

# Visualiseer jouw data

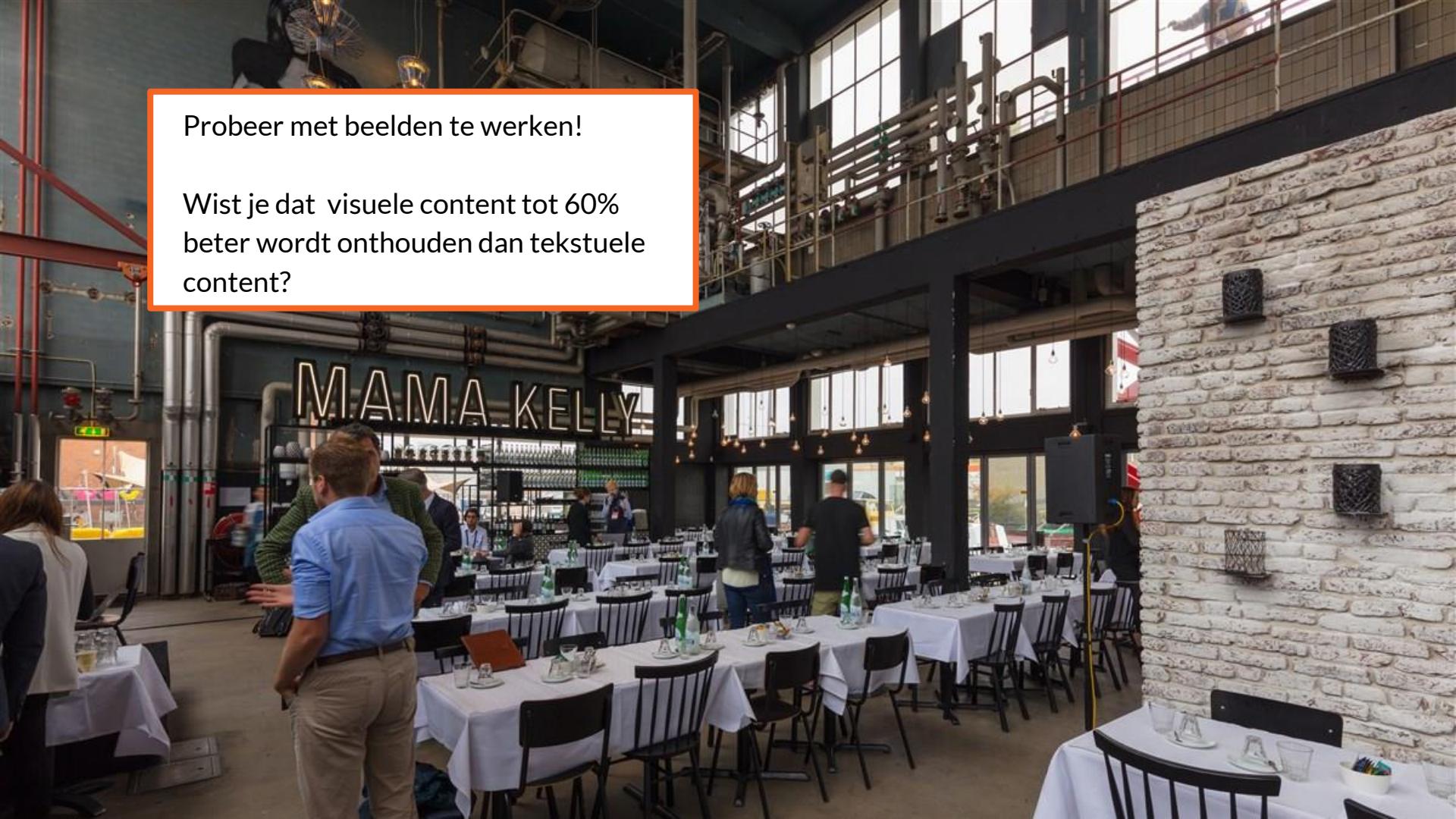
Madelijn Bazen | provincie Zuid-Holland

13 september 2018



#innovatiefZuidHolland





Probeer met beelden te werken!

Wist je dat visuele content tot 60% beter wordt onthouden dan tekstuele content?



REPORTAGE HIGHTECH IN DE STEDEN  
De moderne ambtenaar is een nerd

# Korte introductie

- Beleidsmedewerker Datavisualisatie | Team Data Analytics provincie ZH
- Voormalig statisticus
- Achtergrond in psychologie en statistiek
- Fan van  en  

---

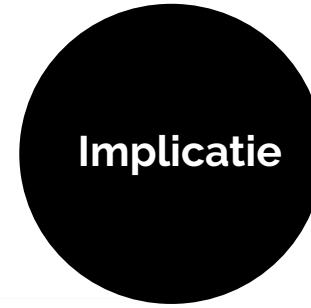
# Programma

- Waarom datavisualisatie?
  - Hoe dan?
    - Visual encodings
    - Edward Tufte's Design Rules
    - Kleur
    - Tools
    - Tips voor een onvergetelijke datavisualisatie
  - Afronding
-

# Waarom datavisualisatie?



# Waarom datavisualisatie?

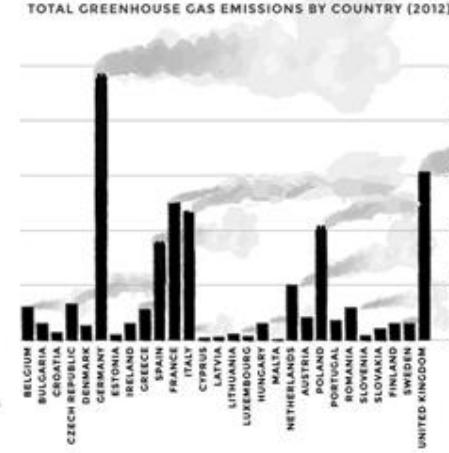


## Doelen:

1. Krijg mensen aan het lezen
2. Trek mensen met een visuele leerstijlen aan
3. Ga mee met de hype
4. Creëer iets moois



## TOWARDS A NEW CLIMATE AGREEMENT: PARIS 2015

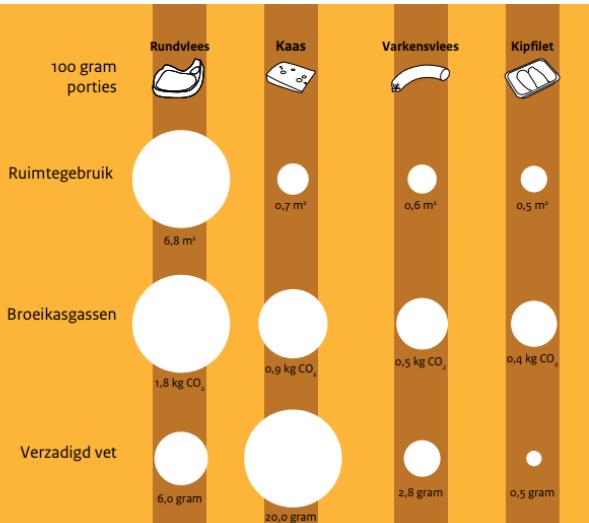


# Waarom datavisualisatie?



## Welk PRODUCT, welke IMPACT?

Als we minder vlees eten, zijn we voor onze eiwitbenodiging op andere voedingsmiddelen. Op de landbouwuniversiteit in Wageningen buigen voedingswetenschappers zich over de potentie van algen, zeewier en insecten en ontwikkelen zij alternatieve voedselproducten, bijvoorbeeld krekelprikketten.



1. Begrijpen
2. Uitleggen
3. Laat mensen zelf ontdekken

Bron: PBL

# Waarom datavisualisatie?

Aandacht &  
Schoonheid

Begrip

Implicatie



1. Overtuigen
2. Denkbeelden corrigeren
3. Gevoelens oproepen
4. Aanzetten tot actie

---

**HOE DAN?!**

---

# Oefening

Bedenk zoveel mogelijk manieren om de dataset [38, 77] te visualiseren

Wat kunnen de cijfers representeren?

- A. Uurloon
- B. Aantal stemmen voor en tegen referendum



# Take-aways

- Er zijn veel manieren om een dataset te visualiseren
- Grenzen tussen types visualisaties zijn vaag
- Welke is het best? Hangt af van situatie

## Which Visualization? A Quick Reference

You have the following data.  
**Categories** in the first two rows  
a single **ordered category**,  
and a few **continuous metrics**.

Here's how to plot them:

Categories		Ordered Cats	Continuous Metrics		
City	Airline	Class	Month	Distance	FlightTime
Alphaville	XeroTrip	\$	1	300	1
Betastan	YoloFly	\$\$	2	500	1
Chicago	ZeusAir	\$\$\$	3	650	2

Metric, split by 1 category

Bar (Row)

Price

A

B

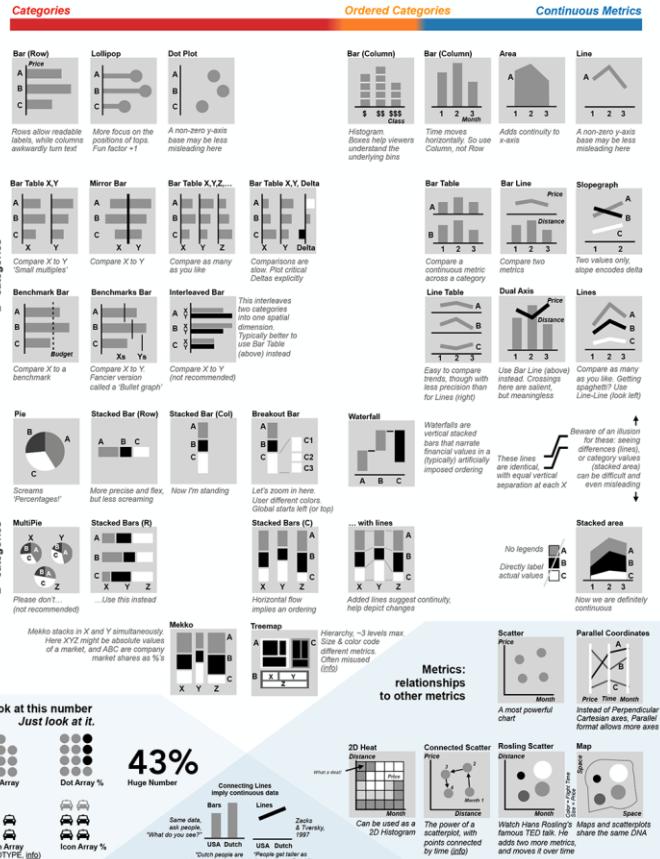
C

Rows allow reading values across columns

... by  
2+ categories

Part-to-Whole, split by 1 category

**Look at this number**



# Grafiekwijzers



## Categorie

### **Ordered Categories**

Metric, split by

by ...

Bar (Row)
A
B
C

A dot plot with three data points labeled A, B, and C.

## Bar (Column)

*Histogram. Boxes help convey the underlying bins.*

A bar chart titled "Bar Table X" with three bars labeled A, B, and C. Bar A has a value of 2, Bar B has a value of 3, and Bar C has a value of 1.

Category	Value
A	2
B	3
C	1

,Y	Mirror Bar
A	
B	
C	

A bar chart titled "Bar Table". The y-axis has three categories labeled A, B, and C. Each category has a corresponding gray bar. Category A is at the top, followed by B, and then C at the bottom.

Category	Value
A	1
B	2
C	3

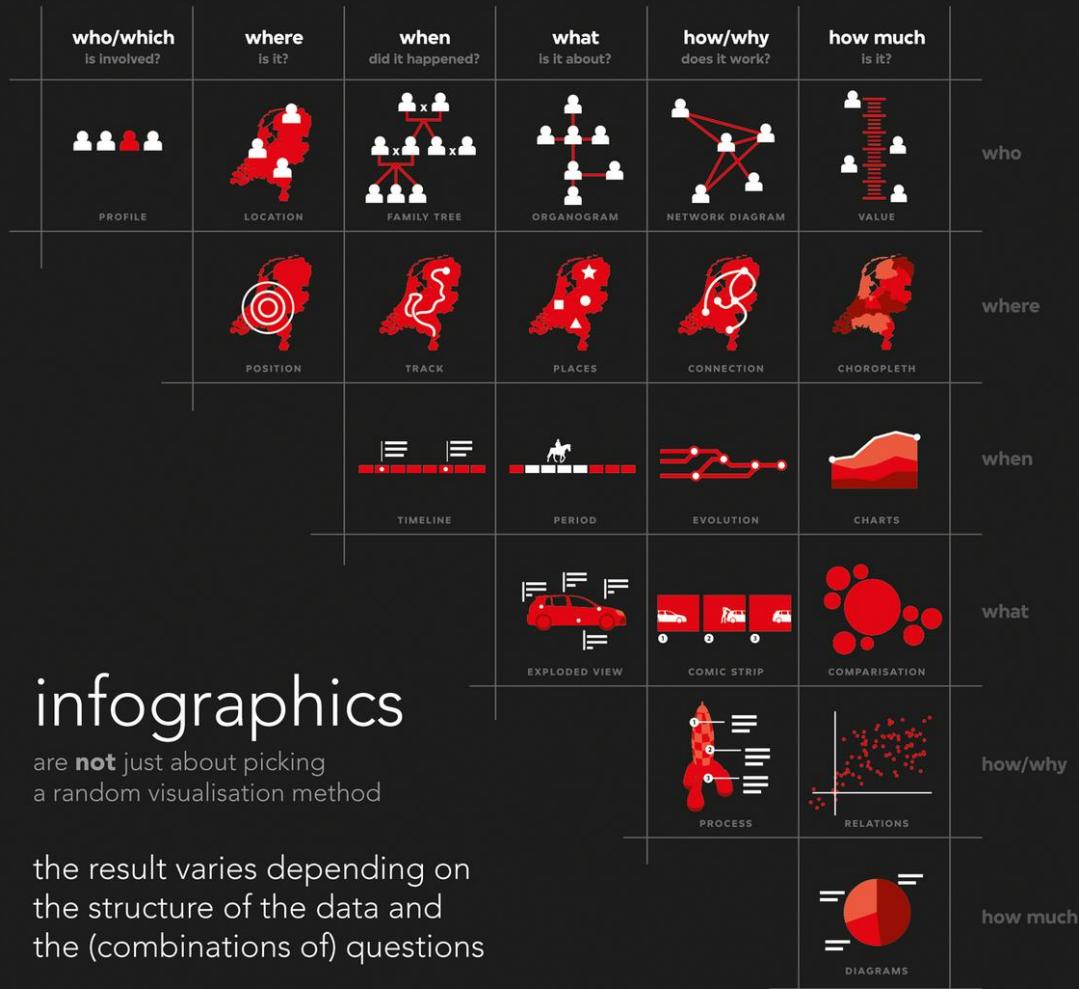
A bar chart titled "Benchmark" with three bars labeled A, B, and C. The y-axis has tick marks for A, B, and C. The x-axis has a vertical dashed line at the right end of the bars. Bar A is gray, Bar B is white, and Bar C is gray.

Category	Benchmark
A	100
B	100
C	100

A horizontal bar chart comparing performance metrics for three benchmarks (A, B, C) across two categories (Bar and Benchmark). The Y-axis represents the metric value, ranging from 0 to 100. The X-axis categories are Bar and Benchmark. For each category, there are three bars representing benchmarks A, B, and C. Benchmark A consistently shows the highest values, followed by Benchmark B, and then Benchmark C.

Category	Benchmark A	Benchmark B	Benchmark C
Bar	~85	~75	~65
Benchmark	~95	~85	~75

Brc



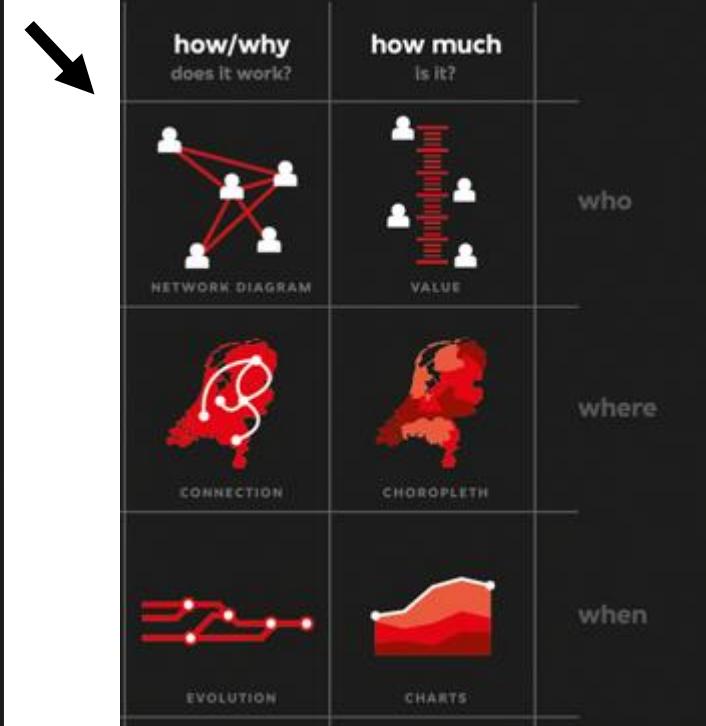
# infographics

are **not** just about picking  
a random visualisation method

the result varies depending on  
the structure of the data and  
the (combinations of) questions

Please feel free to comment @fruys

Frédéric Ruyts, Vizualism 2013.03.13



bron

# Visual Encodings

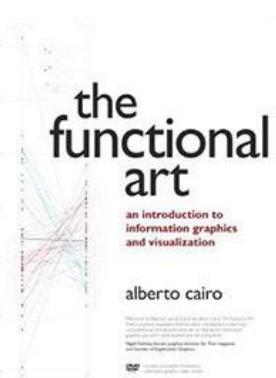
22%,

25%,

34%,

29%,

32%

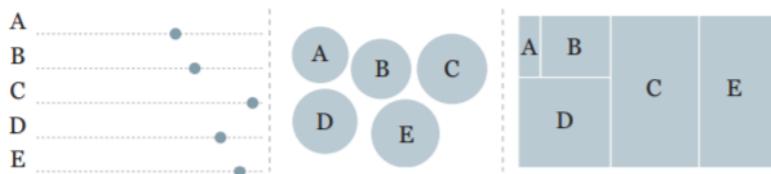


Bron: Alberto Cairo

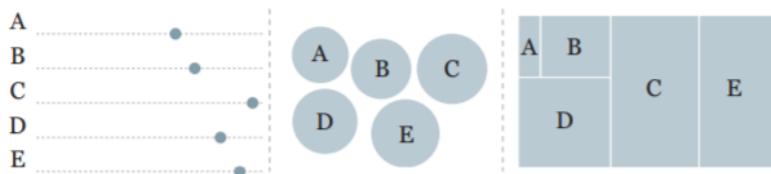
## *Length or height*



## *Position*



## *Area*



## *Angle/area*



## *Line weight*



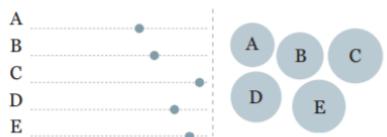
## *Hue and shade*



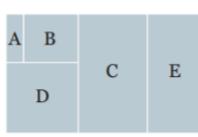
## Length or height



## Position



## Area



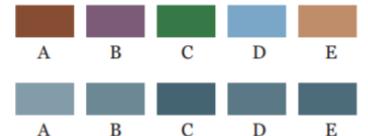
## Angle/area



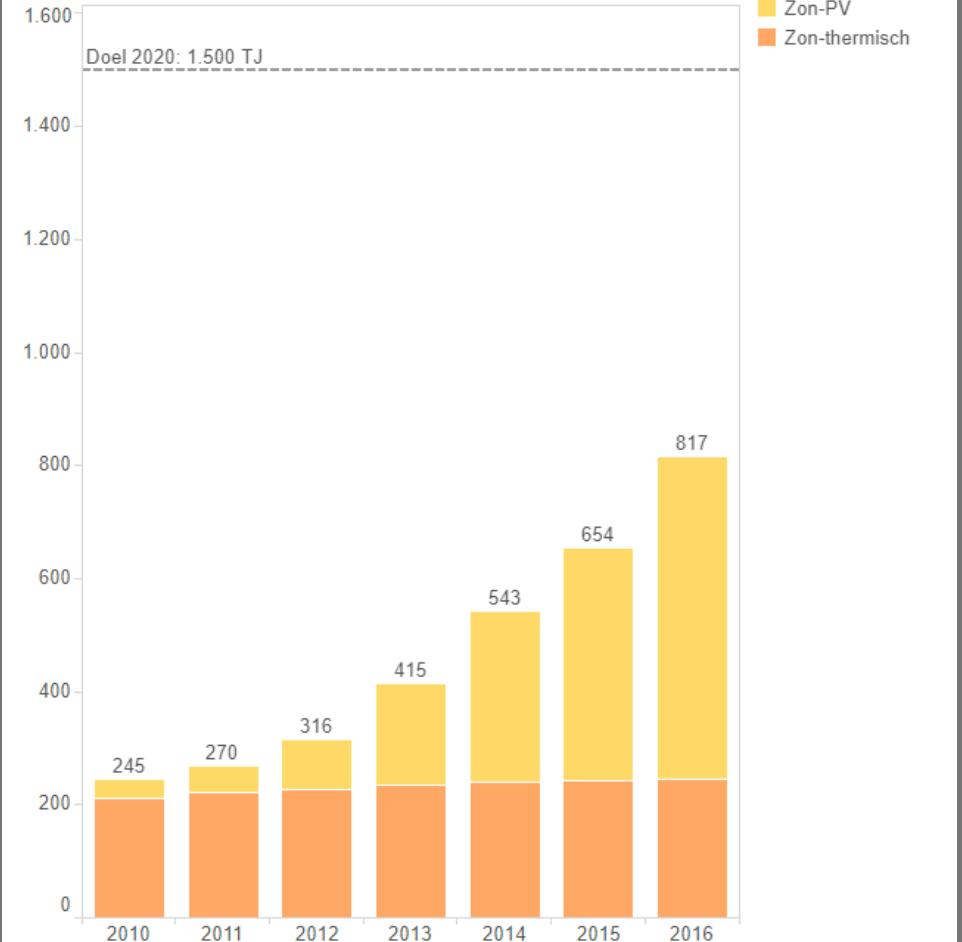
## Line weight



## Hue and shade



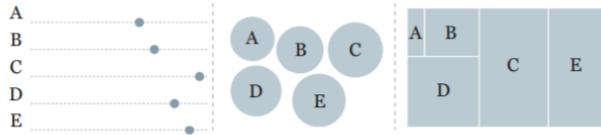
## Zonne-energie 2010-2016 in Zuid-Holland



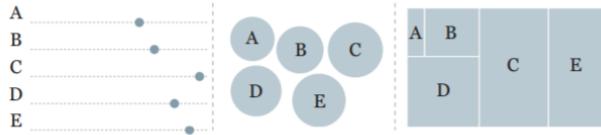
## Length or height



## Position



## Area



## Angle/area



## Line weight

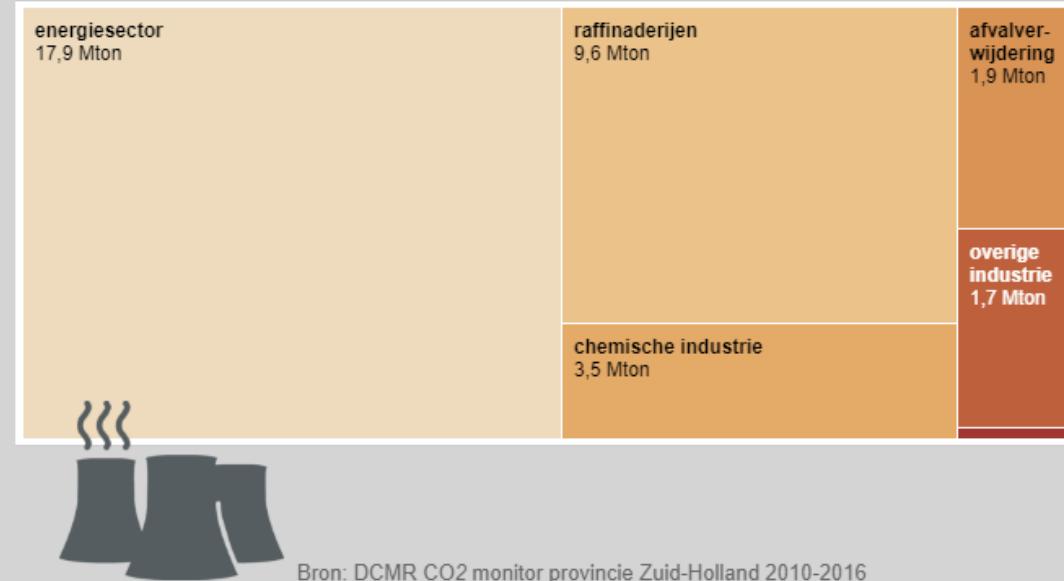


## Hue and shade

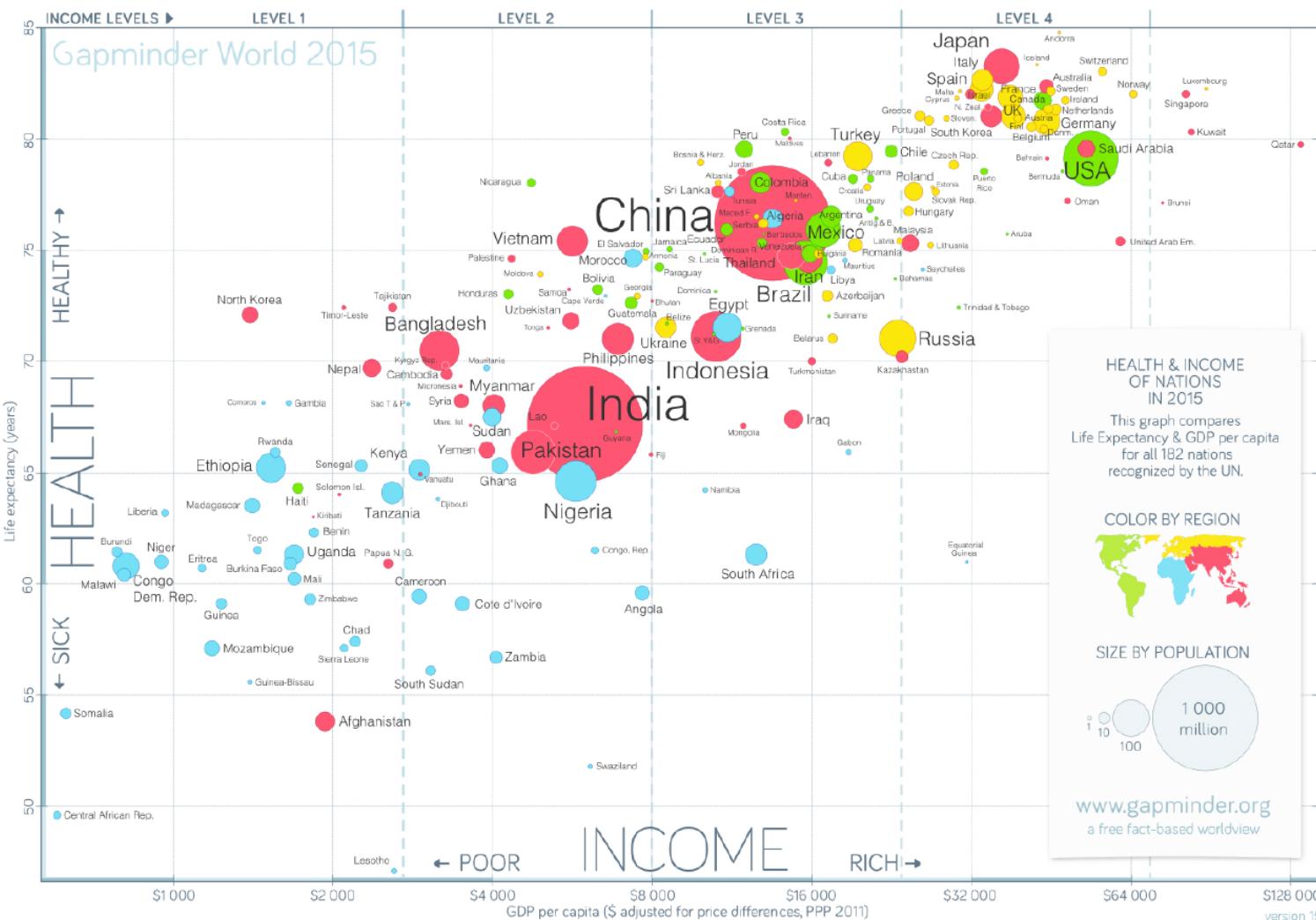


## Energiesector en raffinaderijen grootste aandeel in CO2-uitstoot industrie

Verdeling CO2-uitstoot over de sectoren binnen de industrie Zuid-Holland

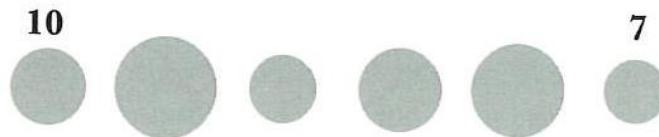
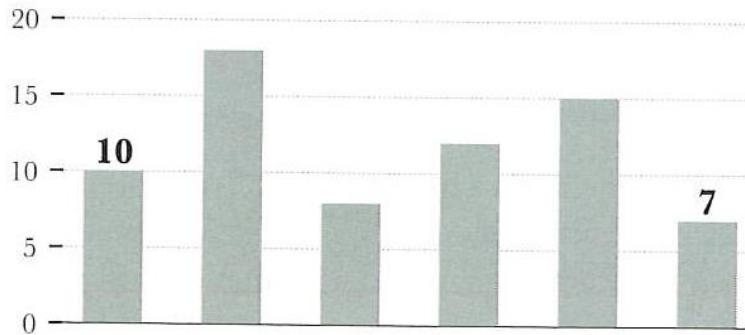


Bron: DCMR CO2 monitor provincie Zuid-Holland 2010-2016

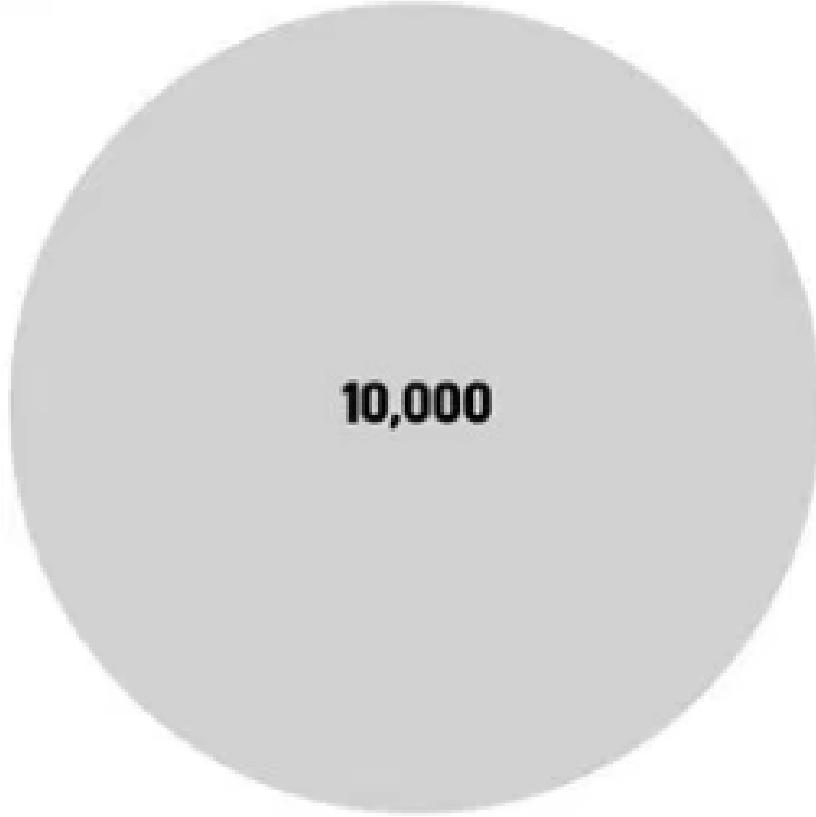


DATA SOURCES—INCOME: World Bank's GDP per capita, PPP (2011 international \$). Income of Syria & Cuba are Gapminder estimates. X-axis uses log-scale to make a doubling income show same distance on all levels. POPULATION: Data from UN Population Division. LIFE EXPECTANCY: HIINE GBD-2015, as of Oct. 2016. ANIMATING GRAPH: Go to [www.gapminder.org/tools](http://www.gapminder.org/tools) to see how this graph changed historically and compare 500 other indicators. LICENSE: Our charts are freely available under Creative Commons Attribution License. Please copy, share, modify, integrate and even sell them, as long as you mention "Based on a free chart from [www.gapminder.org](http://www.gapminder.org)".

# Ons brein ≠ computer



Bron: Alberto Cairo

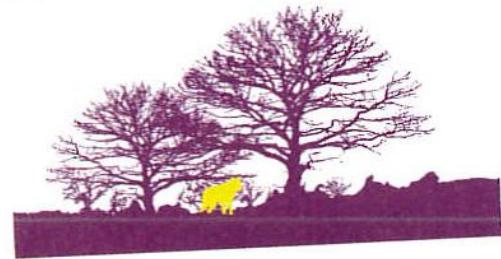
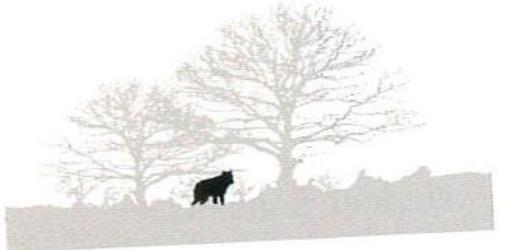


**10,000**



---

# Pre-attentive processing

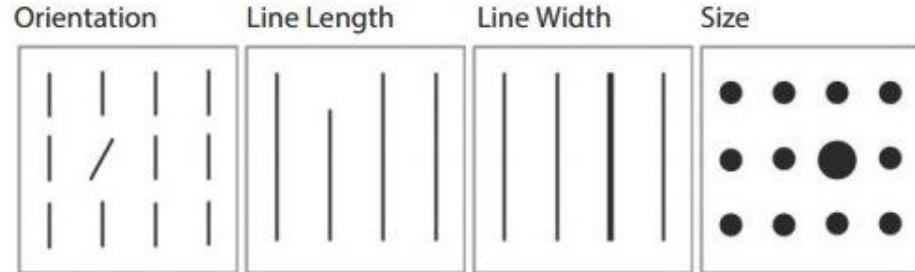


Bron: Alberto Cairo

# Pre-attentive processing

## Form

Orientation

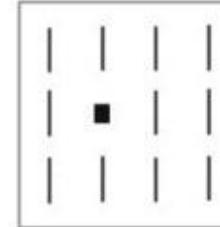


Line Length

Line Width

Size

Shape



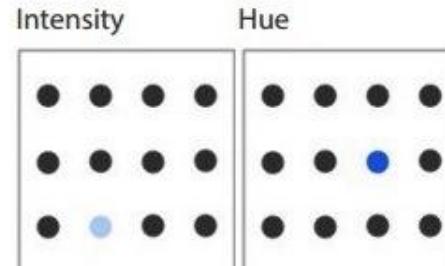
Curvature

Added Marks

Enclosure

## Color

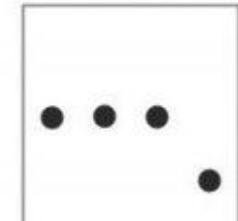
Intensity



Hue

## Spatial Position

2-D Position

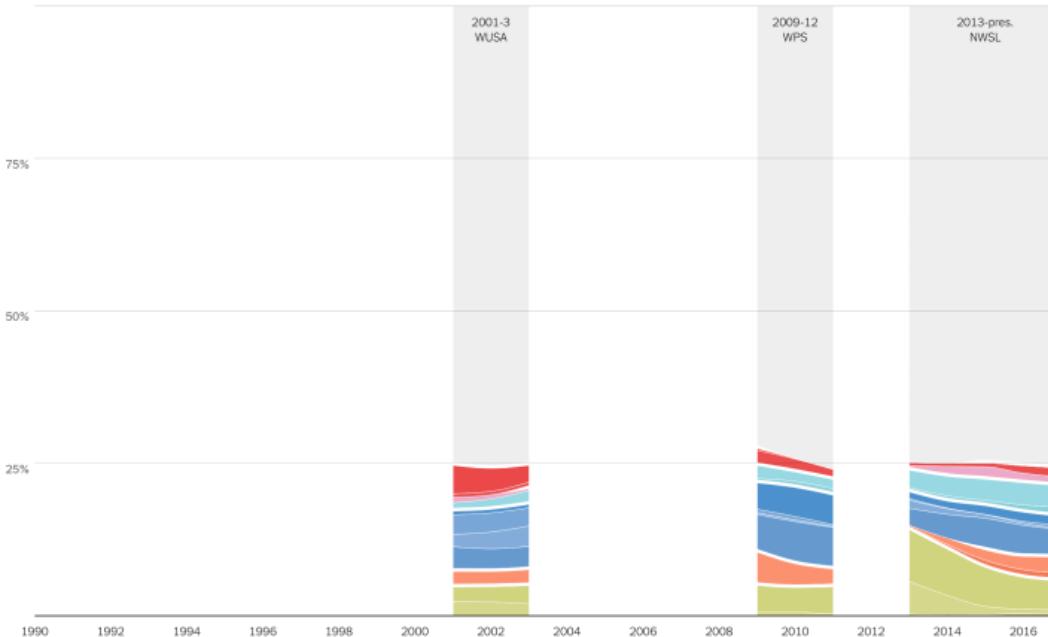




# Edward Tufte's Design Rules

# Tufte: “Let the data speak”

- Vermijd samenvattingen en aggregatie
- Laat zien wanneer data mist



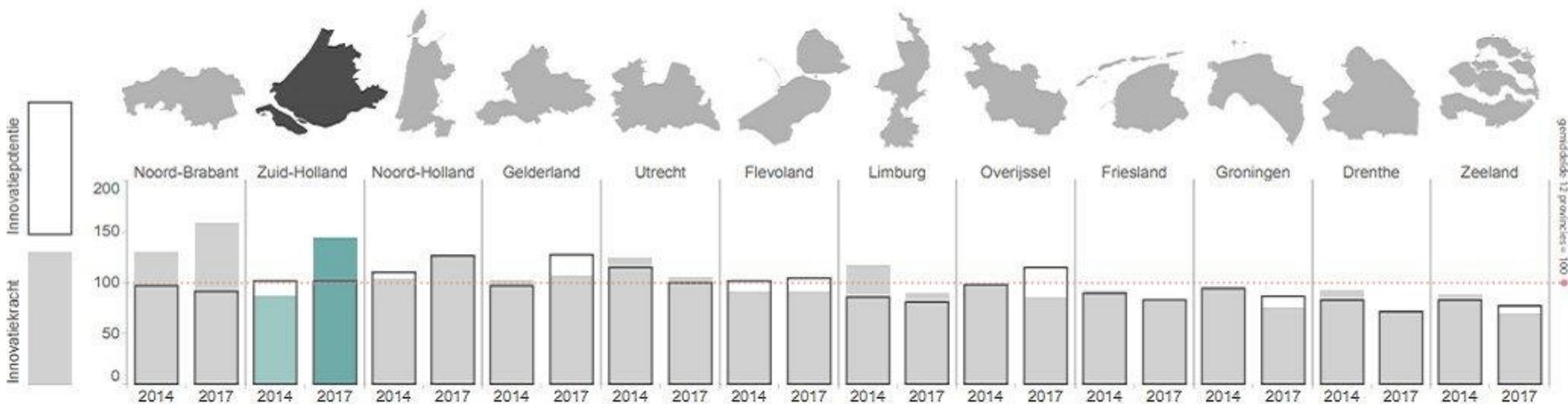
Bron: NY Times

# Tufte: “A picture is worth a thousand Words”



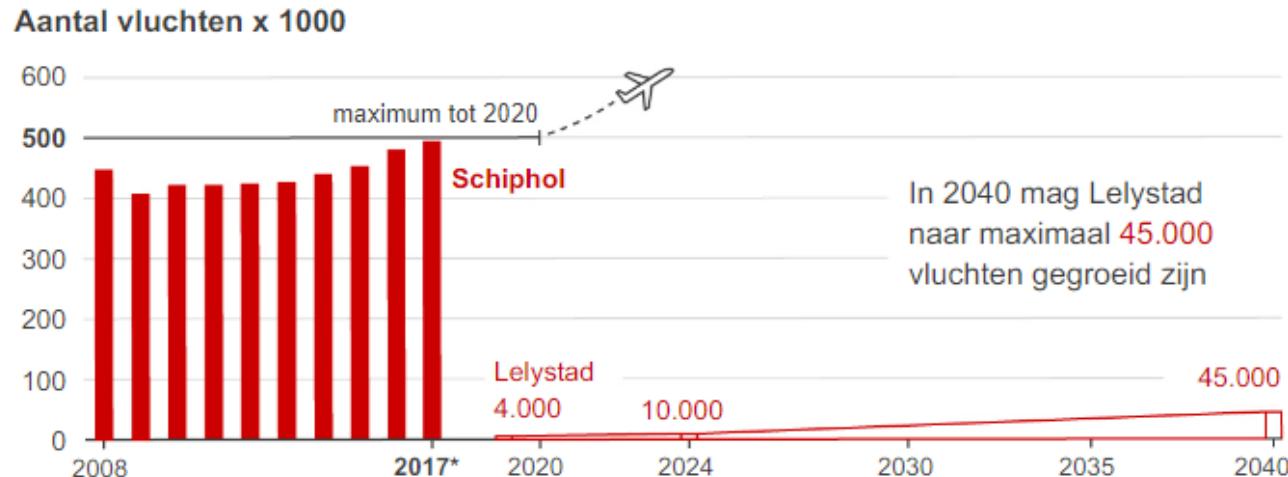
Innovatiekracht-index 2017: Zuid-Holland weet haar innovatiepotentieel beter te benutten

Innovatiekracht en Innovatiepotentie-index, gemiddelde van Nederland = 100. Ranking op innovatiekracht 2017.



# Tufte: annotatie

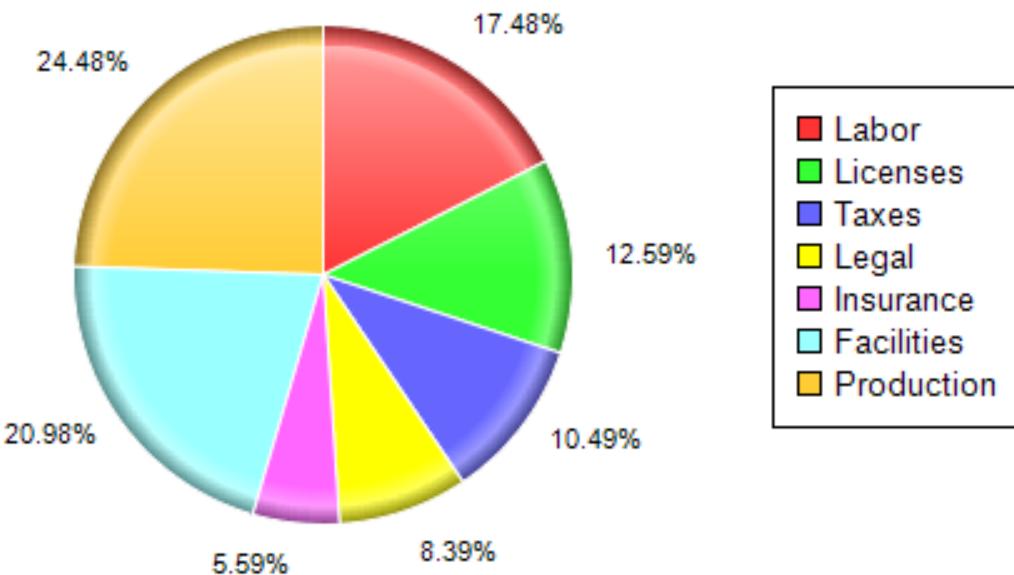
- Label je assen
- Een plaatje heeft nog steeds woorden nodig



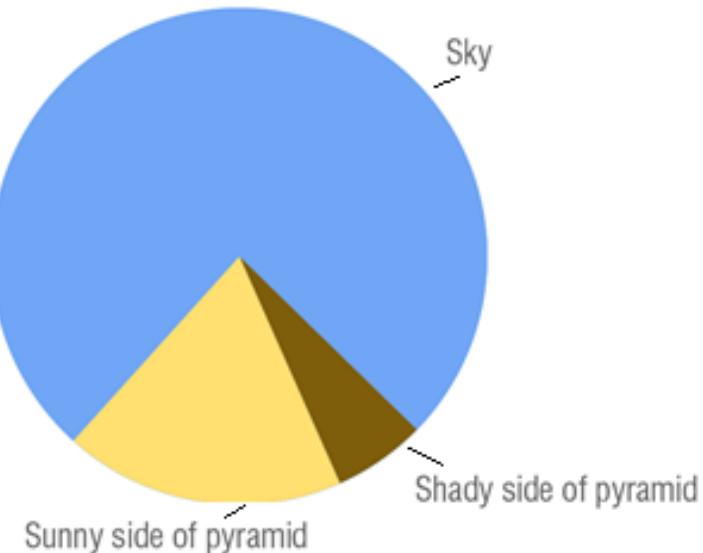
\* Cijfers over het vliegjaar 2017 (november 2016 - oktober 2017). Van het kalenderjaar 2017 zijn nog geen cijfers bekend

Bron: NOS op 3

# Labeltip: zet je legenda zo dicht mogelijk bij de data



Egyptian Pie Chart

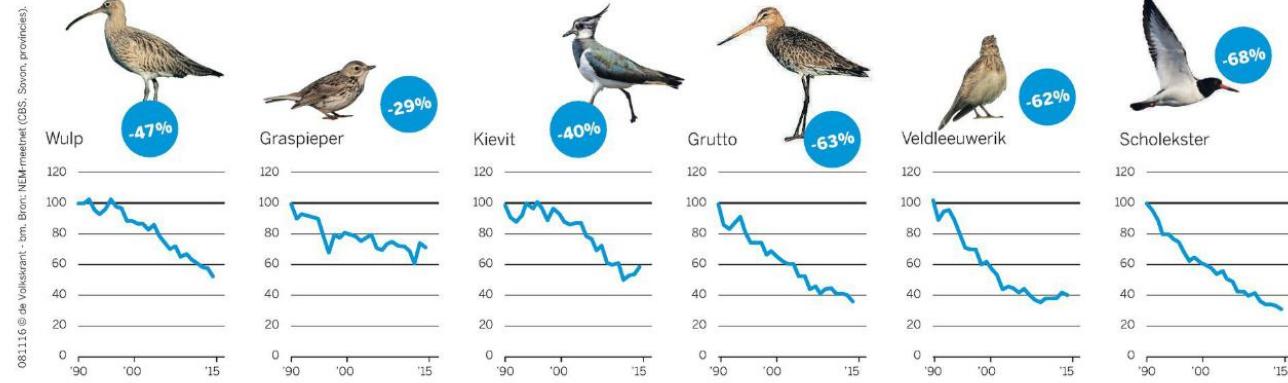
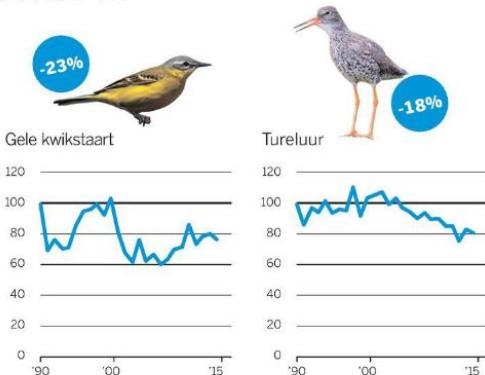


# Tufte: gebruik small multiples

- Constant design legt de nadruk op de data, niet op het design

## FORSE TERUGGANG WEIDEVOGELS

Index, 1990 = 100

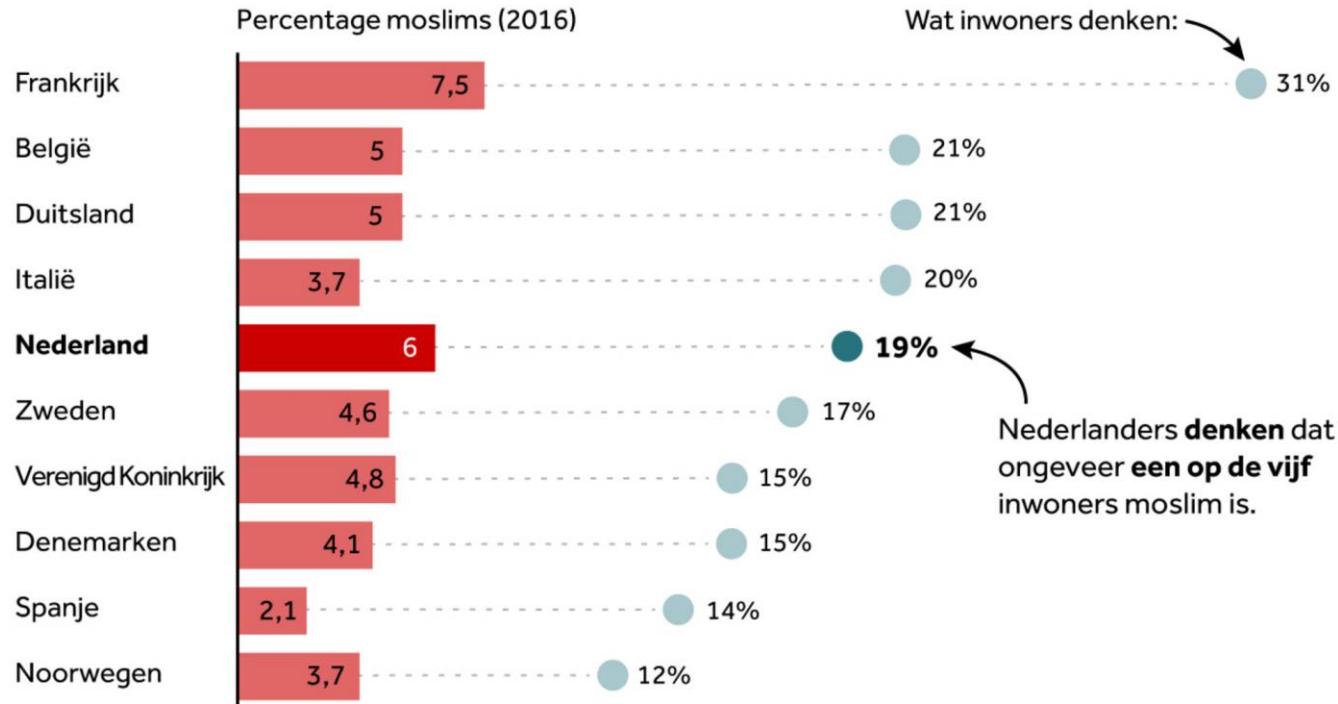


---

KLEUR

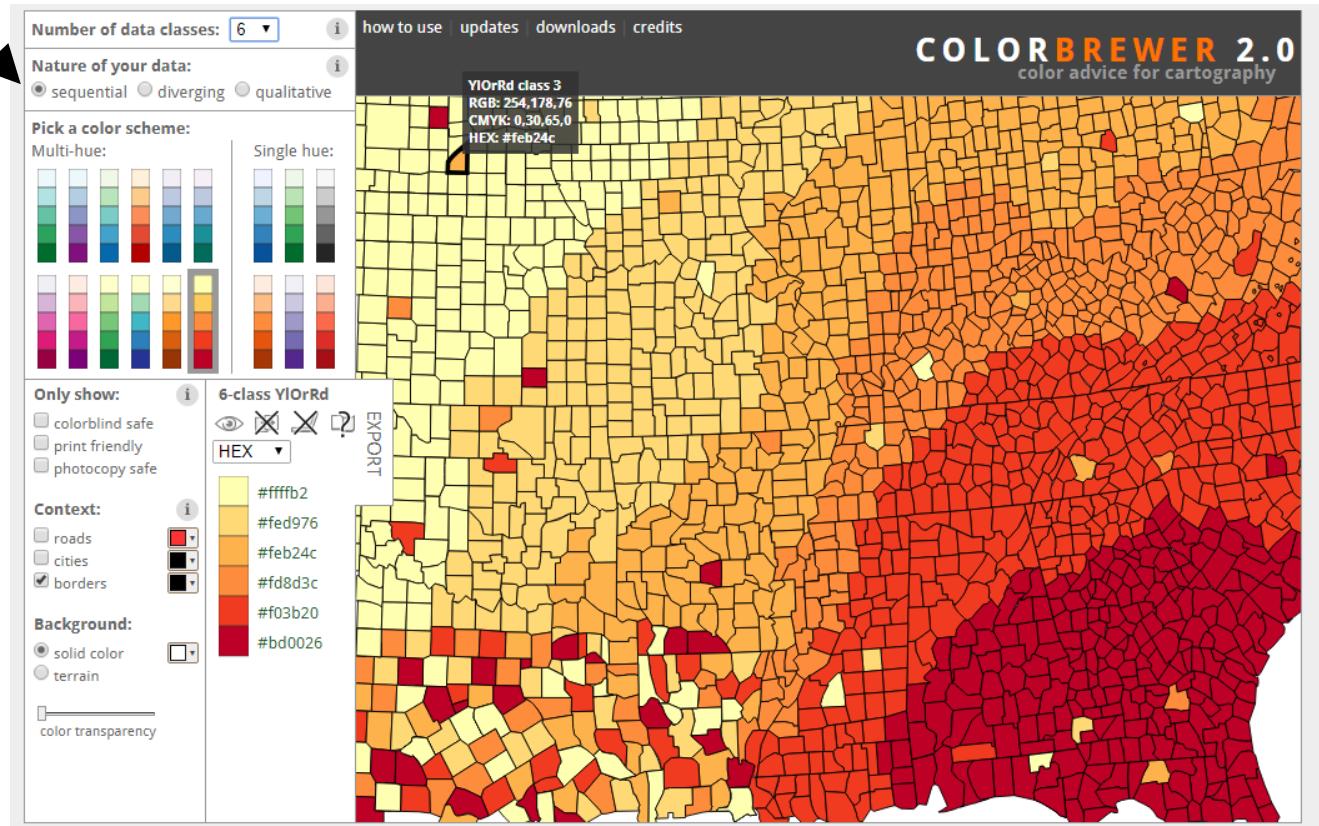
---

# Gebruik felle kleuren met name voor accenten en categorieën



Bron: NOS op 3

# Gebruik een kleurtint (of verwante kleurtinten) voor oplopende waarden



Ordering Direction

→ Sequential



→ Diverging

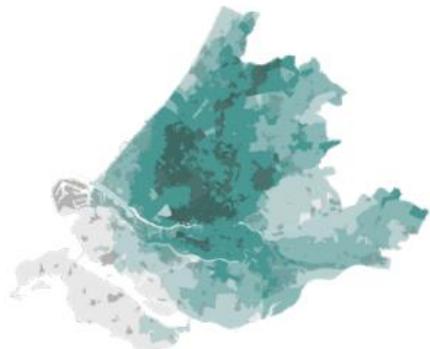


bron



# Voorbeeld oplopende waarden reeks

Bereikbaarheid banen met auto, ov



Auto

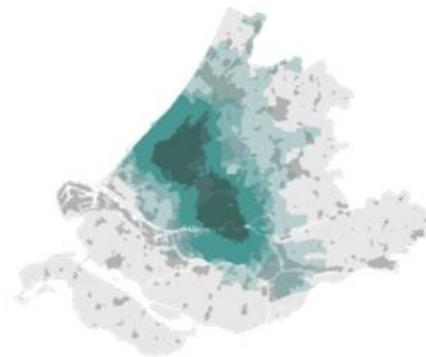
Bereikbaarheid van banen met de auto  
met afstandsvervalcurve

OV

Bereikbaarheid van banen met OV  
met voor- en natranspoort fiets + afstandsvervalcurve

De inwoners van Zuid-Holland moeten vlot van woon- naar werk-  
plek kunnen. Door inzet op alle vervoersmodaliteiten beperken we  
de reistijd. Deze is voor veel inwoners minder dan 1 uur.

en fiets



Fiets

Bereikbaarheid van banen met de fiets  
binnen 45 min / 30 min E-bike

Ordering Direction

→ Sequential



→ Diverging



Bereikbare banen

- zeer goed
- goed
- kansrijk
- matig kansrijk
- beperkt

■ bestaand stads- en dorpsgebied

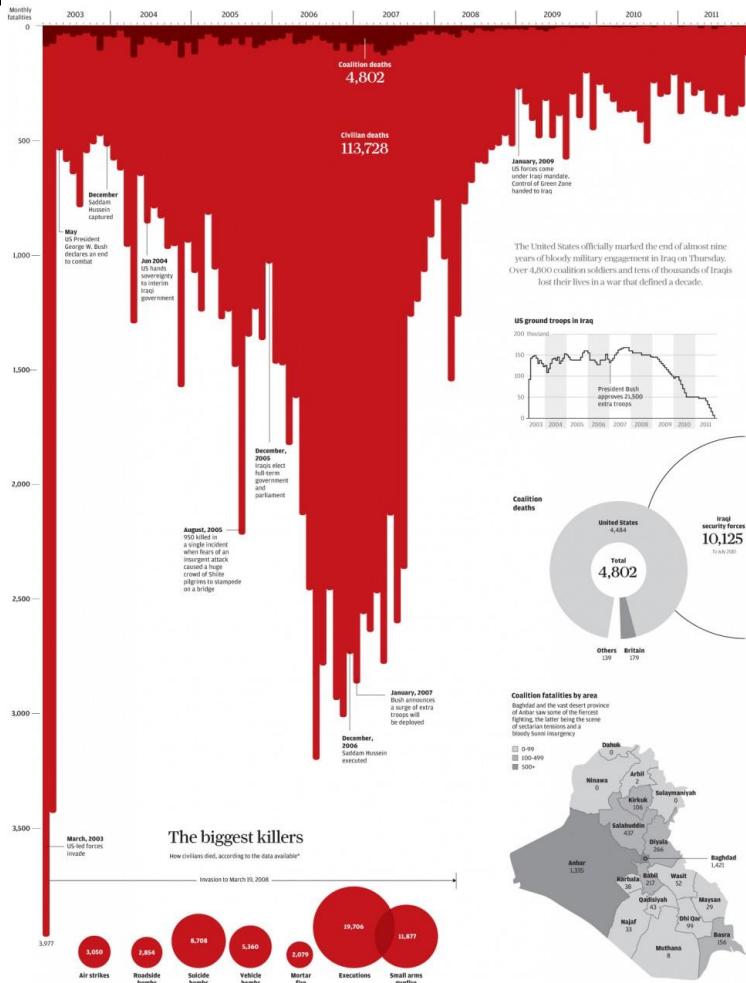
0

10km



# Iraq's bloody toll

## Gebruik semantisch logische kleuren



---

# TOOLS

---

# Mijn datavisualisatie toolbox



ArcMap



GEOGRAFISCHE DATA



BEELDBEWERKING



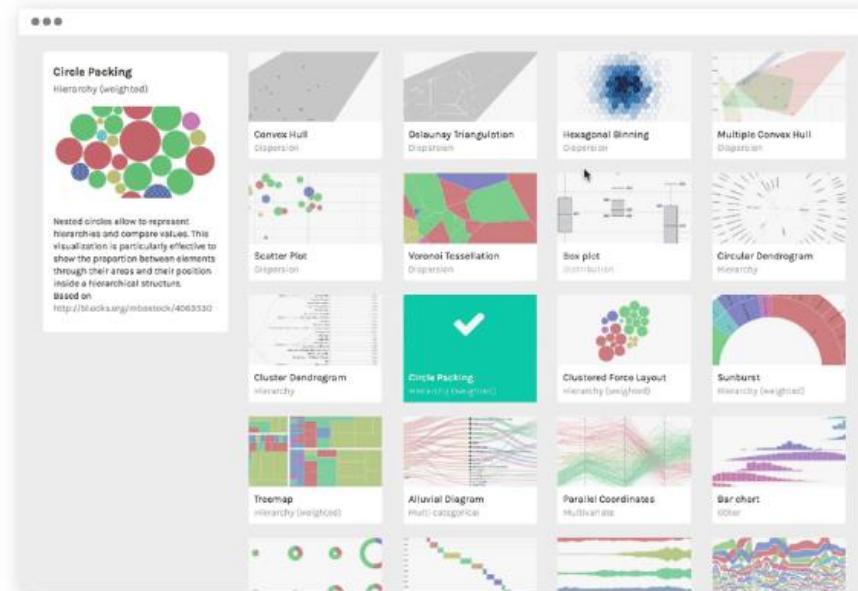
DATABEWERKING & VISUALISATIE



## 2 Choose within a wide range of visual models

### Conventional and unconventional layouts.

We designed and developed RAWGraphs with designers and vis geeks in mind. That's why we focused on providing charts that are not easy to produce with other tools. But don't worry, you can also find bar charts and pies! Something missing? See how easy is to [build your own model](#).



Pages

Page 1

Page 2

Rplc



### Principes van TOD, behaalde score versus maximale score



# FIGMA.COM

HoofdLabels

- 1. Loop
- 2. Fiets
- 3. Verblijf
- 4. Wonen
- 5. Meng
- 6. Verdicht
- 7. Bouw compact
- 8. Verande

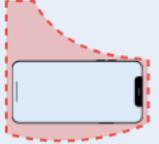


# 3 tips voor een onvergetelijke datavisualisatie

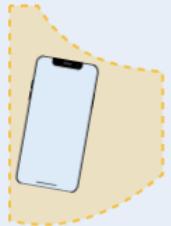


# Gebruik afbeeldingen

WOMEN



MEN



40% of women's pockets and 100% of men's pockets can fit this item.



iPhone X



Samsung Galaxy



Google Pixel



Front Wallet



Pen



Woman's Hand



Man's Hand

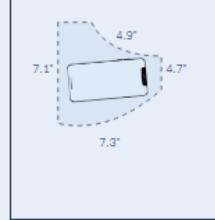
WOMEN'S



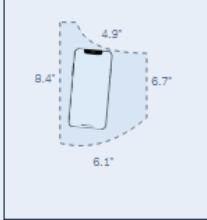
MEN'S



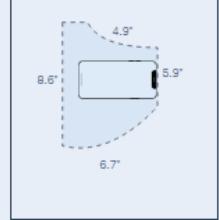
straight  
ABERCROMBIE



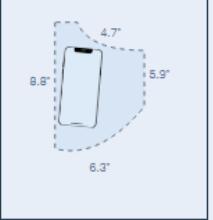
skinny  
ABERCROMBIE



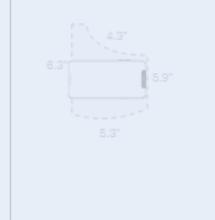
skinny  
ABERCROMBIE



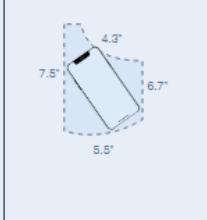
straight  
ABERCROMBIE



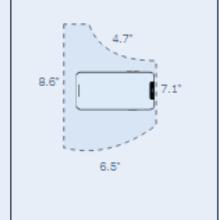
straight  
AMERICAN EAGLE



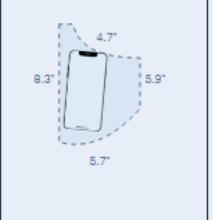
skinny  
AMERICAN EAGLE



straight  
AMERICAN EAGLE



skinny  
AMERICAN EAGLE



straight  
ARIZONA



skinny  
ARIZONA



skinny  
ARIZONA



straight  
ARIZONA





# Gebruik dubbele visual encodings

## Change to Higher Education Funding

Percent change to per-student funding compared to 2008, adjusted for inflation

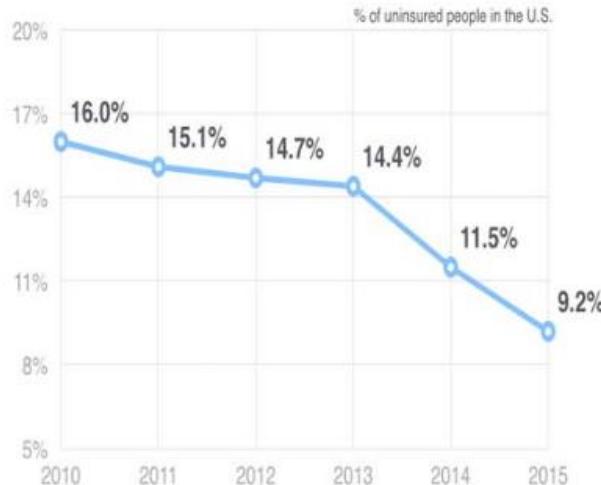


Bron: NY Times



# Lezers besteden de meeste aandacht aan de titel

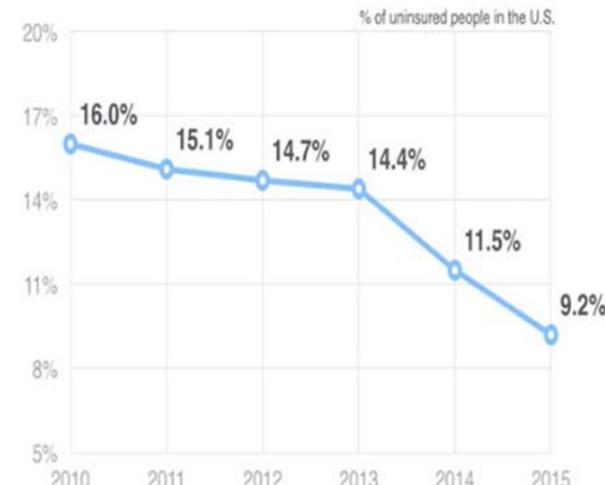
PERCENTAGE OF UNINSURED AMERICANS



Source: CDC/NCHS, National Health Interview Survey, 2010–2015

BAD TITLE

AMERICA'S UNINSURED RATE DIPS BELOW 10%



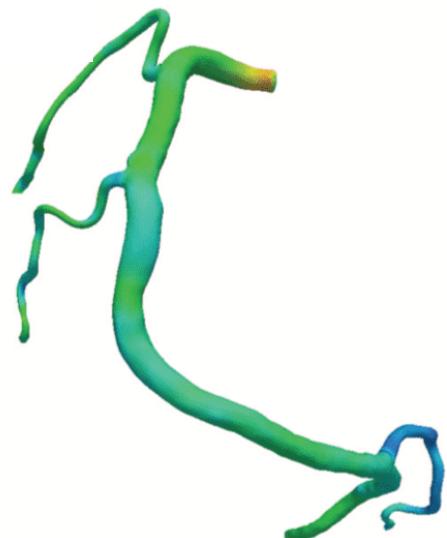
Source: CDC/NCHS, National Health Interview Survey, 2010–2015

GOOD TITLE

---

# Bonusmateriaal

A



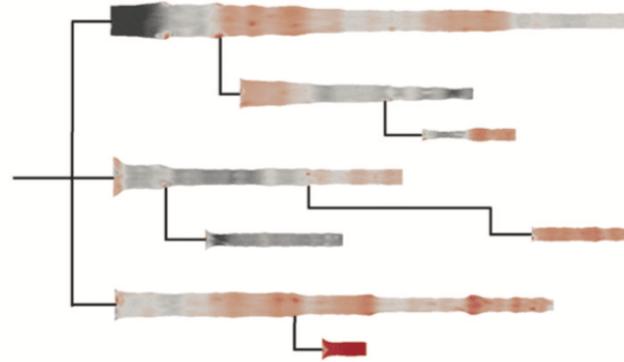
3  
2  
1  
0



B



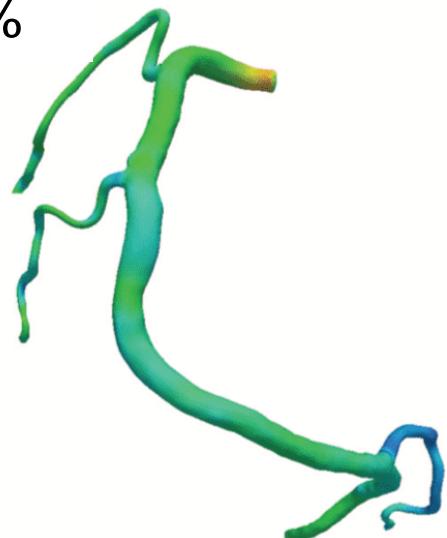
C



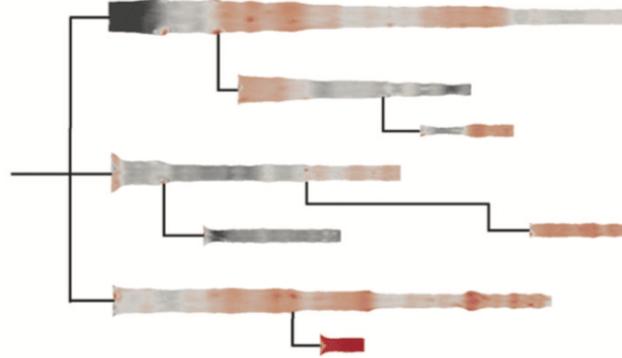
Shear  
Stress (Pa)

3  
2  
1  
0

A 39%

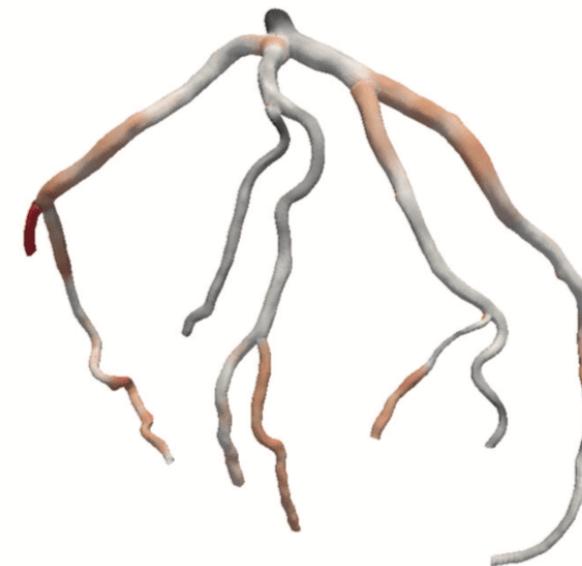


3  
2  
1  
0



C 91%

B 71%



3  
2  
1  
0

Shear Stress (Pa)

---

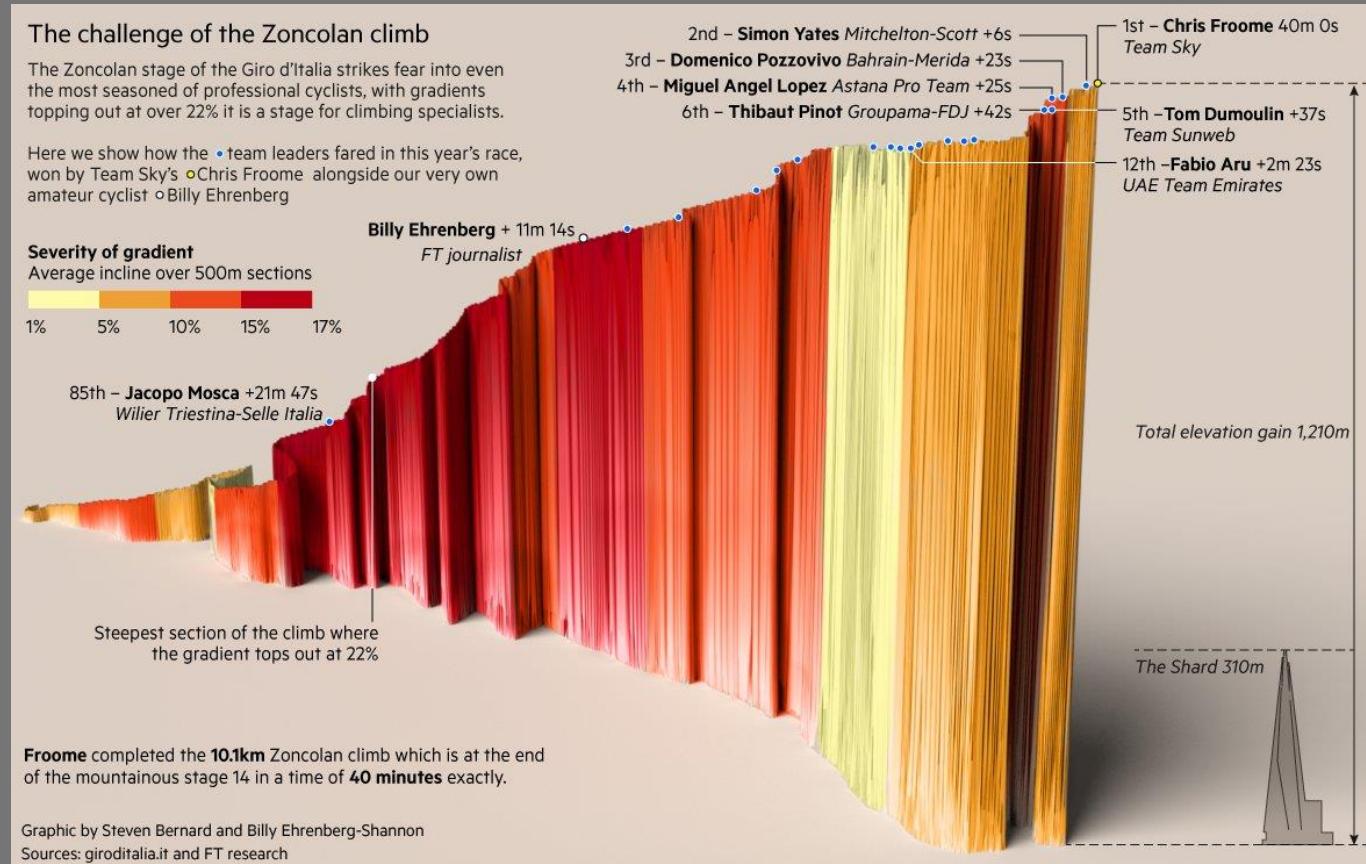
# Afronding

---

- Doe basiskennis + inspiratie op ✓  
..en ga zelf aan de slag!

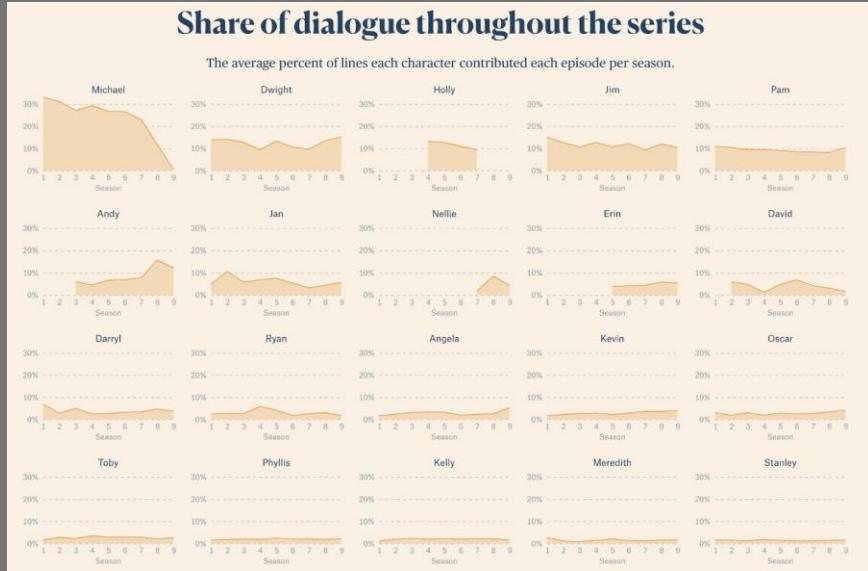


# – Annotatie: toelichtende tekst is belangrijk

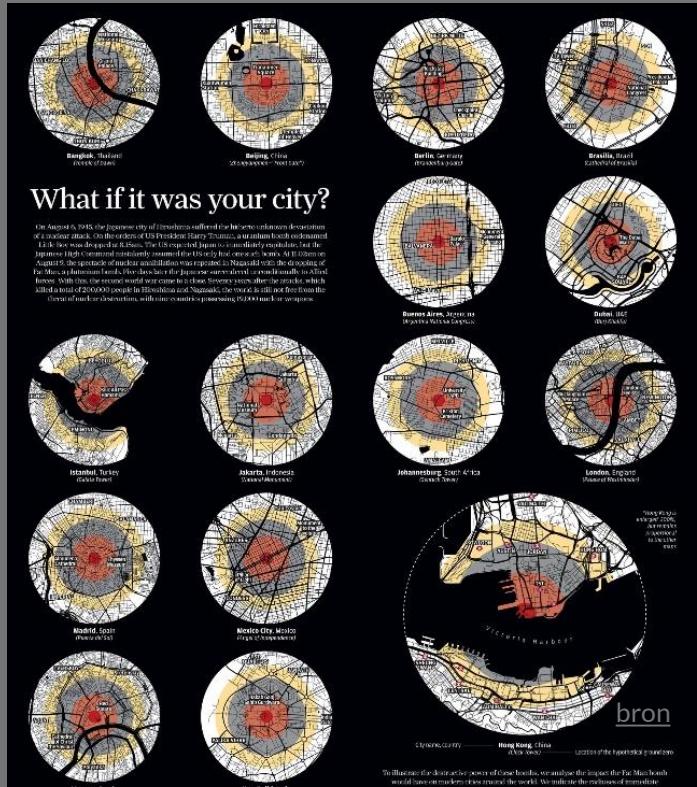


# - Experimenteer met visualisatietypes

## ..zoals small multiples



bron: pudding



---

# Meer weten?

## Podcasts

Data stories

PolicyViz

## Websites

Visualising data

Flowing data

## Online cursussen

Data visualization for storytelling & discovery | Alberto Cairo

Customizing Charts for Beauty & Impact | Nadieh Bremer



Dank!

# Vraag & Antwoord

Slides: <https://github.com/Madelijn/Talks/>

Twitter: @MadelijnBazen