

Data Foundations

Hoofdstuk 2:

Data visualisatie

Hassan Haddouchi



Overzicht

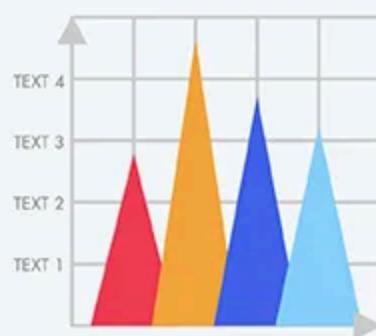
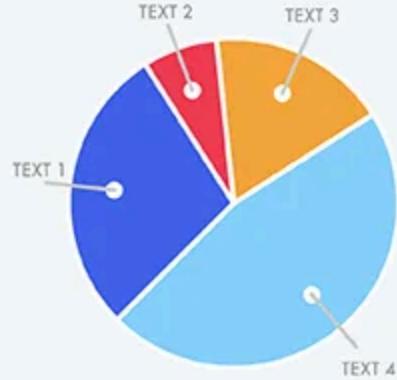
- Waarom data visualiseren?
- Belangrijke vragen
- Business Intelligence
- Soorten visualisaties
- Grammar of Graphics
- Tips voor visualisaties
- Labo
- Quiz



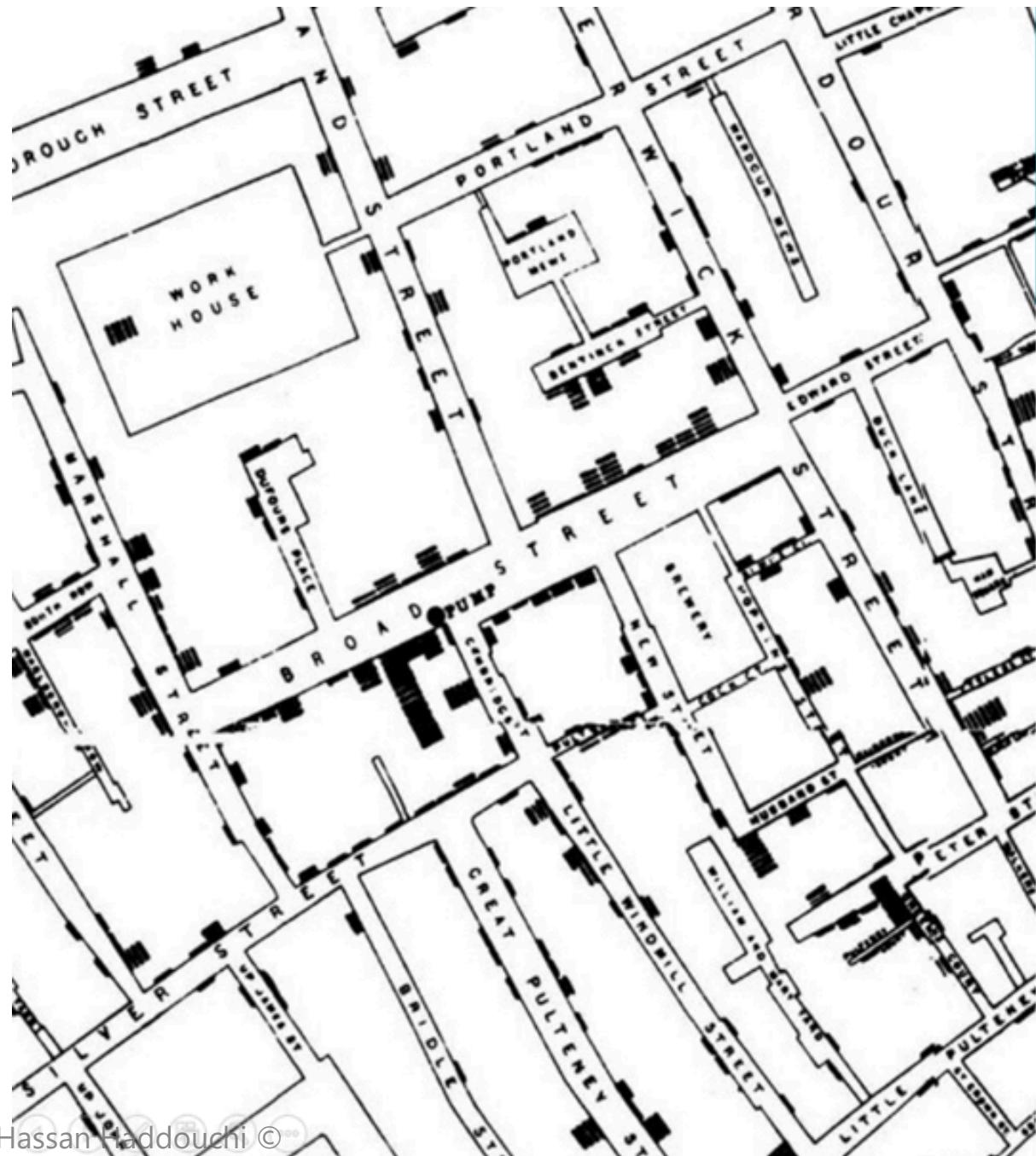


“Data visualization is a general term that describes any effort to help people understand data by placing it in a visual context.”

Voorbeelden



Waarvoor hebben we data visualisaties nodig?



Visualizing the Broad Street Cholera Outbreak

- A severe outbreak of cholera occurred in 1854 near Broad Street in the City of Westminster, London, England, unknowing to people causing over 600 deaths
- Physician Jon Snow identified the source of the outbreak as the public water pump on Broad Street
- Snow used a dot map to illustrate the cluster of cholera cases around the pump
- He also used statistics to illustrate the connection between the quality of the water source and cholera cases.

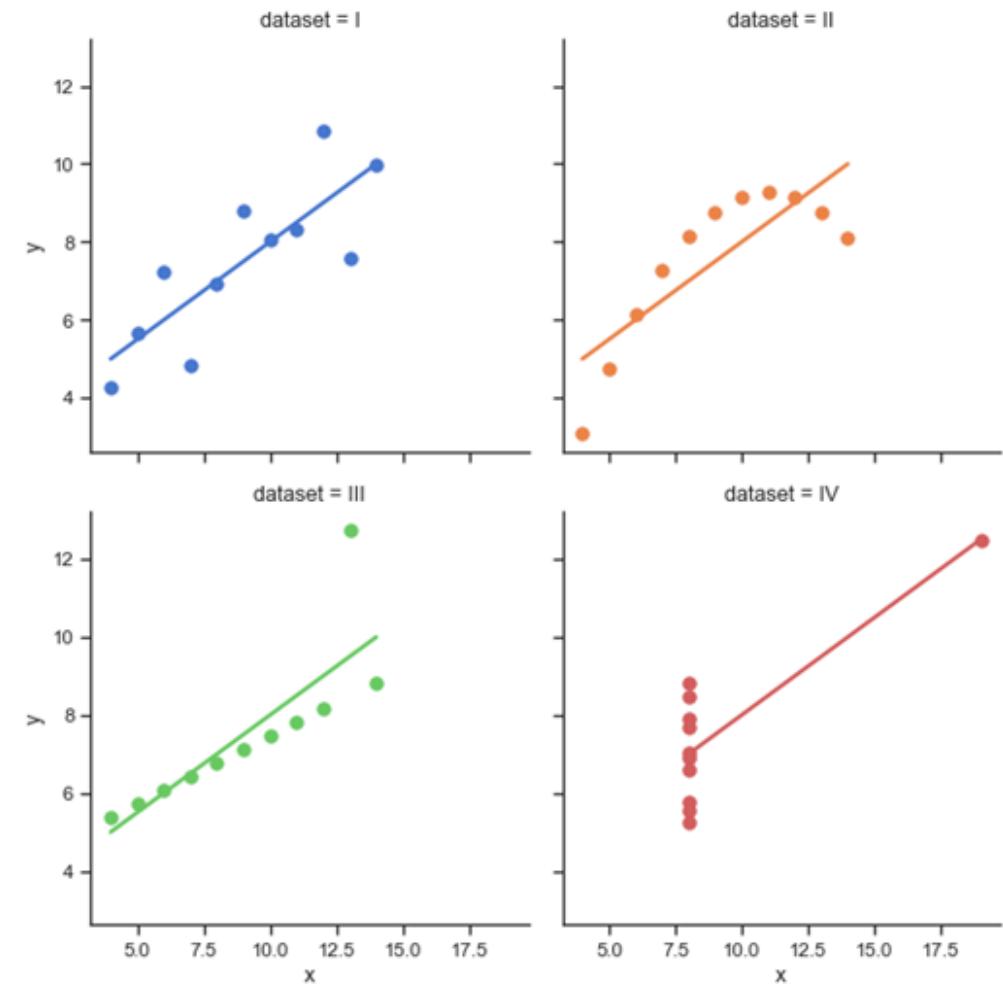
Statistieken geven maar een beperkt beeld.

Het is moeilijk om data in een tabel te interpreteren.

Anscombe's quartet: 4 datasets metzelfde statistieken

I		II		III		IV		
x	y	x	y	x	y	x	y	
10	8,04	10	9,14	10	7,46	8	6,58	
8	6,95	8	8,14	8	6,77	8	5,76	
13	7,58	13	8,74	13	12,74	8	7,71	
9	8,81	9	8,77	9	7,11	8	8,84	
11	8,33	11	9,26	11	7,81	8	8,47	
14	9,96	14	8,1	14	8,84	8	7,04	
6	7,24	6	6,13	6	6,08	8	5,25	
4	4,26	4	3,1	4	5,39	19	12,5	
12	10,84	12	9,13	12	8,15	8	5,56	
7	4,82	7	7,26	7	6,42	8	7,91	
5	5,68	5	4,74	5	5,73	8	6,89	
SUM	99,00	82,51	99,00	82,51	99,00	82,50	99,00	82,51
AVG	9,00	7,50	9,00	7,50	9,00	7,50	9,00	7,50
STDEV	3,32	2,03	3,32	2,03	3,32	2,03	3,32	2,03

Toont het belang aan van
het grafisch weergeven van
data tijdens de analyse,
alsook het effect van
uitschieters.



Belangrijke vragen bij data visualisatie.

1. Wie is je publiek?

Technische mensen, business vertegenwoordigers, managers, ...?

Welke voorkennis hebben ze?

AUDIENCE



2. Wat is je boodschap?

Wat wil je vertellen?

Welke informatie wil je overbrengen?

Welke data wil je presenteren?

3. Welke impact wil je bereiken?

Wil je mensen tot actie overhalen?

Wil je het publiek iets leren?

Wil je ze overtuigen?

Business Intelligence (BI)

Analyseren van data om nuttige inzichten te bekomen die business kunnen ondersteunen.

BUSINESS INTELLIGENCE



Verschillende types rapporten:

Statisch: populair voor presentaties en meetings (bv. PDF of slides)

Interactief (self-service BI): gebruikers kunnen de data filteren, grouperen, sorteren (laat toe om eigen verhaal te maken).

Dashboards (kan interactief zijn): geven een overzicht van belangrijke business data.

Business Intelligence tools

Microsoft Power BI

Kostenefficiënt

Data Analytical Expressions (DAX)

leren kan moeilijk zijn

Mogelijke integratie met Python en R

De tool om mee te starten



Tableau

Meest simpele tool

Vooral gericht op visualisaties

Prijszetting is redelijk complex, geen
gratis versie



QlikView

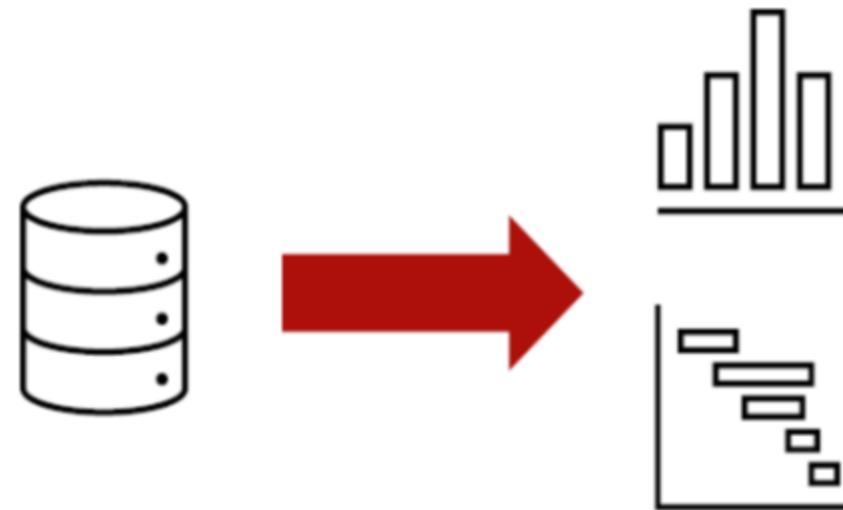
Vergt kennis van scripting

Beperkte integratie met externe tools



Power BI

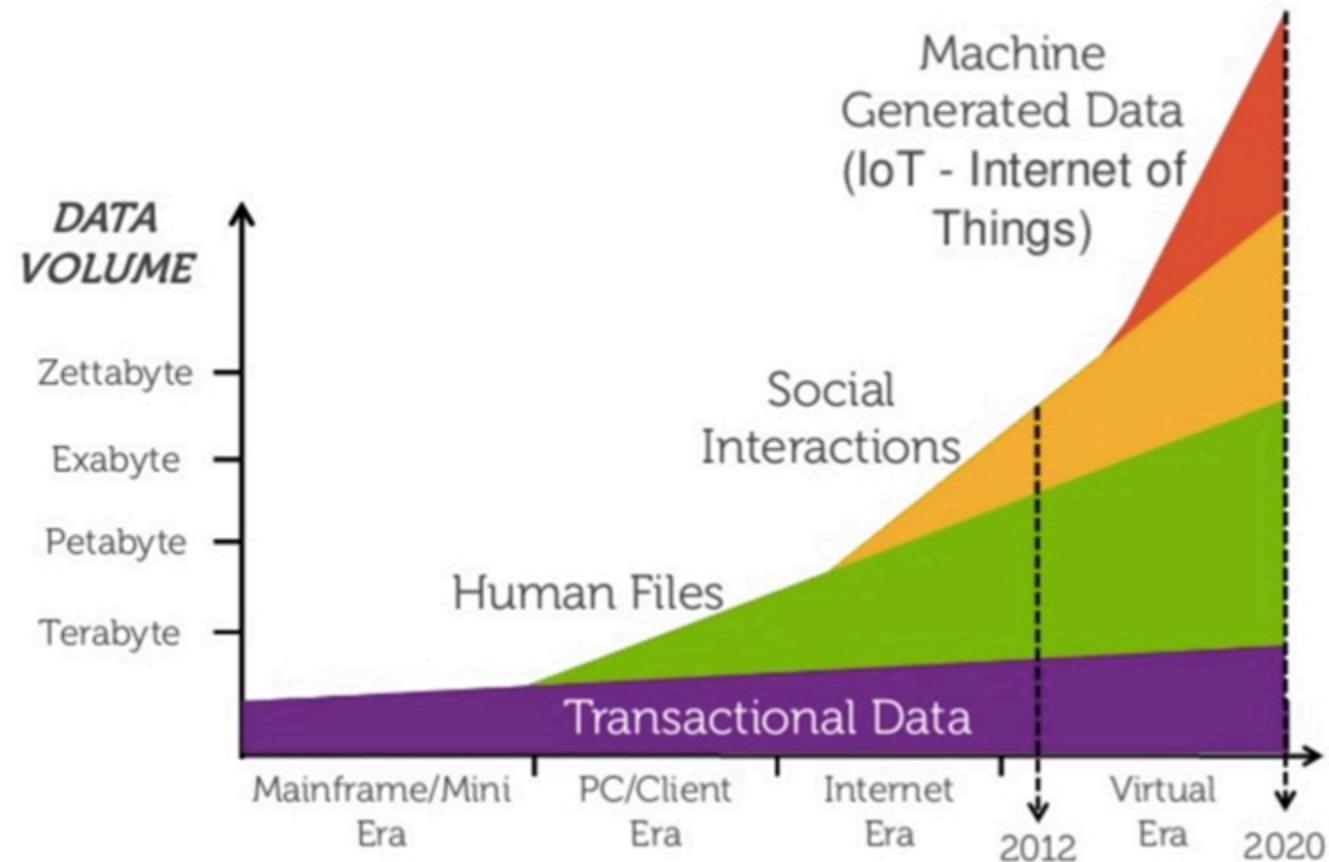
Een krachtige tool die verschillende databronnen omzet in visueel aantrekkelijke inzichten en deze gemakkelijk deelt.



Data explosie

Meer en meer data komen beschikbaar.

Meer nood aan inzichten, vaak d.m.v. visualisaties.



Soorten visualisaties

Welke visualisaties en diagrammen ken je?

Tabel of matrix

Data wordt gepresenteerd in rijen en kolommen.

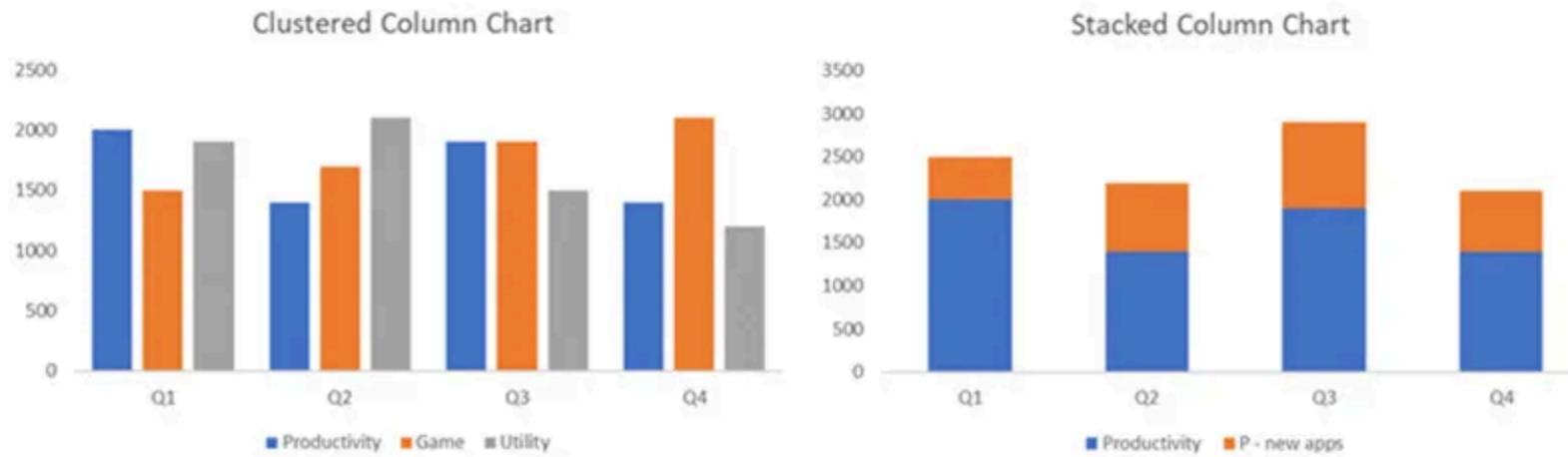
Maximum 2 dimensies (+ totalen).

Country	Sales Amount
United States	\$7,390,464
France	\$773,445
Germany	\$493,628
Total	\$8,550,077

Staafdiagram

Hoogte balken geven hoeveelheid aan.

Stacked: groepen worden boven elkaar gezet.



Lijn- en oppervlaktediagram

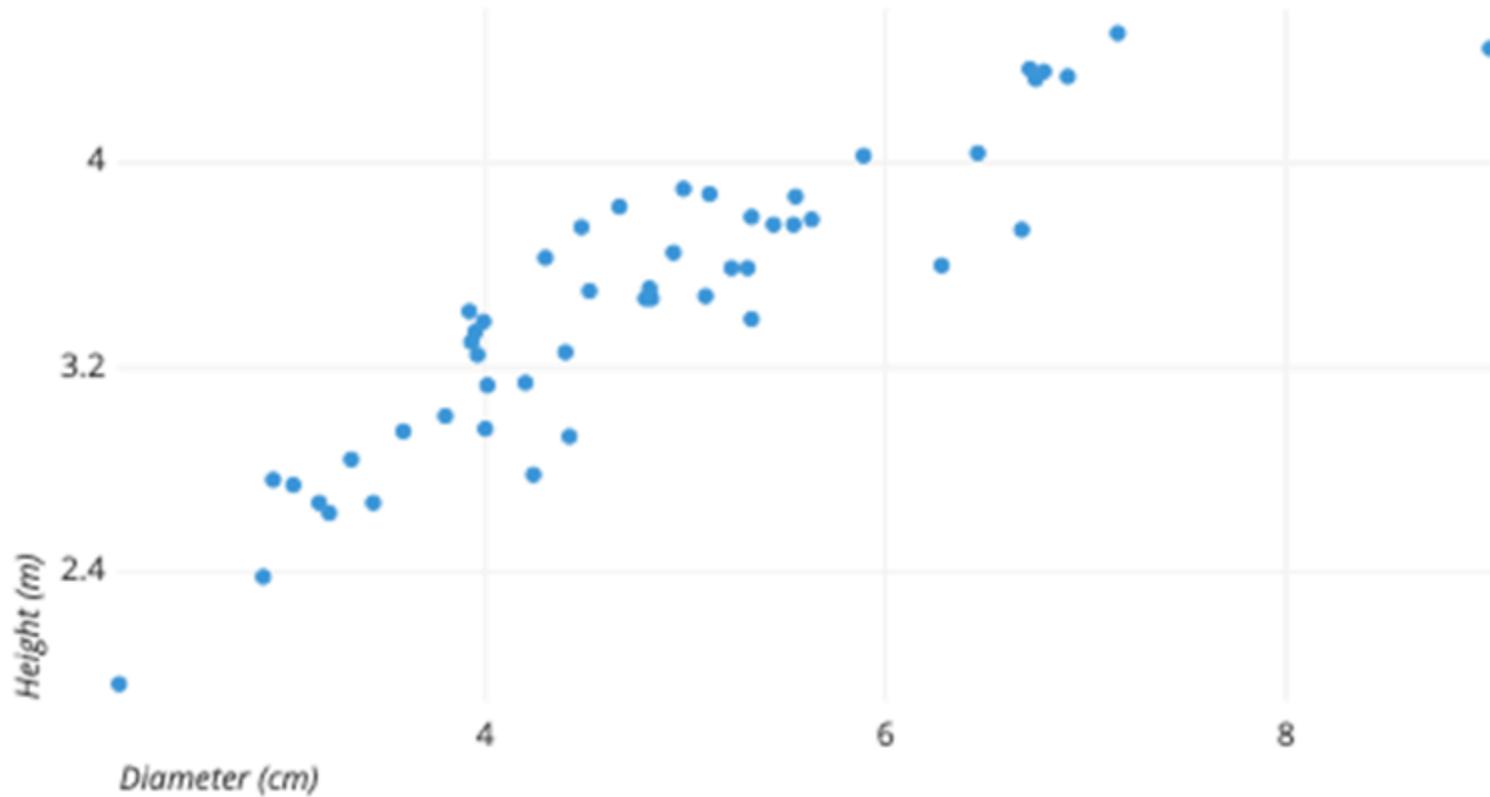
Geven trends in de tijd weer.

Oppervlaktediagram: deel onder de lijn is ingekleurd.



Spreidingsdiagram (scatter plot)

Geeft relatie tussen twee datavelden weer.

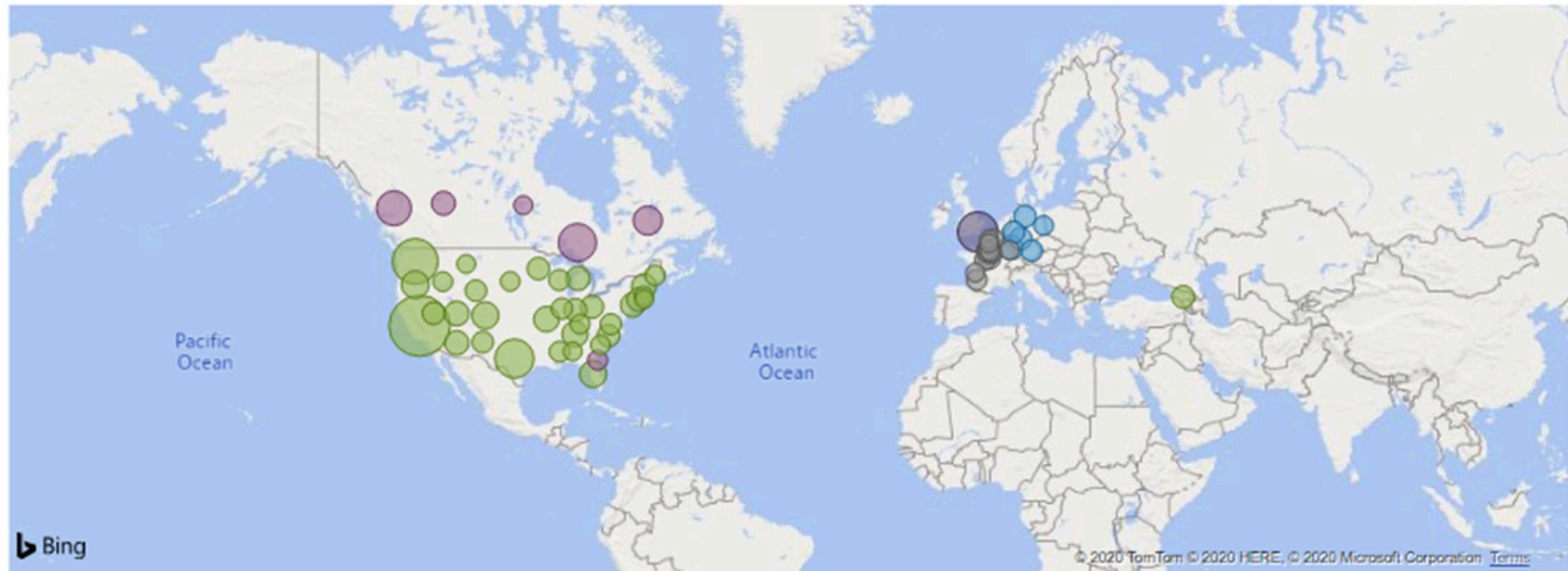


Kaarten

Toont geografische informatie.

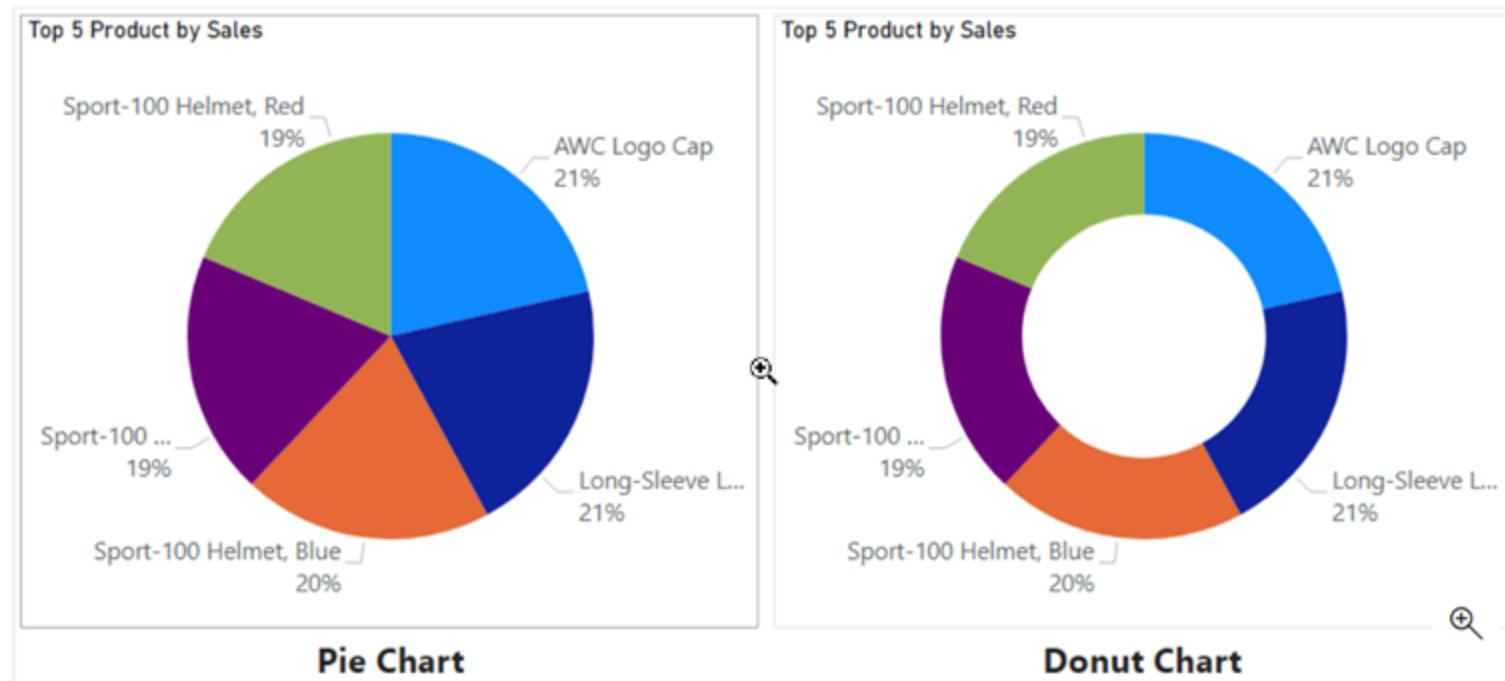
Sales by Country

Country ● Canada ● France ● Germany ● United Kingdom ● United States



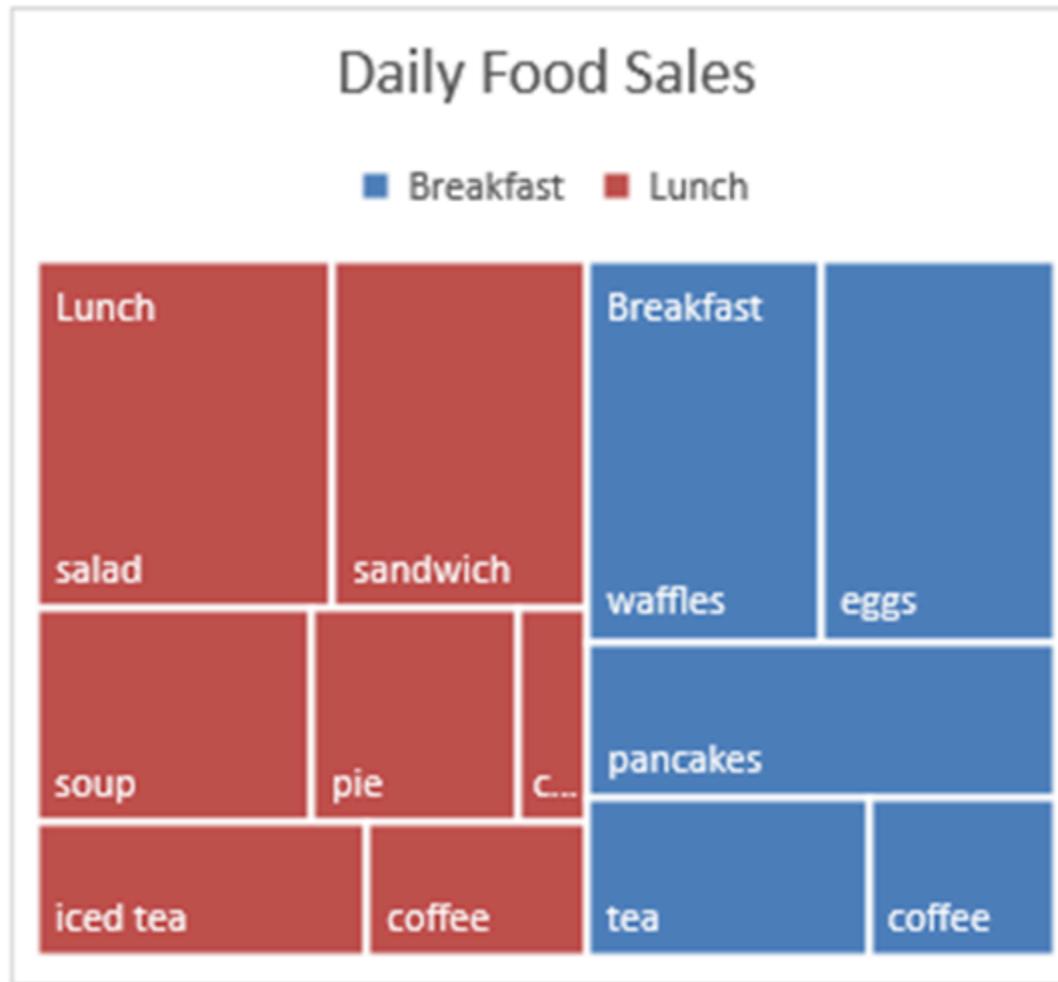
Taart- en donutdiagram

Geeft het aandeel weer door het totaal op te delen in stukken van een taart of donut.
Vaak moeilijk om te lezen (wordt vooral voor esthetische redenen gebruikt).



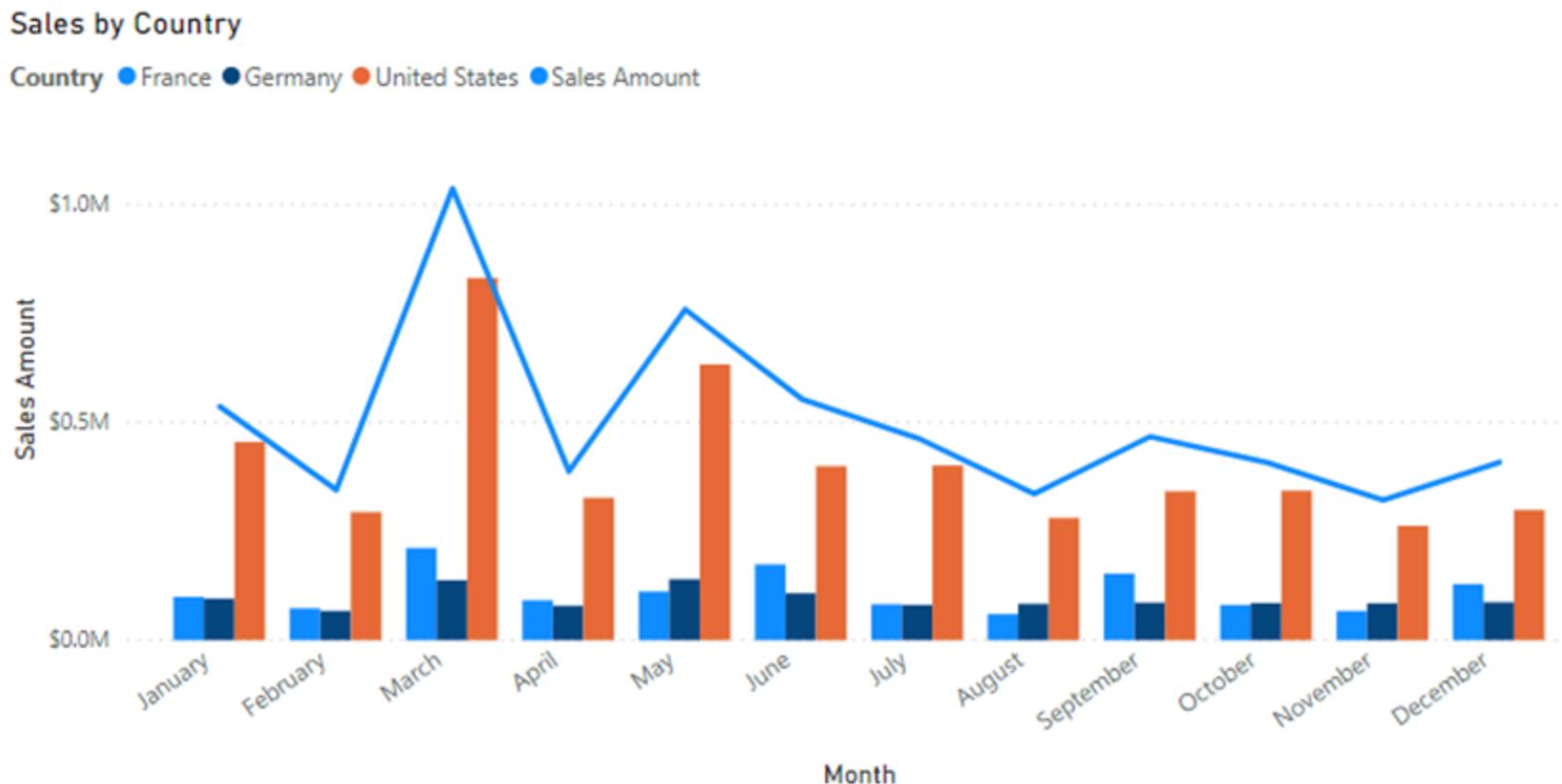
Tree map

Geeft hoeveelheden weer met de grootte van rechthoeken.
Vaak moeilijk te interpreteren.

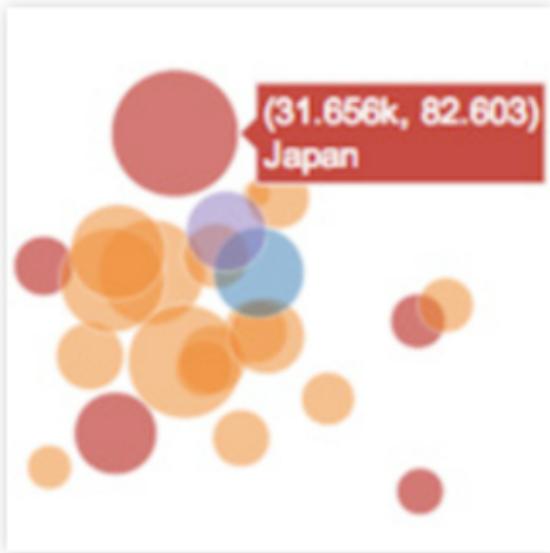


Combo diagram

Combinatie van een staafdiagram en een lijndiagram (2 y-assen mogelijk).
Illustrert correlaties en spaart plaats uit.



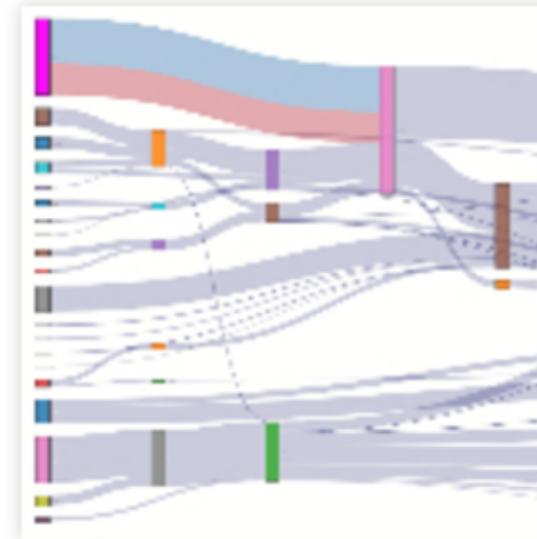
En nog veel andere mogelijkheden



Bubble Charts



Sunburst Charts



Sankey Diagram

95% can be
done with
these chart
types



Relational: scatter plot



Temporal: line chart



Spatial: heat map



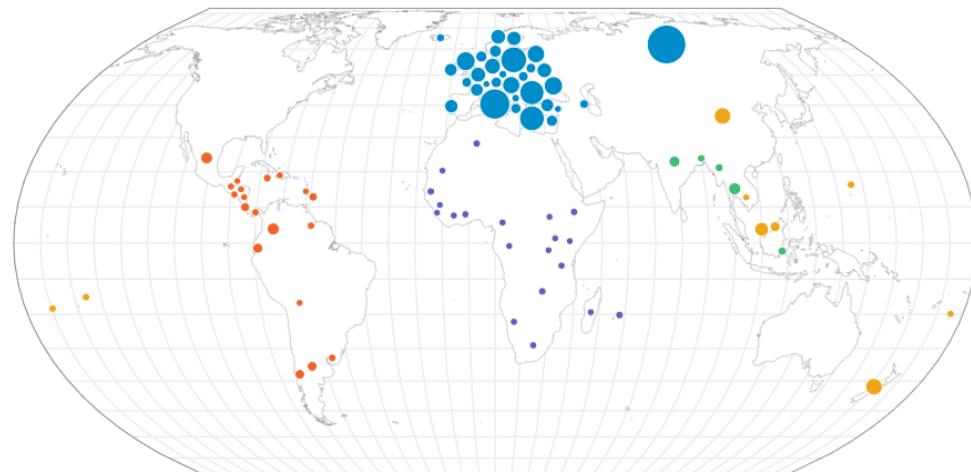
Categorical: bar chart

Interactieve visualisatie

Laat de gebruiker toe om de data te verkennen

Number of COVID-19 cases reported to WHO

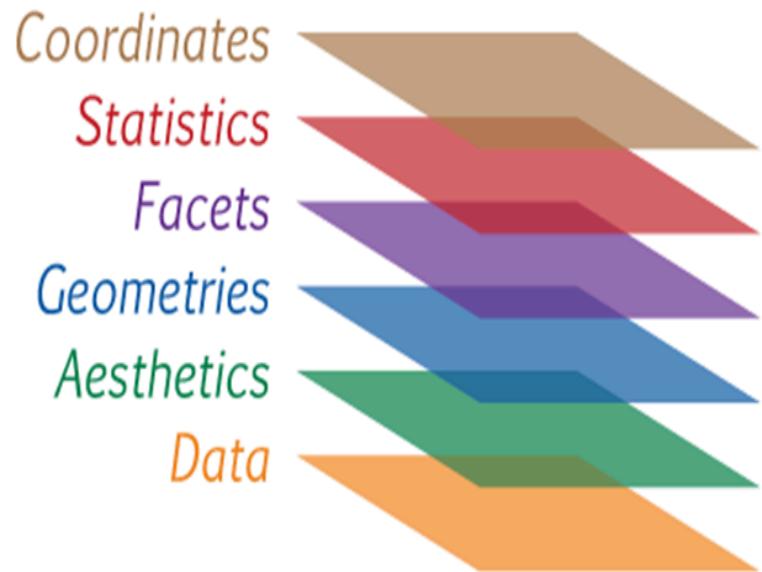
World, 28 days to 1 September 2024



<https://covid19.who.int/>

Grammar of Graphics

De Grammar of Graphics geeft structuur aan visualisaties. Het is vergelijkbaar met grammatica en taal.



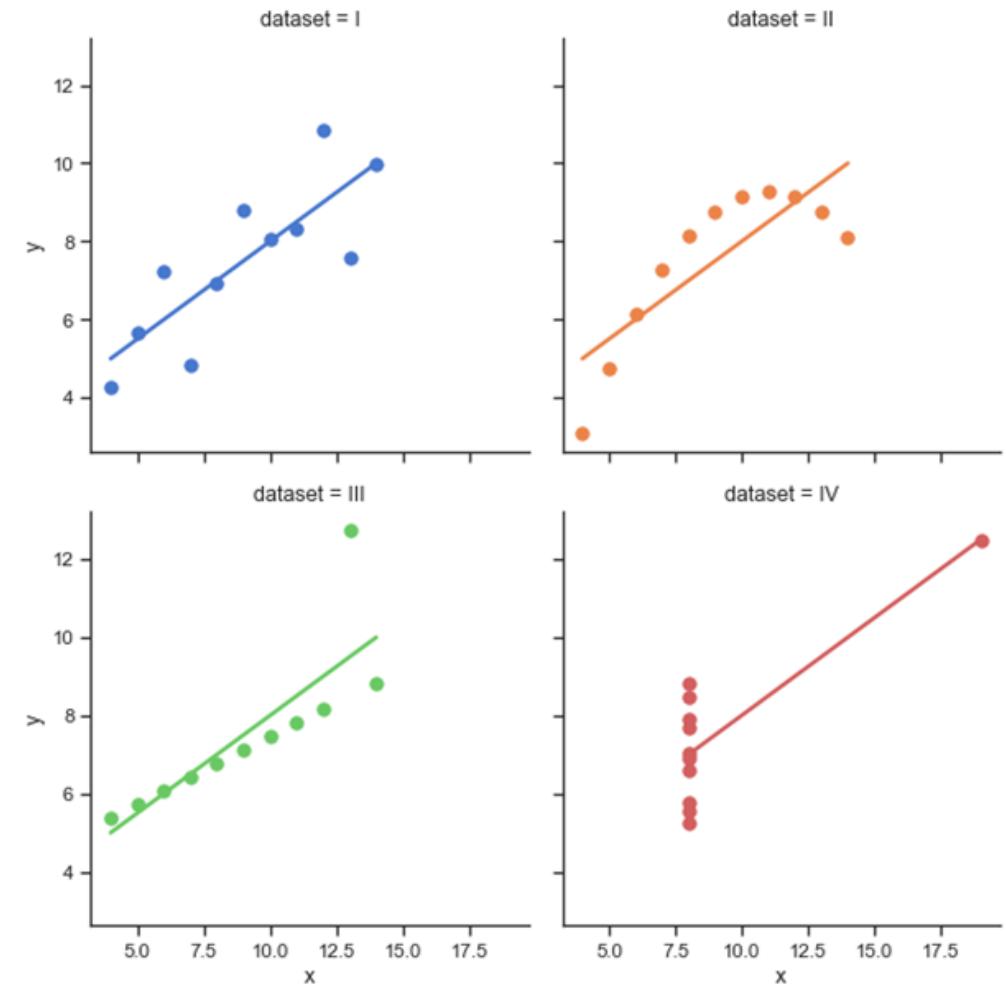
Grammar of Graphics bestaat uit 6 componenten

1. Data
2. Esthetiek: het uitzicht
3. Geometrie: de vormen
4. Facetten: combinatie diagrammen
5. Statistiek: bewerkingen op de data
6. Coördinaten: welke coördinaten gebruik je op de assen

Component 1: data

Welke data wordt getoond?

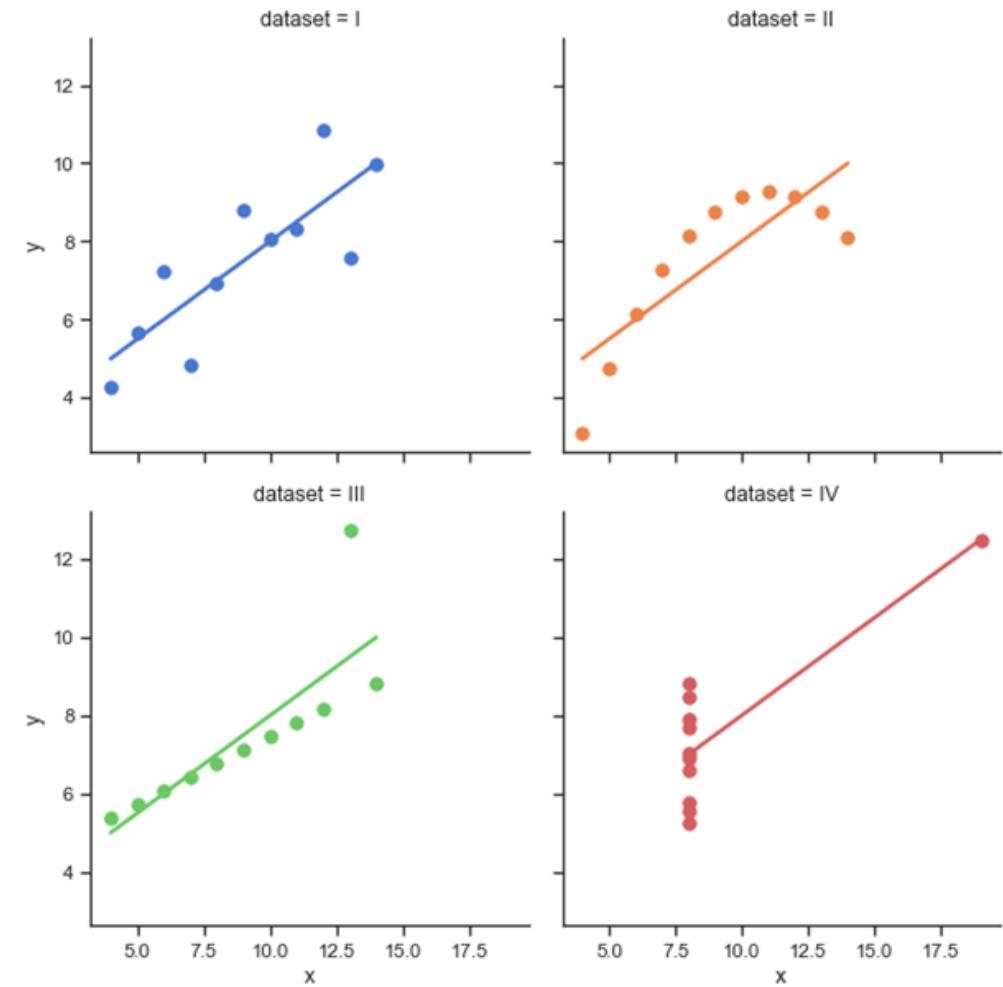
- 4 datasets die elk 13 punten bevatten.



Component 2: esthetiek

Welke visuele elementen worden gebruikt? (assen, relatieve posities, kleur, grootte, ...)

- 4 figuren in een rooster.
- De waarden worden weergegeven door de positie. De kleur wordt gebruikt om verschillende datasets te onderscheiden.

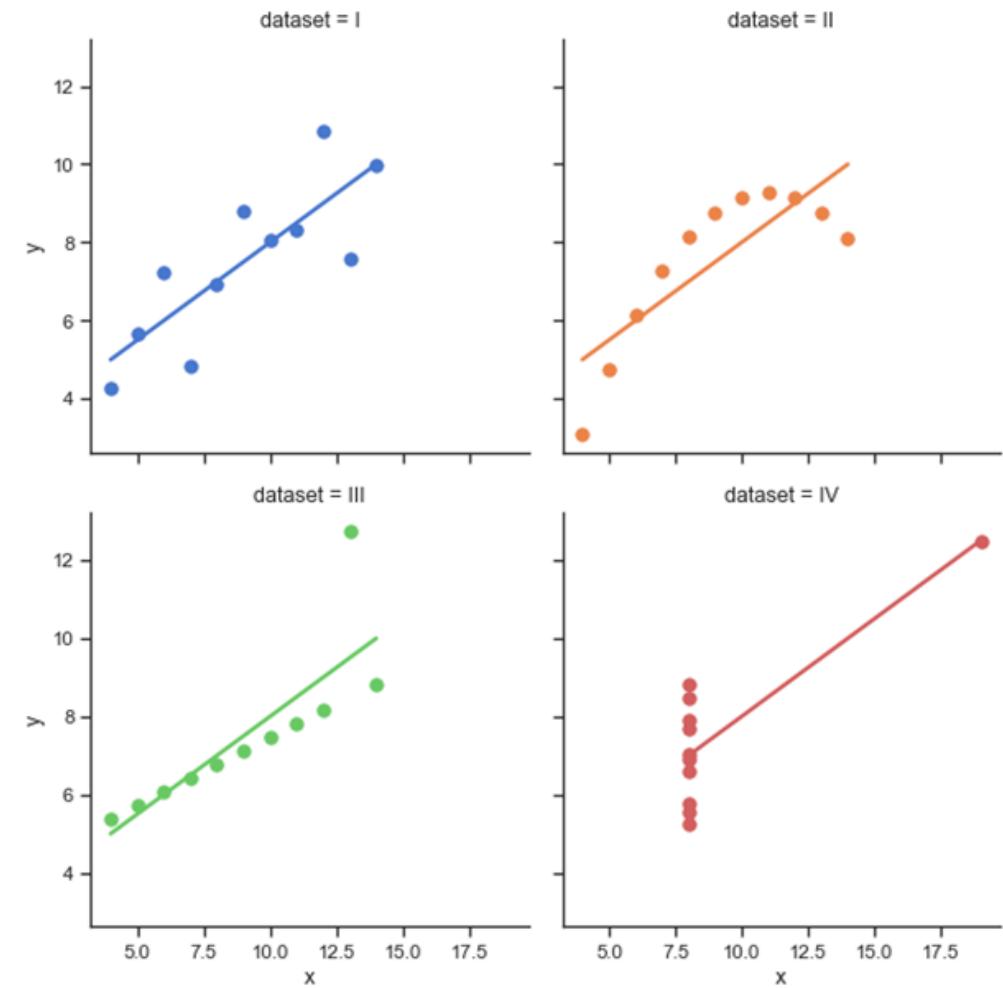


Component 3: geometrie

Welke vormen worden gebruikt?

- De data wordt weergegeven als dikke punten.

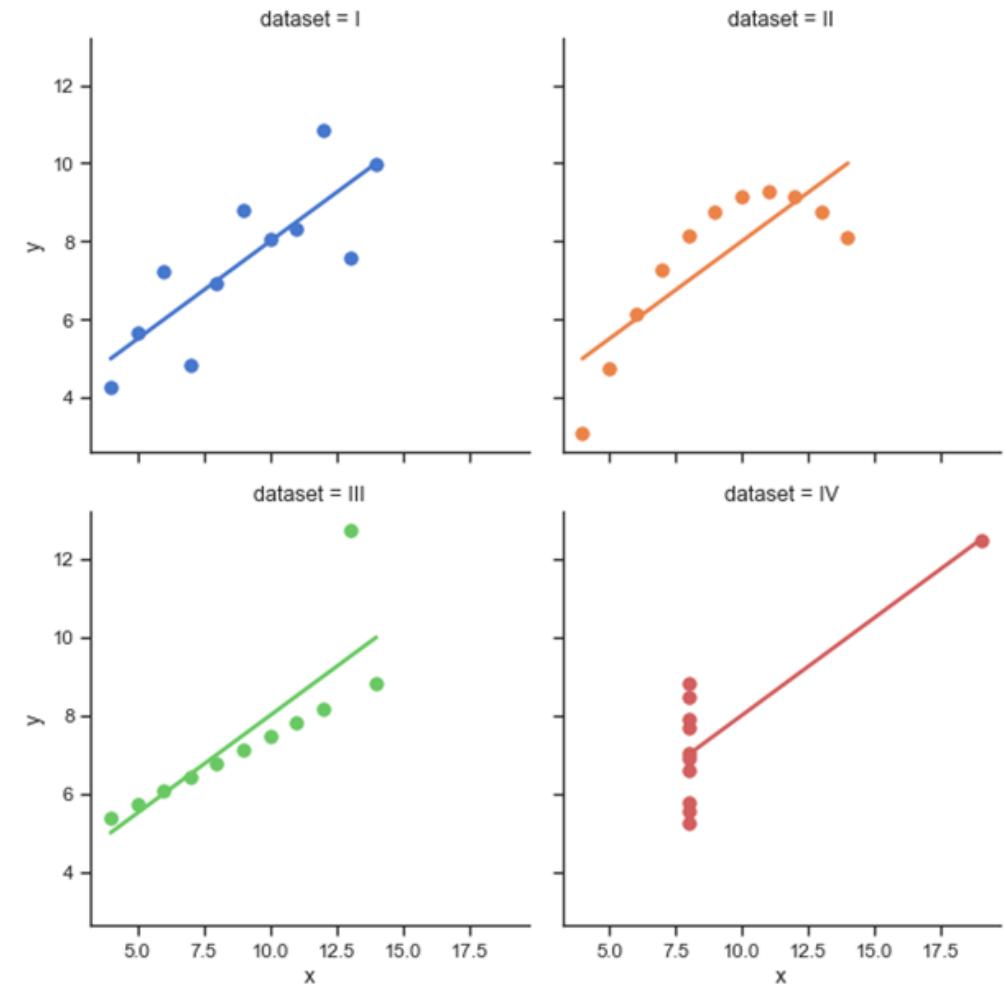
Andere mogelijkheden zijn lijnen, cirkels, vlakken, ...



Component 4: facetten

Wordt er één of meerdere diagrammen gebruikt?

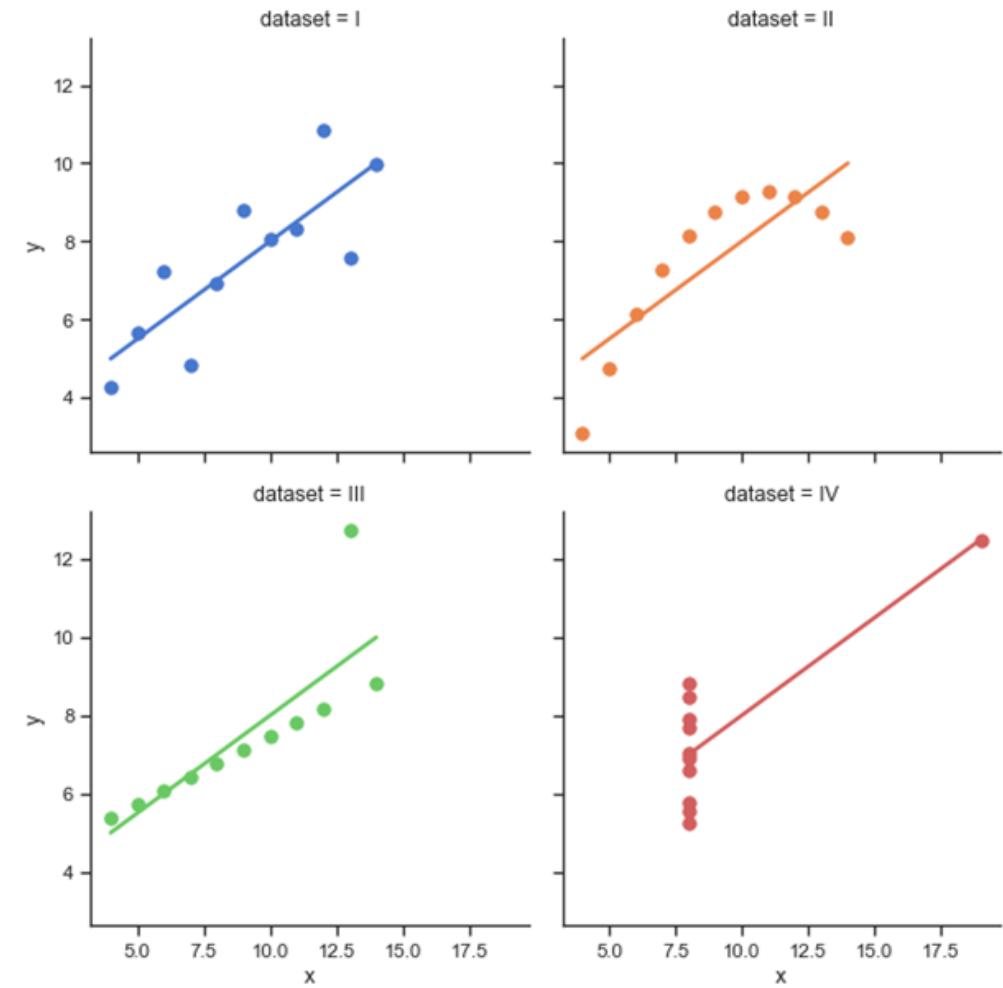
- Voor elke dataset is er een diagram.



Component 5: statistiek

Welke berekende informatie wordt er getoond?

- De onbewerkte data wordt getoond met een lijn die de trend weergeeft.

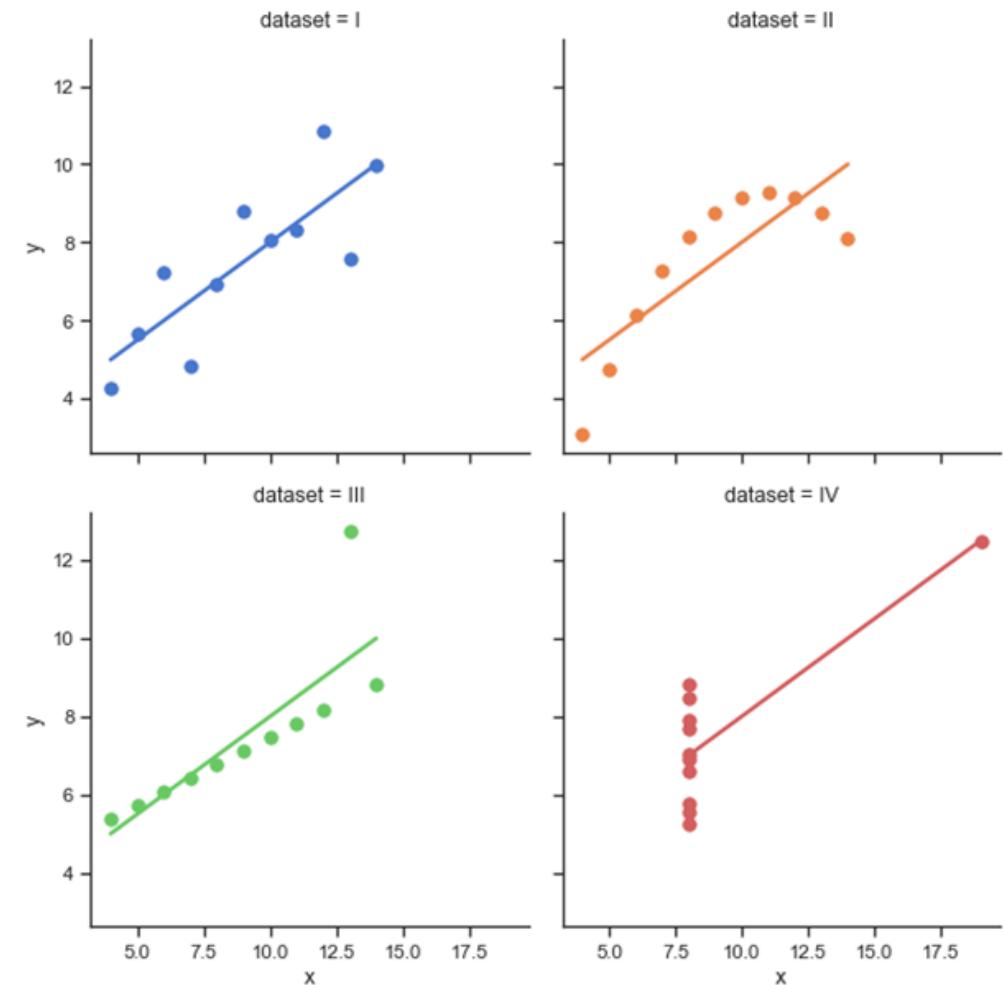


Component 6: coördinaten

Welke coördinaten worden gebruikt op de assen?

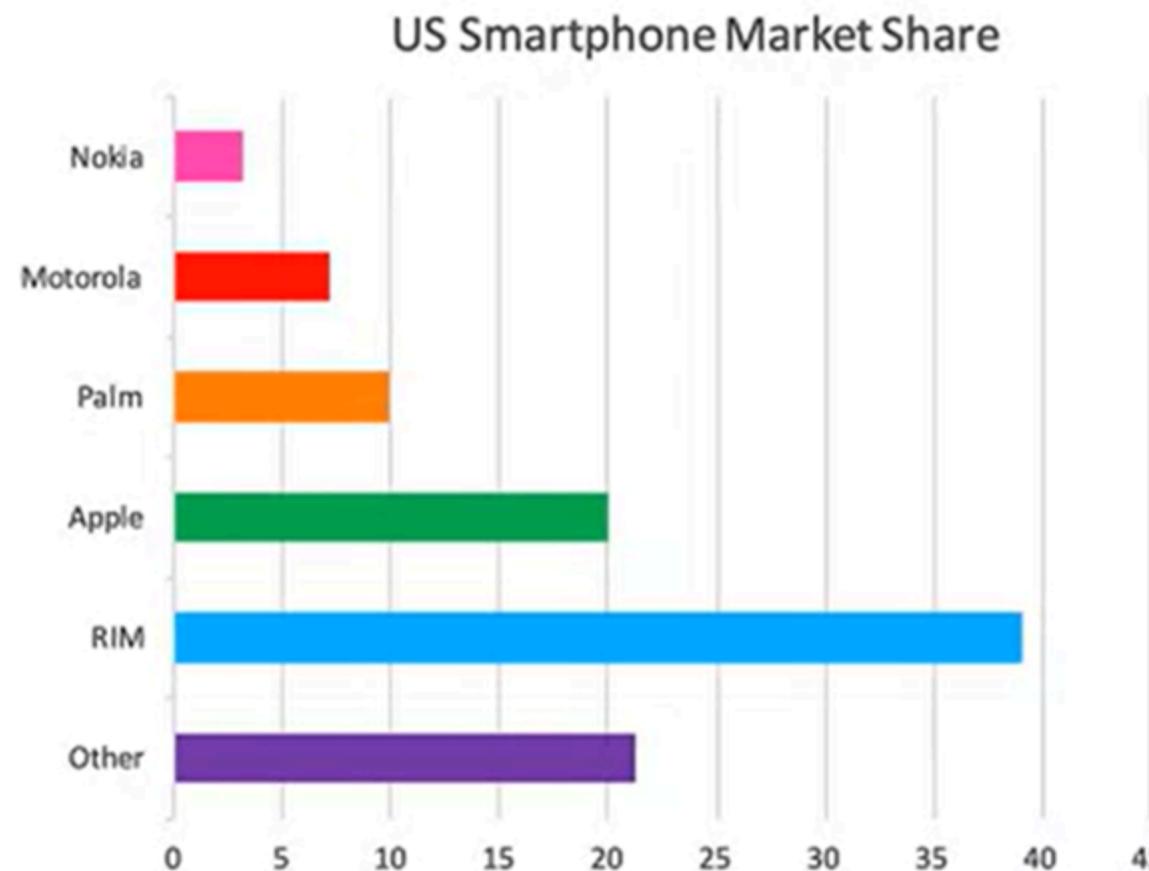
- De data wordt getoond met 2 normale assen.

Andere mogelijkheden zijn categorieën en of polaire coördinaten.



Oefening

Gebruik de Grammar of Graphics om onderstaande figuur te analyseren.

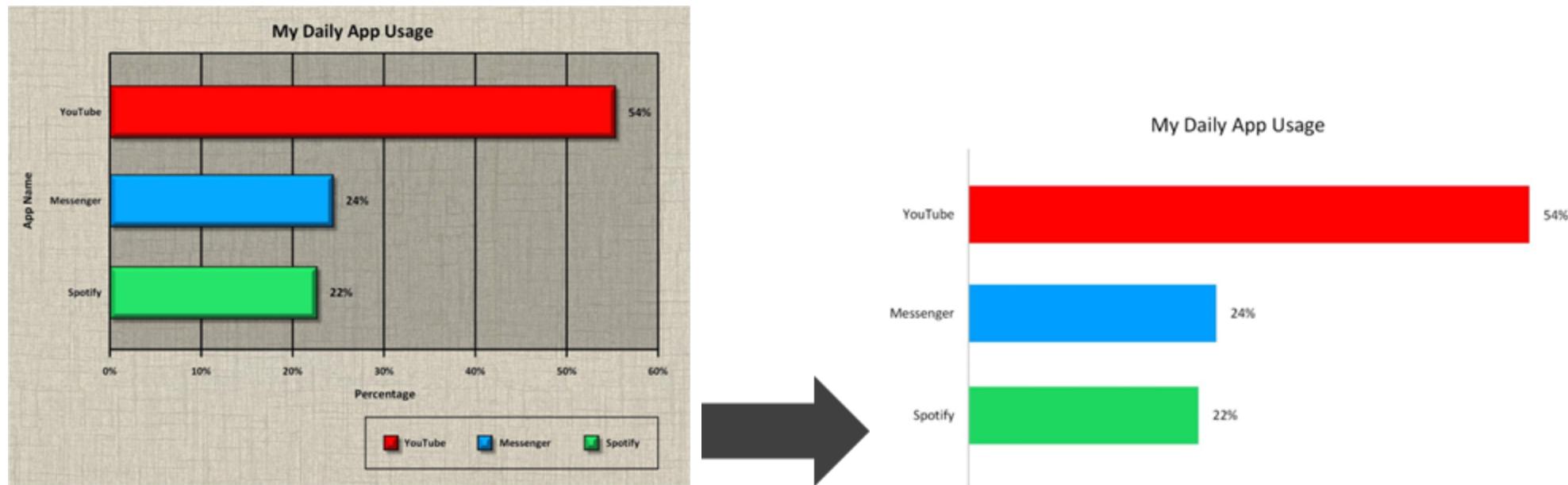


- Data
Marktaandeel smartphones in percentages
- Esthetiek:
Labels en titel weergegeven, kleur om de categorieën te onderscheiden.
- Geometrie
Staven
- Statistiek:
Ruwe data (?)
- Facetten:
Slechts 1 diagram
- Coördinaten
Categorische y-as, continue x-as

Tips voor data visualisatie

Tip 1: less is more

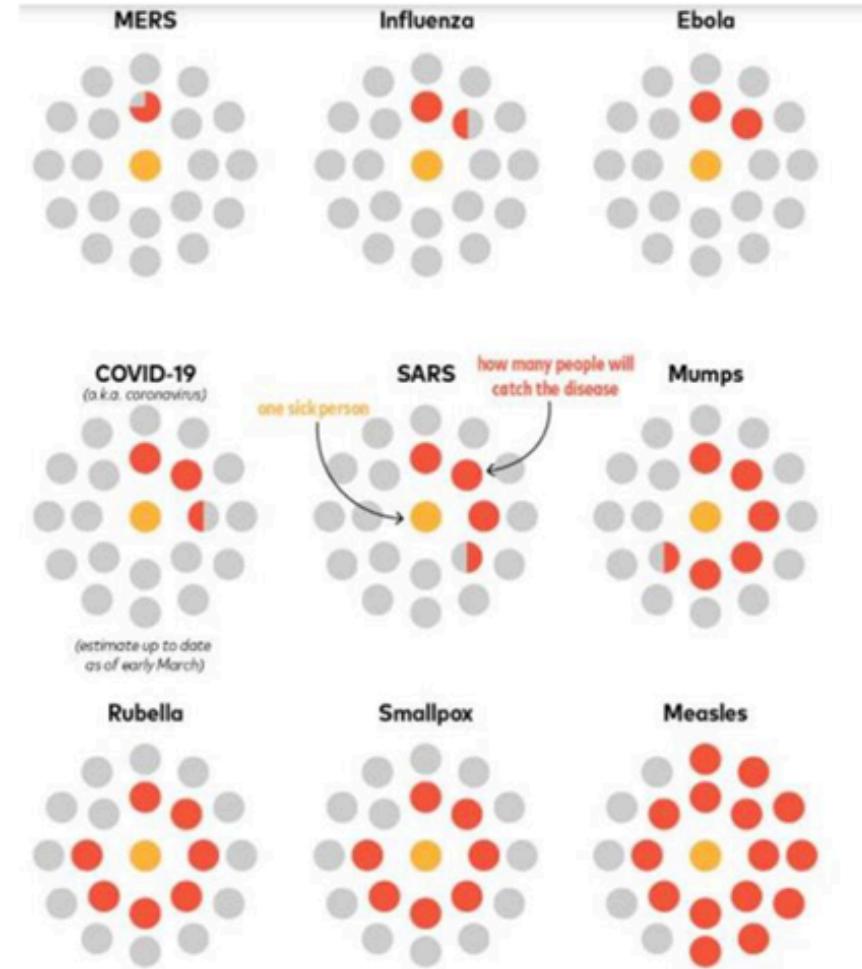
Minimaliseer de data-inkt verhouding.
Hou het zo simpel mogelijk!

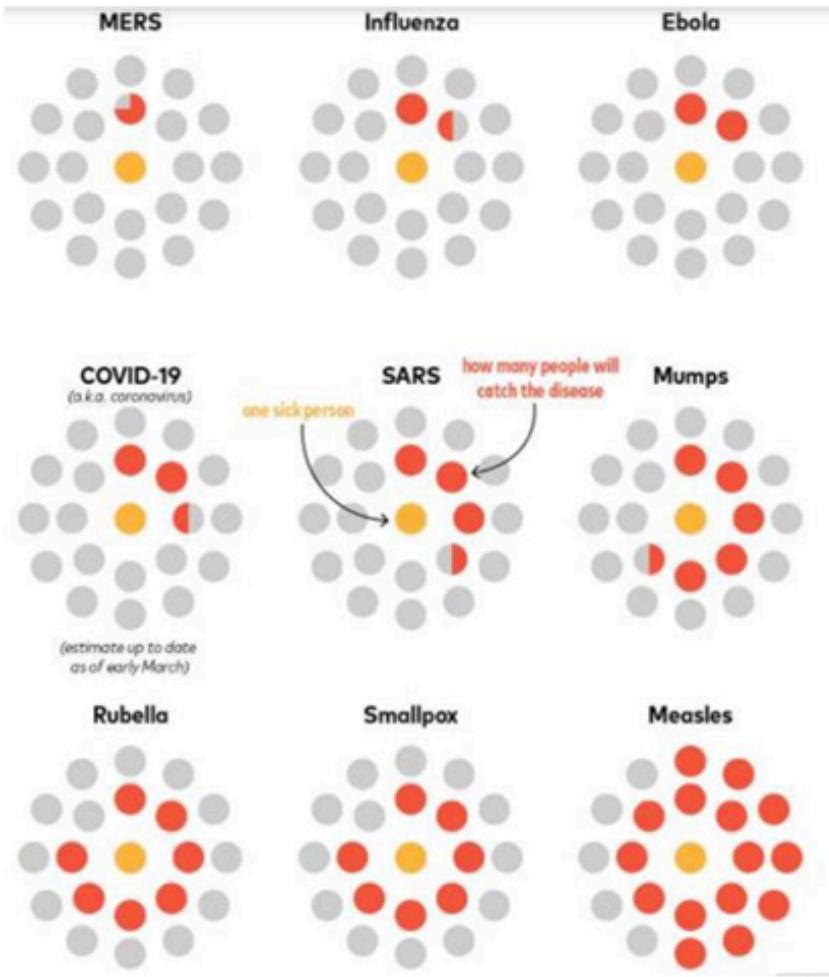


Tip 2: functioneel primeert op 'fancy'

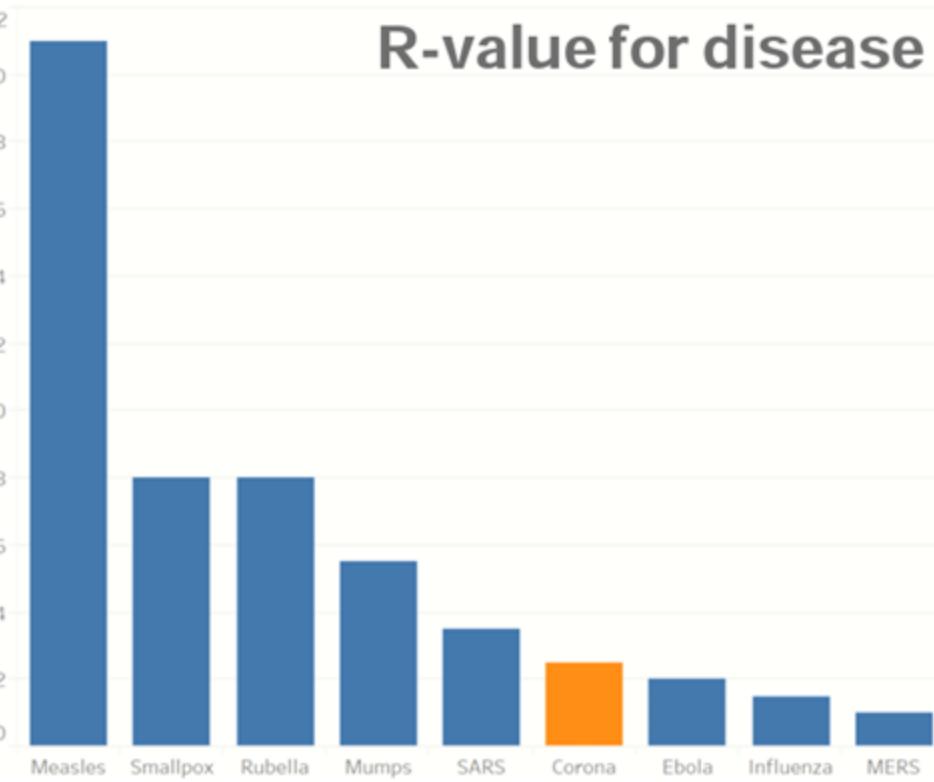
Er zijn veel visualisaties mogelijk. Hou het zakelijk!

Figuur hieronder: geen titel, moeilijk virussen te vergelijken.





R-value for disease



Tip 3: denk na over kleuren

1 op 12 mannen en 1 op 250 vrouwen is kleurenblind. Vermijd bepaalde combinaties.

Gebruik kleur om iets te benadrukken.

Gebruik niet teveel kleuren.

Kleuren kunnen gevoelens oproepen:

- Rood en oranje zijn alarmerend en vaak negatief.
- Blauw of groen is positief.



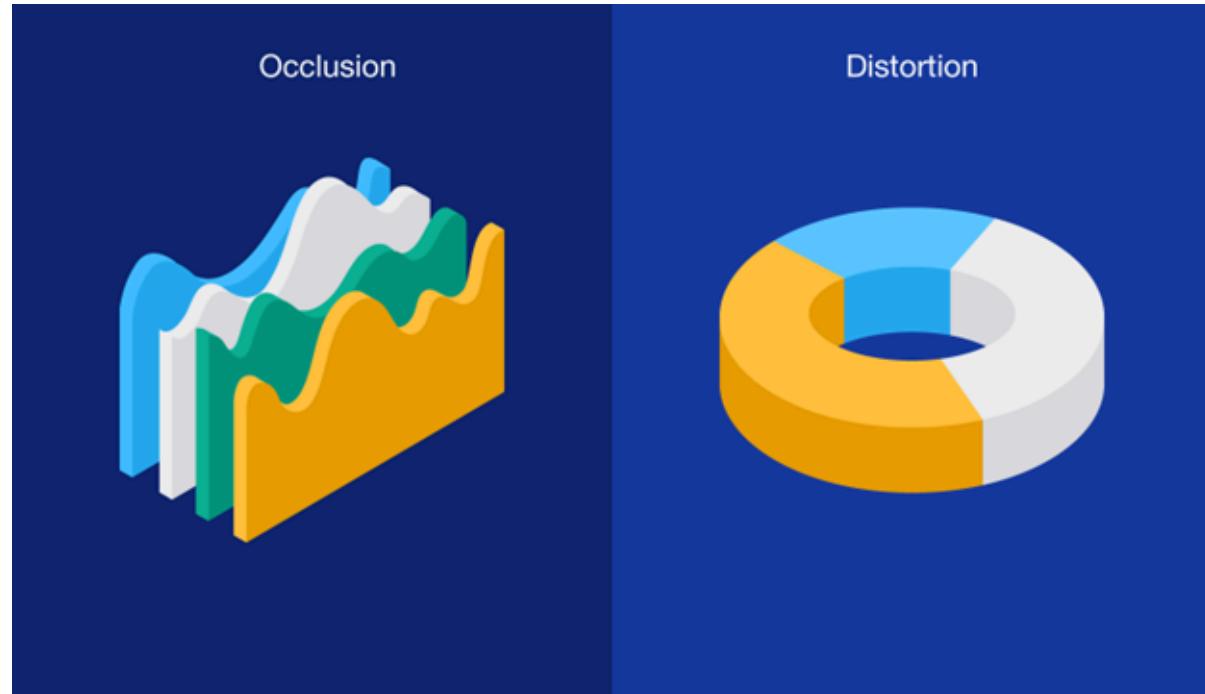
Tip 4: vermijd taartdiagrammen

Ons brein is veel beter in het inschatten van lengtes dan oppervlaktes. Gebruik een staafdiagram.

Tip 5: vermijd 3D grafieken

3D grafieken zijn vaak visueel aantrekkelijk maar moeilijk te lezen.
3D vertekent het perspectief en leidt tot:

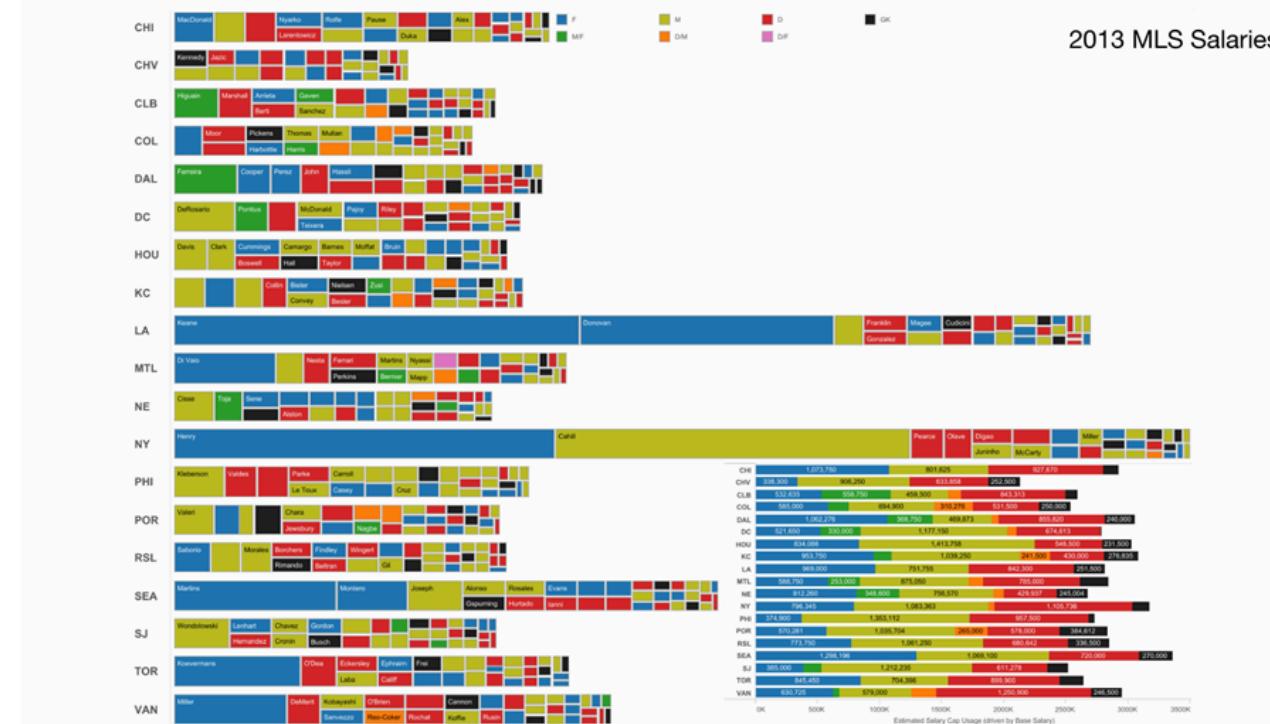
- Vervorming (distortion): delen op de voorgrond lijken groter dan delen op de achtergrond.
- Occlusie: stukken die deels verstopt zitten achter andere stukken lijken minder belangrijk.



Tip 6: informatie overload

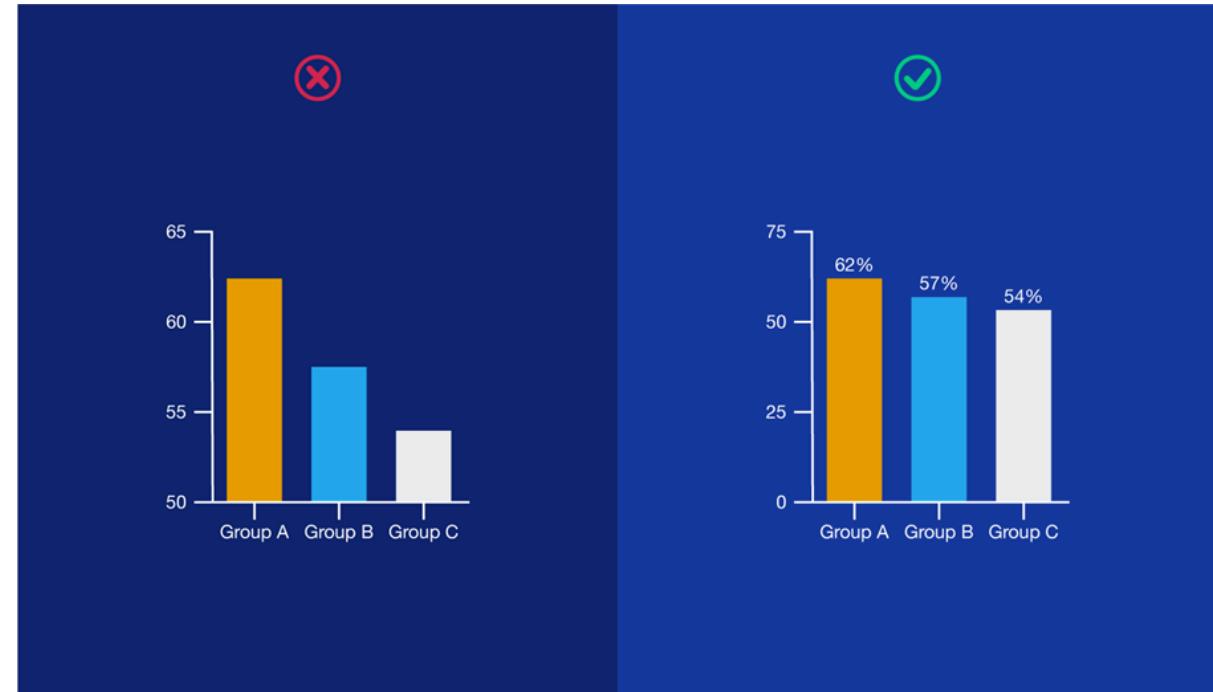
Probeer niet teveel informatie in 1 grafiek te zetten.

Het kan effectiever zijn om te werken met meerdere visualisaties.



Tip 7: toon steeds de volledige as

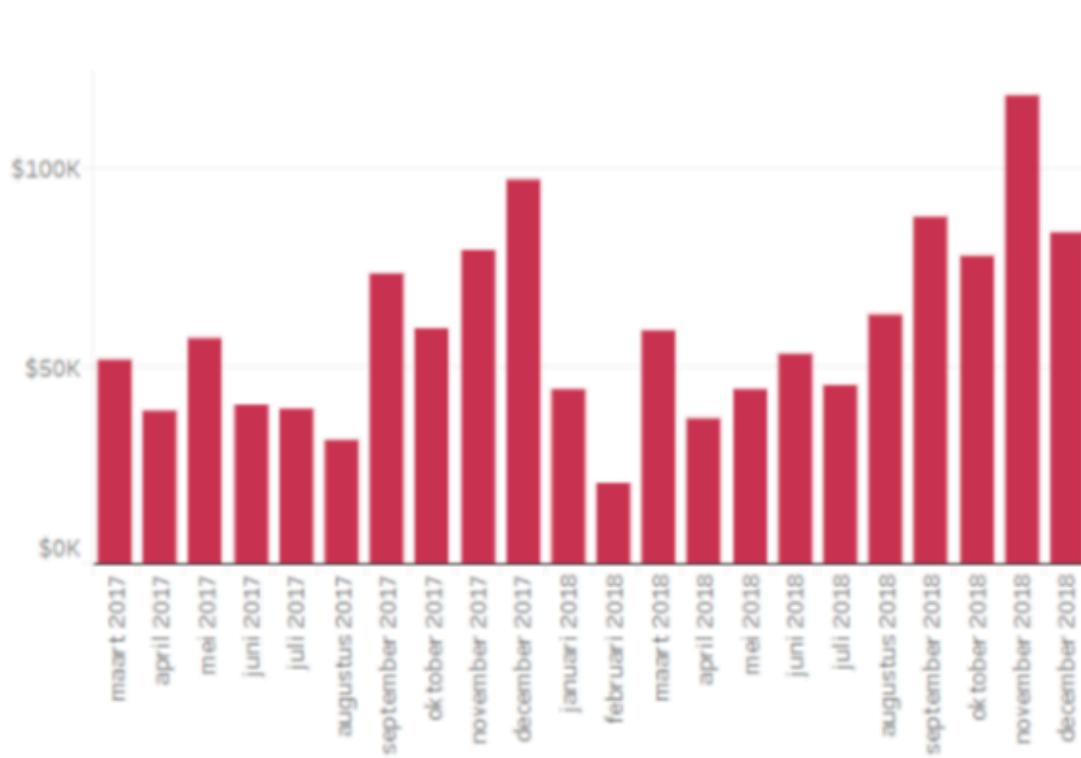
Het weglaten van een deel van de as geeft een vertekend beeld.



Tip 8: gebruik titels, labels en legenda's

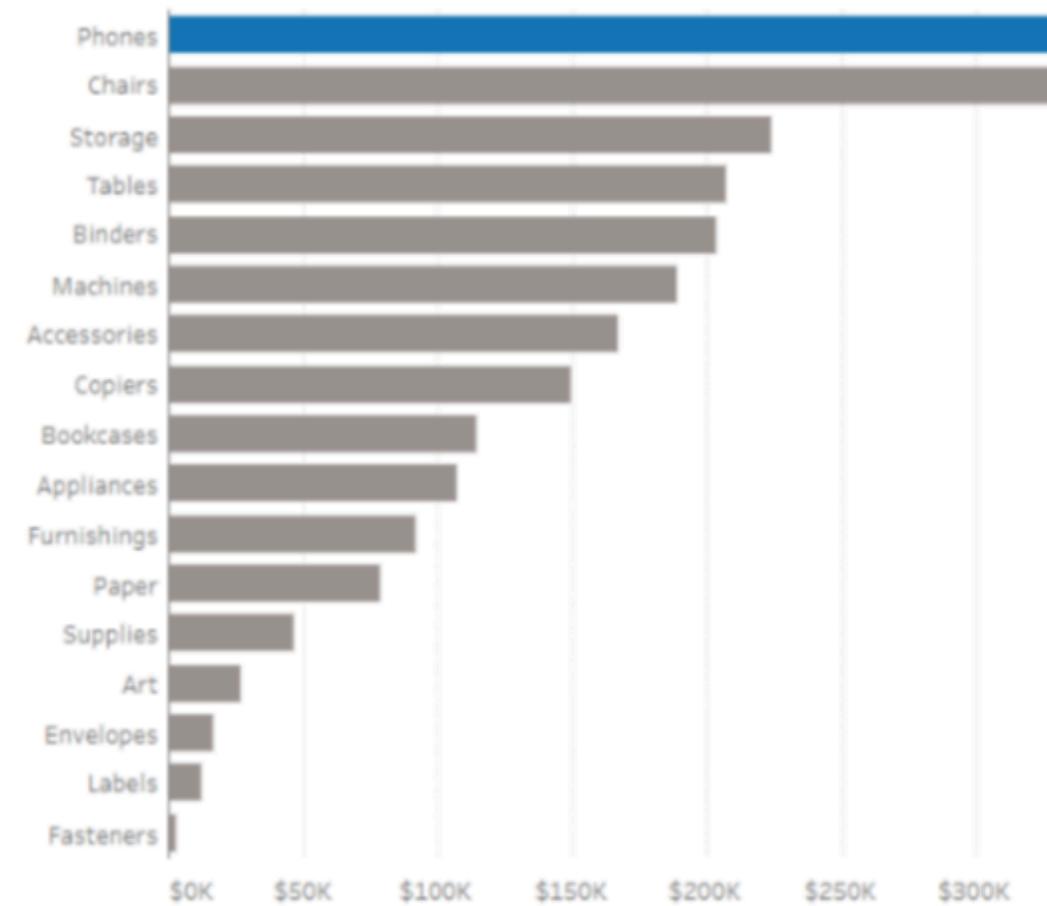
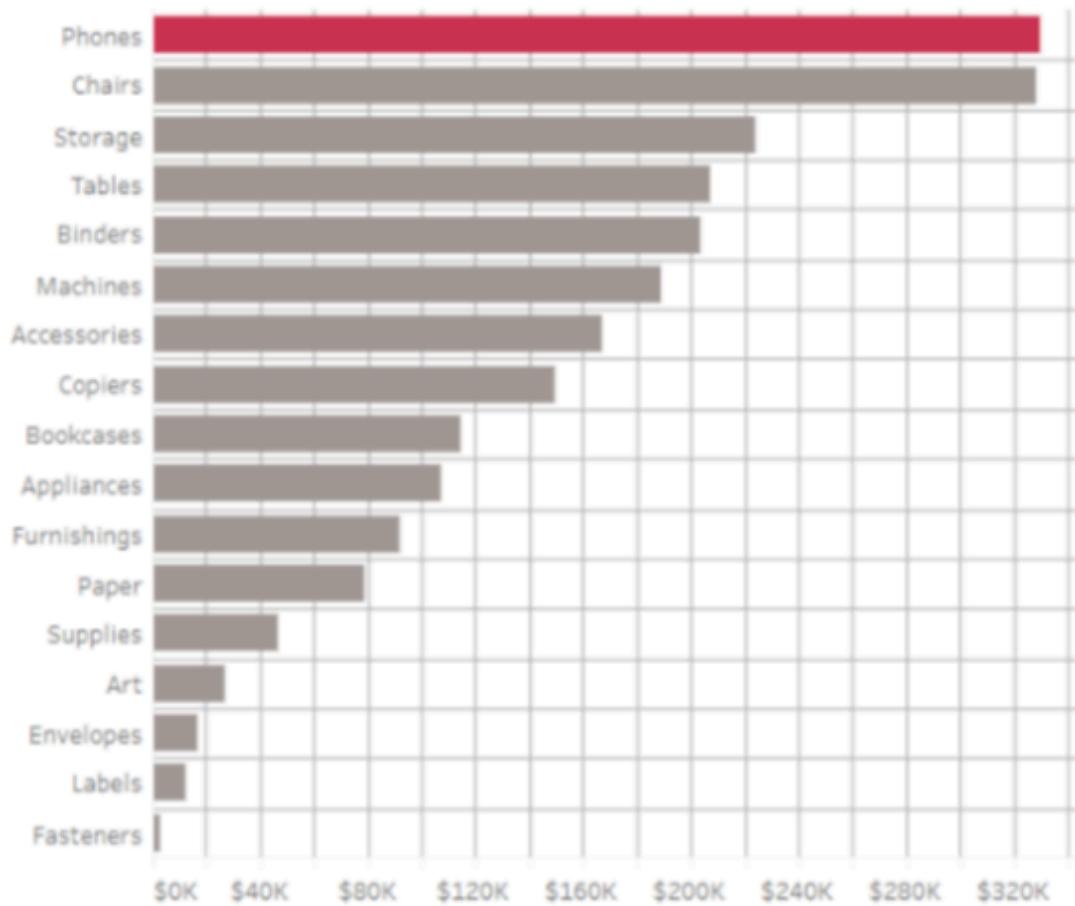
Titels, legenda's en labels maken meteen duidelijk wat er wordt weergegeven (blijf wel letten op de data-inkt ratio!).

Gebruik bij voorkeur geen verticale labels, dat leest moeilijk.

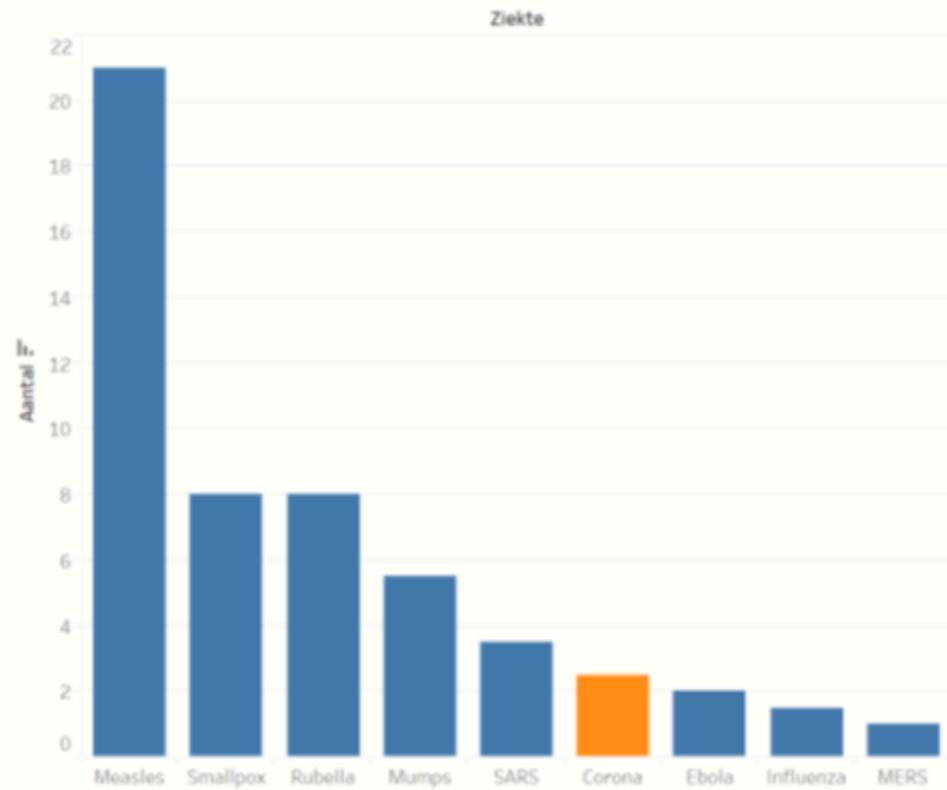
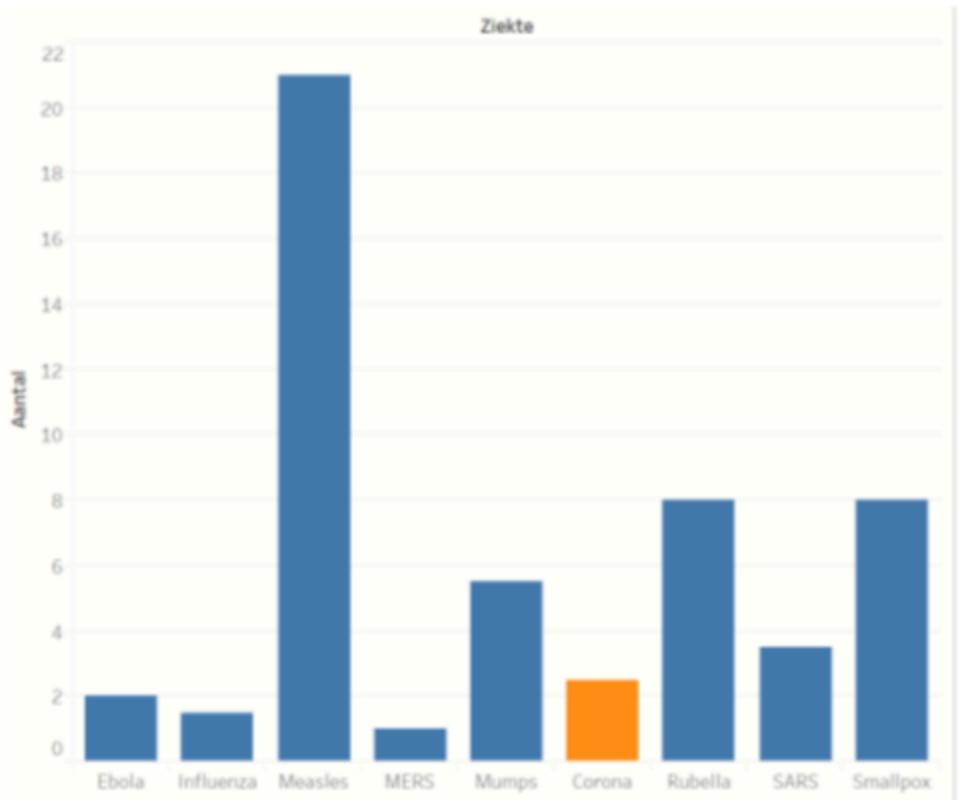


Tip 9: wees zuinig met grid-lijnen

Data-to-ink-ratio: gridlijnen zijn vaak overbodig.



Tip 10: sorteer indien nodig



Tip 11: vertel een verhaal

Dit zijn de ingrediënten om het potentieel van data over te brengen:

- Goede data visualisaties
- Goed verhaal

Tip 12: getallen afronden

38.1382%

SUCCESS RATE

\$8,682,317

TOTAL SALES

38.1%

SUCCESS RATE

\$8.7 M

TOTAL SALES

Tip 13: pre-attentive processing

Bepaalde visuele informatie wordt makkelijker opgepikt door het onderbewuste.



Length



Width



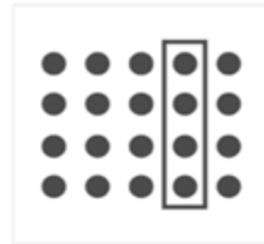
Orientation



Size



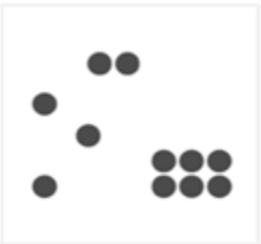
Shape



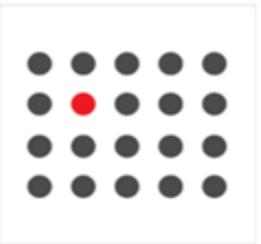
Enclosure



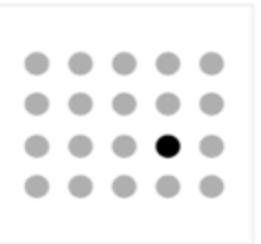
Position



Grouping



Color Hue



Color
Intensity

Lab

Data visualisatie

Lab Deel 1: MS Power BI labs

Maak een Power BI account met je studentenaccount.

Voer vervolgens de volgende stappen uit (klik op de links).

1. Gegevens voorbereiden in Power BI Desktop
2. **Gegevens laden in Power BI Desktop**
3. Modelgegevens in Power BI Desktop deel 1
4. Modelgegevens in Power BI Desktop deel 2
5. DAX-berekeningen maken in Power BI Desktop, deel 1
6. DAX-berekeningen maken in Power BI Desktop deel 2
7. **Een rapport ontwerpen in Power BI Desktop, deel 1**
8. Een rapport ontwerpen in Power BI Desktop deel 2
9. **Een Power BI-dashboard maken**
10. Gegevensanalyse uitvoeren in Power BI Desktop
11. Een gepagineerd Power BI-rapport maken

Lab Deel 2: Netflix dashboard

1. Installeer [Power BI desktop](#) (gebruik je AP Microsoft account)

- Opgelet: er bestaat geen Linux-versie

Linux gebruikers hebben een virtuele Windows omgeving nodig. [Hier kan je een voorbeeld vinden.](#)

Je hebt hiervoor een [Windows 10 Disc Image](#) nodig.

2. We gebruiken Power BI om een dashboard te maken over Netflix gebruik (als je zelf een Netflix account hebt [kan je ook jou geschiedenis downloaden](#)).
3. Maak het Netflix dashboard in Power BI.



223,50

Watch Time (hours)

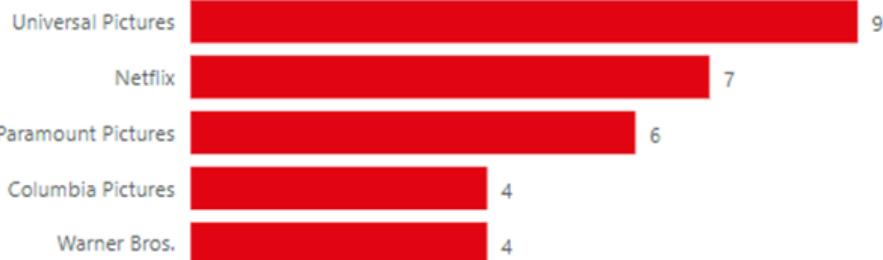
Movies by date



My favorite genres



My favorite producers

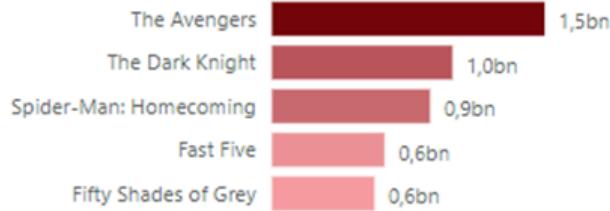


Filters:

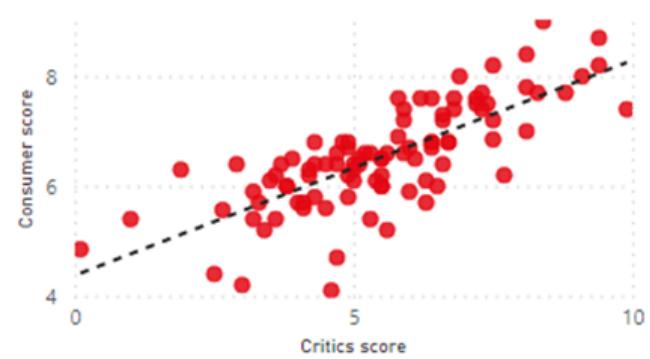
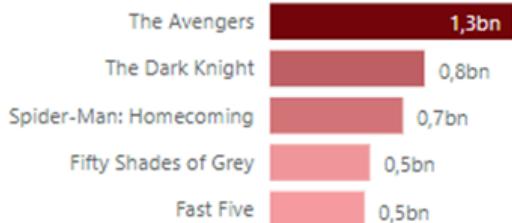
year

Alle

Top Movies by income



Top Movies by profit

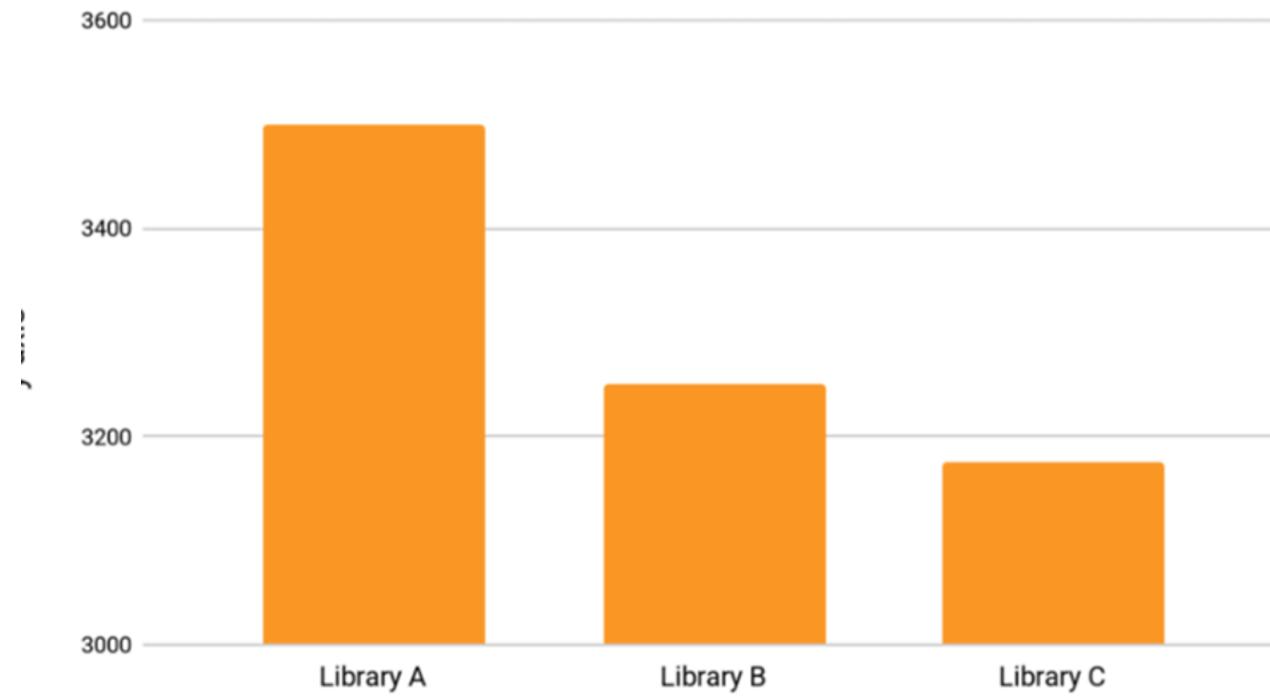


Quiz

Welke visualisatie zou je gebruiken om de winkelverkoop per week weer te geven?

Wat kan er beter?

ooks circulated this month



**Welke term gebruiken we als er teveel informatie op
1 grafiek staat?**

Wat is één van de belangrijke vragen bij data visualisatie?

Einde hoofdstuk 2