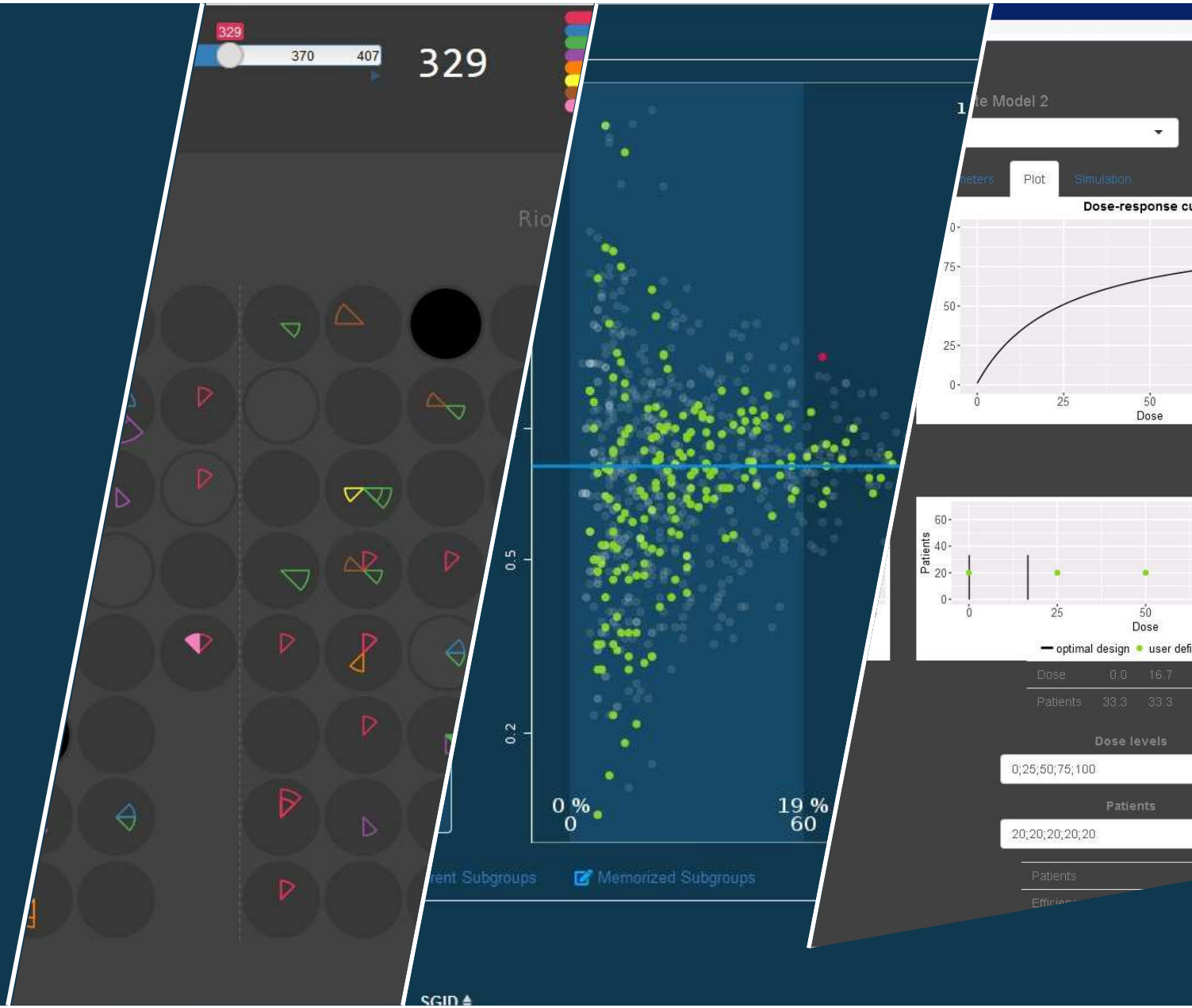
Data Story Telling

Bodo Kirsch

June 2024



Agenda

- Motivation
  - Definition
  - Why visualize data?
- Background
  - Gestalt Principles
  - Pre-attentive processing
  - Hierarchy of perception
- Good data visualization principles

Educational Objectives

Provide an answer to the following questions:

// Why is it important to visualize data?

// What does minimizing the ink to data ratio mean?

// What is the difference between data story telling and a statistical graphic?

Acknowledgements

// Wonderful Wednesday Webinars, VIS SIG of PSI

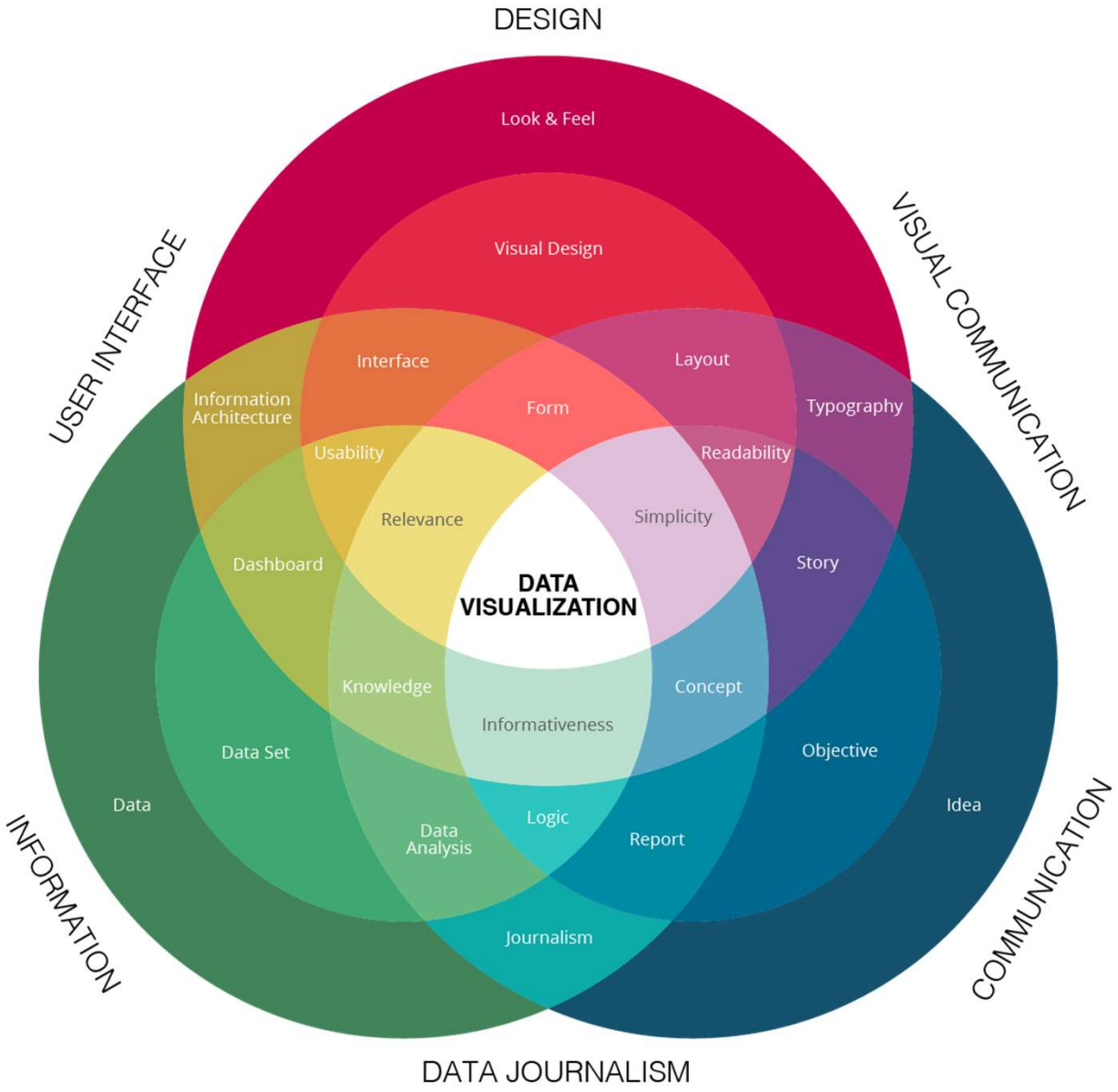
// Abel Rodriguez, Professor in statistics at UCLA

// Tamara Munzner, Professor in information visualization at University of British Columbia

// Brenda Crowe and Zak Skrivanek, Eli Lilly

What is it?

// Size
// Shape
// Color
// Location
// Motion
// Sound



Insights from multiple disciplines….

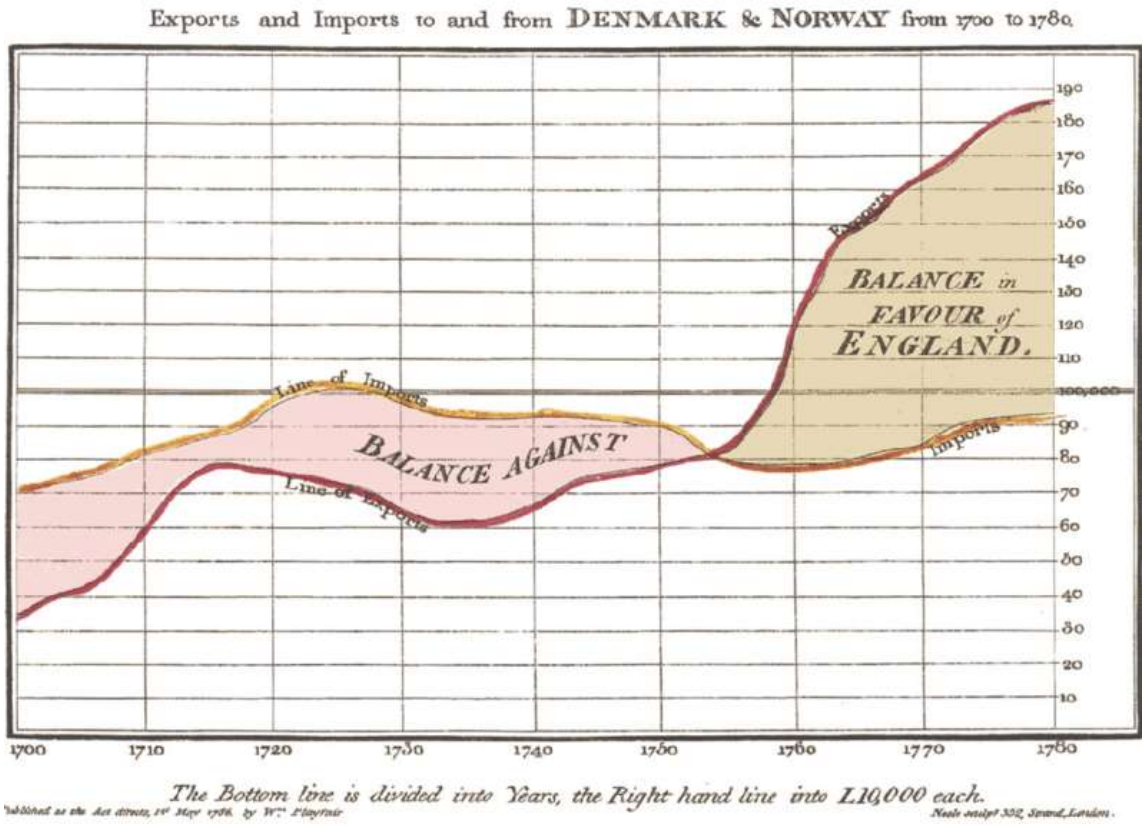
// Graphic design: Emphasizes aesthetics.
// Computer science: Emphasizes algorithms.
// Cognitive psychology: Provides insights into the most effective tools.
// Journalism: Emphasizes storytelling.
// Statistics: Emphasizes quantification of information.

A (Very) Short History of Visualization

- 1637 — Descartes first uses 2D grids to visually encode numbers.
- 1786 — William Playfair's "The Commercial and Political Atlas".
- 1855 — John Snow uses maps to link the 1854 London cholera epidemic to contaminated drinking water.
- 1857 — Florence Nightingale uses stacked bar and pie charts to persuade Queen Victoria to improve conditions on British military hospitals.
- 1954 — Darrel Huff's "How to Lie with Statistics".
- 1977 — John Tukey introduces boxplots.
- 1983 — Edward Tufte's "Visual Displays of Quantitative Information".
- 1994 — William Cleveland's "The Elements of Graphing Data".
- 2004 — Stephen Few "Show me the Numbers".
- Nowadays dominated by computer scientists (on the technical side) and business analytics (on the more applied side).

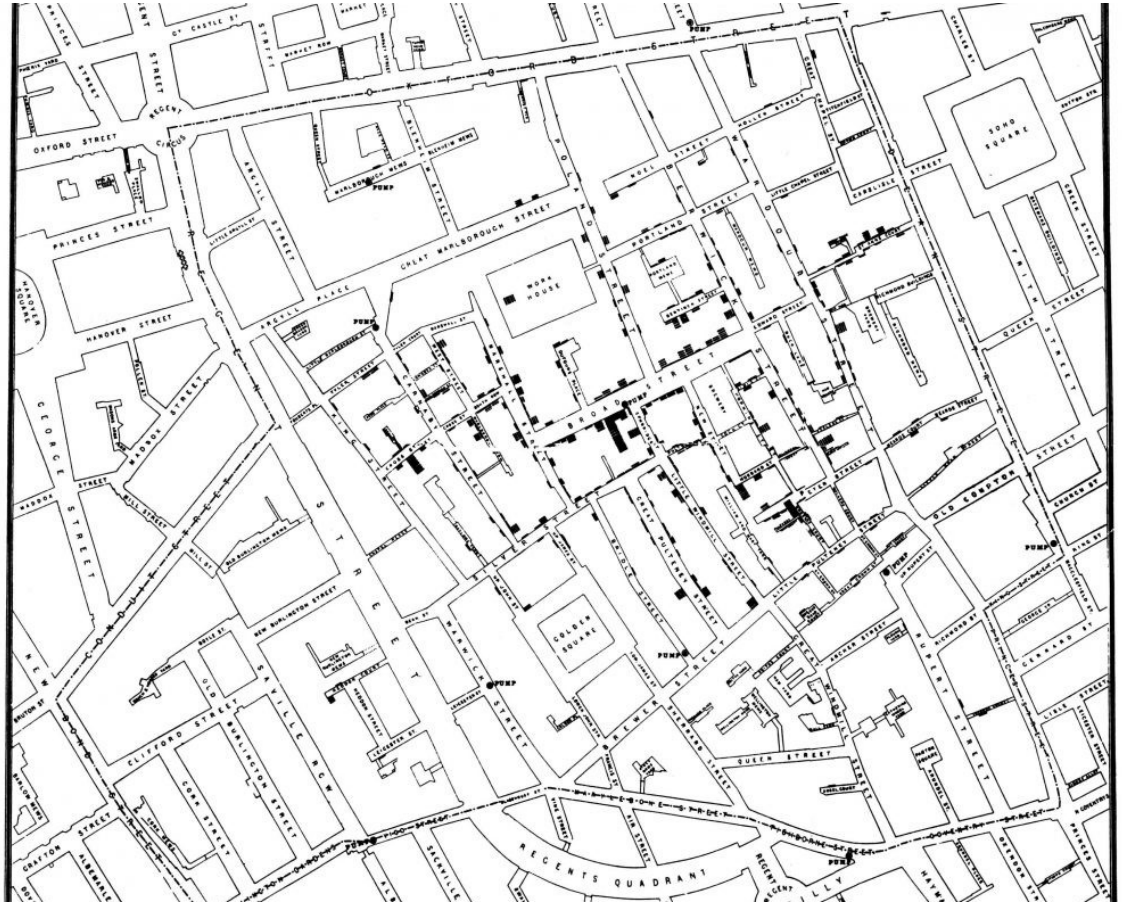
“Invented Statistical graphics”

- 1637 — Descartes first uses 2D grids to visually encode numbers.
- 1786 — William Playfair's "The Commercial and Political Atlas".
- 1855 — John Snow uses maps to link the 1854 London cholera epidemic to contaminated drinking water.
- 1857 — Florence Nightingale uses stacked bar and pie charts to persuade Queen Victoria to improve conditions on British military hospitals.
- 1954 — Darrel Huff's "How to Lie with Statistics".
- 1977 — John Tukey introduces boxplots.
- 1983 — Edward Tufte's "Visual Displays of Quantitative Information".
- 1994 — William Cleveland's "The Elements of Graphing Data".
- 2004 — Stephen Few "Show me the Numbers".
- Nowadays dominated by computer scientists (on the technical side) and business analytics (on the more applied side).



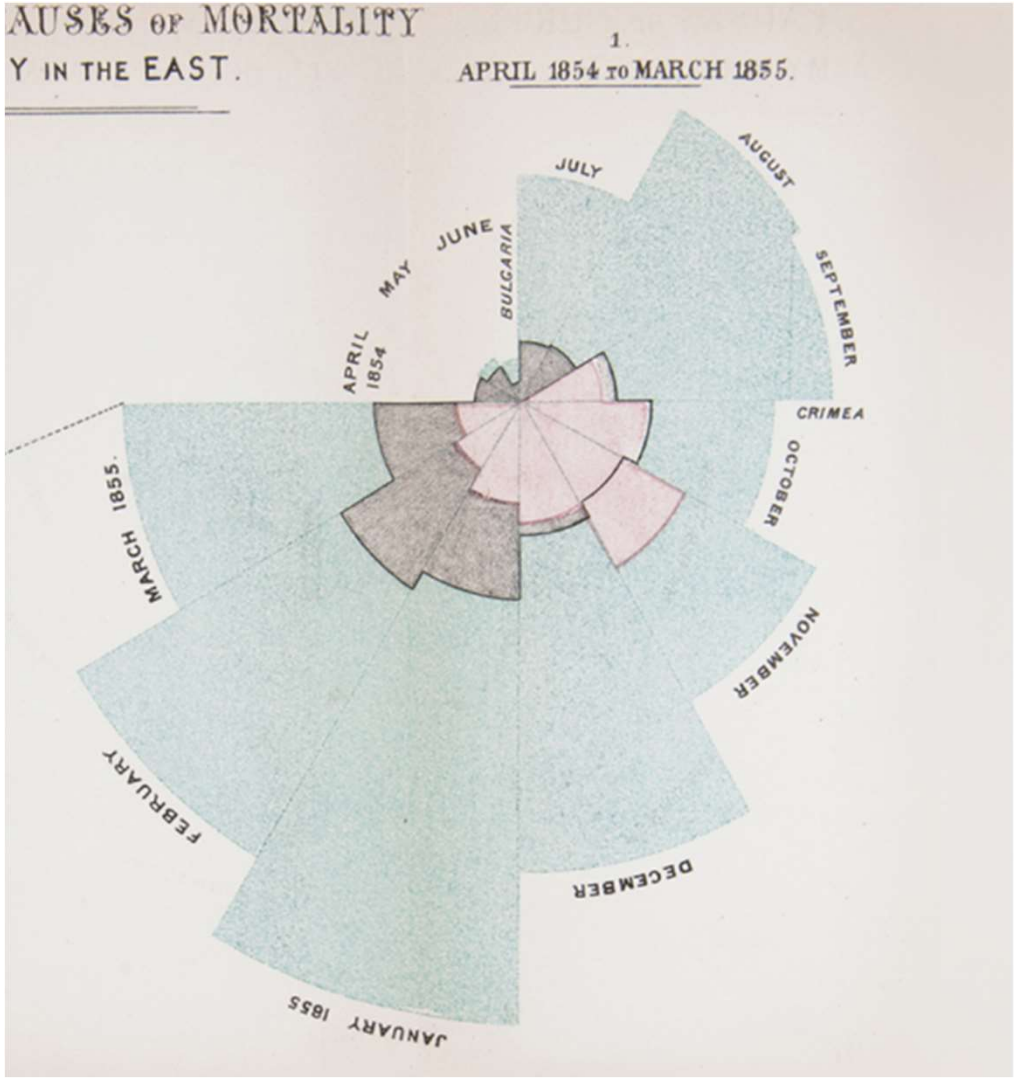
“Map that Revolutionized disease prevention”

- 1637 — Descartes first uses 2D grids to visually encode numbers.
- 1786 — William Playfair's "The Commercial and Political Atlas".
- 1855 — John Snow uses maps to link the 1854 London cholera epidemic to contaminated drinking water.
- 1857 — Florence Nightingale uses stacked bar and pie charts to persuade Queen Victoria to improve conditions on British military hospitals.
- 1954 — Darrel Huff's "How to Lie with Statistics".
- 1977 — John Tukey introduces boxplots.
- 1983 — Edward Tufte's "Visual Displays of Quantitative Information".
- 1994 — William Cleveland's "The Elements of Graphing Data".
- 2004 — Stephen Few "Show me the Numbers".
- Nowadays dominated by computer scientists (on the technical side) and business analytics (on the more applied side).



Data Viz that changed healthcare

- 1637 — Descartes first uses 2D grids to visually encode numbers.
- 1786 — William Playfair's "The Commercial and Political Atlas".
- 1855 — John Snow uses maps to link the 1854 London cholera epidemic to contaminated drinking water.
- 1857 — Florence Nightingale uses stacked bar and pie charts to persuade Queen Victoria to improve conditions on British military hospitals.
- 1954 — Darrel Huff's "How to Lie with Statistics".
- 1977 — John Tukey introduces boxplots.
- 1983 — Edward Tufte's "Visual Displays of Quantitative Information".
- 1994 — William Cleveland's "The Elements of Graphing Data".
- 2004 — Stephen Few "Show me the Numbers".
- Nowadays dominated by computer scientists (on the technical side) and business analytics (on the more applied side).



Data Story Telling

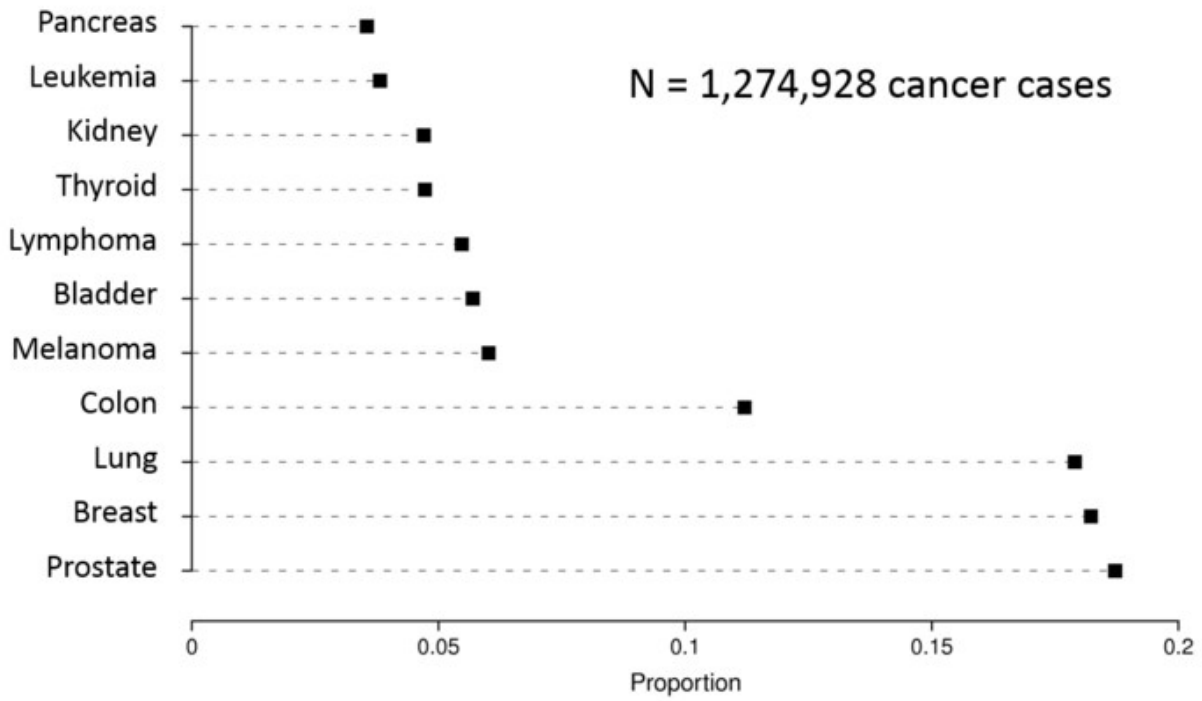
Motivation

Table: Cancer Incidence by Type

Type	Incidence	Proportion
Prostate	238,590	18.7%
Breast	232,340	18.2%
Lung	228,190	17.9%
Colon	142,820	11.2%
Melanoma	76,690	6.0%
Bladder	72,570	5.7%
Lymphoma	69,740	5.5%
Thyroid	60,220	4.7%
Kidney	59,938	4.7%
Leukemia	48,610	3.8%
Pancreas	45,220	3.5%

Data from http://www.cancer.gov/cancertopics/types/commoncancers

Graph: Cancer Incidence by Type

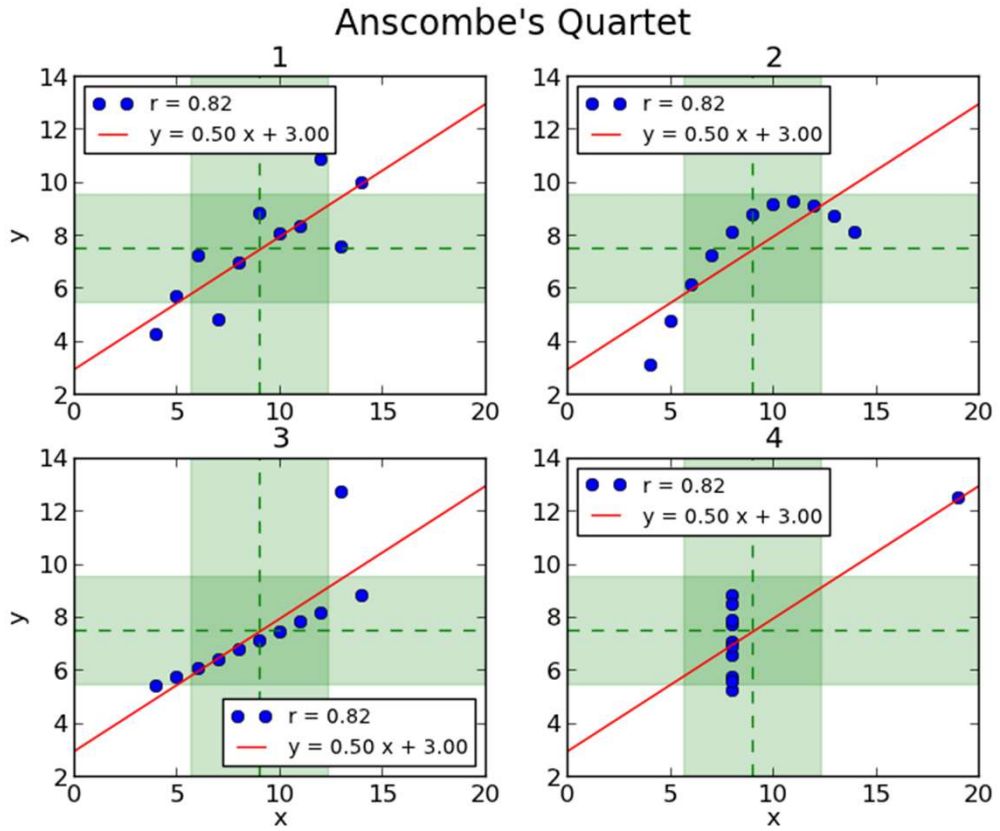


Data from http://www.cancer.gov/cancertopics/types/commoncancers

Are there situations where a table is better than a graph?

- Yes, but these are relative exceptions.
  - To convey a handful of numbers.
  - To report precise values for lookup.
- Tables are usually a bad idea if comparison is important.
- Tables are typically limited to summary or inferential statistics. They do not allow you to see the data.

Ancombe’s Quartet



Data Story Telling

Background

Gestalt Principles

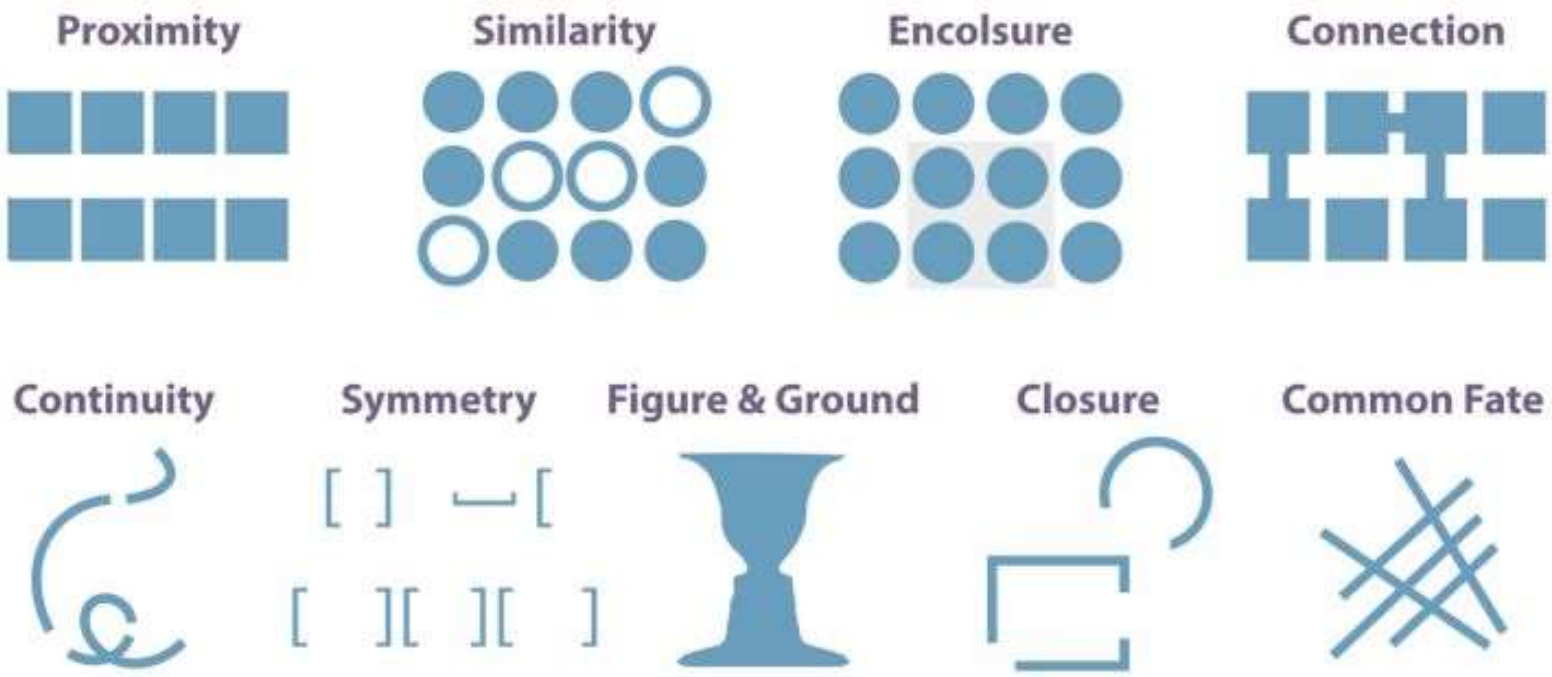
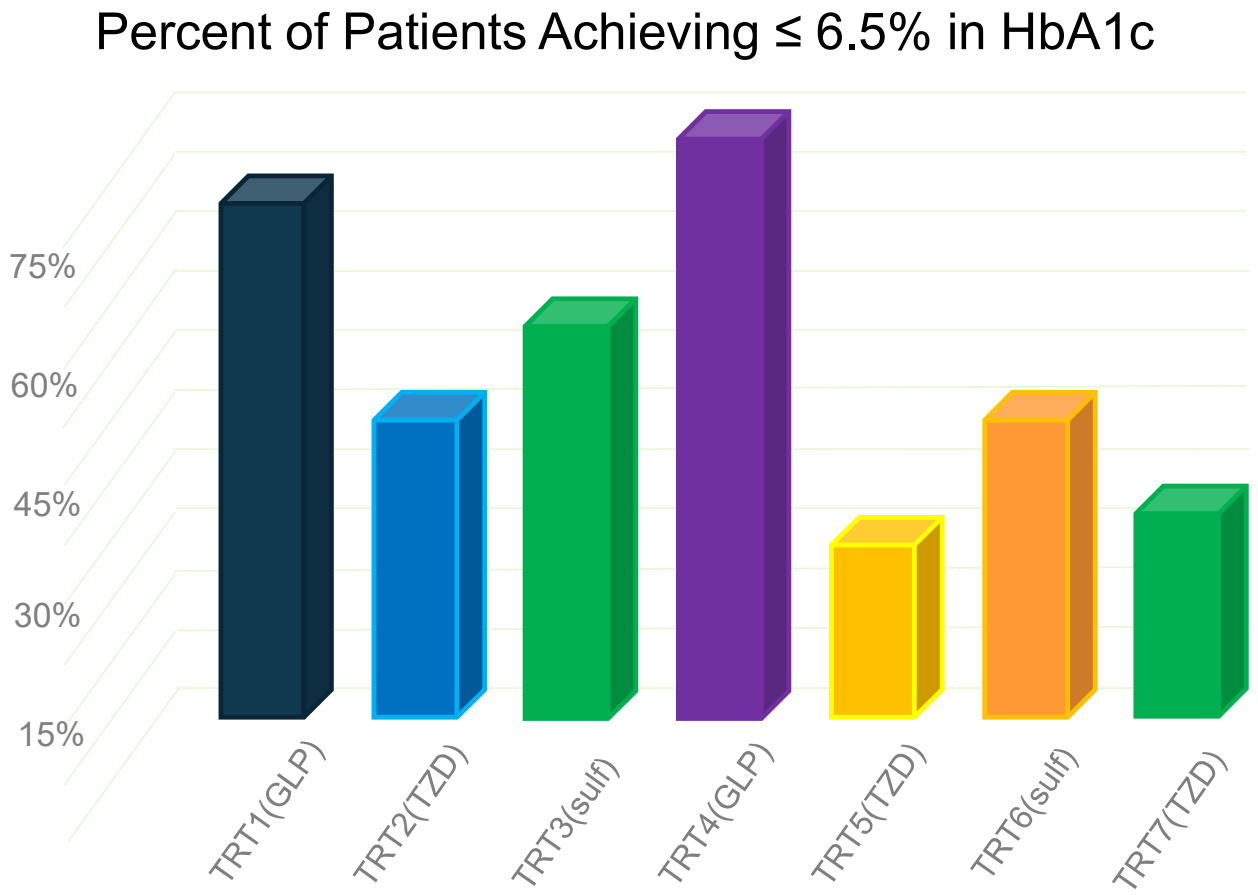
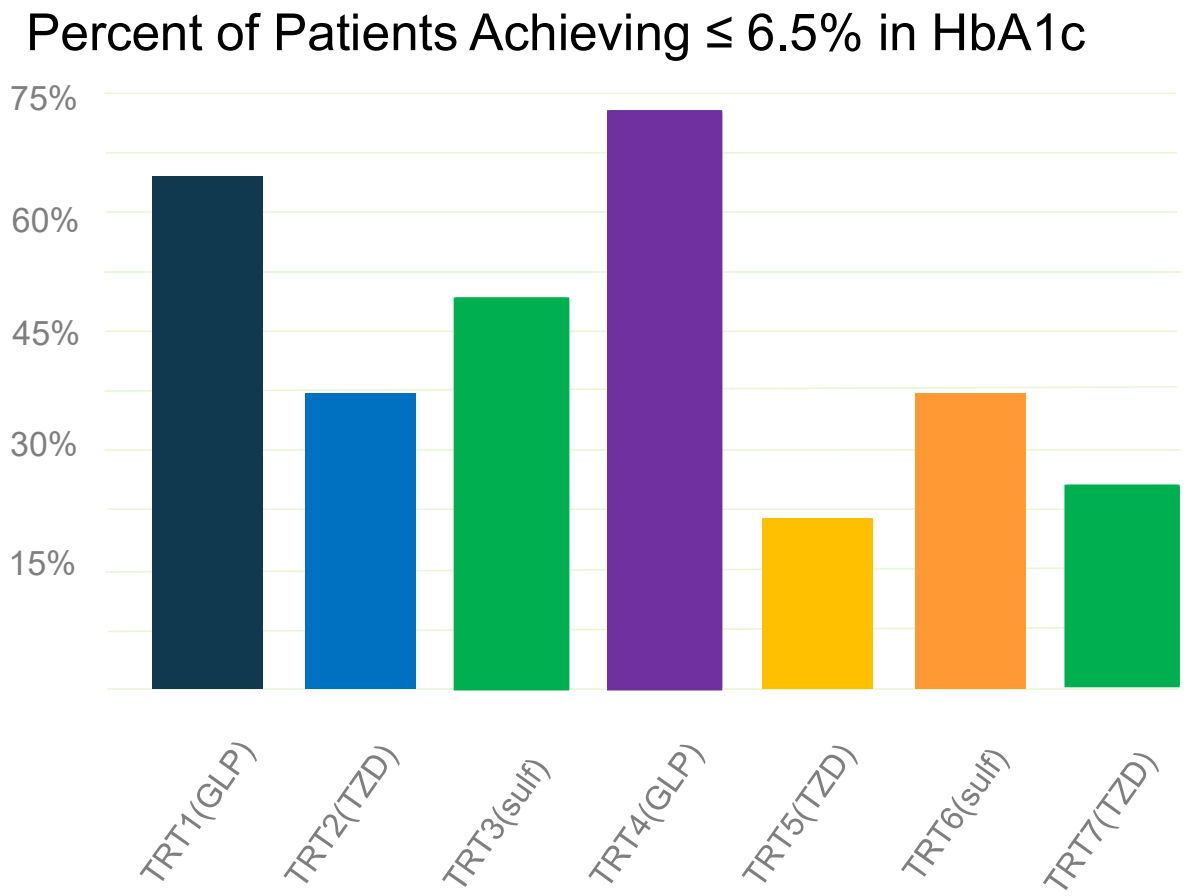


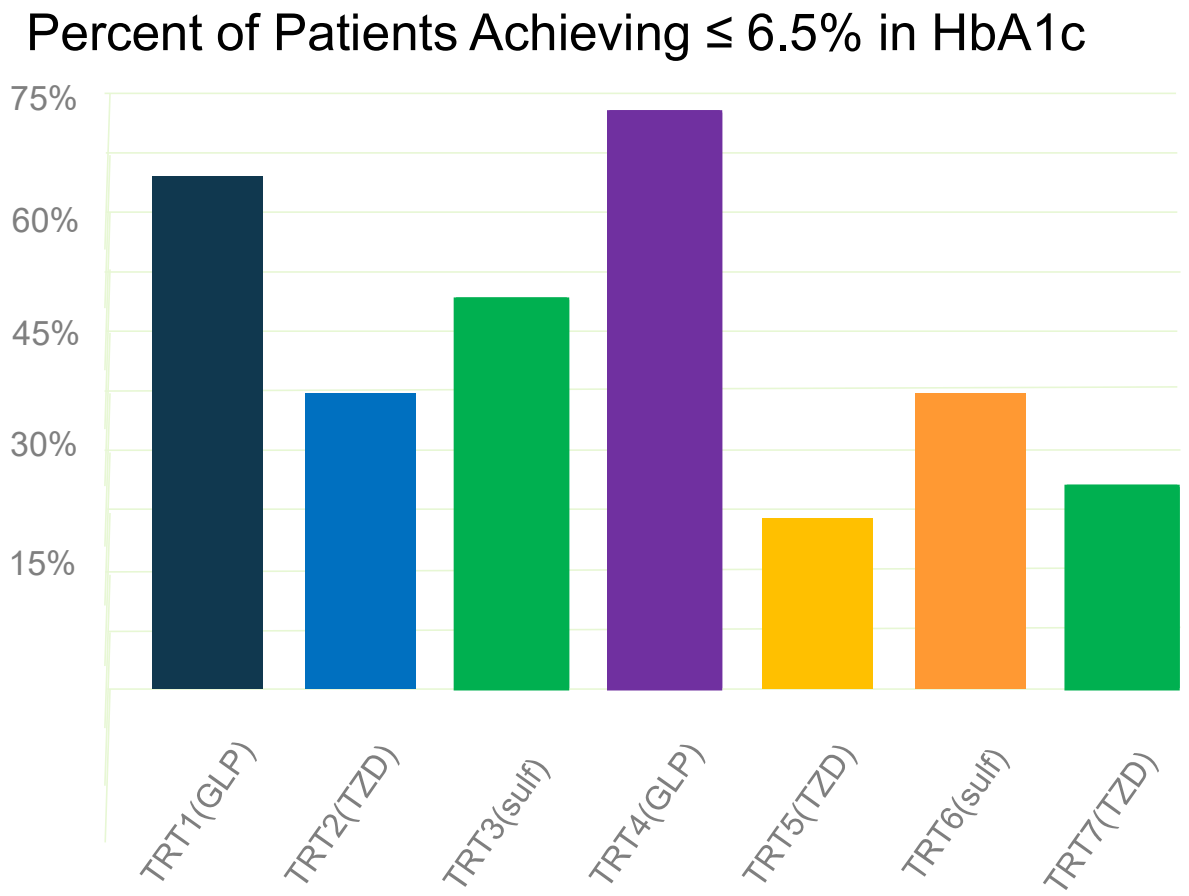
Chart Junk


Percent of Patients Achieving ≤ 6.5% in HbA1c
75%
60%
45%
30%
15%
TRT1(GLP)
TRT2(TZD)
TRT3(sulf)
TRT4(GLP)
TRT5(TZD)
TRT6(sulf)
TRT7(TZD)

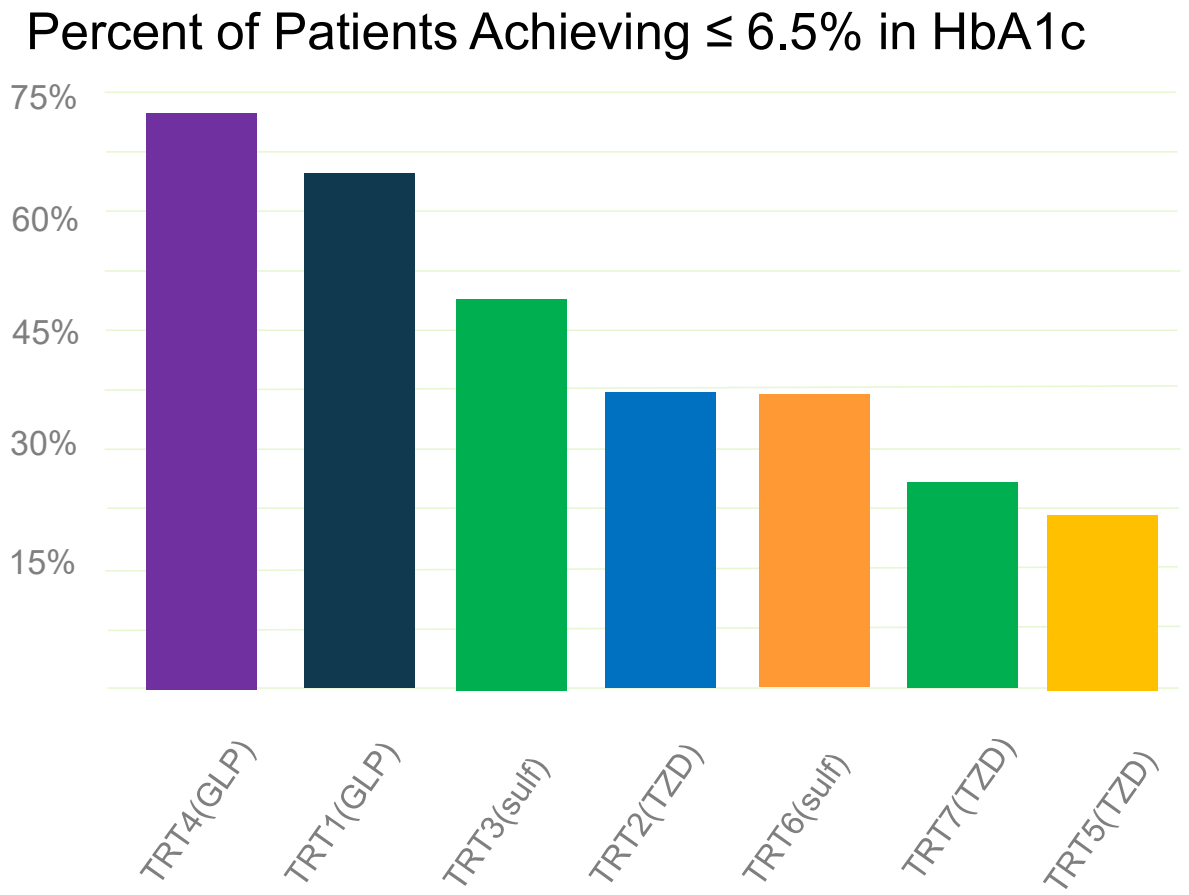
The Gestalt Law of Prägnanz



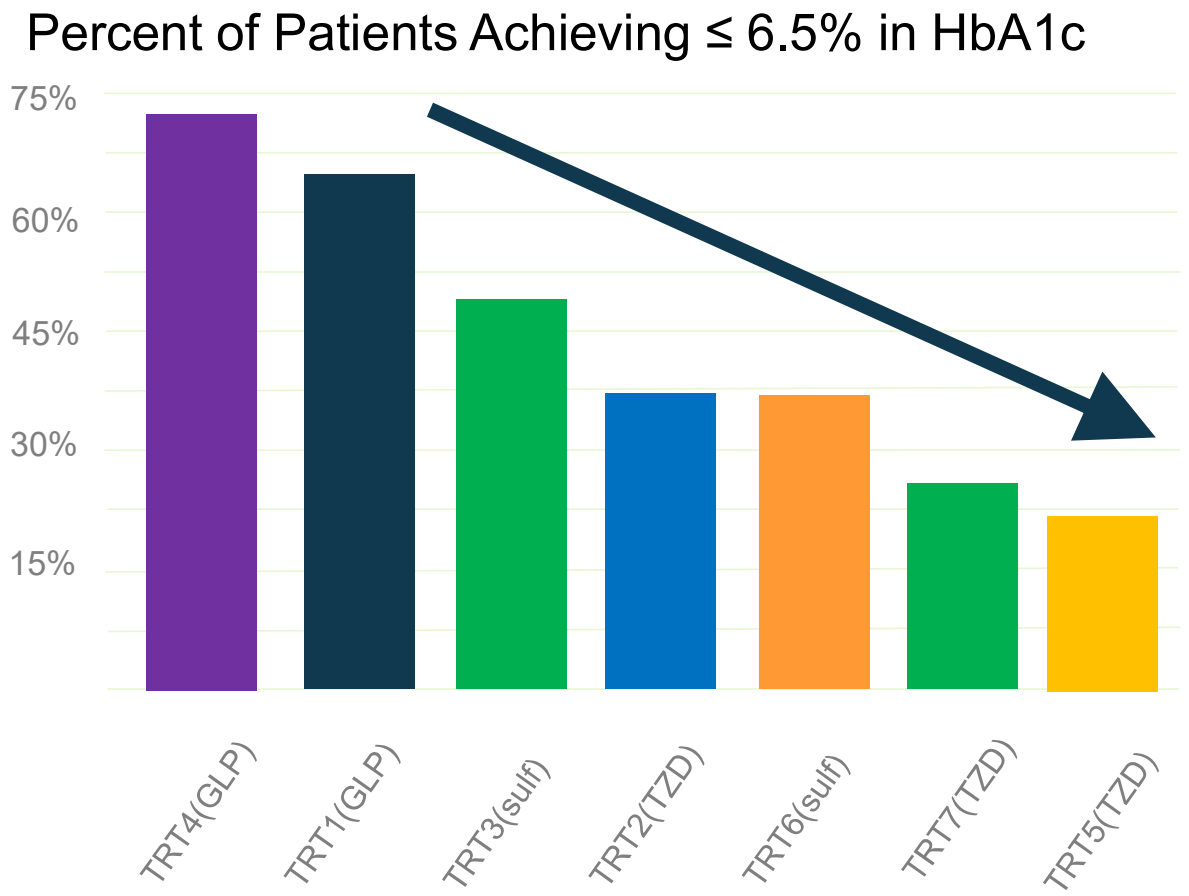
The Gestalt Law of Prägnanz



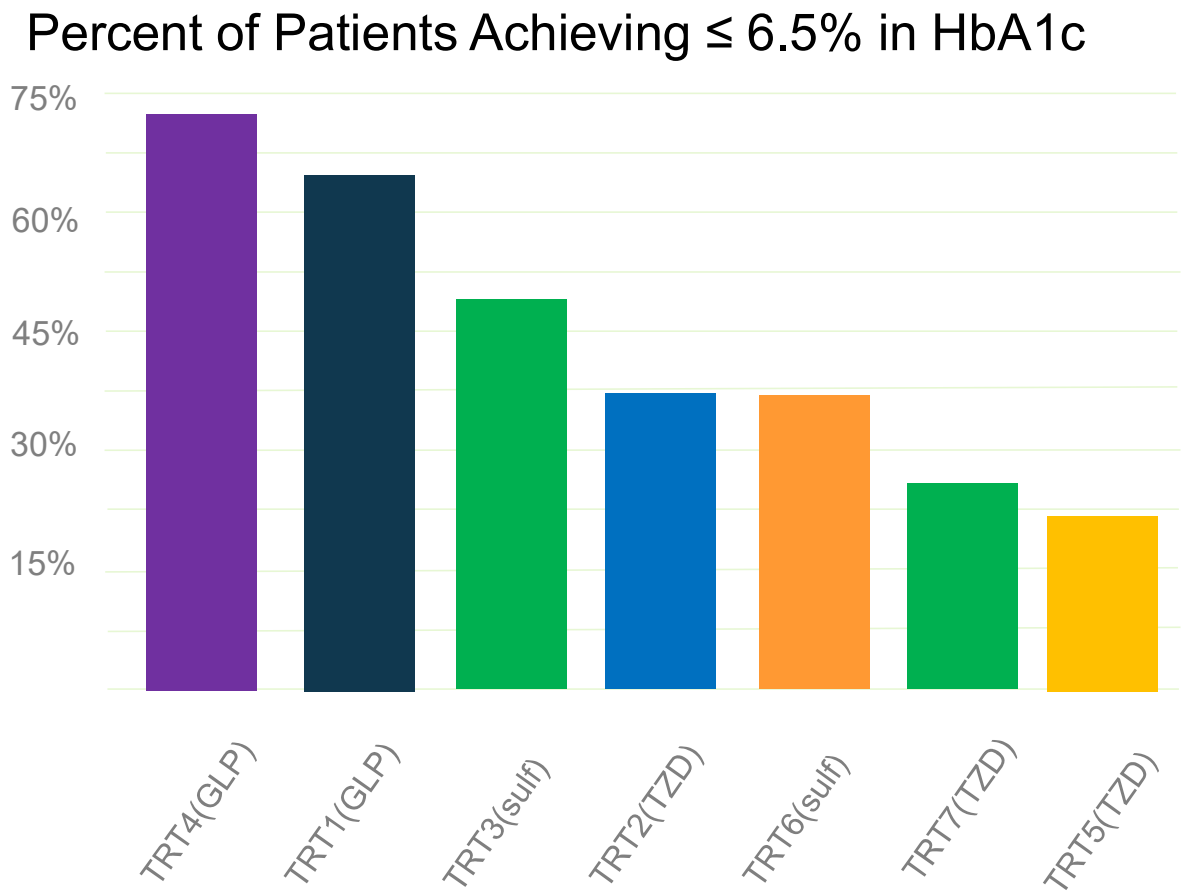
The Gestalt Law of Prägnanz



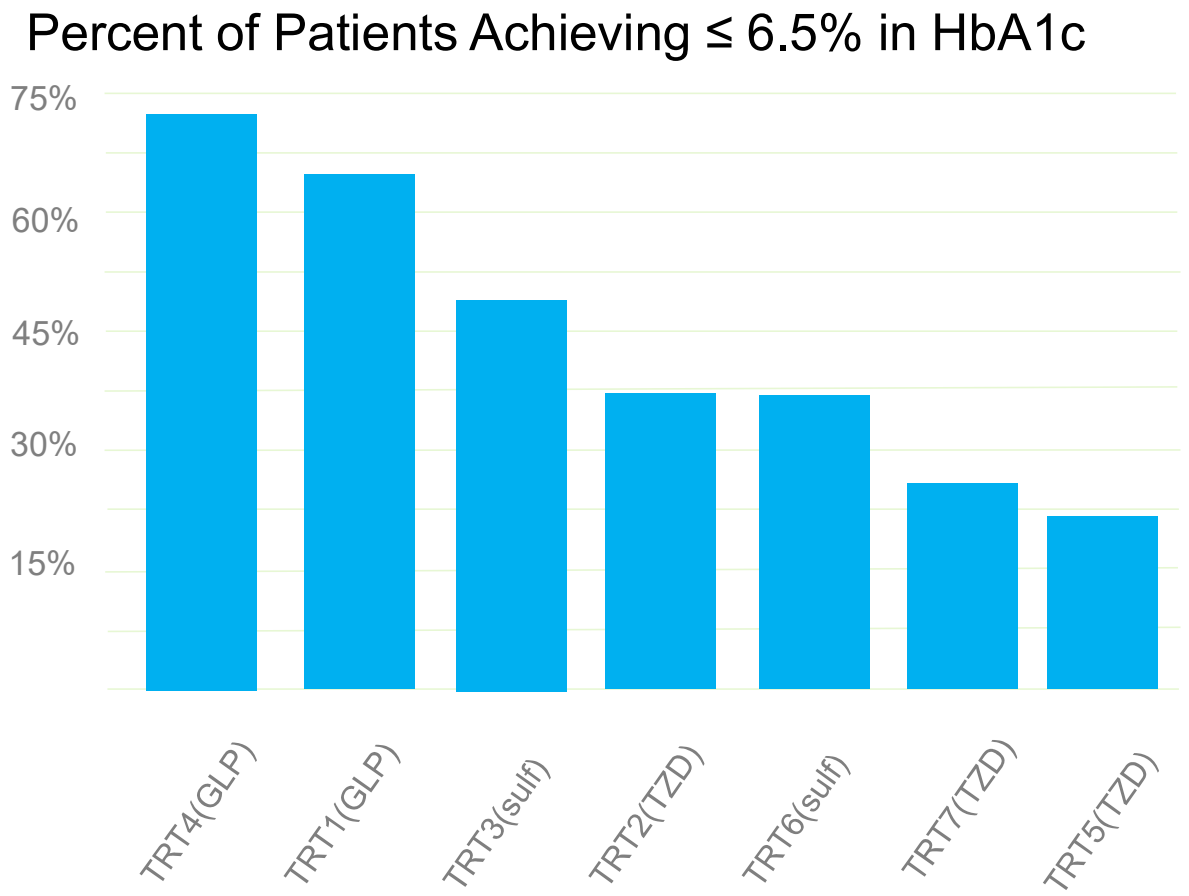
The Gestalt Law of Prägnanz



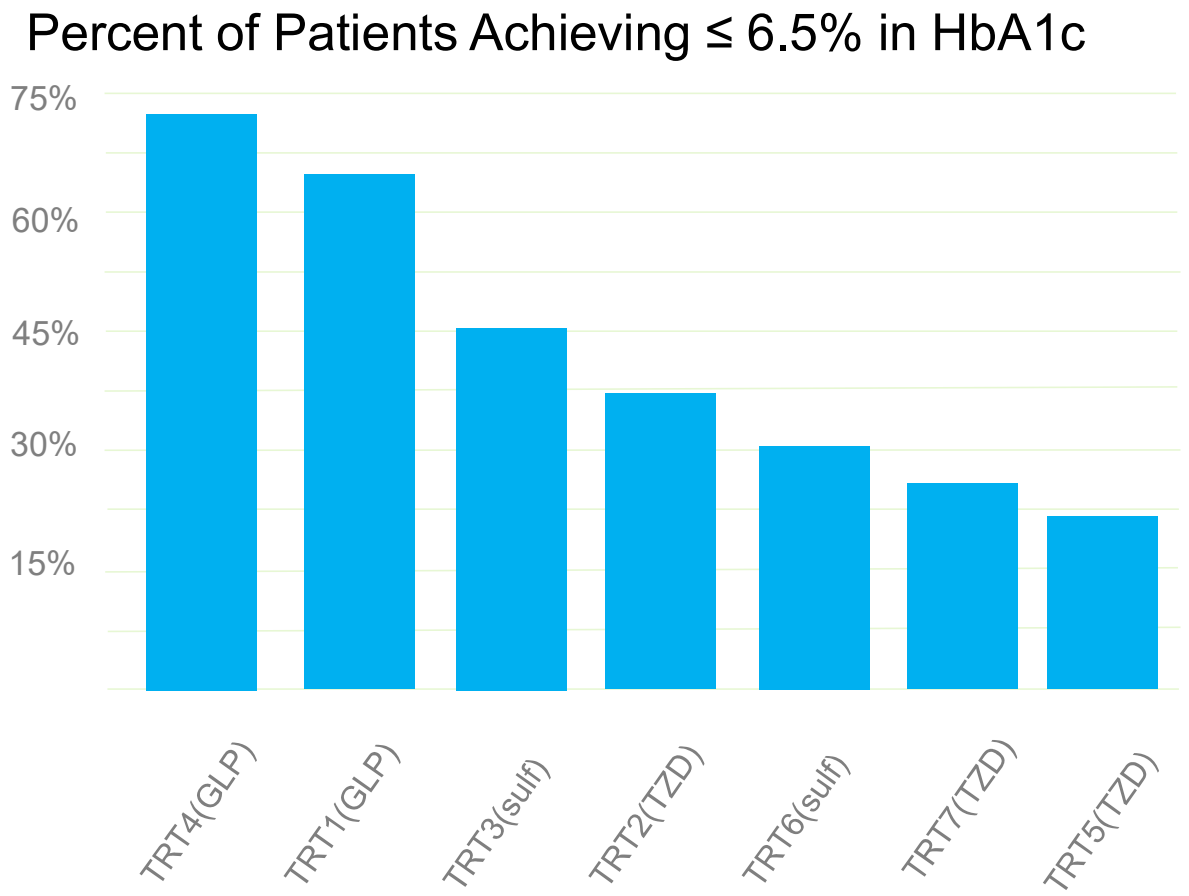
The Gestalt Law of Similarity



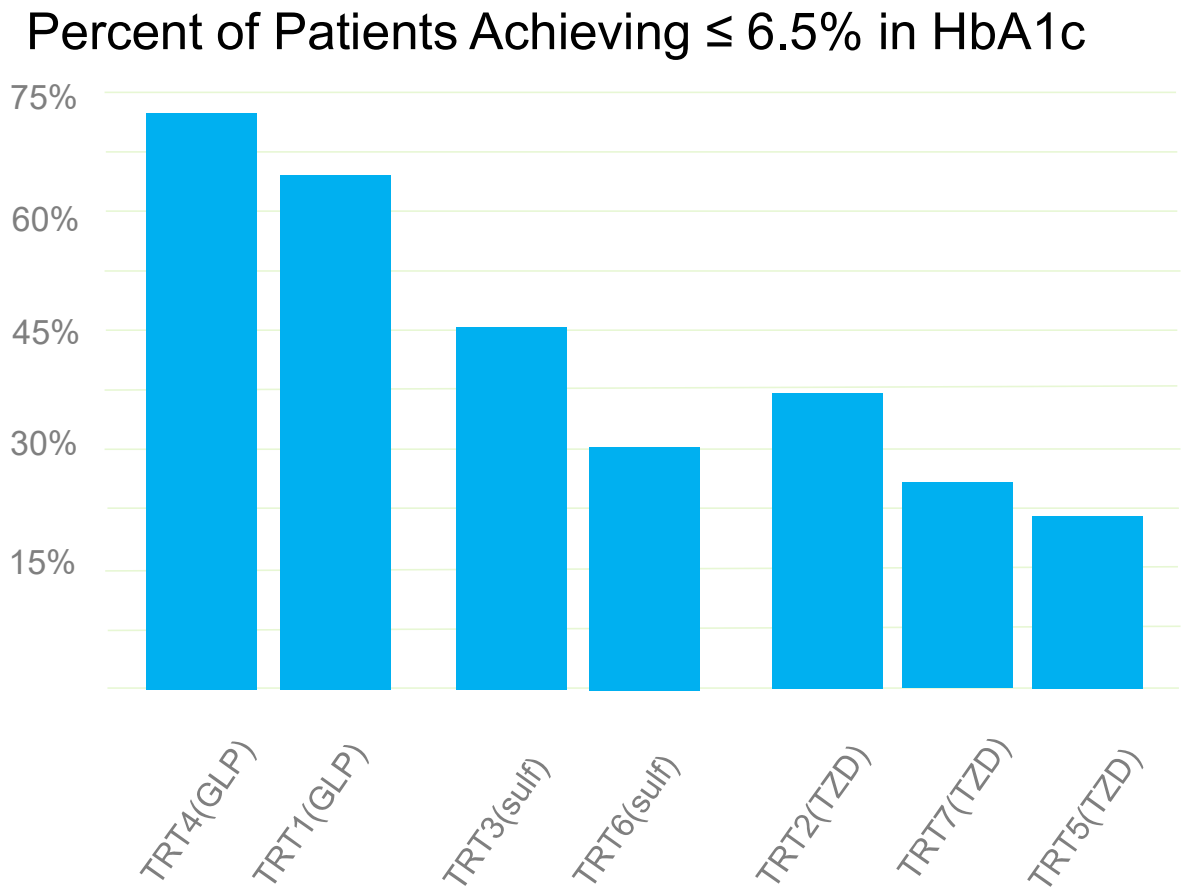
The Gestalt Law of Similarity



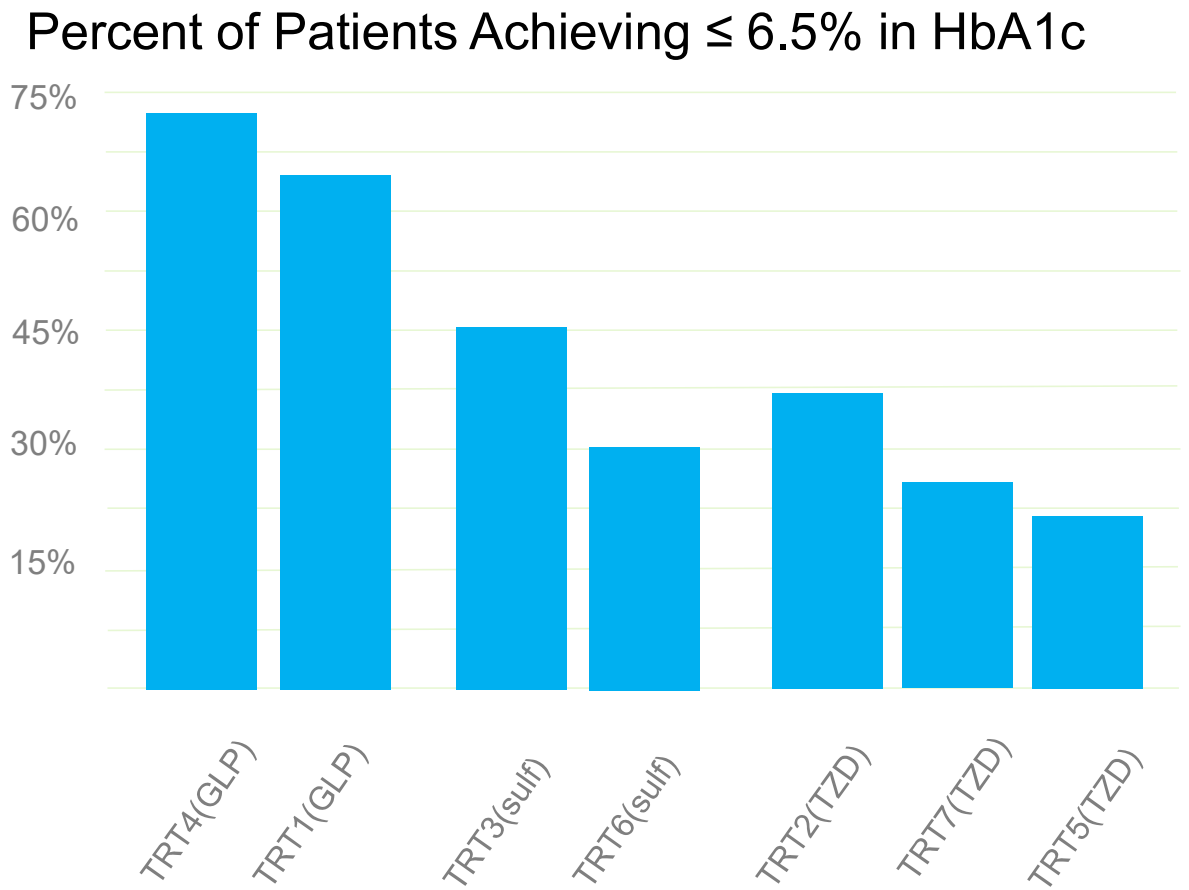
The Gestalt Law of Common Fate



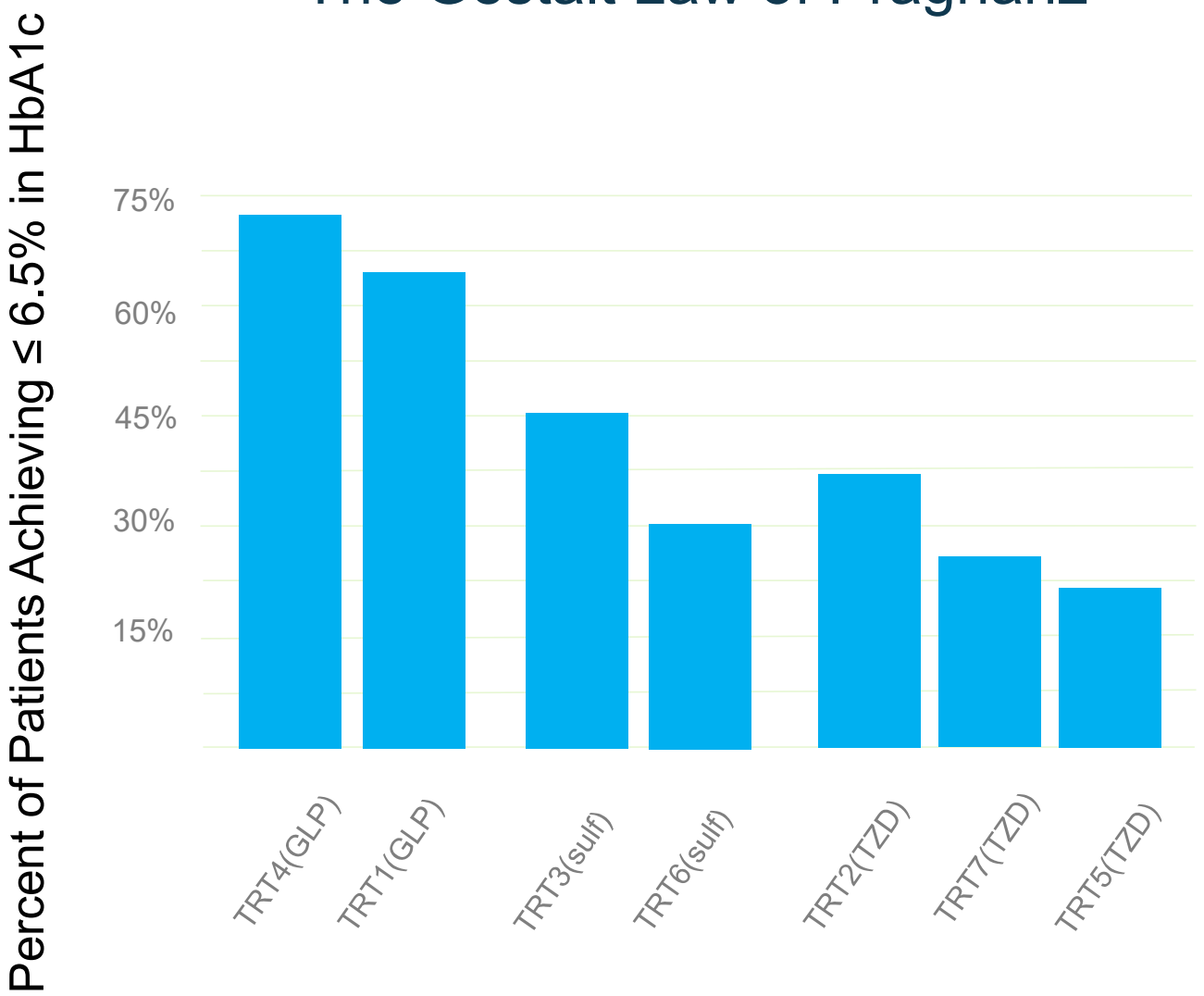
The Gestalt Law of Proximity



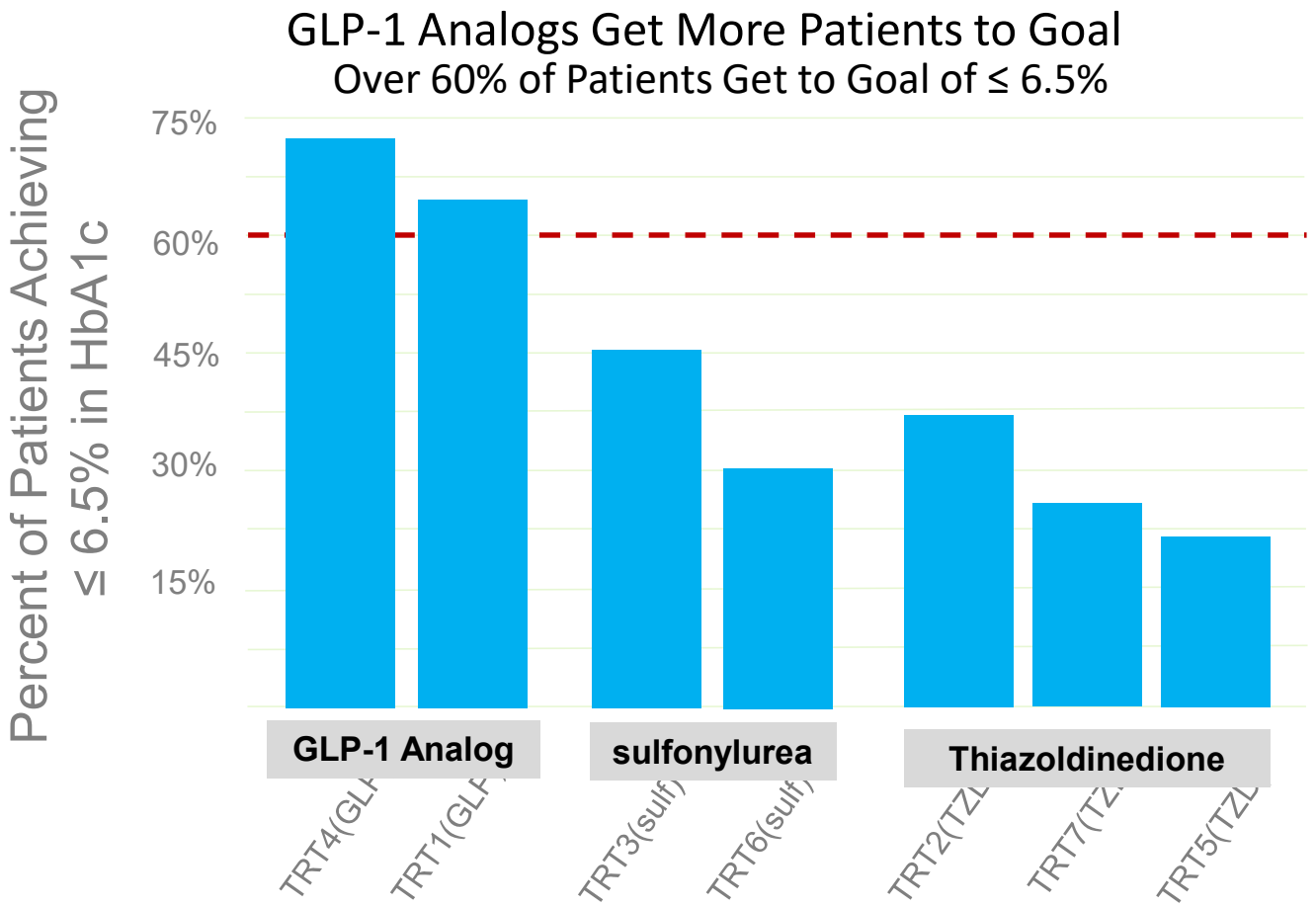
The Gestalt Law of Proximity



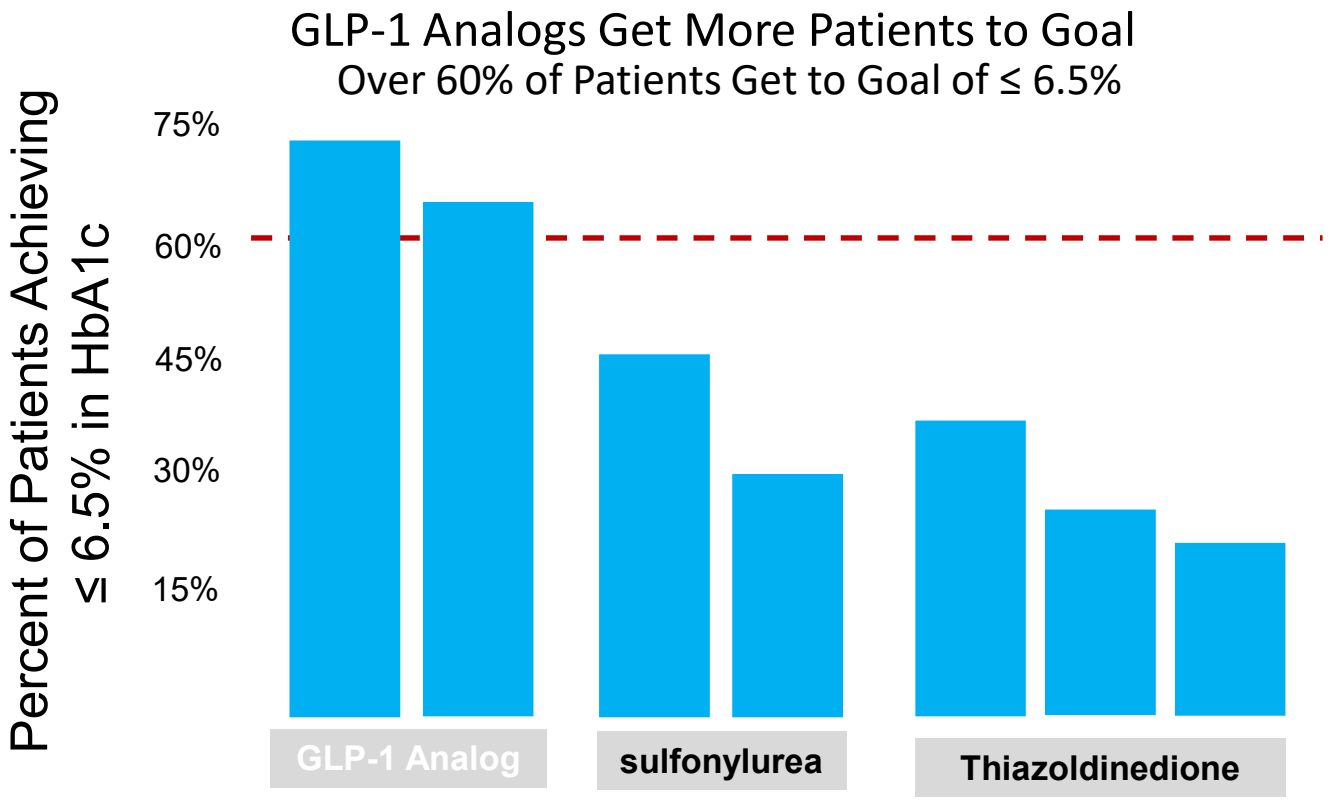
The Gestalt Law of Prägnanz



The Gestalt Law of Prägnanz

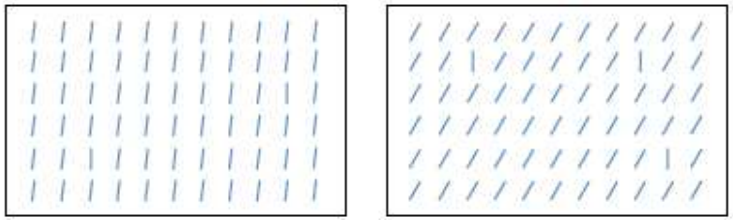


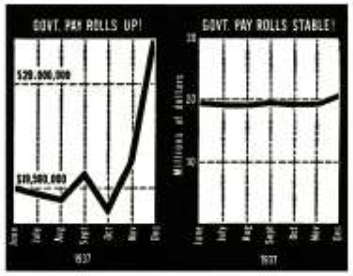
The Gestalt Law of Focal Point



A few lessons from cognitive psychology…

- Attention is drawn to large perceptible differences: humans think in terms of differences.
- People expect changes in properties to carry information.
- Form and meaning need to be compatible.
- People can only hold in mind up to four groups of information at once.
- People automatically group elements into units.
- Try to maximize data/ink ratio.
- When possible, interactivity is your friend.





Blue Red

[_] _][

XXOO XO XK

Hierarchy of Perception

More precise

quantitative

// Position

// Length

// Angle

// Direction

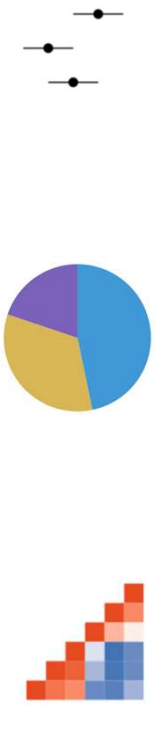
// Shape

// Area

// Volume

// Color saturation

// Color hue



Qualitative

// Color hue

// Orientation

// Shape

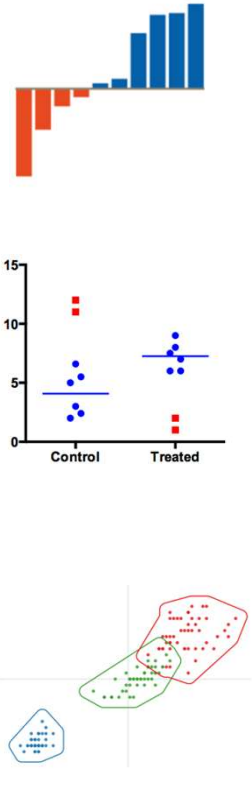
// Color intensity

// Size

// Curvature

// Added marks

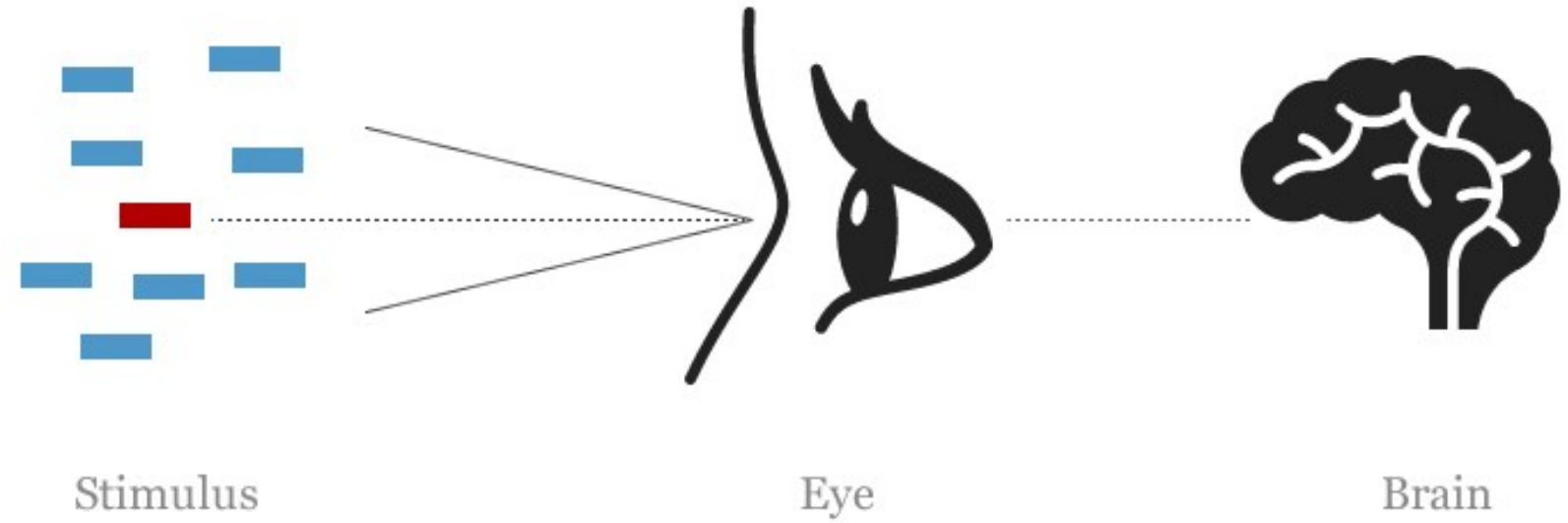
// Closure



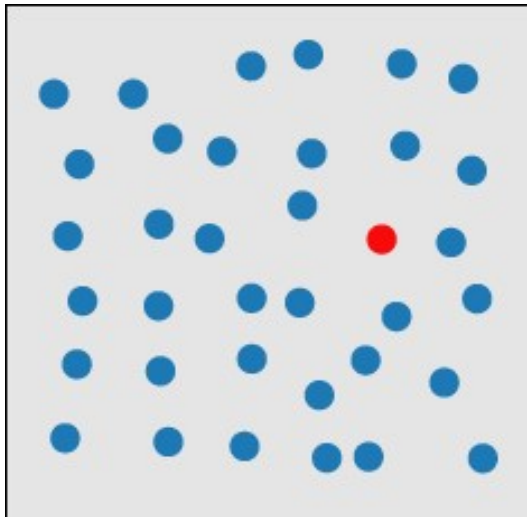
Less precise

Images borrowed with permission from “Graphical Principles Cheat Sheet” – Mark Baillie, Alison Margolskee, Baldur Magnusson, Andrew Wright, Ruquan You, Ivan-Toma Vranesic, Julie Jones, Marc Vandemeulebroecke

Pre-attentive Processing



Target Selection Visual Cue: Color



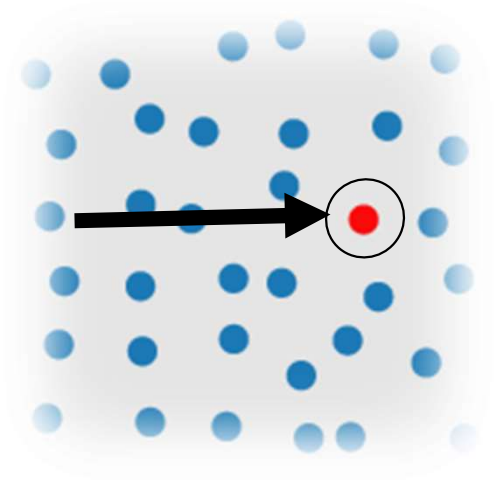
// a) Anomaly present

// b) Anomaly absent

Target Selection Visual Cue: Color

- a) Anomaly present
- b) Anomaly absent

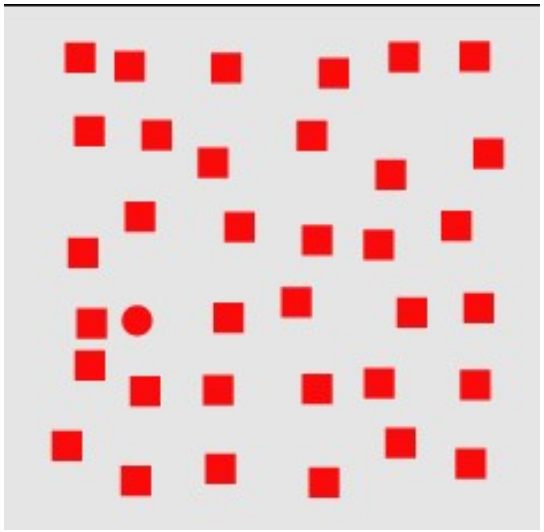
Target Selection Visual Cue: Color



// a) Anomaly present

// b) Anomaly absent

Target Selection Visual Cue: Shape



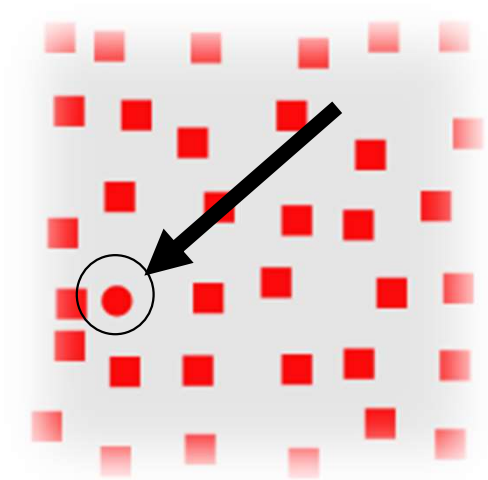
// a) Anomaly present

// b) Anomaly absent

Target Selection Visual Cue: Shape

- a) Anomaly present
- b) Anomaly absent

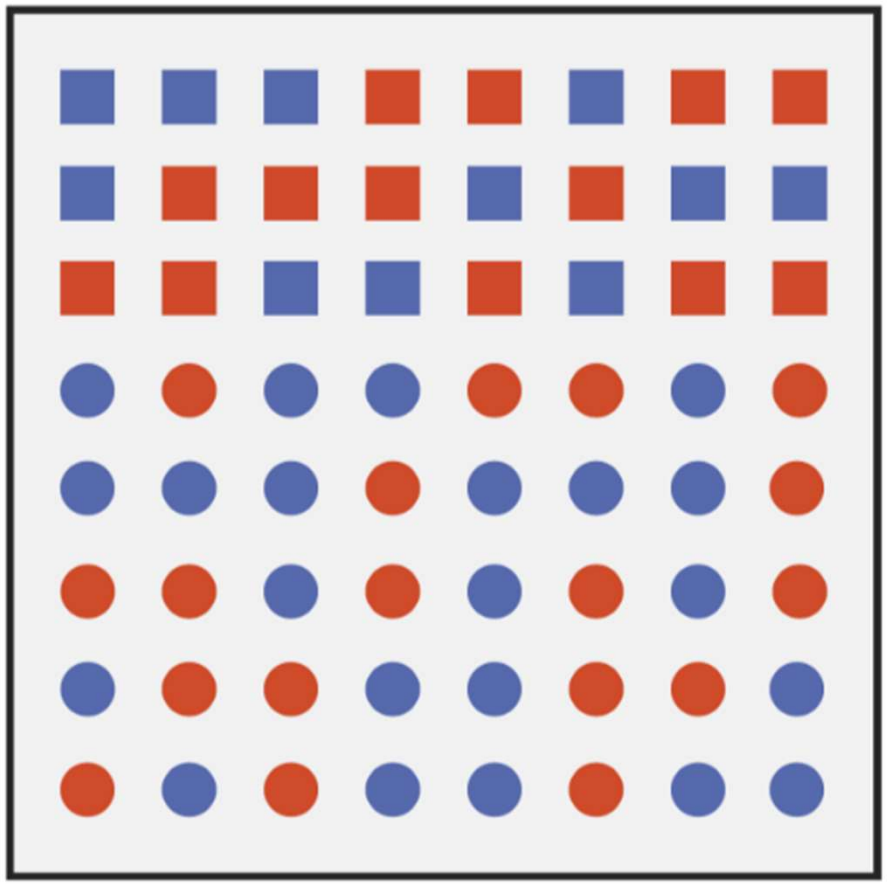
Target Selection Visual Cue: Shape

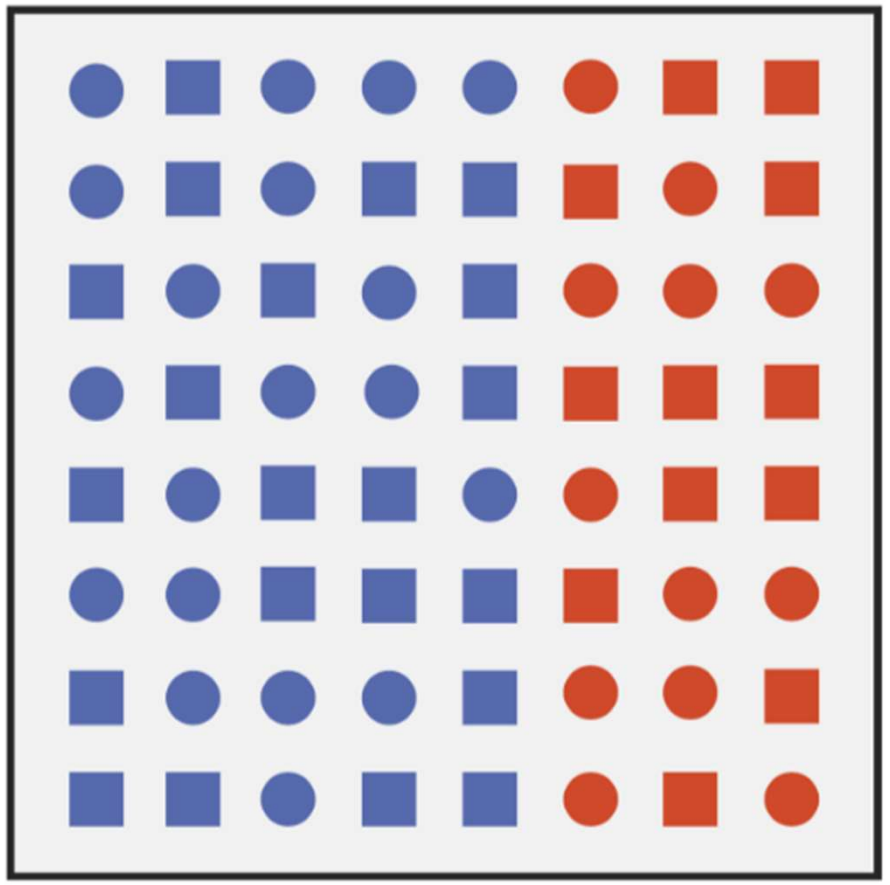


// a) Anomaly present

// b) Anomaly absent

Borders





Pre-attentive Processing

"Pre-attentive symbols become less distinct as the variety of distracters increases. It is easy to spot a single hawk in a sky full of pigeons, but if the sky contains a greater variety of birds, the hawks will be more difficult to see. A number of studies have shown that the immediacy of any pre-attentive cue declines as the variety of alternative patterns increases, even if all the distracting patterns are individually distinct from the target. "

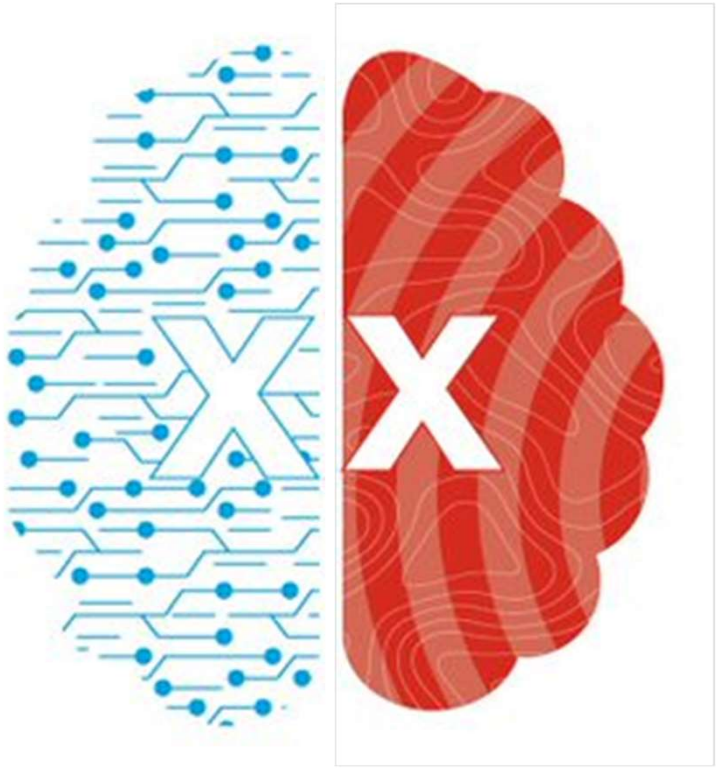
- Colin Ware (2000) "Information Visualization: Perception and Design".

Data Story Telling

Data Visualisation Principles

Visual Analytics and Data Story Telling

Exploration



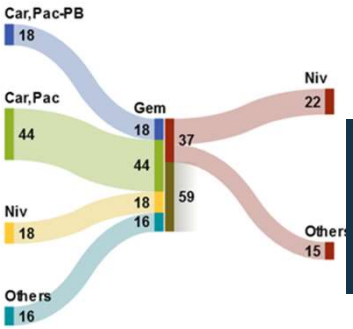
Explanation

Exploration

Interactive Data Visualization

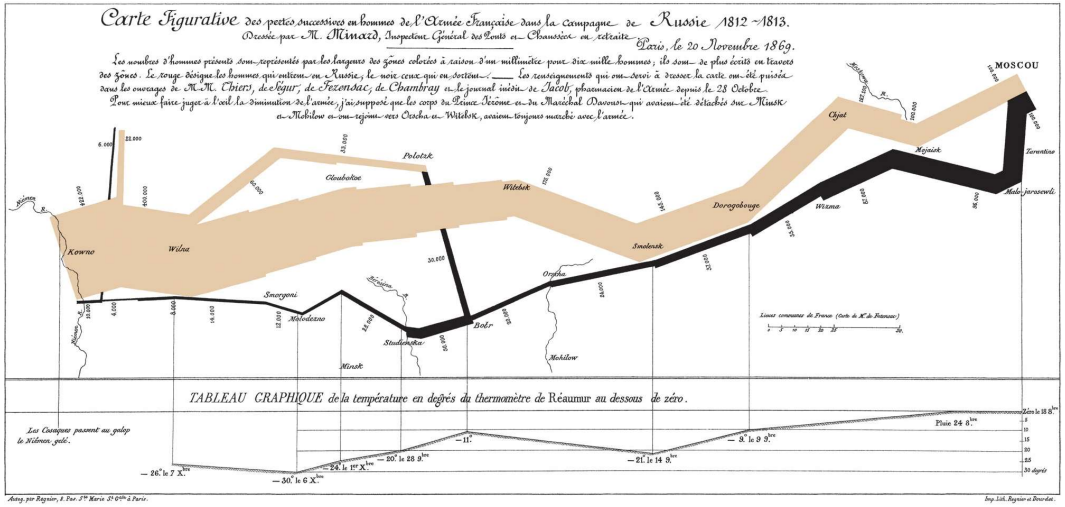
Tibco Spotfire

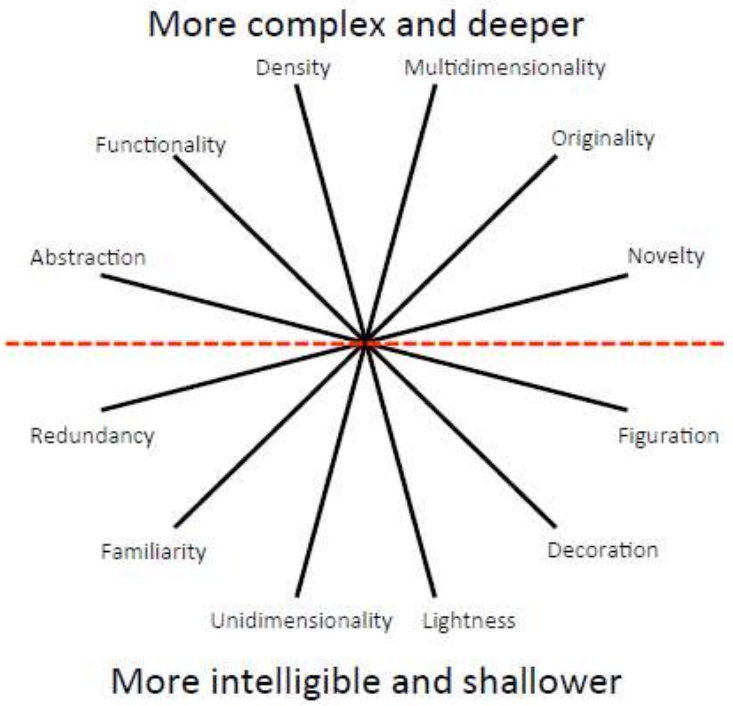
Automated Analyses



Squamous NSCLC

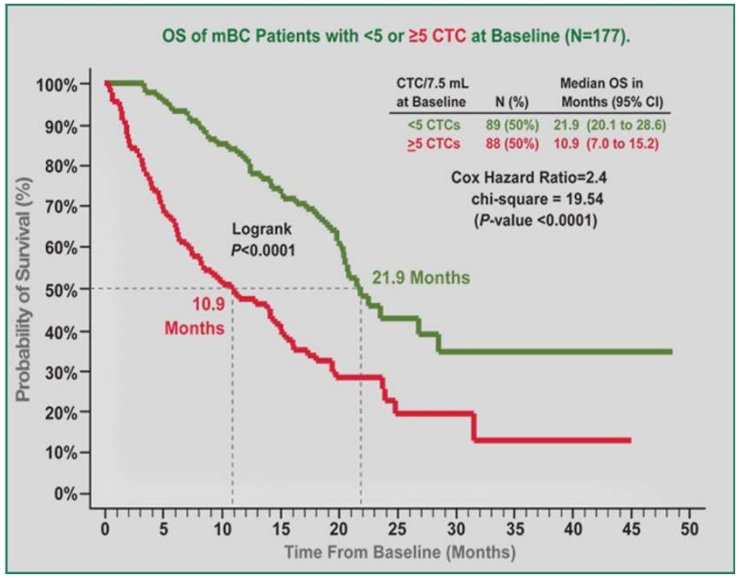
Explanation



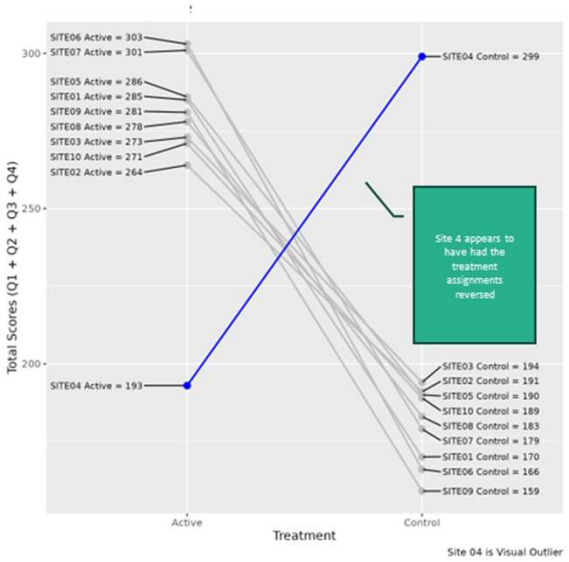


Definitions

// Statistical Graphic



// Data Story



Definitions

Statistical Graphic

- A statistical graph or chart is defined as the pictorial representation of statistical data in graphical form. The statistical graphs are used to represent a set of data to make it easier to understand and interpret statistical information.

Data Story

- Data
- Narrative
- Visualization

Clear Purpose

// Who is your audience?

// What is your venue?

// What is your key message?

Show the Data Clearly

- Minimize ink to data ratio
- Plot the quantity of interest. Don’t make people do mental math
- Use appropriate color schemes
- Do NOT distort the data, be accurate

Make the Message Obvious

// Effective Redundancy
// Judicious Use of Color
// Informative titles
// Annotation
// Apply Gestalt Principles and Pre-attentive processing

Putting it all together

- Group the data: visually (or explicitly!) segment the data into meaningful subsets.
- Plot the raw data whenever possible.
- Prioritize the data: rank the data by importance
- Sequence the data: give direction for the order in which the data should be read. **Storytelling!**
- Vertical and horizontal alignment of figures and/or text is important for clear visual flow and to facilitate comparisons (particularly across multiple graphs)
- Use the same scale for similar variables on different panels (small multiples) to **facilitate comparisons**.

Thank you!

Questions?