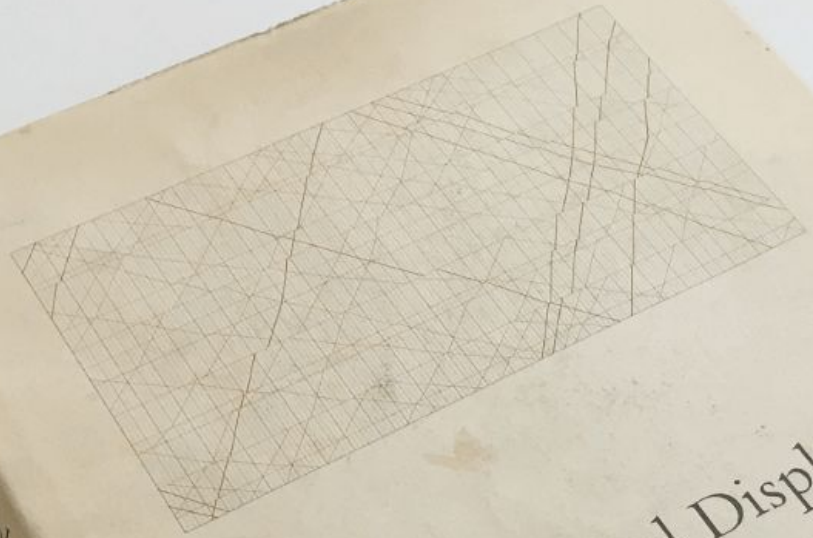
BECOME A DATA VISUALISATION SUPERHERO!

...WITH EDWARD TUFTE, GODFATHER OF DATA VISUALISATION

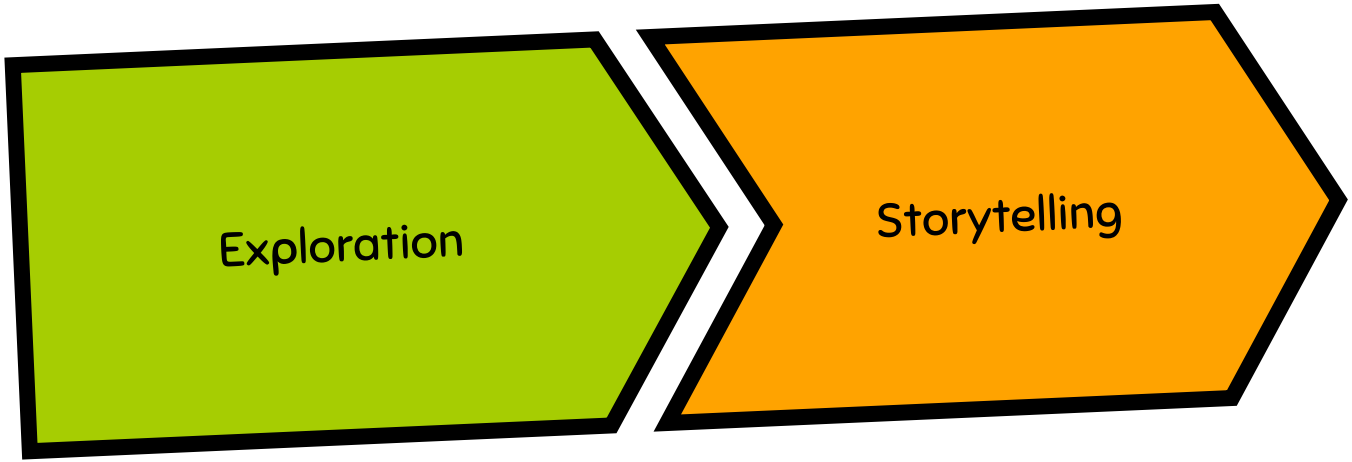


The Visual Display of Quantitative Information

EDWARD R. TUFTE

"EXCELLENCE IN STATISTICAL GRAPHICS CONSISTS OF COMPLEX IDEAS COMMUNICATED WITH CLARITY, PRECISION, AND EFFICIENCY"

WHAT IS DATA VIZ FOR?



1.

EXPLORATION

Table 1. Baseline Characteristics of the Patients.*

Characteristic	Randomization Groups 1 and 2		Randomization Groups 1 and 3	
	PFO Closure Group (N=238)	Antiplatelet-Only Group (N=235)	Anticoagulant Group (N=187)	Antiplatelet-Only Group (N=174)
Age — yr	42.9±10.1	43.8±10.5	43.8±9.5	44.7±10.5
Male sex — no. (%)	137 (57.6)	142 (60.4)	104 (55.6)	102 (58.6)
Medical history				
Hypertension — no. (%)	27 (11.3)	24 (10.2)	15 (8.0)	19 (10.9)
Diabetes mellitus — no. (%)	3 (1.3)	9 (3.8)	2 (1.1)	7 (4.0)
Current smoker — no. (%)	68 (28.6)	69 (29.4)	54 (28.9)	50 (28.7)
Hypercholesterolemia — no. (%)	30 (12.6)	36 (15.3)	22 (11.8)	25 (14.4)
Body mass index ≥30 — no. (%)†	32 (13.4)	27 (11.5)	20 (10.7)	24 (13.8)

Table 1. Baseline Characteristics of the Patients.*

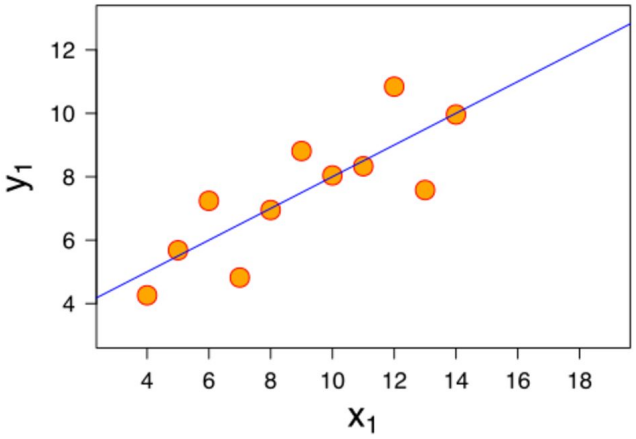
Characteristic	Randomization Groups 1 and 2		Randomization Groups 1 and 3	
	PFO Closure Group (N=238)	Antiplatelet-Only Group (N=235)	Anticoagulant Group (N=187)	Antiplatelet-Only Group (N=174)
Age — yr	42.9±10.1	43.8±10.5	43.8±9.5	44.7±10.5
Male sex — no. (%)	137 (57.6)	142 (60.4)	104 (55.6)	102 (58.6)
Medical history				
Hypertension — no. (%)	27 (11.3)	24 (10.2)	15 (8.0)	19 (10.9)
Diabetes mellitus — no. (%)	3 (1.3)	9 (3.8)	2 (1.1)	7 (4.0)
Current smoker — no. (%)	68 (28.6)	69 (29.4)	54 (28.9)	50 (28.7)
Hypercholesterolemia — no. (%)	30 (12.6)	36 (15.3)	22 (11.8)	25 (14.4)
Body mass index ≥30 — no. (%)†	32 (13.4)	27 (11.5)	20 (10.7)	24 (13.8)

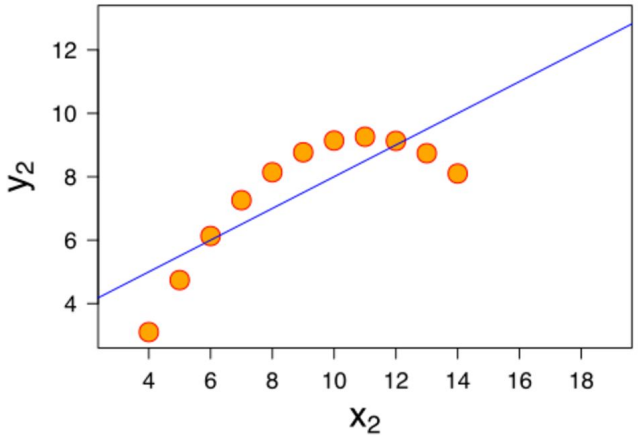
x: 9.0 ± 3.3

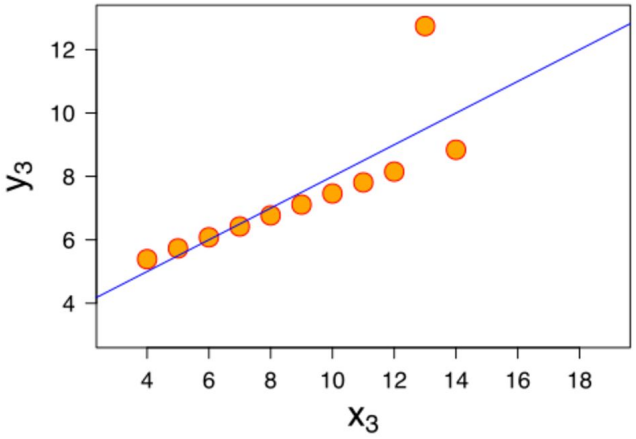
y: 7.5 ± 2.0

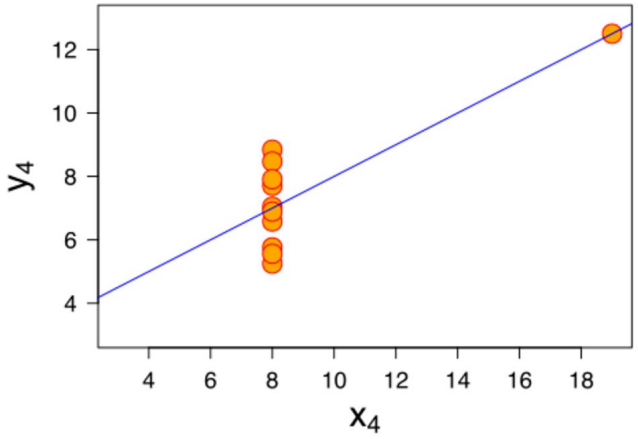
x: 9.0 ± 3.3

y: 7.5 ± 2.0

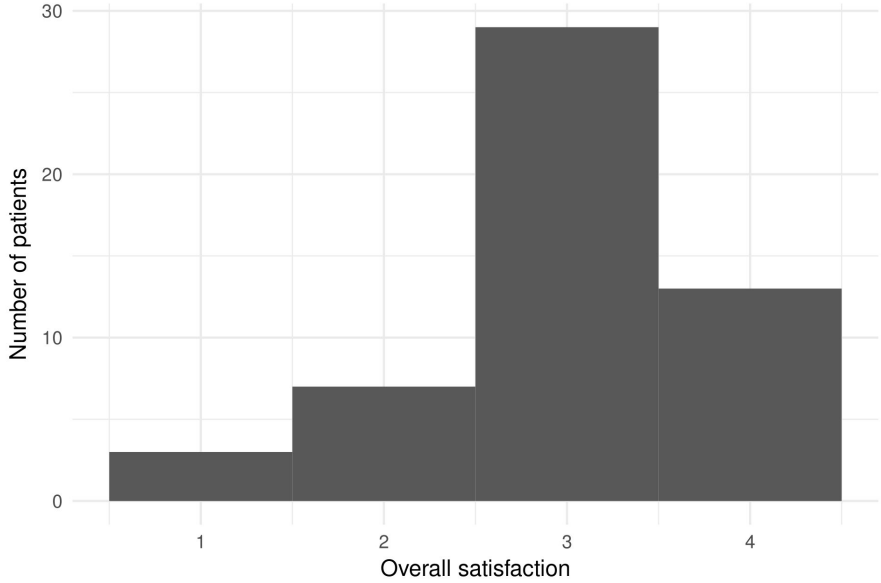

y1
x1
4
6
8
10
12
14
16
18


y2
x2
4
6
8
10
12
14
16
18

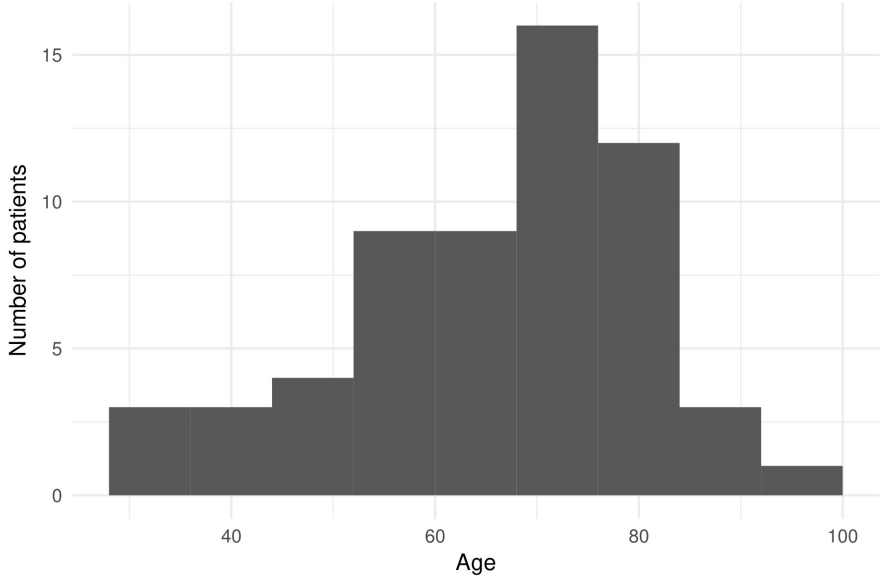

y3
x3
4
6
8
10
12
14
16
18


y4
x4
4
6
8
10
12
14
16
18

DISCRETE VARIABLE? TRY BAR CHARTS

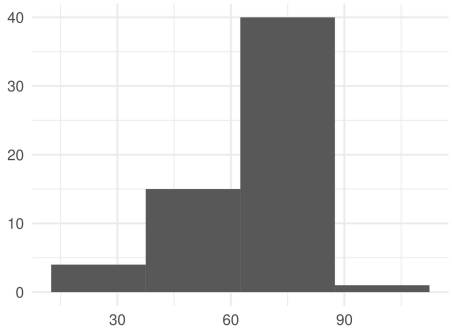


CONTINUOUS VARIABLE? TRY HISTOGRAMS

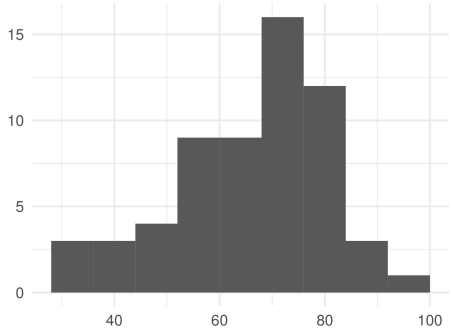


Histograms: How many bins?

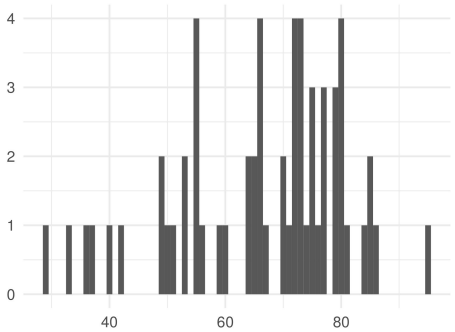
Too few...



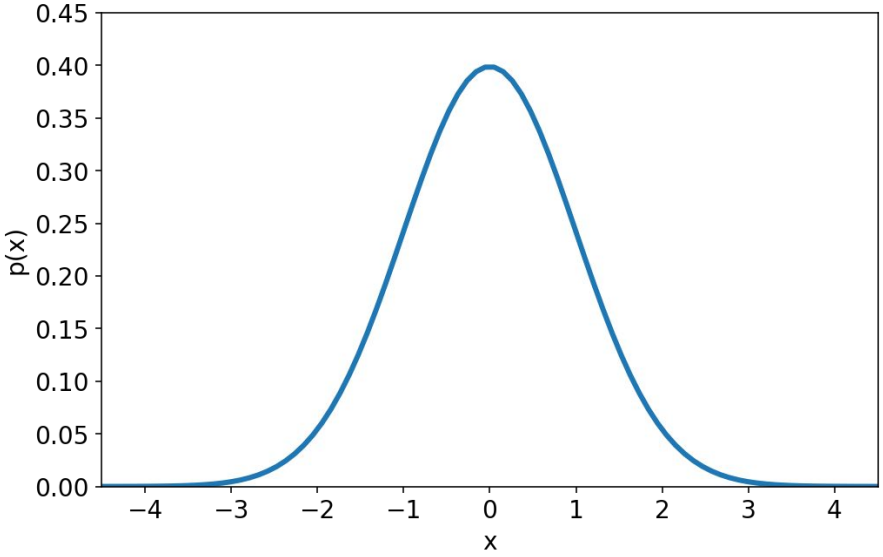
...just right...



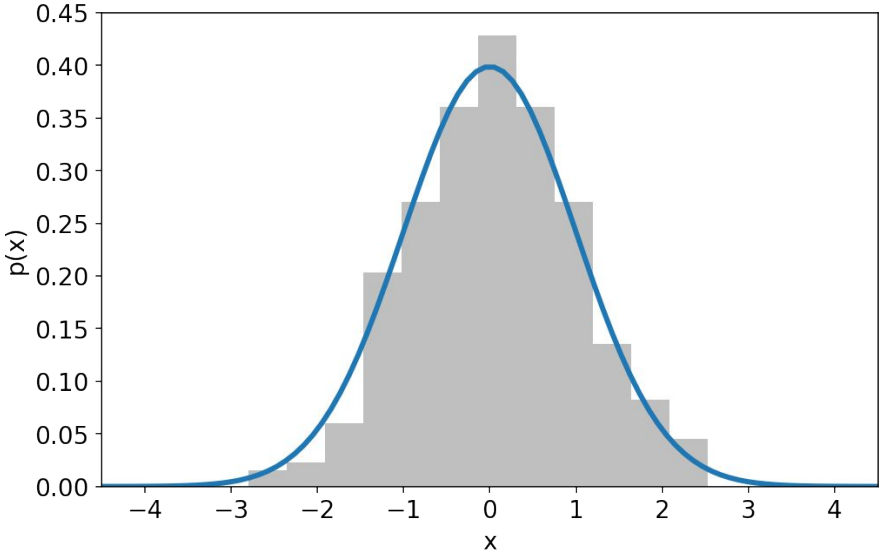
...too many



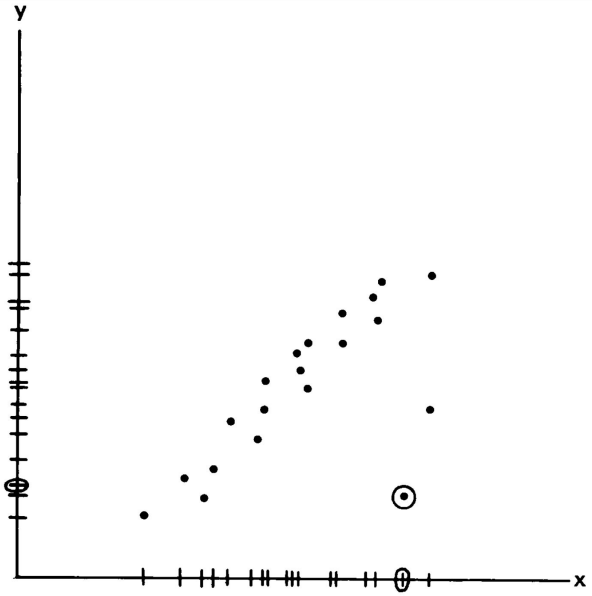
HISTOGRAMS: APPROXIMATING DISTRIBUTIONS



HISTOGRAMS: APPROXIMATING DISTRIBUTIONS

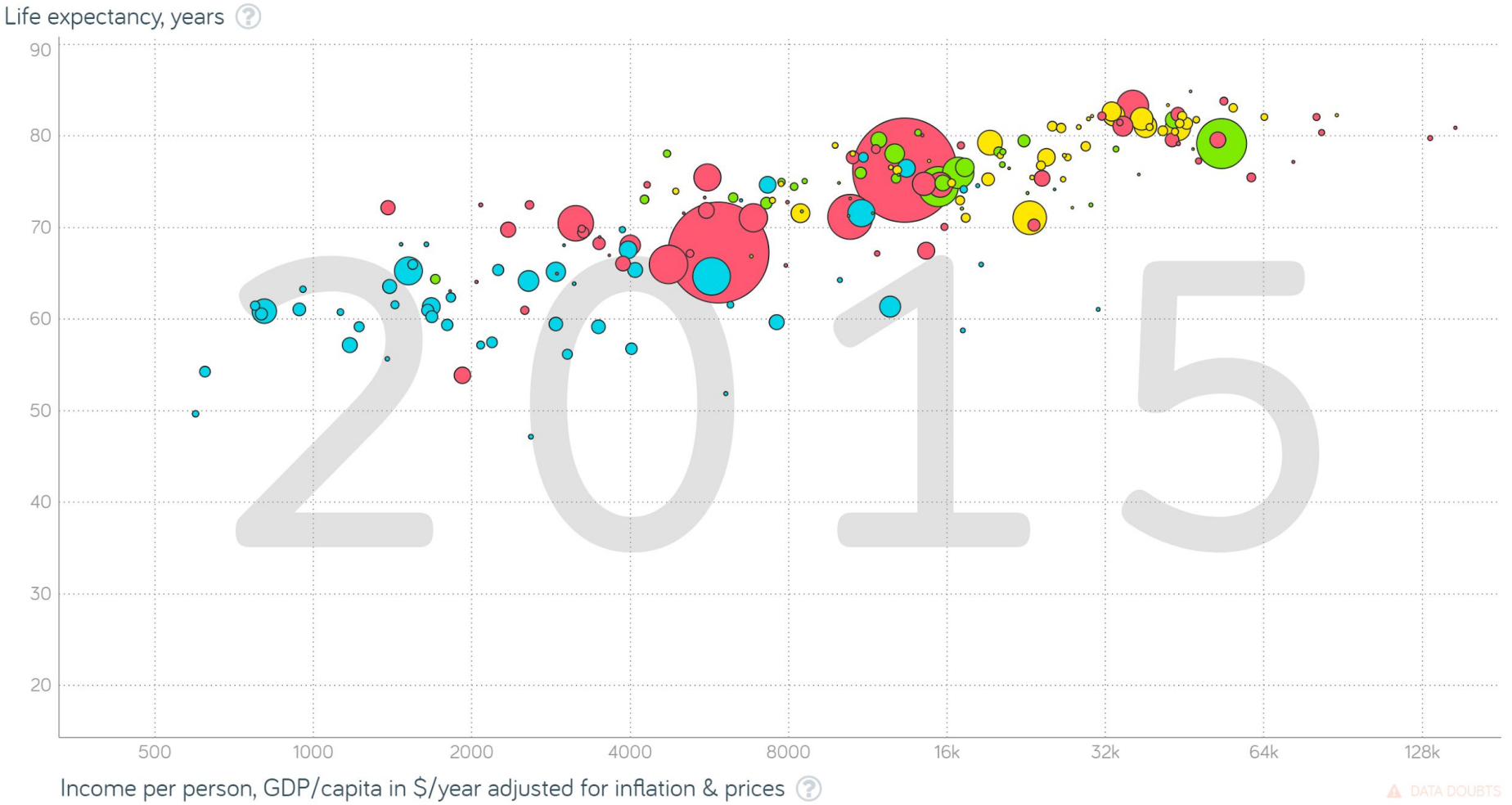


SCATTERPLOTS CAN BE REVEALING!



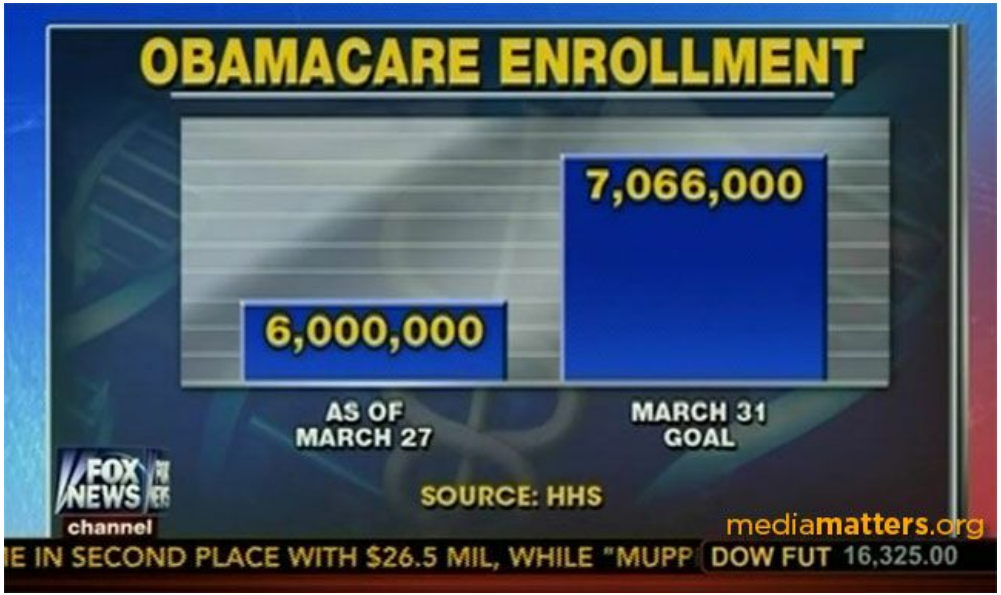
2.

Storytelling (and some words of warning)

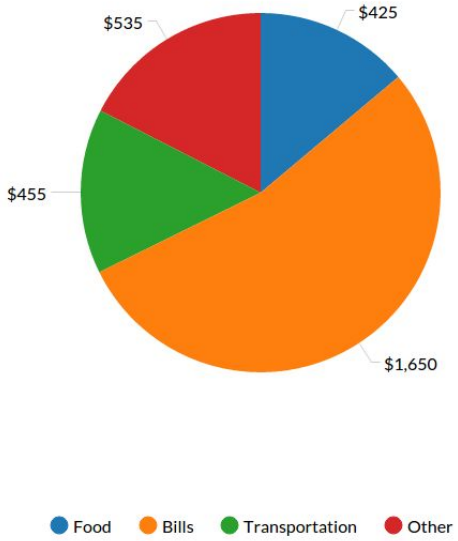

Life expectancy, years
90
80
70
60
50
40
30
20
2015
500
1000
2000
4000
8000
16k
32k
64k
128k
Income per person, GDP/capita in $/year adjusted for inflation & prices
DATA DOUBTS

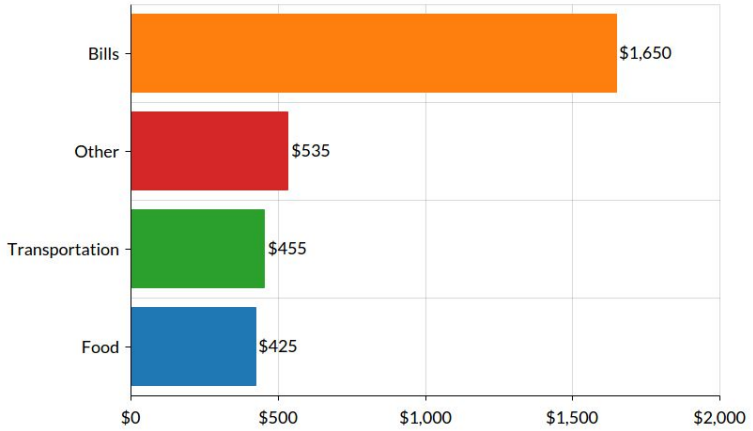
SOME TIPS FROM EDWARD TUSKE...

DON'T MISREPRESENT WHAT THE DATA SAY

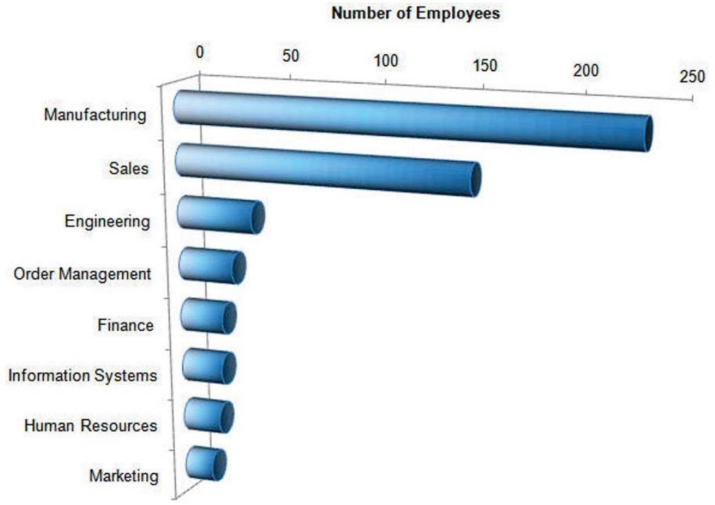


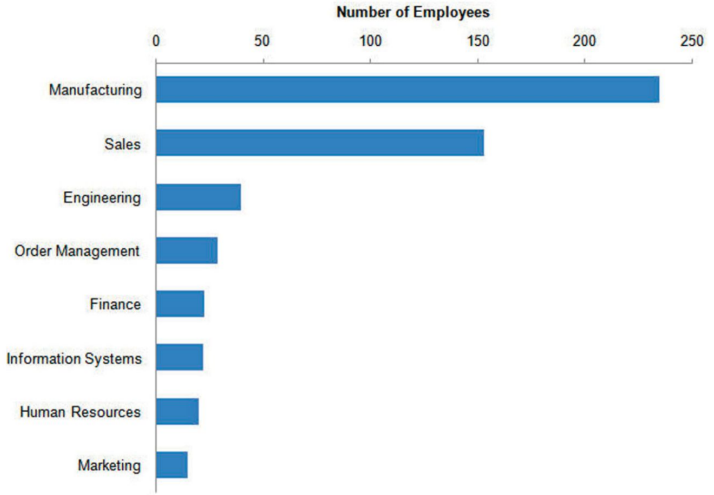
PIE CHARTS: DON'T BOTHER





DON'T USE 3D WHERE 2D WILL DO

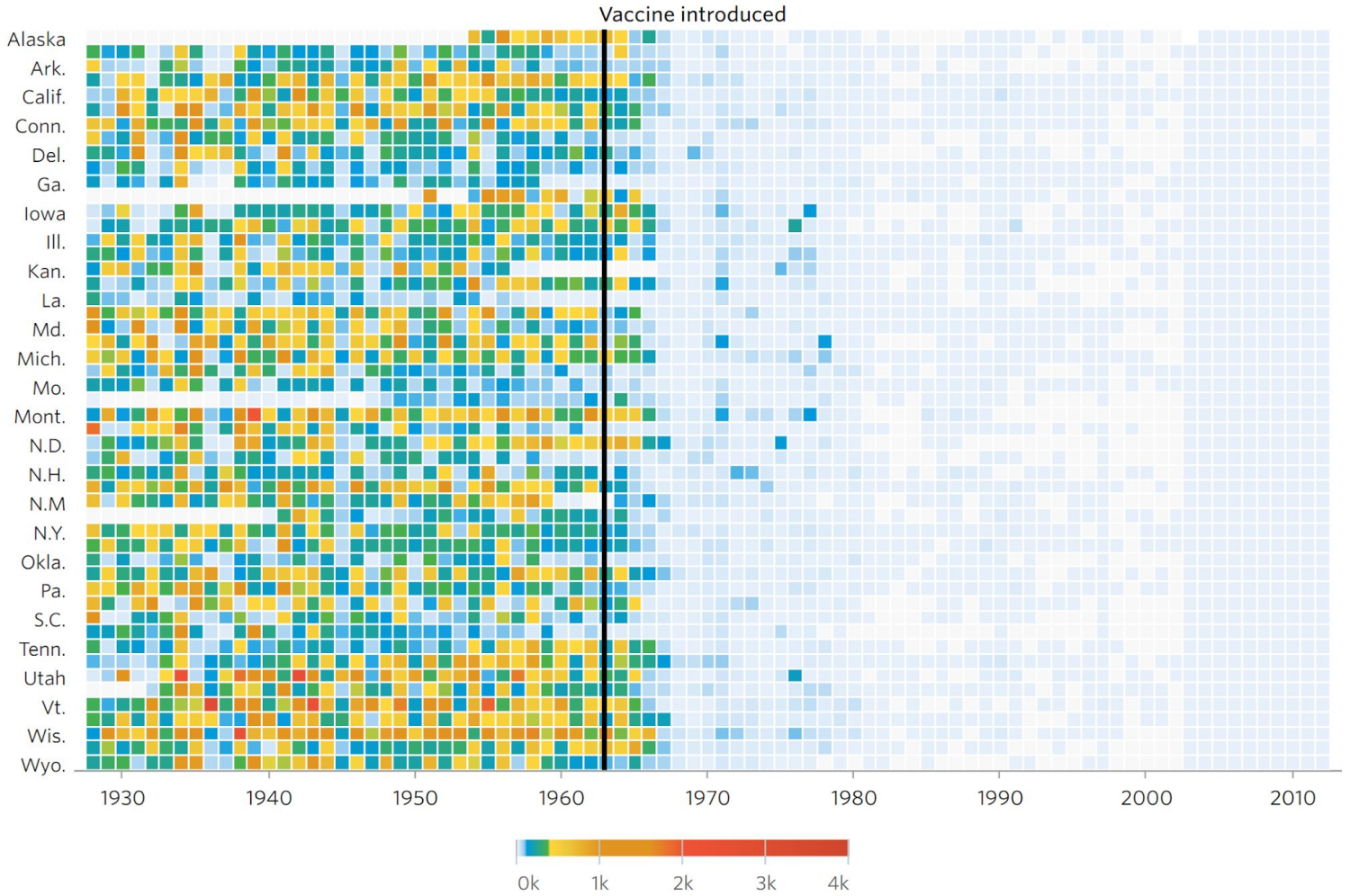




Remove

to improve

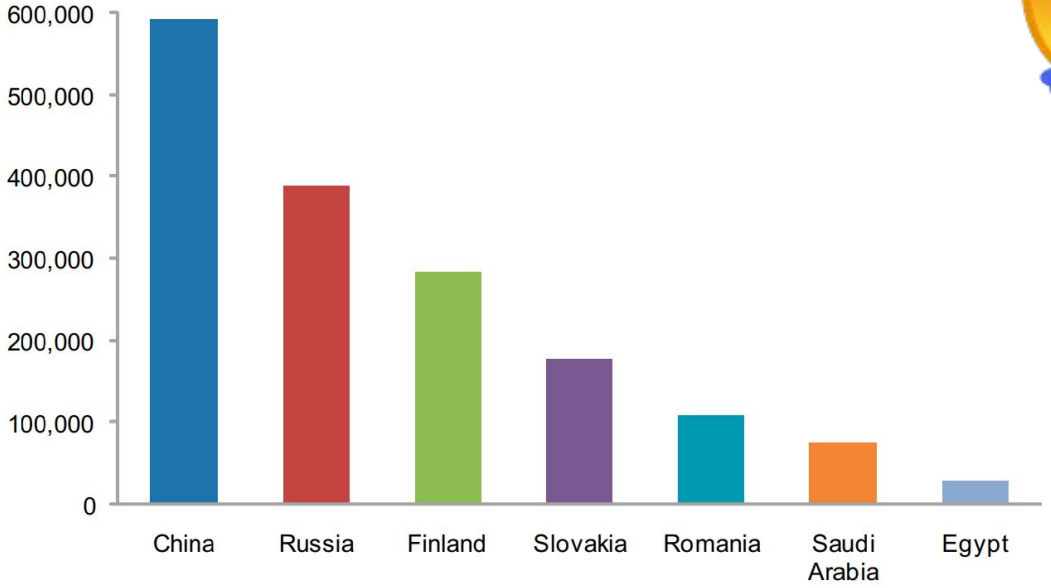
(the **data-ink** ratio)


Vaccine introduced
Alaska
Ark.
Calif.
Conn.
Del.
Ga.
Iowa
Ill.
Kan.
La.
Md.
Mich.
Mo.
Mont.
N.D.
N.H.
N.M
N.Y.
Okla.
Pa.
S.C.
Tenn.
Utah
Vt.
Wis.
Wyo.
1930
1940
1950
1960
1970
1980
1990
2000
2010
0k
1k
2k
3k
4k

3.

USING COLOUR

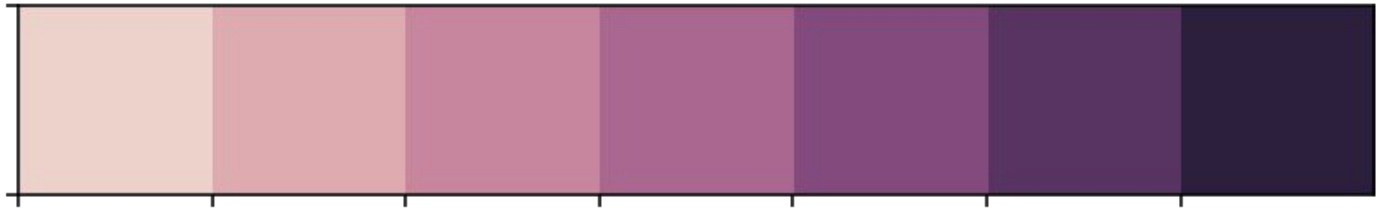
USE COLOUR MEANINGFULLY!



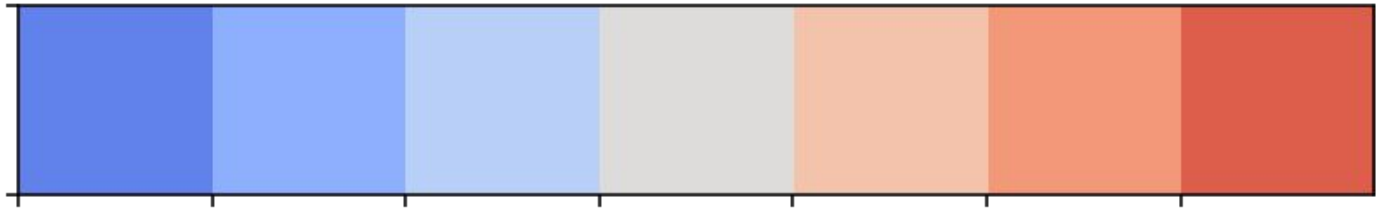
Qualitative palettes: a few categories

Qualitative palettes: more categories

Sequential palettes: emphasise high values



Diverging palettes: emphasise lows and highs



And what about...

Perceived intensity

- Yellows/greens look light
- Blues look dark

Colour blindness

- 8% of men
- 0.5% of women

AND WHAT ABOUT...

Perceived intensity

- Yellows/greens look light
- Blues look dark

Colour blindness

- 8% of men
- 0.5% of women

COLORBREWER 2.0
color advice for cartography

THANKS!

Any questions?