Data visualization
W4-2

Quiz

- What do you find interesting in today's VotW?
- What are the RGB and CMYK system and how are they different from each other?
- Discuss various challenges when using colors in visualization.

Design

Pre-attentive processing

Pre-

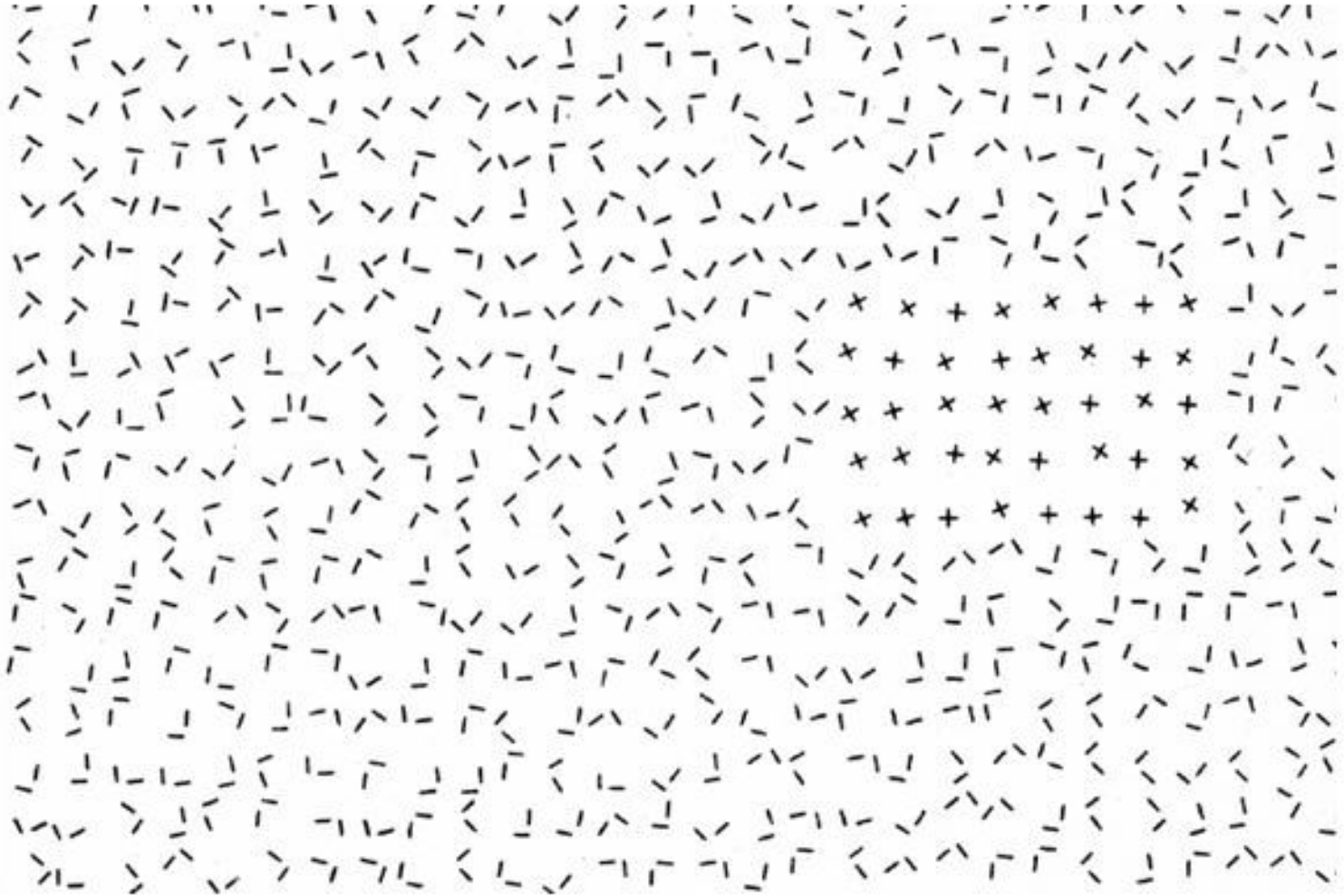
Pre-attentive

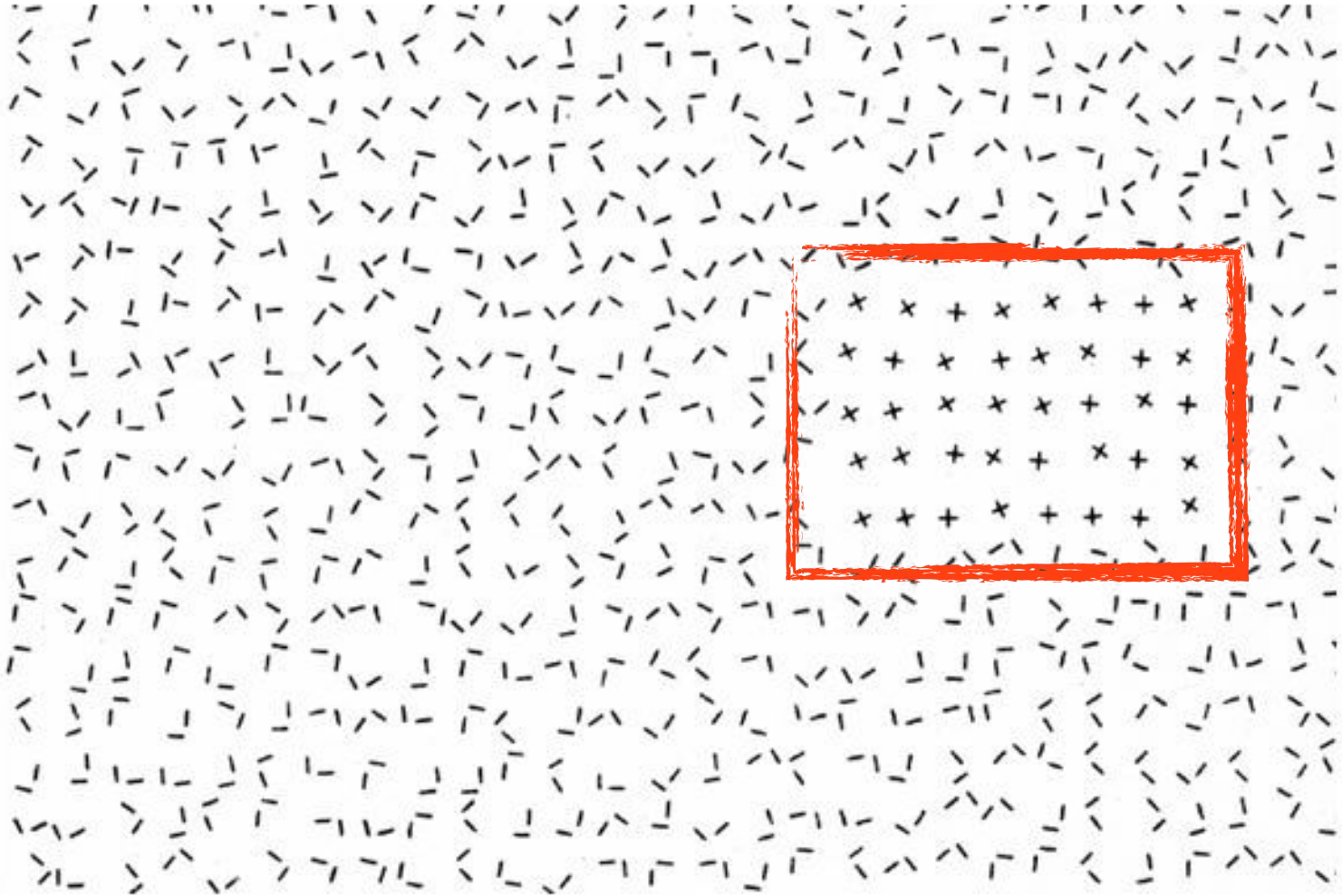
How many 5s?

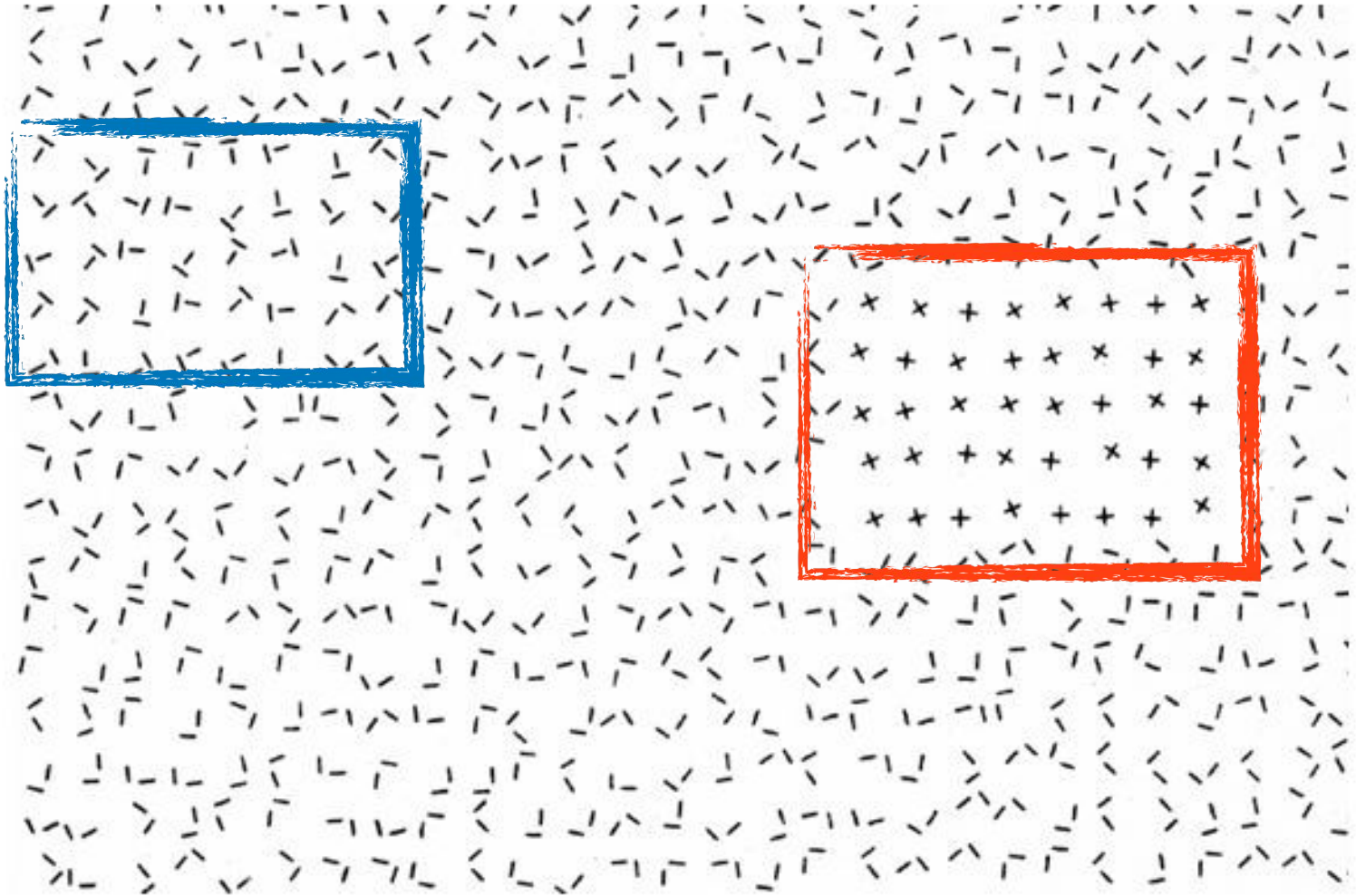
1925687192731635816231529572305192639127017496192361027013750693416294710370126391615

How many 5s?

1925687192731635816231
2957230519263912701749
6
1923610270137506934162
9
4710370126391615



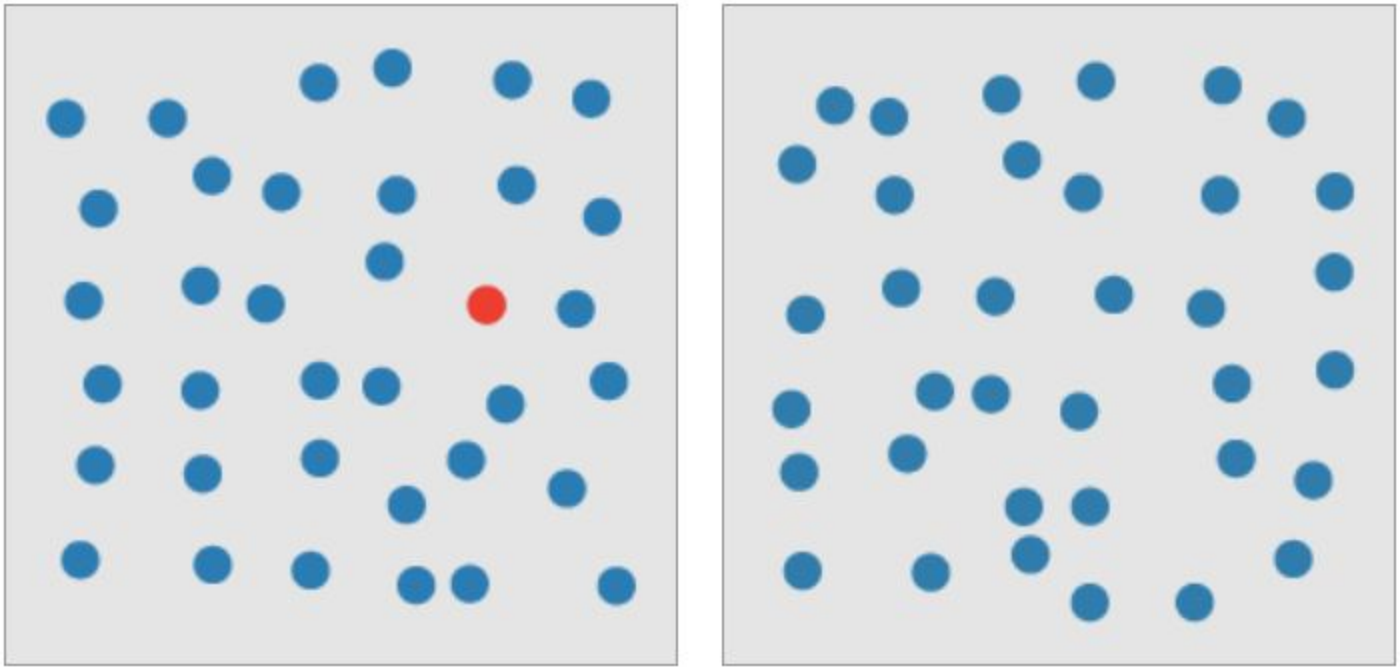




Why should we care?

Clever visual encodings
lead to
Easier-to-read &
powerful visualizations

There are visual encodings that STAND OUT




Schindler’s List

The death toll in the Gaza-Israel conflict

By Lazaro Gamio and Richard Johnson, Updated: Aug 7, 2014

Statistics on death tolls during active conflict are often difficult to track accurately. A note on our sourcing:

For death toll numbers from within Gaza, we receive daily or sometimes twice daily updates from the head of the Information Management Unit in the United Nations Office for the Coordination of Humanitarian Affairs in the Occupied Palestinian Territories field office in the West Bank. The field office gets updates from multiple sources within Gaza, including U.N. agencies and the Palestinian Medical Authorities. They then create and continually update a spreadsheet of all available data. These numbers are often not complete, but represent the best available data and do tend to clarify over time. Israel disputes the numbers provided by the United Nations, saying that a large number of those killed, particularly males over 18, were armed terrorists and not civilians.

For death toll numbers of Israeli soldiers, we rely on Israel Defense Forces press releases and their Twitter account. For Israeli civilian deaths, we rely on news reports and our reporters in Israel and Gaza. These numbers are similarly tentative initially but clarify over time.

Related articles: Reporters grapple with politics, erratic sources in reporting Israeli/Gaza death toll (Washington Post) | Caution needed with Gaza casualty figures (BBC) | Civilian or Not? New Fight in Tallying the Dead From the Gaza Conflict (New York Times)

1,958 Deaths up to **Aug. 6**

65 were Israeli soldiers.

3 were Israeli civilians.

217 were armed Palestinian militants. Of those, **2** were children.

1396 were Palestinian civilians. Of those, **222** were women and **418** were children.

Red: Children

The Washington Post

The death toll in the Gaza-Israel conflict

By Lazaro Gamio and Richard Johnson, Updated: Aug 7, 2014

Statistics on death tolls during active conflict are often difficult to track accurately. A note on our sourcing:

For death toll numbers from within Gaza, we receive daily or sometimes twice daily updates from the head of the Information Management Unit in the United Nations Office for the Coordination of Humanitarian Affairs in the Occupied Palestinian Territories field office in the West Bank. The field office gets updates from multiple sources within Gaza, including U.N. agencies and the Palestinian Medical Authorities. They then create and continually update a spreadsheet of all available data. These numbers are often not complete, but represent the best available data and do tend to clarify over time. Israel disputes the numbers provided by the United Nations, saying that a large number of those killed, particularly males over 18, were armed terrorists and not civilians.

For death toll numbers of Israeli soldiers, we rely on Israel Defense Forces press releases and their Twitter account. For Israeli civilian deaths, we rely on news reports and our reporters in Israel and Gaza. These numbers are similarly tentative initially but clarify over time.

Related articles: Reporters grapple with politics, erratic sources in reporting Israeli/Gaza death toll (Washington Post) | Caution needed with Gaza casualty figures (BBC) | Civilian or Not? New Fight in Tallying the Dead From the Gaza Conflict (New York Times)

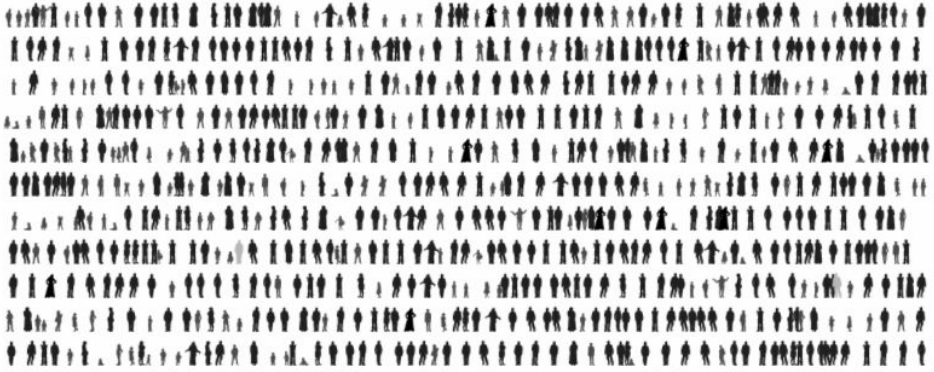
1,958 Deaths up to **Aug. 6**

65 were Israeli soldiers.

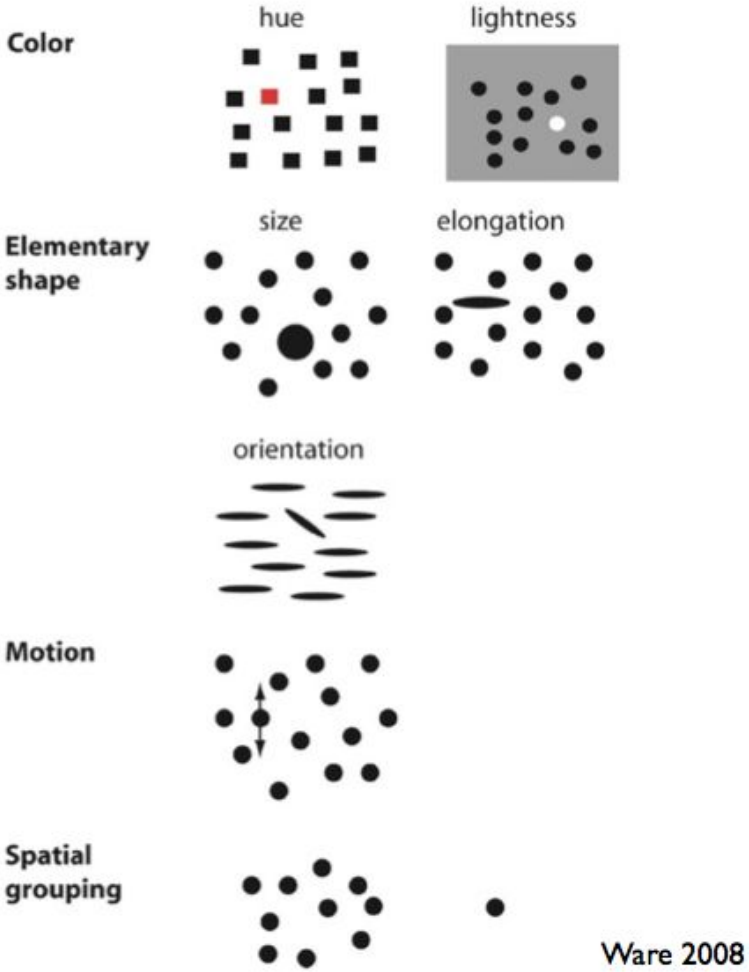
3 were Israeli civilians.

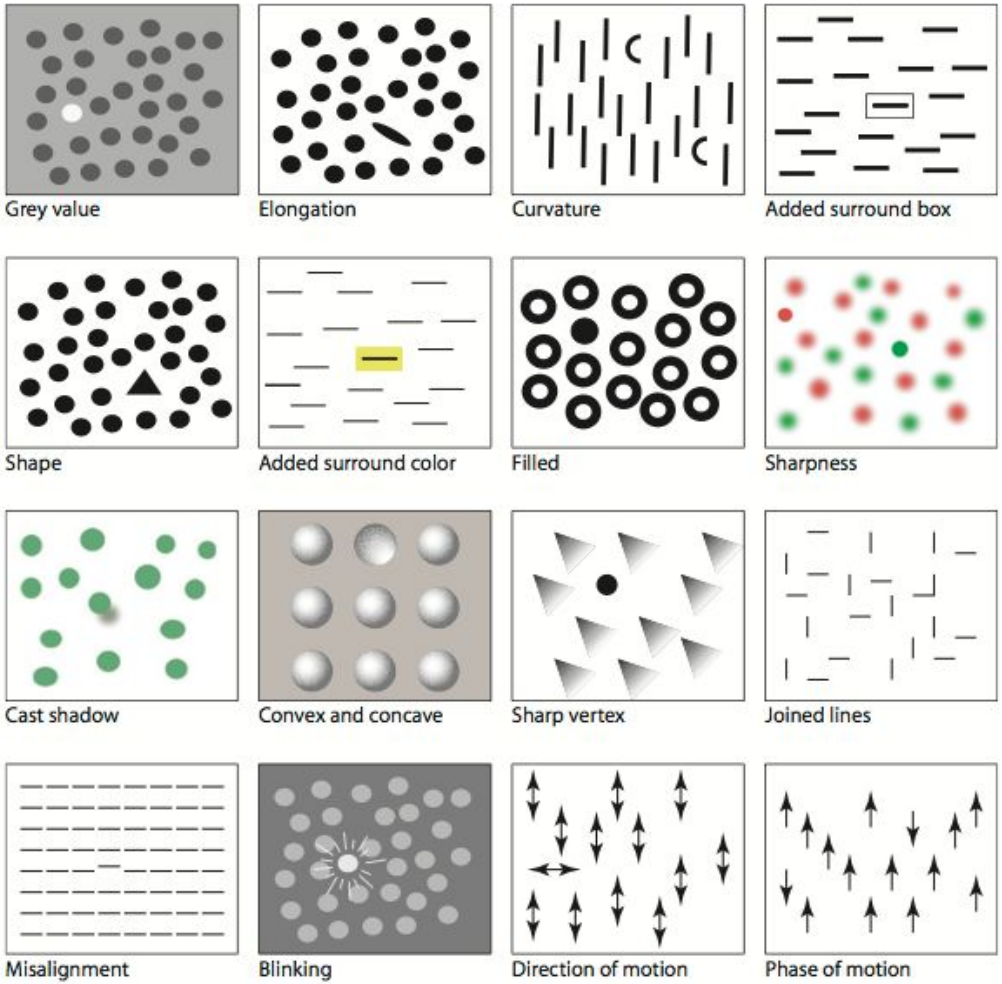
217 were armed Palestinian militants. Of those, **2** were children.

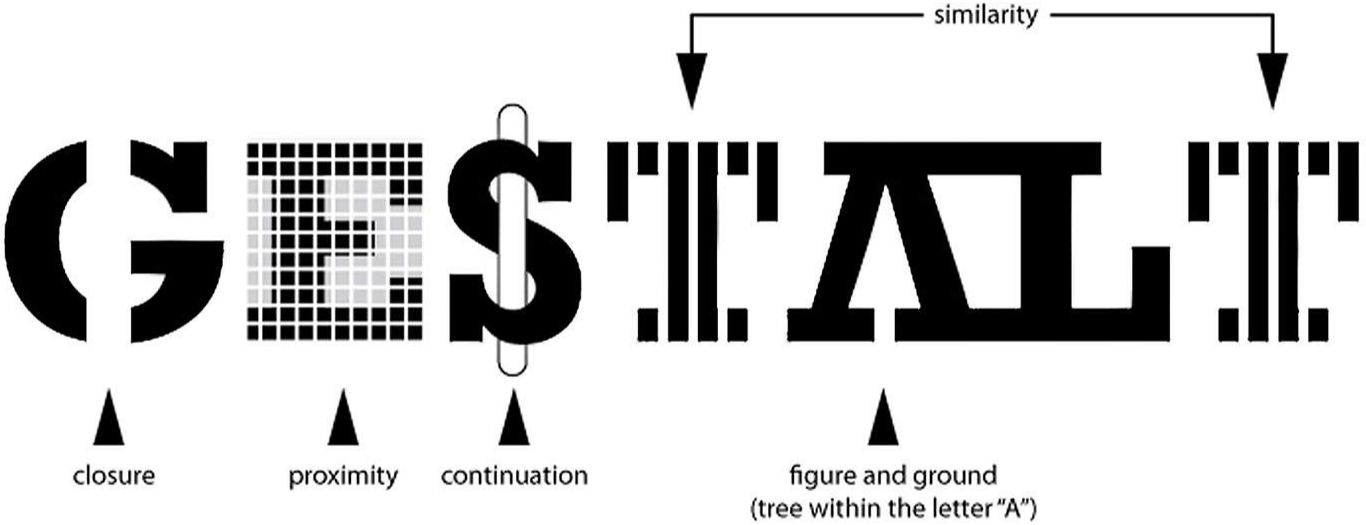
1396 were Palestinian civilians. Of those, **222** were women and **418** were children.



The Washington Post

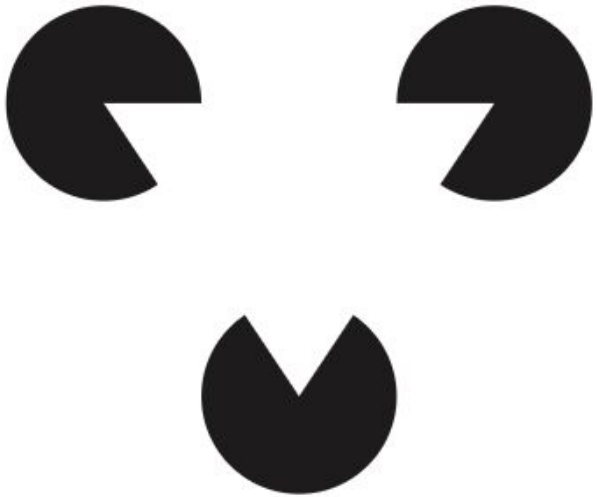

Color
hue
lightness
Elementary shape
size
elongation
orientation
Motion
Spatial grouping
Ware 2008

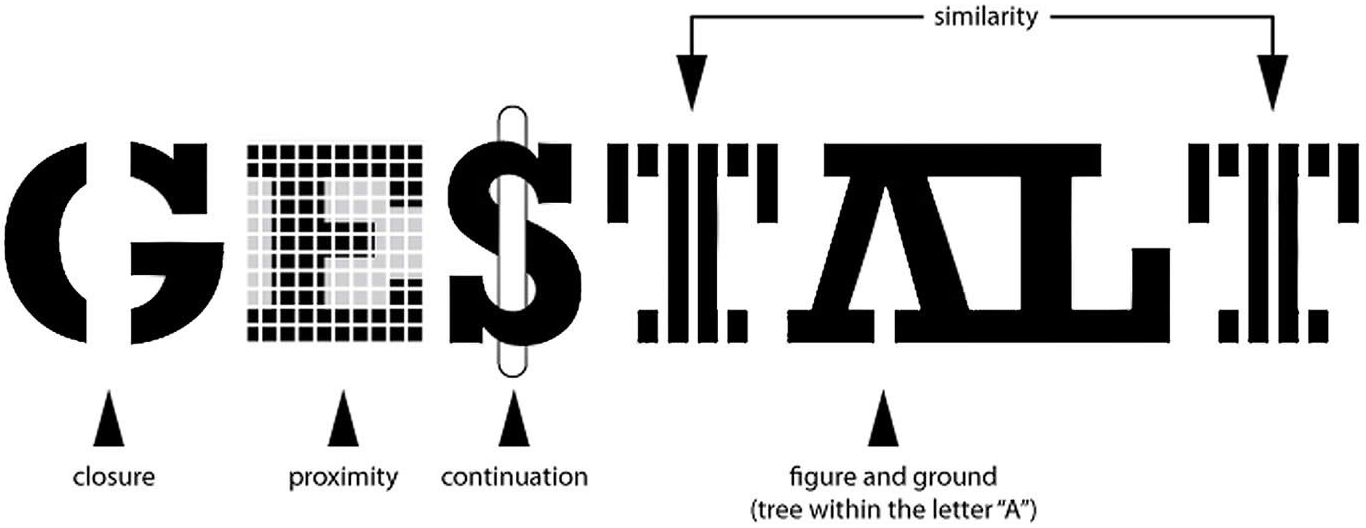

Grey value
Elongation
Curvature
Added surround box
Shape
Added surround color
Filled
Sharpness
Cast shadow
Convex and concave
Sharp vertex
Joined lines
Misalignment
Blinking
Direction of motion
Phase of motion



an organized whole ("gestalt") that is perceived as more than the sum of its parts.

“Gestalt” is about how we visually assemble objects




similarity
closure
proximity
continuation
figure and ground
(tree within the letter "A")

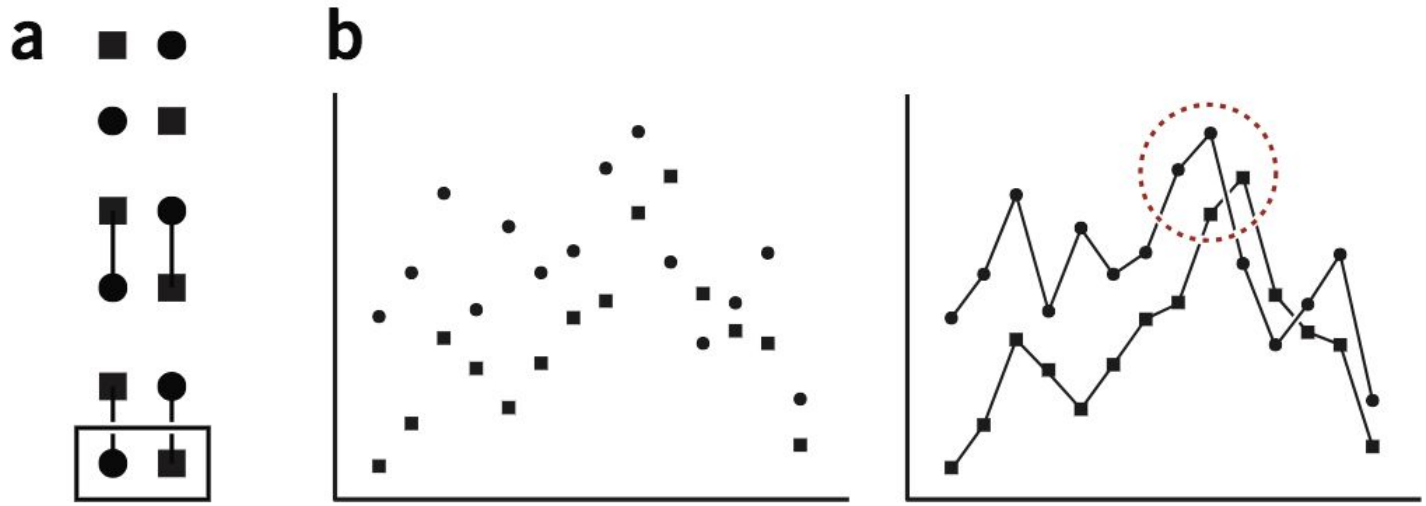


Figure 2 | Principles of grouping. (**a**) Relative strength of grouping by similarity, proximity, connection and enclosure. (**b**) Lines in graphs create clear connection. Enclosure is an effective way to draw attention to a group of objects.

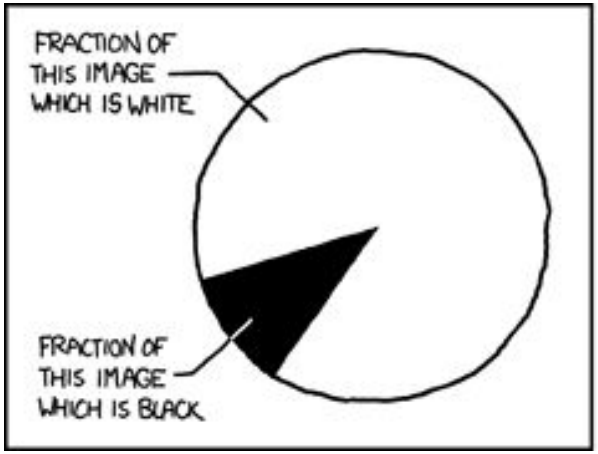
Some design principles

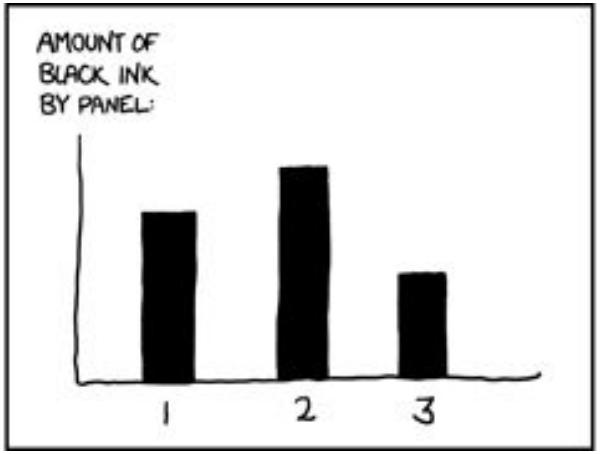
E. R. Tufte:
Maximize Data-Ink Ratio
(minimalism)

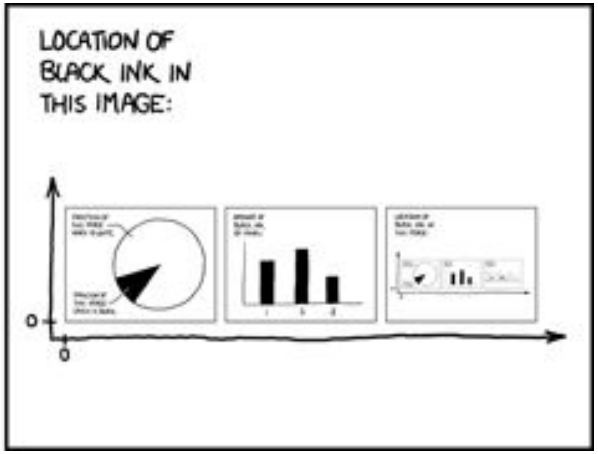
$$\text{Data-ink ratio} = \frac{\text{data-ink}}{\text{total ink used to print the graphic}}$$

$$= \text{proportion of a graphic's ink devoted to the non-redundant display of data-information}$$

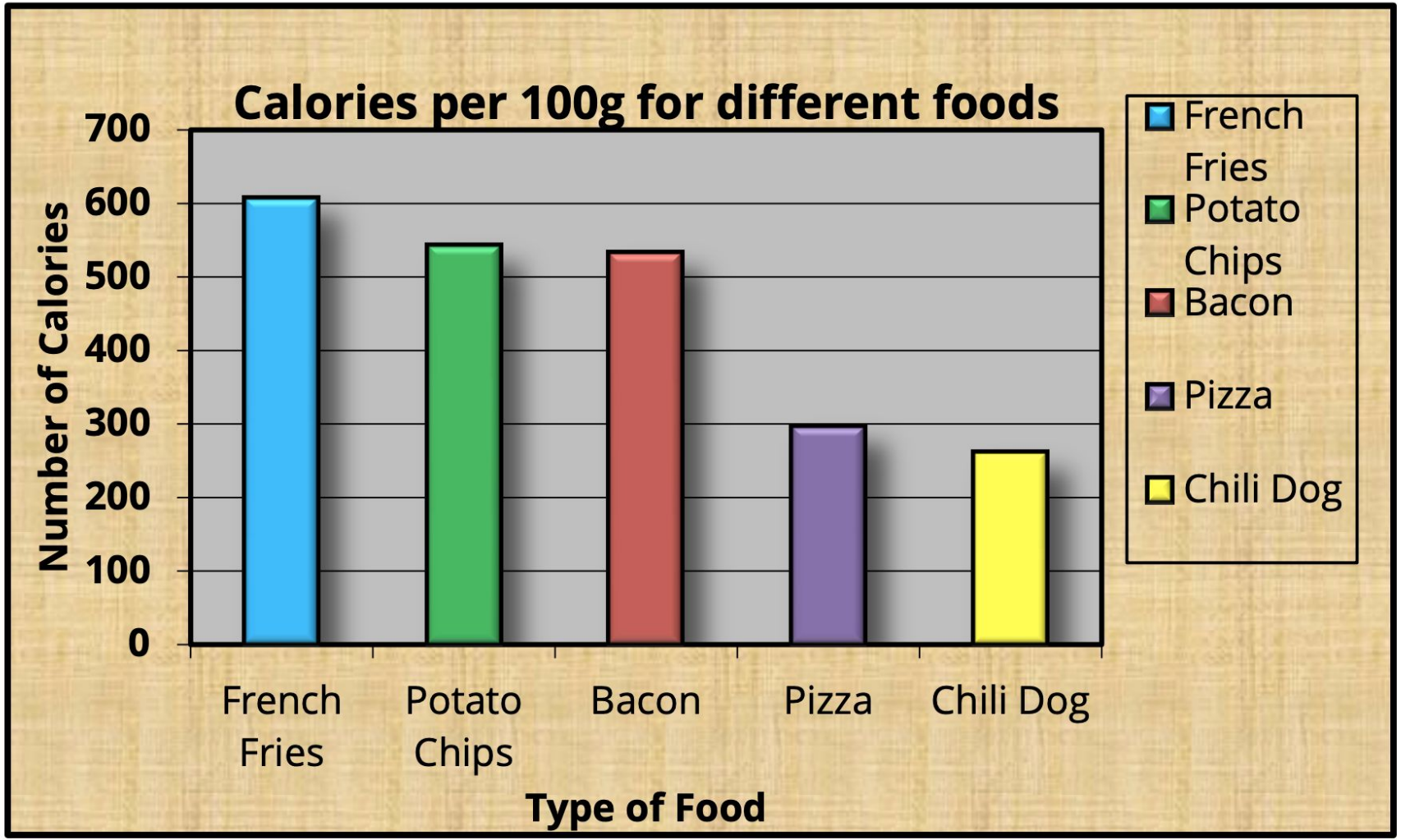
$$= 1.0 - \text{proportion of a graphic that can be erased without loss of data-information.}$$



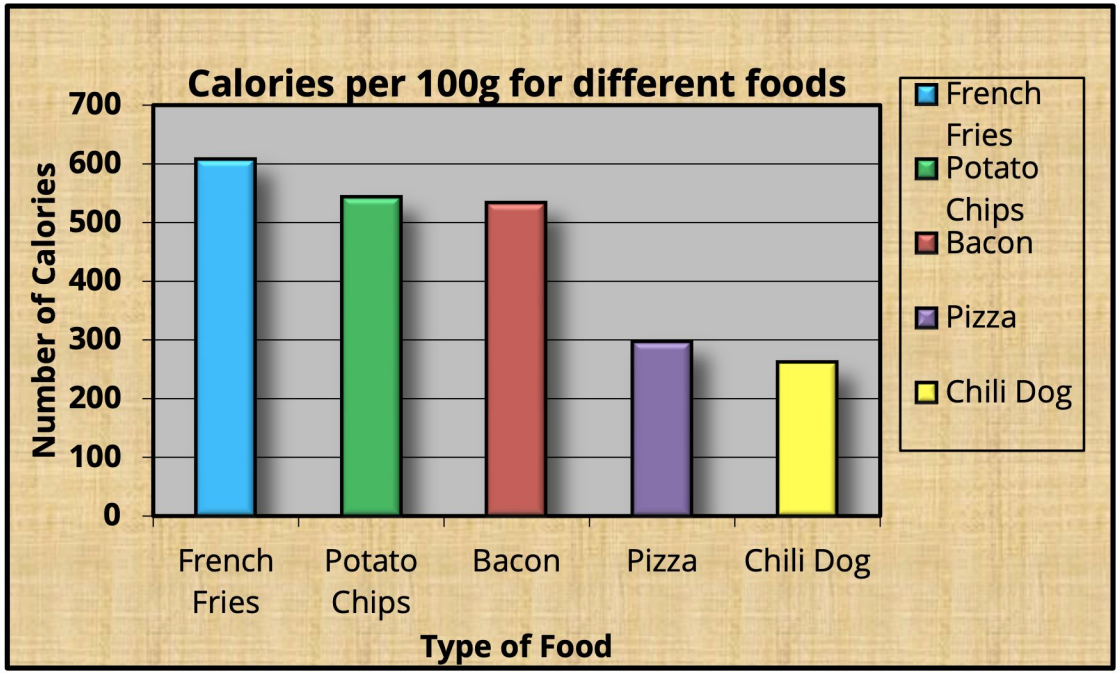


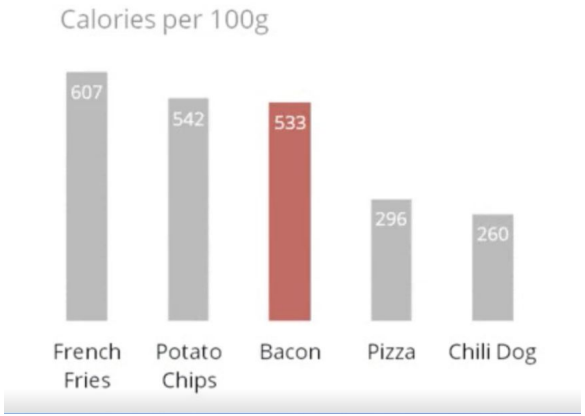


https://xkcd.com/688/


Calories per 100g for different foods
Number of Calories
700
600
500
400
300
200
100
0
French Fries
Potato Chips
Bacon
Pizza
Chili Dog
Type of Food
French Fries
Potato Chips
Bacon
Pizza
Chili Dog

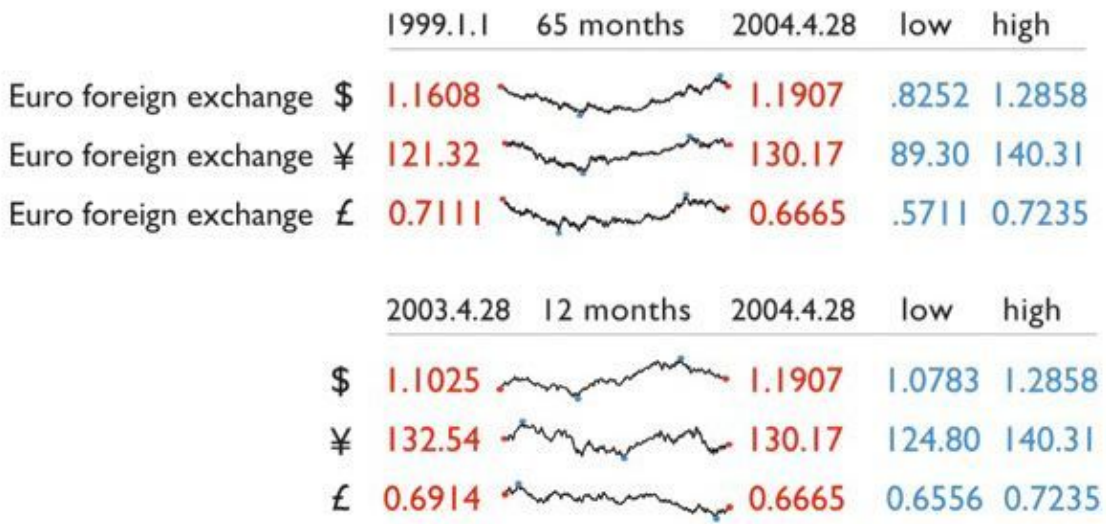
Maximize the data-ink ratio!





Check: https://i0.wp.com/socialmediaguerilla.com/wp-content/uploads/2013/10/data-ink.gif
https://files.speakerdeck.com/presentations/87bb9f00ec1e01308020727faa1f9e72/data-ink.pdf

An example: Sparklines



http://projects.fivethirtyeight.com/2016-endorsement-primary/

Sparklines

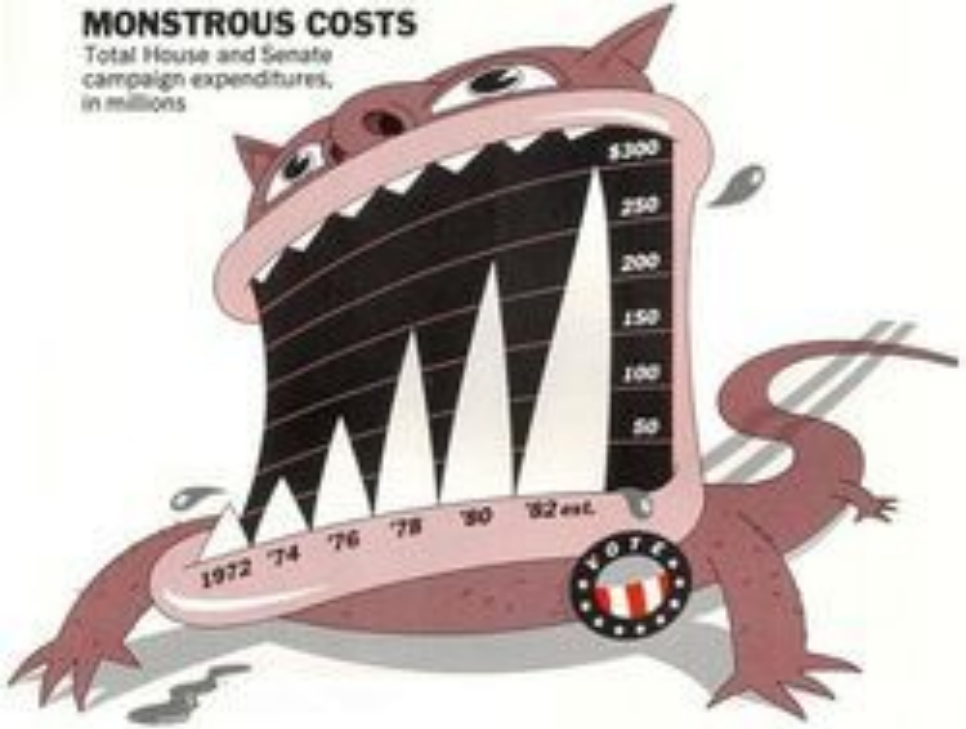
Dequantification In exchange for an enormous increase in graphical resolving power, the wordlike size of sparklines precludes the overt labels and scaling of conventional statistical displays. Most of our examples have, however, depicted *contextual methods* for quantifying sparklines: the gray bar for normal limits and the red encoding to link data points in sparklines to exact numbers  glucose 6.6 ; global scale bars and labels for sparkline clusters; and, probably best of all, surrounding a sparkline with an implicit data-scaling box formed by nearby numbers that label key data points (such as beginning/end, high/low) 1.1025  1.1907 1.0783 1.2858. And now and then sparklines might be scaled by very small type: 

Mauricio Pochettino has lead Spurs on their best run 8TH  2ND in 24 years of the Premier League

Alibaba stock is at 5 yr high 93.89 

The FTSE100 Brexit bounce 5562
from the vote last su

Tufte: remove “Chart junk”

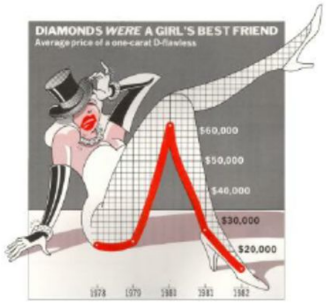


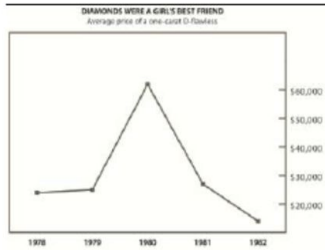
Nigel Holmes

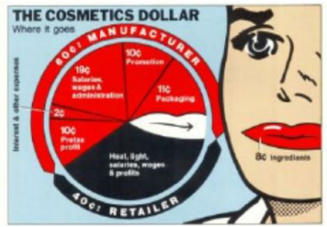
Two pioneers of **modern data** visualization.

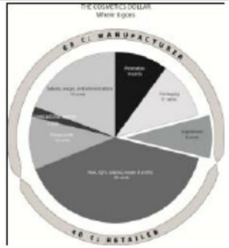
- *Edward Tufte*

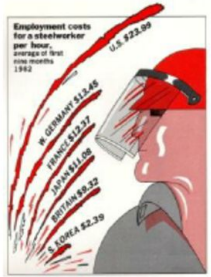
- *Nigel Holmes*

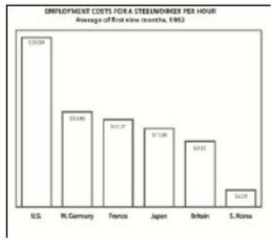


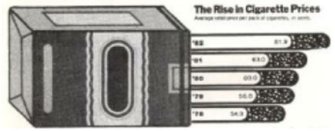


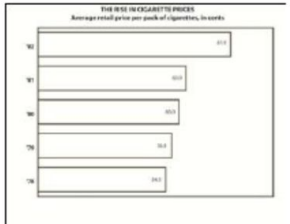


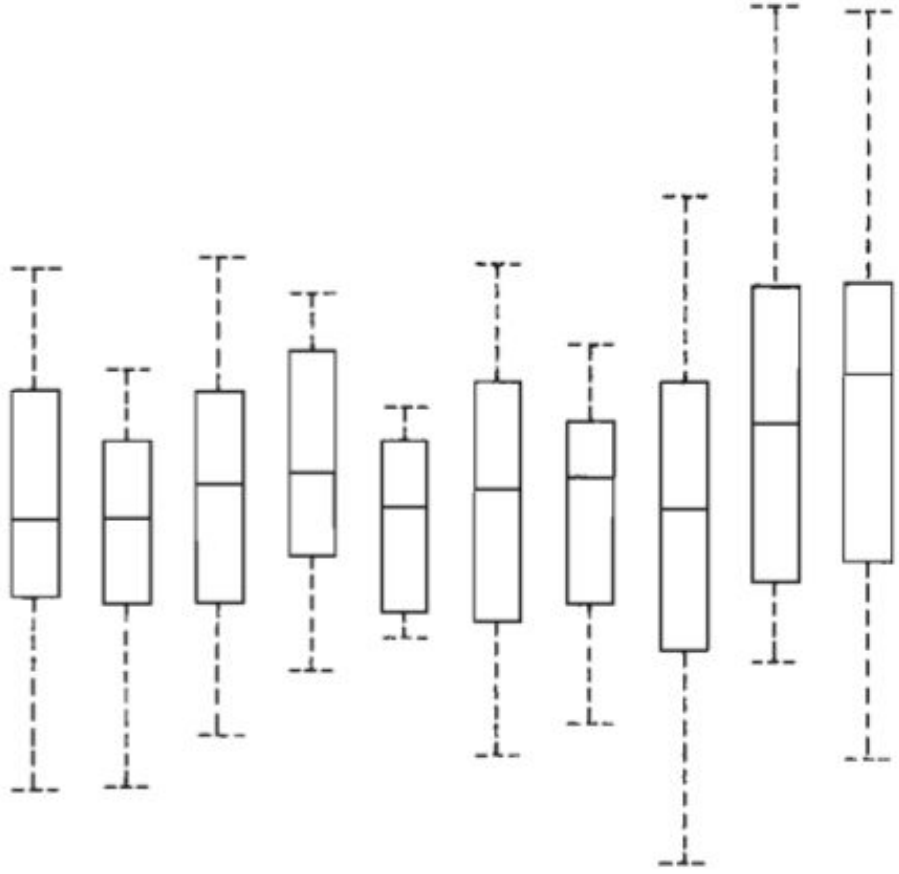






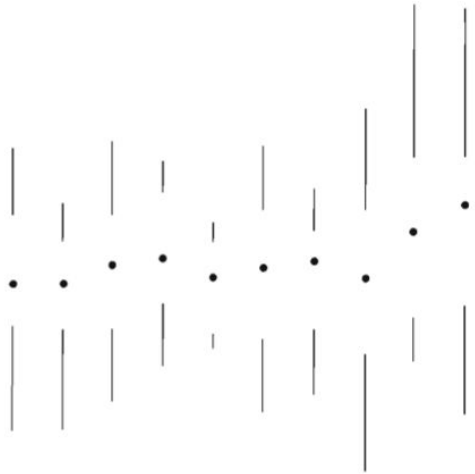


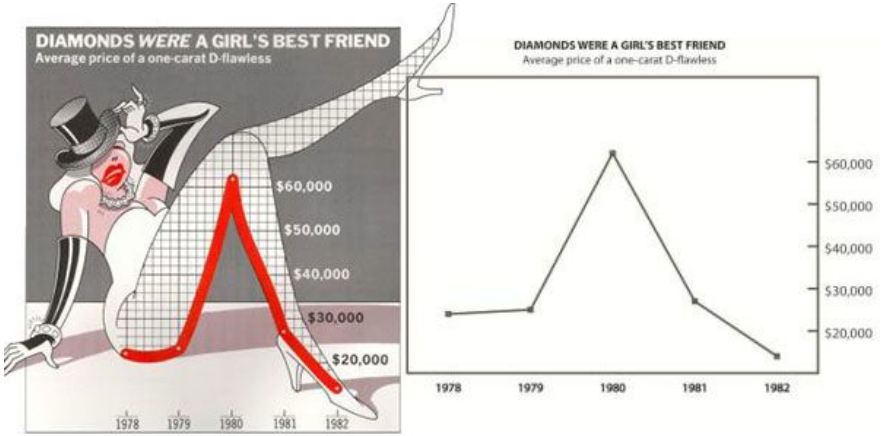




remove “Chart junk”

Maximize data-ink ratio





What do you think?

Useful Junk? The Effects of Visual Embellishment on Comprehension and Memorability of Charts

Scott Bateman, Regan L. Mandryk, Carl Gutwin,
Aaron Genest, David McDine, Christopher Brooks
Department of Computer Science, University of Saskatchewan, Saskatoon, Saskatchewan, Canada
scott.bateman@usask.ca, regan@cs.usask.ca, gutwin@cs.usask.ca,
aaron.genest@usask.ca, dam085@mail.usask.ca, cab938@mail.usask.ca

ABSTRACT
Guidelines for designing information charts often state that the presentation should reduce 'chart junk' – visual embellishments that are not essential to understanding the data. In contrast, some popular chart designers wrap the presented data in detailed and elaborate imagery, raising the questions of whether this imagery is really as detrimental to understanding as has been proposed, and whether the visual embellishment may have other benefits. To investigate these issues, we conducted an experiment that compared embellished charts with plain ones, and measured both interpretation accuracy and long-term recall. We found that people's accuracy in describing the embellished charts was no worse than for plain charts, and that their recall after a two-to-three-week gap was significantly better. Although we are cautious about recommending that all charts be produced in this style, our results question some of the premises of the minimalist approach to chart design.

Author Keywords
Charts, information visualization, imagery, memorability.

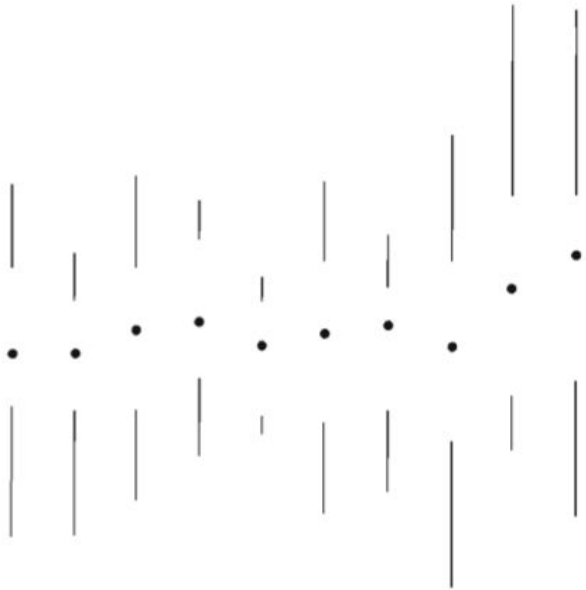
ACM Classification Keywords
H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms
Design, Human Factors

INTRODUCTION
Many experts in the area of chart design, such as Edward Tufte, criticize the inclusion of visual embellishment in charts and graphs; their guidelines for good chart design often suggest that the addition of *chart junk,* decorations and other kinds of non-essential imagery, to a chart can *data-ink* – or the ink in the chart used to represent data.

Despite these minimalist guidelines, many designers include a wide variety of visual embellishments in their charts, from small decorations to large images and visual backgrounds. One well-known proponent of visual embellishment in charts is the graphic artist Nigel Holmes, whose work regularly incorporates strong visual imagery into the fabric of the chart [7] (e.g., Figure 1).

http://hci.usask.ca/uploads/173-pap0297-bateman.pdf



Steve Haroz 📊👁🧠
@sharoz

People perform worse (more inaccurate) with Tufte's mid-gap plot compared with standard box and whisker plots.

(Stock, W. A., & Behrens, J. T. 1991)
doi:10.3102/10769986016001001
via @hadleywickham

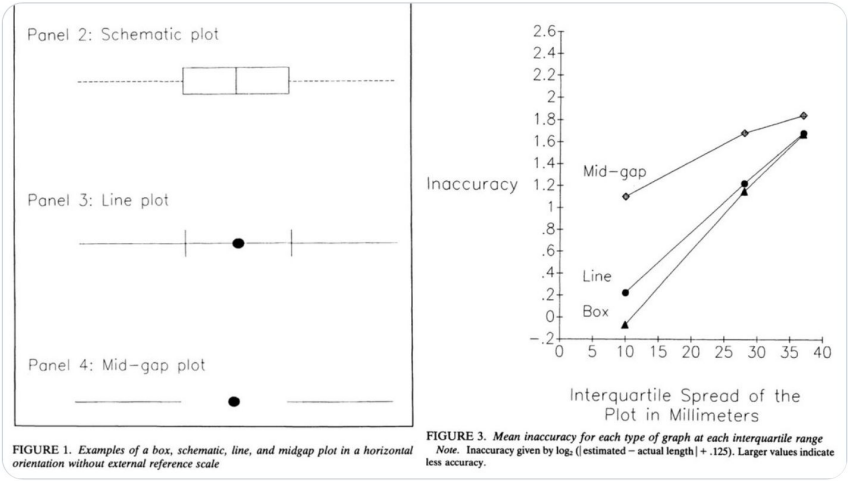


FIGURE 1. *Examples of a box, schematic, line, and midgap plot in a horizontal orientation without external reference scale*

FIGURE 3. *Mean inaccuracy for each type of graph at each interquartile range*
Note. Inaccuracy given by $\log_2$ (| estimated − actual length | + .125). Larger values indicate less accuracy.

7:49 AM · Sep 23, 2018 · Twitter Web Client

What’s your thoughts about “chart junks”?