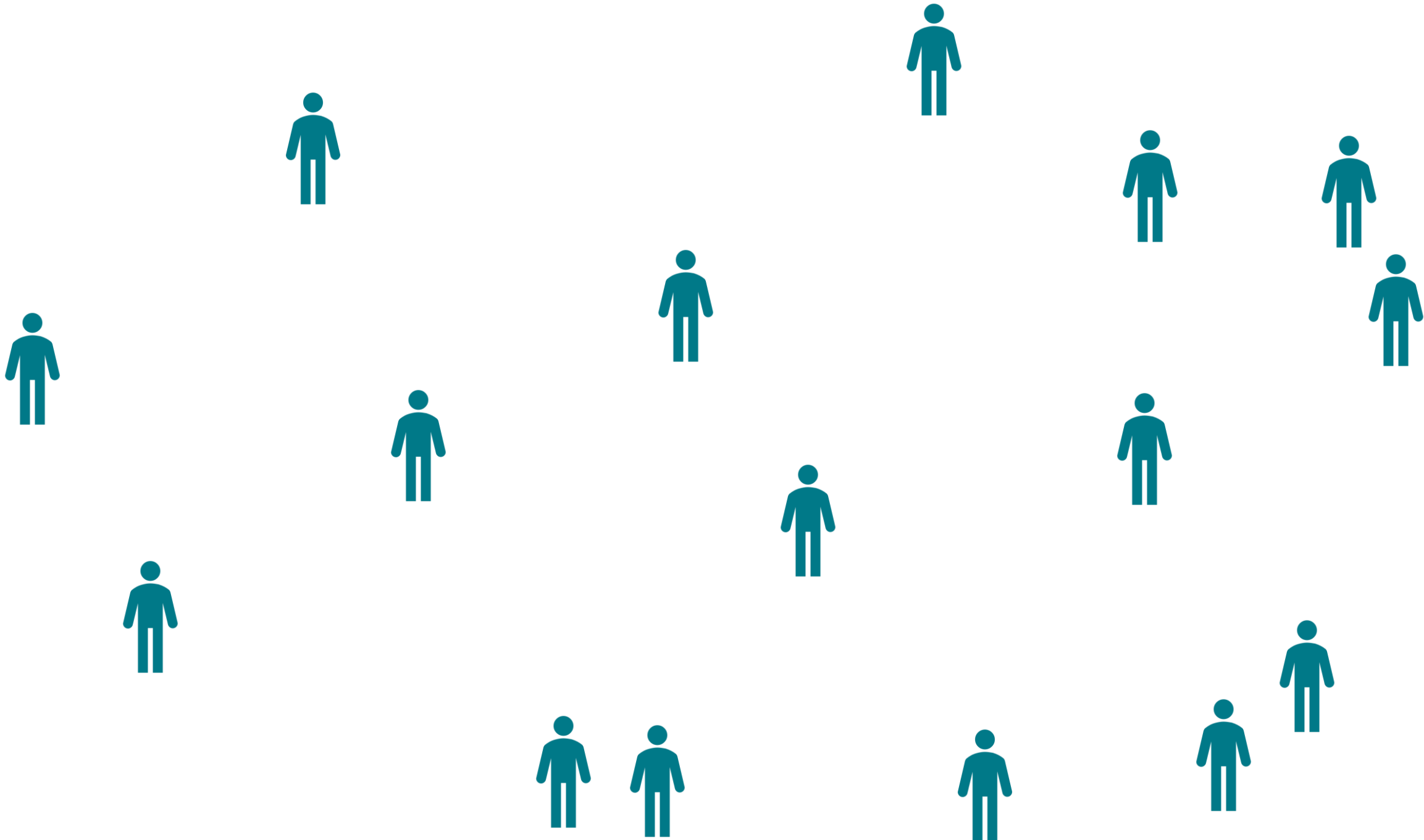
Netwerkdag datageletterdheid

30 augustus 2023

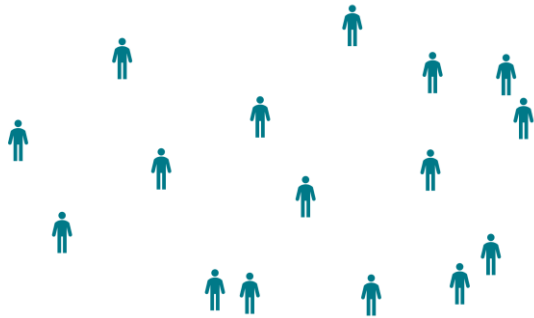
# POPULATIE EN STEEKPROEF



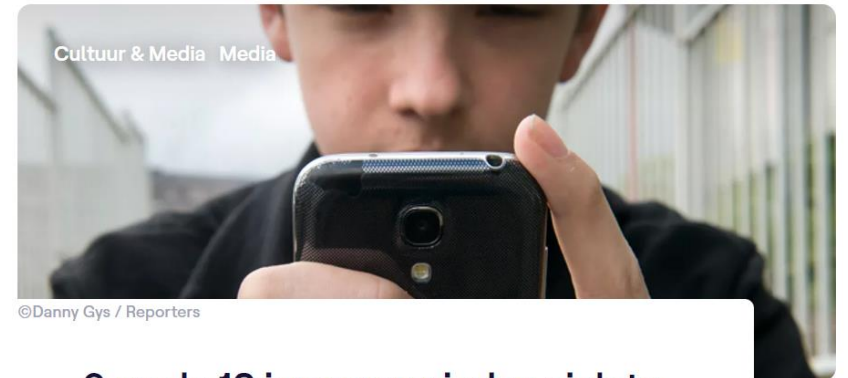




# Digimeter 2022



vrtnws



©Danny Gys / Reporters

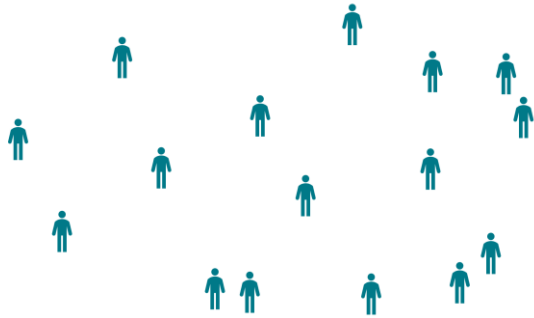
**6 op de 10 jongeren vinden zich te afhankelijk van smartphone:  
"Technologie is als een wild dier in huis, je moet dat temmen"**

Meer dan 6 op de 10 Vlaamse jongeren tussen 18 en 24 jaar oud vinden dat ze te afhankelijk geworden zijn van hun smartphone.

<https://www.vrt.be/vrtnws/nl/2023/03/06/digimeter-2023/>  
Inhoud van de webpagina gebruikt met toestemming.

Bron: Imec.Digimeter

# Digimeter 2022



## Bevraging

Ik vind van mezelf dat ik te veel tijd  
spendeer op mijn smartphone

Ik voel me verslaafd aan mijn smartphone

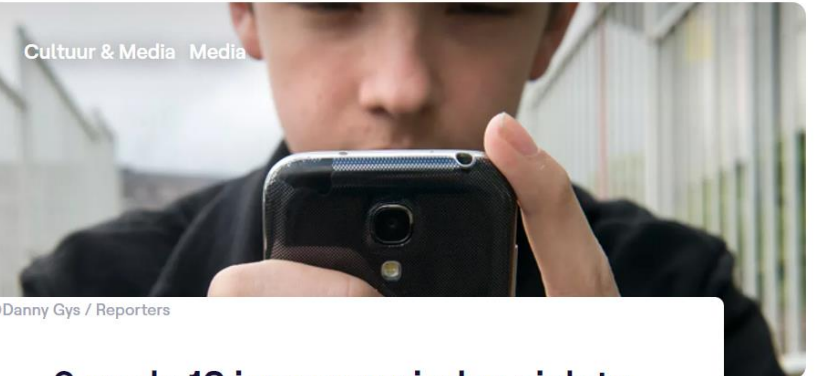
**62%**

<https://www.vrt.be/vrtnws/nl/2023/03/06/digimeter-2023/>  
<https://www.imec.be/nl/kennisuitwisseling/techmeters/digimeter/digimeter-2022>

Bron: Imec.Digimeter



vrtnws



Cultuur & Media Media

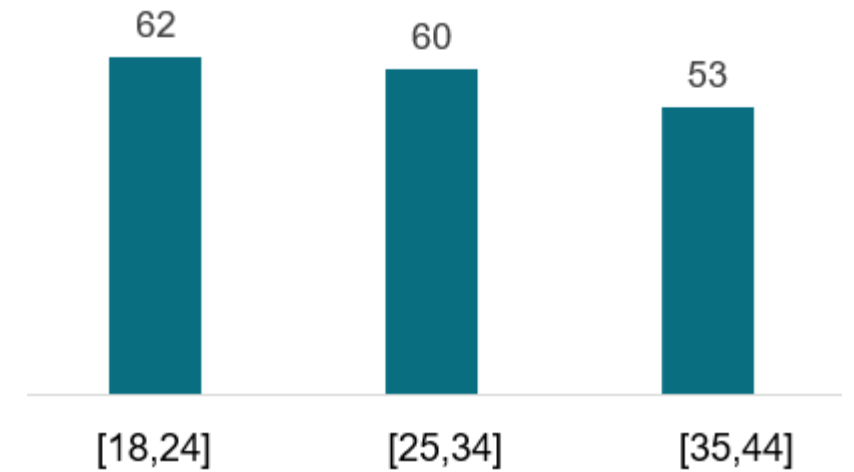
©Danny Gys / Reporters

**6 op de 10 jongeren vinden zich te afhankelijk van smartphone:**  
"Technologie is als een wild dier in huis, je moet dat temmen"

Meer dan 6 op de 10 Vlaamse jongeren tussen 18 en 24 jaar oud vinden dat ze te afhankelijk geworden zijn van hun smartphone.

# VISUALISATIE

EEN BEELD ZEGT MEER DAN 1000 WOORDEN



# Digimeter 2022 Smartphone-afhankelijkheid

Bron: Imec.Digimeter

		16-24	25-34	35-44	45-54	55-64	65-74	75+	Totaal
Smartphone-afhankelijk	2018	48%	49%	40%	22%	11%	10%		<b>31%</b>
	2019	52%	48%	46%	29%	17%	12%		<b>33%</b>
	2020	53%	60%	49%	31%	16%	14%		<b>35%</b>
	2021	51%	60%	43%	31%	15%	9%		<b>33%</b>
	<b>2022</b>	<b>62%</b>	<b>60%</b>	<b>53%</b>	<b>42%</b>	<b>20%</b>	<b>18%</b>	<b>15%</b>	<b>40%</b>

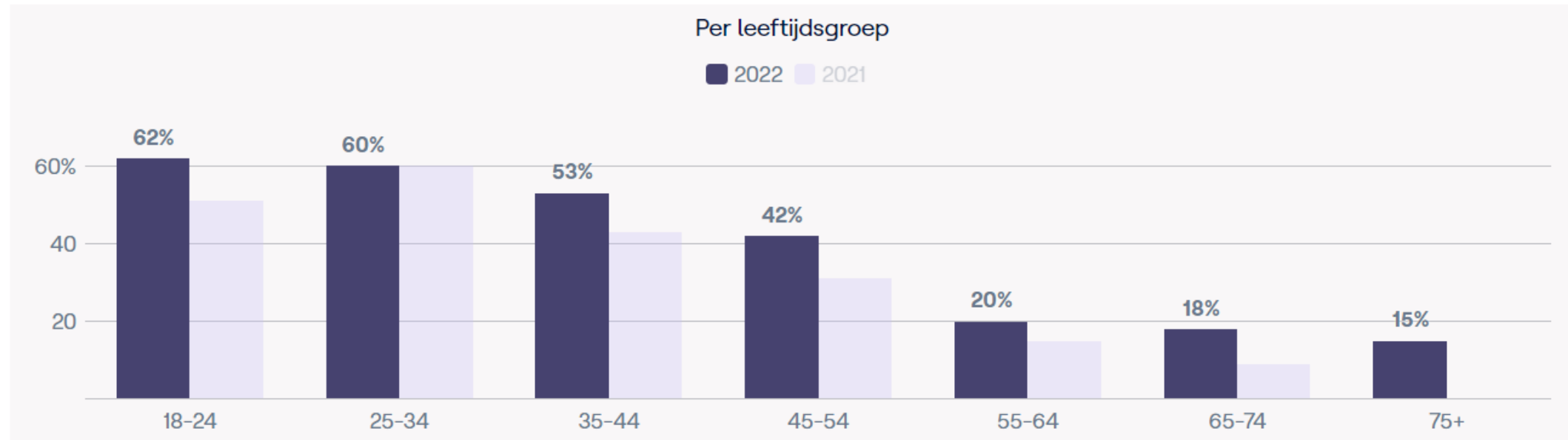




# Digimeter 2022 Smartphone-afhankelijkheid

Bron: Imec.Digimeter

		16-24	25-34	35-44	45-54	55-64	65-74	75+	Totaal
Smartphone-afhankelijk	2018	48%	49%	40%	22%	11%	10%		31%
	2019	52%	48%	46%	29%	17%	12%		33%
	2020	53%	60%	49%	31%	16%	14%		35%
	2021	51%	60%	43%	31%	15%	9%		33%
	2022	62%	60%	53%	42%	20%	18%	15%	40%



# Originele vormgeving - Belgische post



Staafdiagram in originele vorm.

Leerlingen zouden in de les kunnen narekenen of de lengte van de staven verhoudingsgewijs juist zijn.

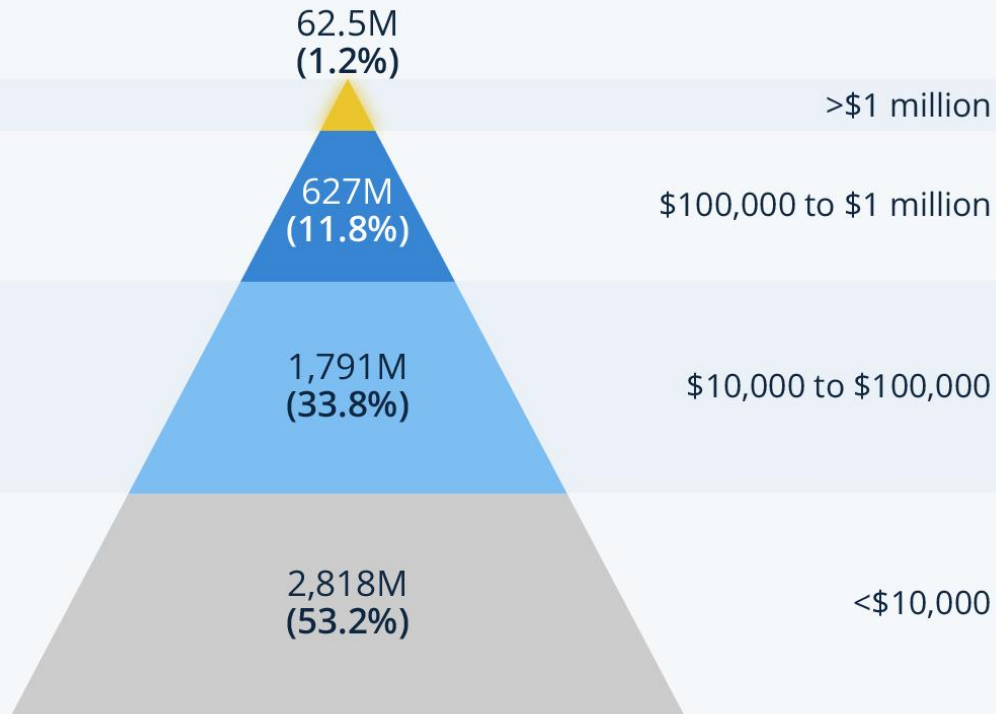
Het nulpunt op de y-as is niet gegeven. Ligt het waar je op het eerste zicht zou vermoeden?

Afbeelding gebruikt op basis van citaatrecht.

# Originele vormgeving - Global Wealth

## The Global Wealth Pyramid

An overview of global wealth distribution in 2021



Number of adults (% of world adults)

Source: Credit Suisse 2022 Global Wealth Report

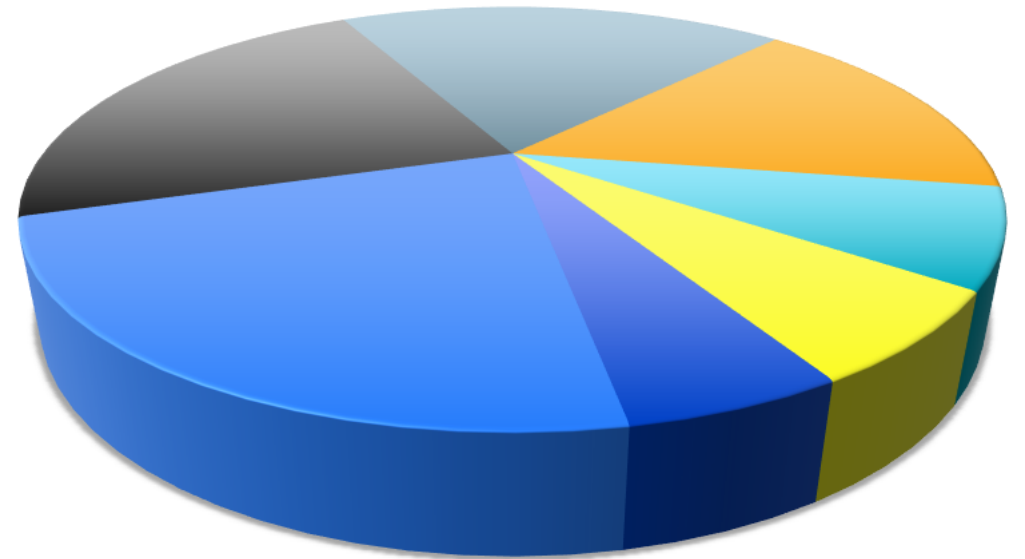


statista

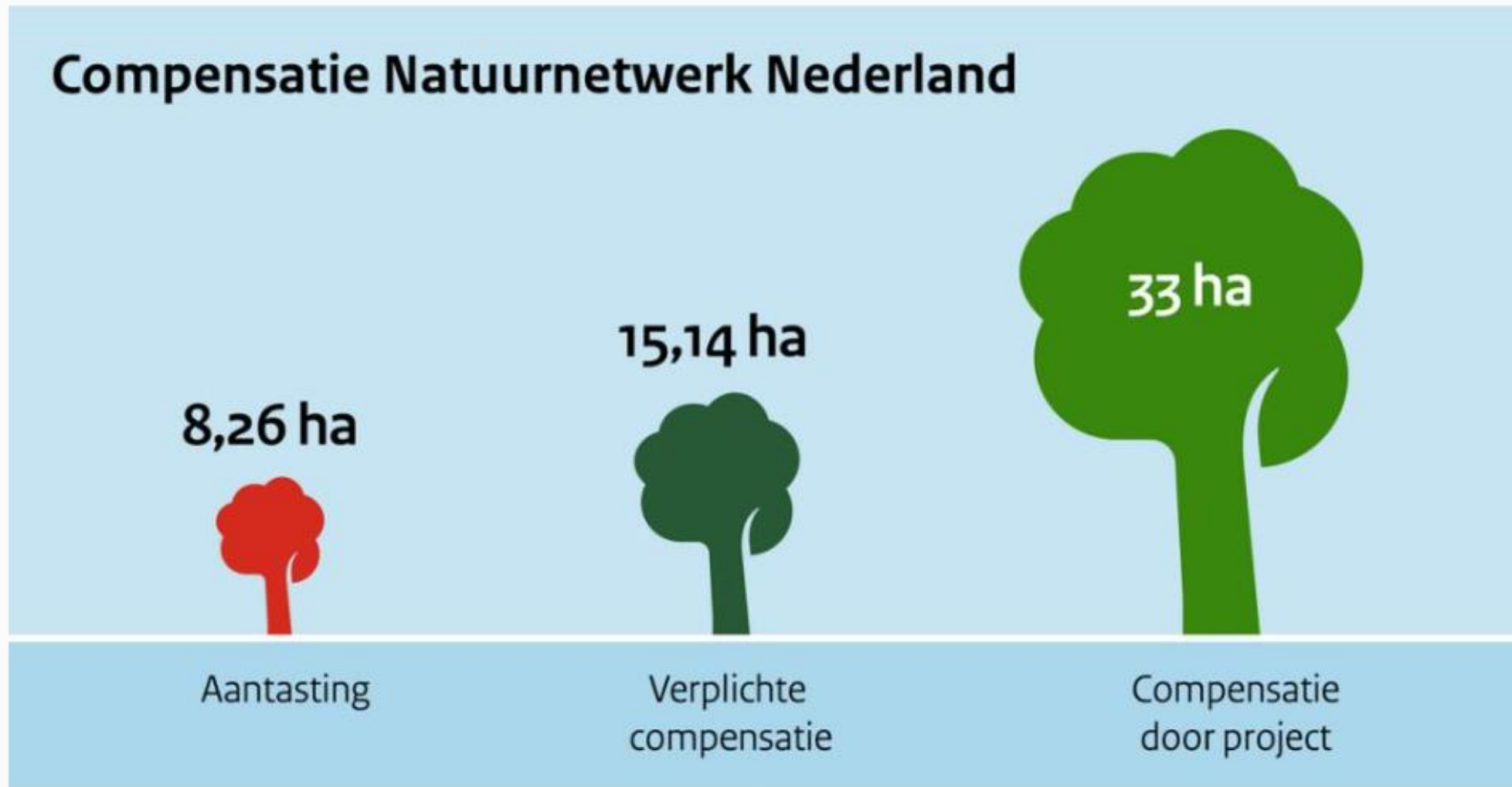
Er werd gekozen voor een driehoek omwille van de gelijkenis met een piramide.

Leerlingen zouden in de les kunnen narekenen of de oppervlaktes van de gebieden verhoudingsgewijs juist zijn. Ze maken daarbij gebruik van de formules voor de oppervlaktes van een driehoek en een trapezium.

## MISLEIDENDE GRAFIEKEN



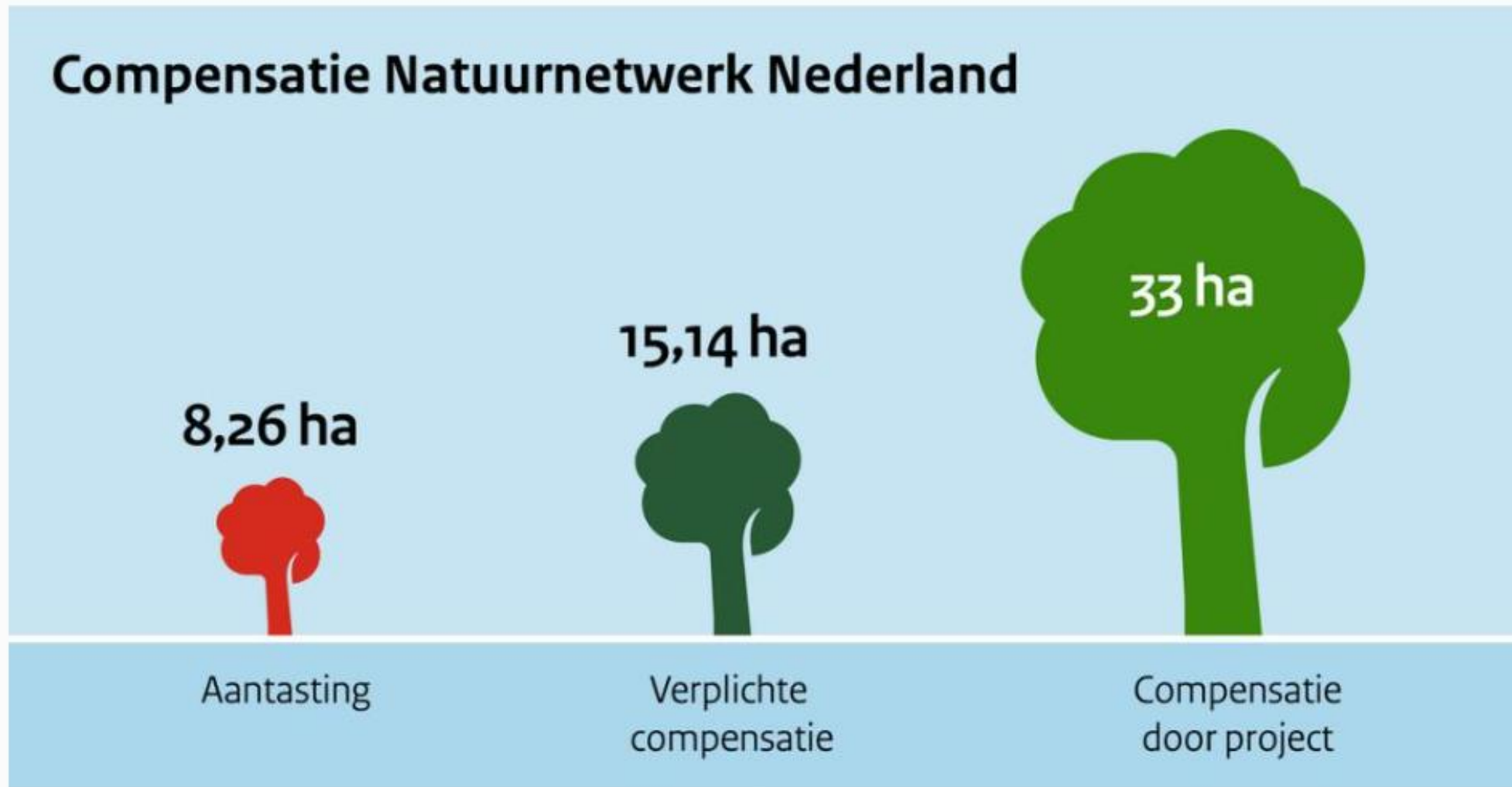
# Misleidende grafieken – Foute grafieken



De originele grafiek

Inhoud van de webpagina gebruikt met toestemming.

# Misleidende grafieken - Rijkswaterstaat



De originele grafiek

Inhoud van de webpagina gebruikt met toestemming.



# CORRECTIE

## Compensatie Natuurnetwerk Nederland

8,26 ha



Aantasting

15,14 ha



Verplichte  
compensatie

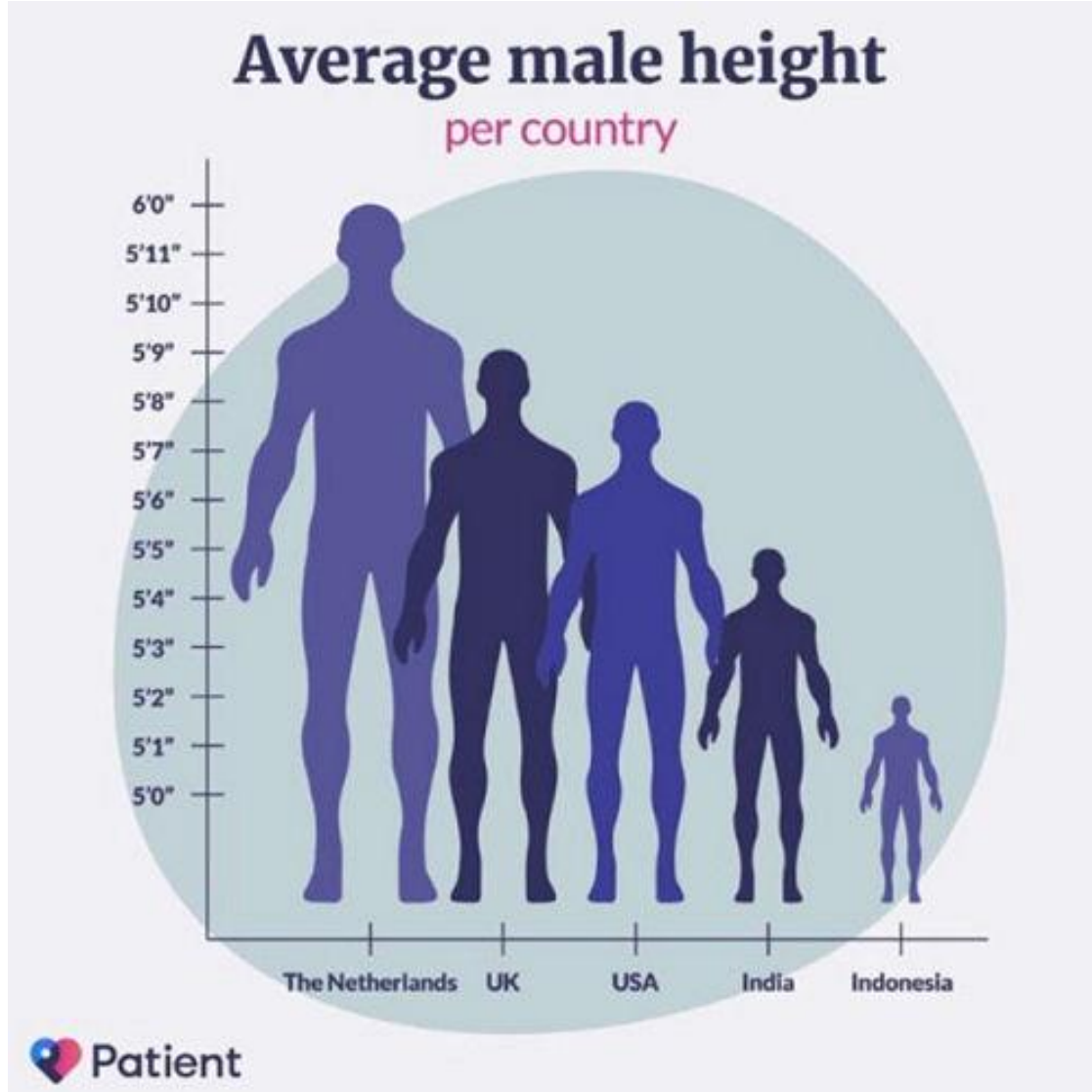
33 ha



Compensatie  
door project

De gecorrigeerde grafiek

# Misleidende grafieken - Gemiddelde lengte man



Inhoud van de webpagina gebruikt met toestemming.

<https://patient.info/news-and-features/whats-the-average-height-for-men>  
Afbeelding gebruikt op basis van citaatrecht.

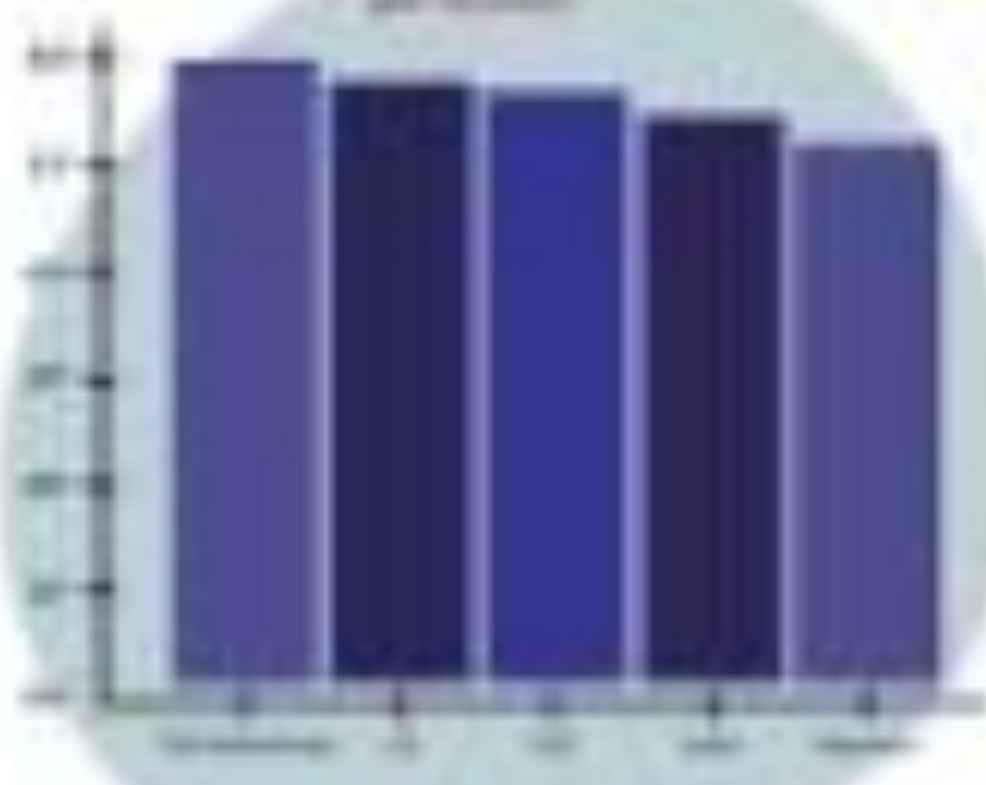




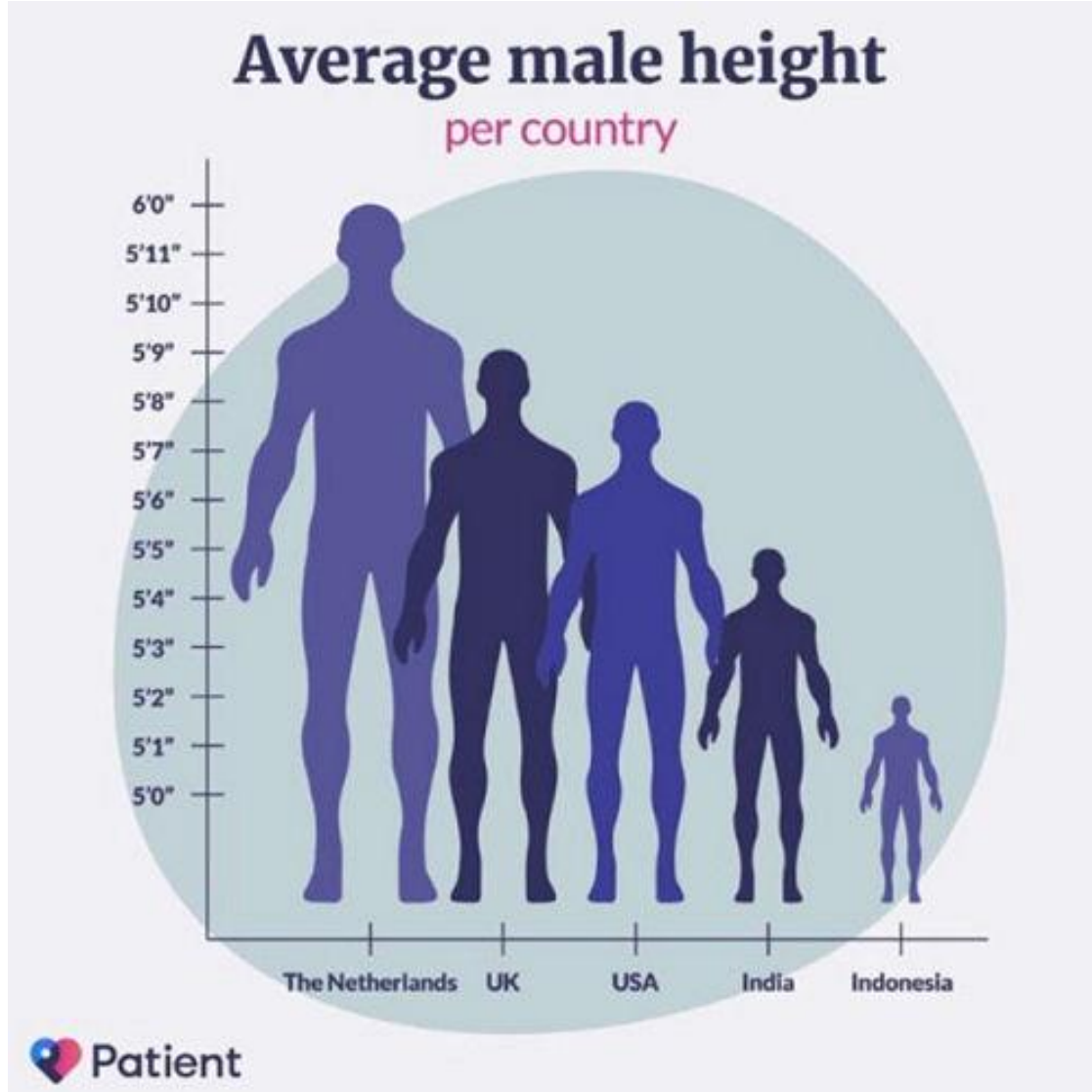
Average male height



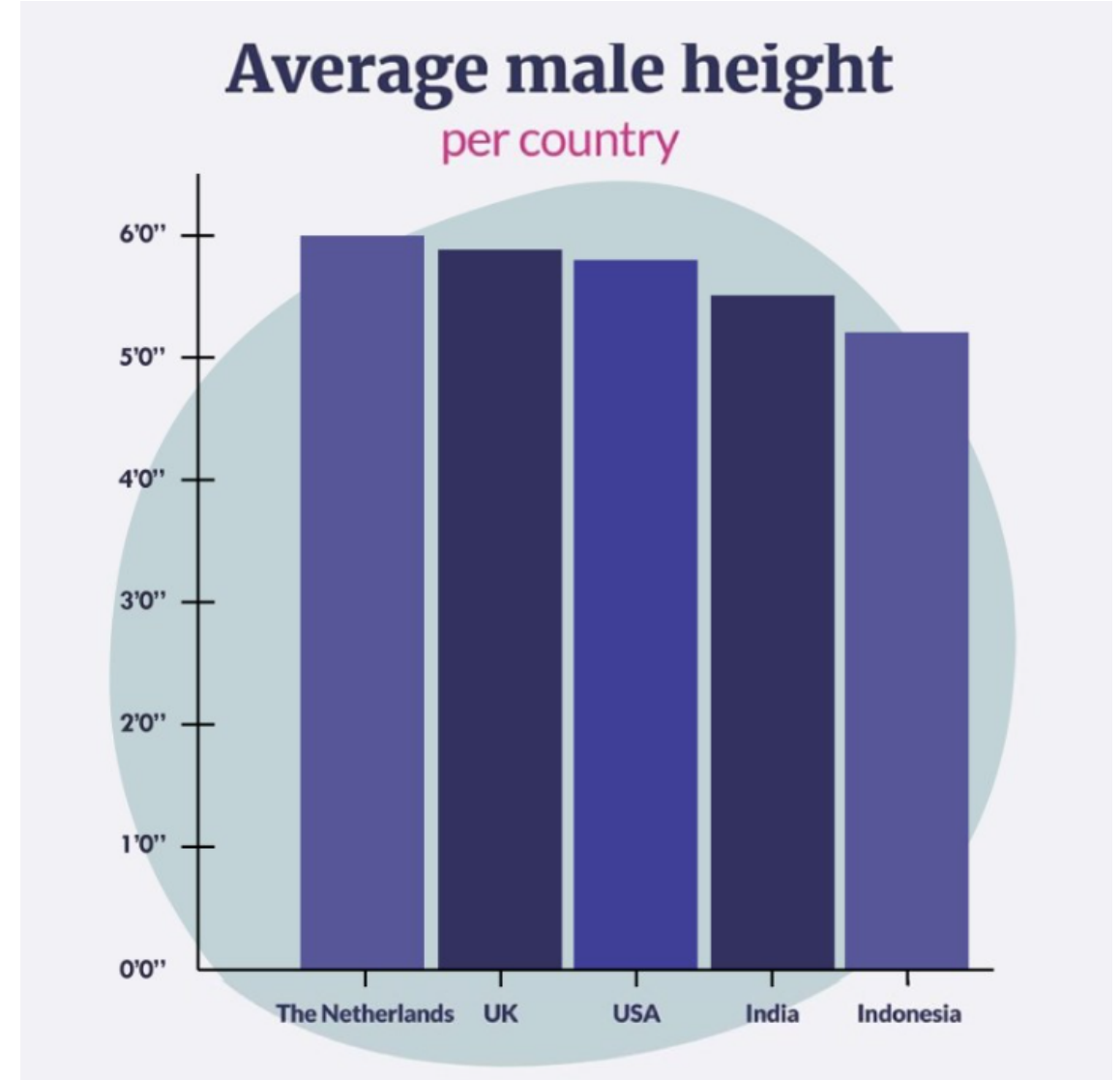
Average male height



# Misleidende grafieken - Gemiddelde lengte man

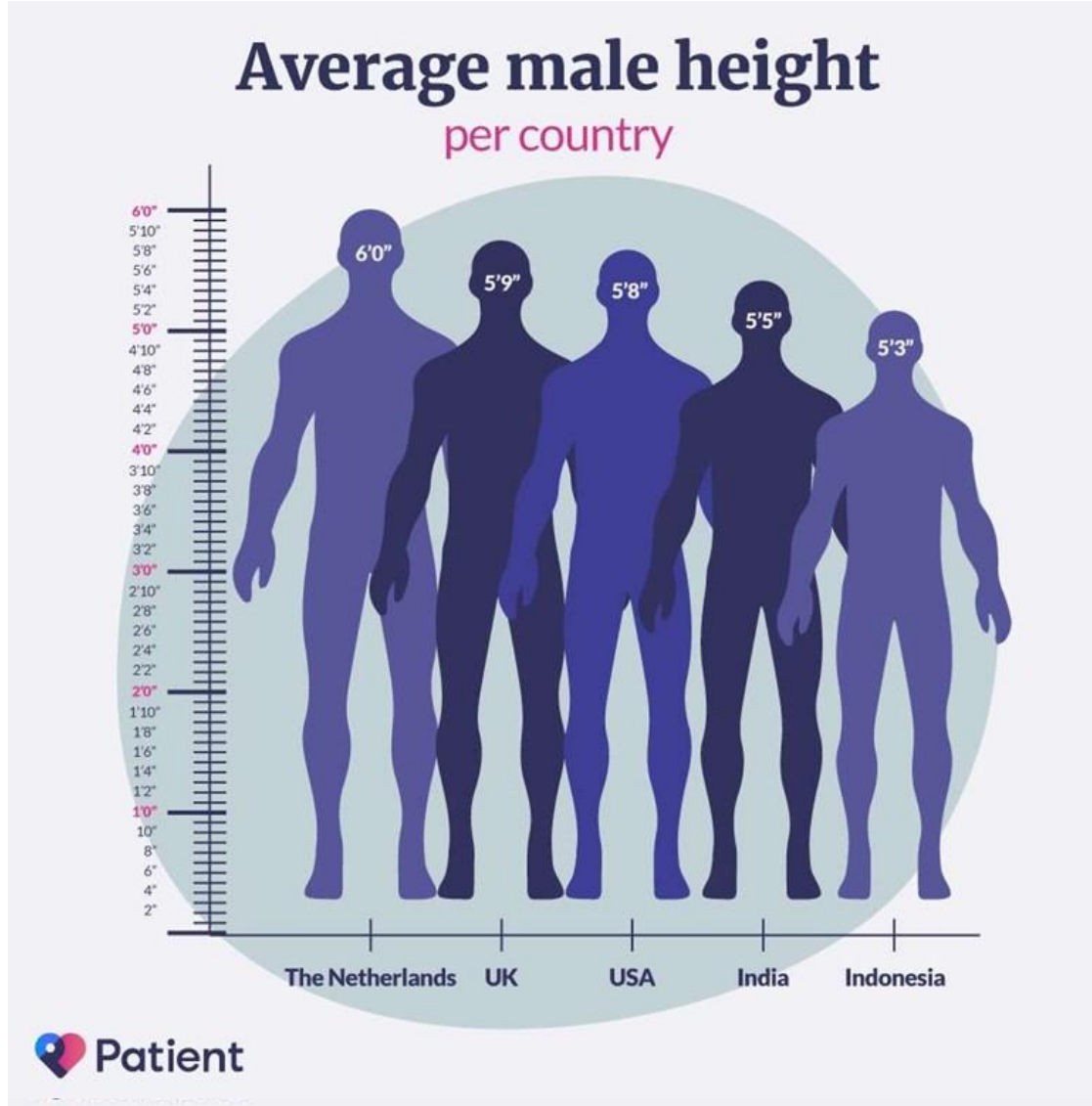


<https://patient.info/news-and-features/whats-the-average-height-for-men>  
Afbeelding gebruikt op basis van citaatrecht.

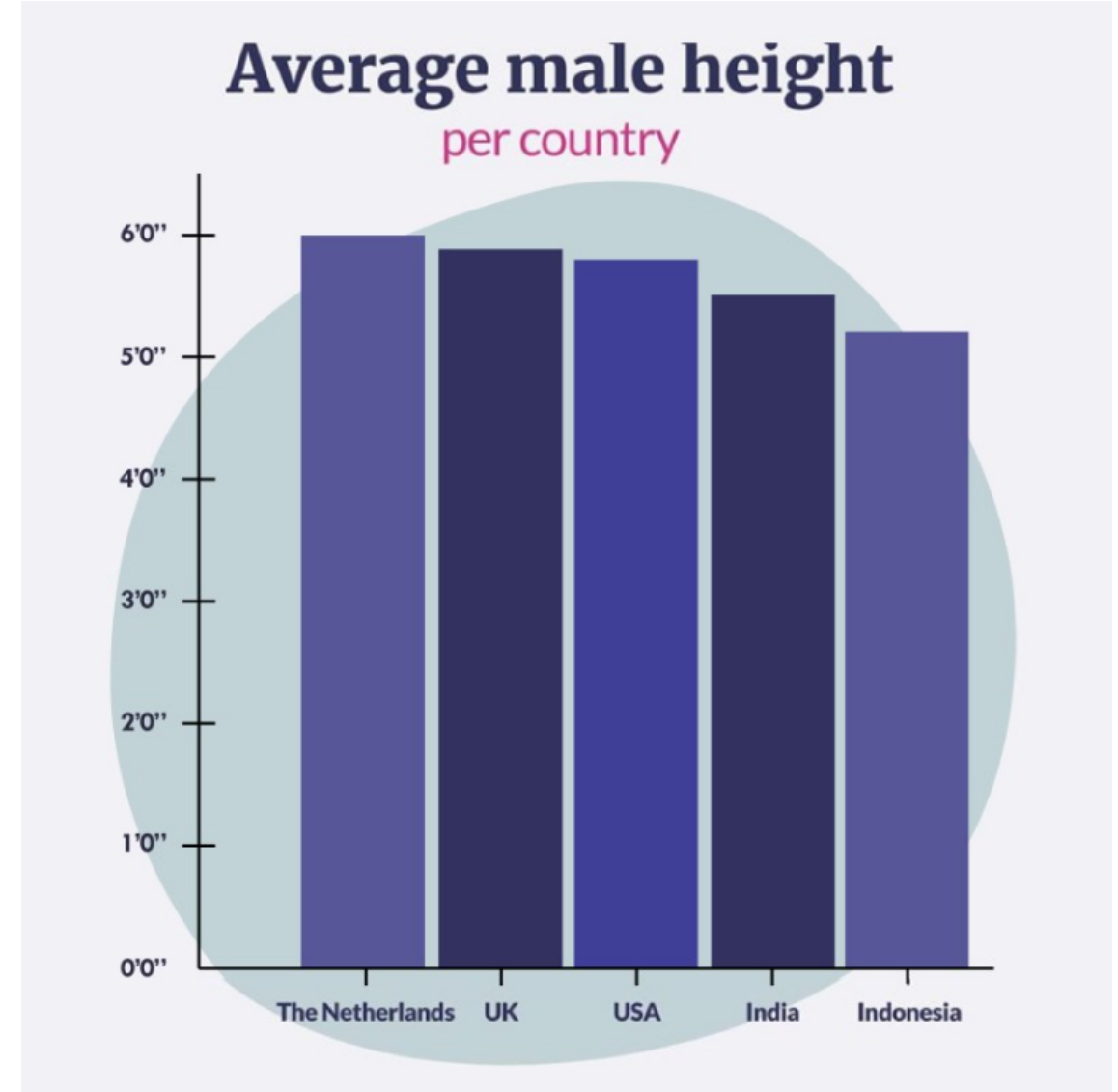


<https://nieuwscheckers.nl/gezondheidsorganisatie-patient-uk-maakt-nederlanders-reusachtig-groot/>

# Misleidende grafieken - Gemiddelde lengte man



<https://patient.info/news-and-features/whats-the-average-height-for-men>  
Afbeelding gebruikt op basis van citaatrecht.

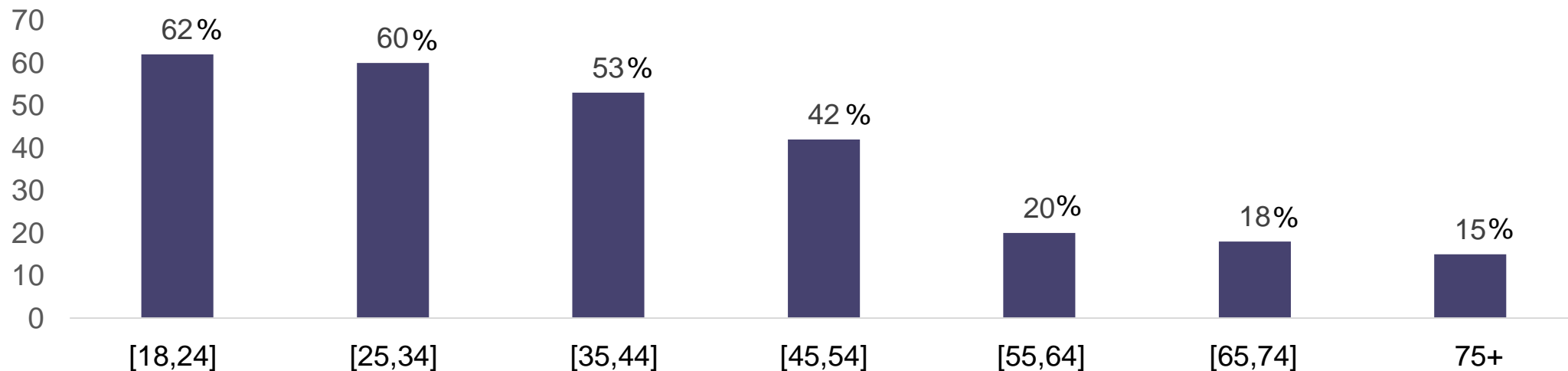


<https://nieuwscheckers.nl/gezondheidsorganisatie-patient-uk-maakt-nederlanders-reusachtig-groot/>

# Misleidende grafieken – Foute grafieken

Bron: Imec.Digimeter

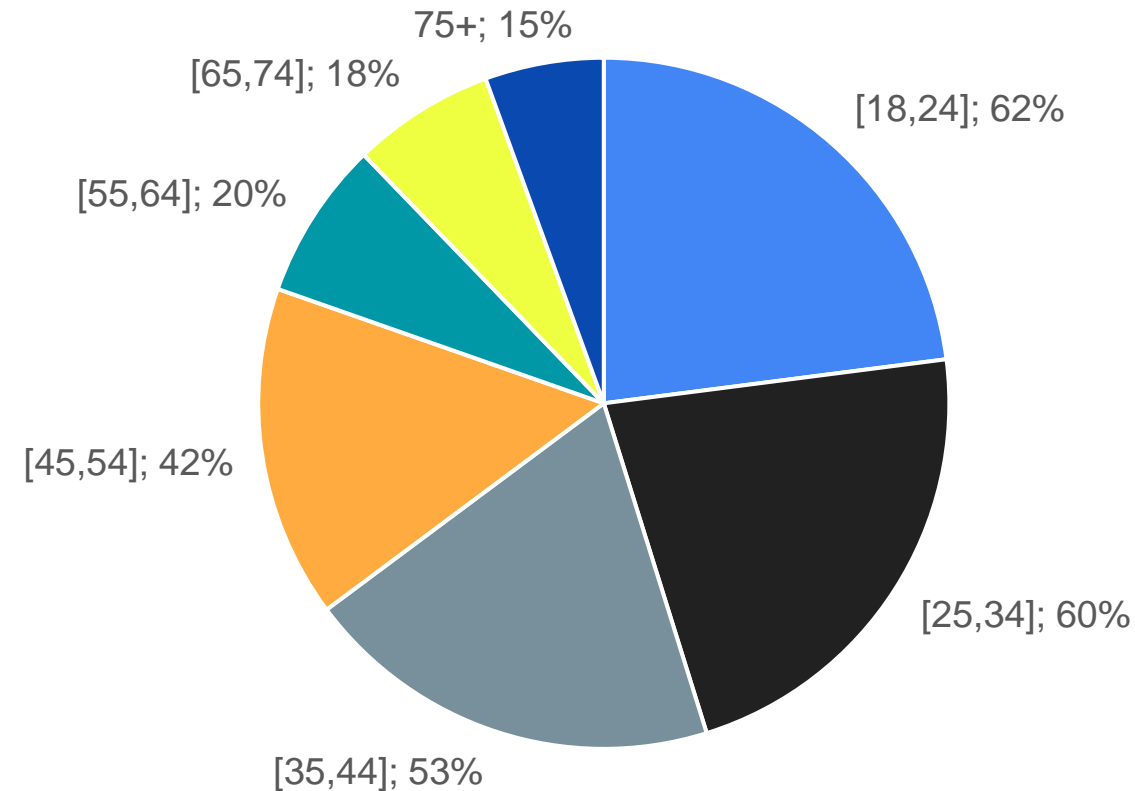
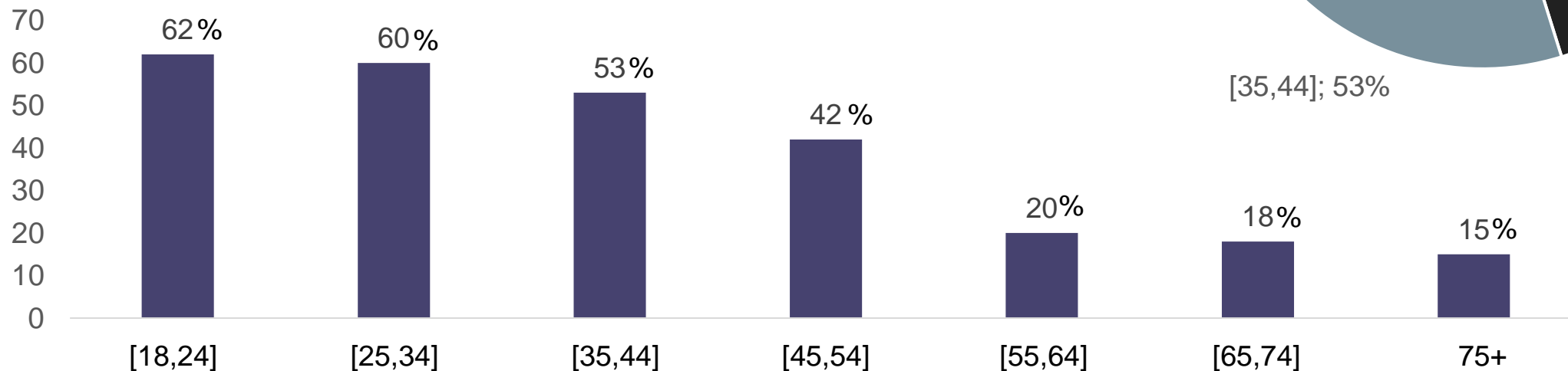
		16-24	25-34	35-44	45-54	55-64	65-74	75+	Totaal
Smartphone-afhankelijk	2018	48%	49%	40%	22%	11%	10%		31%
	2019	52%	48%	46%	29%	17%	12%		33%
	2020	53%	60%	49%	31%	16%	14%		35%
	2021	51%	60%	43%	31%	15%	9%		33%
	2022	62%	60%	53%	42%	20%	18%	15%	40%



# Misleidende grafieken – Foute grafieken

Bron: Imec.Digimeter

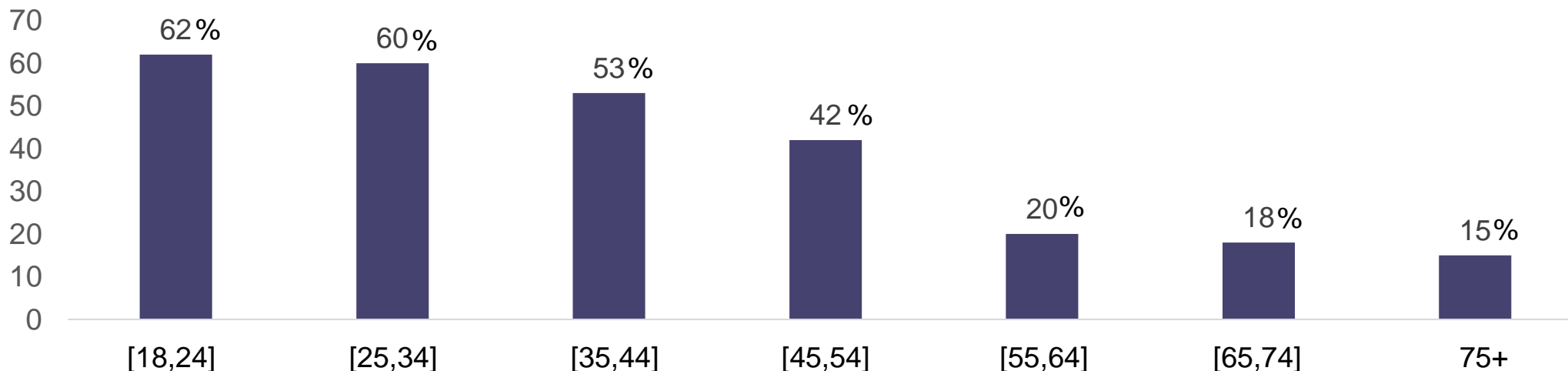
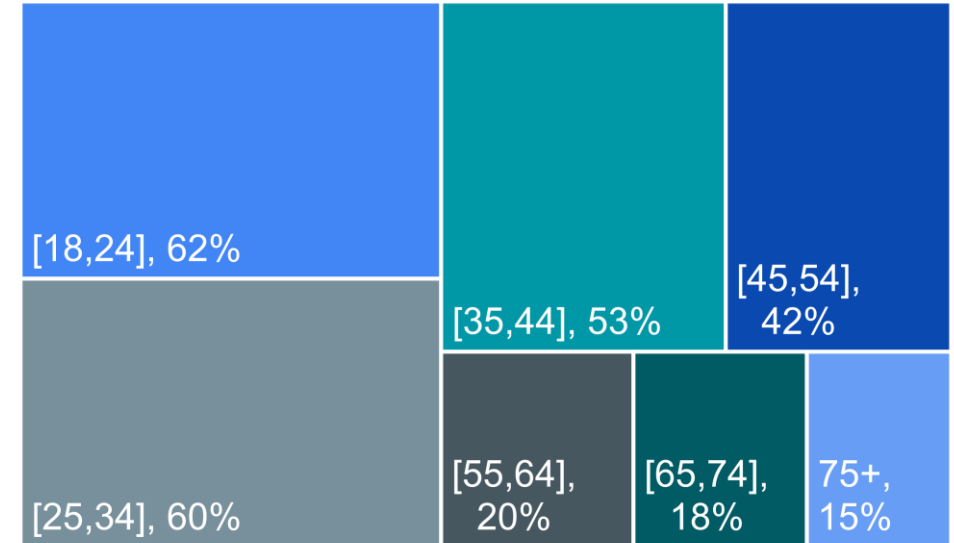
Met Excel kan je van deze data heel gemakkelijk een foutieve voorstelling maken: een taartdiagram (schijfdiagram) is hier niet van toepassing.



# Misleidende grafieken – Foute grafieken

Bron: Imec.Digimeter

Met Excel kan je van deze data heel gemakkelijk een foutieve voorstelling maken: een treemapdiagram is hier niet van toepassing.





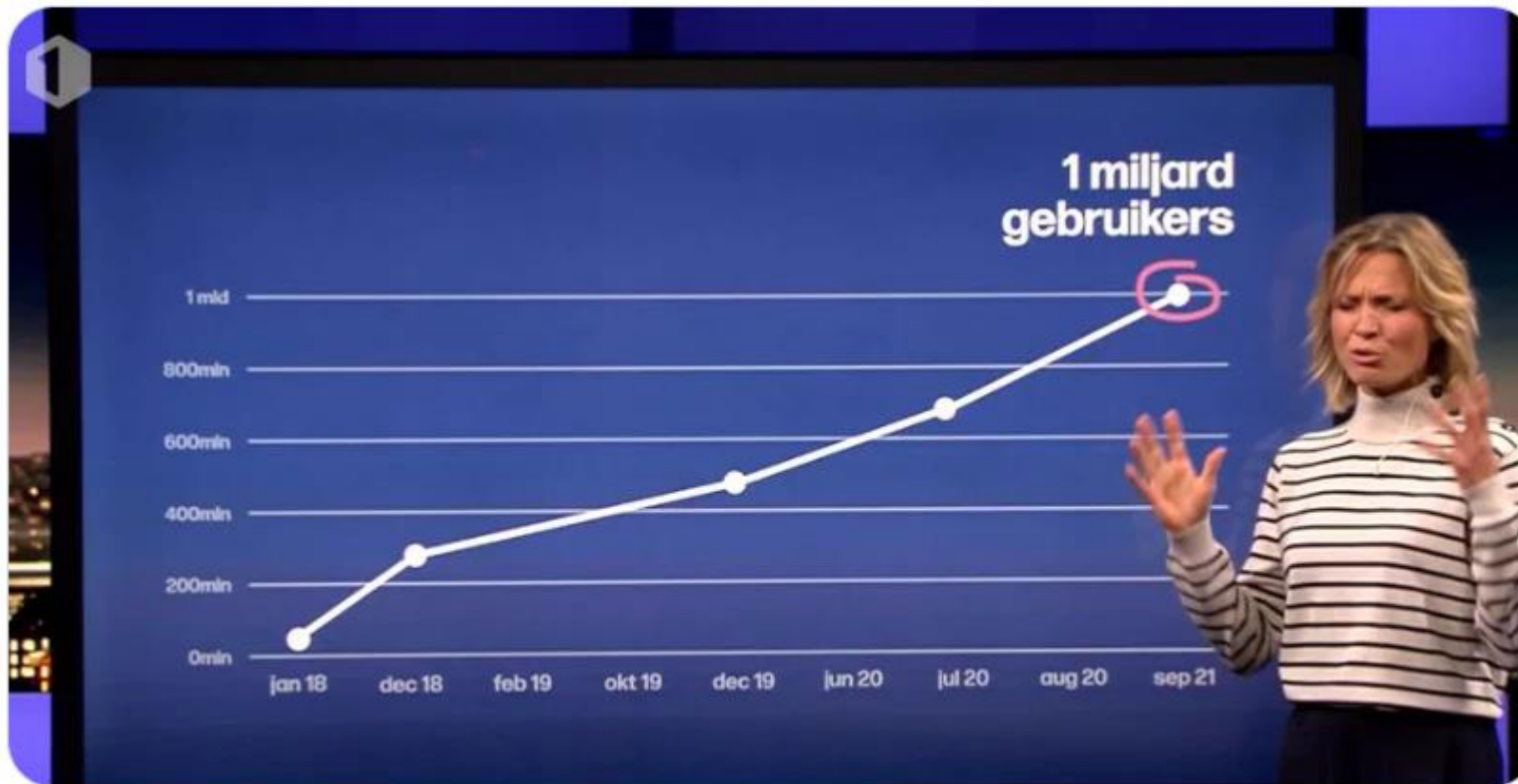
# Misleidende grafieken – Foute grafieken



**Philippe Smet** @pfsmet · Sep 28, 2021

...

Ik zou hier liever niet over moeten zeuren, maar willen jullie de [#grafieken](#) altijd dubbelchecken, [@vrtnews](#)? In deze grafiek is de tijd-as niet gelijkmatig: die verspringt soms met 13 maanden, soms maar met 1 maand. De expressie van [@GoedeleWachters](#) is toeval, maar treffend. 📈 1/



Tweet gebruikt op basis van citaatrecht.

# Misleidende grafieken – Foute grafieken



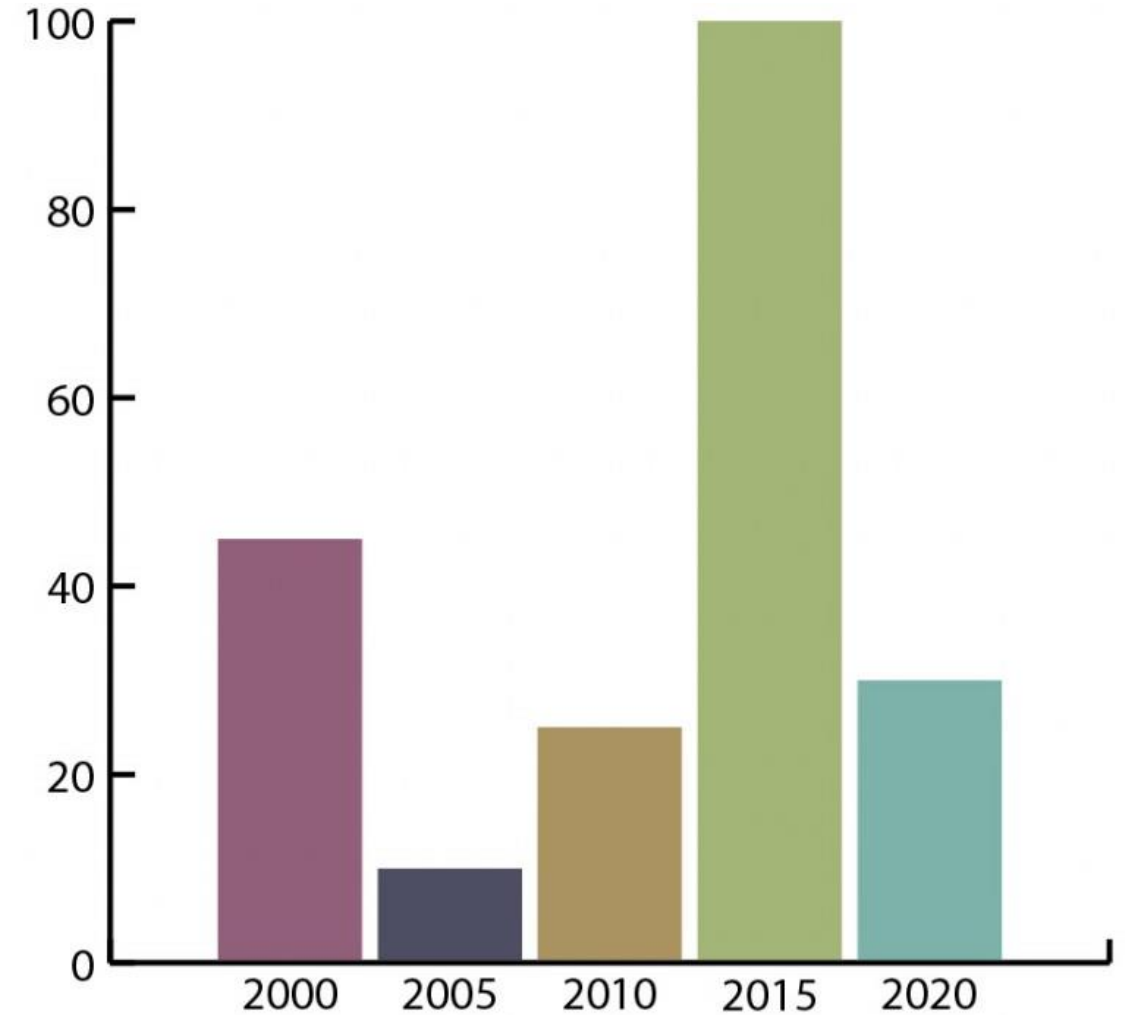
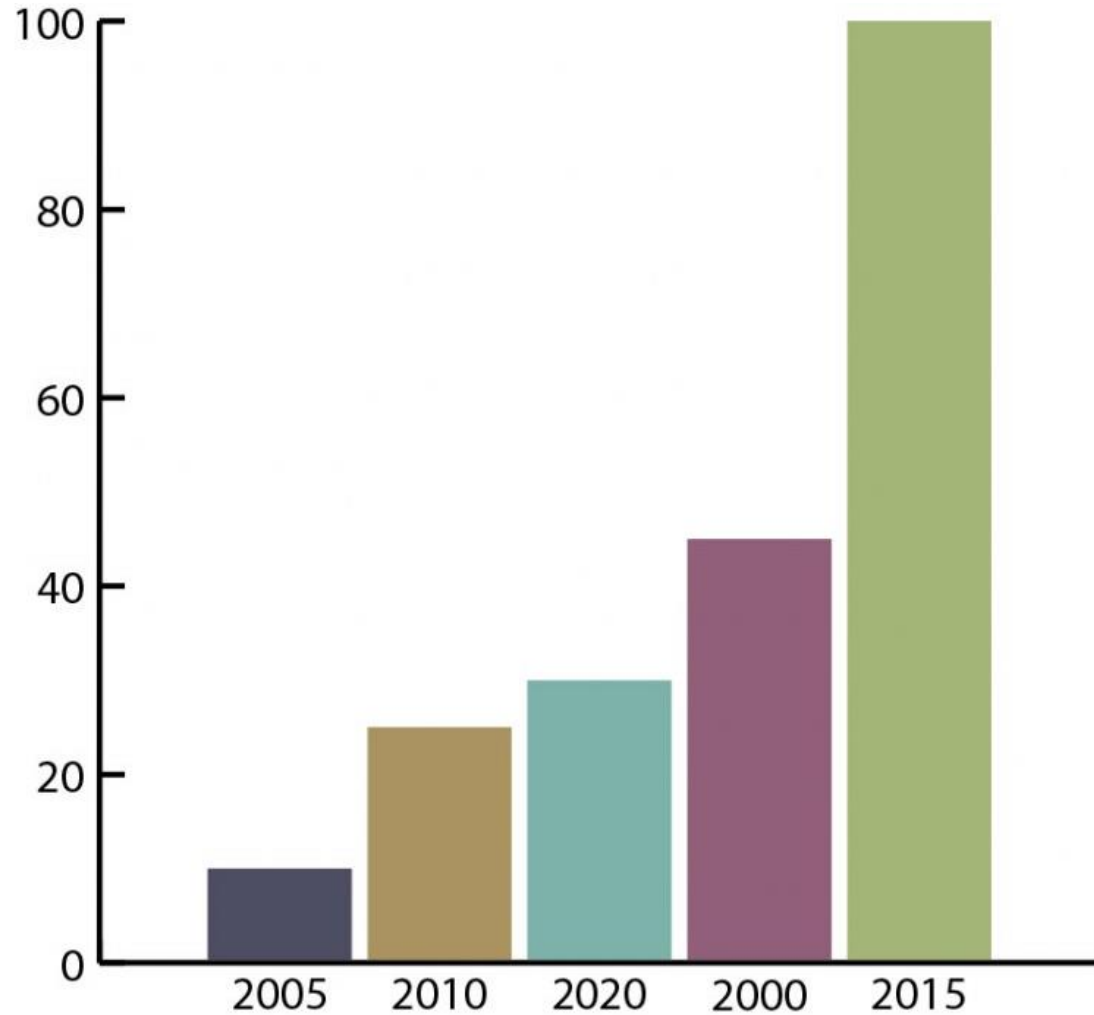
Afbeelding gebruikt op basis van citaatrecht.



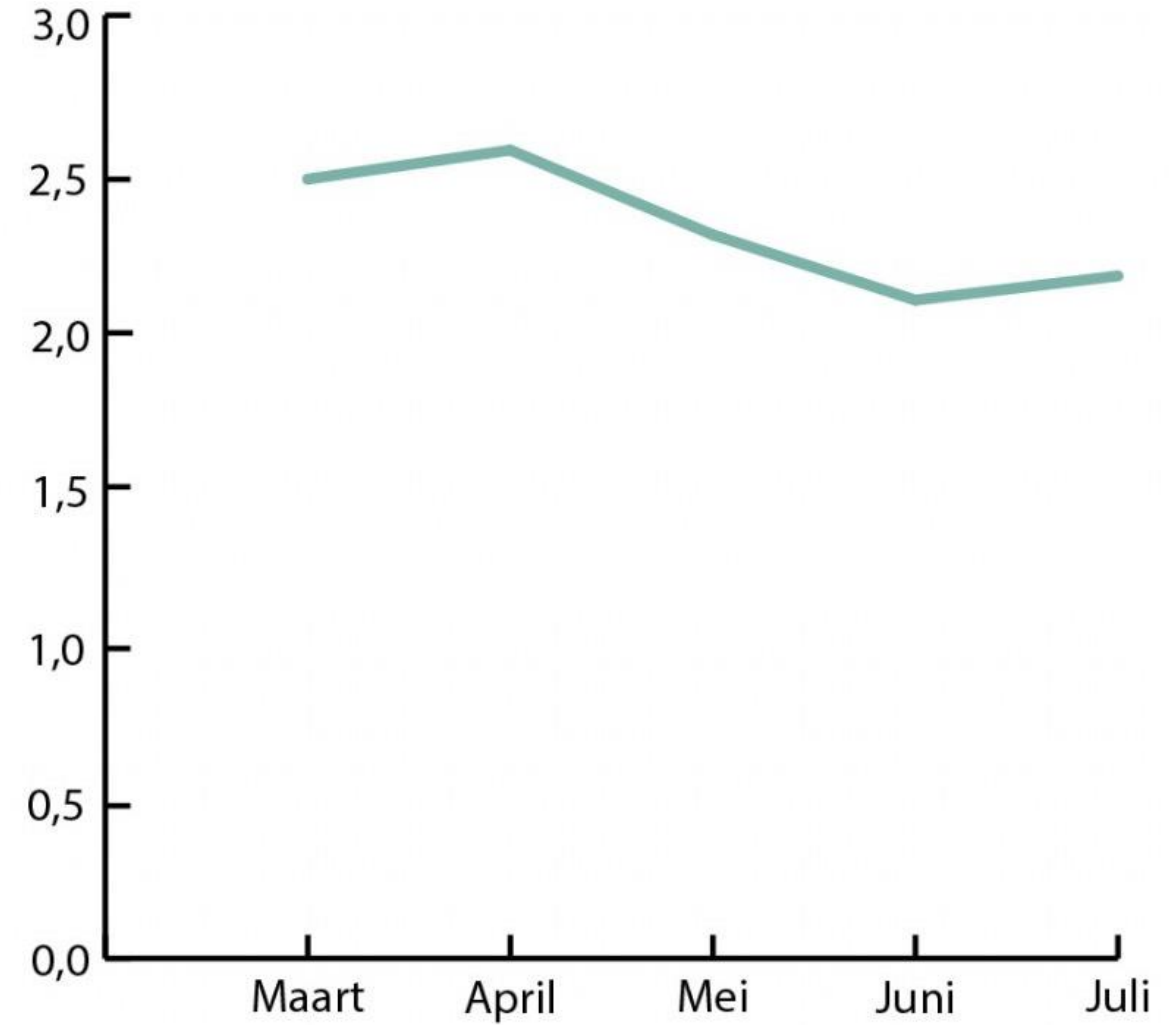
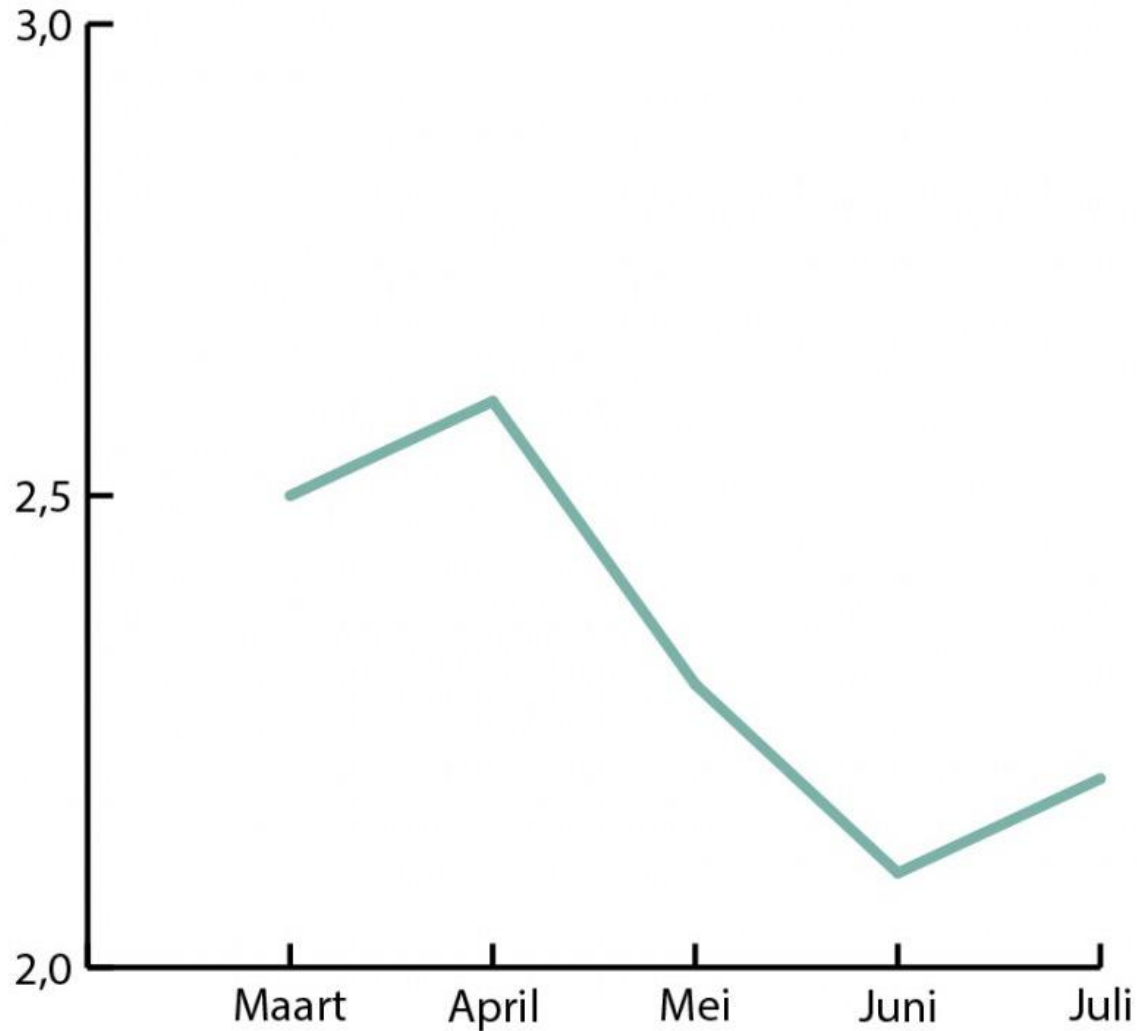


[https://youtu.be/4meeZifCVro?si=g7lcluLTZ\\_G14EBt](https://youtu.be/4meeZifCVro?si=g7lcluLTZ_G14EBt) LOVEFIELD (Short Film by Mathieu Ratthe, 26 april 2008)

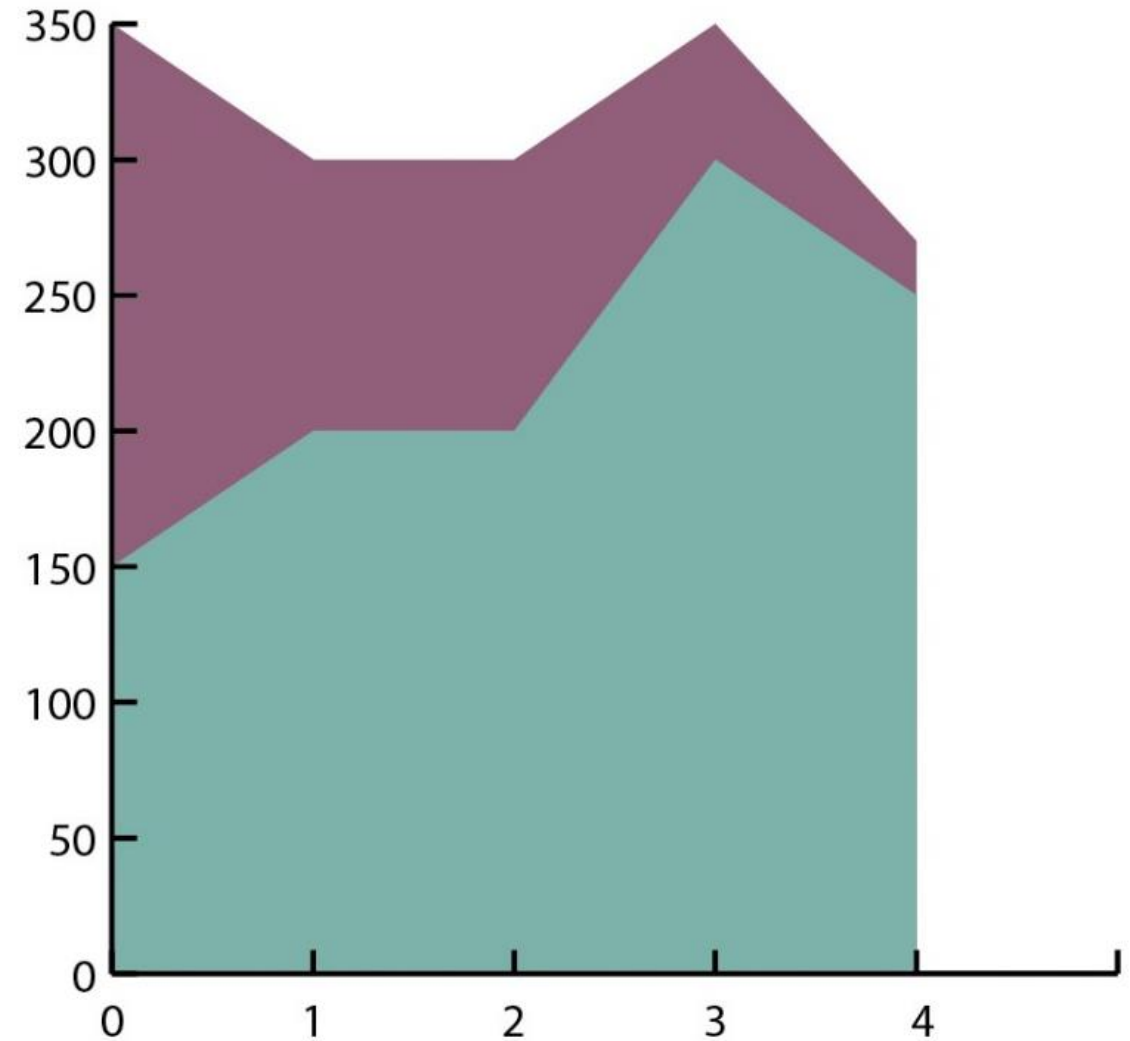
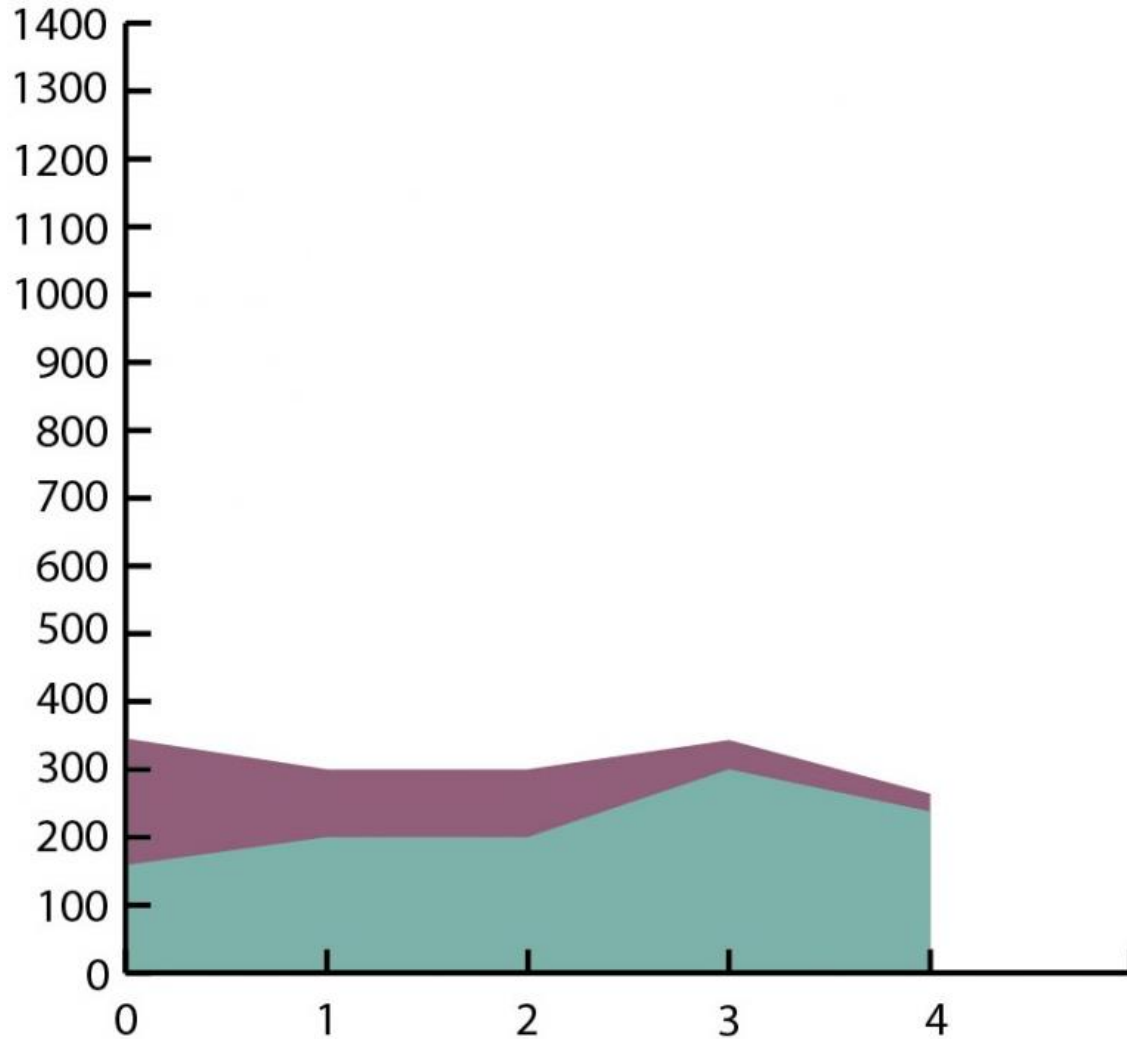
# Misleidende grafieken



# Misleidende grafieken

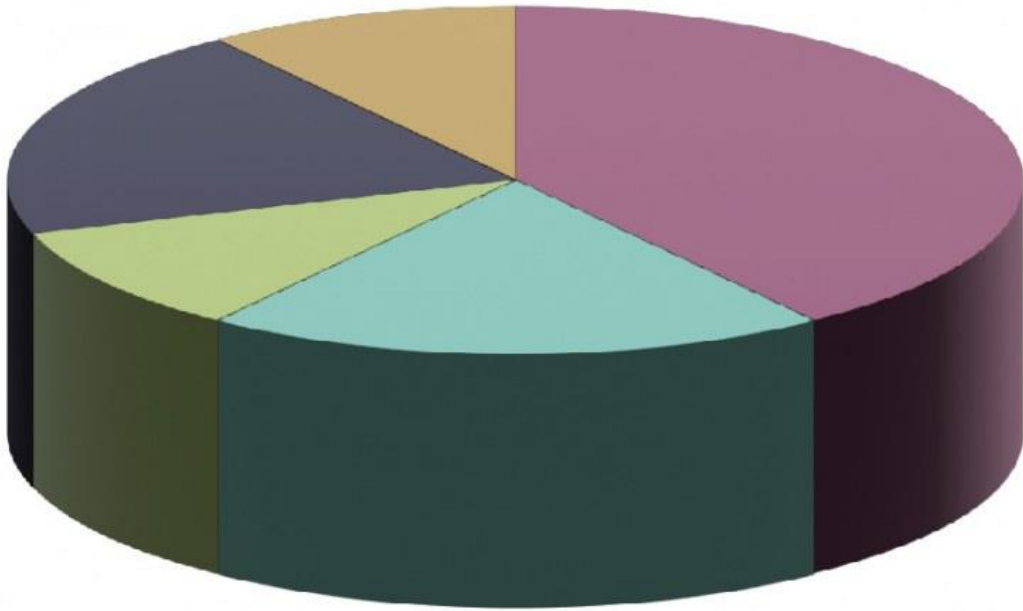


# Misleidende grafieken



<https://nieuwscheckers.nl/grafieken-drie-soorten-misleiding/>  
Inhoud van de webpagina gebruikt met toestemming.

# Misleidende grafieken



# Misleidende grafieken – Cherry picking



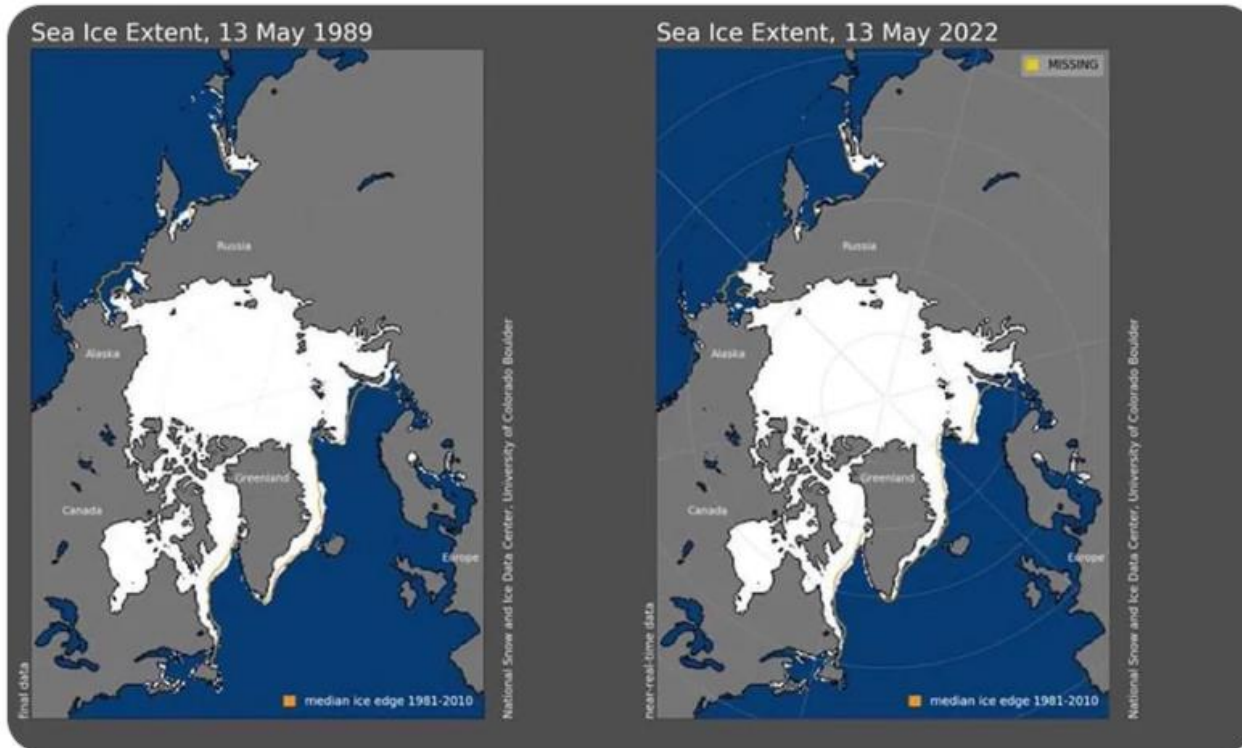
ferdinand meeus ✓  
@fmeeus1



Weet u nog, het Noordpool ijs ging verdwijnen...Euh...er ligt vandaag evenveel ijs als in 1989.

Zoek het verschil in beide beelden...

[Translate post](#)



10:39 AM · May 15, 2022

Tweet gebruikt op basis  
van citaatrecht.

<https://www.vrtnws.nl/2022/05/16/check-neen-er-ligt-niet-evenveel-ijs-op-de-noordpool-als-in-19/>



# Misleidende grafieken – Cherry picking

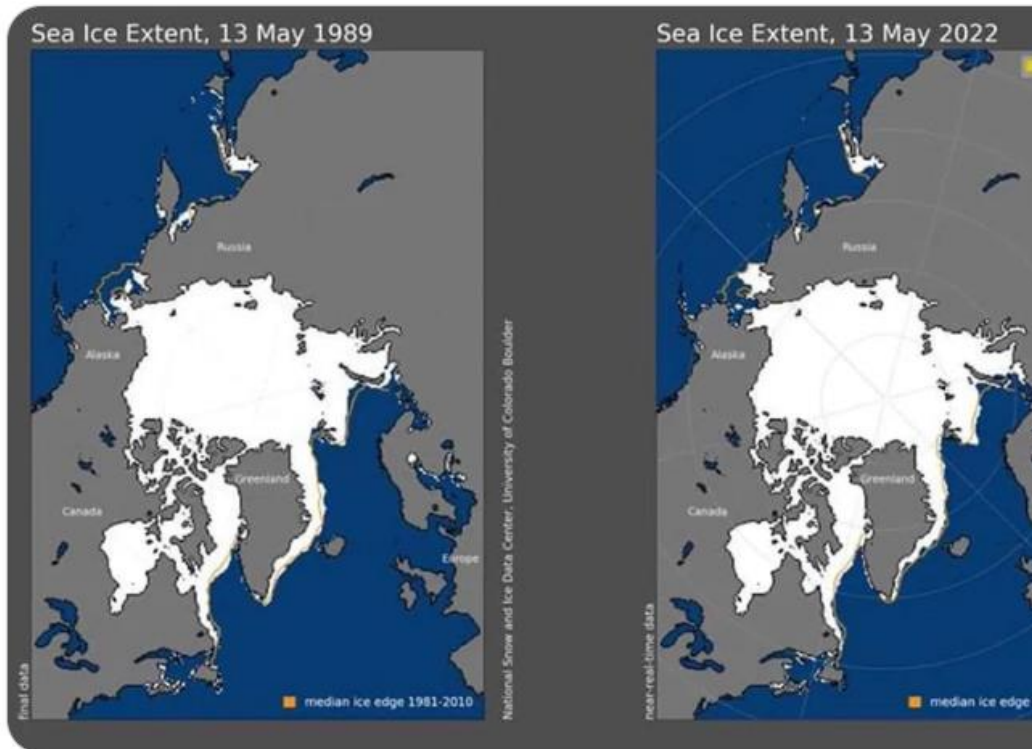


ferdinand meeus ✓  
@fmeeus1

Weet u nog, het Noordpool ijs ging verdwijnen...Euh...er ligt va  
evenveel ijs als in 1989.

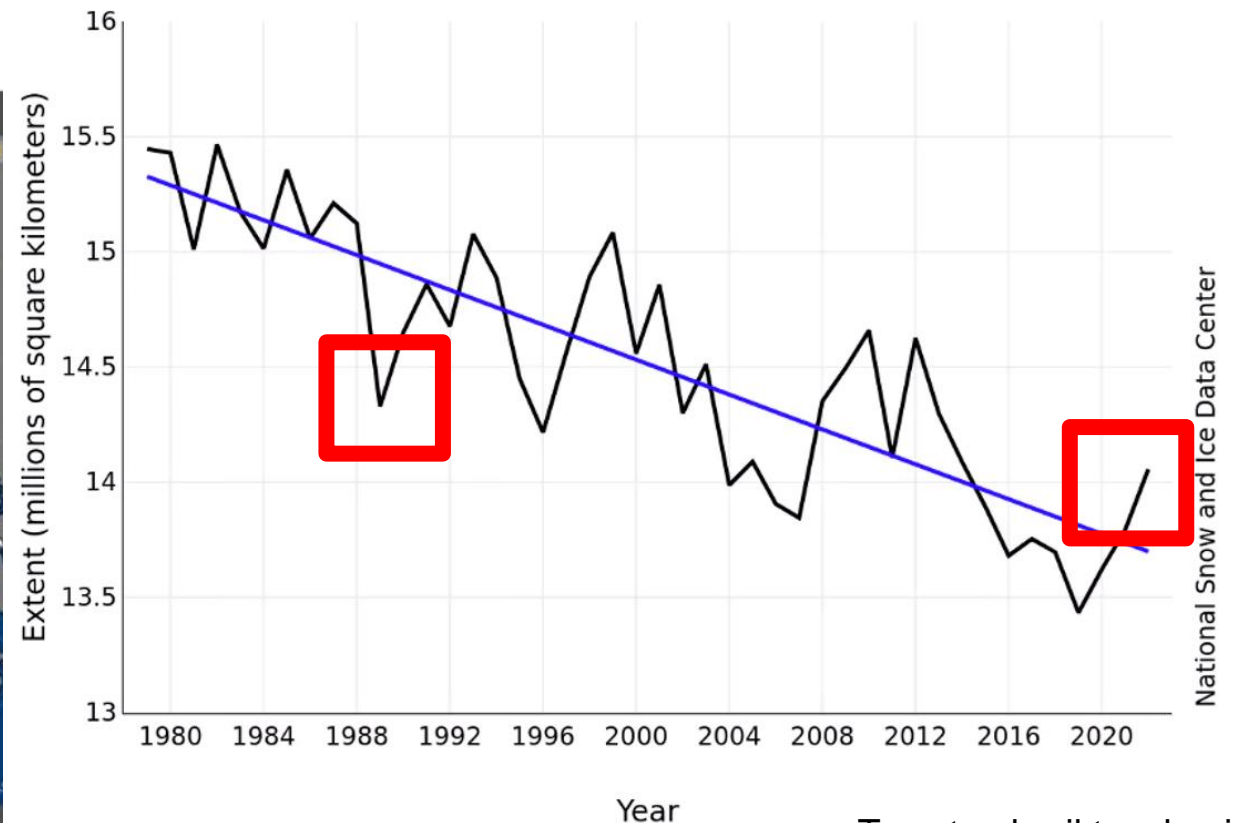
Zoek het verschil in beide beelden...

[Translate post](#)



10:39 AM · May 15, 2022

Average Monthly Arctic Sea Ice Extent  
April 1979 - 2022

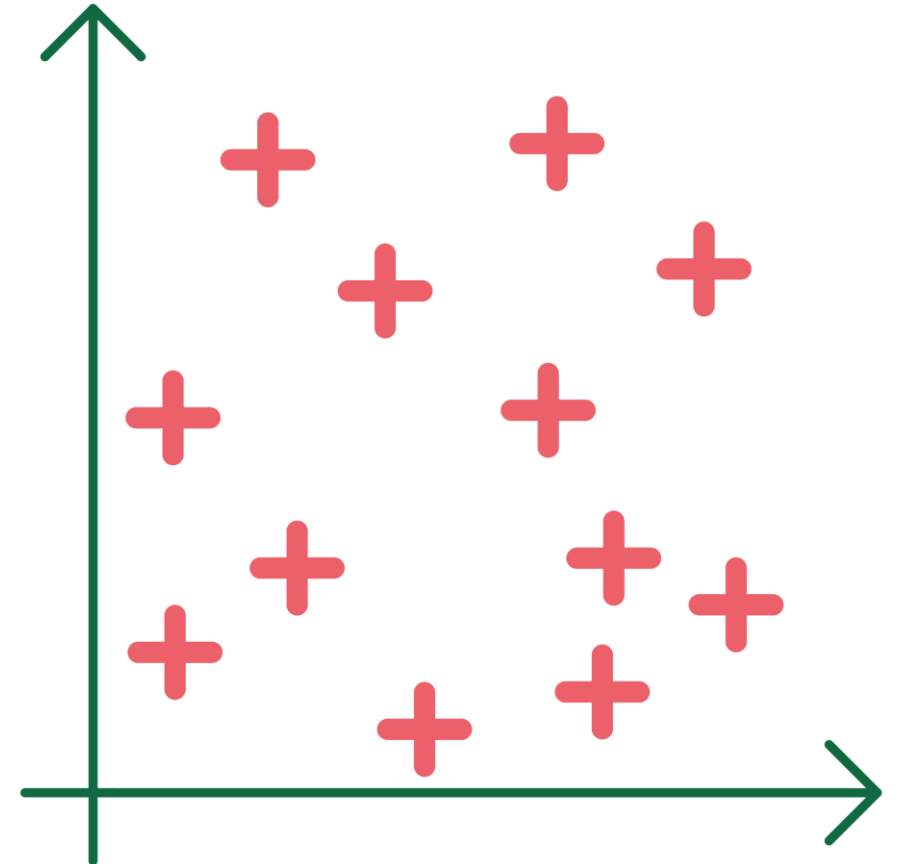


Tweet gebruikt op basis  
van citaatrecht.

<https://www.vrt.nw.nl/2022/05/16/check-neen-er-ligt-niet-evenveel-ijs-op-de-noordpool-als-in-19/>

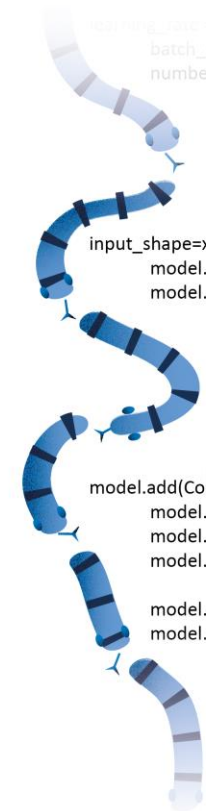
# ZELF DATA VISUALISEREN

## PUNTENWOLK BIVARIATE DATASETS





# PYTHON NOTEBOOKS



```
learning_rate = 0.001
batch_size = 32
number_dense_neurons = 32
number_filters = 8
filter_size_1 = 3
filter_size_2 = 3

model = Sequential()
model.add(Conv2D(number_filters, (filter_size_1, filter_size_1), padding='valid',
input_shape=x_train.shape[1:]))
model.add(Activation('relu'))
model.add(MaxPooling2D(pool_size=(2, 2)))
model.add(Dropout(0.1))

model.add(Conv2D(number_filters*2, (filter_size_2, filter_size_2), padding='valid'))
model.add(Activation('relu'))
model.add(MaxPooling2D(pool_size=(2, 2)))
model.add(Dropout(0.25))

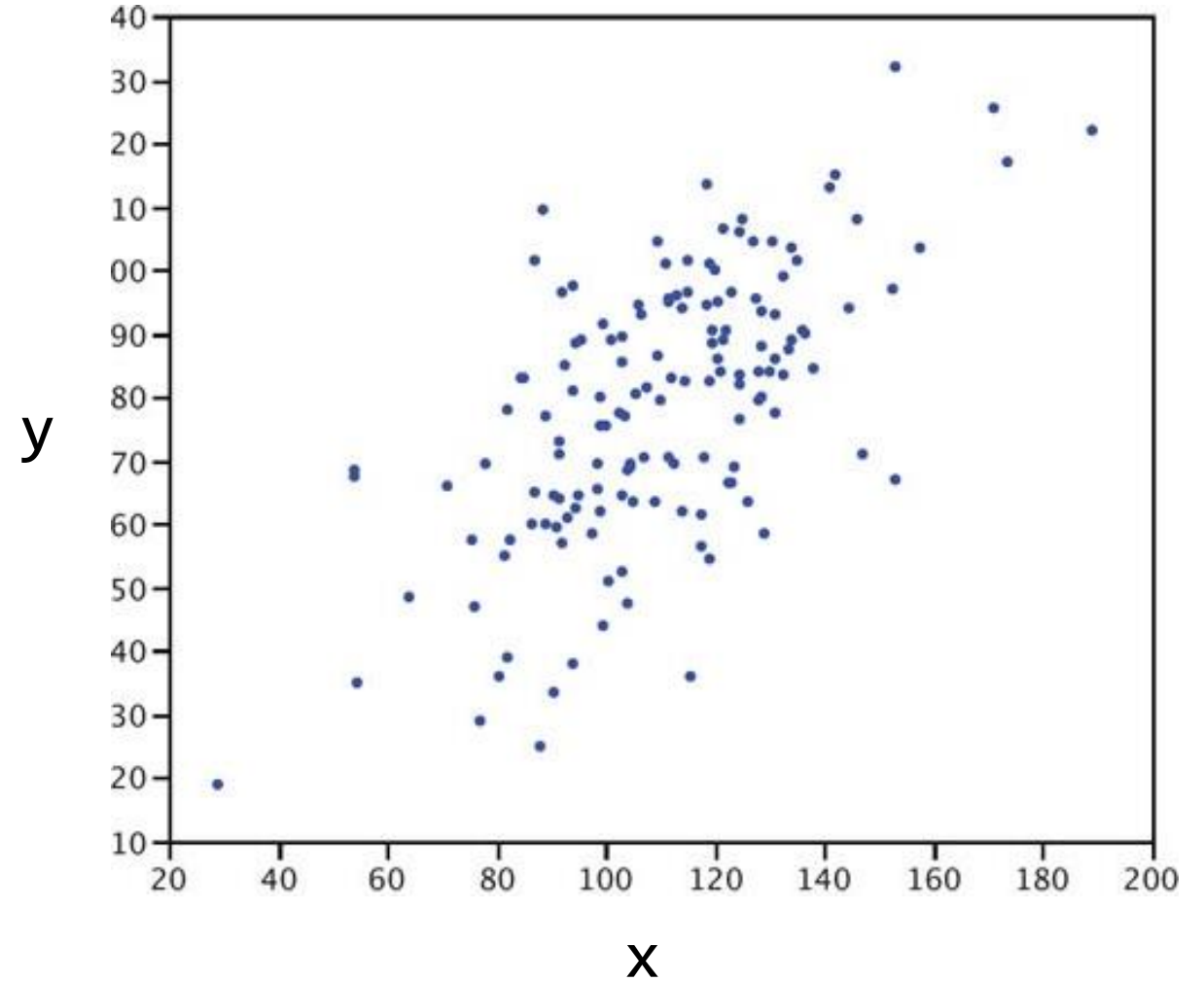
model.add(Conv2D(number_filters*4, (3, 3), padding='valid'))
model.add(Activation('relu'))
model.add(MaxPooling2D(pool_size=(2, 2)))
model.add(Dropout(0.25))

model.add(Flatten()) # Convert 3D feature maps to a 1D feature vector
model.add(Dense(number_dense_neurons))
model.add(Activation('relu'))
model.add(Dropout(0.5))
model.add(Dense(number_dense_neurons))
model.add(Activation('relu'))
model.add(Dropout(0.5))
```

# Spreidingsdiagram

CSV *comma separated value*

```
x,y  
105.5,80.5  
106.5,93  
94,81  
90.5,33.5  
104,47.5  
171,125.5  
107.5,81.5  
124 5 83 5
```



## KEELINGCURVE

In deze notebook bekijk je hoe de  $\text{CO}_2$ -concentratie op een bepaalde plaats gedurende een dag verandert en hoe ze gedurende een jaar evolueert.

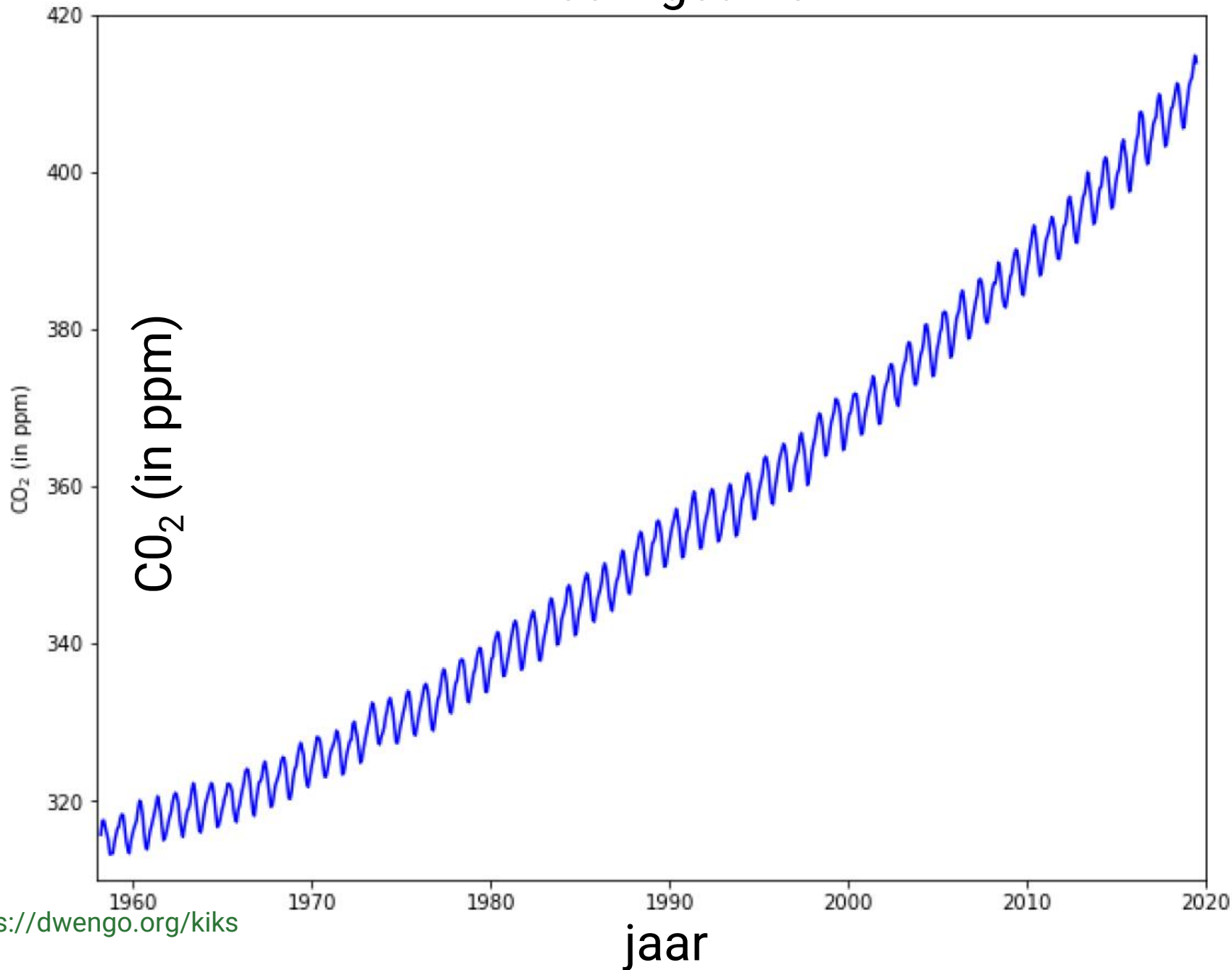
Daarnaast maak je met de data van  $\text{CO}_2$ -metingen op Mauna Loa, zelf een grafiek van de Keelingcurve. Je zal daarvoor de datapunten uit een *csv-bestand* verbinden met een gebroken lijn.

De Keelingcurve is een van de belangrijkste krommen in het klimaatonderzoek. De kromme is genoemd naar de wetenschapper David Keeling.

$\text{CO}_2$ -concentratie wordt uitgedrukt in **ppm**: parts per million.

<https://dwengo.org/kiks>

# Keelingcurve



<https://dwengo.org/kiks>

Keeling, C.D., Piper, S.C., Bacastow, R.B., Wahlen, M., Whorf, T.P., Heimann, M., & Meijer, H.A. (2001).

Exchanges of atmospheric CO<sub>2</sub> and <sup>13</sup>CO<sub>2</sub> with the terrestrial biosphere and oceans from 1978 to 2000.

*I. Global aspects, SIO Reference Series. No. 01-06, Scripps Institution of Oceanography, San Diego, 88 pages, 2001.*

Geraadpleegd op 5 september 2019 via [http://scrippsco2.ucsd.edu/data/atmospheric\\_co2/primary\\_mlo\\_co2\\_record](http://scrippsco2.ucsd.edu/data/atmospheric_co2/primary_mlo_co2_record)

# TOEPASSING SPREIDINGSDIAGRAM: TREINSTATIONS

In deze notebook zal je data m.b.t. de locatie van Belgische treinstations op een degelijke wijze proberen voor te stellen. Hiervoor zal je alle modules, functies en methodes nodig hebben die je tot nog toe hebt leren kennen!

## Nodige modules importeren

```
In [1]: import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
```

## 1. Data inlezen

Je zal de data visualiseren van treinstations die bediend worden door de NMBS.

Het bestand `nmbs-stations.csv` bevat de geografische coördinaten van al deze treinstations [1].

<https://dwengo.org/stem/>

```
In [2]: # voorbeeldscript

csv = pd.read_csv("data/nmbsstations.csv")

csv
```

Out[2]:

	name	breedtegraad	lengtegraad
0	Verviers Palais	50.590961	5.865262
1	Fraipont	50.565414	5.723475
2	Stockem	49.690736	5.767712
3	Courriere	50.387727	4.995732
4	Marchienne Zone	50.397166	4.389017
...	...	...	...
554	Aarsele	50.984458	3.418361
555	Furnes	51.073897	2.669742
556	Wichelen	51.001903	3.974897
557	Aalst Kerrebroek	50.948316	4.024773
558	KOLN HBF	50.943706	6.958122

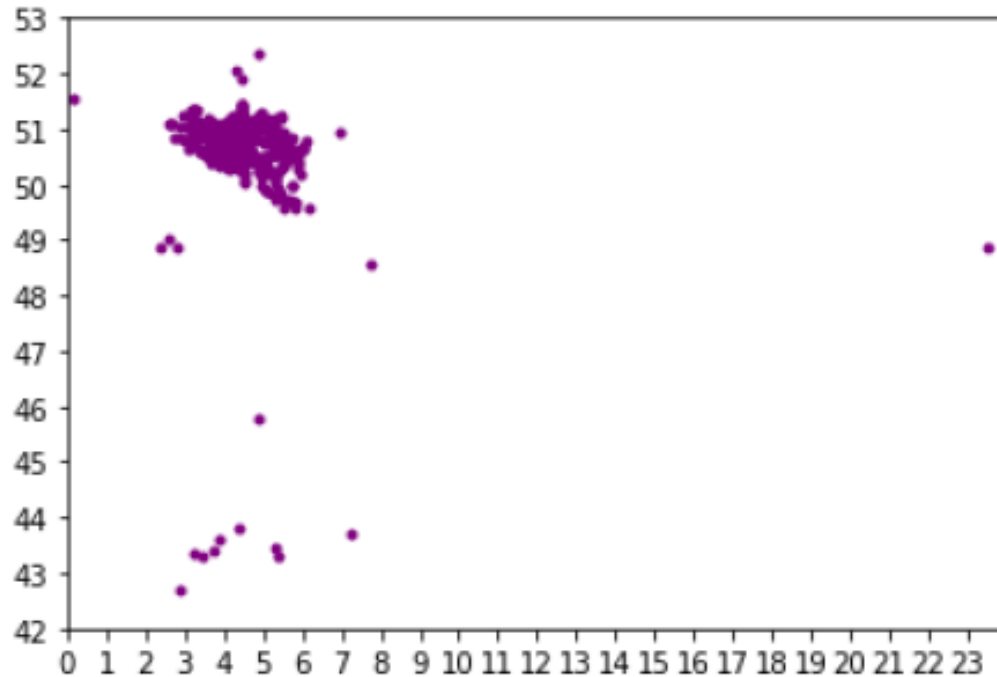
In [6]: *# voorbeeldscript*

```
plt.figure()                                # creëer tekenvenster

plt.axis(xmin=0, xmax=24, ymin=42, ymax=53)  # vensterinstellingen
plt.xticks(np.arange(0, 24, step=1))
plt.yticks(np.arange(42, 54, step=1))

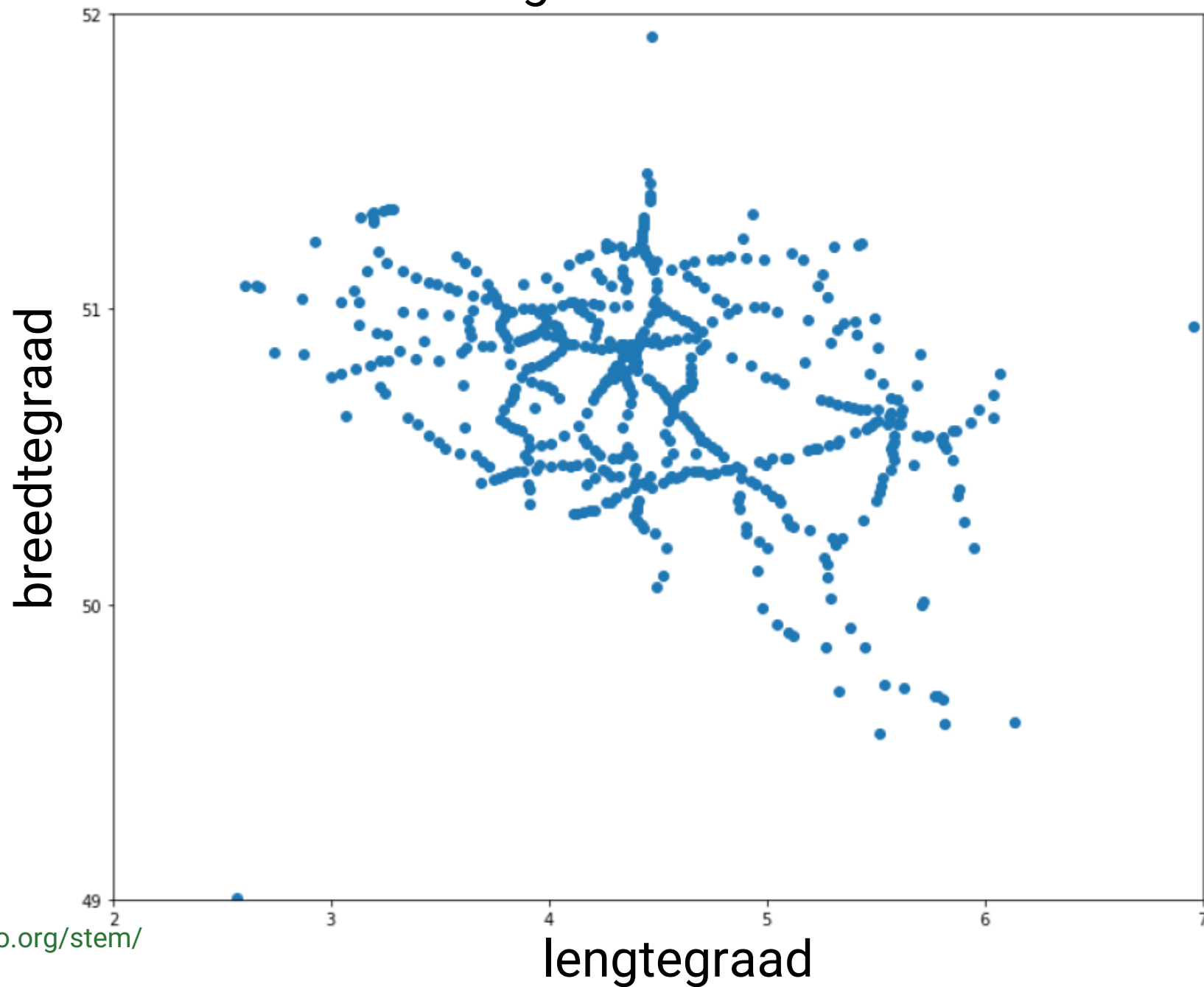
plt.scatter(x, y, color="purple", marker=".") # plot punten

plt.show()                                  # toon grafiek
```



De leerlingen creëren deze visualisatie en passen deze nadien aan: focus op België, assen benoemen en een gepaste titel voorzien.

# Belgische treinstations





# Digimeter 2022 Smartphone-afhankelijkheid

		16-24	25-34	35-44	45-54	55-64	65-74	75+	Totaal
Smartphone-afhankelijk	2018	48%	49%	40%	22%	11%	10%		<b>31%</b>
	2019	52%	48%	46%	29%	17%	12%		<b>33%</b>
	2020	53%	60%	49%	31%	16%	14%		<b>35%</b>
	2021	51%	60%	43%	31%	15%	9%		<b>33%</b>
	<b>2022</b>	<b>62%</b>	<b>60%</b>	<b>53%</b>	<b>42%</b>	<b>20%</b>	<b>18%</b>	<b>15%</b>	<b>40%</b>

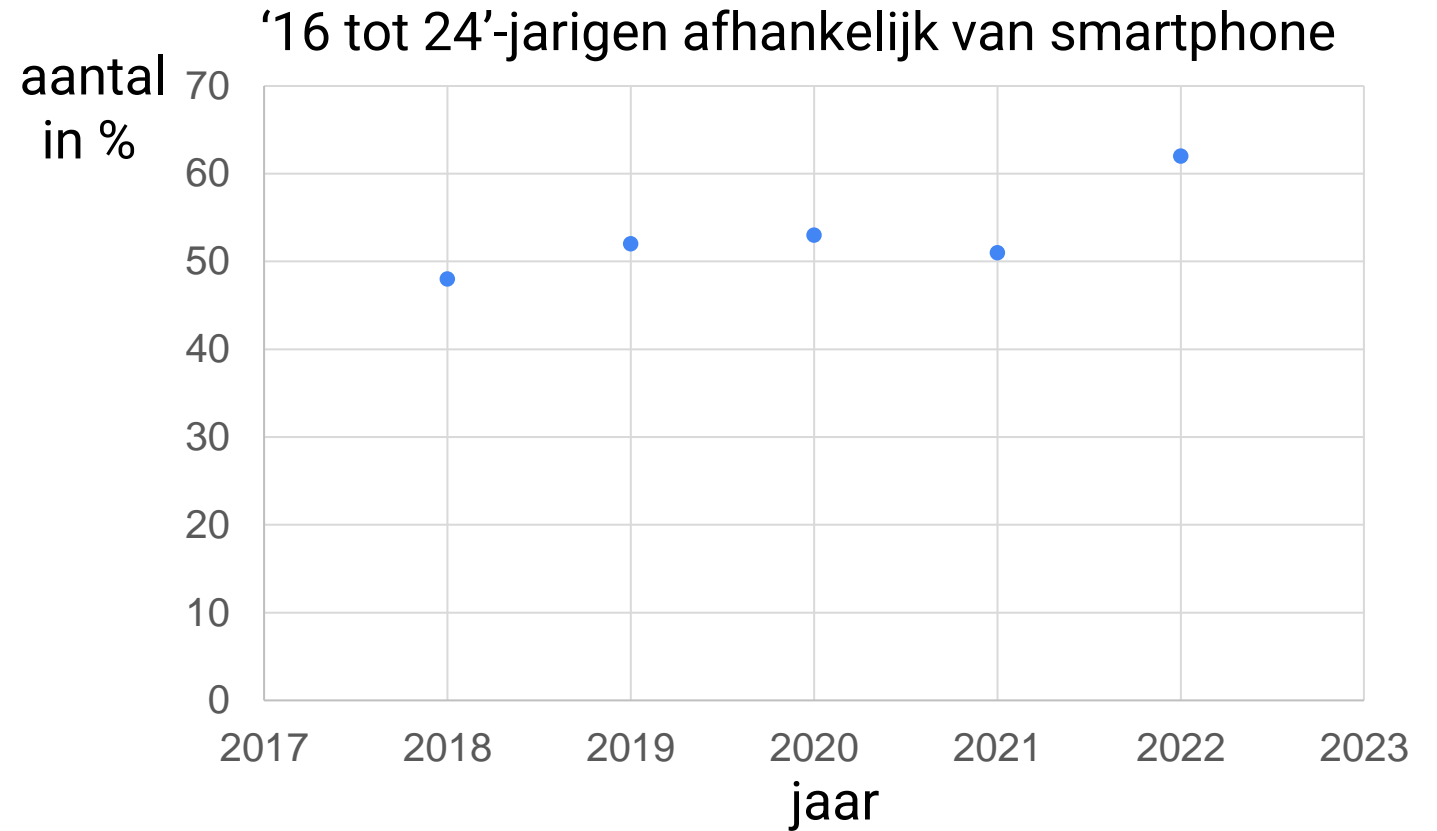
# Digimeter 2022 Smartphone-afhankelijkheid

		16-24	25-34	35-44	45-54	55-64	65-74	75+	Totaal
Smartphone-afhankelijk	2018	48%	49%	40%	22%	11%	10%		31%
	2019	52%	48%	46%	29%	17%	12%		33%
	2020	53%	60%	49%	31%	16%	14%		35%
	2021	51%	60%	43%	31%	15%	9%		33%
	2022	62%	60%	53%	42%	20%	18%	15%	40%



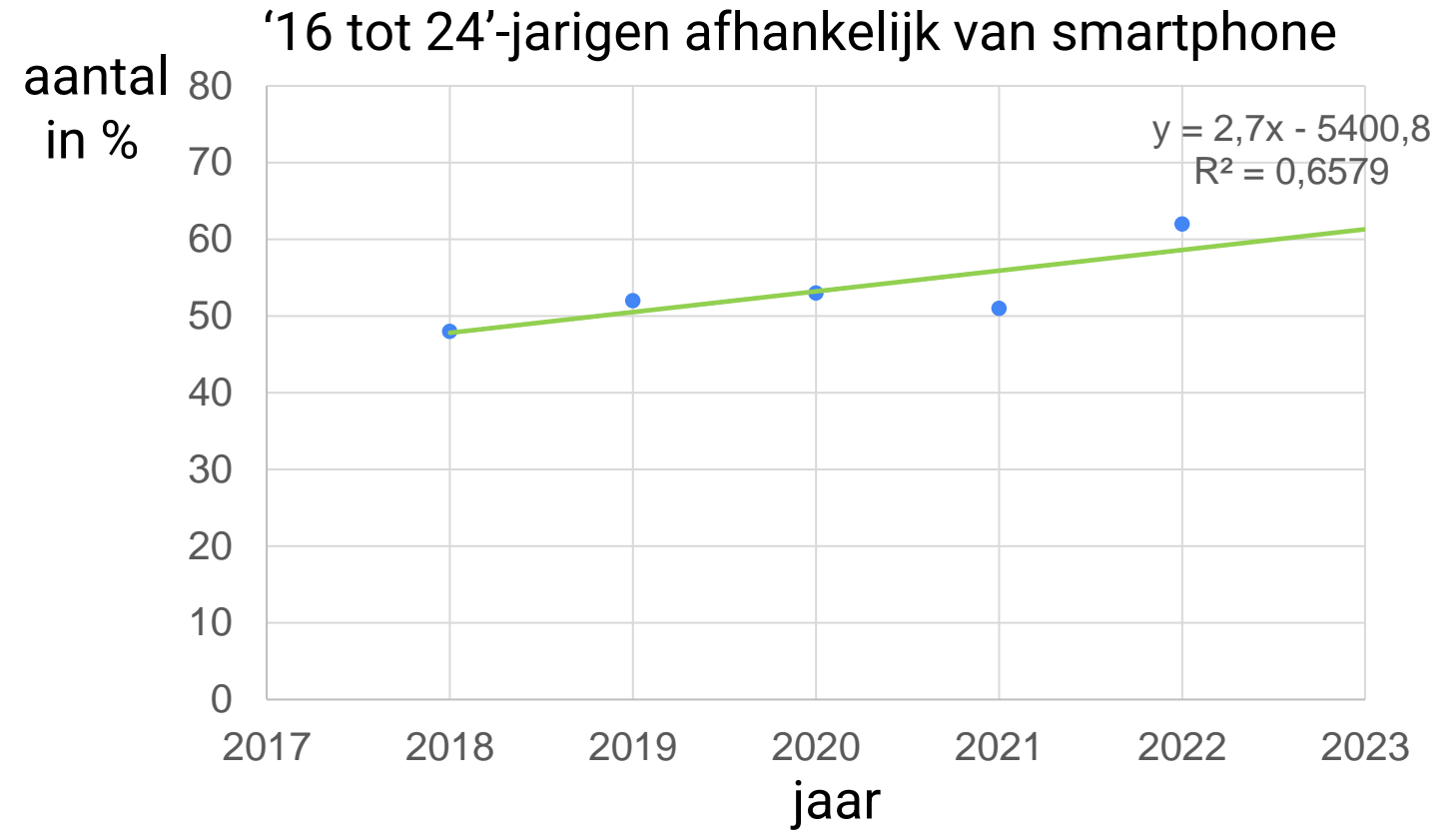
# Digimeter 2022 Smartphone-afhankelijkheid

		16-24
Smartphone-afhankelijk	2018	48%
	2019	52%
	2020	53%
	2021	51%
	<b>2022</b>	<b>62%</b>



# Digimeter 2022 Smartphone-afhankelijkheid

		16-24
Smartphone-afhankelijk	2018	48%
	2019	52%
	2020	53%
	2021	51%
	<b>2022</b>	<b>62%</b>

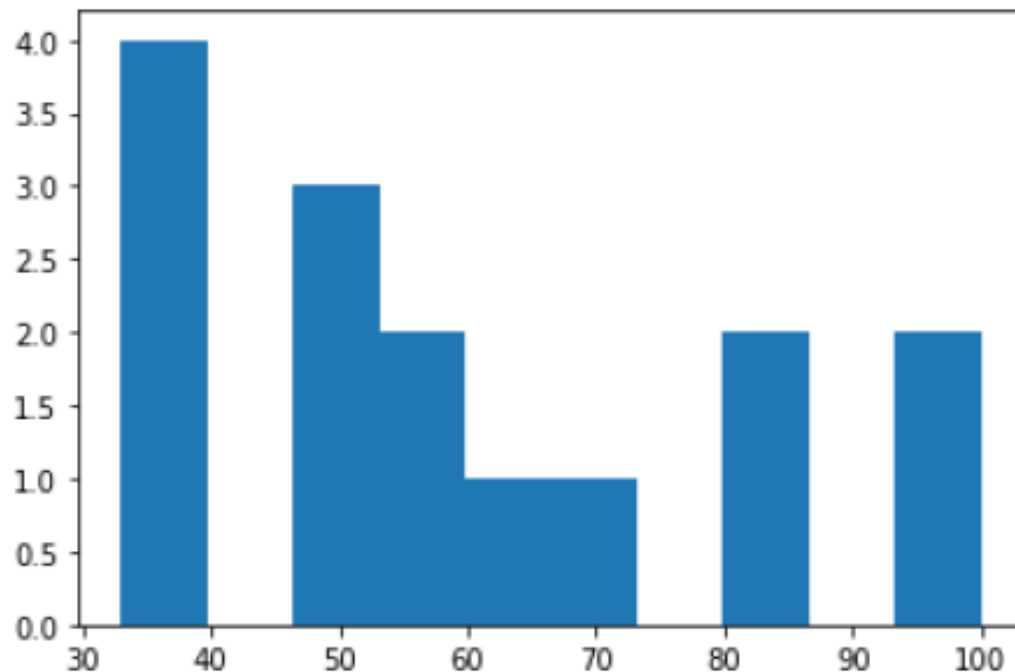


```
In [18]: print(y)
print(np.min(y), np.max(y))
```

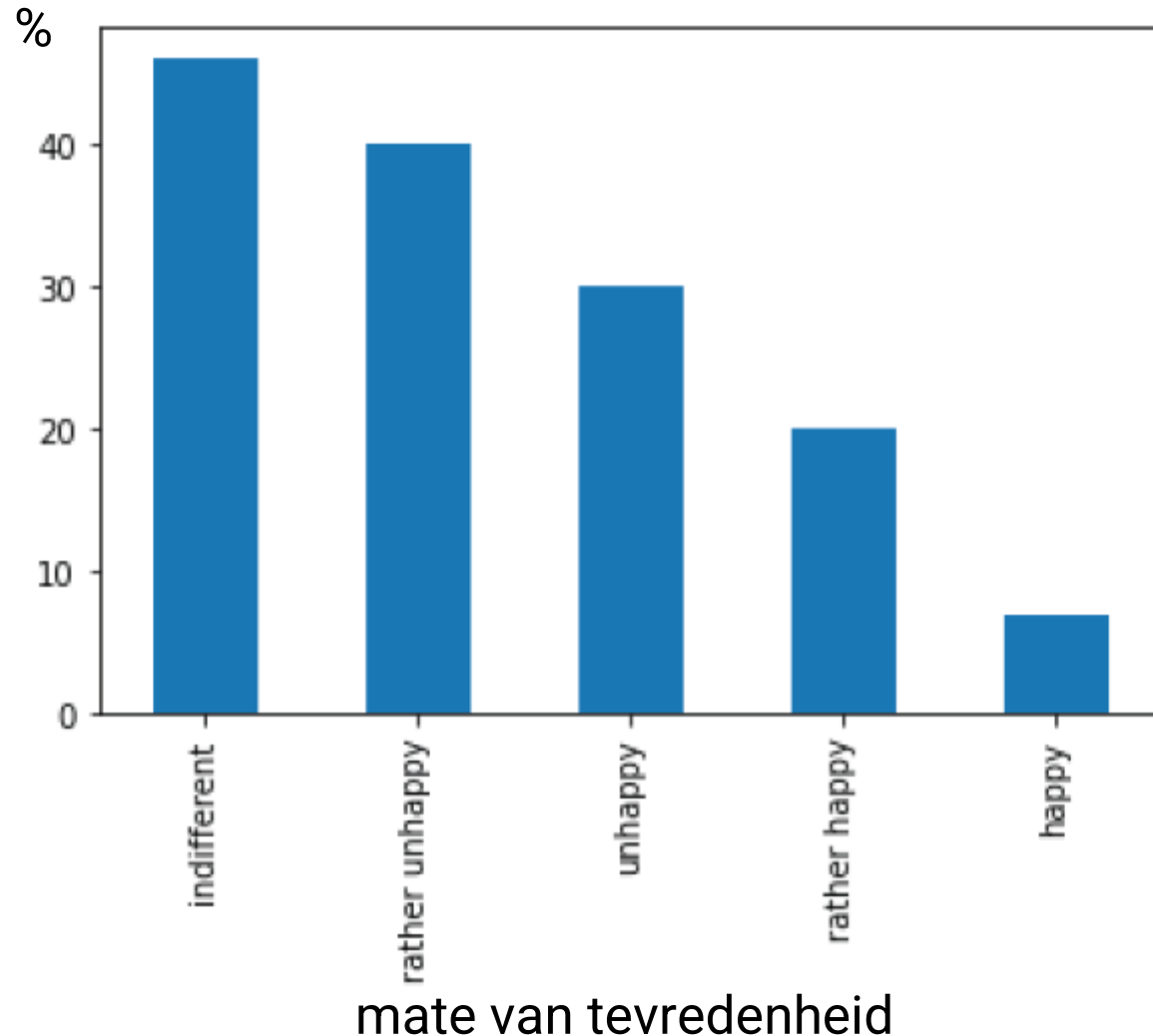
```
[ 33  35  47  35  55  56  50  50  70  65  35  82  80  95 100]
33 100
```

```
In [19]: import matplotlib.pyplot as plt
plt.hist(y, bins=10)
```

```
Out[19]: (array([4., 0., 3., 2., 1., 1., 0., 2., 0., 2.]),
array([ 33.,  39.7,  46.4,  53.1,  59.8,  66.5,  73.2,  79.9,  86.6,
        93.3, 100. ]),
<a list of 10 Patch objects>)
```



## Tevredenheid over vastgoedprijzen

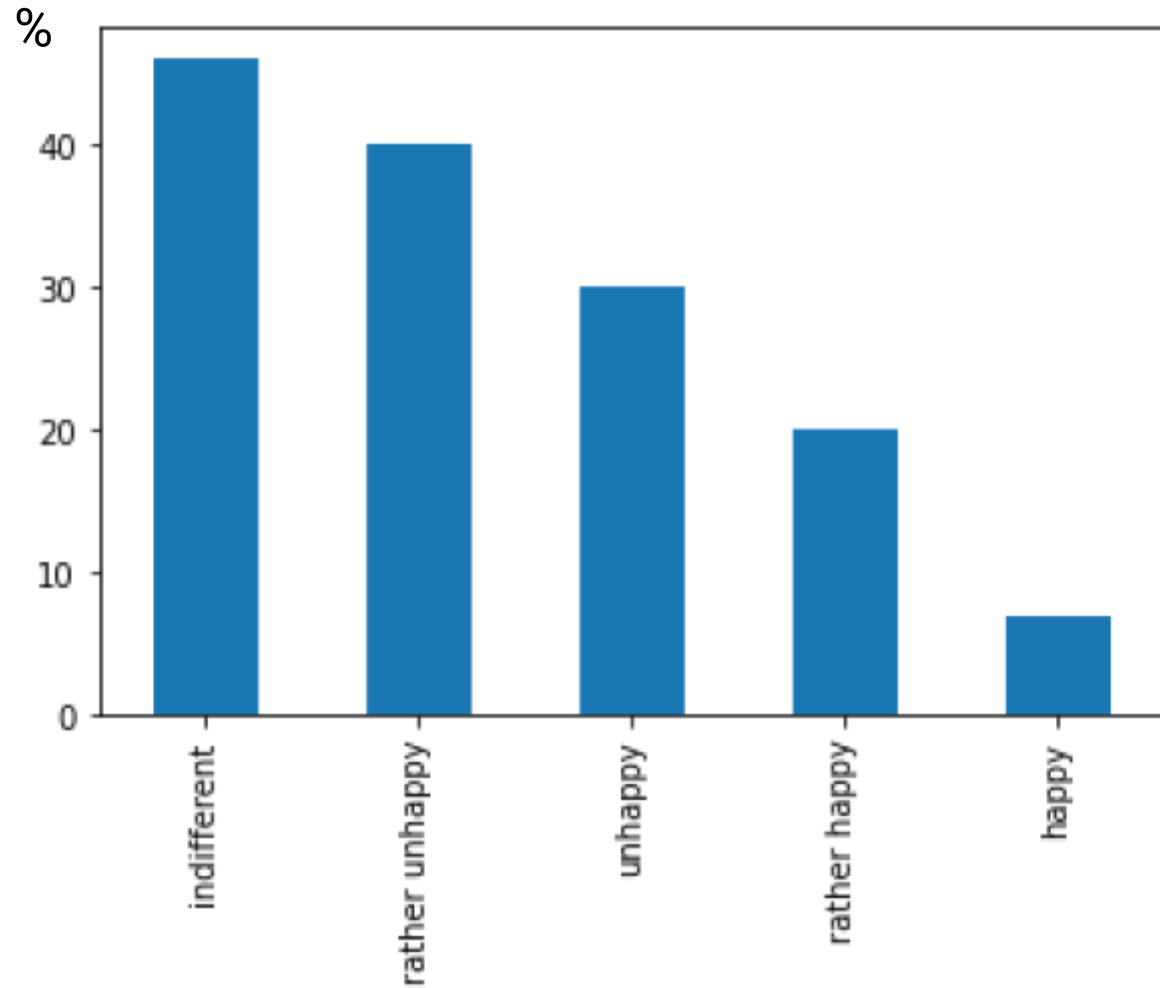


Priyanshu Sethi. Happiness Classification Dataset. <https://www.kaggle.com/datasets/priyanshusethi/happiness-classification-dataset>  
Deze dataset is gebaseerd op [2].

[2] Waldemar Koczkodaj. Somerville Happiness Survey. CC BY 4.0. <https://archive.ics.uci.edu/dataset/479/somerville+happiness+survey>  
Deze dataset is gebaseerd op [3].

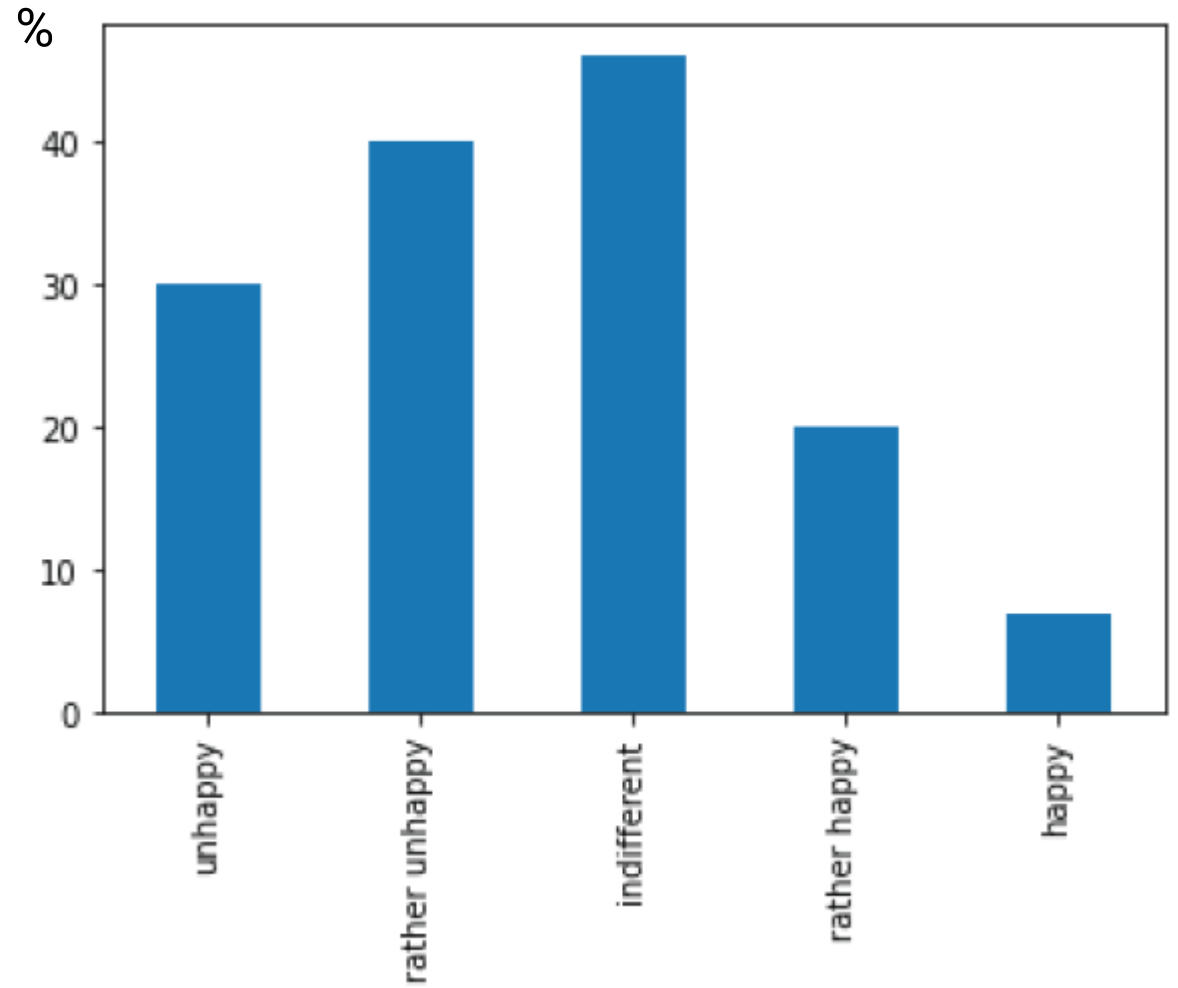
[3] Somerville Happiness Survey Responses. Public access and use. <https://catalog.data.gov/dataset/somerville-happiness-survey-responses>

Tevredenheid over vastgoedprijzen



mate van tevredenheid

Tevredenheid over vastgoedprijzen

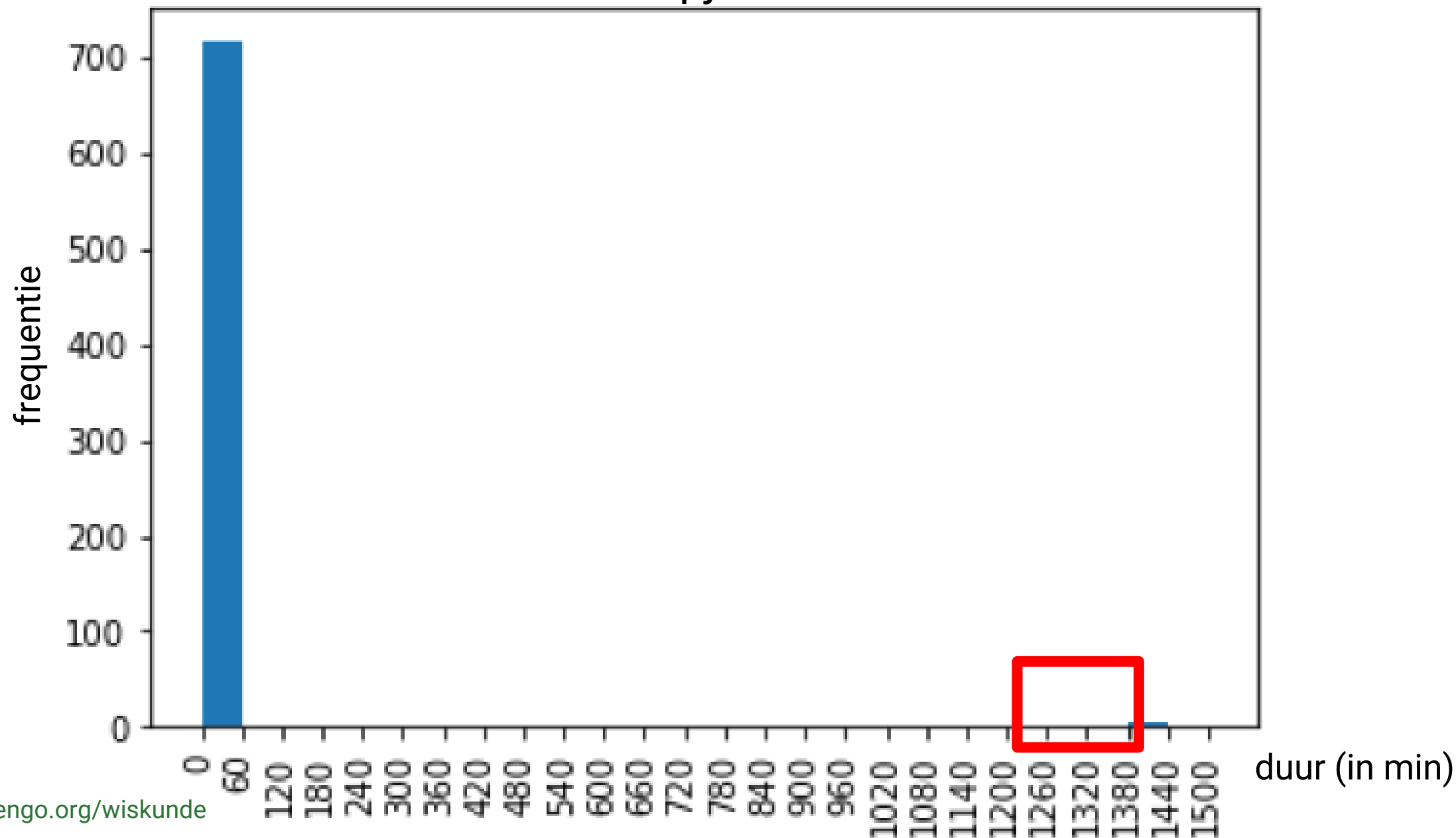


mate van tevredenheid



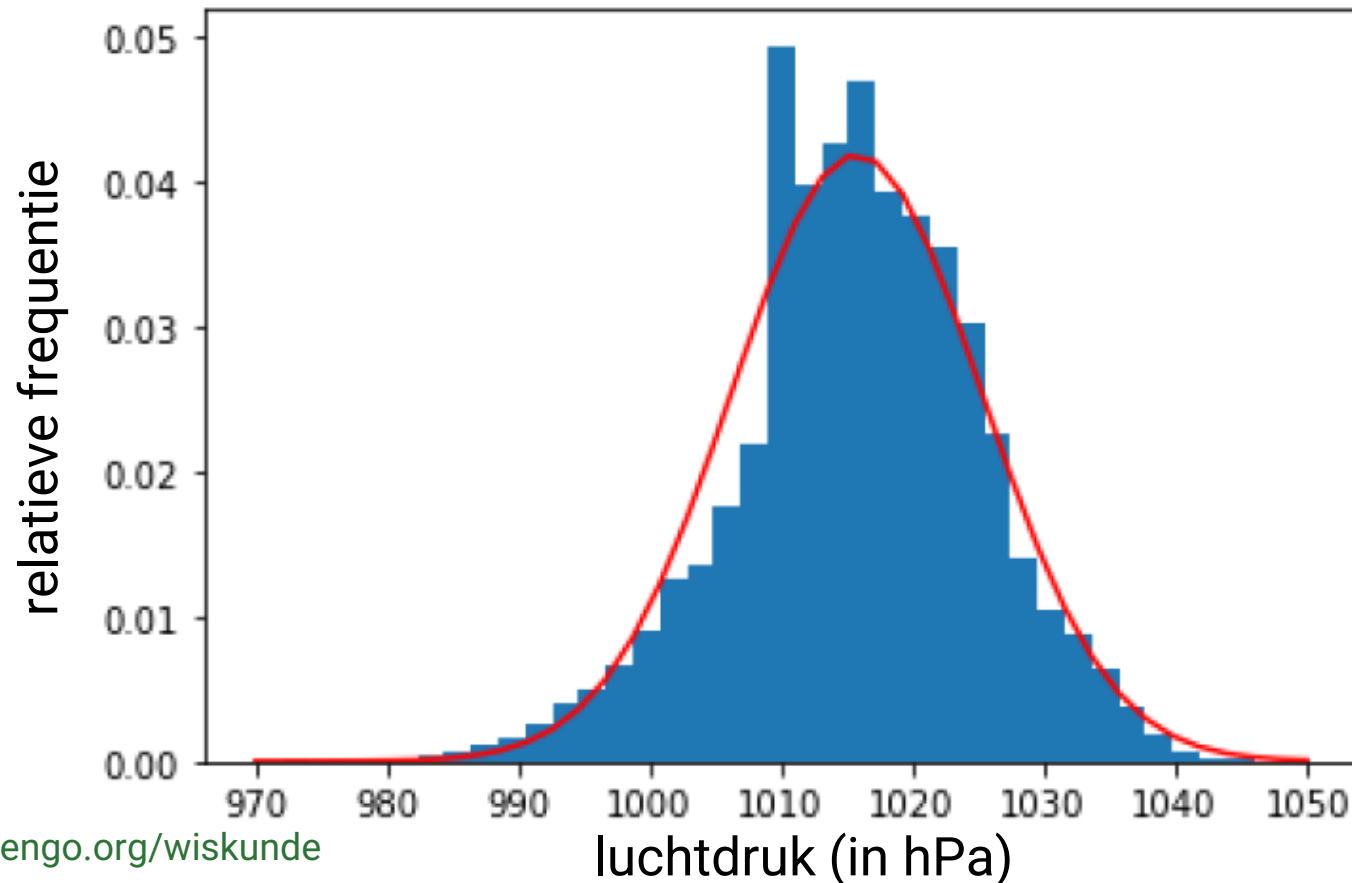
bron van de dataset: <a href="https://www.kaggle.com/datasets/dissfya/mr-beast-youtube-stats-and-subtitles">https://www.kaggle.com/datasets/dissfya/mr-beast-youtube-stats-and-subtitles</a>					17	13.066667
<a href="https://www.kaggle.com/datasets/dissfya/mr-beast-youtube-stats-and-subtitles">https://www.kaggle.com/datasets/dissfya/mr-beast-youtube-stats-and-subtitles</a> )					18	17.016667
YouTubekanaal van MrBeast: <a href="https://www.youtube.com/@MrBeast">https://www.youtube.com/@MrBeast</a> ( <a href="https://www.youtube.com/@MrBeast">https://www.youtube.com/@MrBeast</a> )					19	13.983333
					20	9.016667
den eerst de dataset in en slaan ze op in een variabele.					21	8.016667
					22	12.150000
<pre>data = pd.read_csv("data/MrBeast.csv") # De dataset uitlezen</pre>					23	12.083333
<pre>data # De dataset op het scherm tonen</pre>					24	11.033333
					...	
title	id	published_at	description	viewCo		
I Paid A Real Assassin To Try To Kill Me	YLt73w6criQ	2023-03-18 20:00:01	File with TurboTax today to get your biggest r...	65789	703	19.900000
					704	2.633333
					705	1.466667
Do Men Lie About Their Height?	IgIIQkCHO0g	2023-02-14 21:00:14	New Merch - <a href="https://shopmrbeast.com/">https://shopmrbeast.com/</a> SUBSCR...	105469	706	0.966667
					707	1.066667
1,000 Blind People See For The First Time	TJ2ifmkGGus	2023-01-28 21:00:00	If you would like to support more of this sigh...	126742	708	3.216667
					709	2.566667
					710	5.316667
I Survived 50 Hours In Antarctica	7IKab3HcfFk	2022-12-24 20:59:59	Check out my interview! <a href="http://bit.ly/Shopify-...">http://bit.ly/Shopify-...</a>	104436	711	5.450000
					712	4.350000
Hydraulic Press Vs Lamborghini	h5NvTTOIOtI	2022-12-10 21:00:01	Hello Neighbor 2 is out now on PC & Consoles!...	135219	713	3.583333
					714	4.466667

# Duur MrBeast filmpjes in minuten



```
In [8]: x = np.linspace(970,1050,40)
y = norm.pdf(x,barometer_gemiddelde,barometer_standaardafwijking) # Bereken de
weather['barometer'].plot(kind='hist',bins=x, density=True) # Teken het histogr
plt.plot(x,y,color='red') # Teken de dichtheidsfunctie van normale verdeling
```

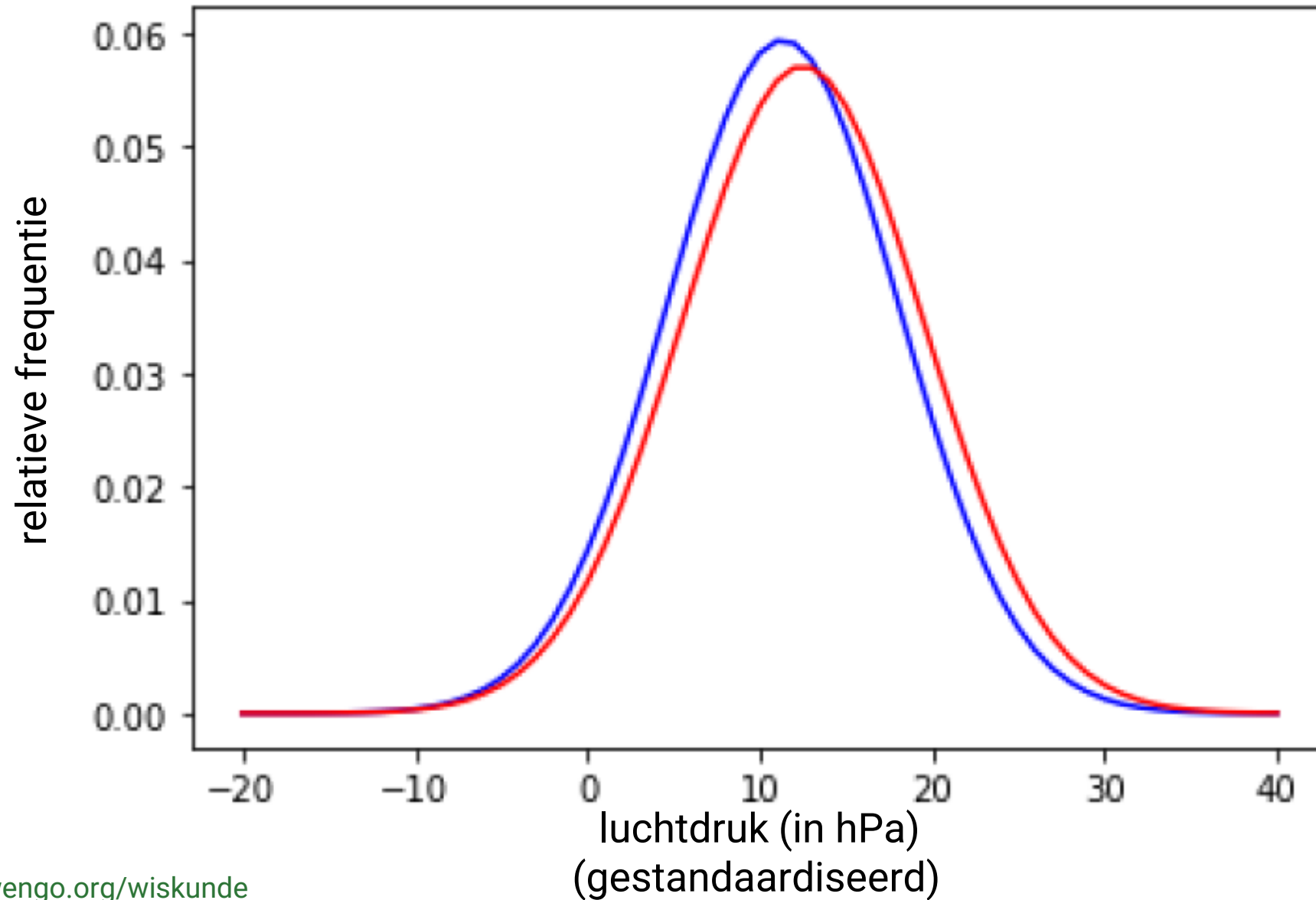
Out[8]: [<matplotlib.lines.Line2D at 0x223e82cd370>]



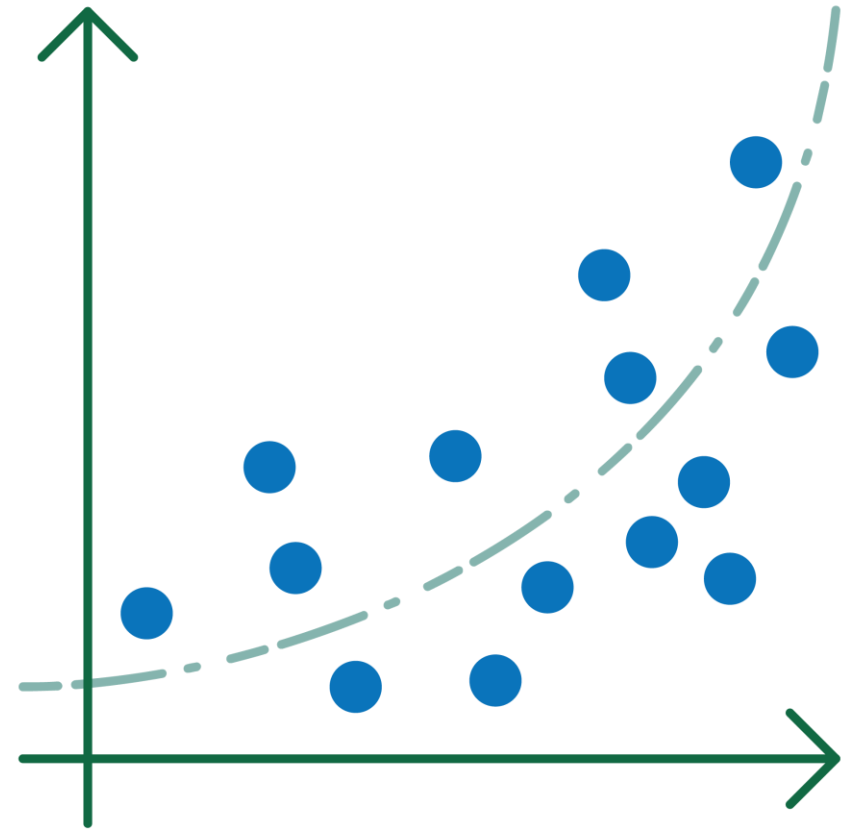
<https://dwengo.org/wiskunde>

Weather dataset in Antwerp, Belgium. <https://www.kaggle.com/datasets/ramima/weather-dataset-in-antwerp-belgium>  
Data gebruikt met toestemming van CustomWeather, eigenaar van de data [www.customweather.com](http://www.customweather.com) gebruikt voor deze dataset.

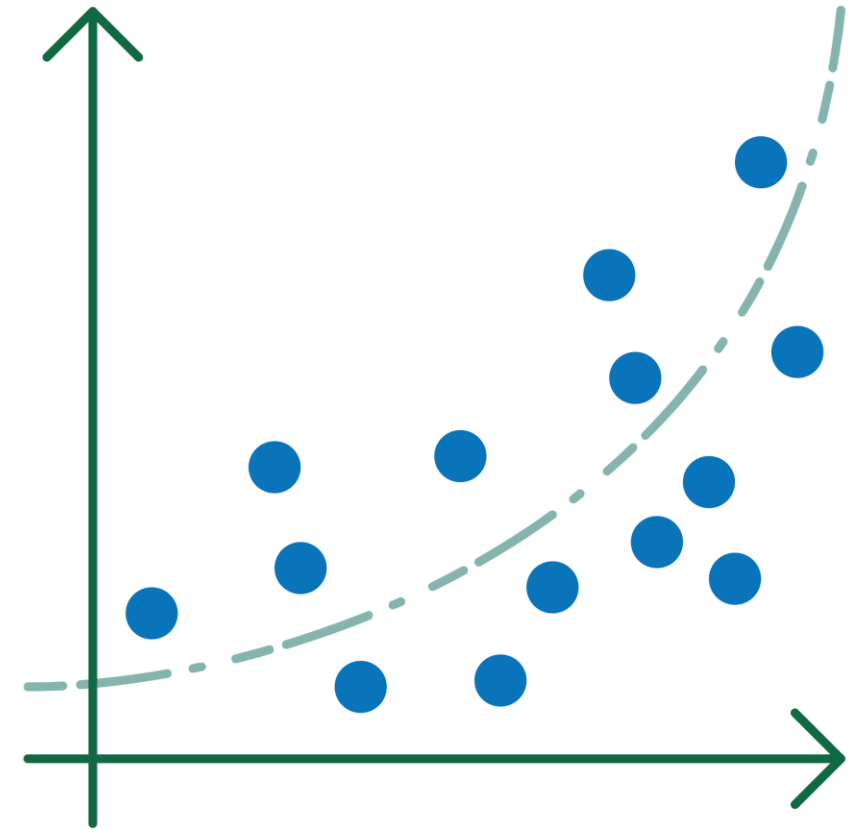
## Weer in Antwerpen 2012-2016



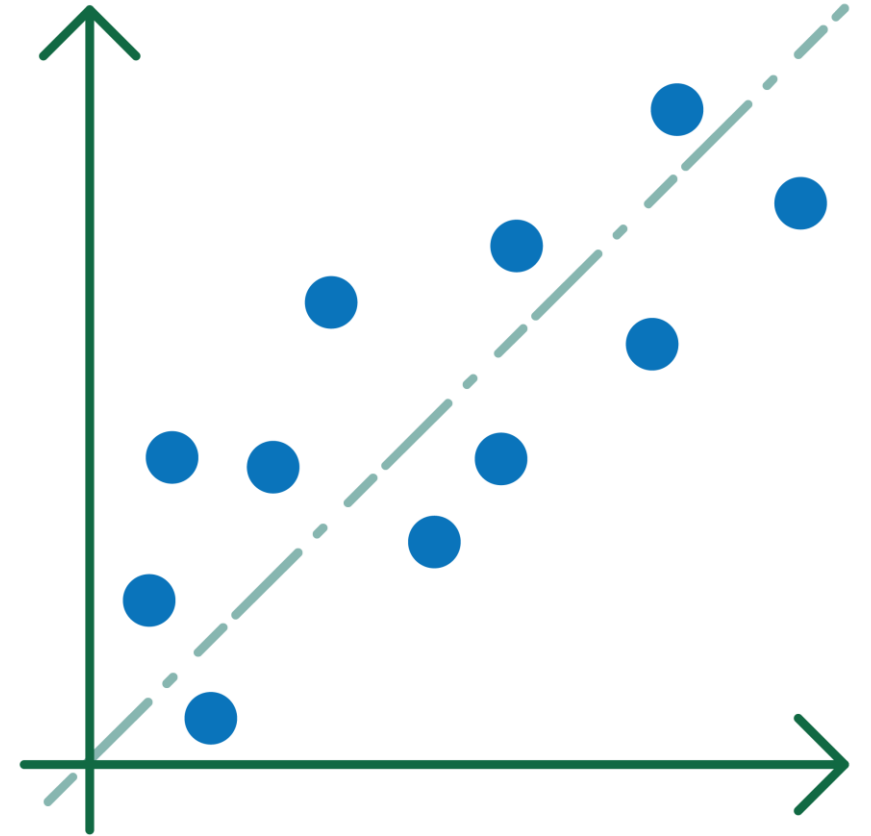
## REGRESSIE - TRENDLIJN



# WISKUNDIGE MODELLEN MODELLEREN EN SIMULEREN A.D.H.V. TRENDLIJN

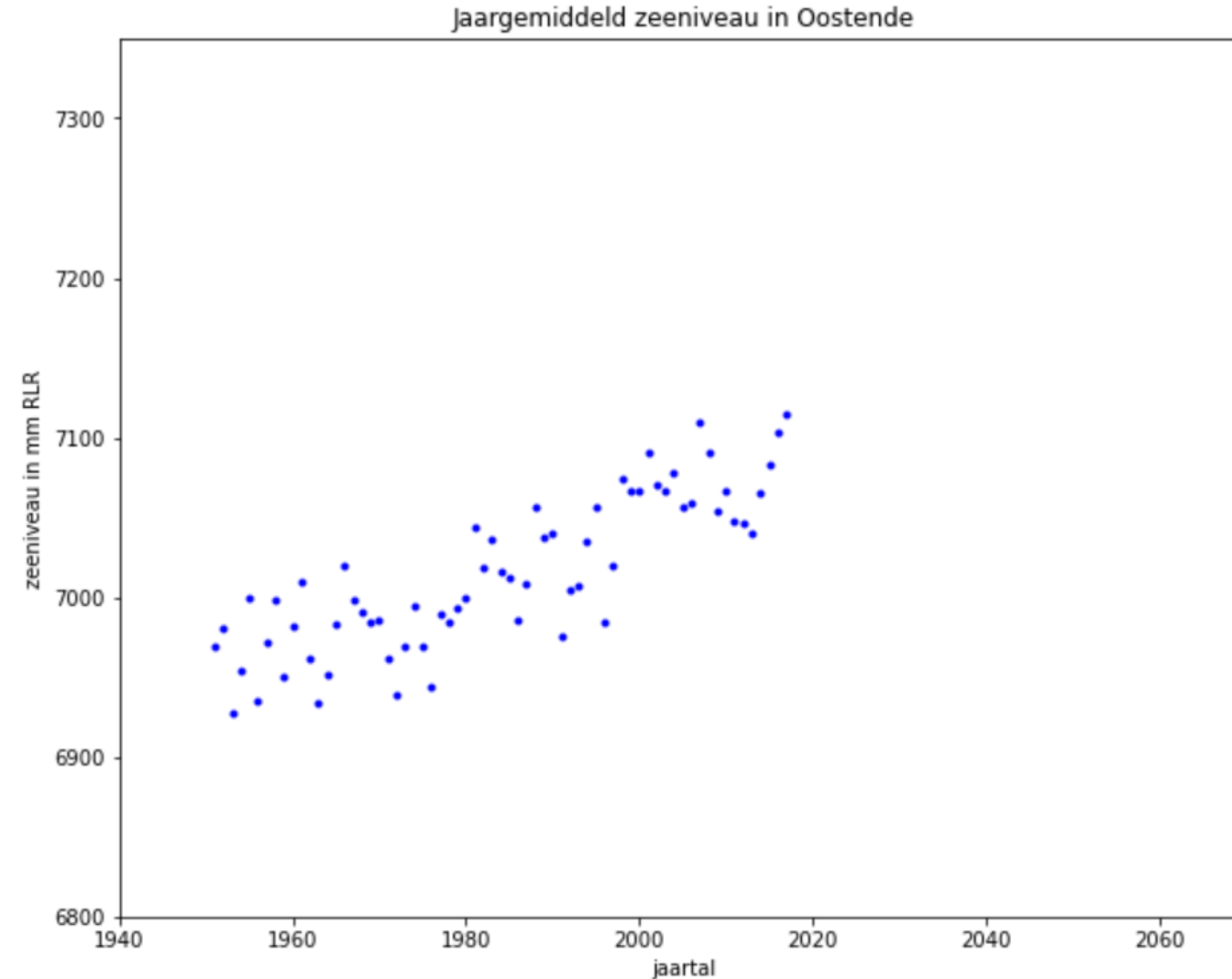


# LINEAIRE REGRESSIE





# Zeeniveau Oostende

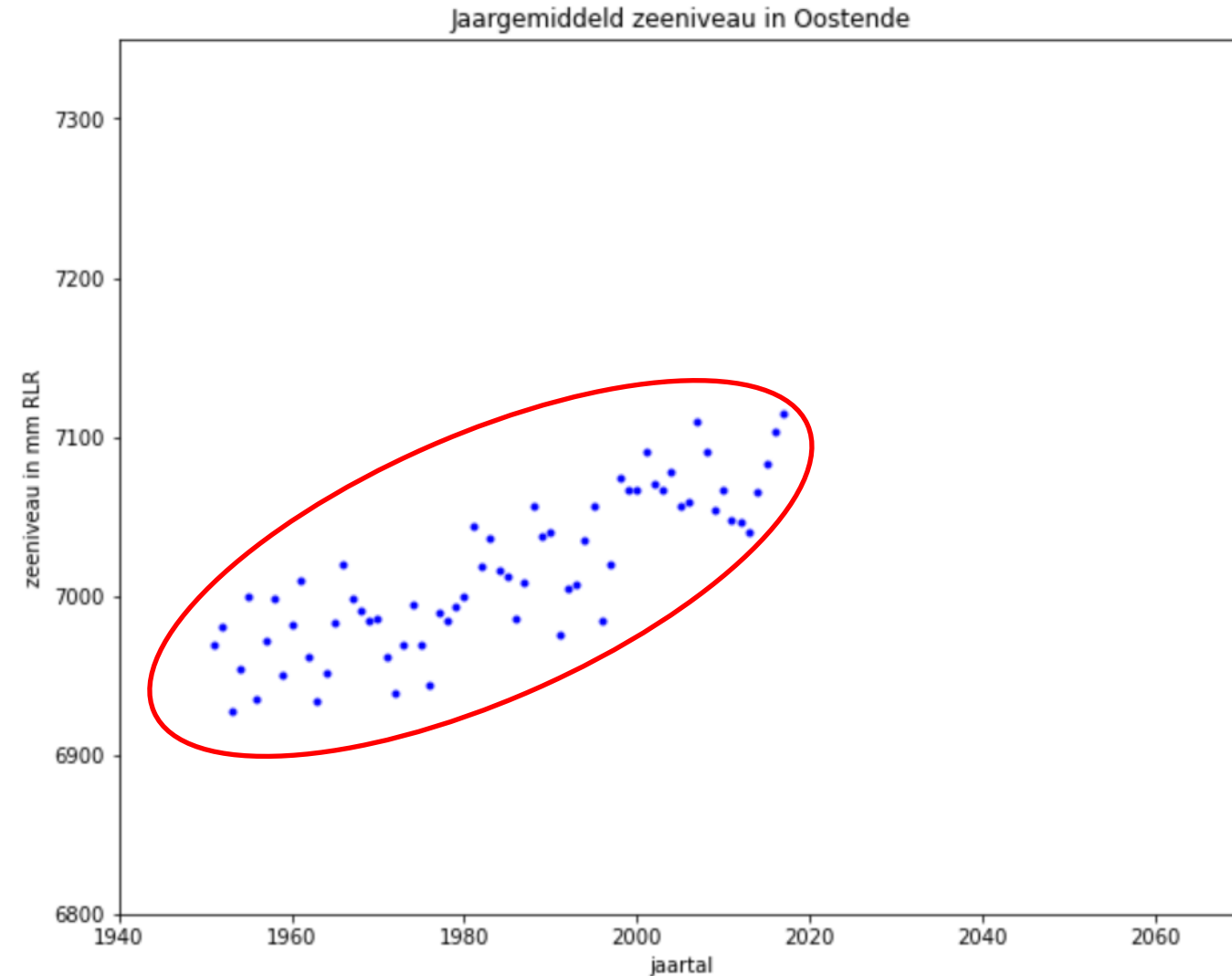


Bron: MIRA - zie [dwengo.org](https://dwengo.org) leerpad Lineaire Regressie > notebook Zeeniveau

<https://dwengo.org/stem/>

Vlaamse Milieumaatschappij (2019). Zeeniveau.  
Geraadpleegd op 21 januari 2020 via  
<https://www.milieurapport.be/milieuthemas/klimaat/verandering/zeeklimaat/zeeniveau/zeeniveau>

# Zeeniveau Oostende

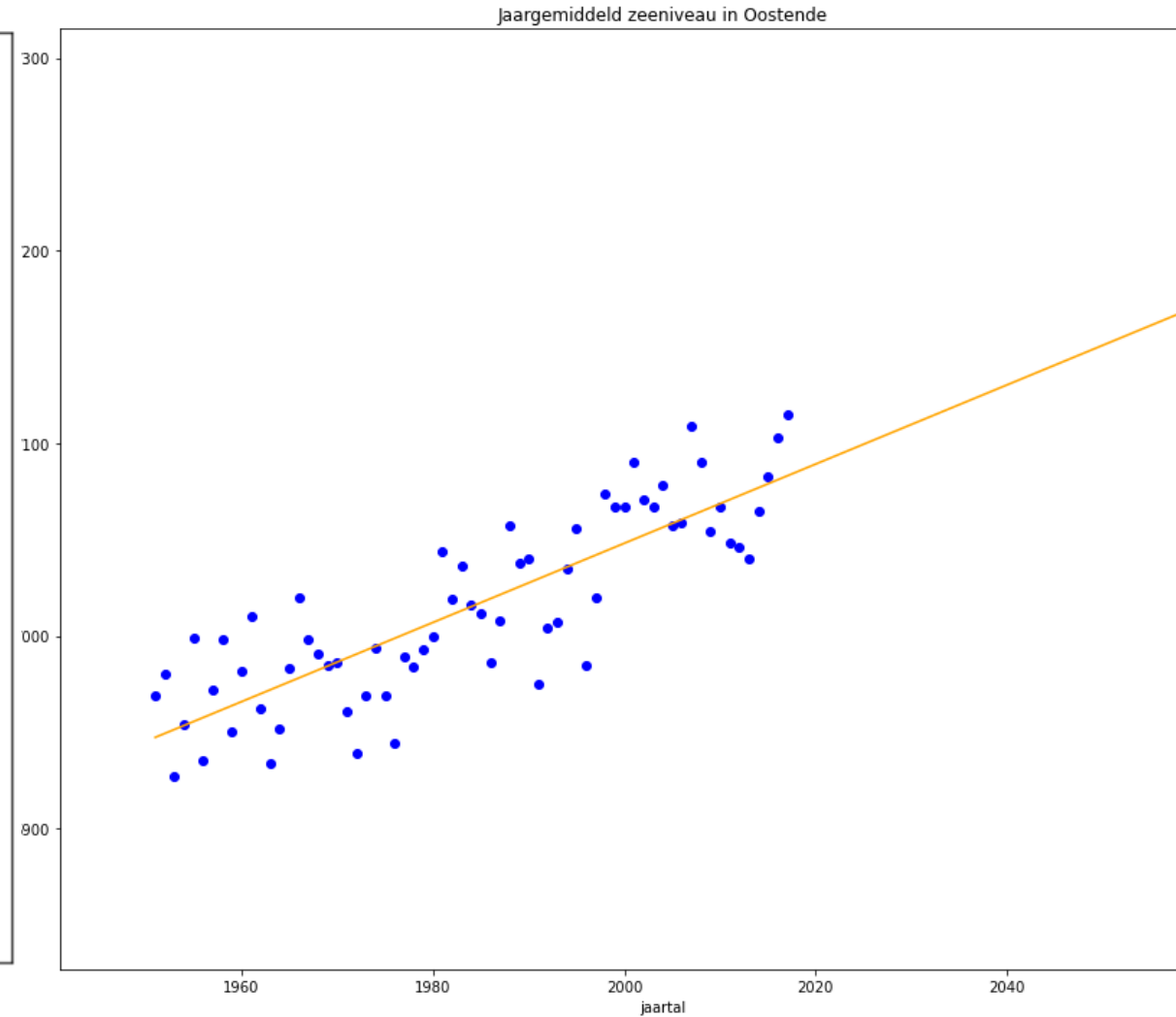
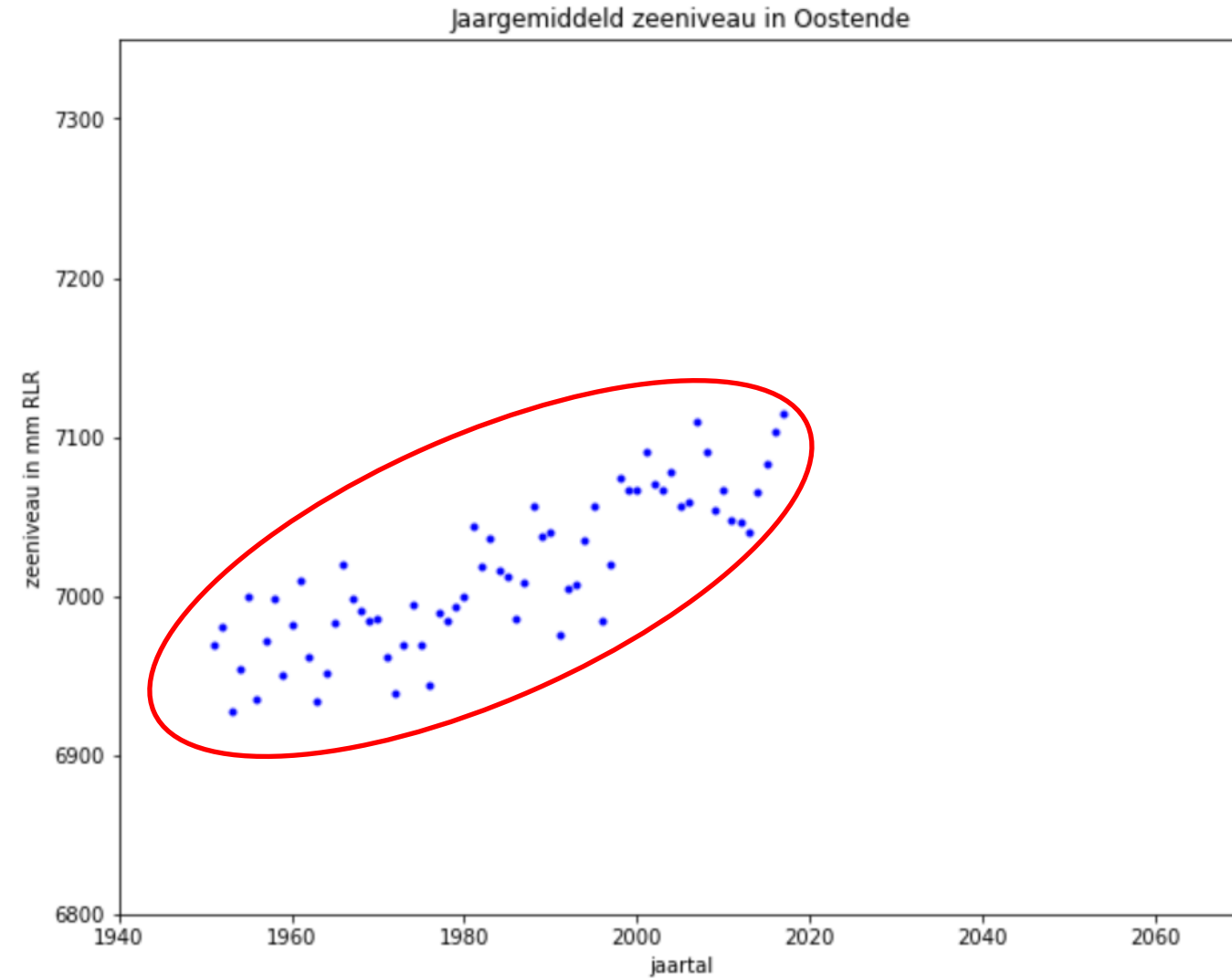


Bron: MIRA - zie [dwengo.org](https://dwengo.org) leerpad Lineaire Regressie > notebook Zeeniveau

<https://dwengo.org/stem/>

Vlaamse Milieumaatschappij (2019). Zeeniveau.  
Geraadpleegd op 21 januari 2020 via  
<https://www.milieurapport.be/milieuthemas/klimaat/verandering/zeeklimaat/zeeniveau/zeeniveau>

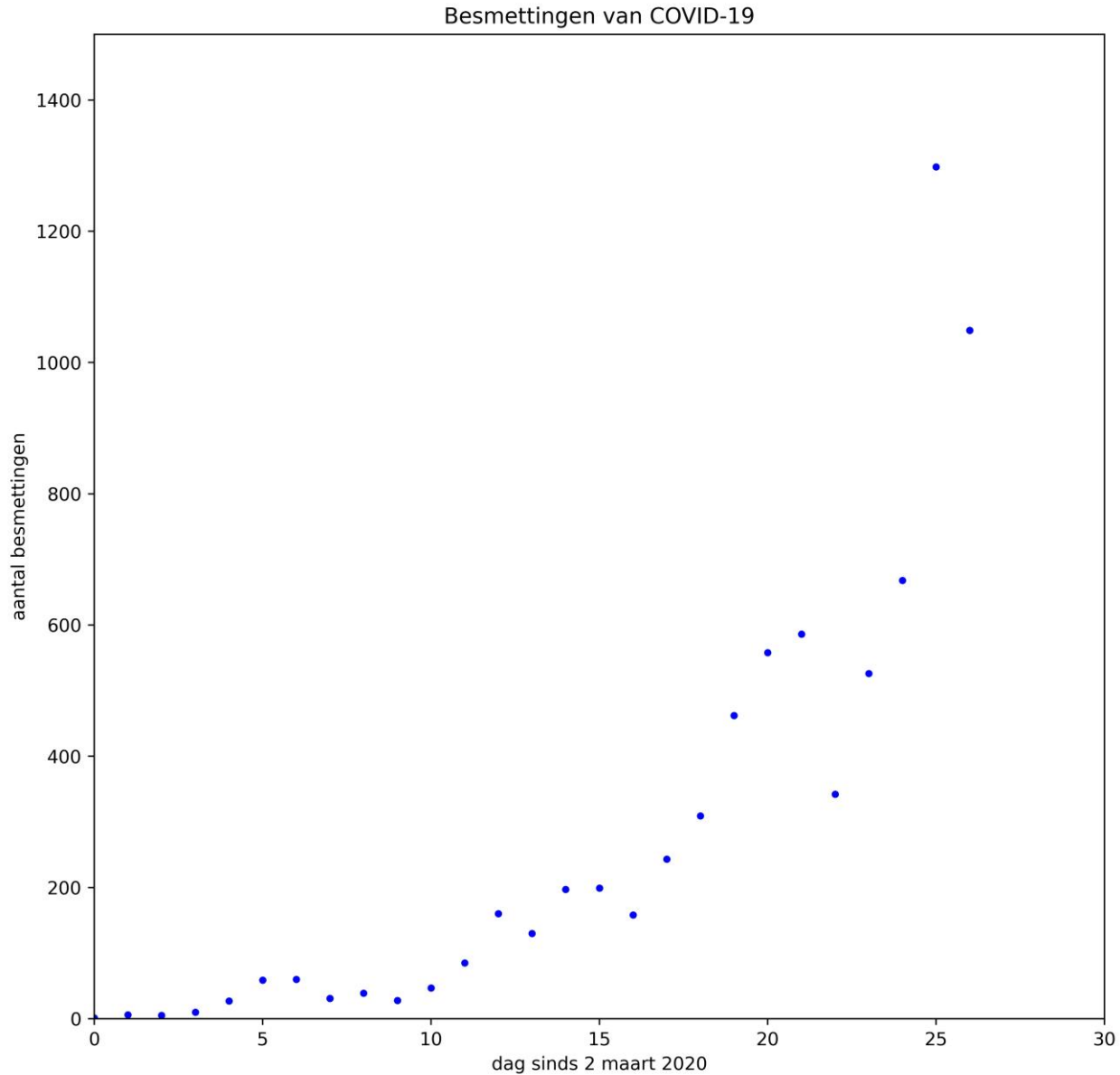
# Zeeniveau Oostende



Bron: MIRA - zie [dwengo.org](https://dwengo.org) leerpad Lineaire Regressie > notebook Zeeniveau

<https://dwengo.org/stem/>

# COVID-19

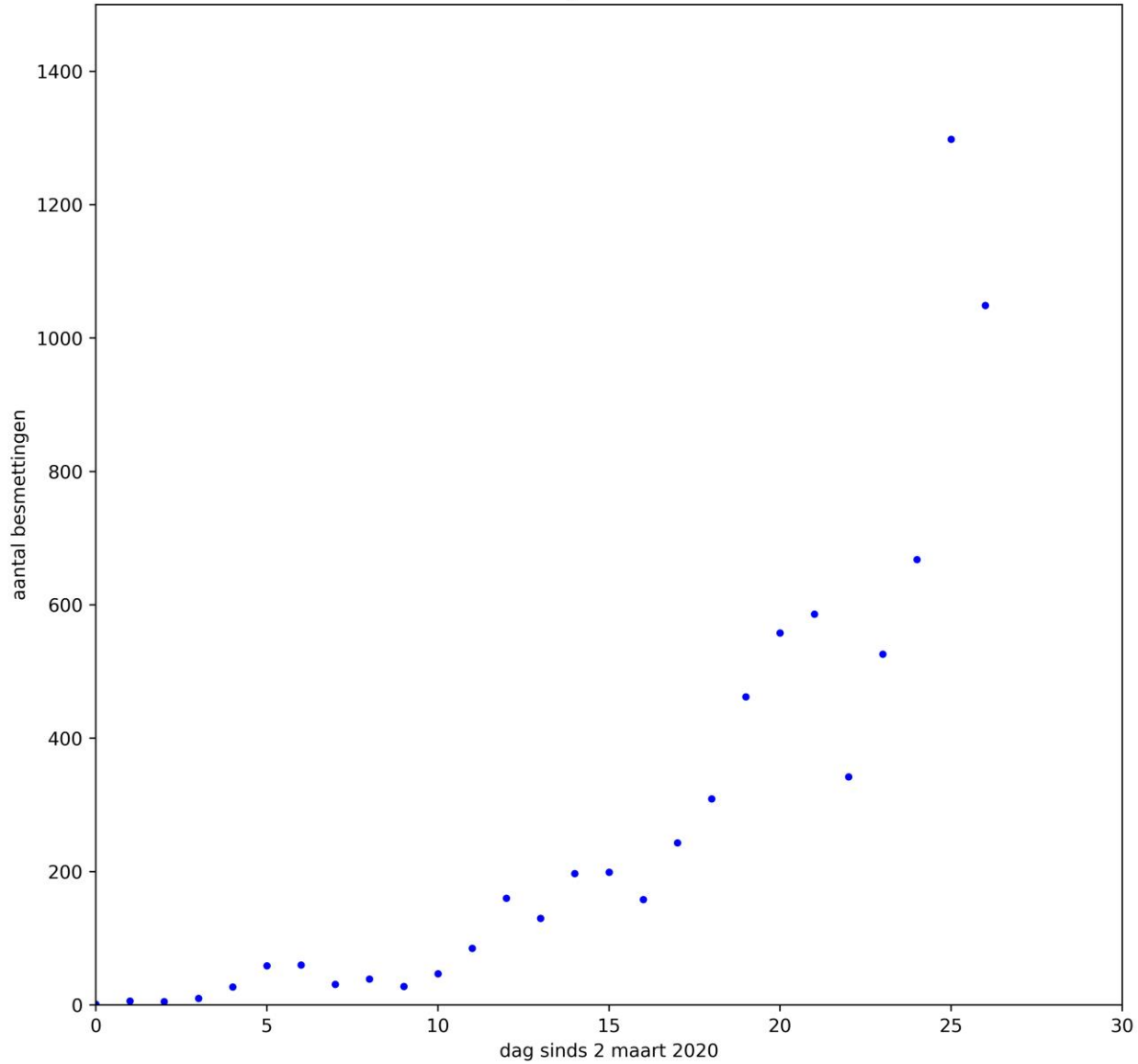


Datawrapper. Corona in België.

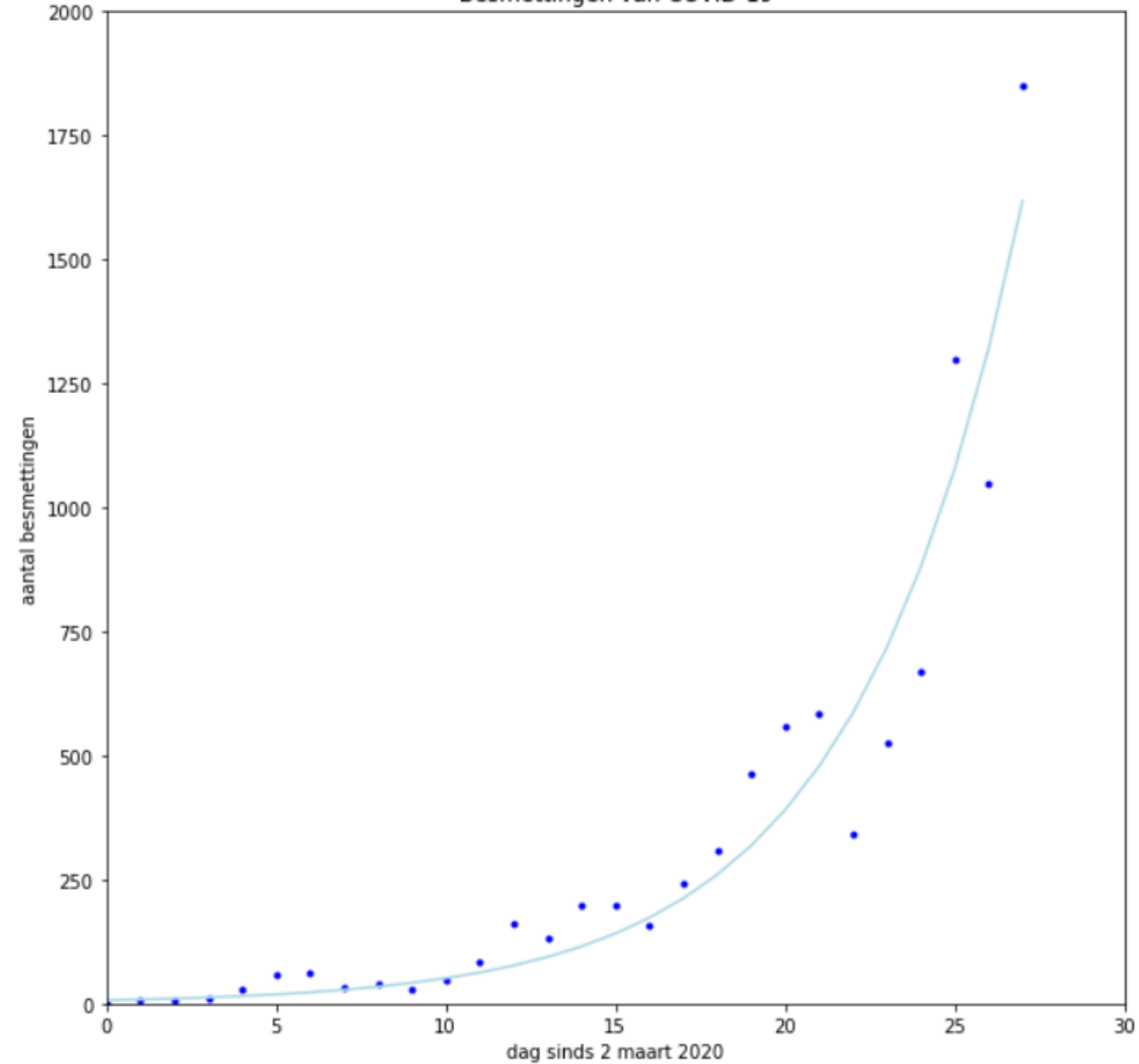
[https://www.datawrapper.de/\\_/dFAEh/](https://www.datawrapper.de/_/dFAEh/) opgesteld door Het Laatste Nieuws met cijfers van de FOD Gezondheid met het aantal gehospitaliseerden, mensen op intensieve zorg en sterfgevallen.

# COVID-19

Besmettingen van COVID-19



Besmettingen van COVID-19

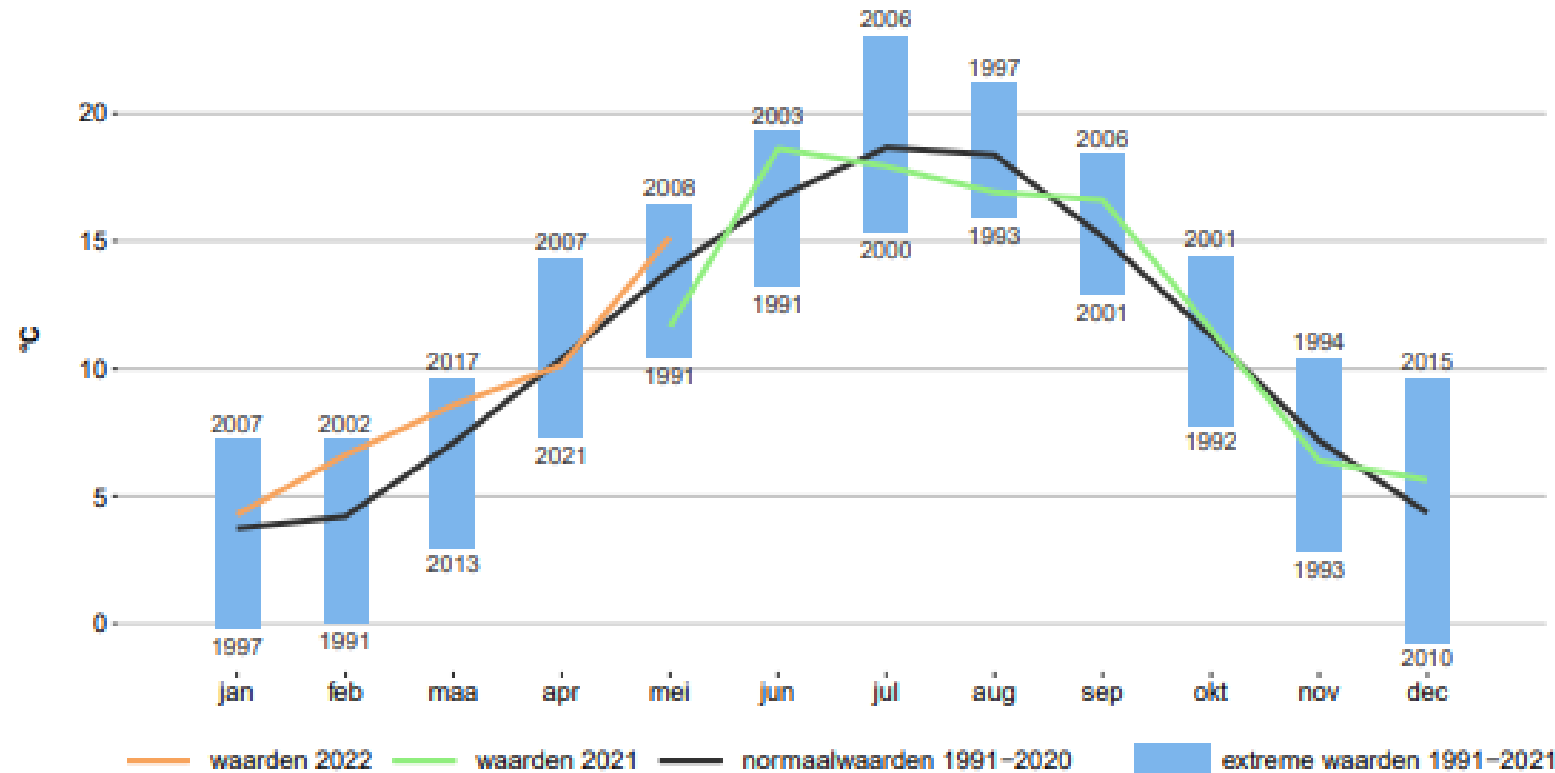


# Gemiddelde maandtemperatuur, Ukkel



## Gemiddelde maandtemperatuur, Ukkel

recente waarden, normaalwaarden (1991-2020) en extreme waarden (1991-2021)



<https://www.meteo.be/nl/klimaat/klimaat-van-belgie/klimatologisch-overzicht/2022/mei#&gid=1&pid=5>

<https://www.meteo.be/nl/klimaat/klimaat-van-belgie/klimaatnormalen-te-ukkel/luchttemperatuur/gemiddelde-temperatuur>

# Gemiddelde maandtemperatuur, Ukkel

maand,gemiddeldetemp

1,3.7

2,4.2

3,7.1

4,10.4

5,13.9

6,16.7

7,18.7

8,18.4

9,15.2

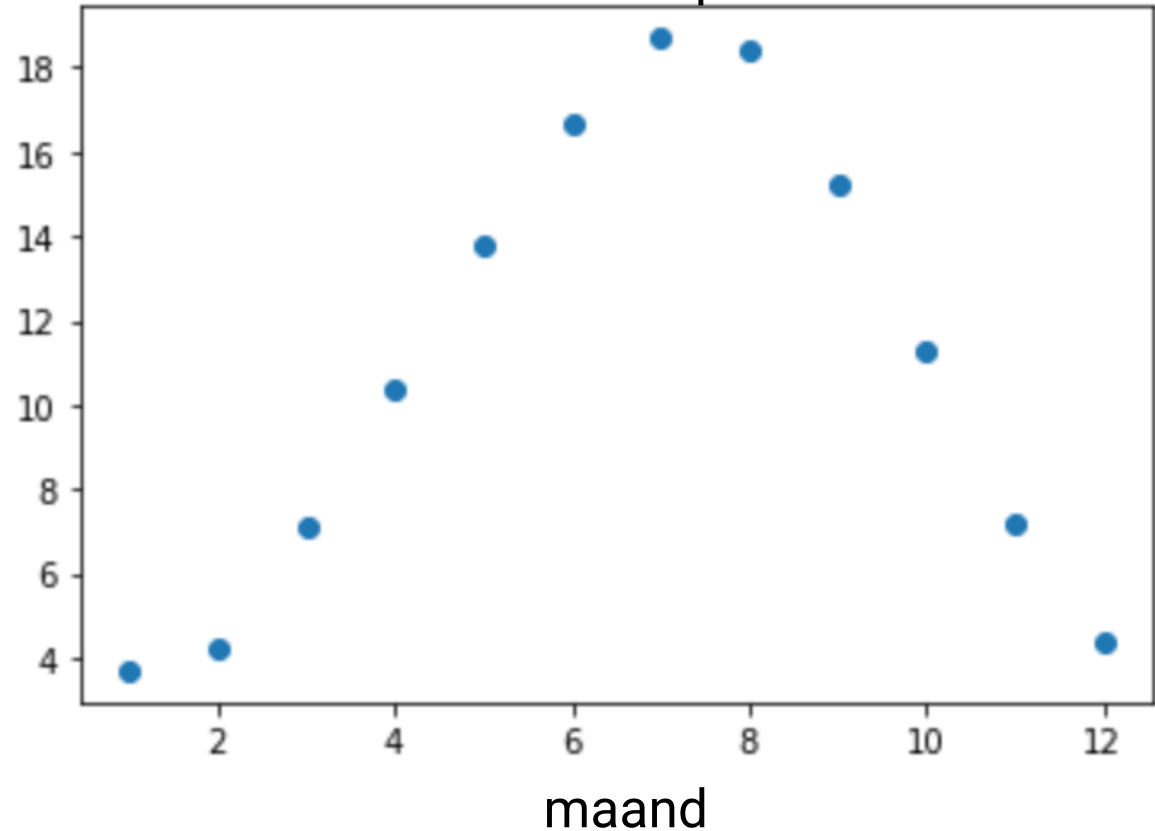
10,11.3

11,7.2

12,4.3

temperatuur  
in °C

Gemiddelde maandtemperatuur in Ukkel



<https://www.meteo.be/nl/klimaat/klimaat-van-belgie/klimaatnormalen-te-ukkel/luchttemperatuur/gemiddelde-temperatuur>

<https://dwengo.org/wiskunde>



# Gemiddelde maandtemperatuur, Ukkel

maand,gemiddeldetemp

1,3.7

2,4.2

3,7.1

4,10.4

5,13.9

6,16.7

7,18.7

8,18.4

9,15.2

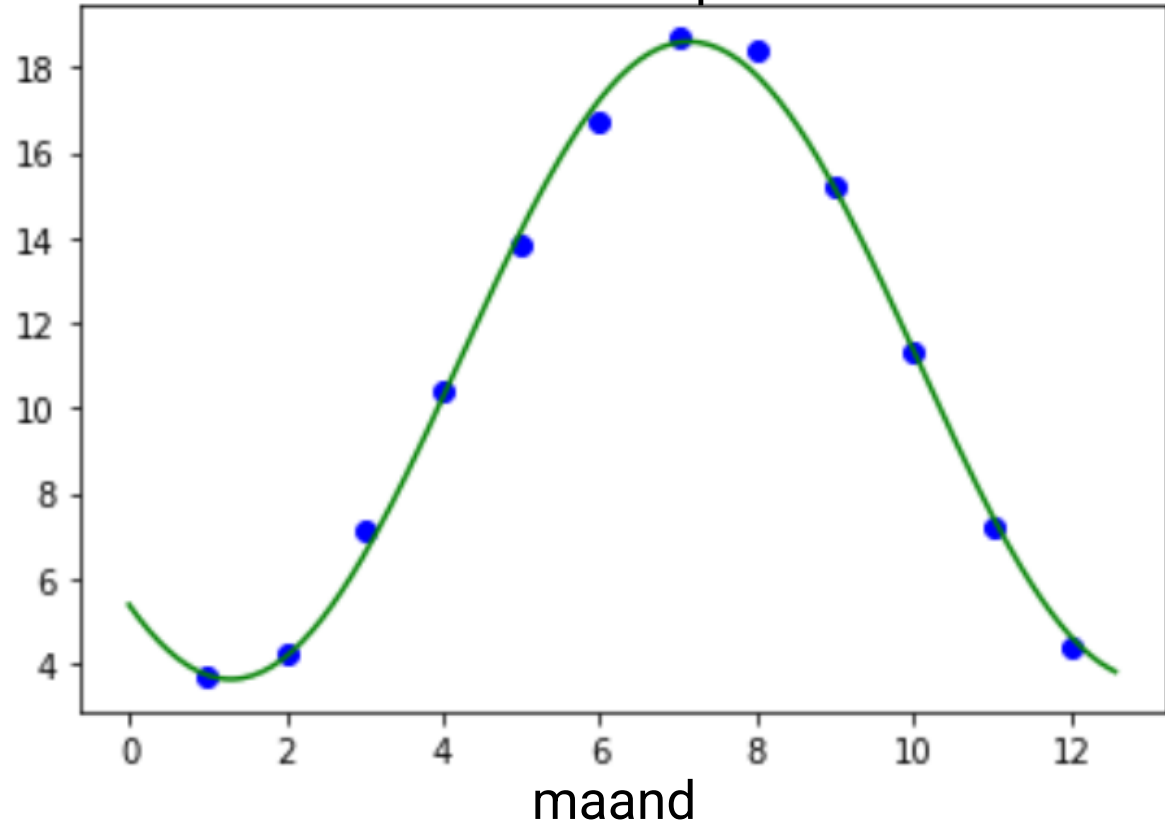
10,11.3

11,7.2

12,4.3

temperatuur  
in °C

Gemiddelde maandtemperatuur in Ukkel

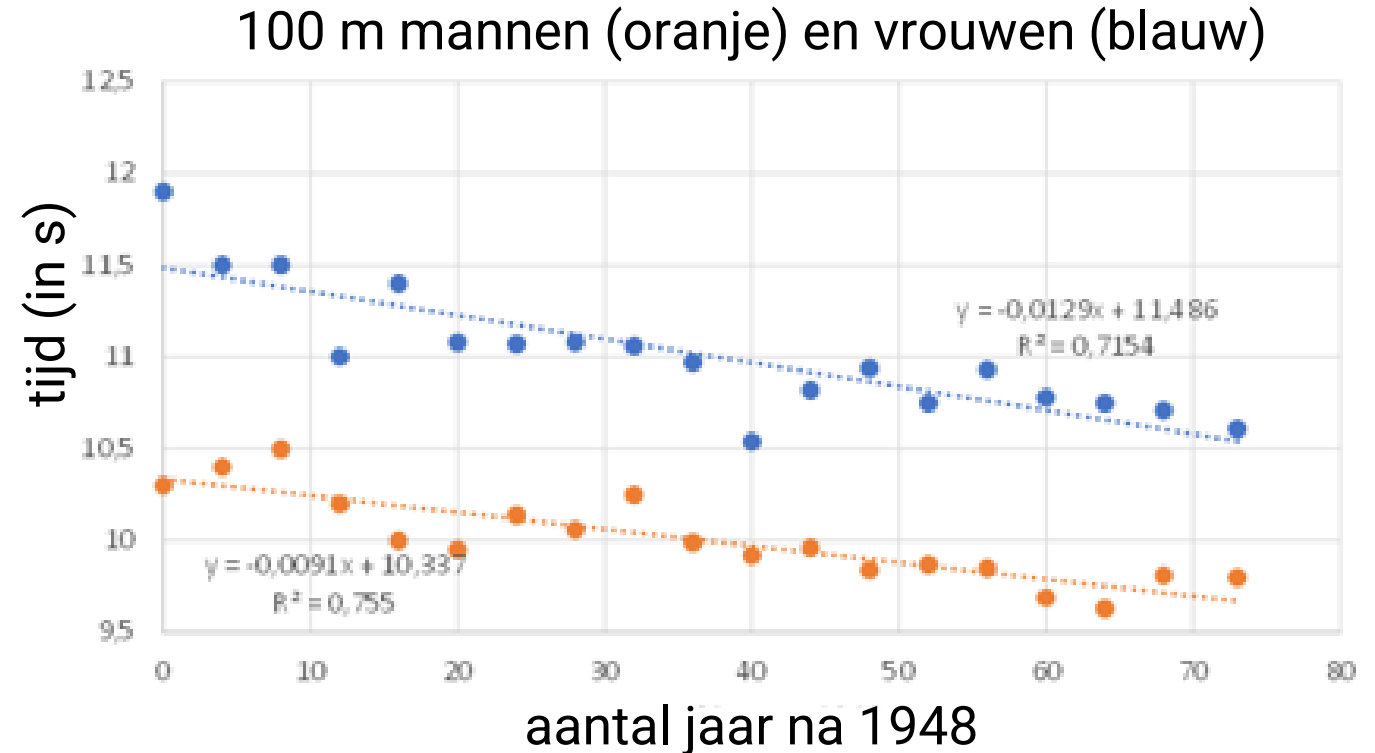


<https://www.meteo.be/nl/klimaat/klimaat-van-belgie/klimaatnormalen-te-ukkel/luchttemperatuur/gemiddelde-temperatuur>

<https://dwengo.org/wiskunde>

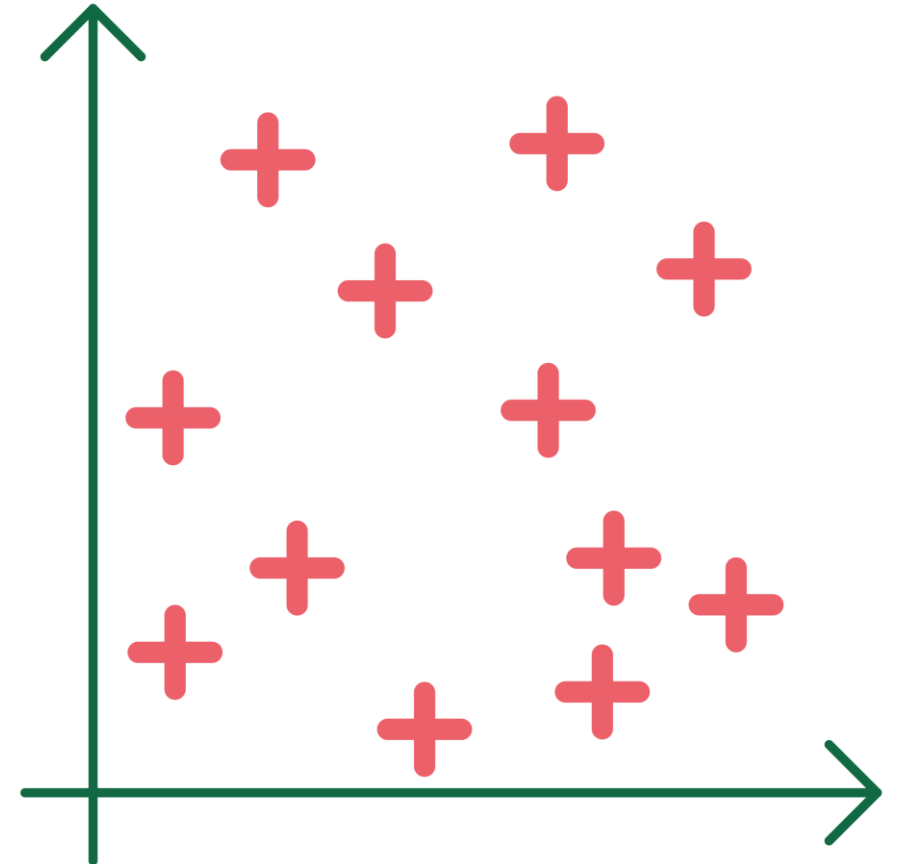
# Winnaars 100 meter sprint op Olympische Spelen

- vrouwen:  $y = -0,0129x + 11,486$
- mannen:  $y = -0,0091x + 10,337$
- wanneer lopen de vrouwen sneller dan de mannen?  
editie van 2252
- wanneer lopen de vrouwen de 100 meter in 0 s?  
editie van 2886
- deze antwoorden zijn natuurlijk niet zinvol meer...
- verregaande extrapolatie: is een rechte trendlijn dan nog wel gerechtvaardigd?



# CORRELATIE

VISUALISEREN EN STANDAARDISEREN  
CORRELATIECOEFFICIENT  
OORZAKELIJK VERBAND



# Oorzakelijk verband?

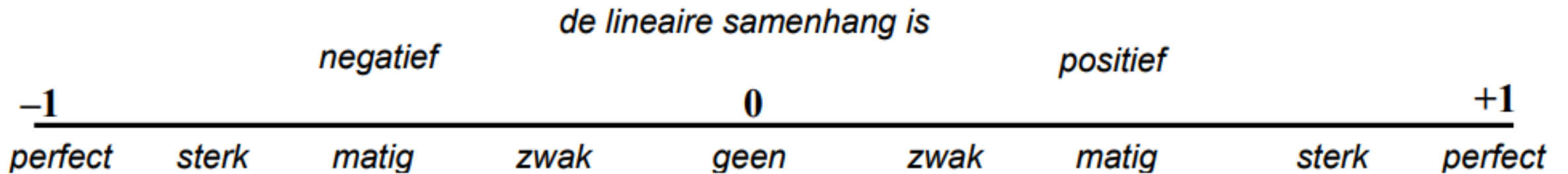
Omwille van copyrightredenen is deze afbeelding verwijderd. Via de onderstaand links kan je de afbeelding gemakkelijk terugvinden.

Vraag: er is een positieve correlatie tussen het chocoladeverbruik in een land en het aantal Nobelprijzen. Is dit een causal verband?

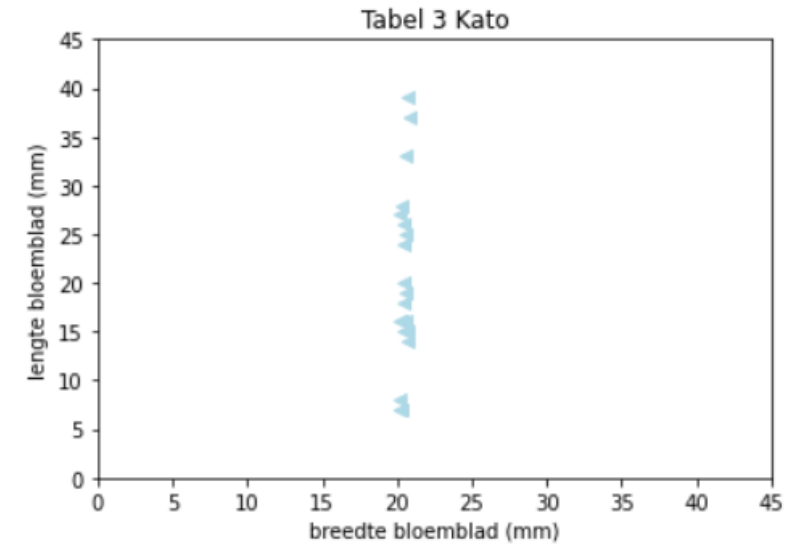
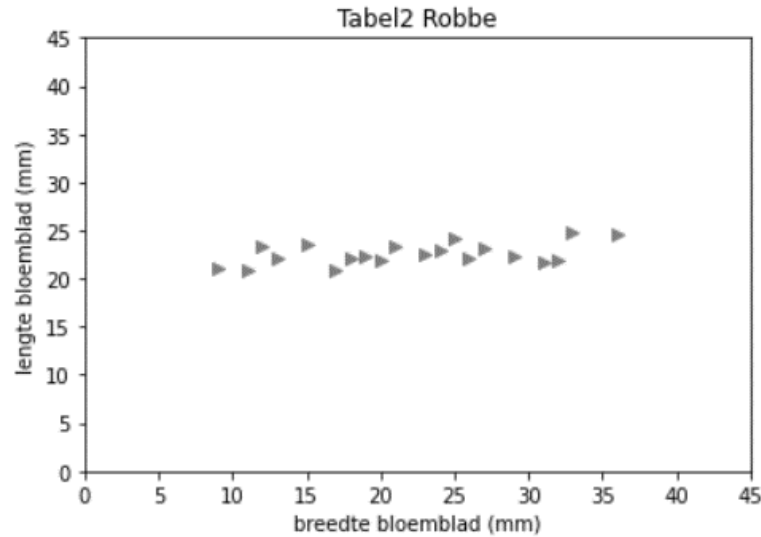
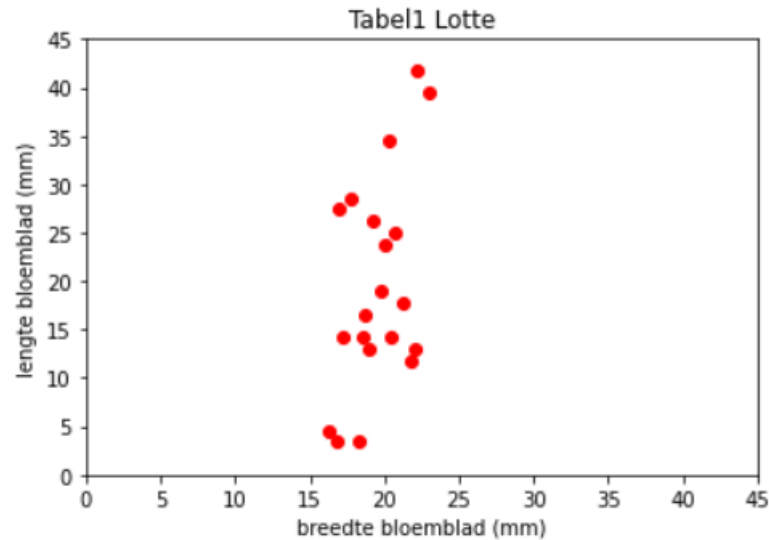
[https://www.biostat.jhsph.edu/courses/bio621/misc/Chocolate%20consumption%20cognitive%20function%20and%20nobel%20laurates%20\(NJM\).pdf](https://www.biostat.jhsph.edu/courses/bio621/misc/Chocolate%20consumption%20cognitive%20function%20and%20nobel%20laurates%20(NJM).pdf)

Messerli, Franz. (2012). Chocolate Consumption, Cognitive Function, and Nobel Laureates. The New England journal of medicine. 367. 1562-4.

# Lineaire samenhang - Correlatiecoëfficiënt



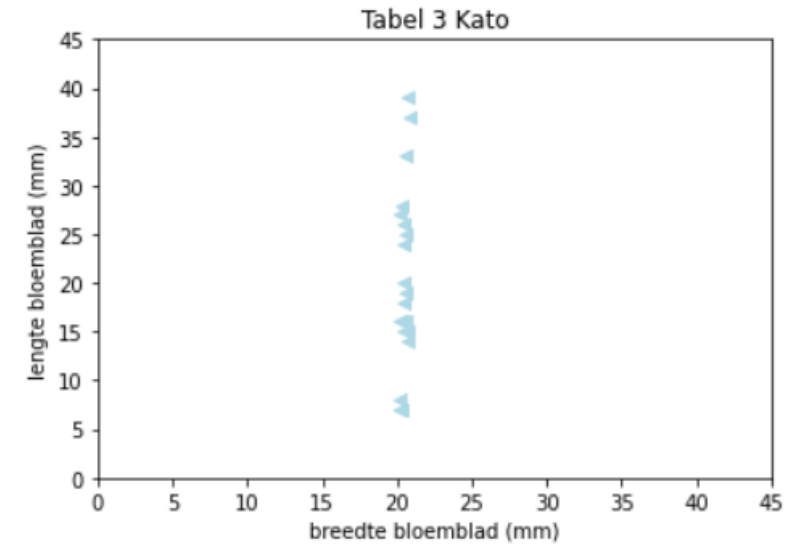
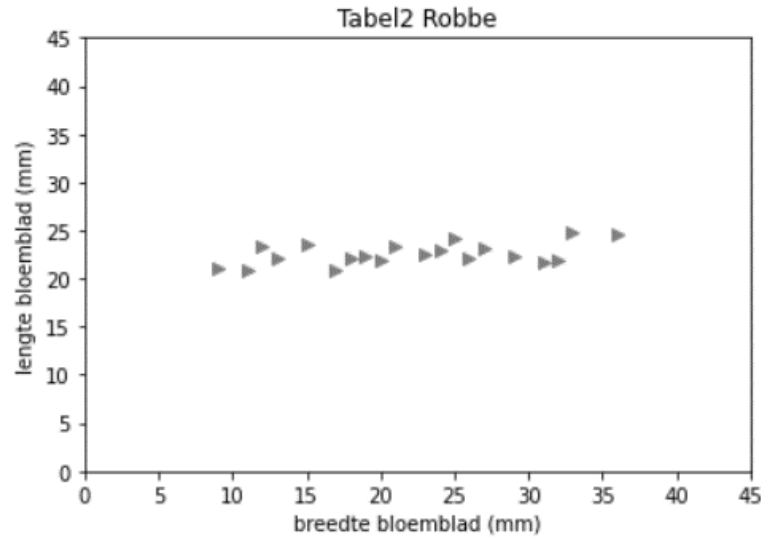
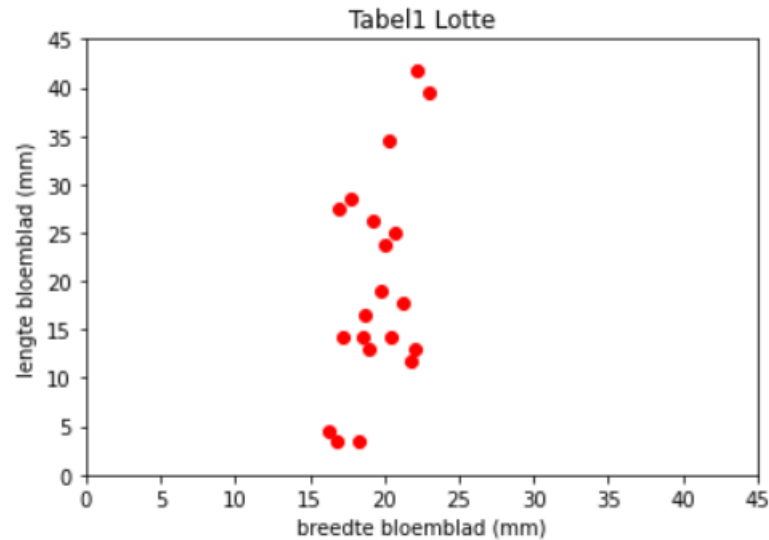
# Voorbeeld - Meeste lineaire samenhang?



Callaert, H., & Bogaerts, S. (2004). Statistische Intelligentie: de samenhang ontdekken: exploratie van bivariaat cijfermateriaal. Universiteit Hasselt. Geraadpleegd op 15 april 2019 via <https://docplayer.nl/32671814-Statistische-intelligentie.html>

<https://dwengo.org/wiskunde>

# Voorbeeld - Meeste lineaire samenhang?



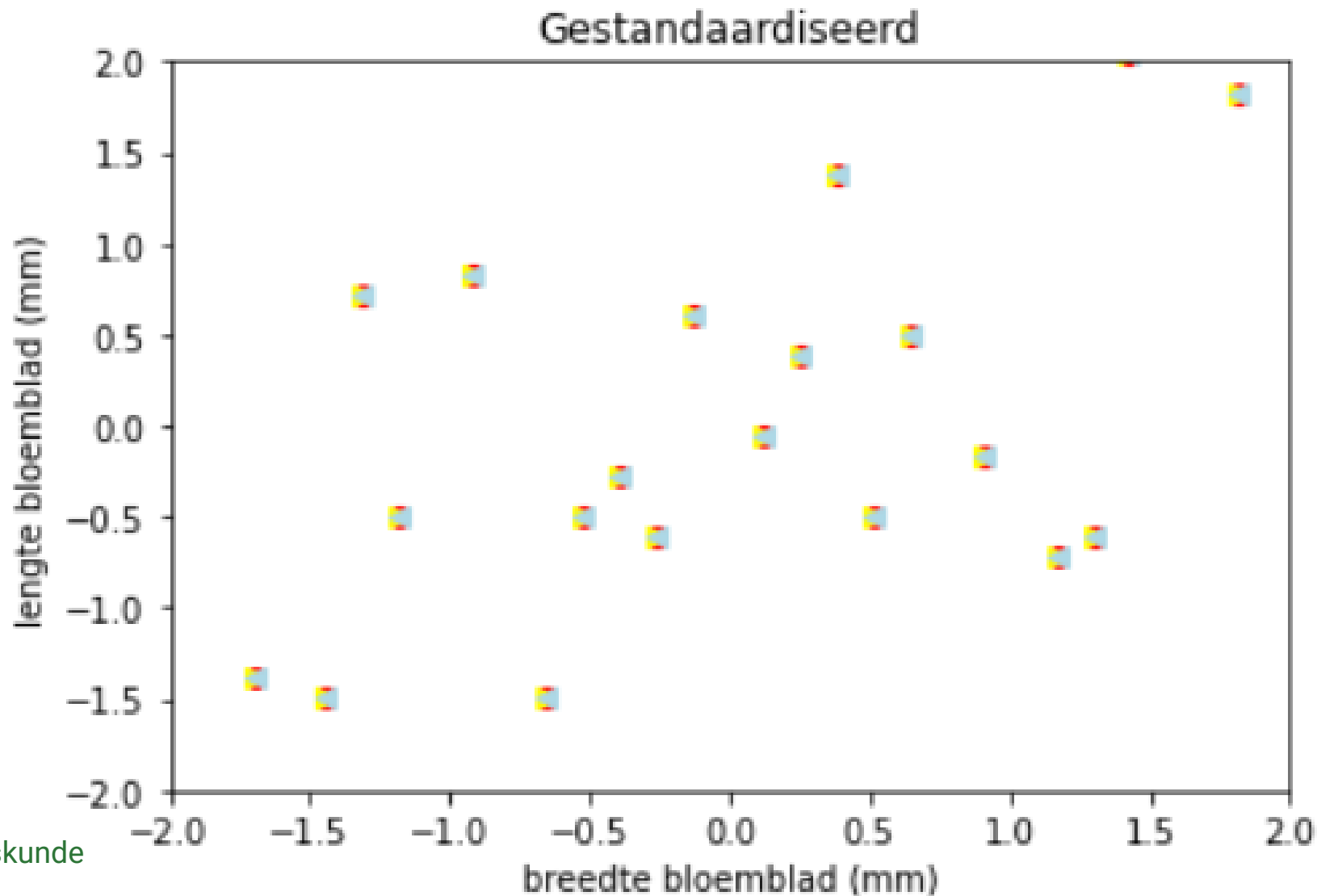
Dezelfde correlatiecoëfficiënt!

Callaert, H., & Bogaerts, S. (2004). Statistische Intelligentie: de samenhang ontdekken: exploratie van bivariaat cijfermateriaal. Universiteit Hasselt. Geraadpleegd op 15 april 2019 via <https://docplayer.nl/32671814-Statistische-intelligentie.html>

<https://dwengo.org/wiskunde>



# Belang van standaardiseren

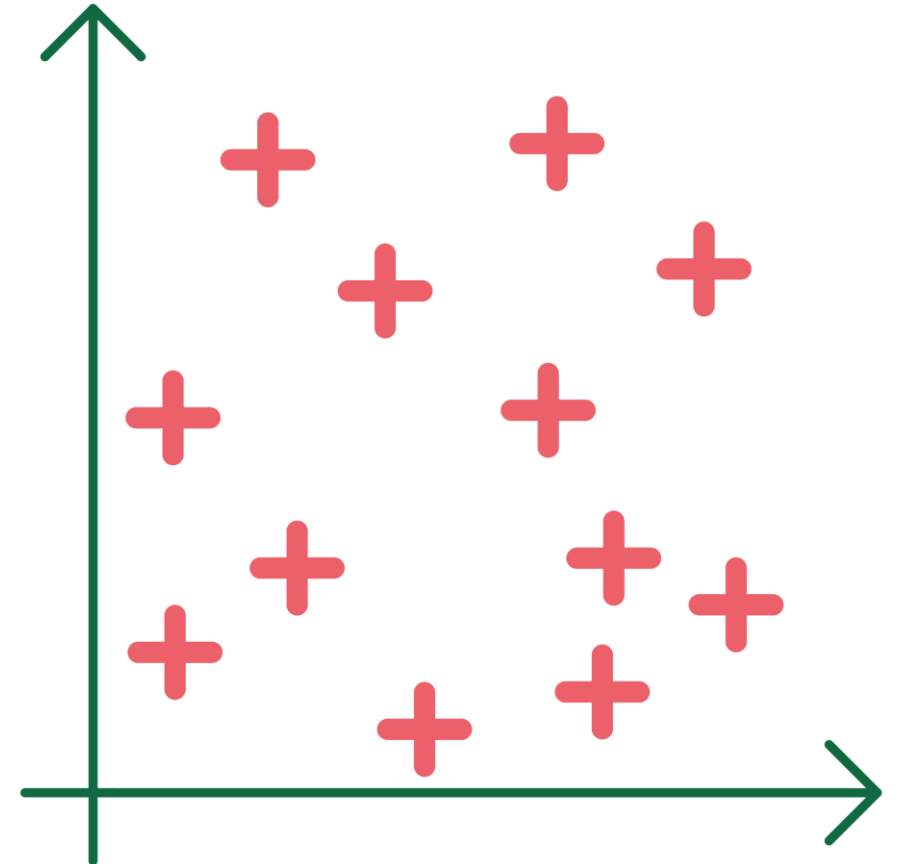


$$x \rightarrow \frac{x - \mu}{\sigma}$$

<https://dwengo.org/wiskunde>

Callaert, H., & Bogaerts, S. (2004). Statistische Intelligentie: de samenhang ontdekken: exploratie van bivariaat cijfermateriaal. Universiteit Hasselt. Geraadpleegd op 15 april 2019 via <https://docplayer.nl/32671814-Statistische-intelligentie.html>.

## BELANG VAN VISUALISATIE



# Voorbeeld – Datasets met vergelijkbare data

	a	h
0	55.3846	97.1795
1	51.5385	96.0256
2	46.1538	94.4872
3	42.8205	91.4103
4	40.7692	88.3333
...	...	...
137	39.4872	25.3846
138	91.2821	41.5385
139	50.0000	95.7692
140	47.9487	95.0000
141	44.1026	92.6923

```
print(x_gemiddelde, y_gemiddelde)  
print(x_variantie, y_variantie)
```

```
54.26327323943662 47.832252816901416  
279.0906214819599 720.4066936731963
```

Getallen van dezelfde  
grootte-orde.

a en b hebben ongeveer  
hetzelfde gemiddelde  
en dezelfde variantie,  
h en k hebben ongeveer  
hetzelfde gemiddelde  
en dezelfde variantie.

Correlatiecoëfficiënten  
zijn -0,0645 en -0,0686.

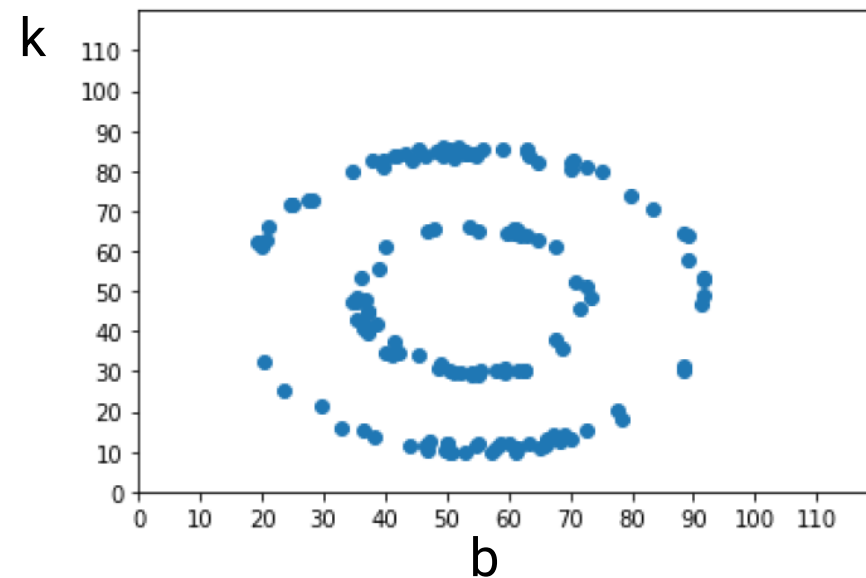
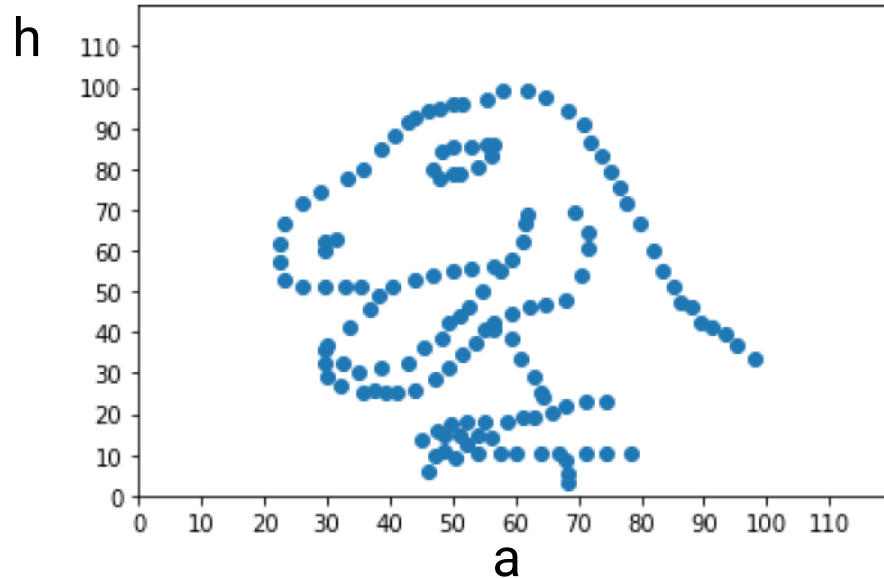
	b	k
0	51.203891	83.339777
1	58.974470	85.499818
2	51.872073	85.829738
3	48.179931	85.045117
4	41.683200	84.017941
...	...	...
137	50.967748	29.679774
138	91.191054	46.674343
139	55.863768	85.336487
140	49.280595	84.048823
141	43.368502	84.332177

```
print(x2_gemiddelde, y2_gemiddelde)  
print(x2_variantie, y2_variantie)
```

```
54.268730022378875 47.830823155217814  
279.227059433059 720.4239823663735
```

```
R1 = -0.06447185270095167  
R2 = -0.06858639425794844
```

# Voorbeeld – Datasets met vergelijkbare data



Cairo, A. (2016). *Download the Datasaurus: Never trust summary statistics alone; always visualize your data* [Blogpost].

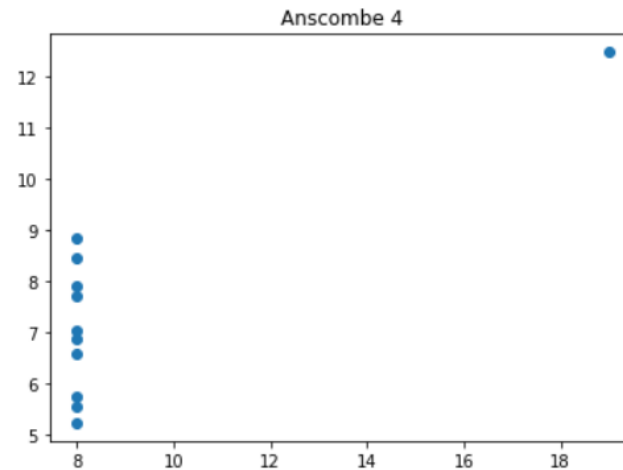
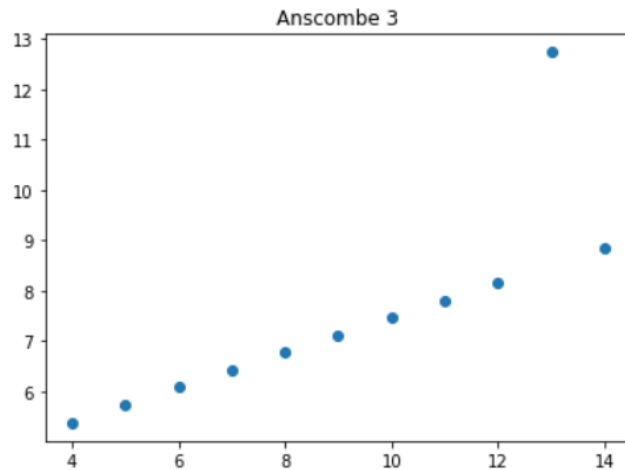
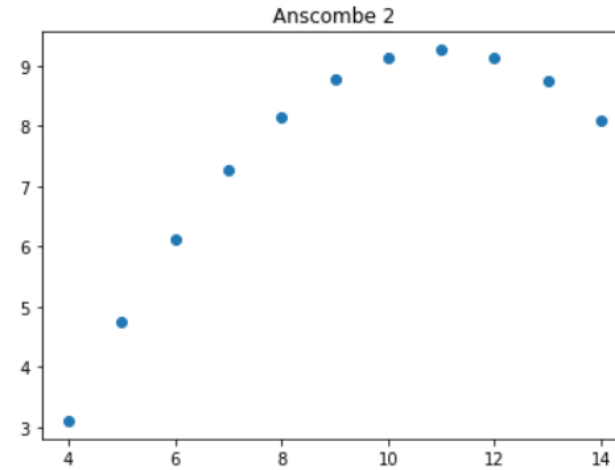
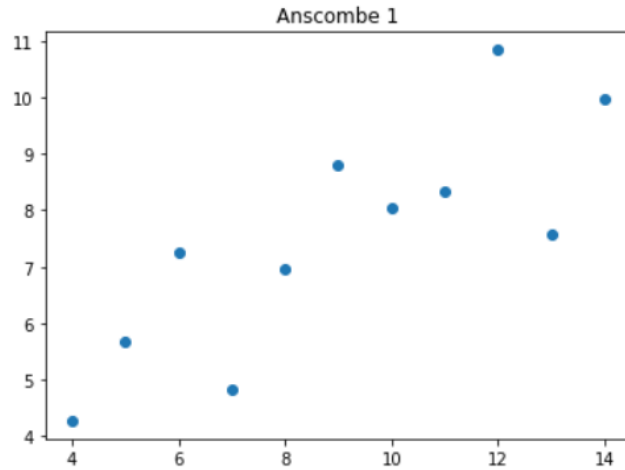
<http://helloworld.cc/datasaurus>

Matejka, J. & Fitzmaurice, G. (2017). Same Stats, Different Graphs: Generating Datasets with Varied Appearance and Identical Statistics through Simulated Annealing. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1290-1294.

<https://doi.org/10.1145/3025453.3025912>

# Belang van visualiseren

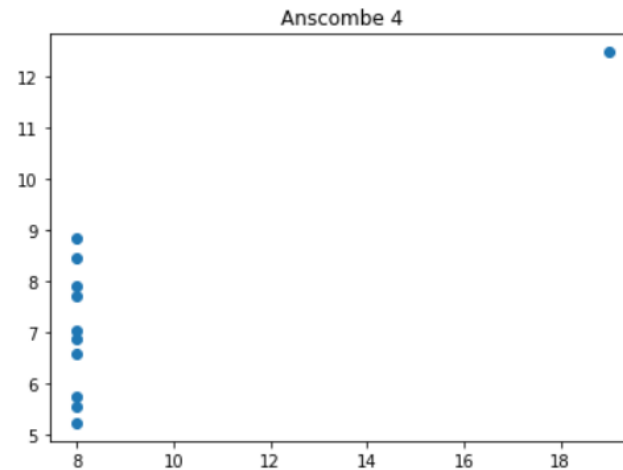
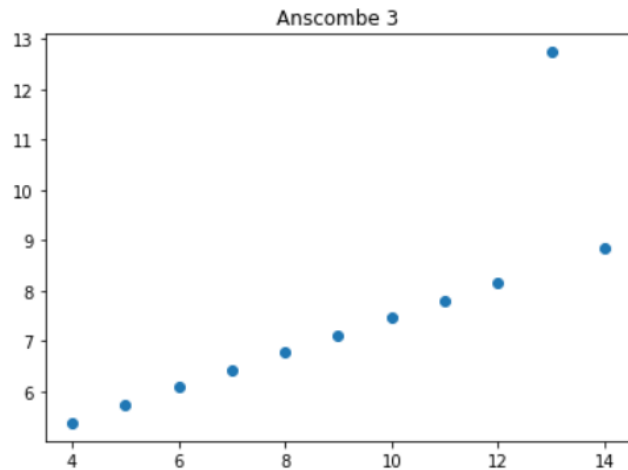
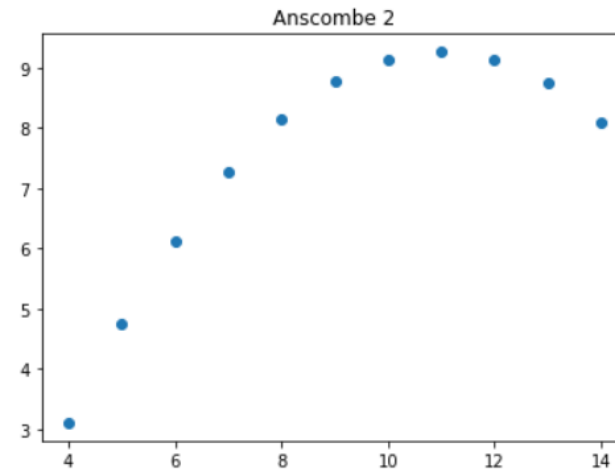
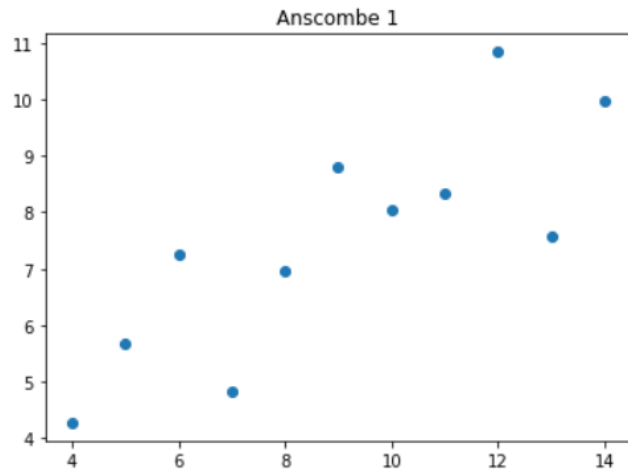
## Anscombe dataset



Hetzelfde gemiddelde voor x-waarden,  
hetzelfde gemiddelde voor y-waarden,  
dezelfde variantie voor de x-waarden,  
dezelfde variantie voor de y-waarden.

Getallen van dezelfde grootte-orde.

# Belang van visualiseren

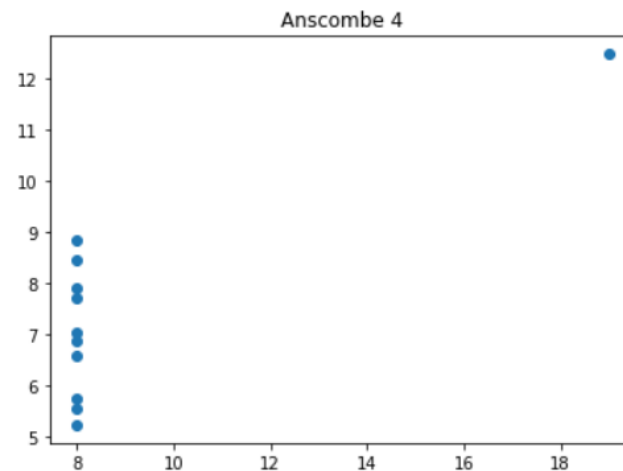
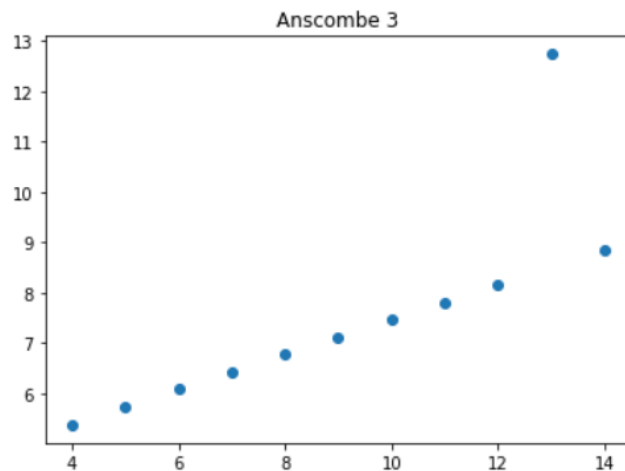
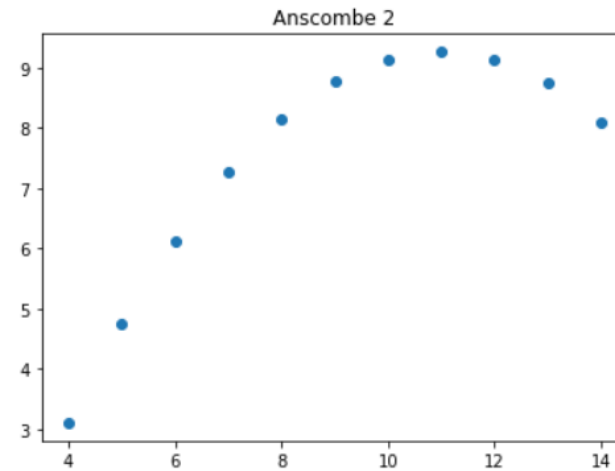
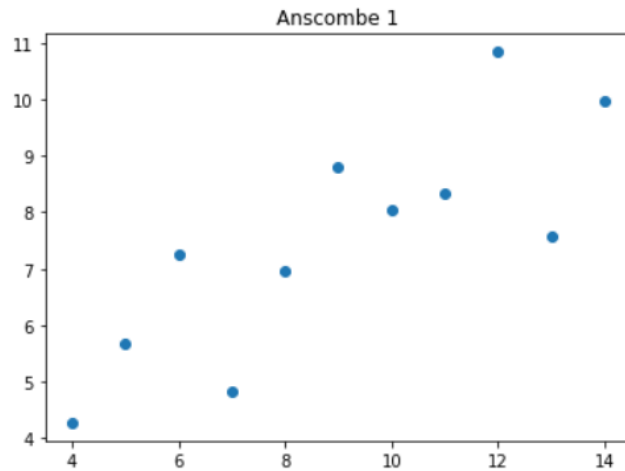


Hetzelfde gemiddelde voor x-waarden,  
hetzelfde gemiddelde voor y-waarden,  
dezelfde variantie voor de x-waarden,  
dezelfde variantie voor de y-waarden.

Getallen van dezelfde grootte-orde.

Welke dataset vertoont de meeste  
lineaire samenhang?

# Belang van visualiseren



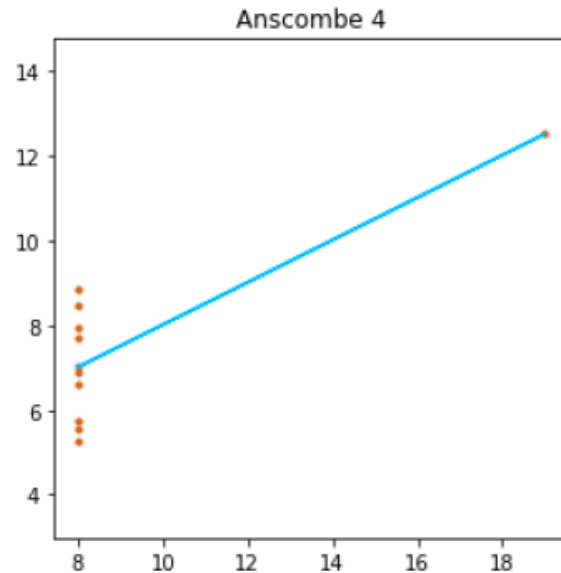
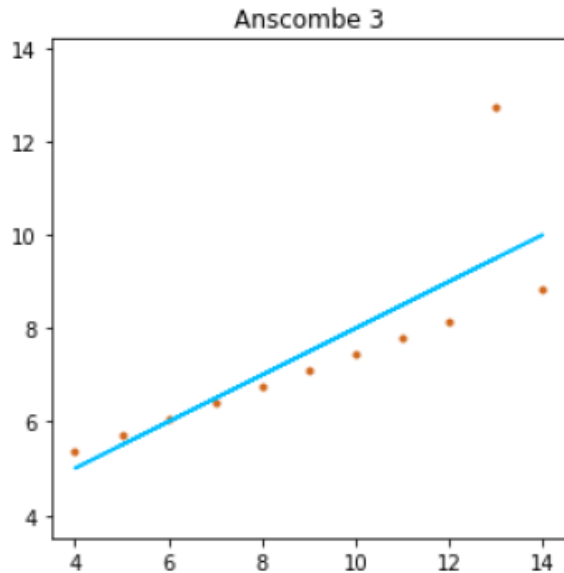
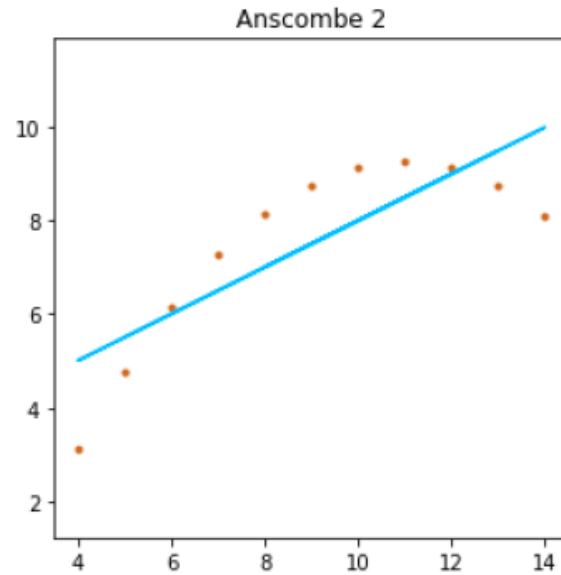
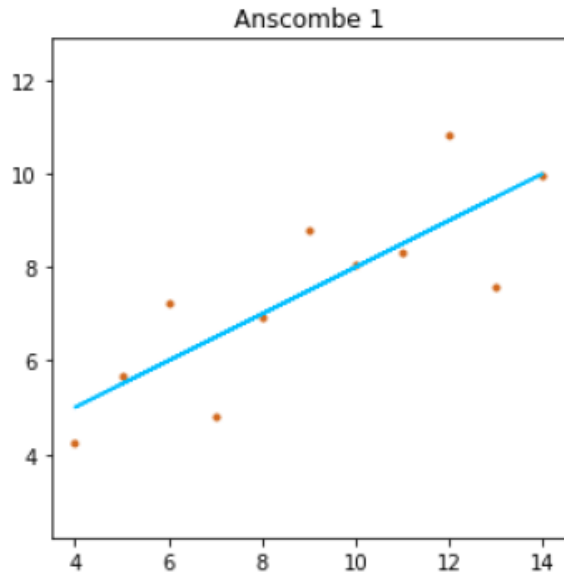
Hetzelfde gemiddelde voor x-waarden,  
hetzelfde gemiddelde voor y-waarden,  
dezelfde variantie voor de x-waarden,  
dezelfde variantie voor de y-waarden.

Getallen van dezelfde grootte-orde.

Welke dataset vertoont de meeste  
lineaire samenhang?

Dezelfde correlatiecoëfficiënt!

# Belang van visualiseren



Hetzelfde gemiddelde voor x-waarden,  
hetzelfde gemiddelde voor y-waarden,  
dezelfde variantie voor de x-waarden,  
dezelfde variantie voor de y-waarden.

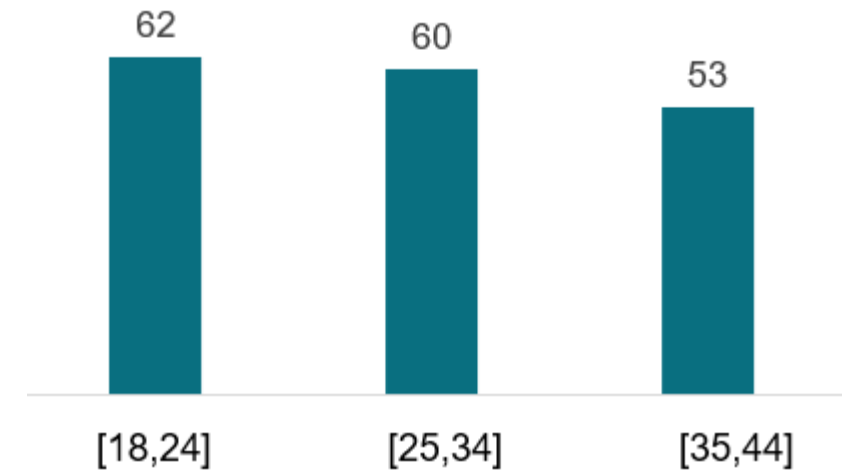
Getallen van dezelfde grootte-orde.

Welke dataset vertoont de meeste  
lineaire samenhang?

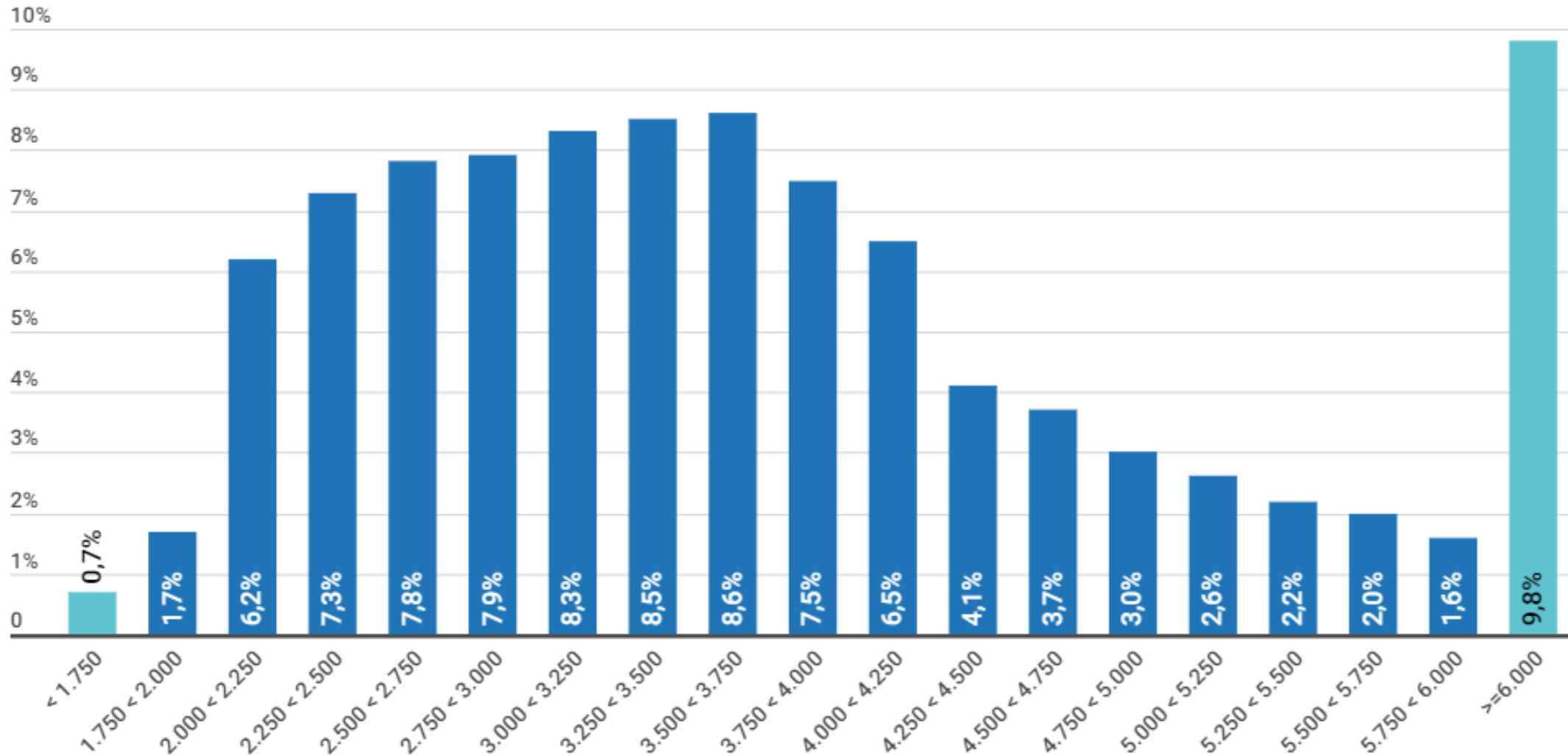
Dezelfde correlatiecoëfficiënt!



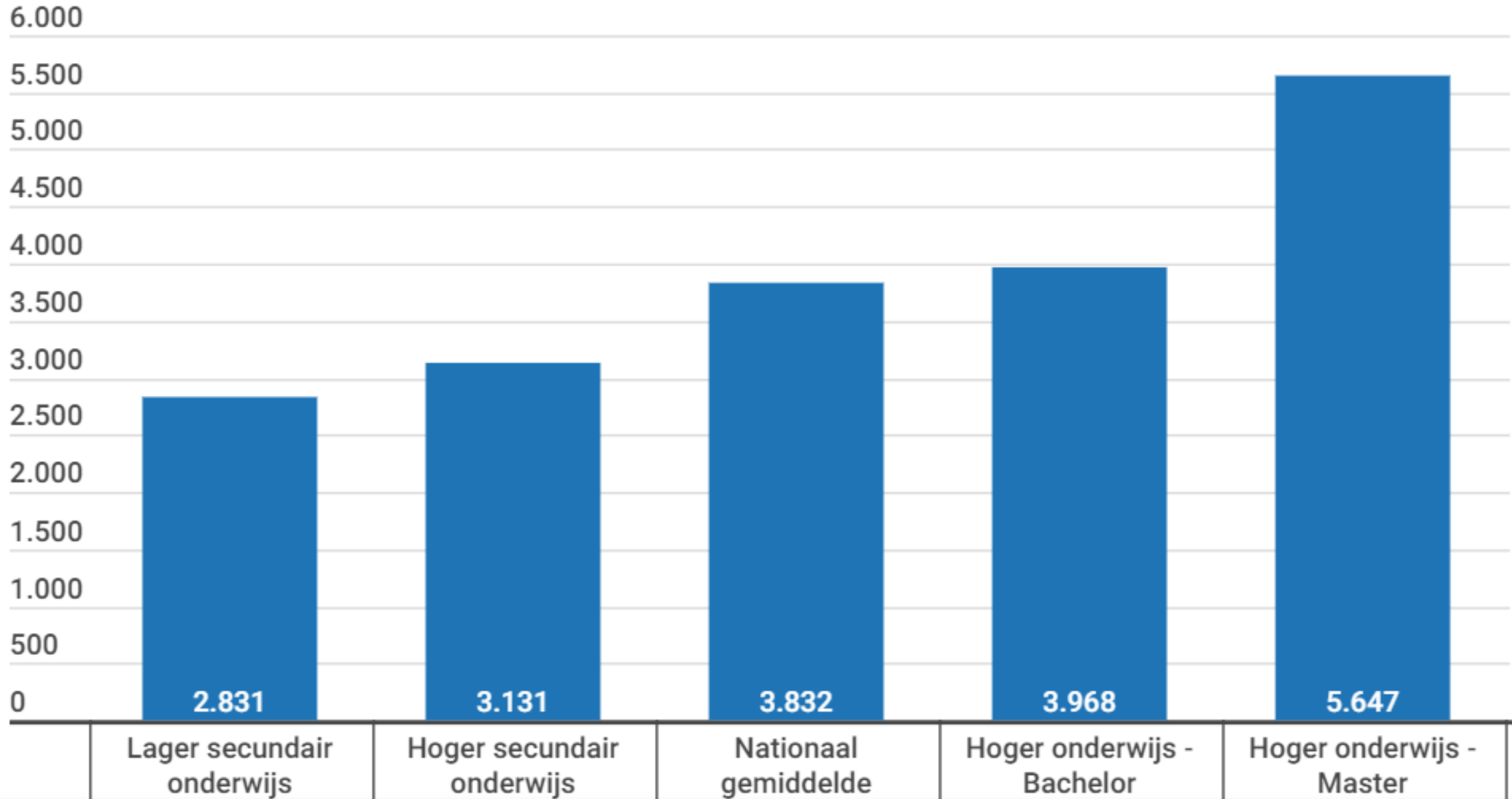
# HISTOGRAM EN STAAFDIAGRAM



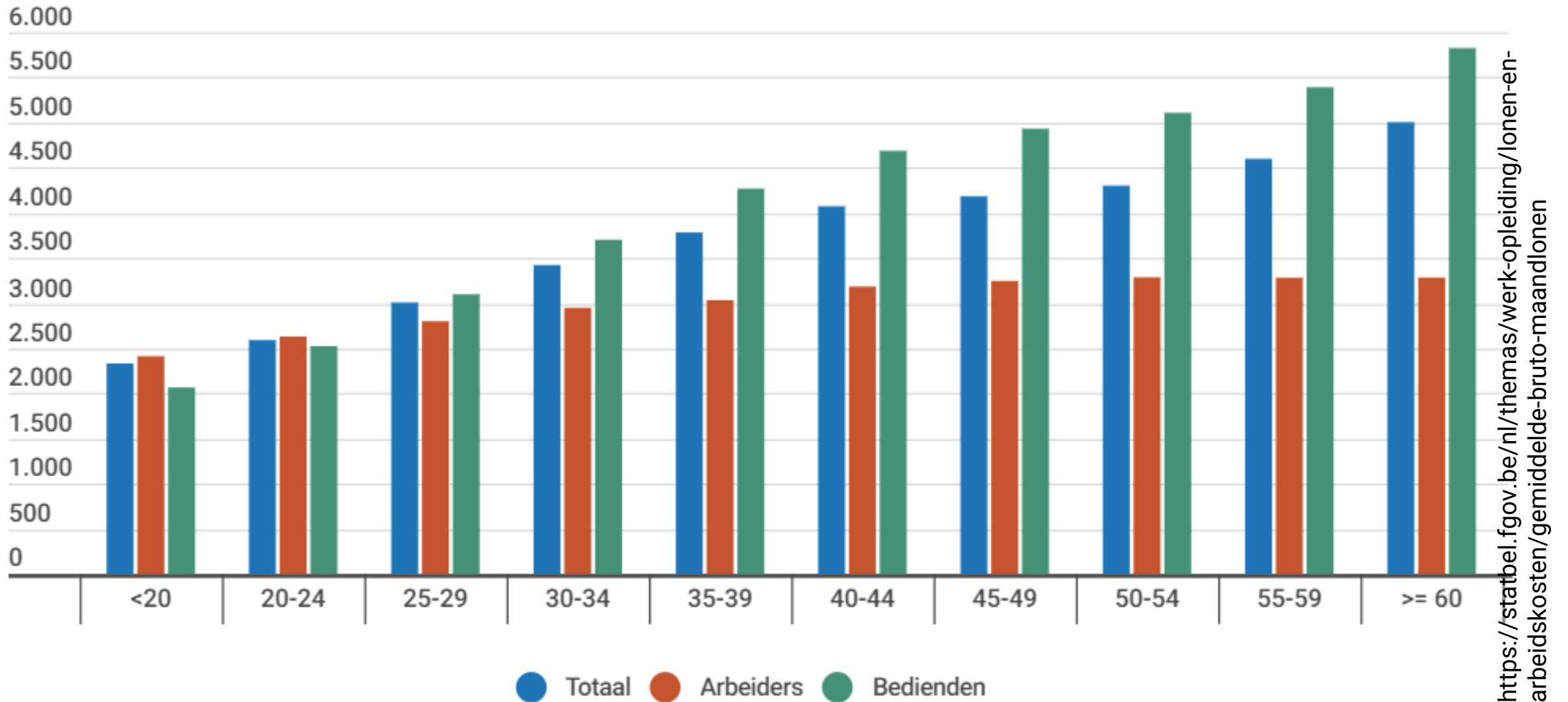
# Staafdiagram of histogram?



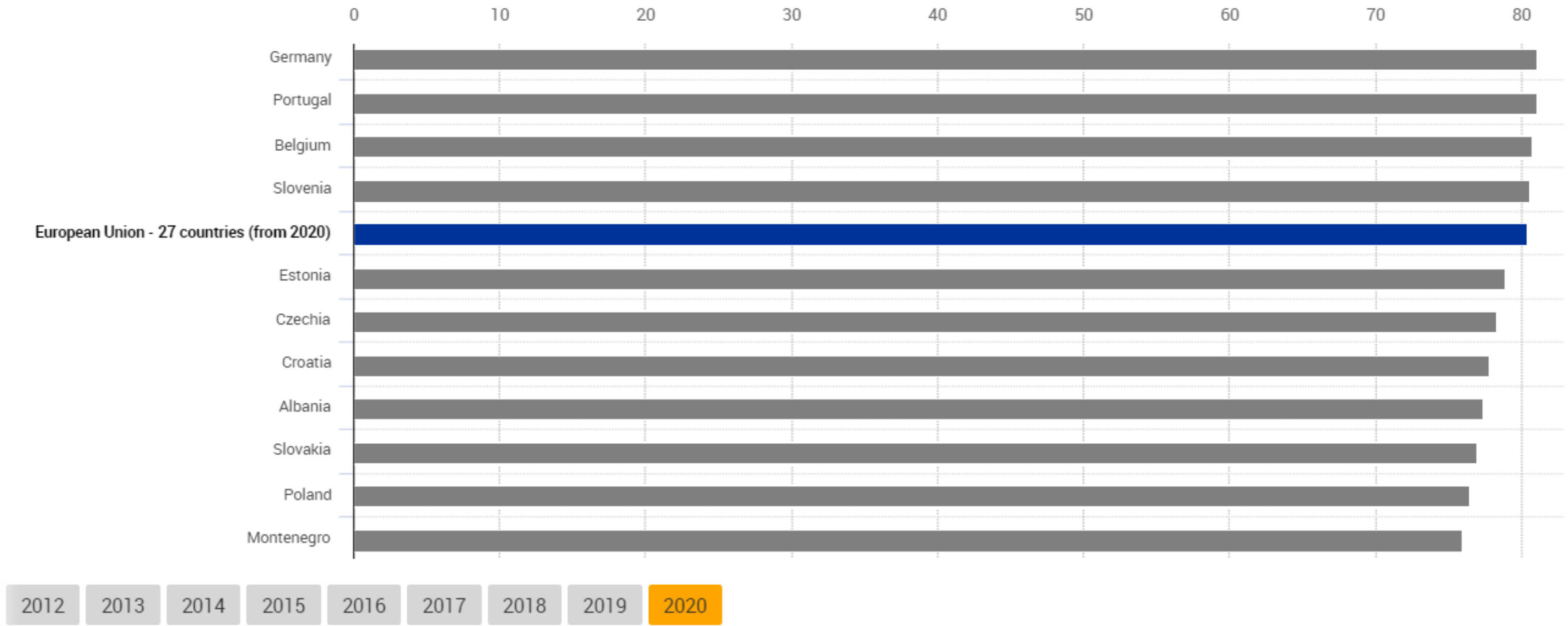
# Staafdiagram of histogram?



# Staafdiagram of histogram?



# Staafdiagram of histogram?



# Waar vind je dit materiaal?

[dwengo.org](http://dwengo.org)

[dwengo.org/wiskunde](http://dwengo.org/wiskunde)

[blockly.dwengo.org](http://blockly.dwengo.org)

[dwengo.org/agenda](http://dwengo.org/agenda)

[istem.be](http://istem.be)

 [facebook.com/dwengo](https://facebook.com/dwengo)

 [@dwengo](https://twitter.com/dwengo)

 [@NachaGesquiere](https://twitter.com/NachaGesquiere)





- Ontwikkeling projecten en bijbehorende lesmaterialen
- STEM, computationeel denken, physical computing
- Leerpaden op dwengo.org
- Software & hardware
- Extra lesmaterialen (bv. notebooks, technische fiches, handleidingen)
- Didactiek
- Evaluatiekader

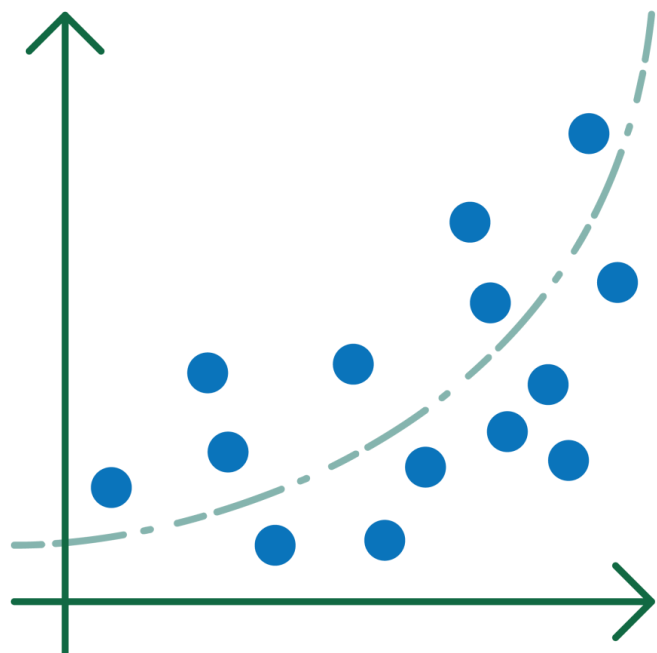


- Ondersteuning STEM-teams in scholen (Dwengo – UGent – iSTEM)
- Communicatie over projecten via istem.be



- Computationeel denken concreet in de klas





# Data visualiseren

Netwerkdag datageletterdheid

30 augustus 2023