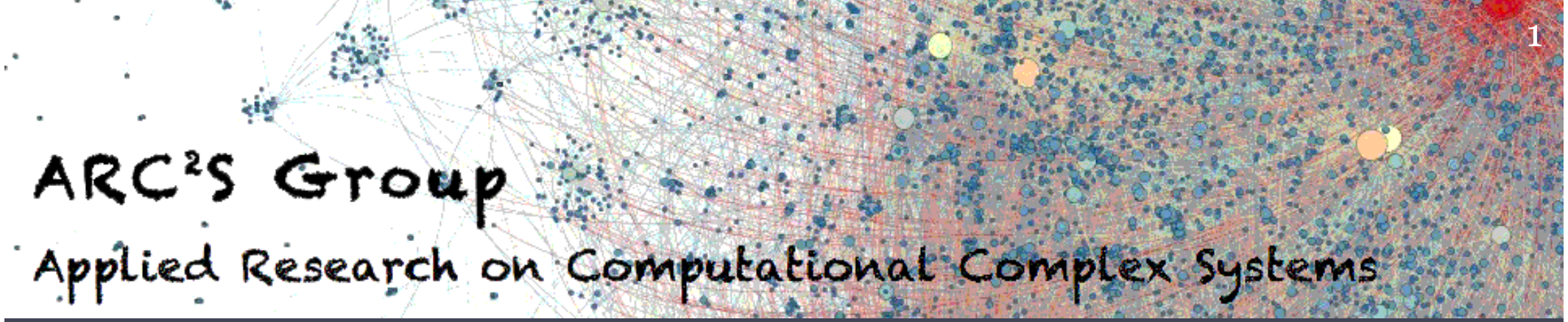


Introduction to Data Viz

Prof. Giancarlo **Ruffo**

"Analisi e Visualizzazione di Reti Complesse" (9 credits)
Laurea Magistrale in **Informatica**
Università degli Studi di Torino

@giaruffo

Agenda

Visualizzare dati

Il test dello Scimpanzé

Percezioni, inclinazioni, ignoranza degli strumenti matematici

Misurare le bugie nei grafici

Riferimenti

Visualizzare dati

Il potere della visualizzazione dati

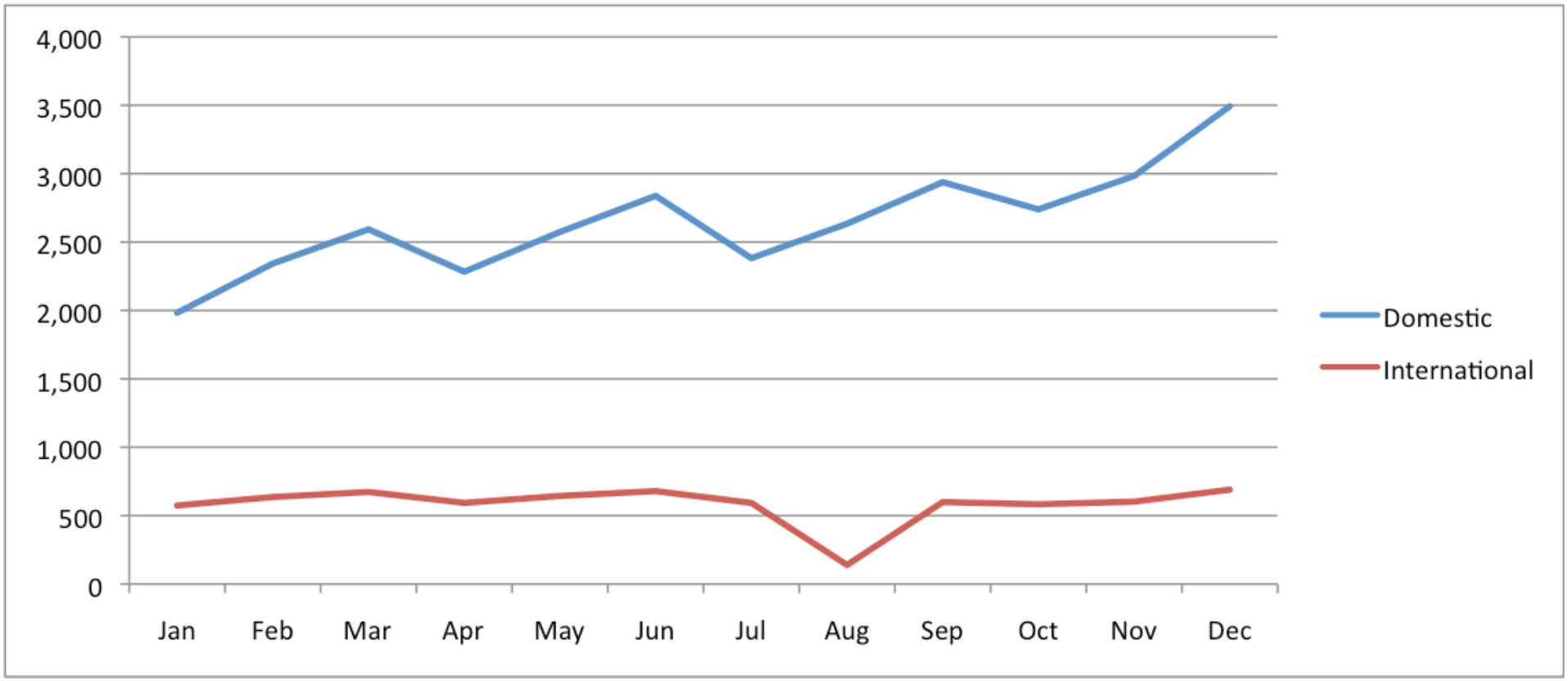
Le tabelle piene di informazioni funzionano perfettamente se abbiamo dei valori precisi e rendono facile l'identificazione di dati individuali

2007 Sales Revenue
(U.S. dollars in thousands)

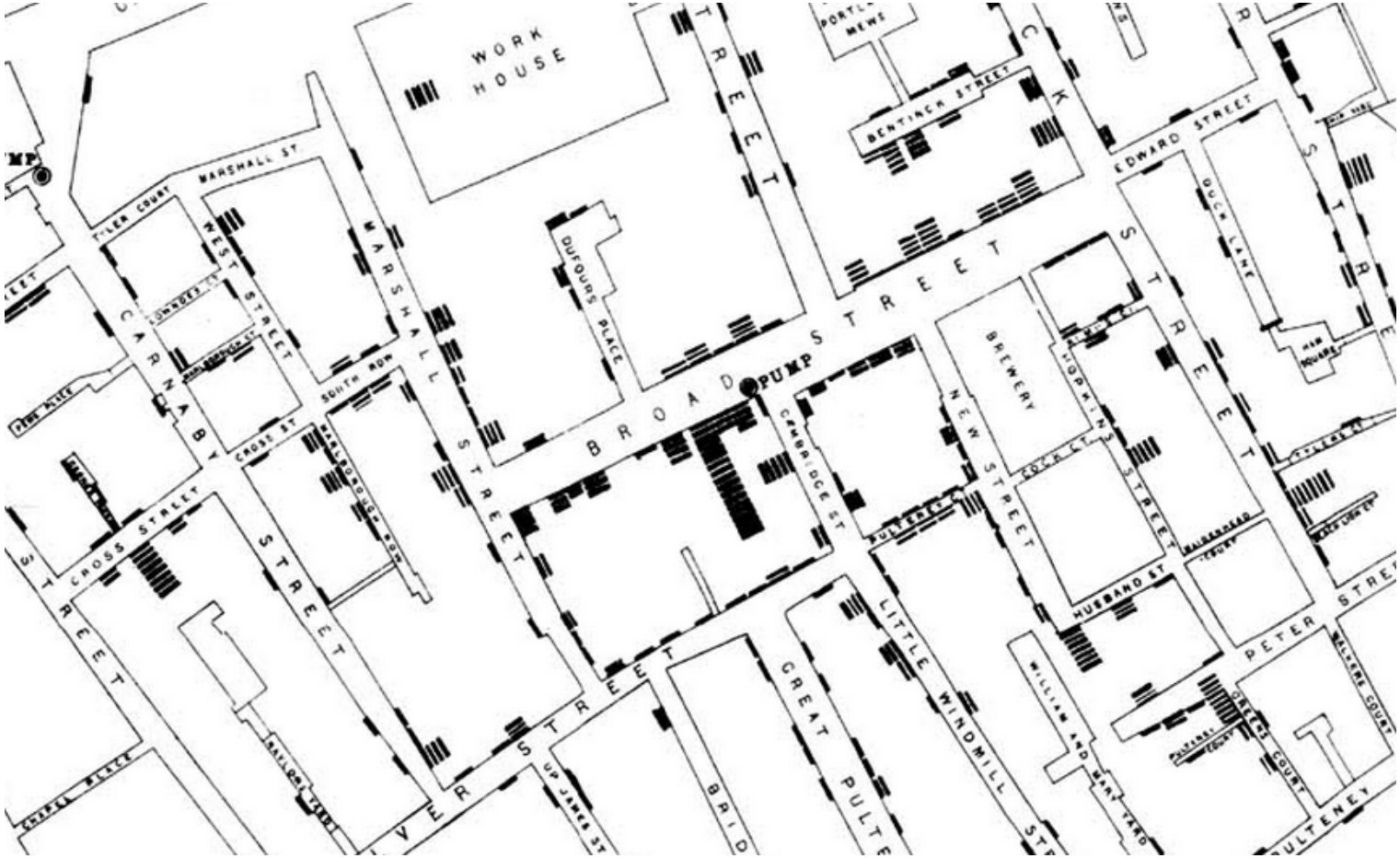
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Domestic	1983	2343	2593	2283	2574	2838	2382	2634	2938	2739	2983	3493
International	574	636	673	593	644	679	593	139	599	583	602	690
	$2,557	$2,979	$3,266	$2,876	$3,218	$3,517	$2,975	$2,773	$3,537	$3,322	$3,585	$4,183

Cosa fare quando vogliamo confrontare i dati, trovare "pattern", tendenze, relazioni, etc.?

2007 Sales Revenue (esempio)



John Snow's data journalism: the cholera map that changed the world



http://www.theguardian.com/news/datablog/2013/mar/15/john-snow-cholera-map

Le “visualizzazioni” ci aiutano ad interpretare dati e numeri.

I numeri sono “fatti”, “osservazioni”.
Non possono mentire

Ma…

The best stats you've ever seen
Hans Rosling

https://www.ted.com/talks/hans_rosling_shows_the_best_stats_you_ve_ever_seen

The chimpanze test: I nostri pregiudizi peggiorano la nostra capacità di analisi

Dobbiamo approcciare la visualizzazione e la comunicazione scientifica nel modo più “onesto” possibile: prima ancora di mentire agli altri, rischiamo di mentire a noi stessi.

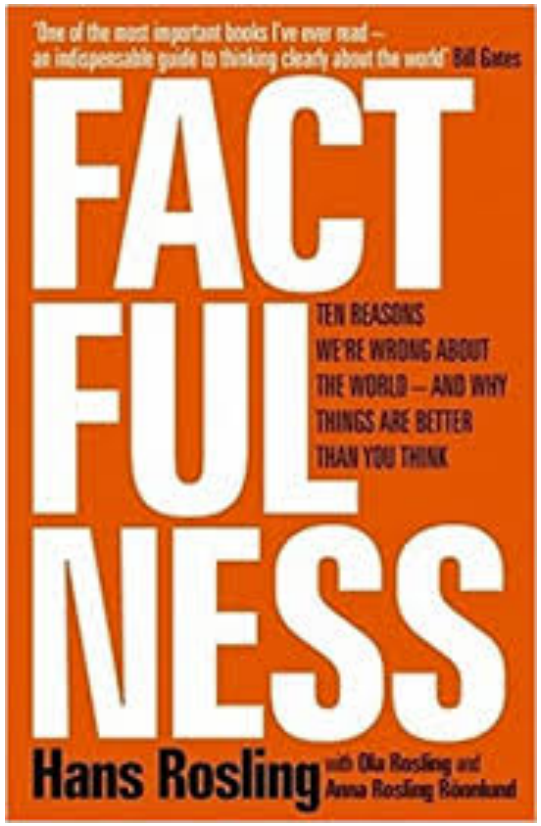
… non abbiamo neanche iniziato a parlare dei limiti della nostra percezione.

Lezione #1

Non fidatevi dei vostri pregiudizi e fate parlare i dati (senza torturarli)

"LET MY DATASET CHANGE YOUR MINDSET"

Hans Rosling
(1948-2017)



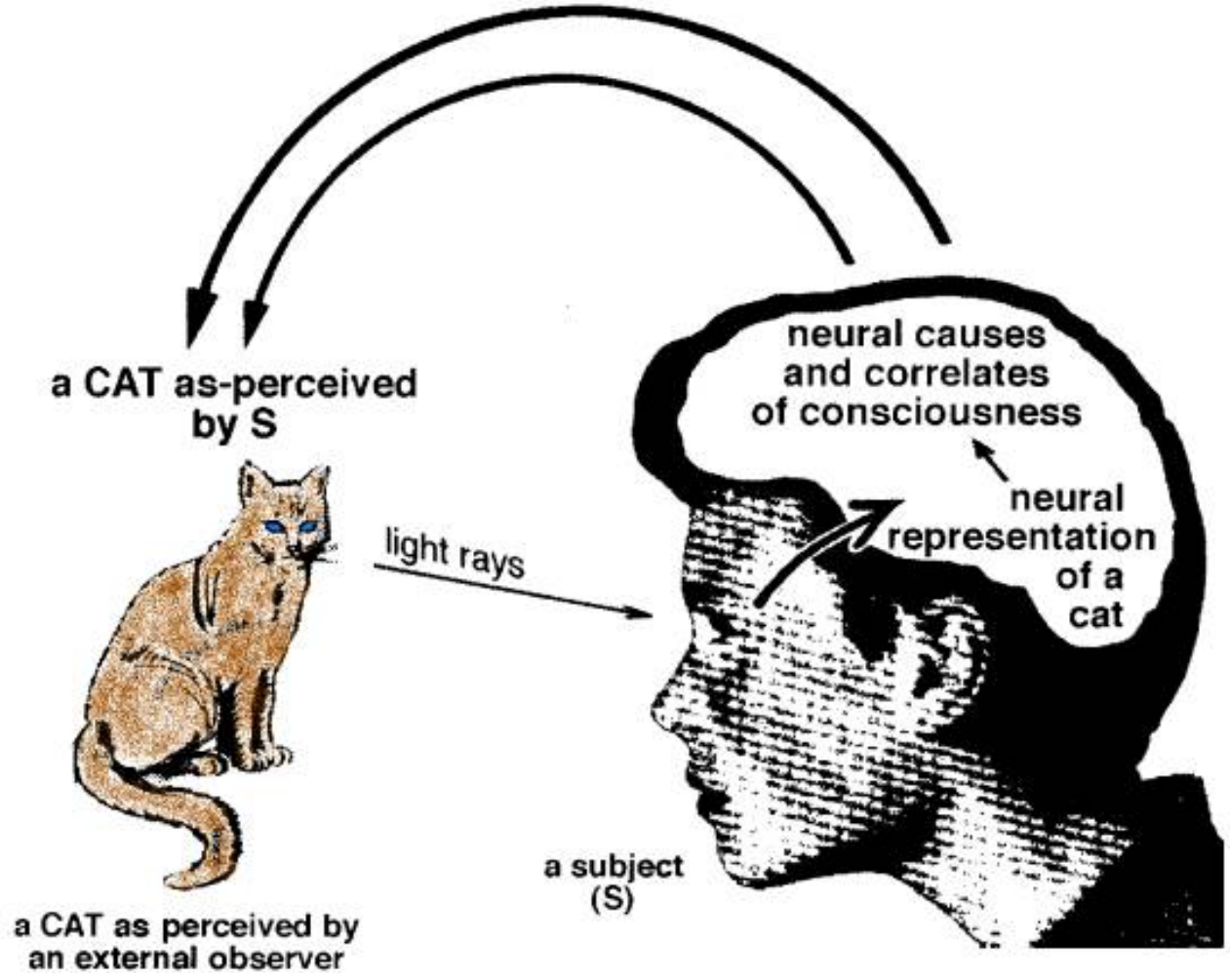
Usare le visualizzazione in modo efficace

"We must do more than simply display data graphically: we must understand how visual perception works and then present data visually in ways that follow the rules"

Stephen Few, Now You See It, Analytics Press
Chapter 3: Thinking with our eyes

Percezione, inclinazioni, ignoranza degli strumenti matematici

Come funziona la percezione visiva



Cosa possiamo percepire?

percepiamo un intero oggetto nella sua totalità, ma possiamo anche percepirlo come composto di diverse proprietà visive

posizione 2D, lunghezze, ampiezze, profondità, forma, colore orientamento, …

Possiamo potenzialmente “catturare” molti valori in una sola volta (i valori possono essere combinati in “pattern”)

Quanti 3 vedete nella sequenza sottostante?

248134811871167155413881984437713479156415318453058486
234757894114841222388146916135480484078908770786787512
865842340443771347915641531845305848641234757894114841
238814691613548048407890877078678751211865842340188742

Quanti 3 vedete nella sequenza sottostante?

2481348118711671554138819844377134791564153184530584864123475789411484122238814691613548048407890877078678751211865842340443771347915641531845305848641234757894114841222388146916135480484078908770786787512118658423401887427623881469161354804840789087707867875121186584234018874276

Lezione #2

"Non limitiamoci a tutto quello che possiamo vedere. La percezione visiva è selettiva ed è giusto che sia cosí, dato che ci obbliga a concentrarci su ciò che potrebbe sopraffarci. La nostra attenzione si focalizza spesso su ciò che si differenzia dall'ordinario"

Cosa vedete qui?

Cos’altro vedete qui?

Okay, c’è una rosa. Cos’altro vedete dentro la rosa?

Okay, facciamo tutti uno sforzo per vedere un **delfino** dentro la rosa…

Lezione #3

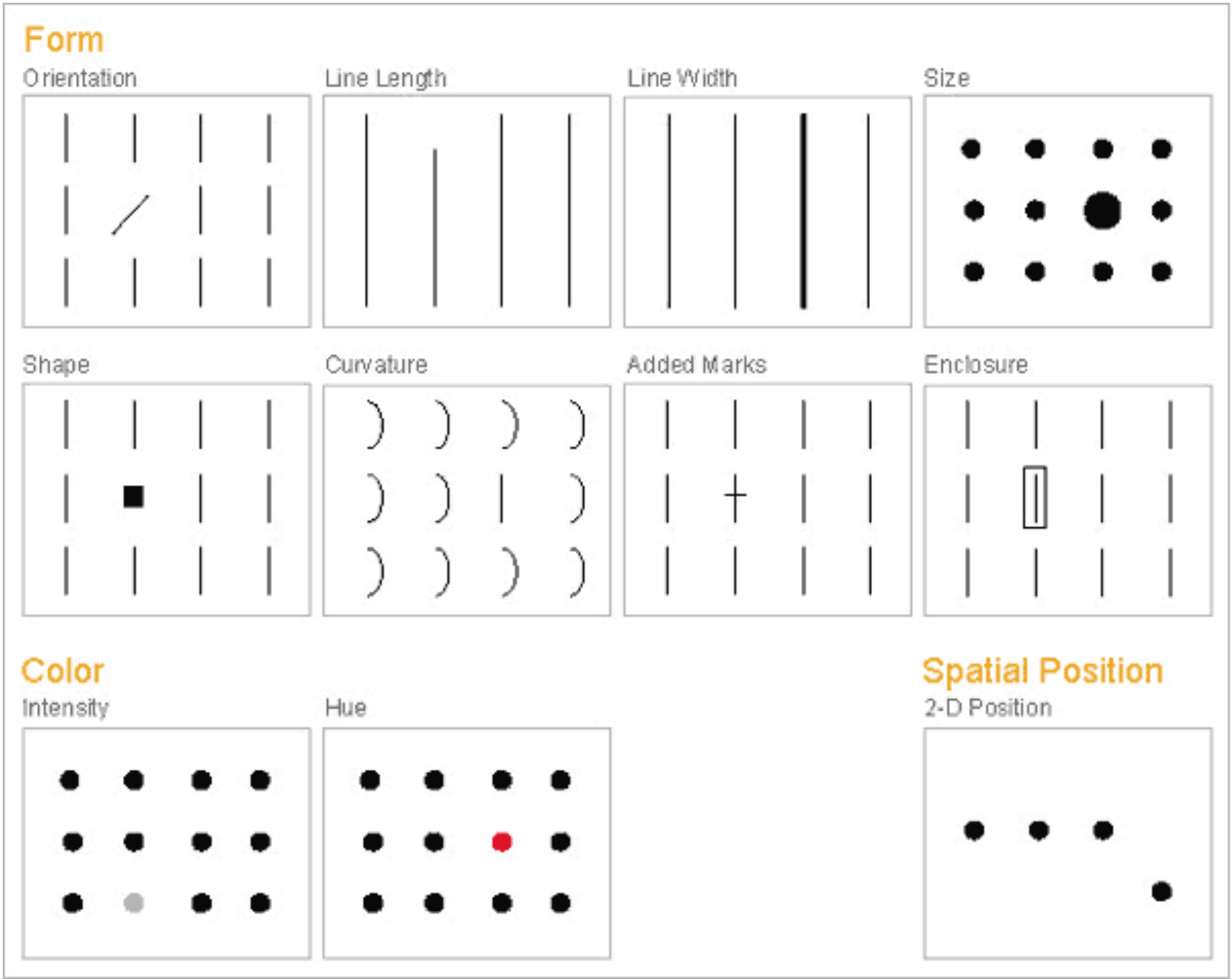
"I nostri occhi sono "guidati" da figure familiari. Vediamo ciò che conosciamo e ci aspettiamo di vedere"

Lezione #4

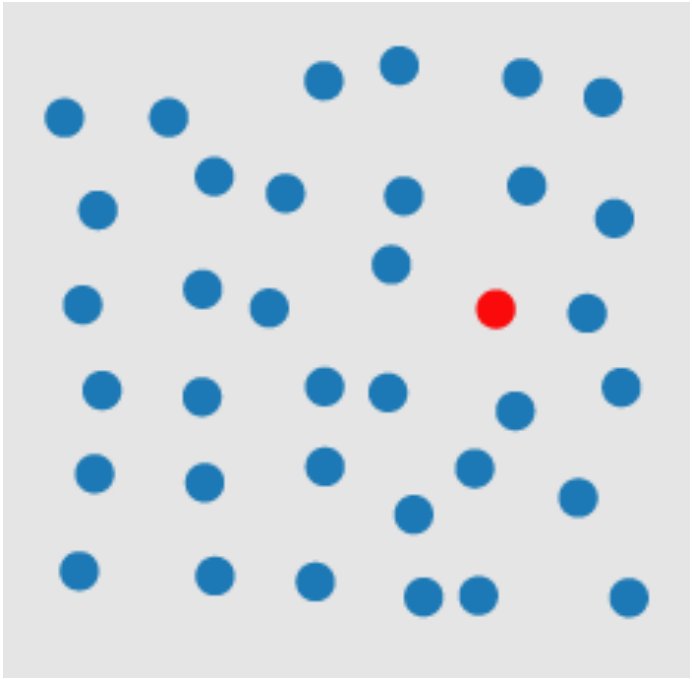
"La memoria gioca un ruolo fondamentale nella cognizione umana, ma la memoria utile è estremamente limitata"

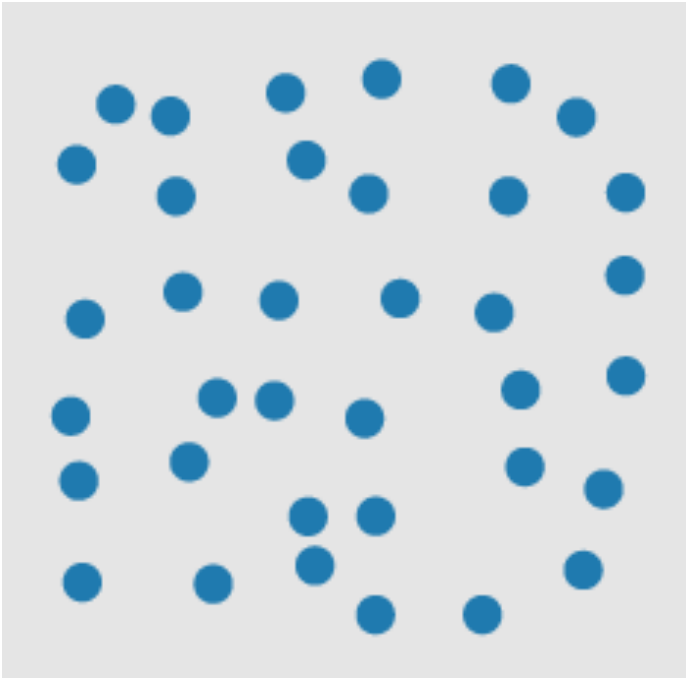
Attributi “pre-attentive”

- Quattro categorie
  - Colori
  - Forme
  - Posizioni
  - Animazioni
- Associa valori e dati ad attributi “pre-attentive” con cautela:
  - Non creare DISTRAZIONI dai dati
  - Tieni in considerazione la memoria a breve termine

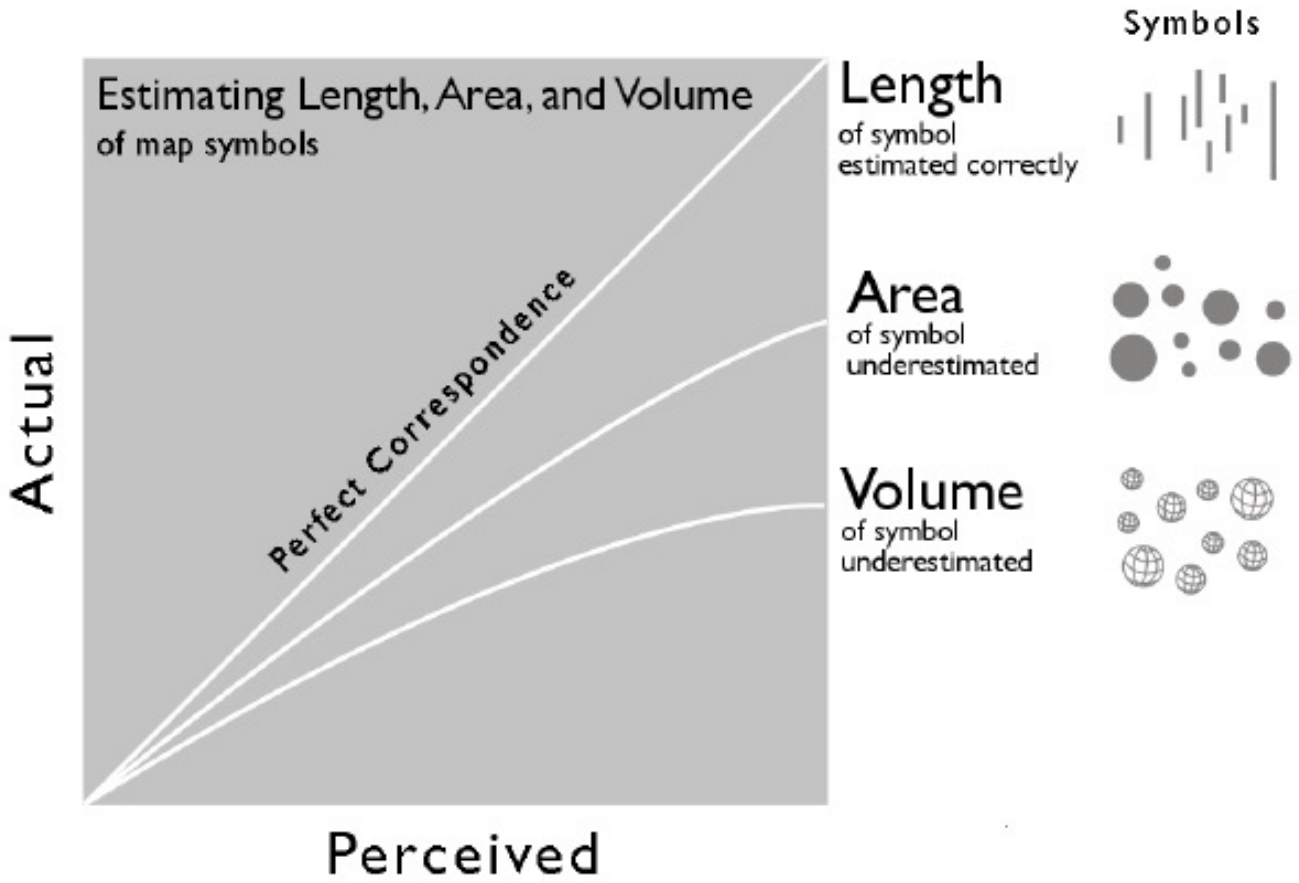

Form
Orientation
Line Length
Line Width
Size
Shape
Curvature
Added Marks
Enclosure
Color
Intensity
Hue
Spatial Position
2-D Position

Esempio





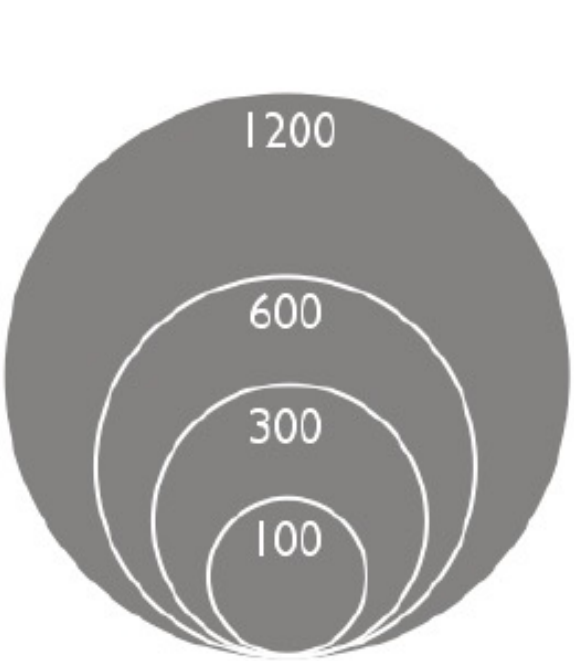
Forme



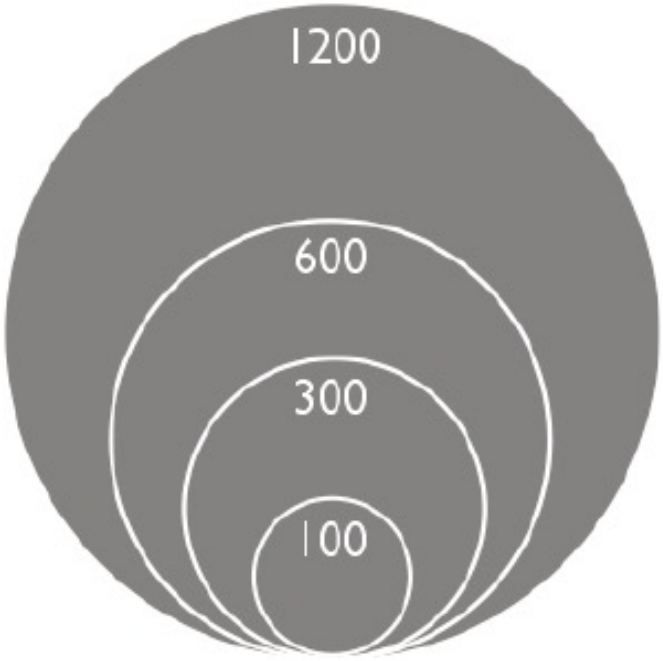
http://makingmaps.net/2007/08/28/perceptual-scaling-of-map-symbols/

Forme

- Dimensioni effettive e percepite nei cerchi



Absolute Scaling

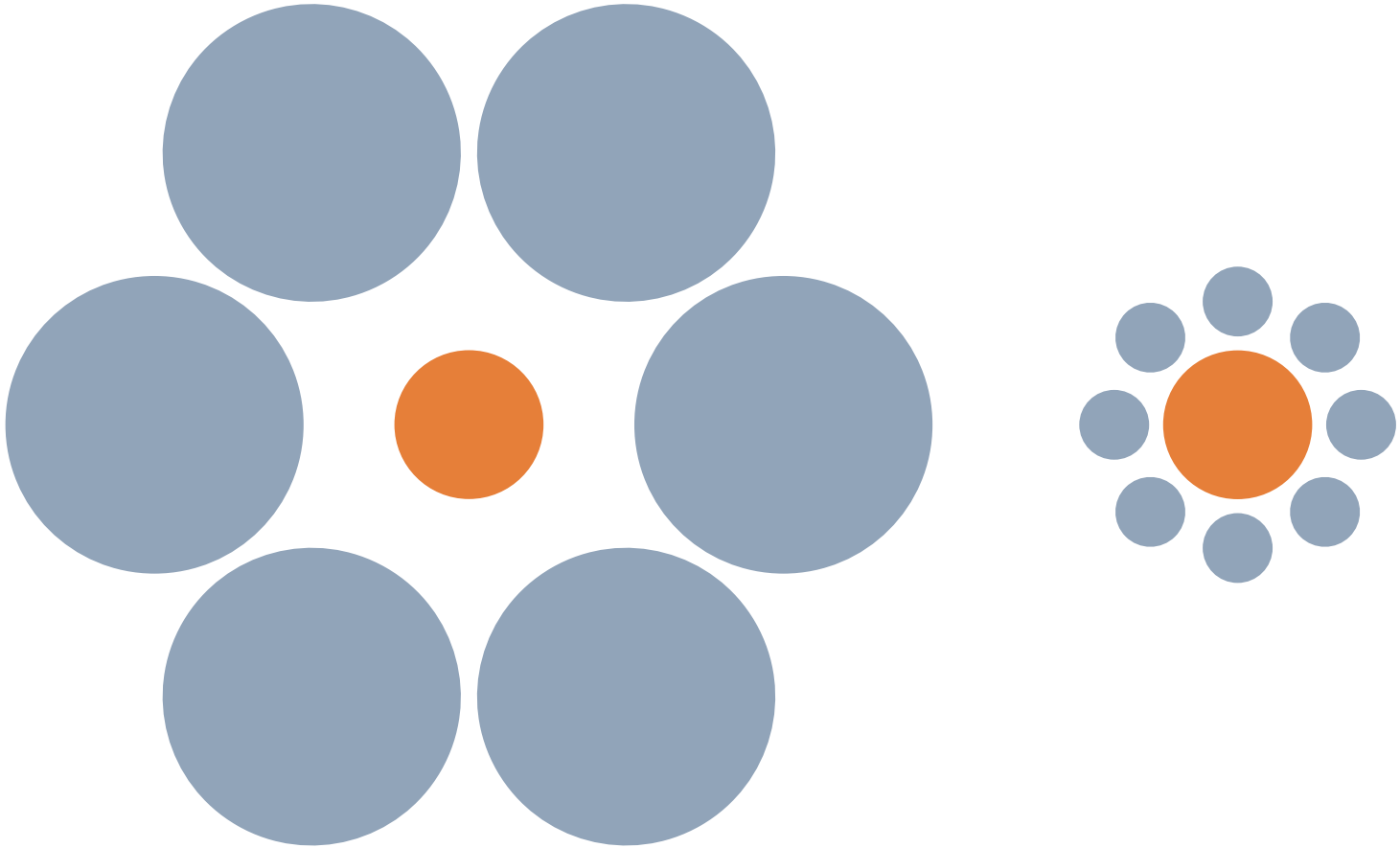


Apparent Scaling
(Flannery's Compensation)

http://makingmaps.net/2007/08/28/perceptual-scaling-of-map-symbols/

Forme

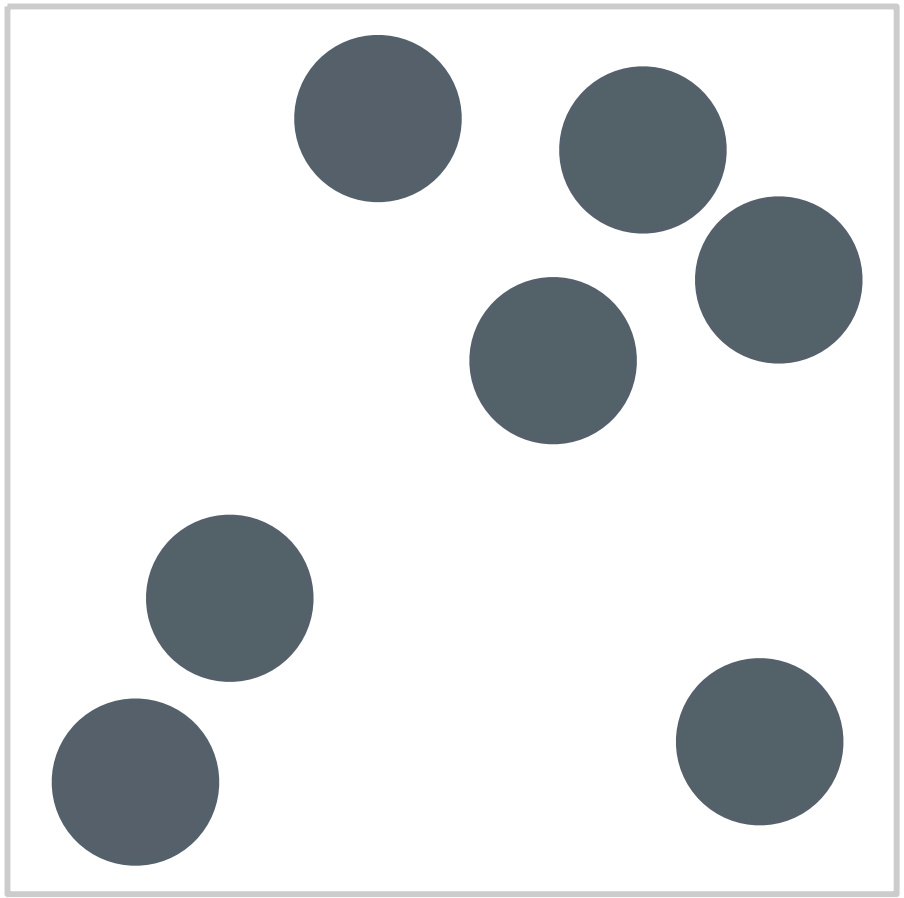
- Dimensioni effettive e percepite nei cerchi

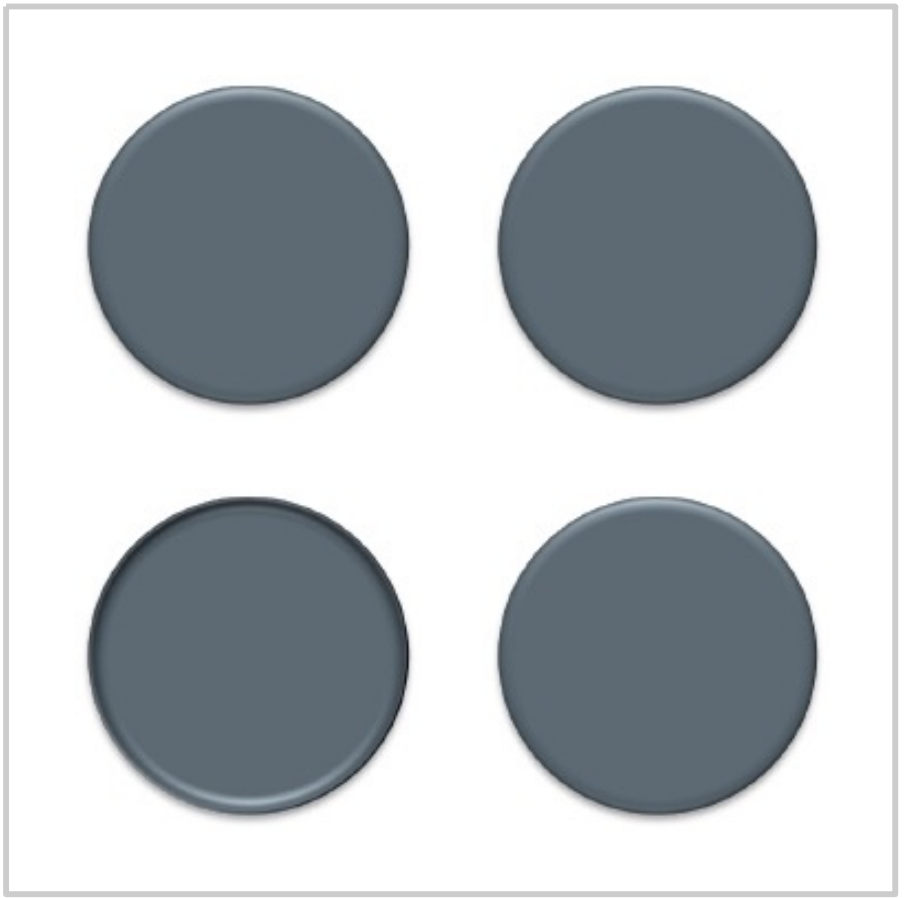


http://en.wikipedia.org/wiki/Ebbinghaus_illusion

Posizioni

- Posizioni 2d, figure concave e convesse





Animazioni

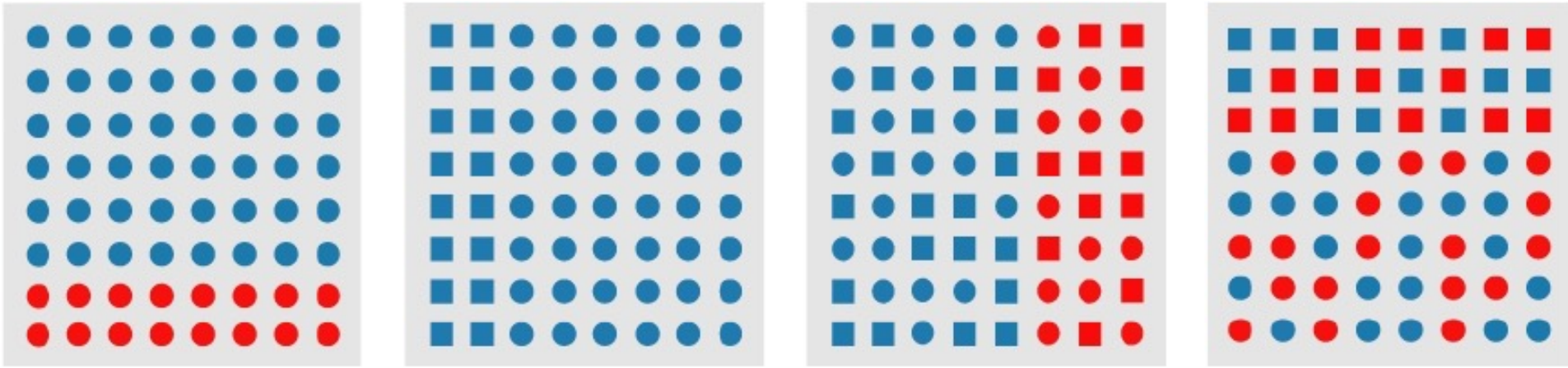
- Due attributi
  - Intermittenza (compare/scompare)
  - Movimento (da una posizione all'altra)
- Uno dei modi più efficaci per attirare l'attenzione
- Molto spesso abusato nel marketing

Attenzione e movimento

https://www.youtube.com/watch?time_continue=2&v=vJG698U2Mvo

https://www.youtube.com/watch?v=IGQmdoK_ZfY

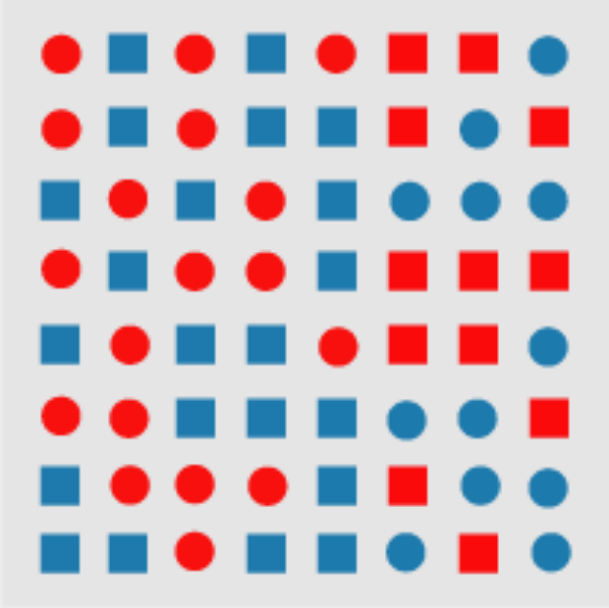
Contorni e margini



http://www.csc.ncsu.edu/faculty/healey/PP/

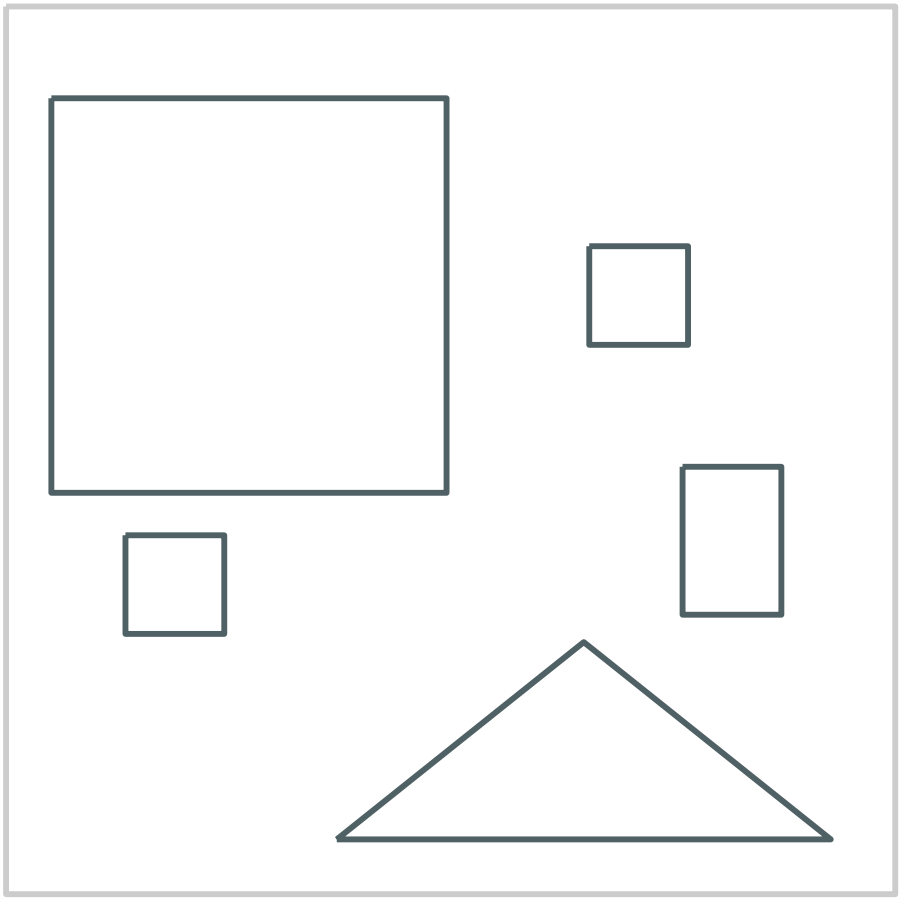
Contorni e margini

Un contorno definito da due o più attributi diventa difficile da identificare

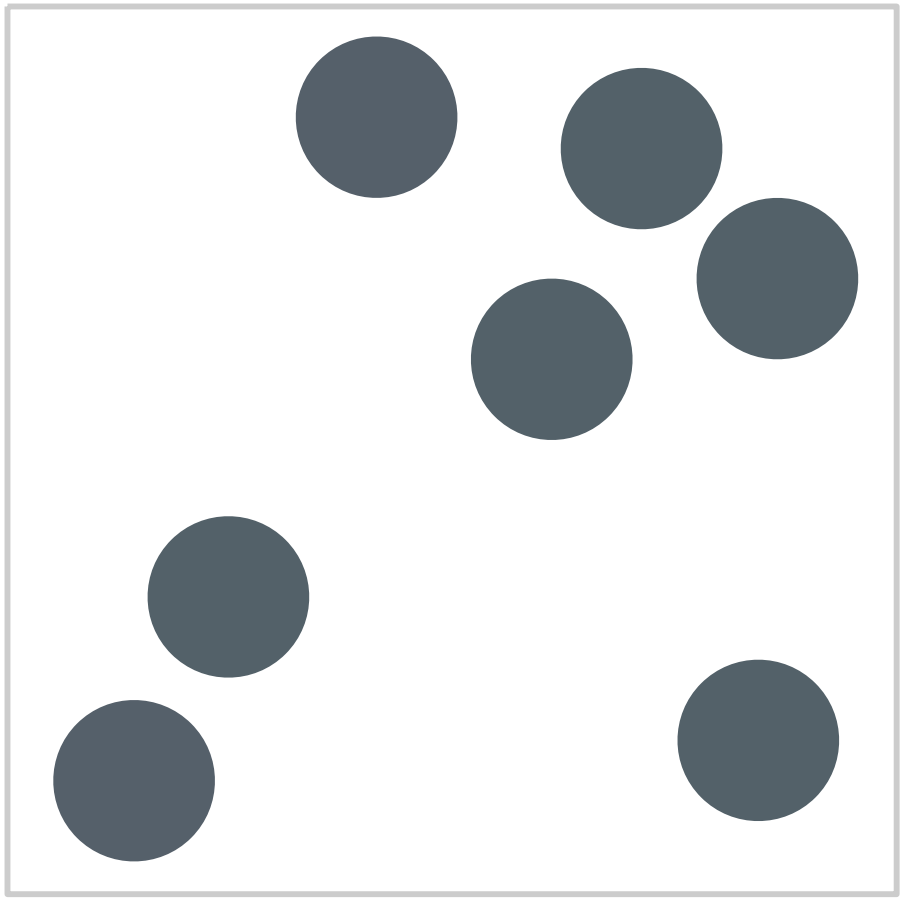


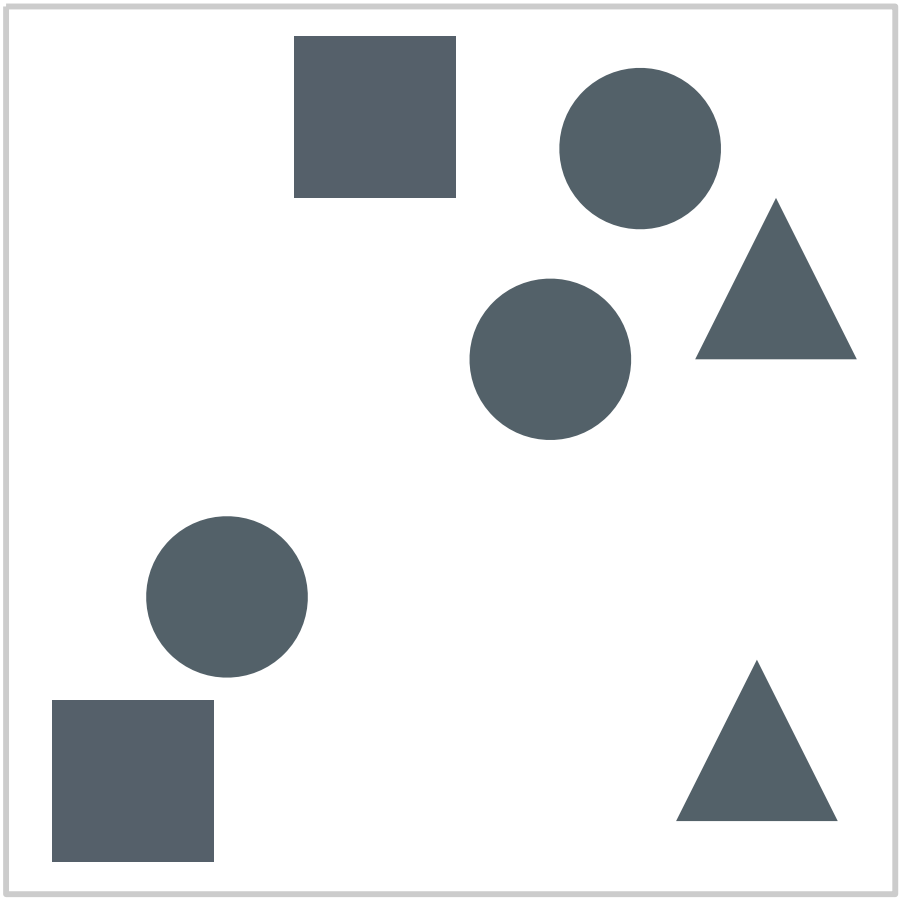
Gestalt

- Tendenza a paercepire oggetti come parti di un tutto

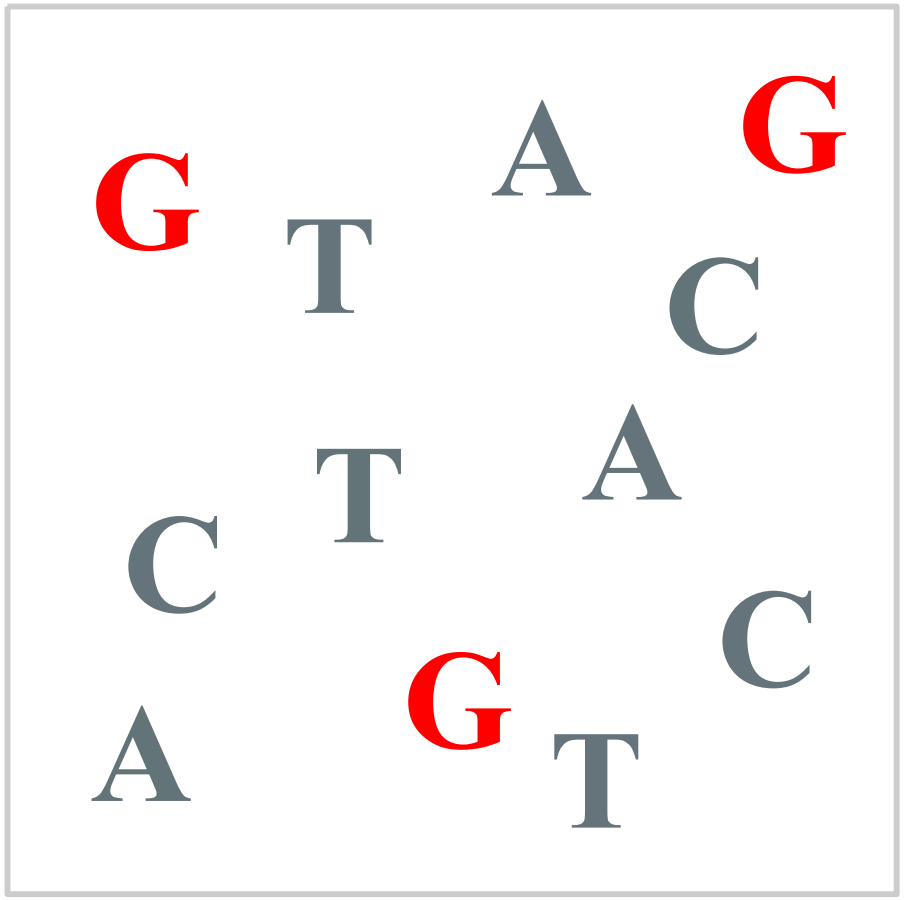


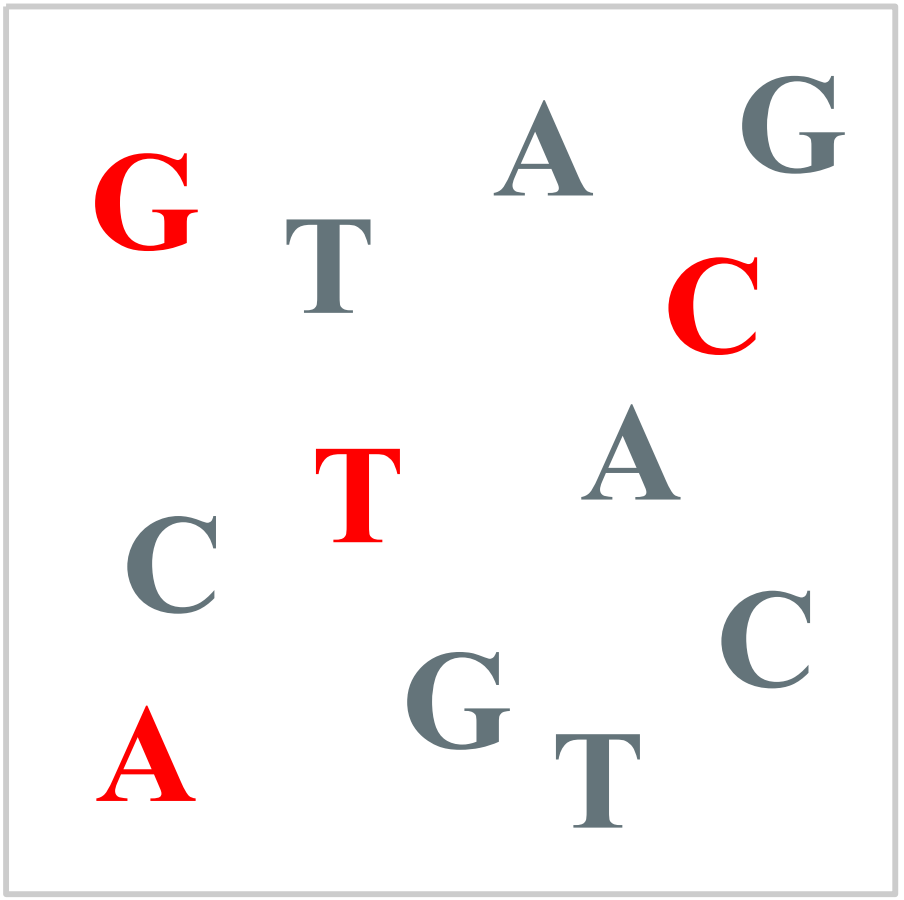
Gruppi





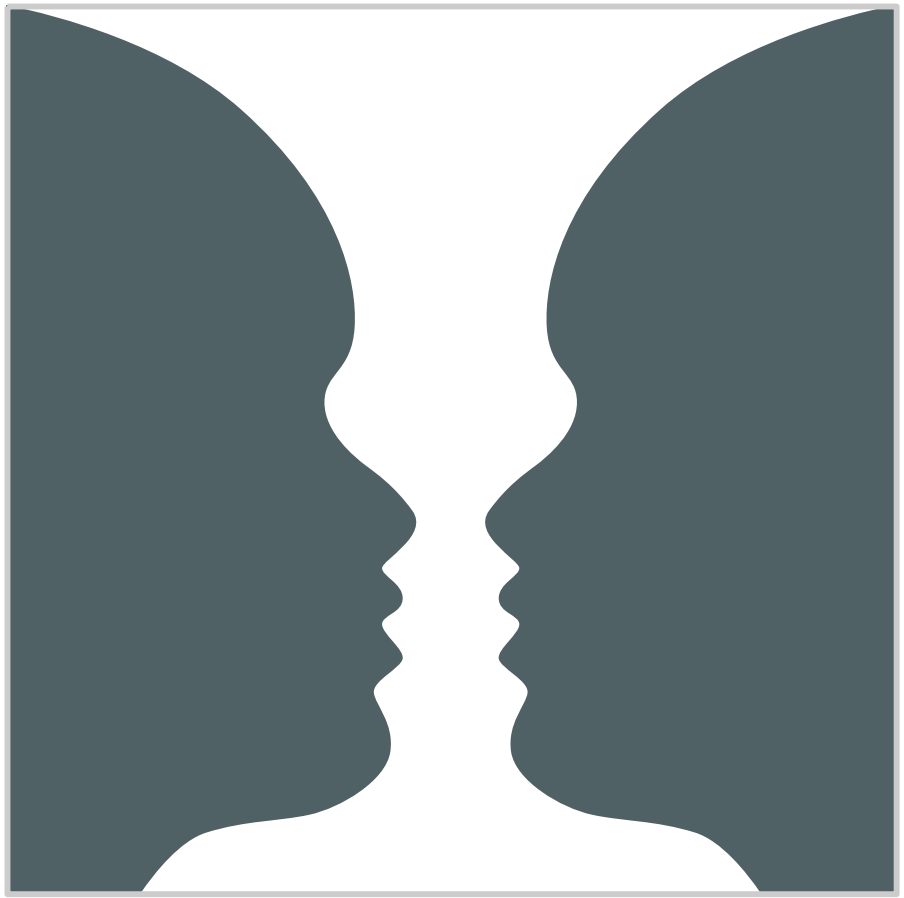
Similitudini

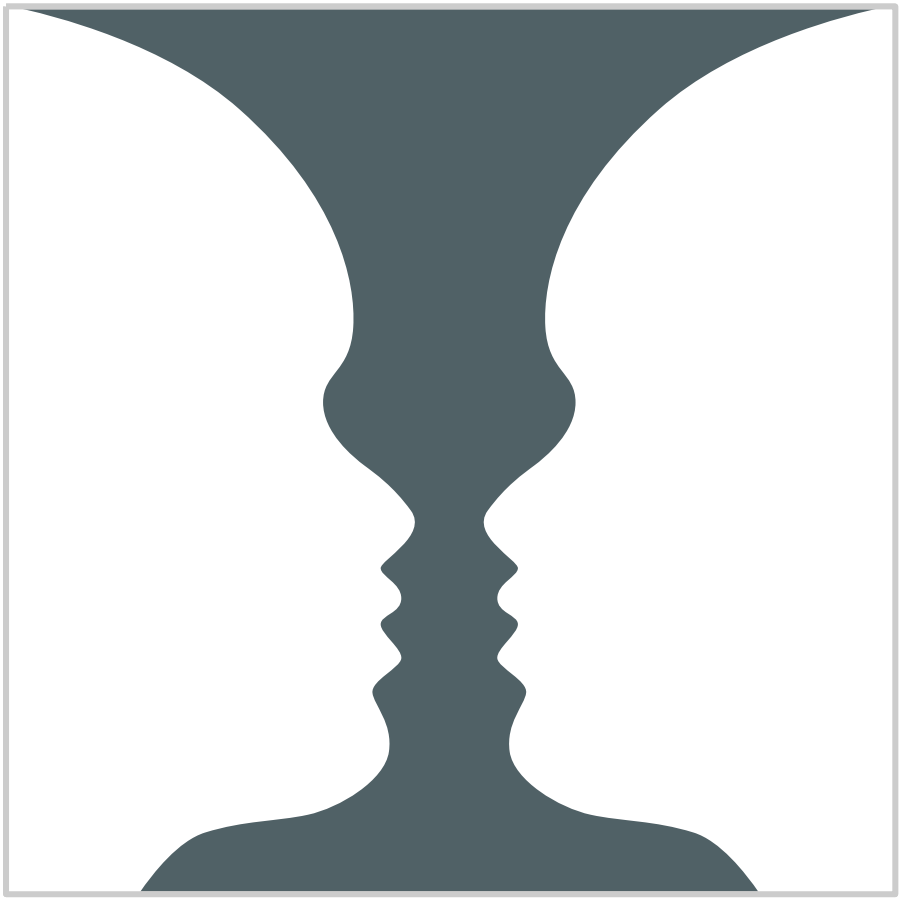




Chiusura

Figure e sfondo

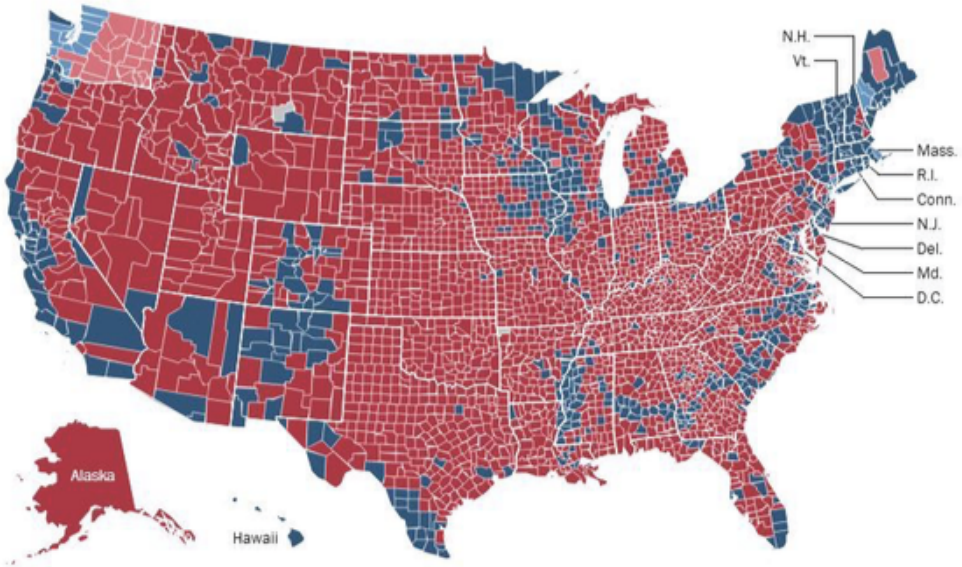


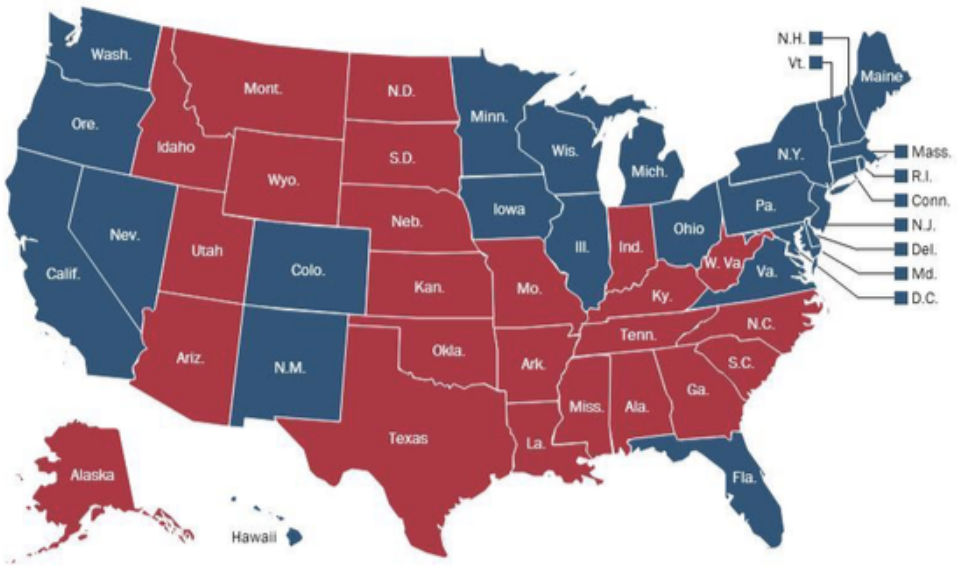


Lezione #5

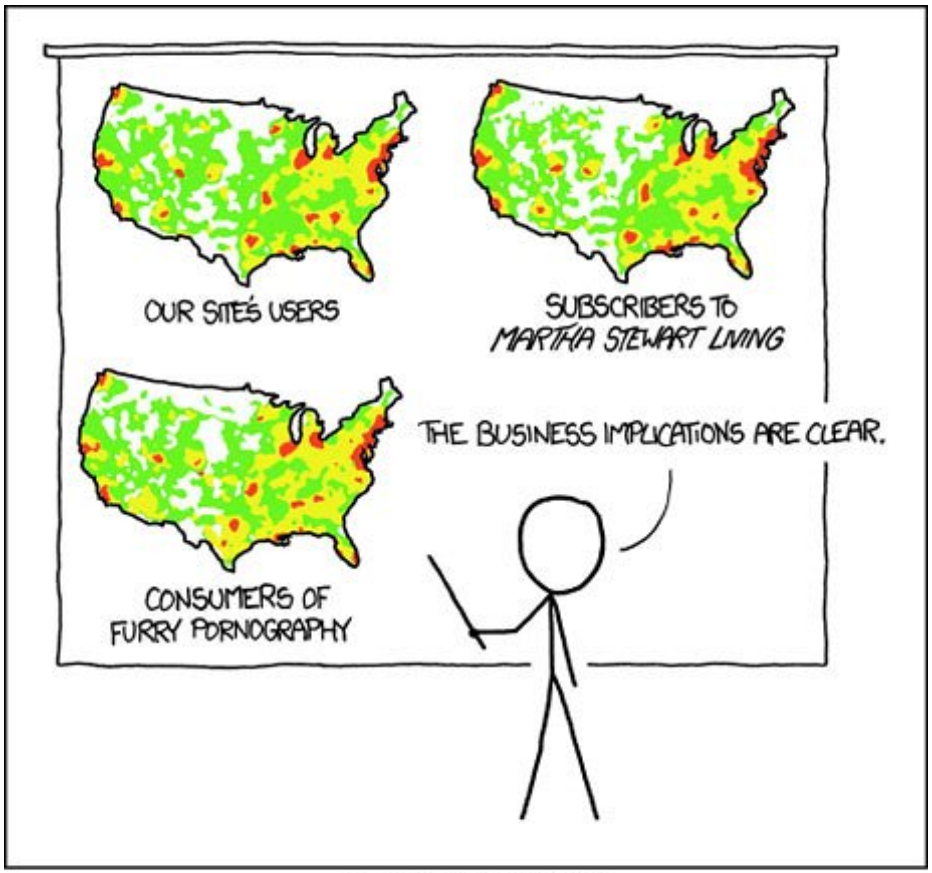
"Usare attributi pre-attentive per valorizzare ed esaltare strutture che emergono dai dati, ma non esagerare. Inoltre, non fidatevi troppo di quello che l'occhio vede e vuole vedere"

Raccontare storie diverse con stessi dati



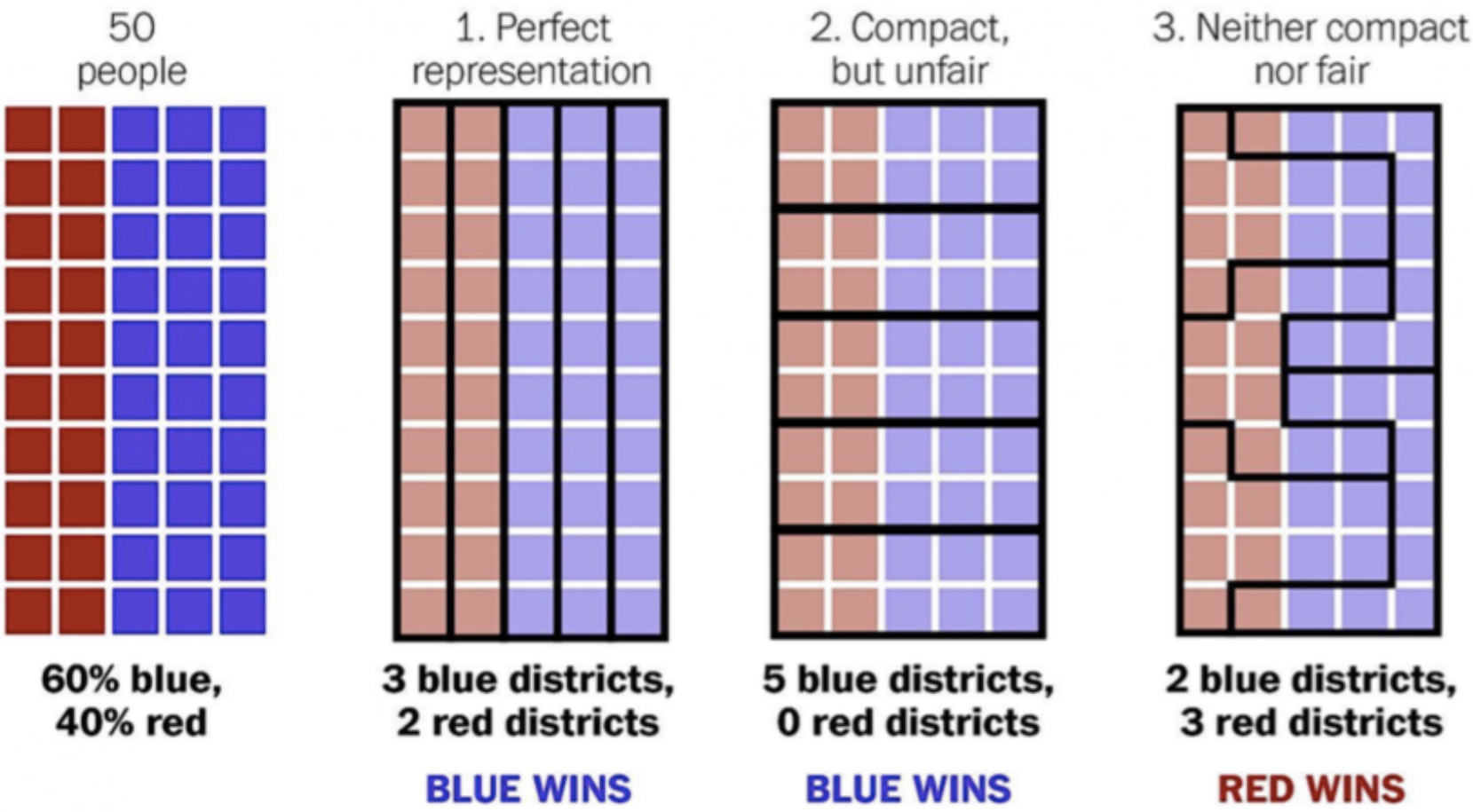


also heatmaps and dotmaps have flaws

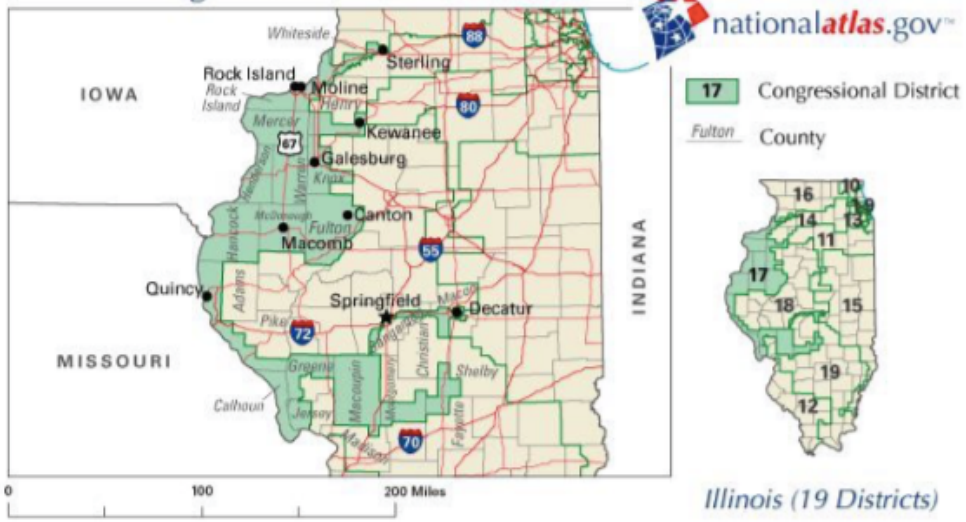


PET PEEVE #208:
GEOGRAPHIC PROFILE MAPS WHICH ARE BASICALLY JUST POPULATION MAPS

gerrymandering

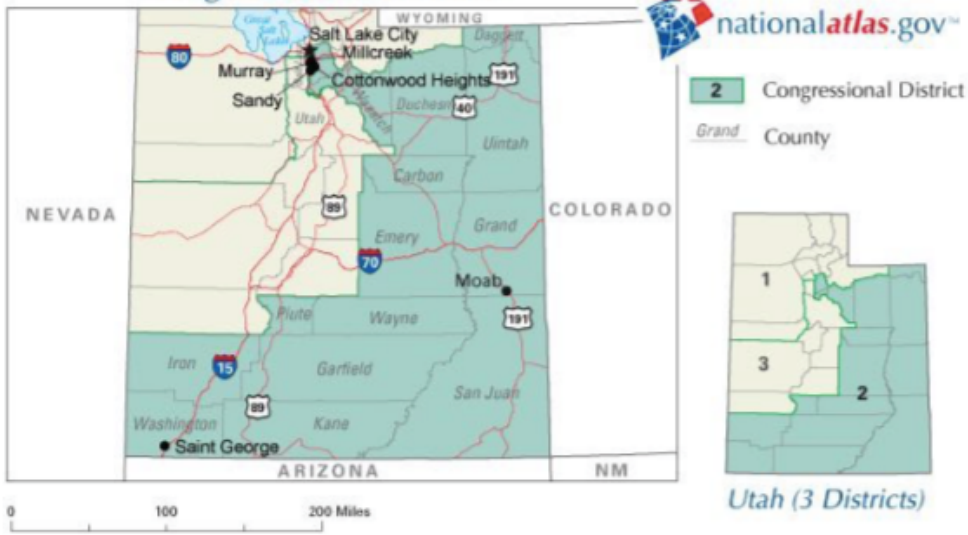


Congressional District 17



Illinois (19 Districts)

Congressional District 2



Utah (3 Districts)

Congressional District 22

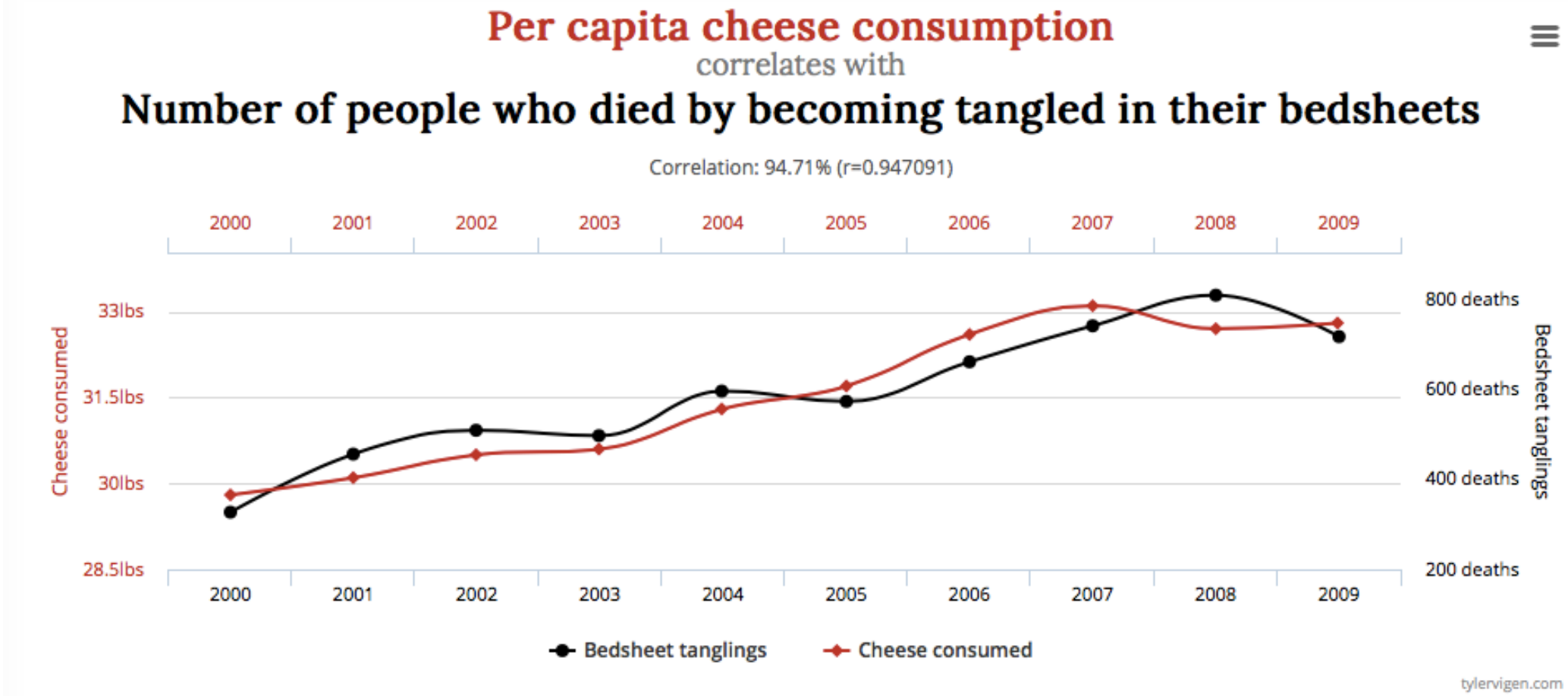
Texas (32 Districts)

Congressional District 12

North Carolina (13 Districts)

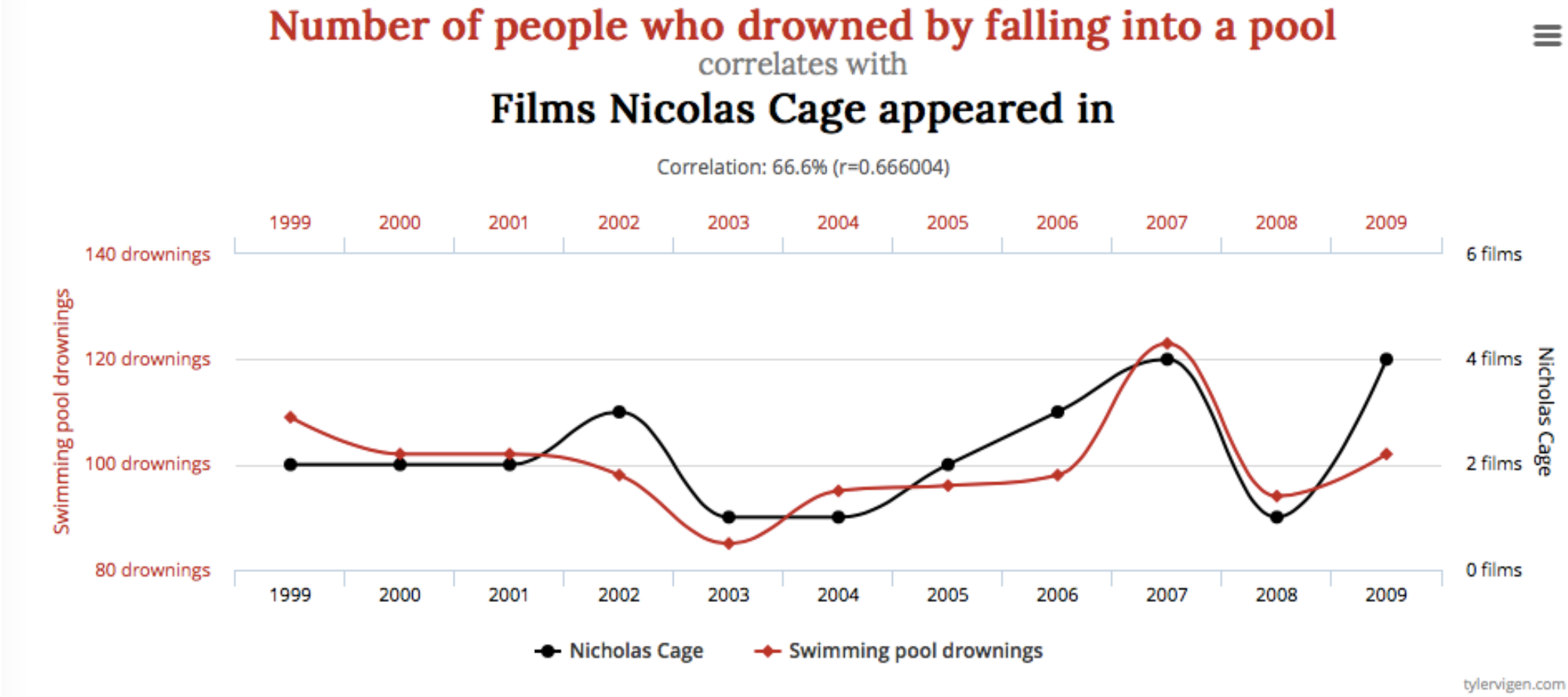
Correlazioni spurie

http://www.tylervigen.com/spurious-correlations



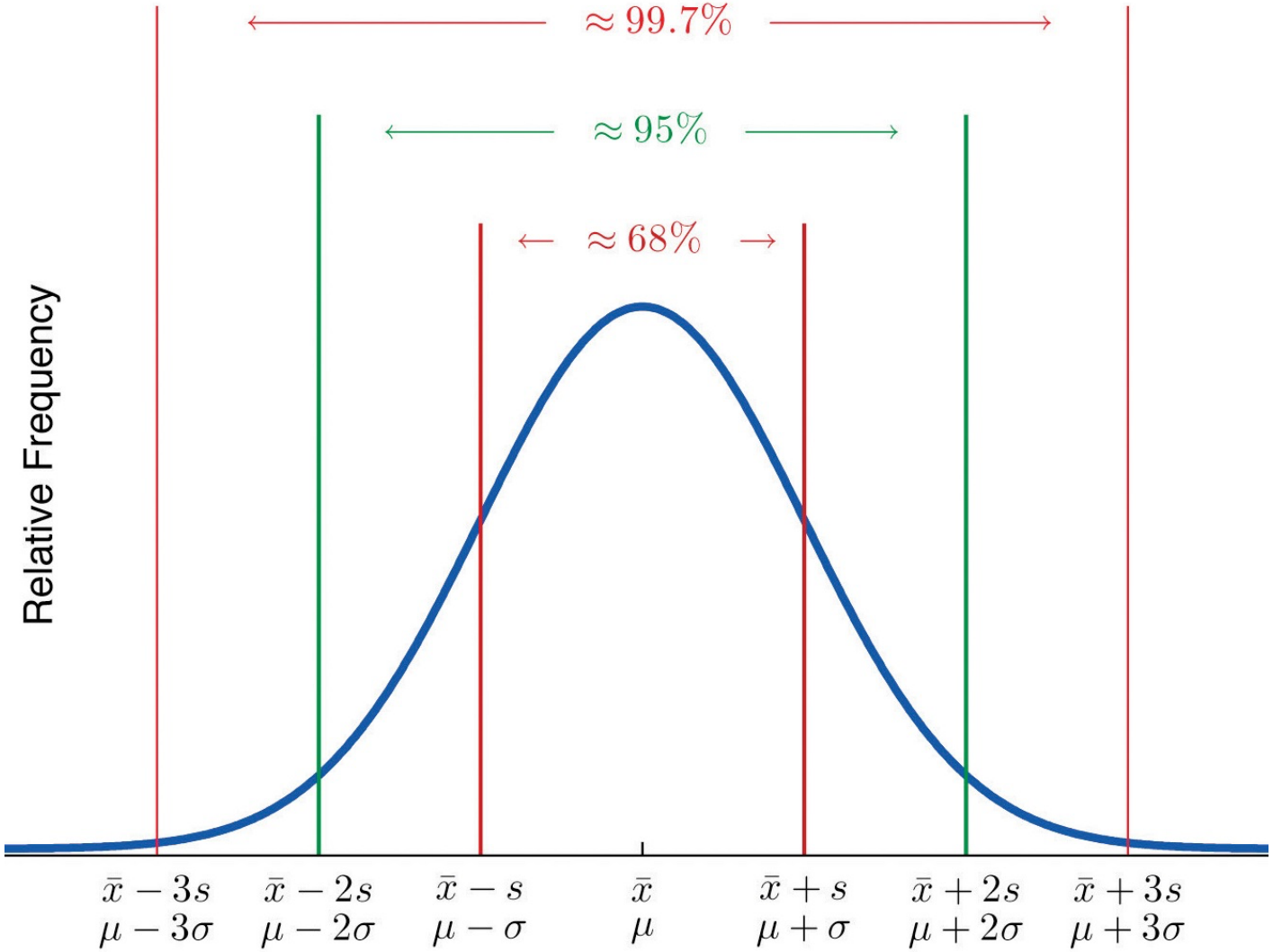
Correlazioni spurie

http://www.tylervigen.com/spurious-correlations

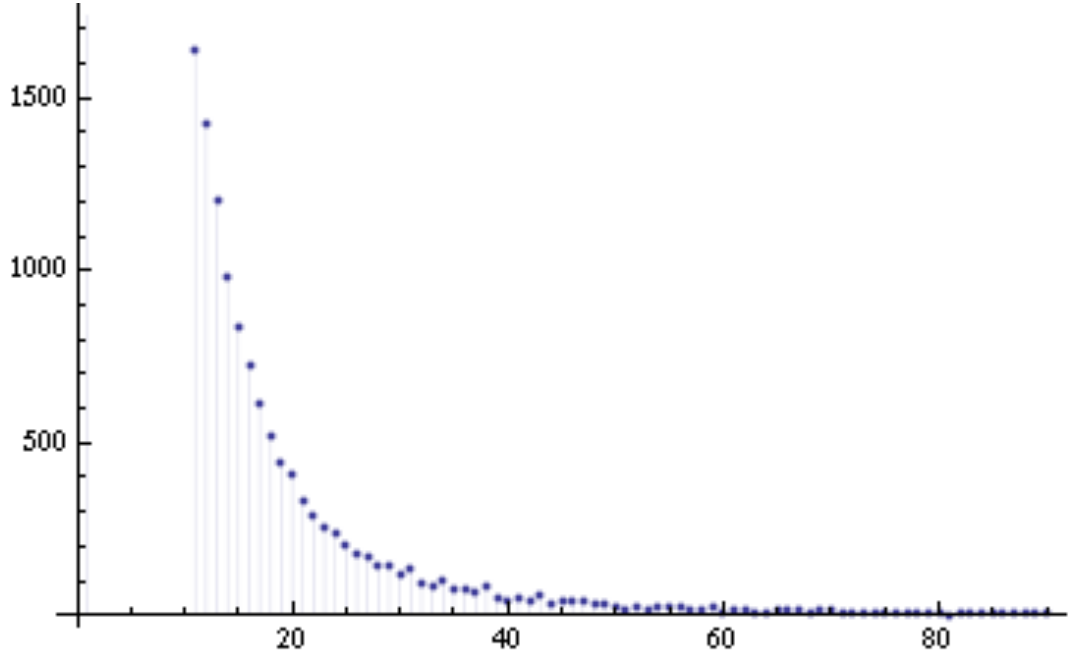


La media statistica

Distribuzione delle altezze negli individui:
la media ci aiuta a fare predizioni



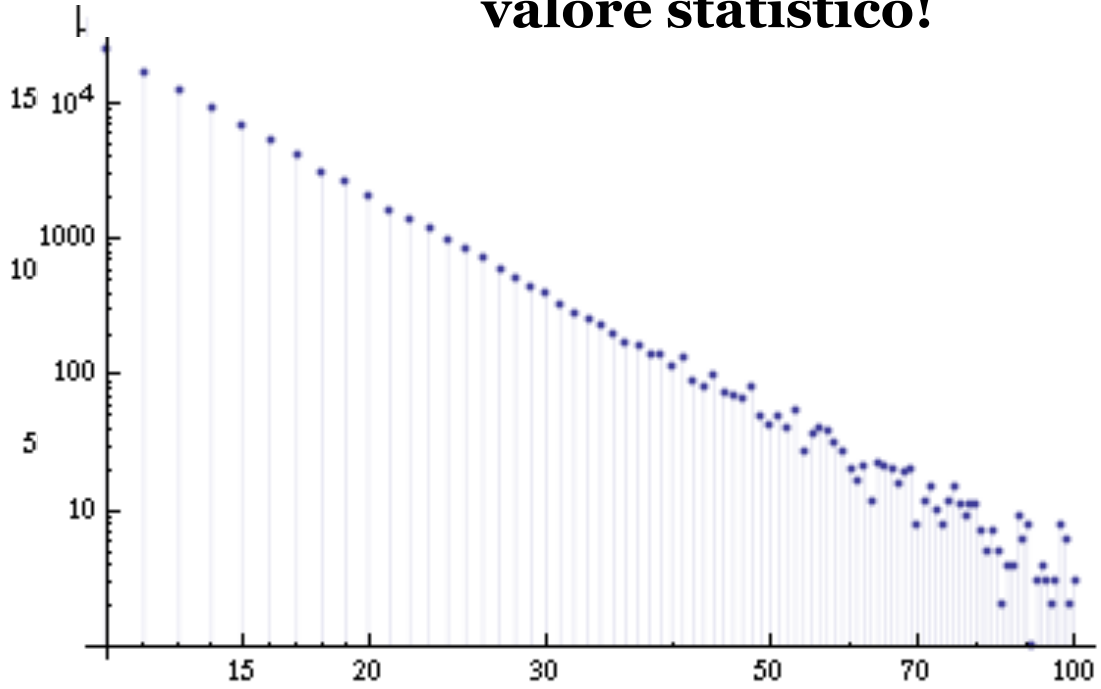
La media statistica



esempio:
Distribuzione della ricchezza

La media statistica

La media in questo caso non ha alcun valore statistico!

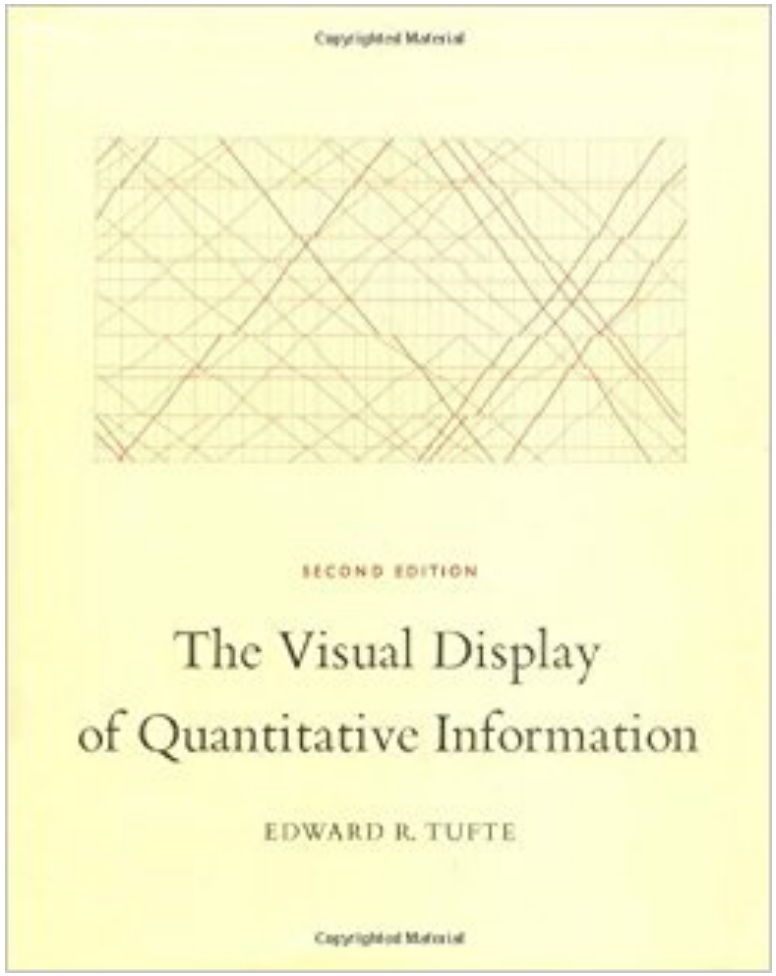


esempio:
Distribuzione della ricchezza

Lezione #6

"Non torturare i dati con una matematica sbagliata. Indici, misure, rappresentazioni e modelli per fare predizioni sono spesso utili, ma il rischio di usare strumenti non fondati matematicamente è altissimo e quasi mai i risultati sono ovvi"

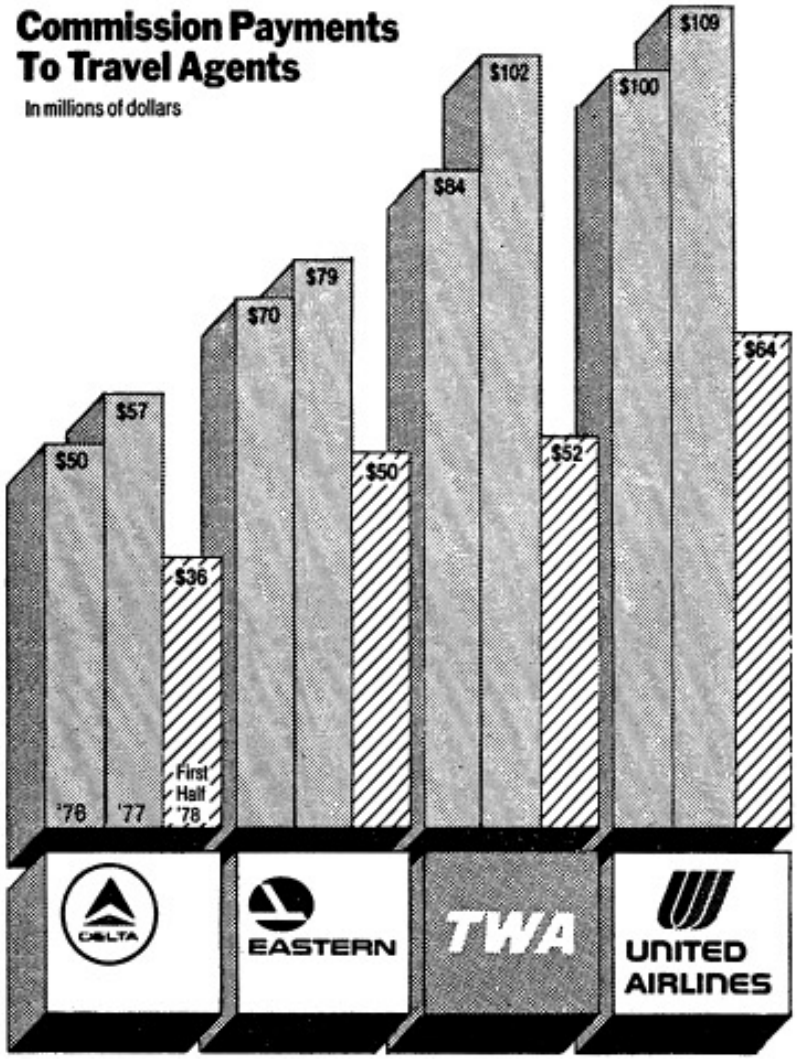
Misurare le bugie nei grafici



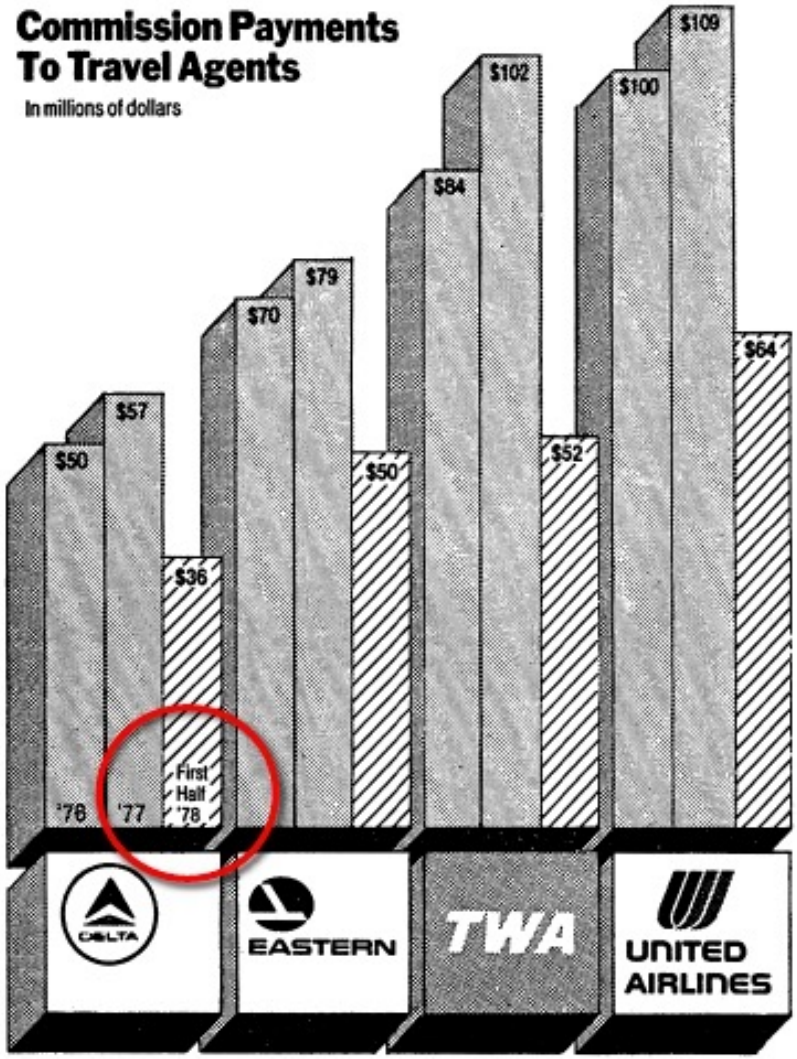
The Visual Display of Quantitative Information
Edward R. Tufte

https://www.edwardtufte.com/tufte/books_vdqi

Integrità grafica

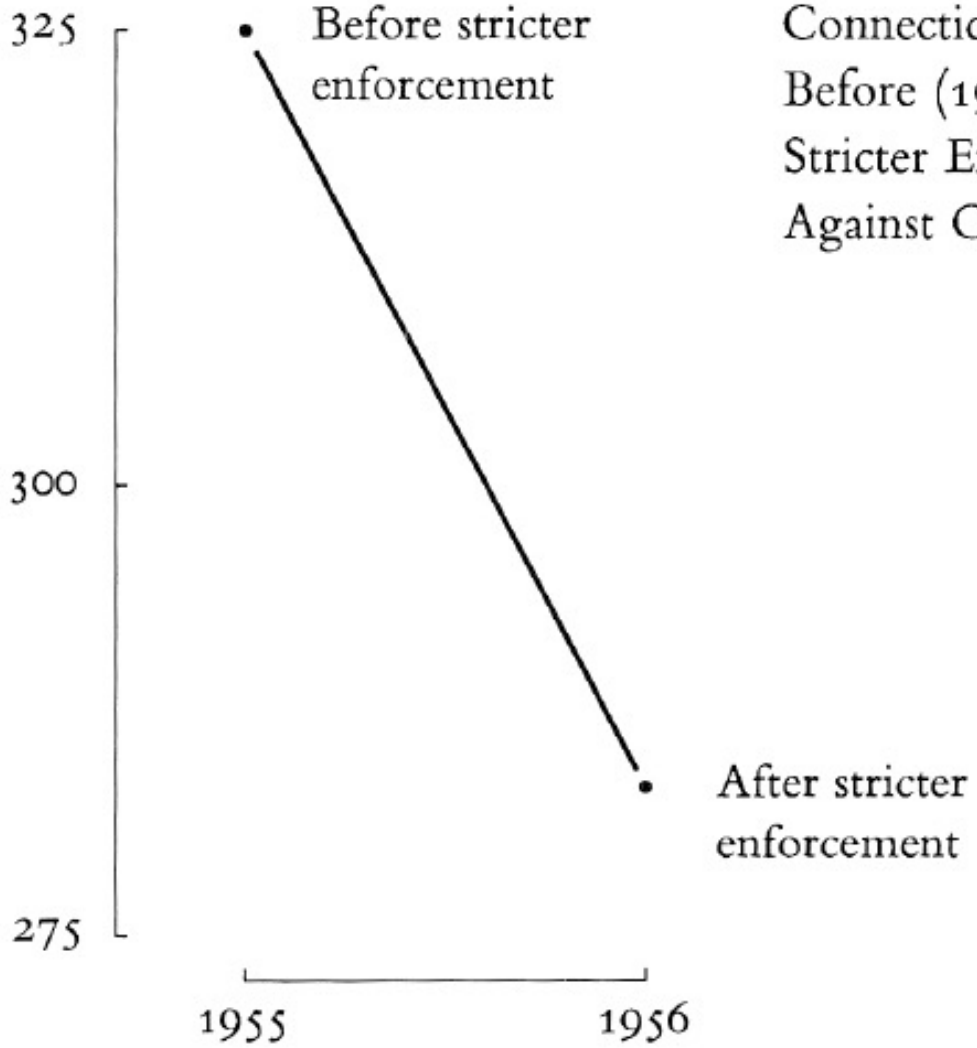


Integrità grafica



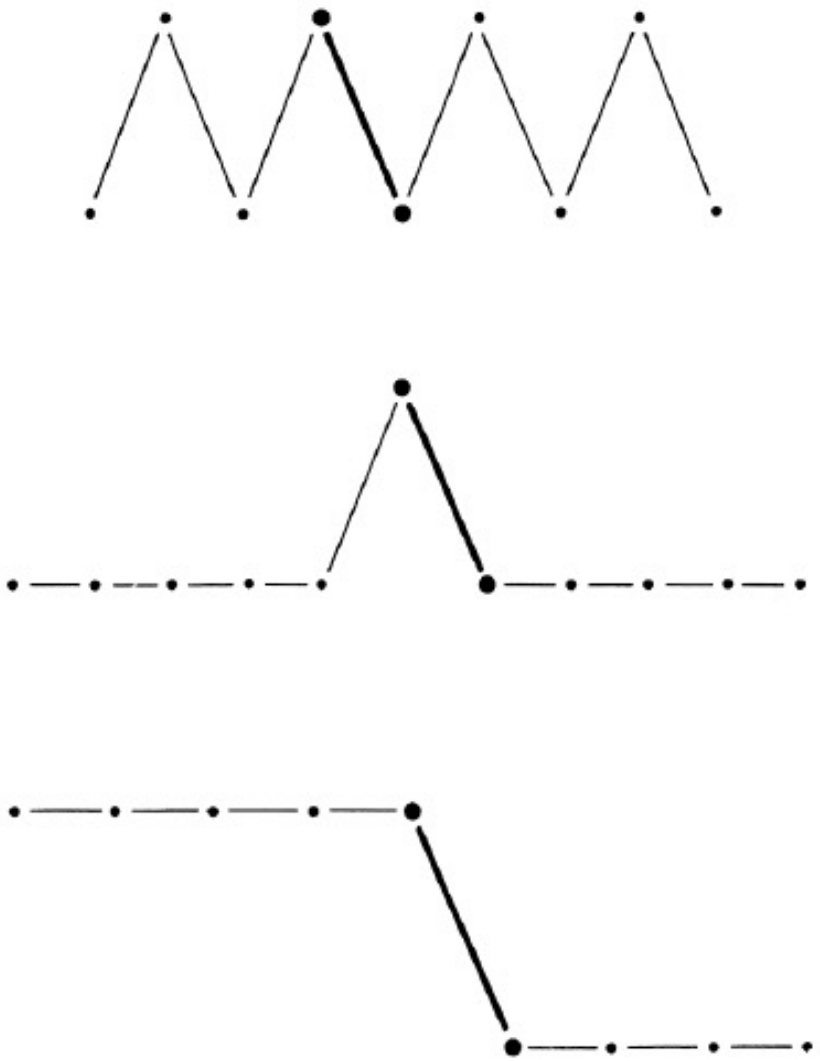
- "Pseudo-decline"
- Comparing full years (1976 and 1977) to half year (1978)

Integrità grafica



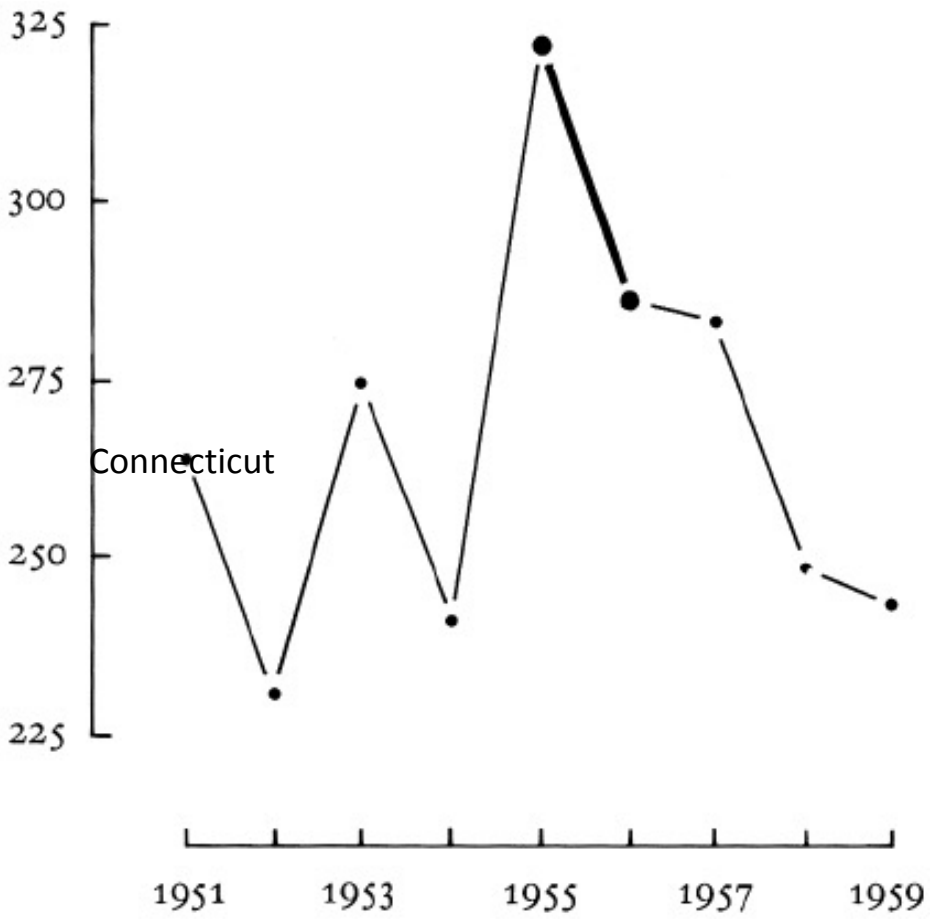
Connecticut Traffic Deaths,
Before (1955) and After (1956)
Stricter Enforcement by the Police
Against Cars Exceeding Speed limit

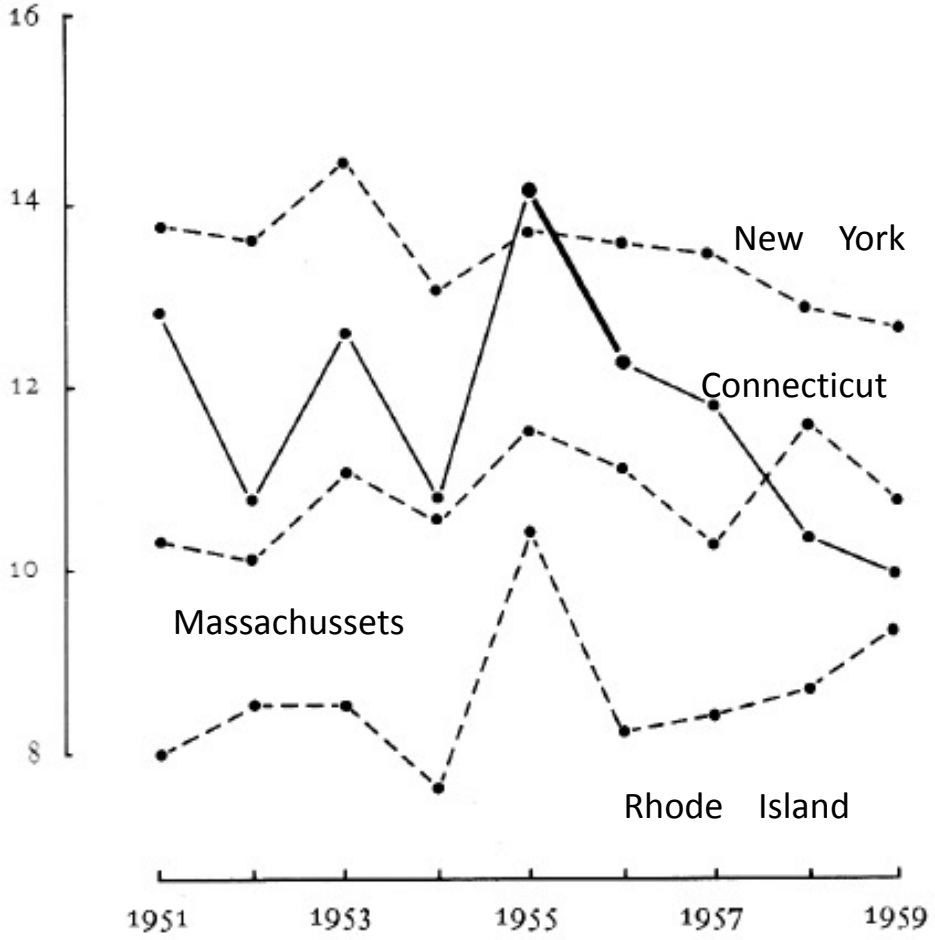
Integrità grafica



- Missing context
- Which pattern did this segment come from?

Integrità grafica





Lezione #7

"Non introdurre bugie nascondendosi dietro l'idea di migliorare l'estetica della infografica. Esistono misure molto efficaci ed oggettive che svelano l'inganno e la reputazione si perde in un baleno"

Terminology

- Data Visualization
- Scientific Visualization
- Information Visualization
- Statistical Graphics
- Visual Analytics
- Information Dashboards
- Infographics
- Informative Art

Terminology

- Differences between terms are often fuzzy
  - Information visualization versus infographics
- Differences between terms are sometimes highly contested
  - Information visualization versus statistical graphics
- Differences often come down to two aspects
  - Type of data being visualized
  - Why data is being visualized

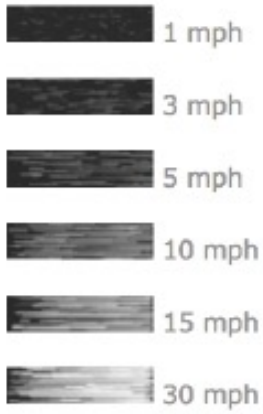
Data Visualization

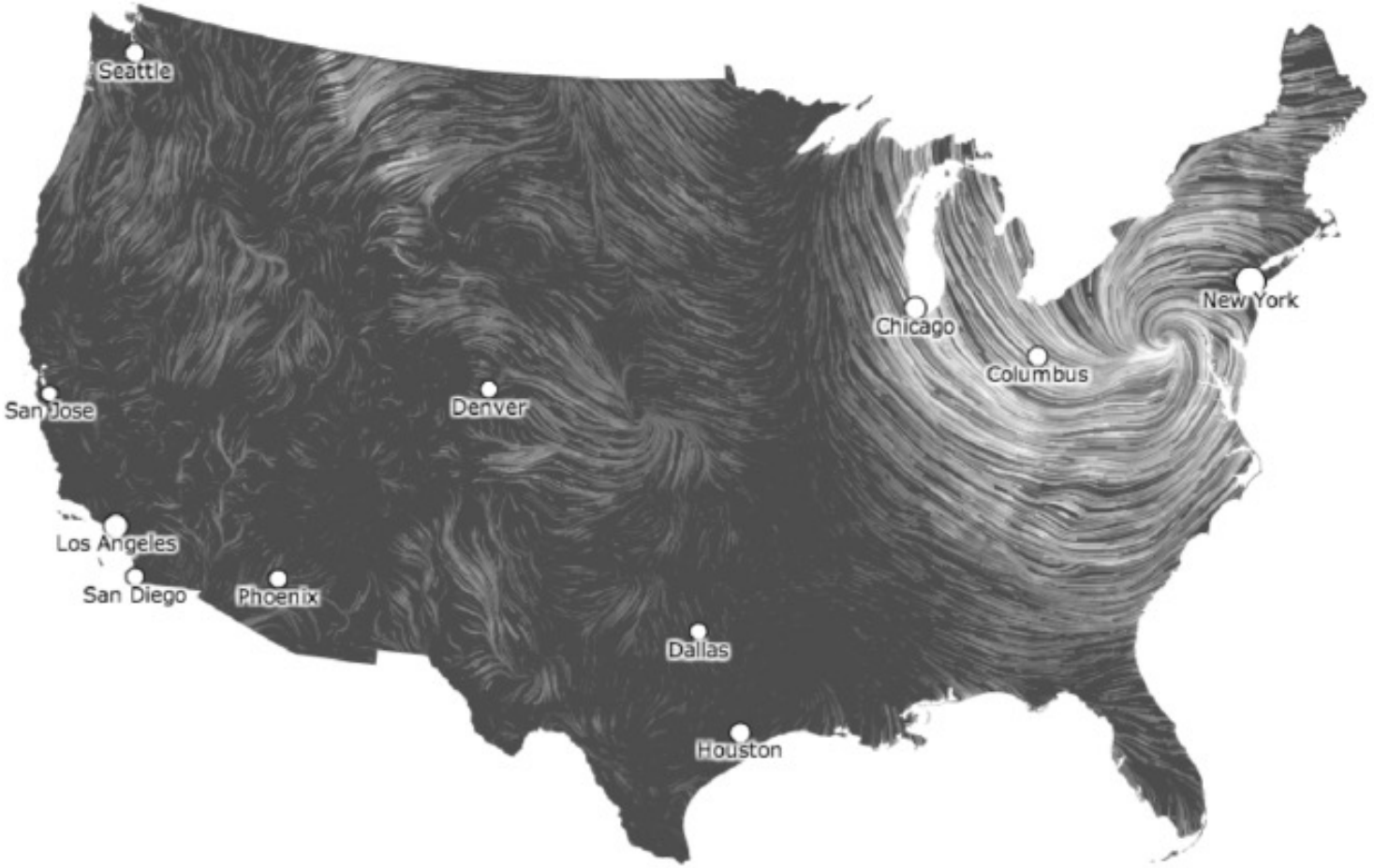
- Definition
  - **Communicates non-visual data visually**
  - Result should be readable and recognizable*
- Two Subfields
  - Scientific visualization
  - Information visualization
- Transforms raw **data** into **information**

* http://eagereyes.org/criticism/definition-of-visualization

October 30, 2012
6:59 am EST
(time of forecast download)

top speed: **39.7 mph**
average: **8.4 mph**

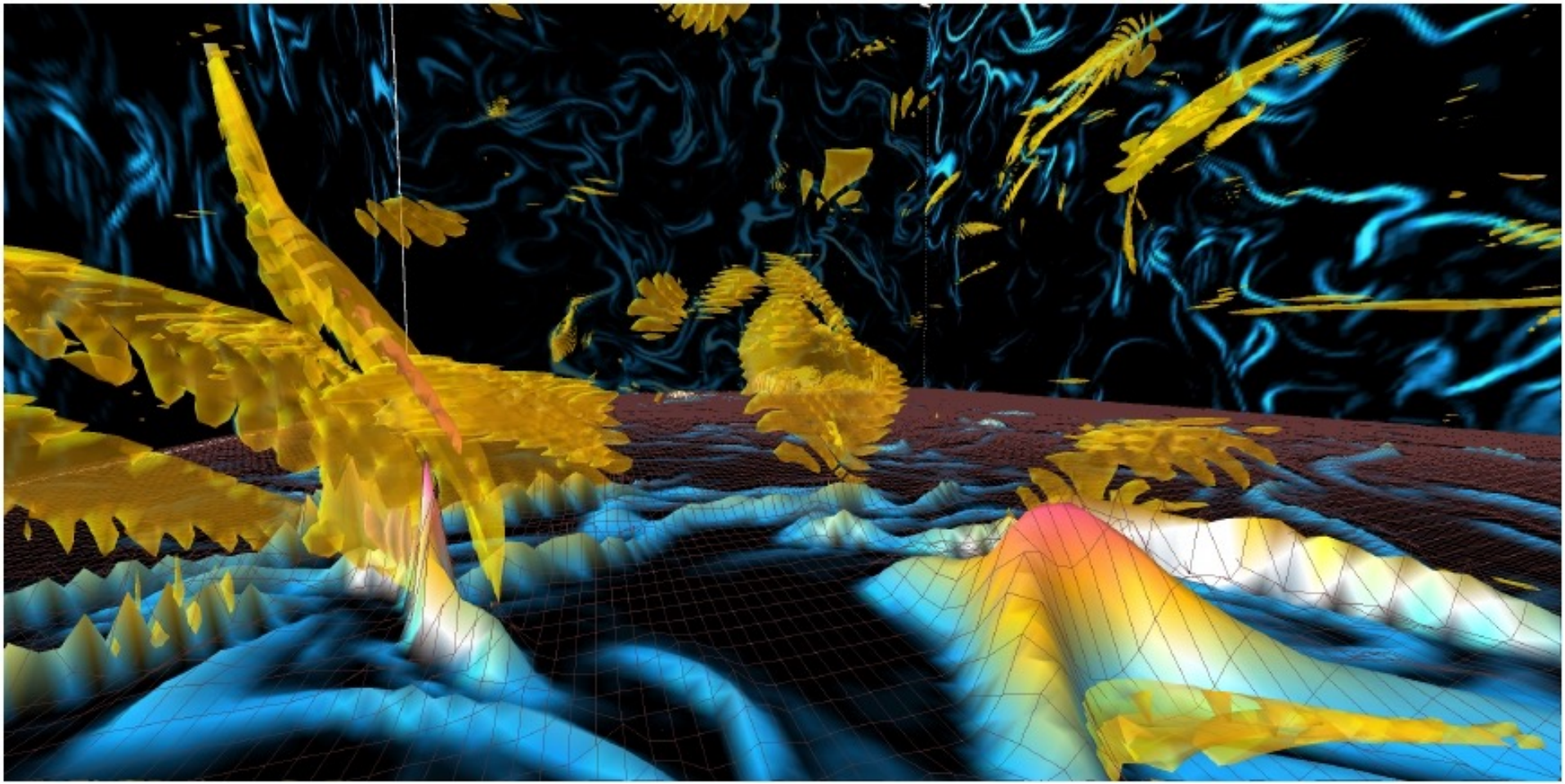




http://hint.fm/wind/gallery/oct-30.js.html

Scientific Visualization

- Type of Data
  - Scientific data (objects exist in 1D, 2D, or 3D space)
  - Often scalar or vector fields from computer simulations
- Primary Purpose
  - Aims to convey **scientific** data accurately
  - Aims to reveal underlying **structure** in data
  - Aims to encourage **exploration** of data (interactivity)



http://vis.lbl.gov/Events/SC04/Incite3/index.html

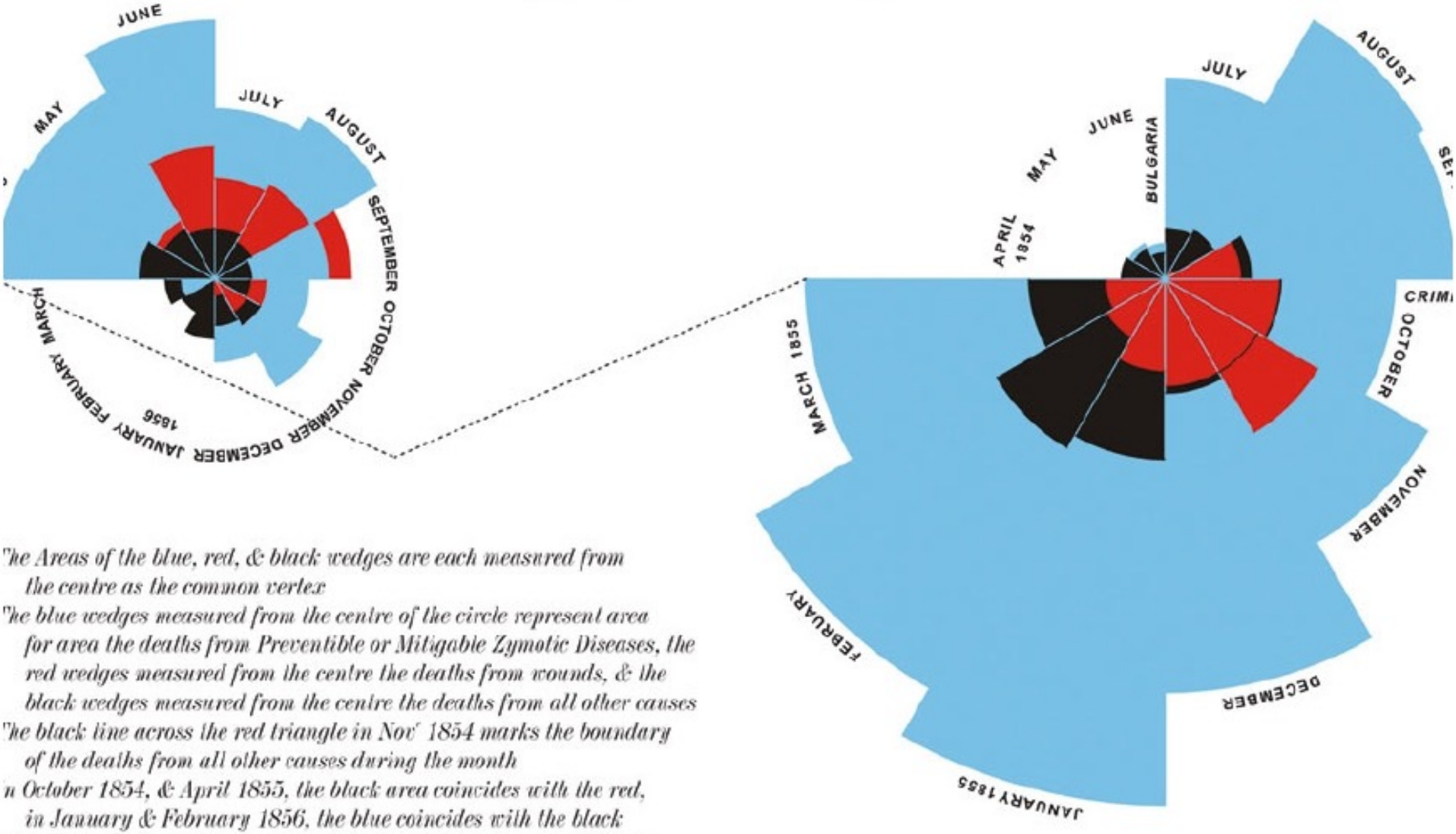
Information Visualization

- Type of Data
  - Abstract data (has no inherent physical form)
  - May be numerical, categorical, temporal, geospatial, or text data
- Primary Purpose
  - Aims to convey **abstract** data accurately
  - Aims to reveal underlying **structure** in data
  - Aims to encourage **exploration** of data (interactivity)
  - Aims to display data **aesthetically**

DIAGRAM OF THE CAUSES OF MORTALITY
IN THE ARMY IN THE EAST.

2.
APRIL 1855 TO MARCH 1856.

1.
APRIL 1854 TO MARCH 1855

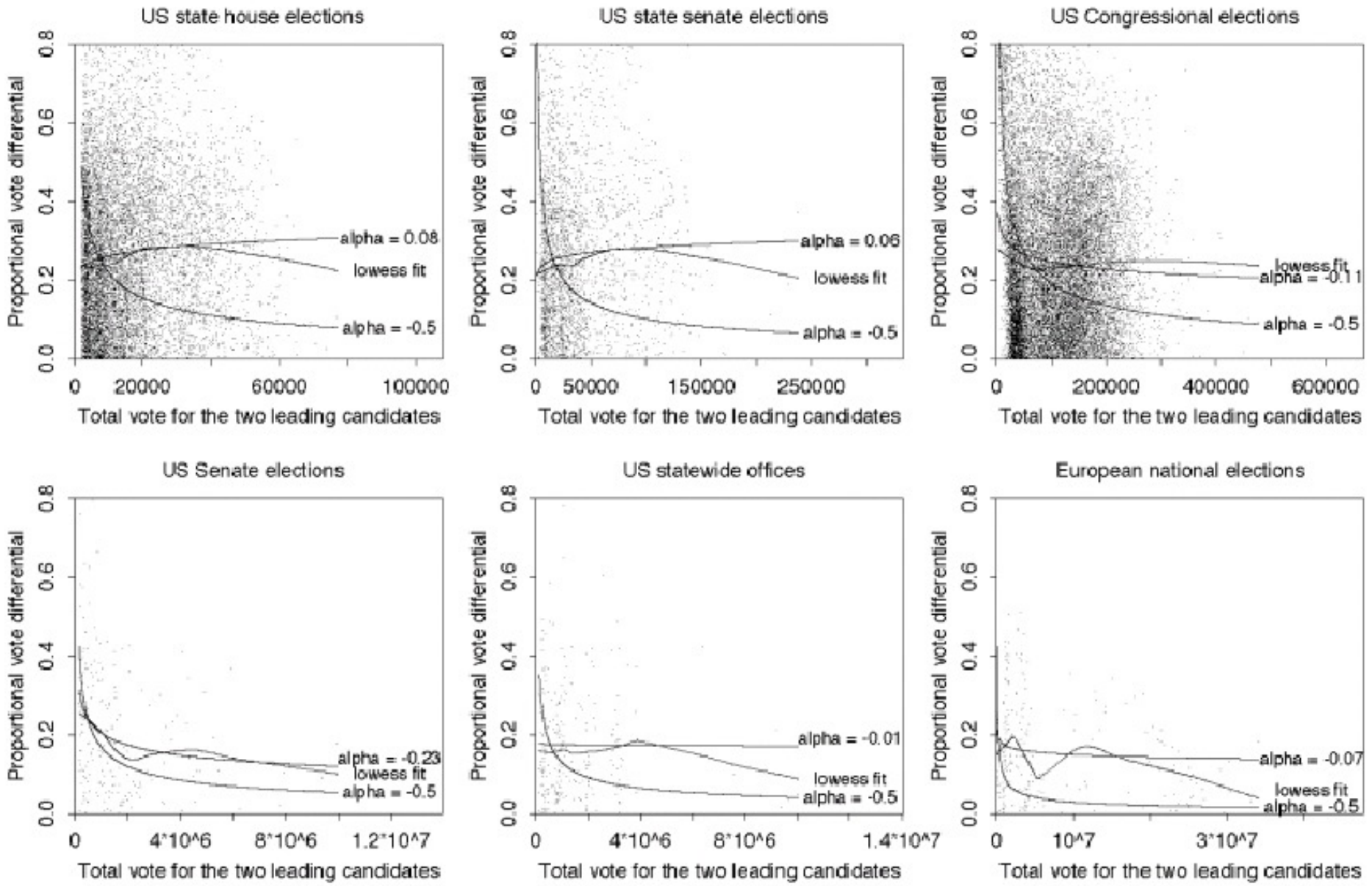


The Areas of the blue, red, & black wedges are each measured from the centre as the common vertex
The blue wedges measured from the centre of the circle represent area for area the deaths from Preventible or Mitigable Zymotic Diseases, the red wedges measured from the centre the deaths from wounds, & the black wedges measured from the centre the deaths from all other causes
The black line across the red triangle in Nov' 1854 marks the boundary of the deaths from all other causes during the month
n October 1854, & April 1855, the black area coincides with the red, in January & February 1856, the blue coincides with the black
The entire areas may be compared by following the blue, the red & the black lines enclosing them. ©hugh-small.co.uk

Created by Florence Nightingale, 1858. Image from http://www.economist.com/node/10278643 online

Statistical Graphics

- Type of Data
  - Abstract data
  - Mostly statistical, quantitative, or numerical data
- Primary Purpose
  - Aims to convey data accurately
  - Aims to convey underlying structure in data
- Not Emphasized
  - May not be aesthetically pleasing
  - May not encourage exploration or be interactive



http://andrewgelman.com/2007/10/why_the_squarer/

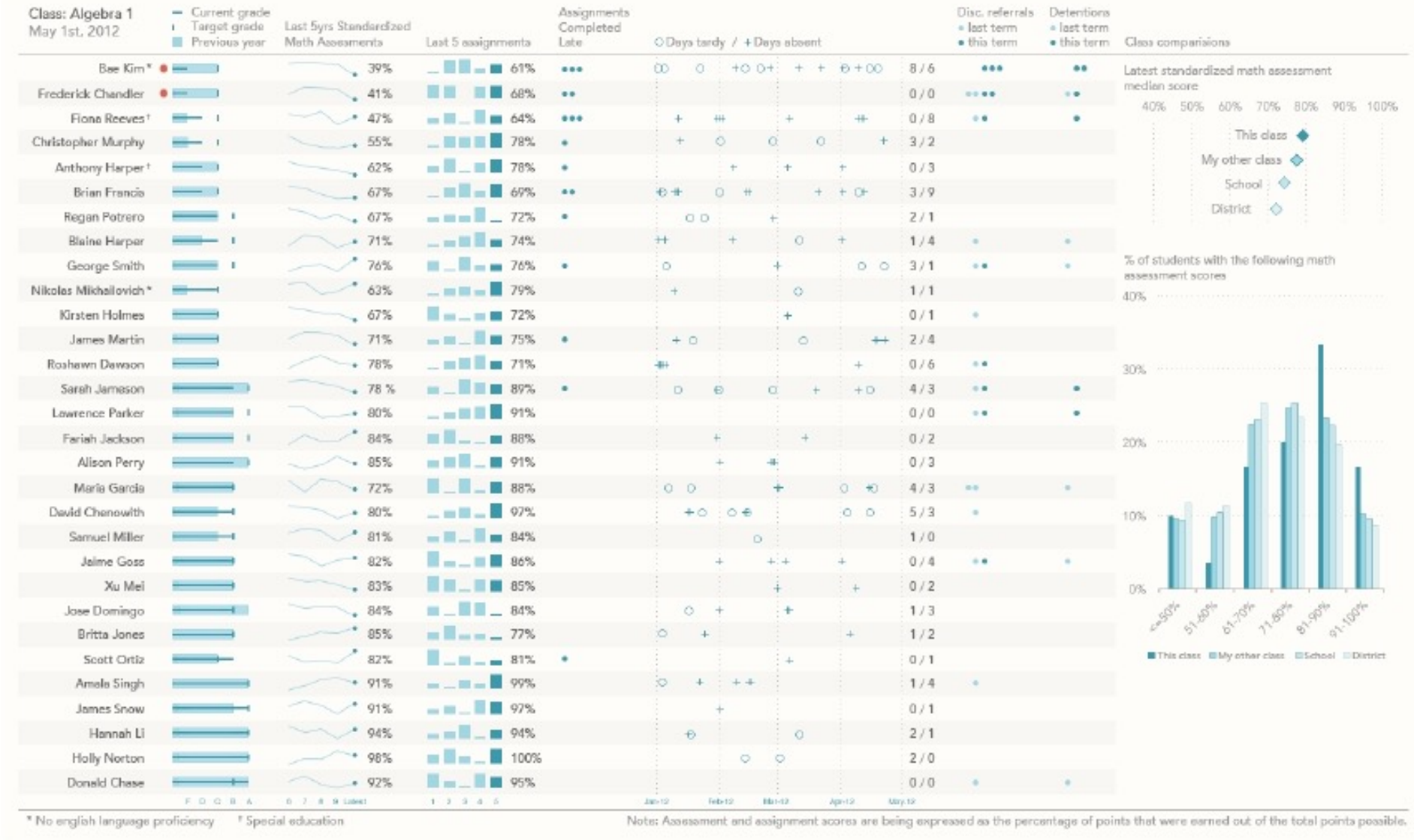
Visual Analytics

- Type of Data
  - Abstract data
- Primary Purpose
  - Aims to answer a specific question (goal-oriented)
  - Aims to support analytical reasoning with interactive visual interfaces
- Not Emphasized
  - May not be aesthetically pleasing
  - May not be constrained by a single display

http://ff.cx/vast-challenge-2012/

Information Dashboards

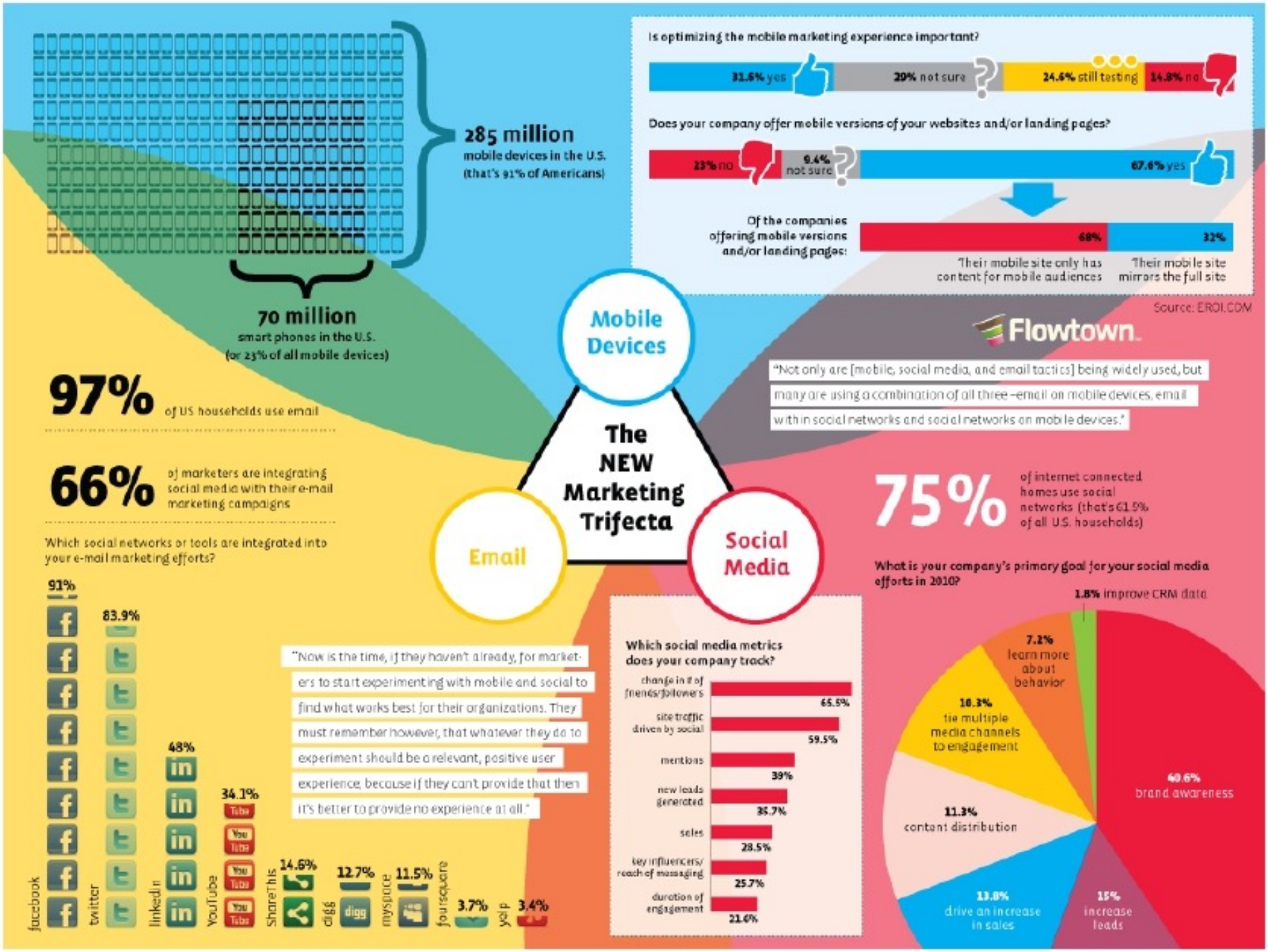
- Type of Data
  - Abstract data
  - Temporal data or time series
- Primary Purpose
  - Aims to convey large amount of information quickly
  - Aims to convey outliers and trends at a glance
- Not Emphasized
  - May be extremely dense



http://www.perceptualedge.com/blog/?p=1374

Infographics

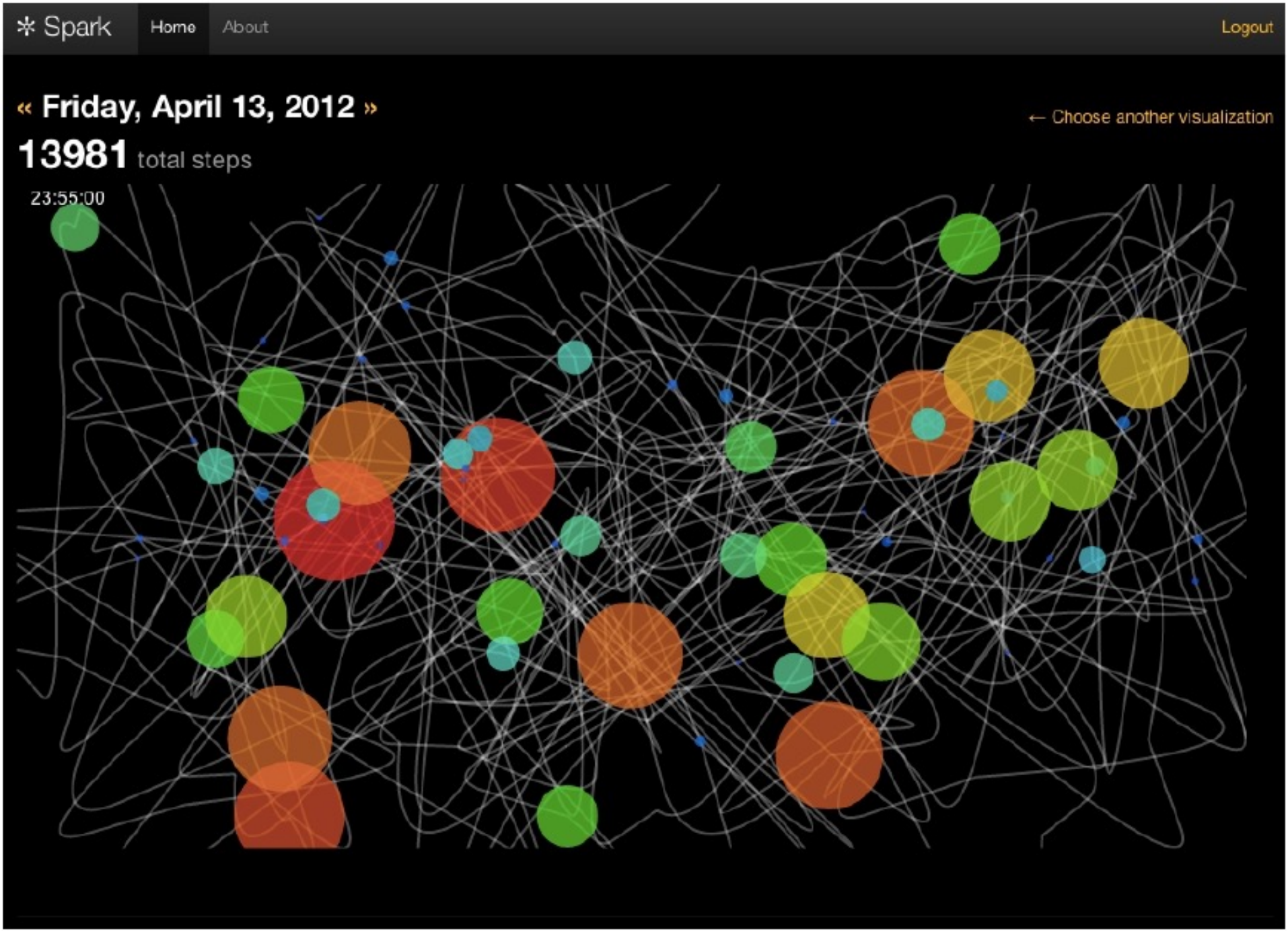
- Type of Data
  - Abstract data
- Primary Purpose
  - Aims to be eye-catching and capture attention
  - Aims to convey information quickly
- Not Emphasized
  - May not be accurate
  - May not use space efficiently
  - May not encourage exploration of data



http://visual.ly/new-world-marketing

Informative Art

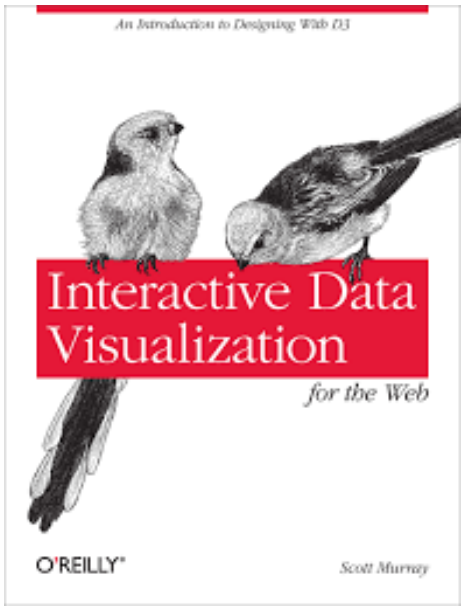
- Type of Data
  - Abstract data
- Primary Purpose
  - Aims to make visualization ambient or part of everyday life
  - Aims to be aesthetically pleasing
- Not Emphasized
  - May not be informative
  - May not be eye-catching



http://quantifiedself.com/2012/05/spark-visualizing-physical-activity-using-abstract-ambient-art/

Resources and books

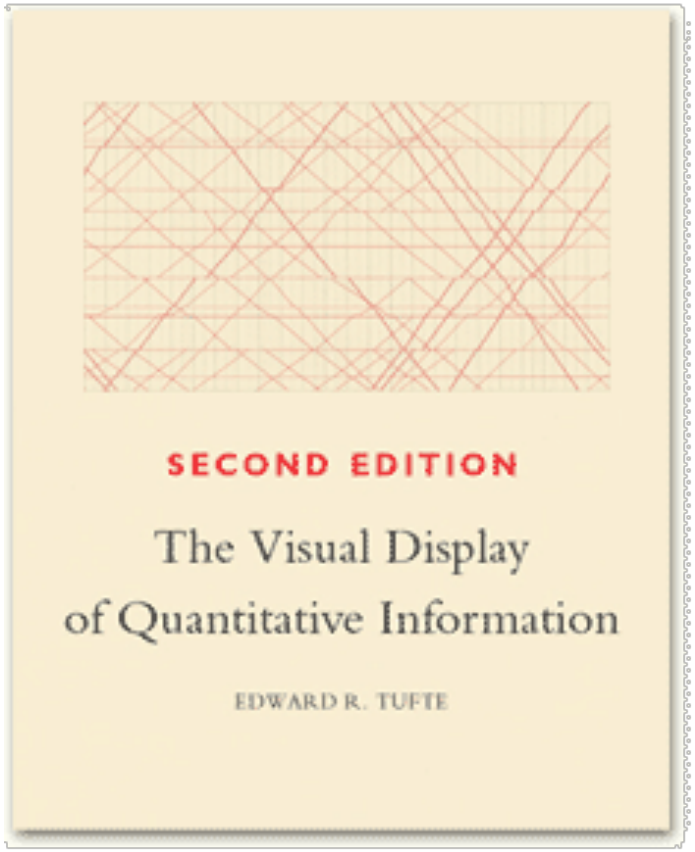
Practical textbooks



Scott Murray

Interactive Data Visualization for the Web
An Introduction to Designing with D3
O'Reilly, 2013

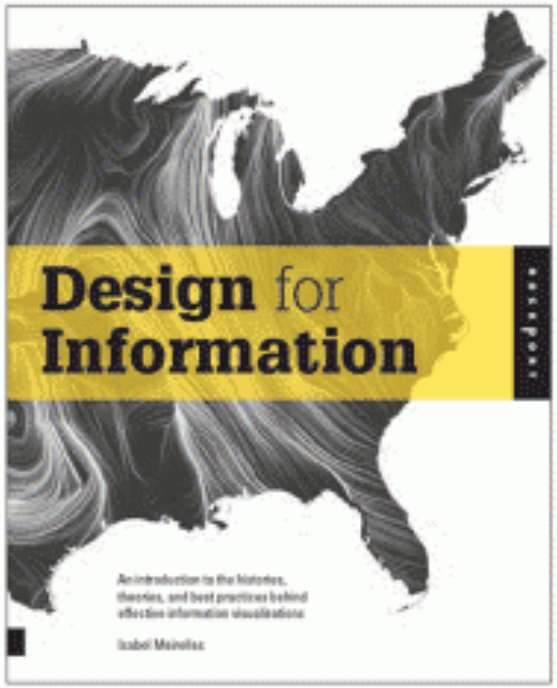
http://chimera.labs.oreilly.com/books/1230000000345/index.html



Quantitative Information: the "Bible"

Edward R. Tufte
The Visual Display of Quantitative Information
Graphics Press, 2013

http://www.edwardtufte.com/tufte/books_vdqi



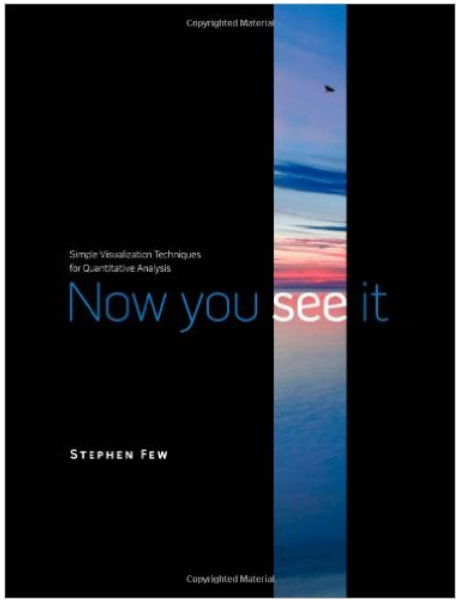
Information Visualization

Isabel Meirelles

Design for Information
Rockport, 2013

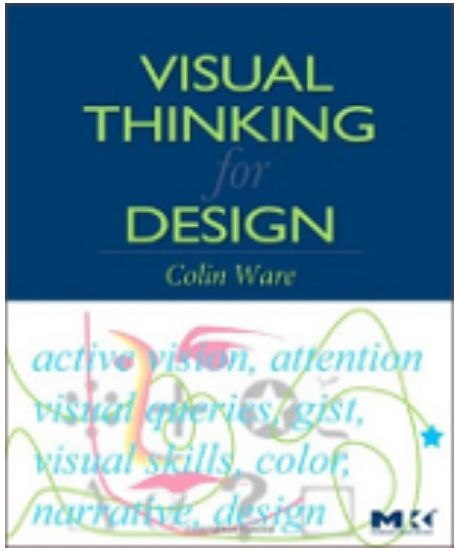
http://isabelmeirelles.com/book-design-for-information/

Perceptual principles of visualization



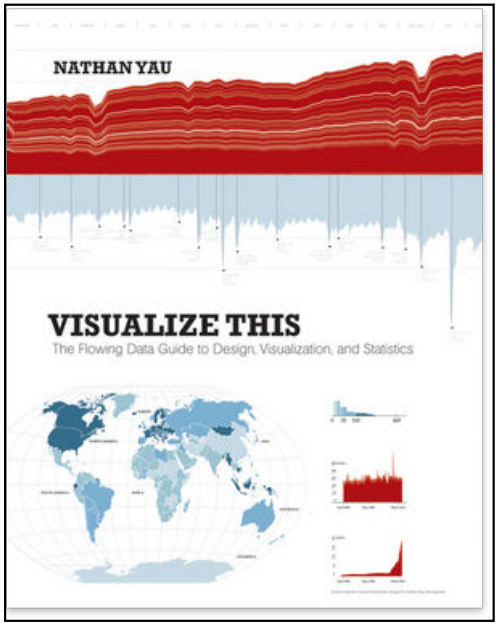
Stephen Few

Now You See It
Analytics Press, 2009



Colin Ware

Visual Thinking for Design
Morgan Kaufman, 2008

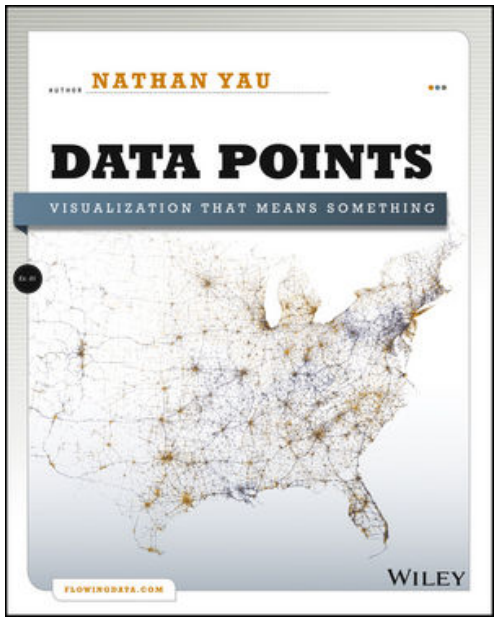


Data Visualization

Nathan Yau

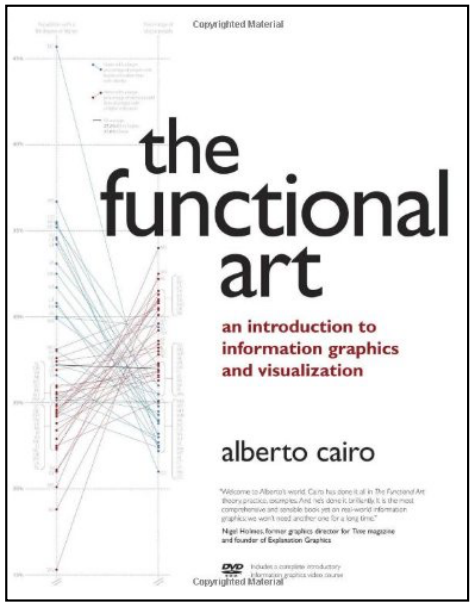
Visualize This
The FlowingData Guide to Design,
Visualization, and Statistics
Wiley, 2011

http://flowingdata.com/



Nathan Yau

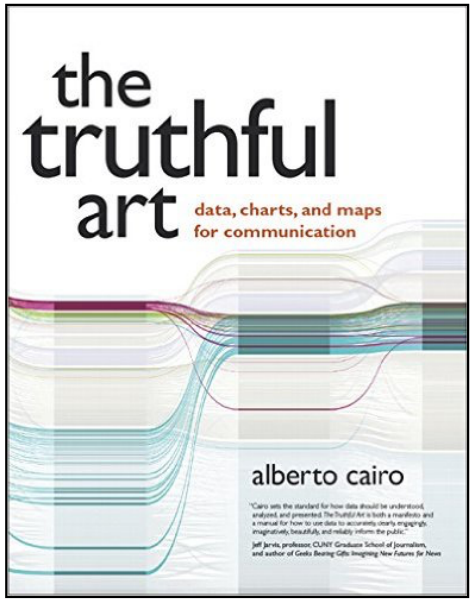
Data Points
Visualization That Means Something
Wiley, 2013



Data Journalism

Alberto Cairo

The Functional Art
An Introduction to Information Graphics and Visualization
New Riders, 2013
http://www.thefunctionalart.com/



Alberto Cairo

The Truthful Art
Data, Charts, and Maps for Communication
New Riders, 2016

Questions?