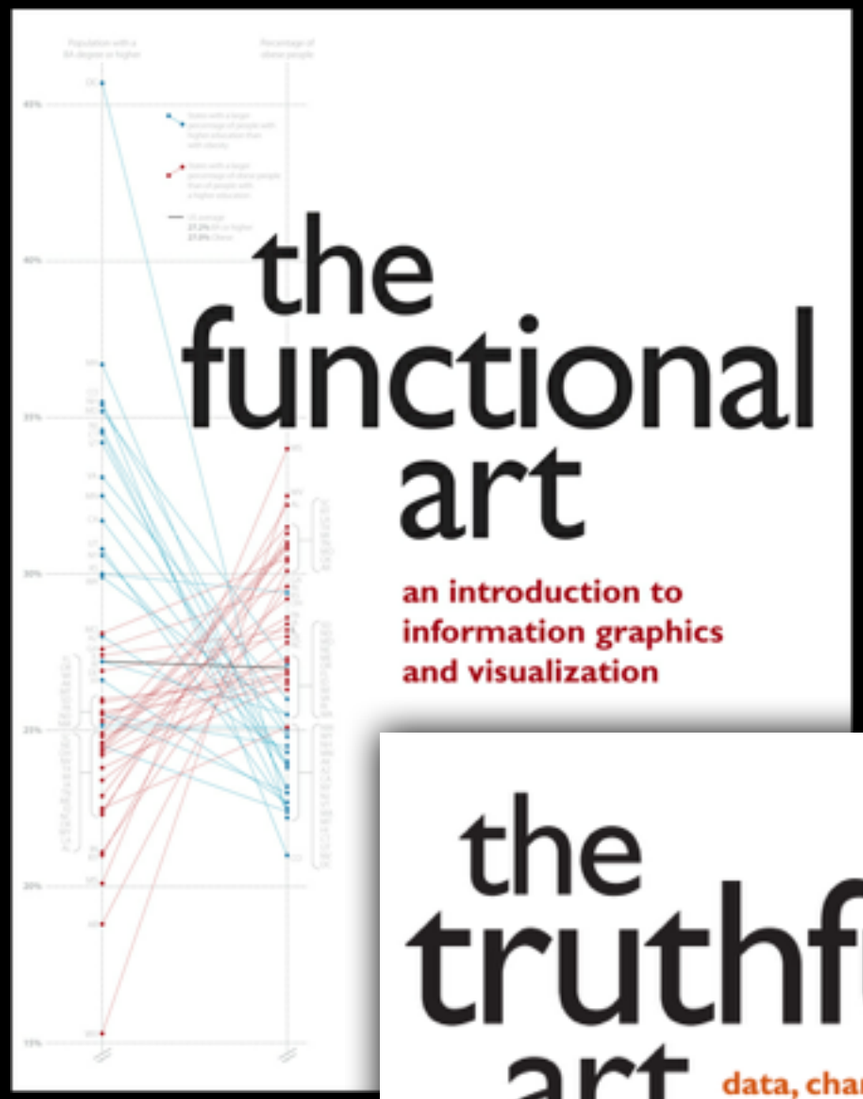
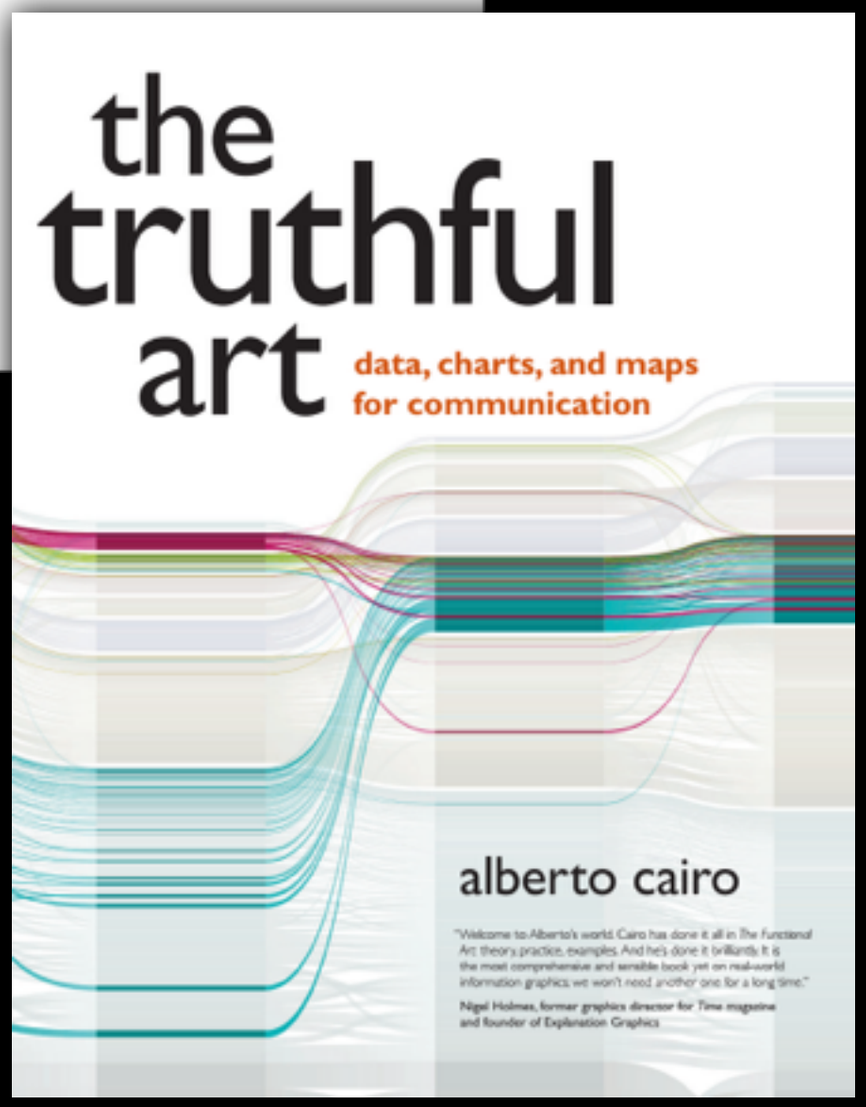
Visualization for exploration



2012

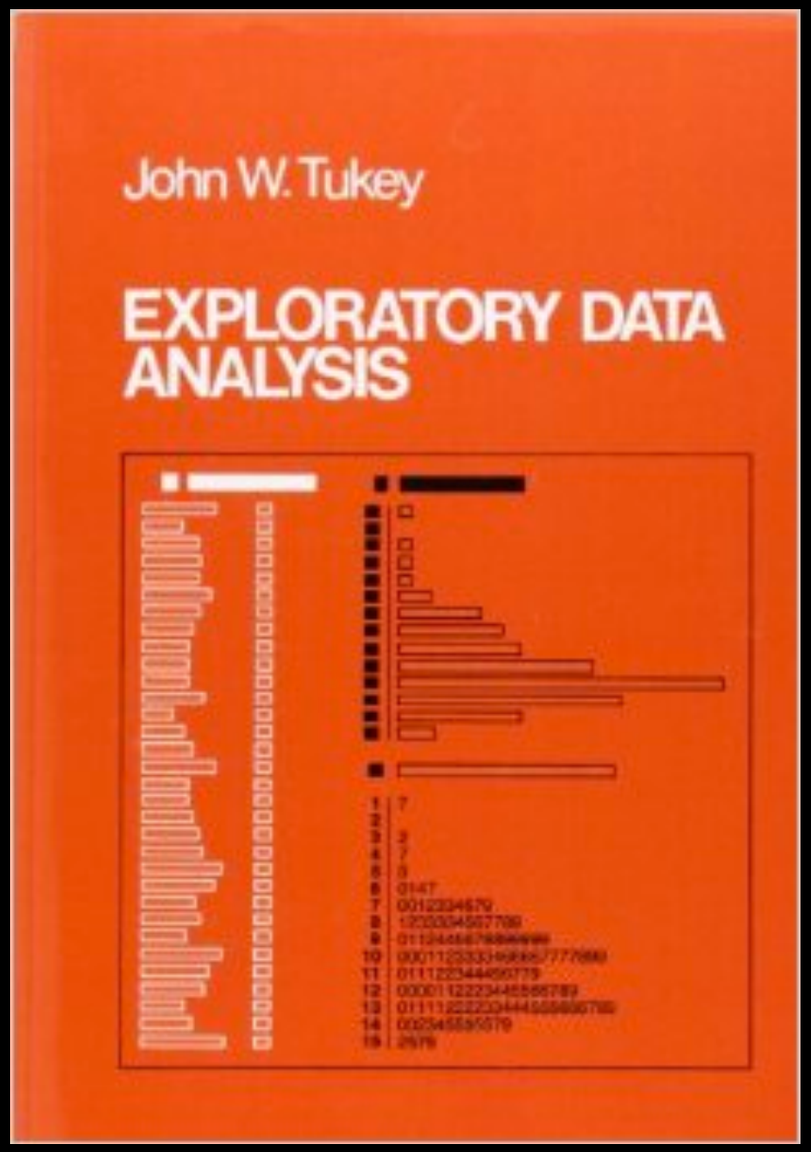


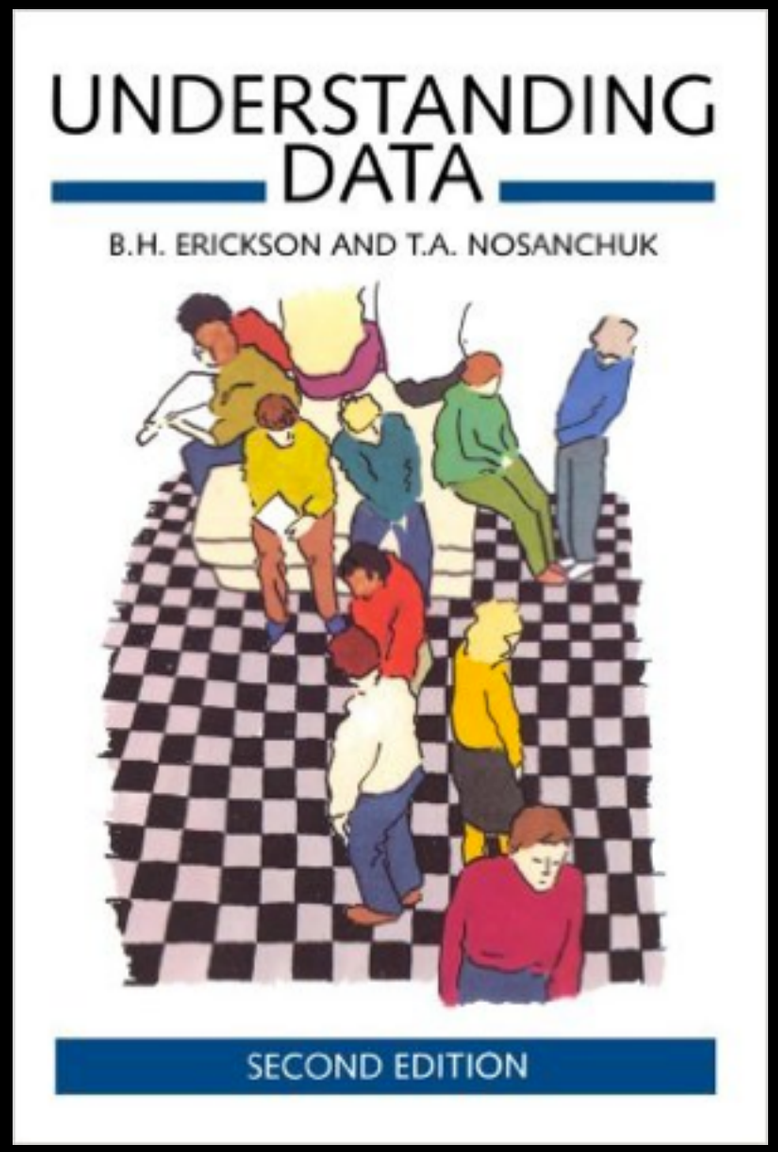
2016

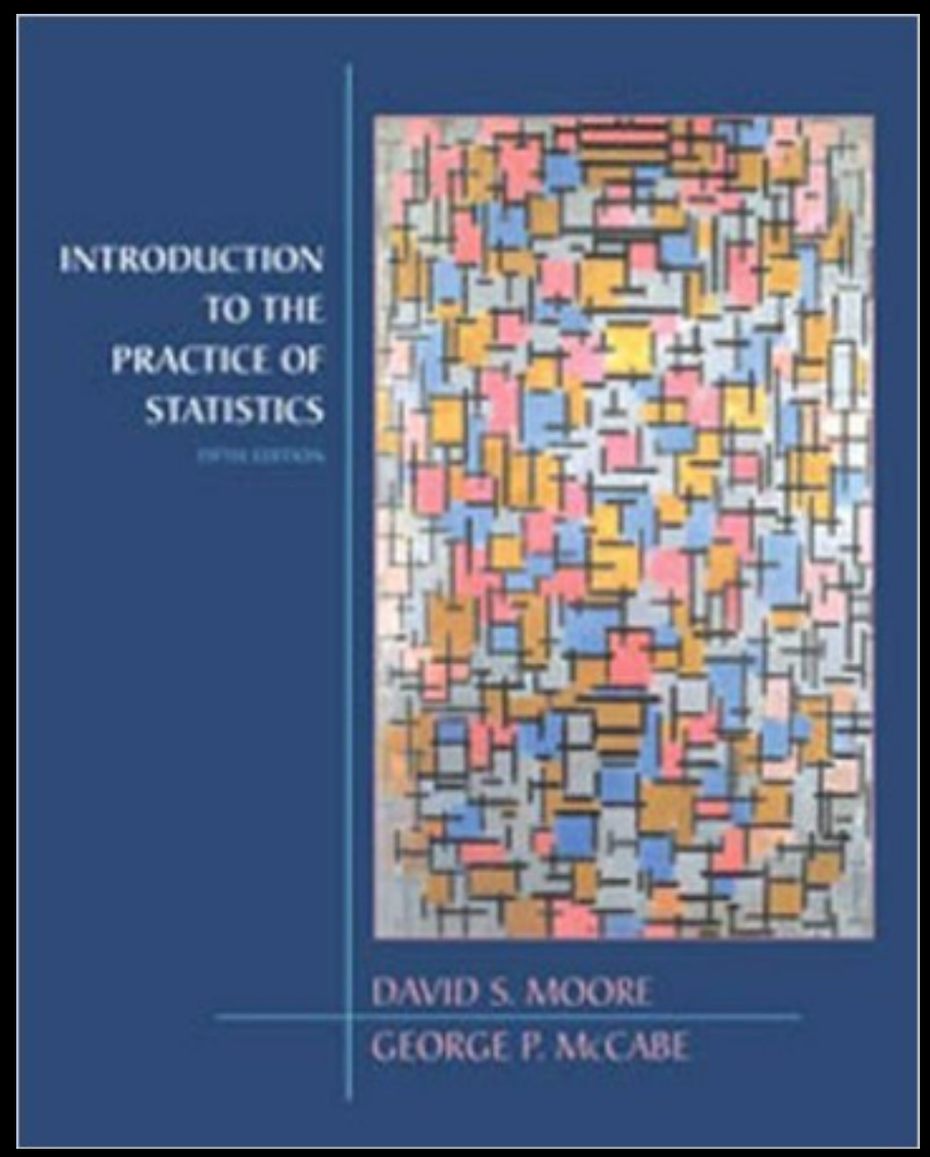
A visualization is a graphical representation designed to enable exploration, analysis, or communication

Alberto Cairo • University of Miami • www.thefunctionalart.com • Twitter: @albertocairo

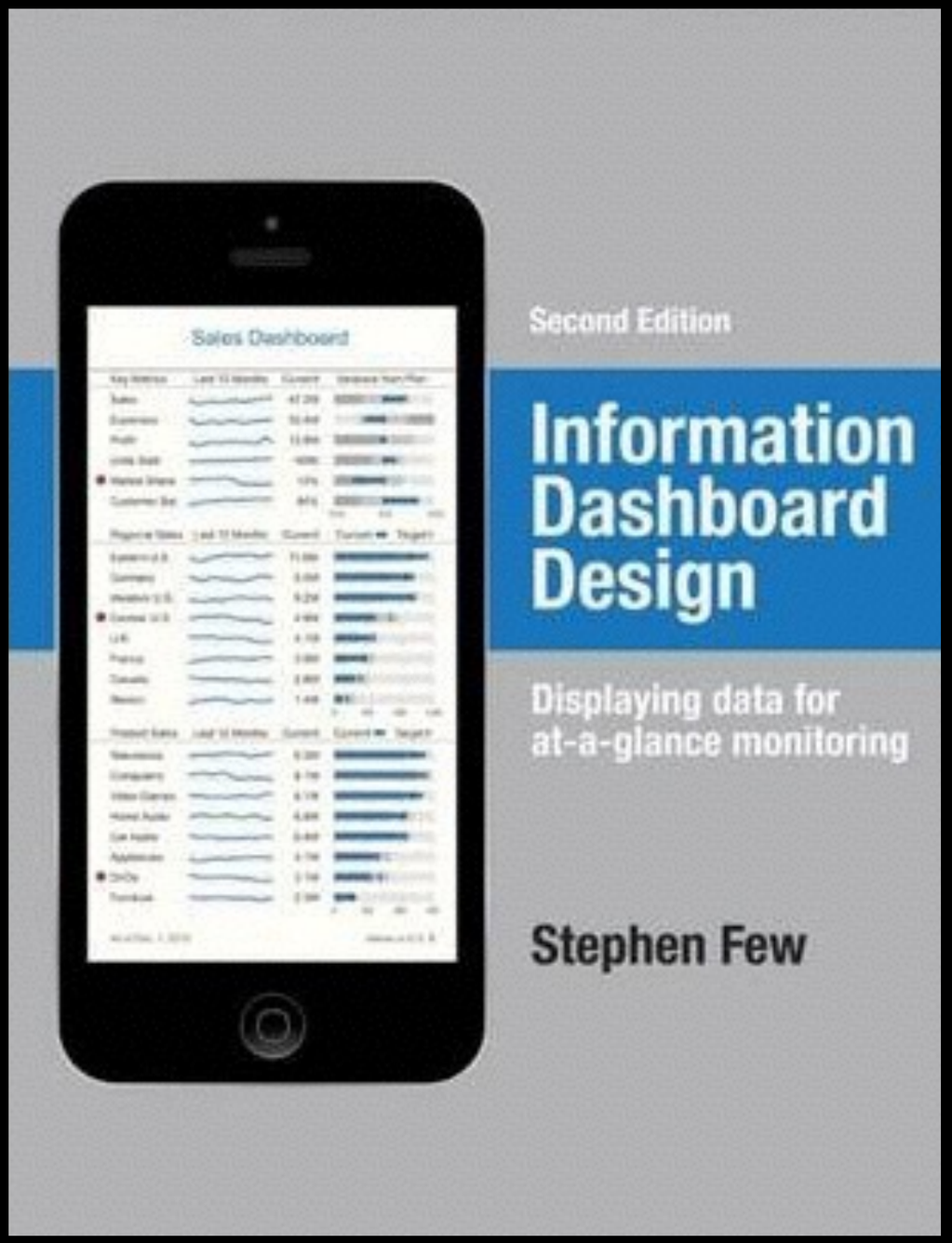
A specific branch of visualization, called data visualization, was originally created for exploration and analysis

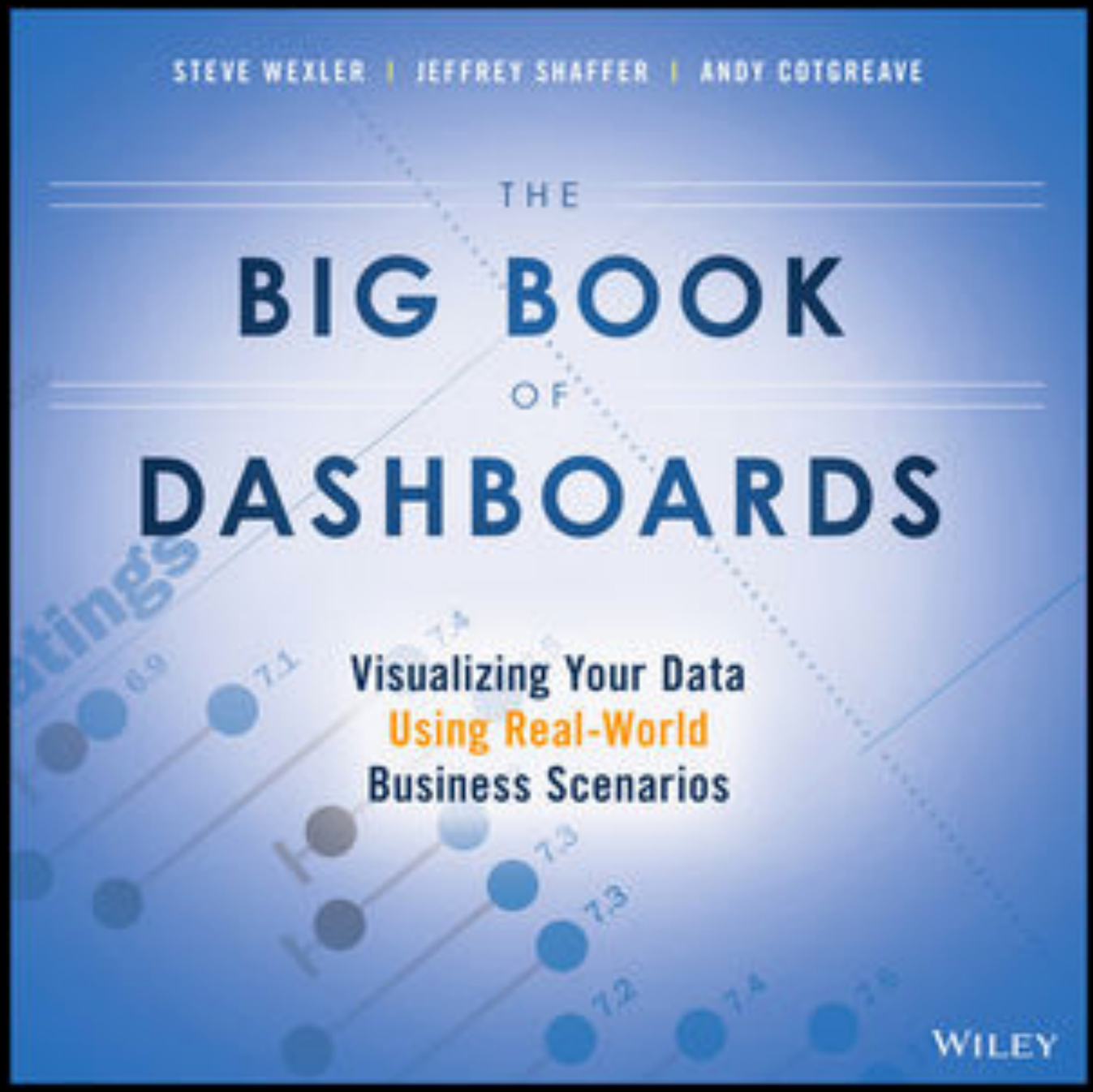




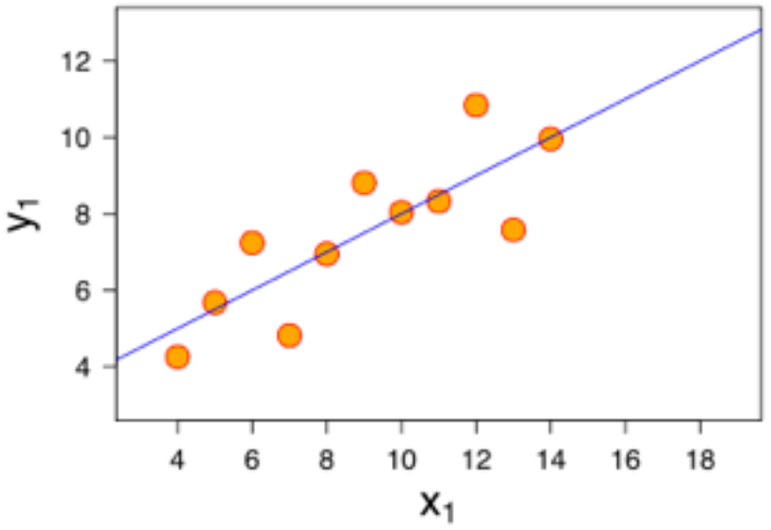


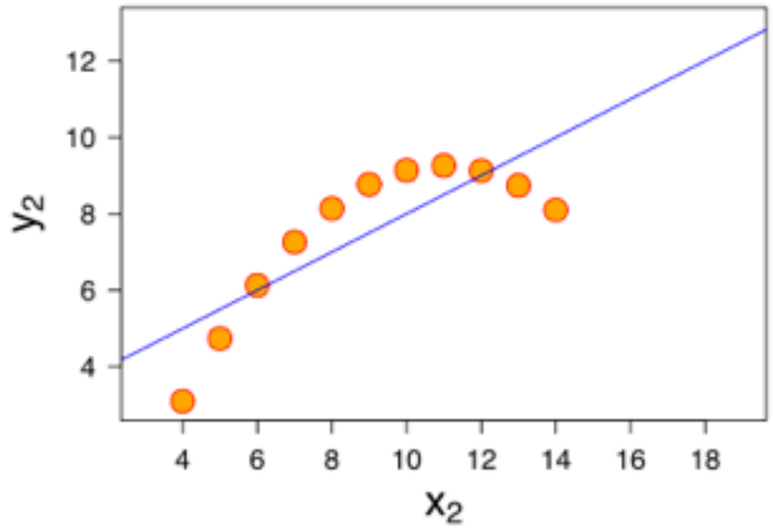
A specific branch of visualization, called data visualization, was originally created for exploration and analysis

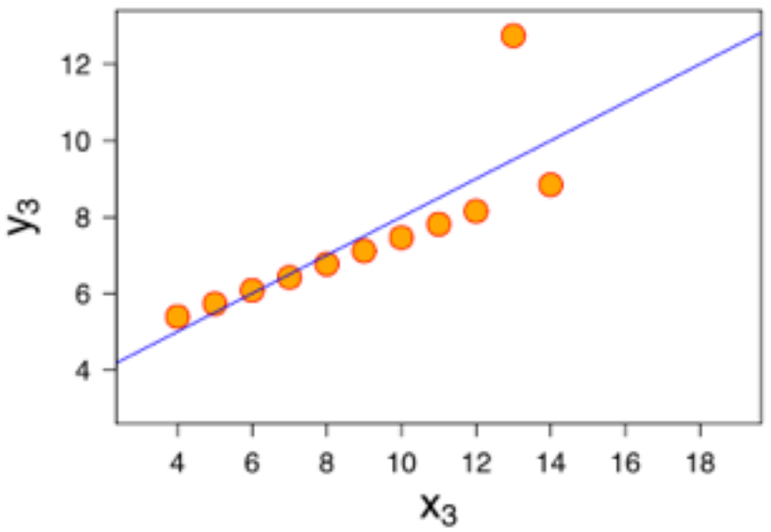


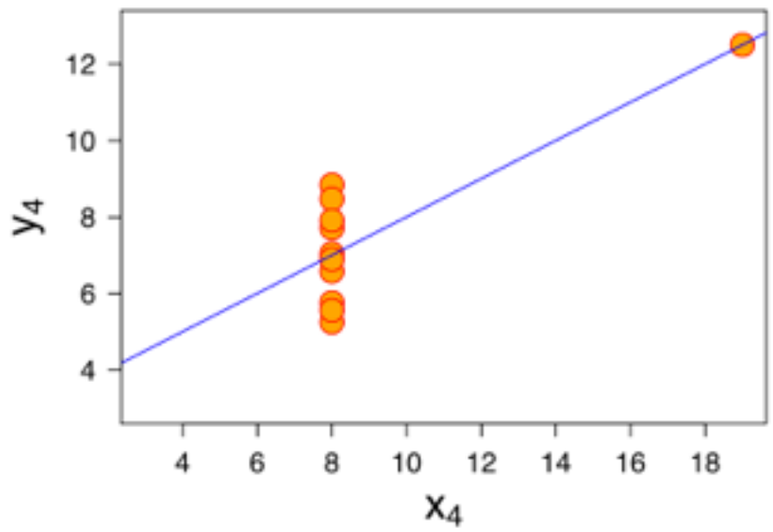


http://www.wiley.com/WileyCDA/WileyTitle/productCd-1119282713.html









Francis Anscombe's Quartet

https://en.wikipedia.org/wiki/Anscombe%27s_quartet

For all four datasets:

Property	Value
Mean of *x* in each case	9 (exact)
Sample variance of *x* in each case	11 (exact)
Mean of *y* in each case	7.50 (to 2 decimal places)
Sample variance of *y* in each case	4.122 or 4.127 (to 3 decimal places)
Correlation between *x* and *y* in each case	0.816 (to 3 decimal places)
Linear regression line in each case	$y = 3.00 + 0.500x$ (to 2 and 3 decimal places, respectively)

http://robertgrantstats.co.uk/drawmydata.html

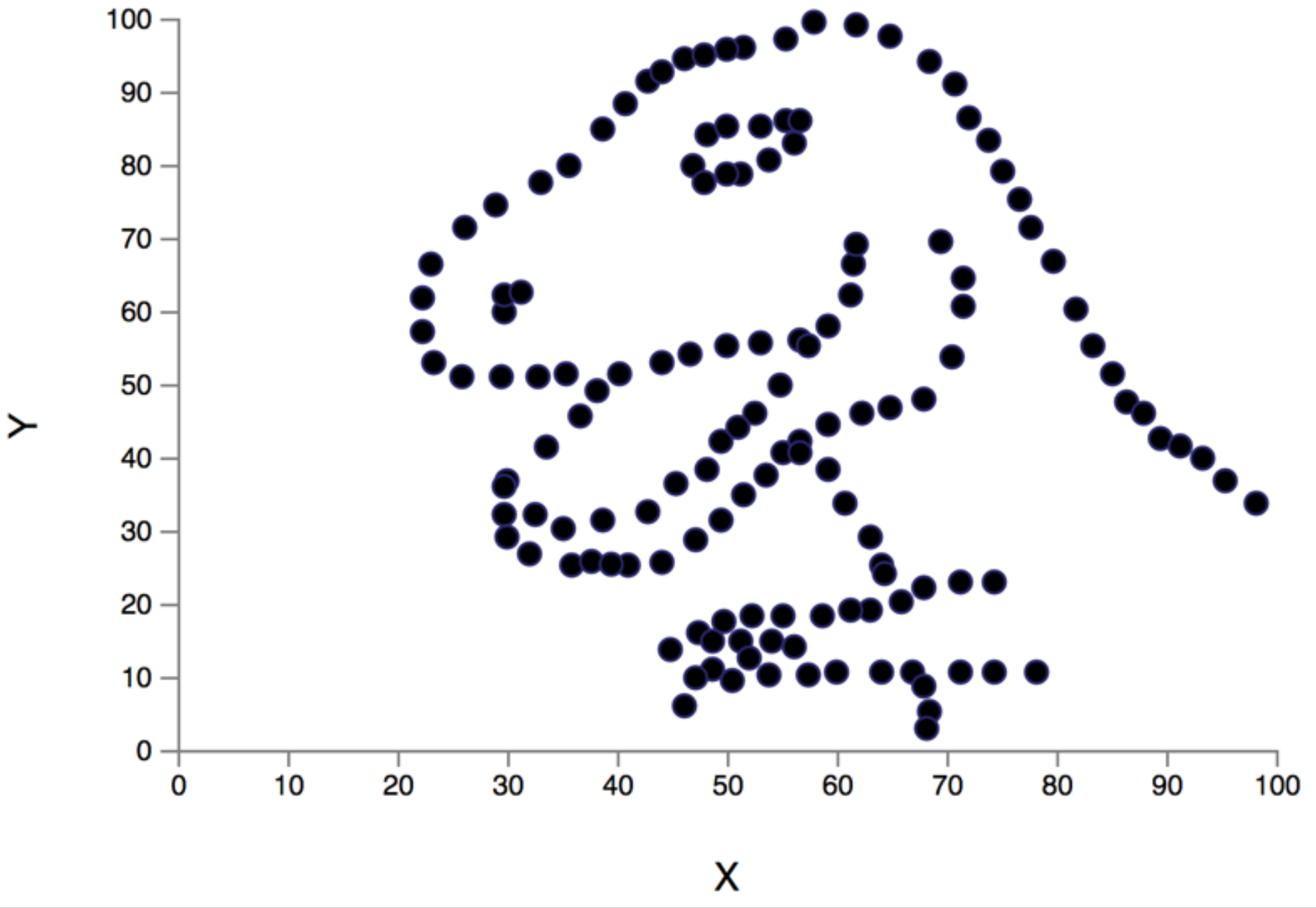
N = 142 ; X mean = 54.2633 ; X SD = 16.7651 ; Y mean = 47.8323 ; Y SD = 26.9354 ; Pearson correlation = -0.0645

55.3846	97.1795
51.5385	96.0256
46.1538	94.4872
42.8205	91.4103
40.7692	88.3333
38.7179	84.8718
35.641	79.8718
33.0769	77.5641
28.9744	74.4872
26.1538	71.4103
23.0769	66.4103
22.3077	61.7949
22.3077	57.1795
23.3333	52.9487
25.8974	51.0256
29.4872	51.0256
32.8205	51.0256
35.3846	51.4103
40.2564	51.4103

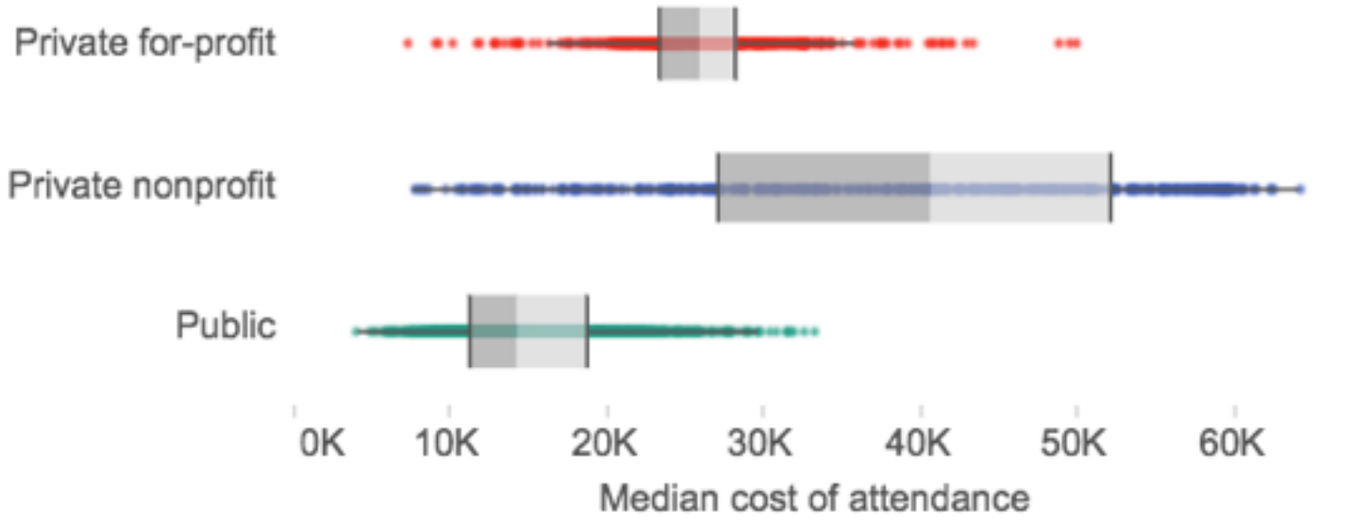
http://robertgrantstats.co.uk/drawmydata.html

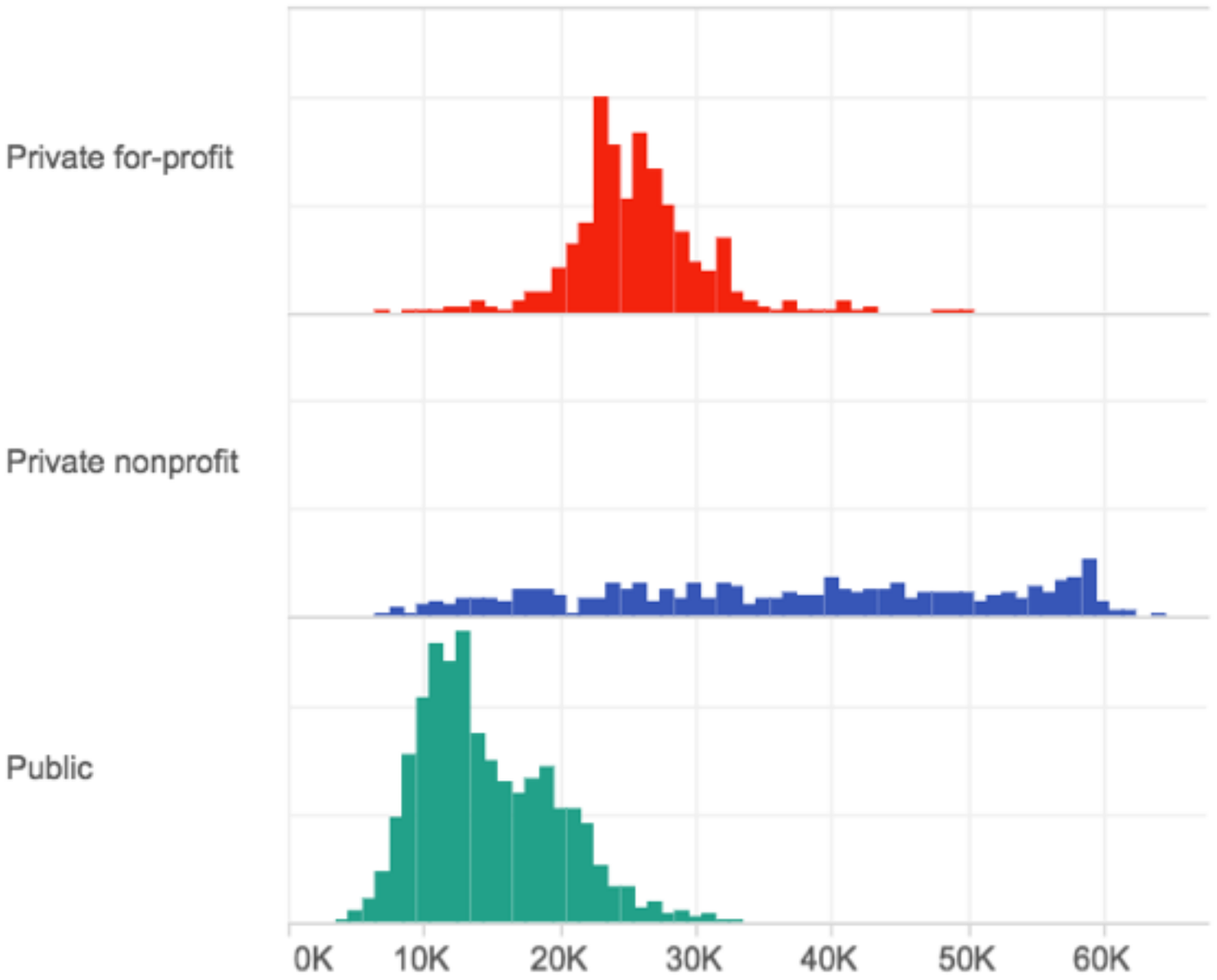
N = 142 ; X mean = 54.2633 ; X SD = 16.7651 ; Y mean = 47.8323 ; Y SD = 26.9354 ; Pearson correlation = -0.0645

55.3846	97.1795
51.5385	96.0256
46.1538	94.4872
42.8205	91.4103
40.7692	88.3333
38.7179	84.8718
35.641	79.8718
33.0769	77.5641
28.9744	74.4872
26.1538	71.4103
23.0769	66.4103
22.3077	61.7949
22.3077	57.1795
23.3333	52.9487
25.8974	51.0256
29.4872	51.0256
32.8205	51.0256
35.3846	51.4103
40.2564	51.4103

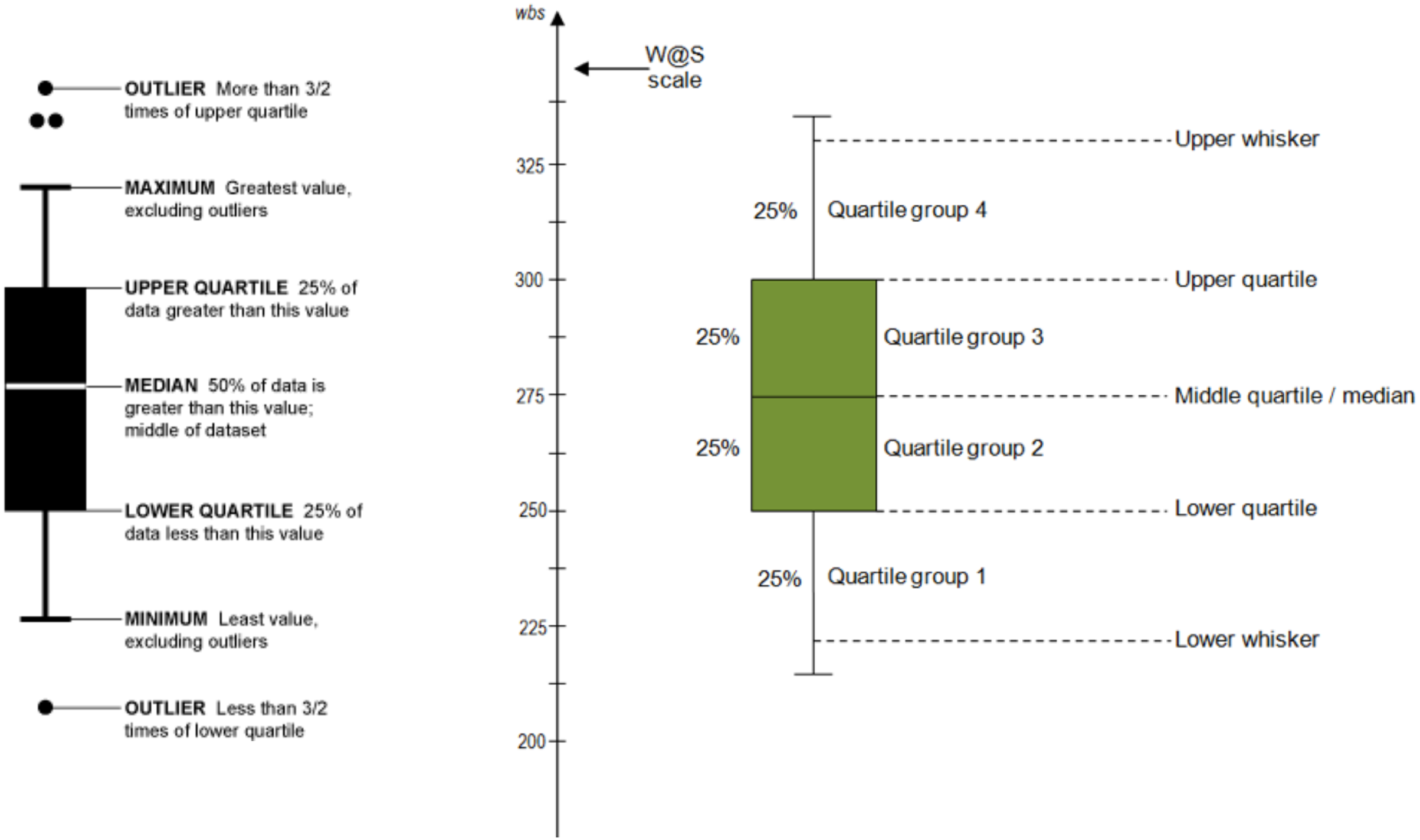


High-level cost





Box plot


OUTLIER More than 3/2 times of upper quartile
MAXIMUM Greatest value, excluding outliers
UPPER QUARTILE 25% of data greater than this value
MEDIAN 50% of data is greater than this value; middle of dataset
LOWER QUARTILE 25% of data less than this value
MINIMUM Least value, excluding outliers
OUTLIER Less than 3/2 times of lower quartile
wbs
W@S scale
325
300
275
250
225
200
Upper whisker
25%
Quartile group 4
Upper quartile
25%
Quartile group 3
Middle quartile / median
25%
Quartile group 2
Lower quartile
25%
Quartile group 1
Lower whisker