

Giancarlo Ruffo - Università degli Studi di Torino

Mentire con i dati

26 Marzo 2021, Comunicazione ICT e Media, Torino

http://www.di.unito.it/~ruffo/talks/2021_Mar_LabICT.pdf



<http://www.di.unito.it/~ruffo>

giancarlo.ruffo@unito.it

@giaruffo



i rischi della data visualization e della comunicazione scientifica

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

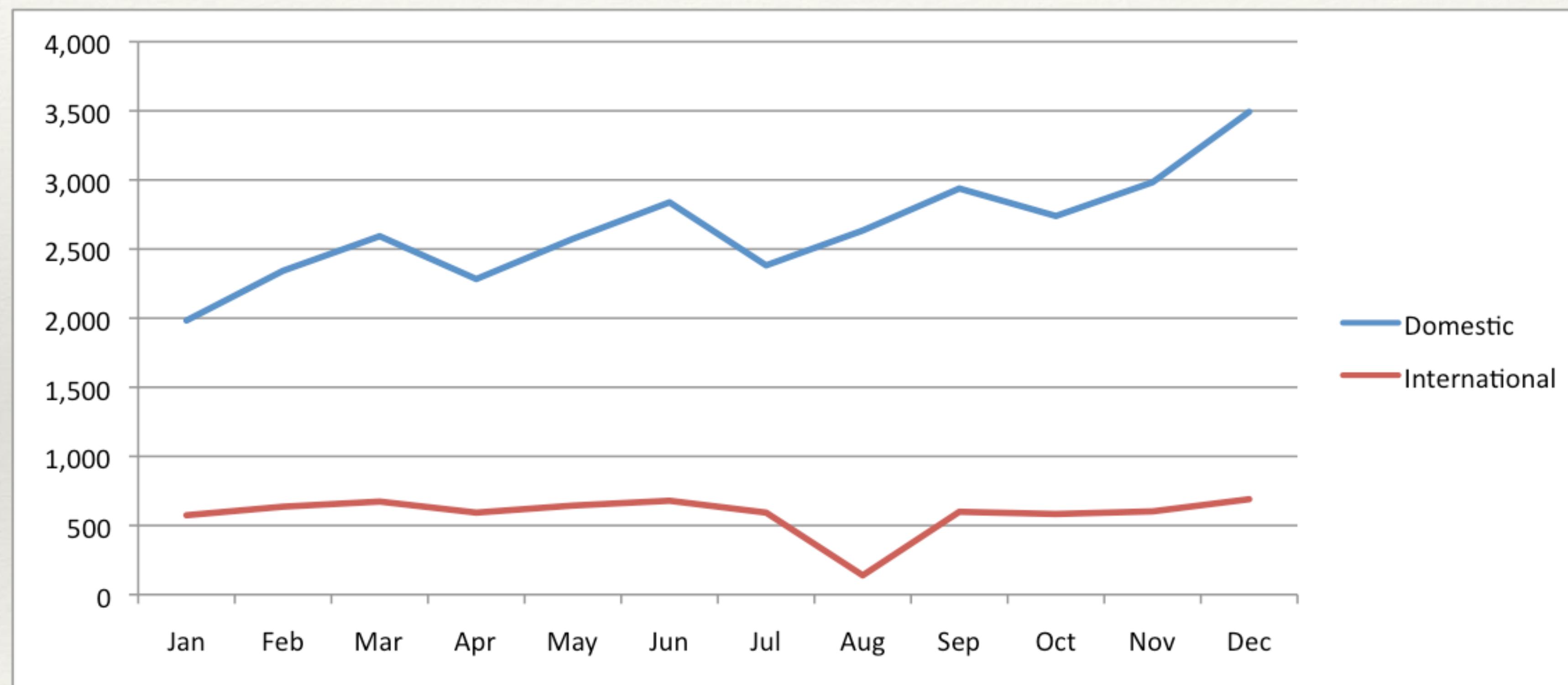
Le tabelle piene di informazioni funzionano perfettamente se abbiamo dei valori precisi e rendono facile l'identificazione di dati individuali

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

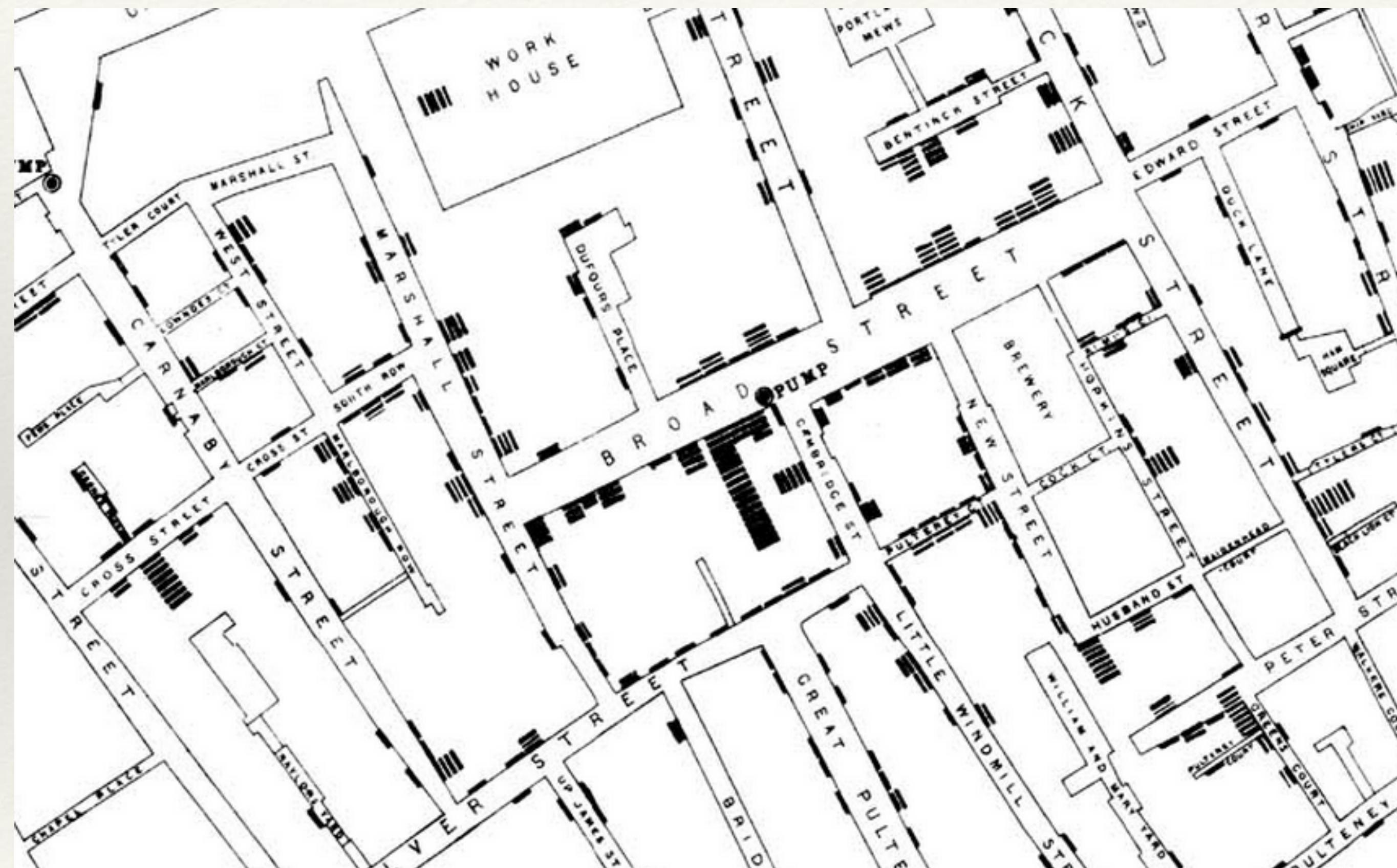
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Domestic	1983	2343	2593	2283	2574	2838	2382	2634	2938	2739	2983	3493
International	574	636	673	593	644	679	593	139	599	583	602	690
	\$2,557	\$2,979	\$3,266	\$2,876	\$3,218	\$3,517	\$2,975	\$2,773	\$3,537	\$3,322	\$3,585	\$4,183

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://tinyurl.com/jvt4mgd>

Il test dello scimpanzè

- ❖ 13 domande di geopolitica
- ❖ 3 risposte possibili per ognuna

Test dello scimpanzè

Prova la tua cultura generale sul mondo!

RICORDA:

Non sarai giudicato se sbagli.

Rispondi con la risposta che credi sia giusta, evitando di "cercare su Internet".

Quando tutti avrete terminato vedremo insieme le risposte corrette e faremo qualche riflessione sui risultati.

147 responses

Accepting responses

Summary

Question

Individual

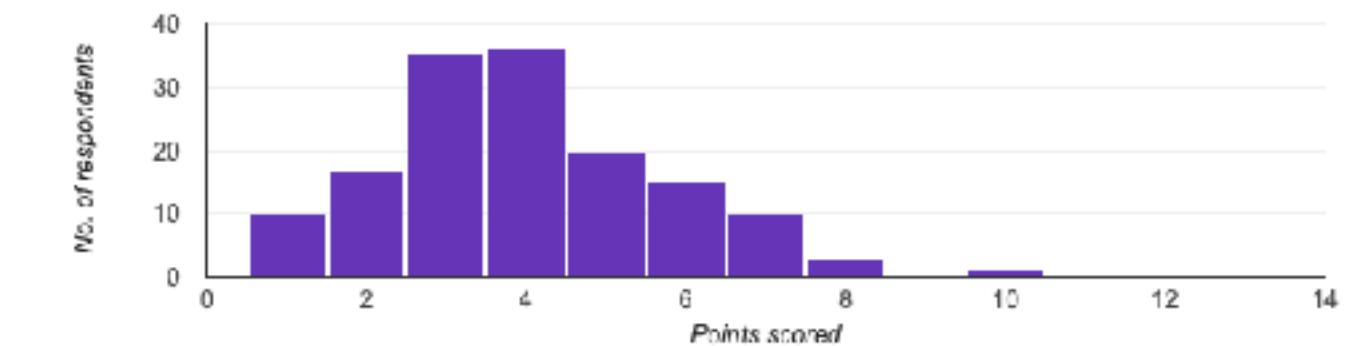
Insights

Average
3.99/13 points

Median
4/13 points

Range
1-10 points

Total points distribution



Frequently missed questions

Question

Correct responses

1. Nelle nazioni che oggi nel mondo sono considerate a basso reddito, quante ragazze terminano le scuole elementari? 20/147

2. Dove vive la maggioranza della popolazione mondiale? 55/147

3. Negli ultimi 20 anni, la proporzione della popolazione mondiale che vive in estrema povertà è 29/147

4. Qual è oggi l'aspettativa di vita nel mondo? 66/147

5. Ci sono circa 2 miliardi di bambini oggi nel mondo, tra i 0 ed i 15 anni. Quanti bambini ci saranno nel 2100, secondo le Nazioni Unite? 49/147

The chimpanze test

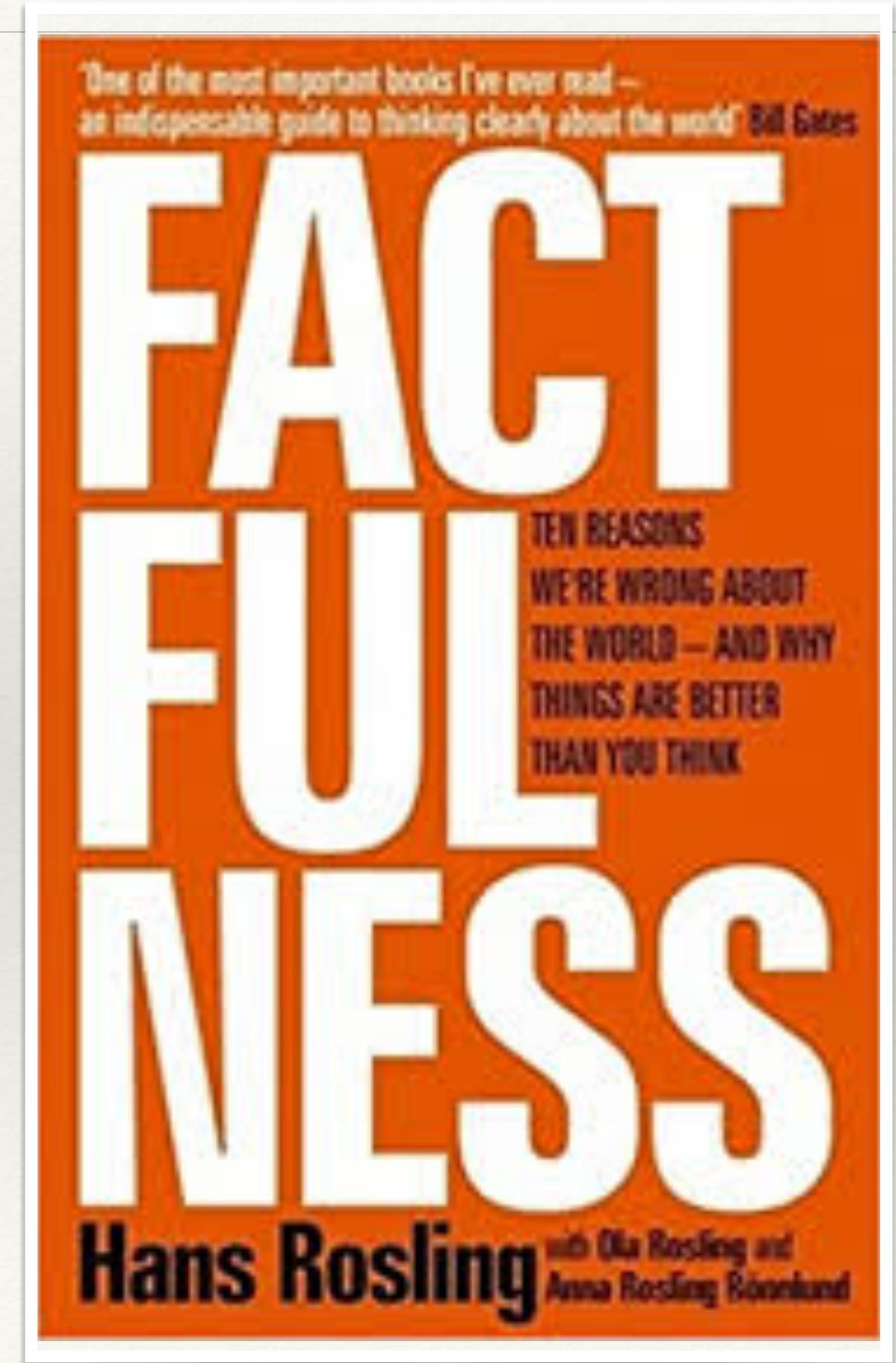
- ❖ I nostri pregiudizi peggiorano la nostra capacità di analisi
- ❖ Dobbiamo approcciare la visualizzazione e la comunicazione scientifica nel modo più “onesto” possibile: prima ancora di mentire agli altri, rischiamo di mentire a noi stessi.
- ❖ ... non abbiamo neanche iniziato a parlare dei limiti 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 visualizzazioni in modo efficace

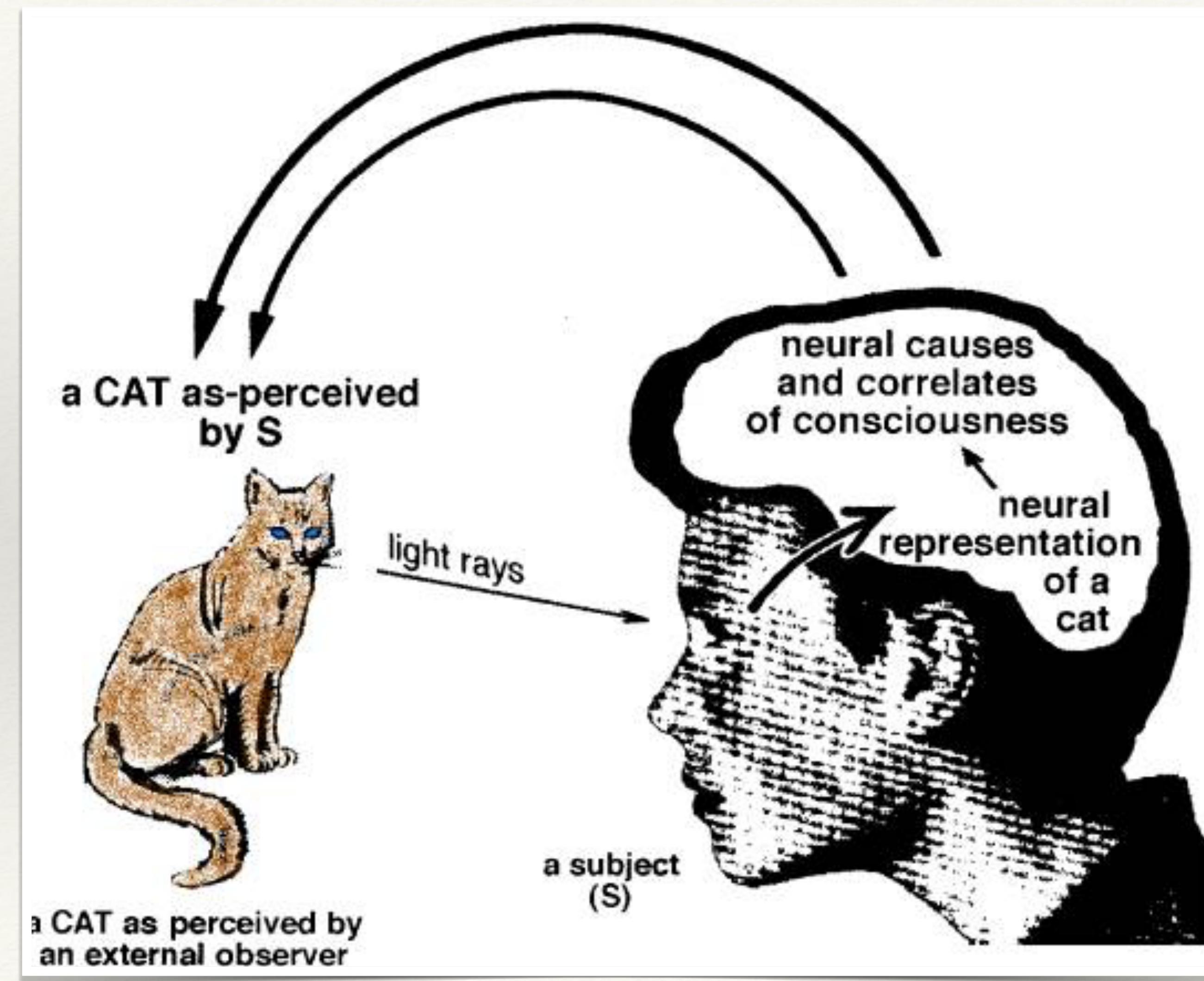
“We must do more than simply display data graphically: we must understand how visual perception works and then present data visually in ways that follow the rules”

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 percepirllo 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?

24813481187116715541388198443771347915641531845305848641
23475789411484122238814691613548048407890877078678751211
86584234044377134791564153184530584864123475789411484122
23881469161354804840789087707867875121186584234018874276

Quanti 3 vedete nella sequenza sottostante?

24813481187116715541388198443771347915641531845305848641
23475789411484122238814691613548048407890877078678751211
86584234044377134791564153184530584864123475789411484122
23881469161354804840789087707867875121186584234018874276

Lezione #2

“Non limitiamoci a tutto quello che possiamo vedere. La percezione visiva è selettiva ed è giusto che sia così, 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?

“Our eyes are drawn to familiar patterns.
We see what we know and expect”



Cos'altro vedete qui?

“Our eyes are drawn to familiar patterns.
We see what we know and expect”



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

“Our eyes are drawn to familiar patterns.
We see what we know and expect”



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

“Our eyes are drawn to familiar patterns.
We see what we know and expect”



Lezione #3

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



<https://www.youtube.com/watch?v=oH6yOc606uQ>

Lezione #4

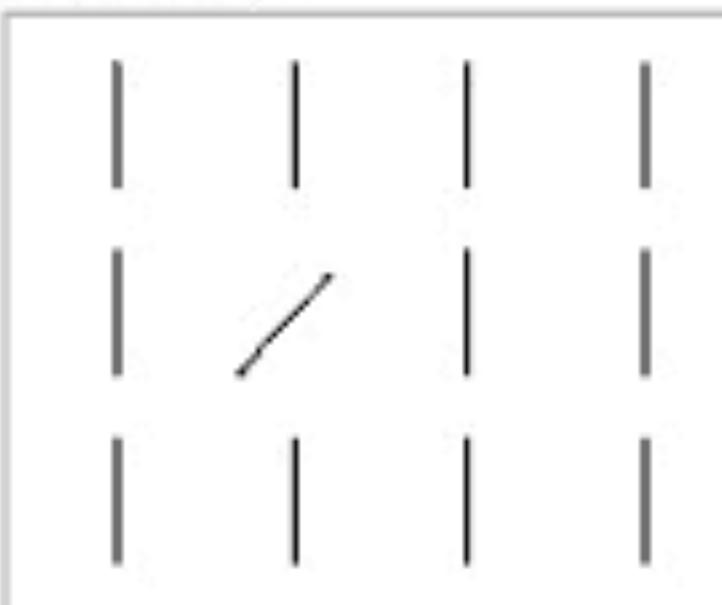
“La memoria gioca un ruolo fondamentale nella cognizione umana, ma la memoria utile è estremamente limitata”

Attributi “pre-attentive”

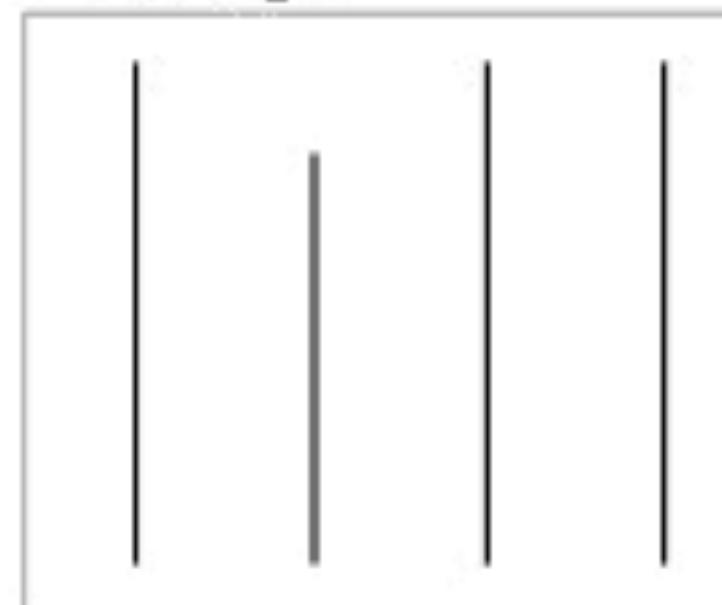
- ❖ Quattro categorie
 - ❖ Colori
 - ❖ Forme
 - ❖ Posizioni
 - ❖ Animazioni
- ❖ Associa valori e dati ad attributi “pre-attentive” con cautela:
 - ❖ Non creare DISTRAZIONI dai dati
 - ❖ Tieni in considerazione la memoria a breve termine

Form

Orientation



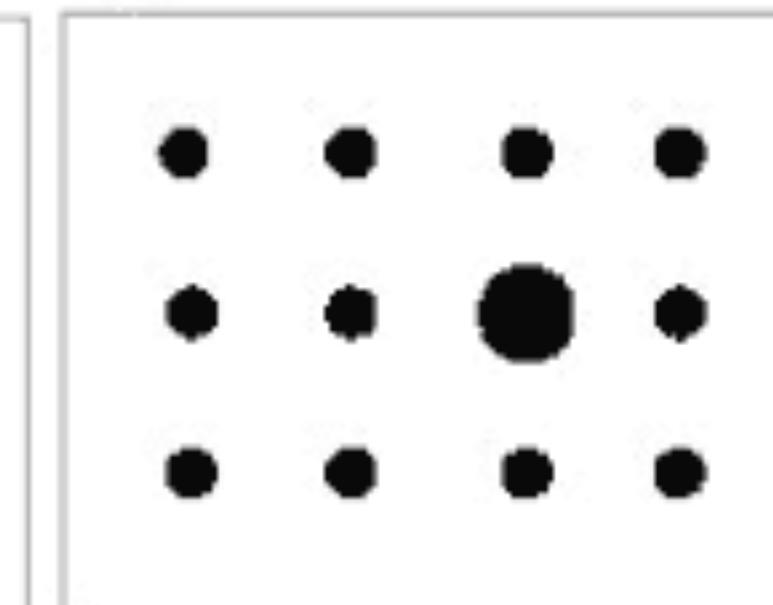
Line Length



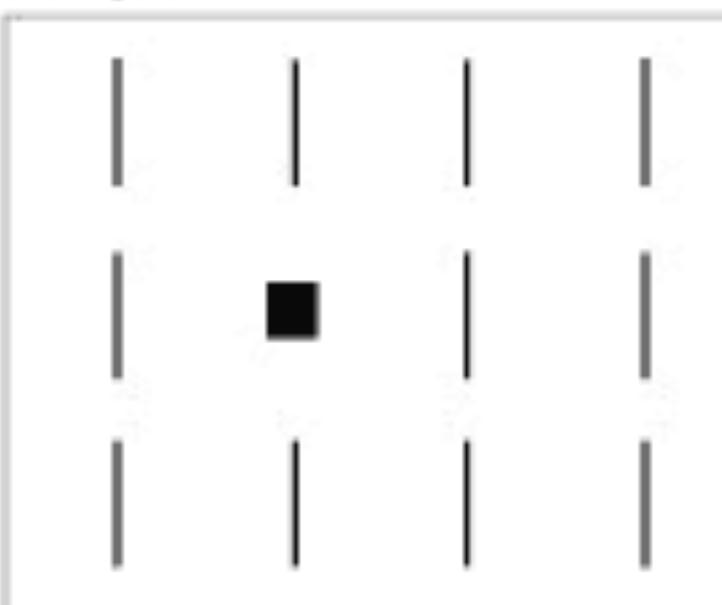
Line Width



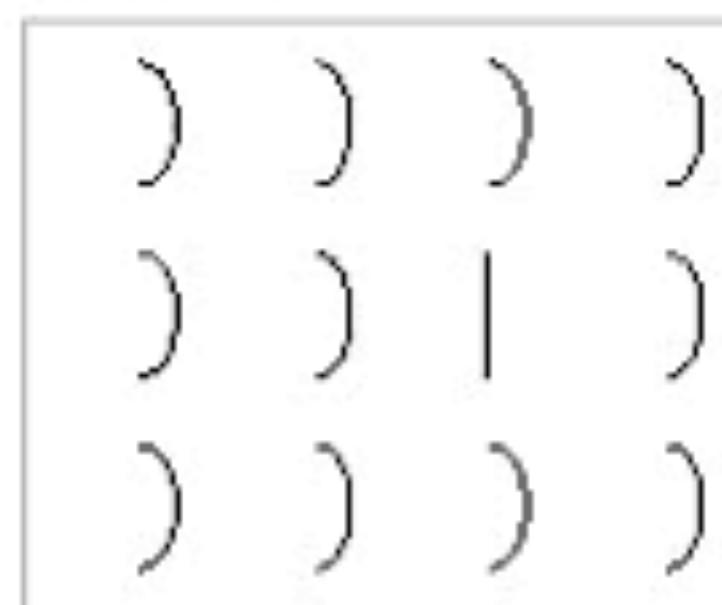
Size



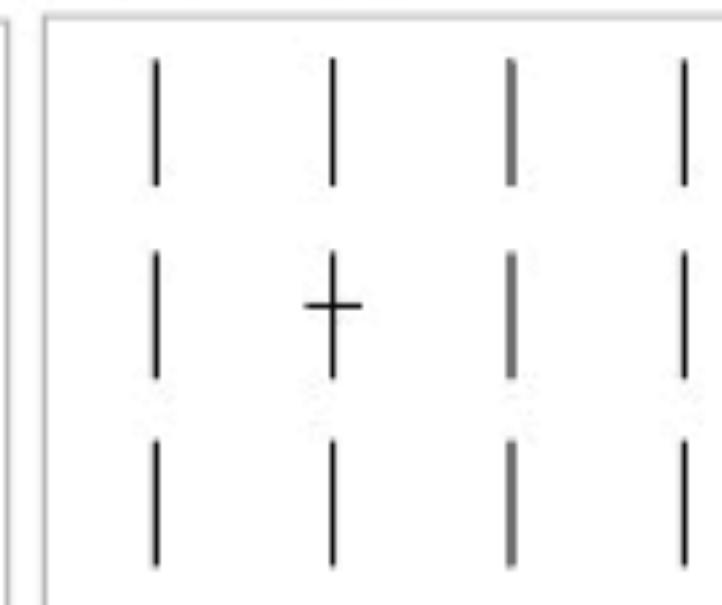
Shape



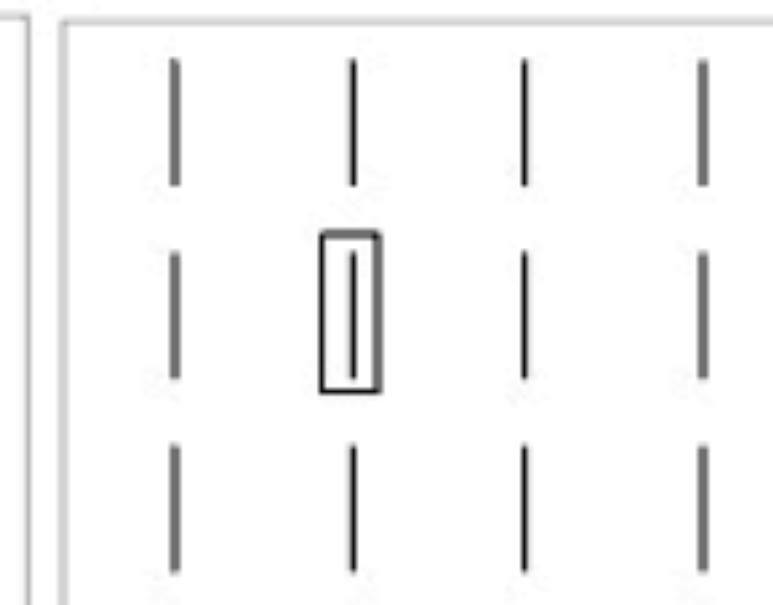
Curvature



Added Marks

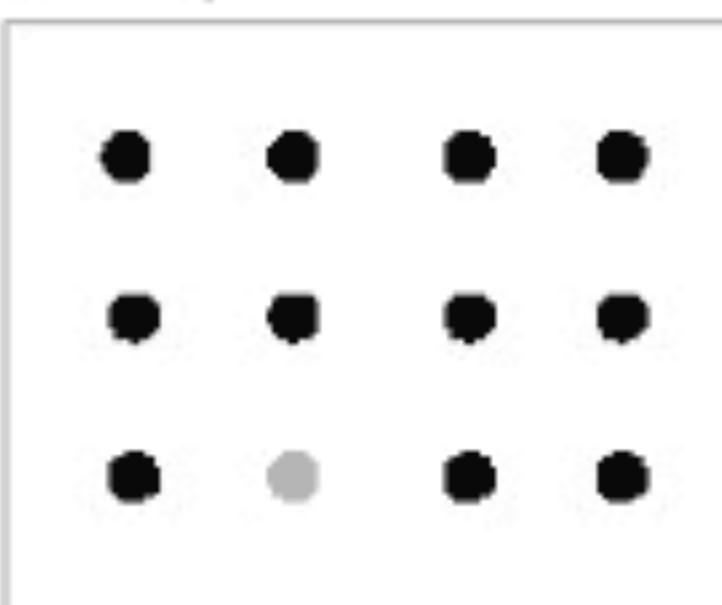


Enclosure

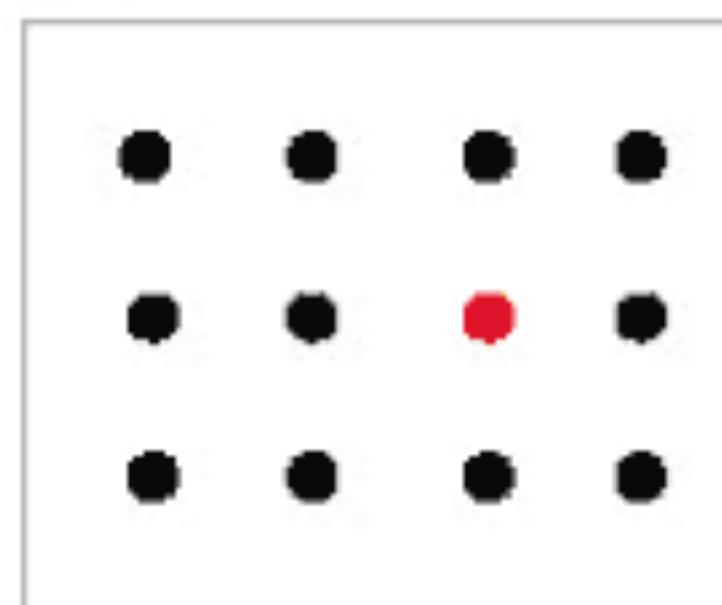


Color

Intensity



Hue

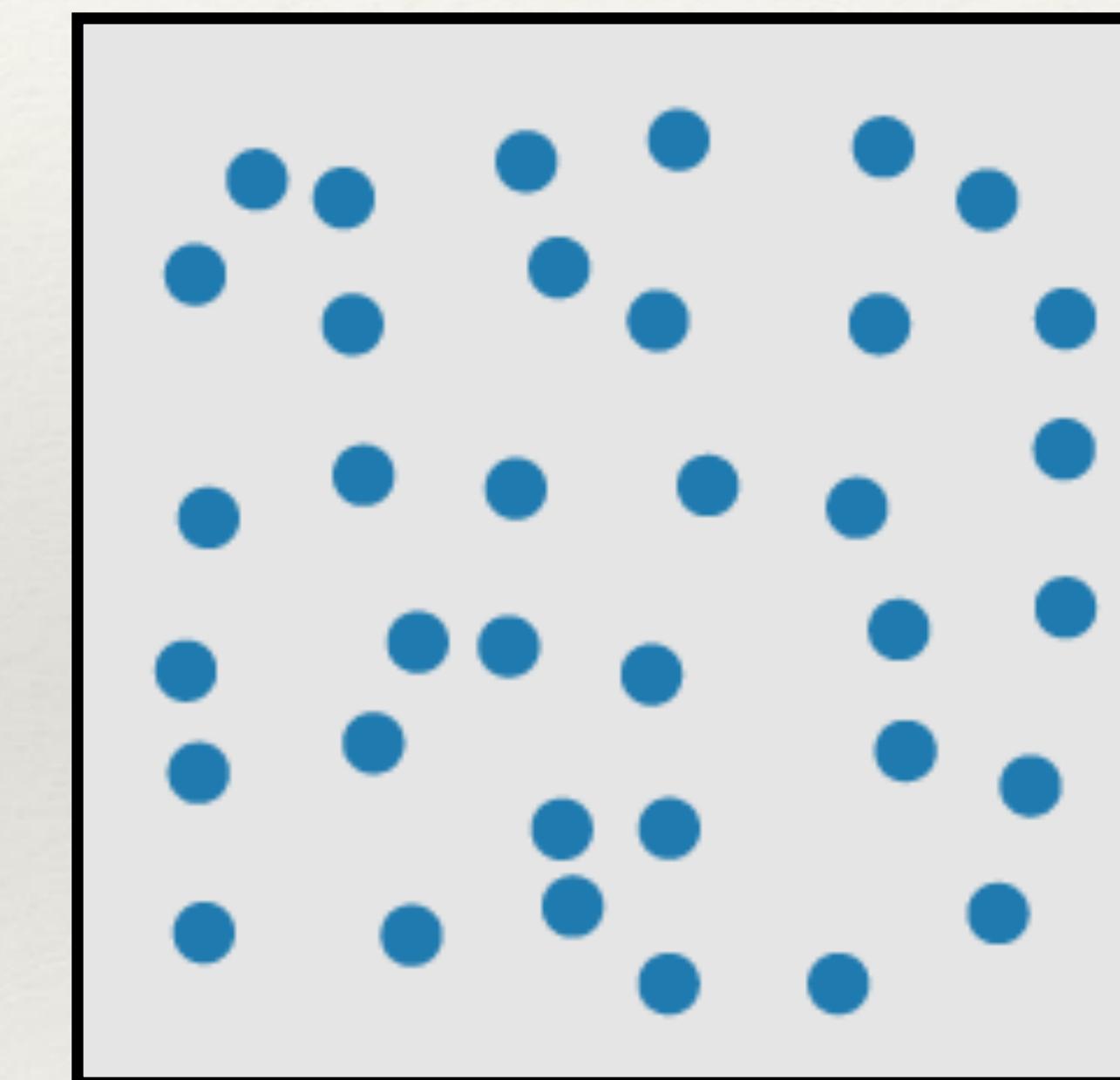
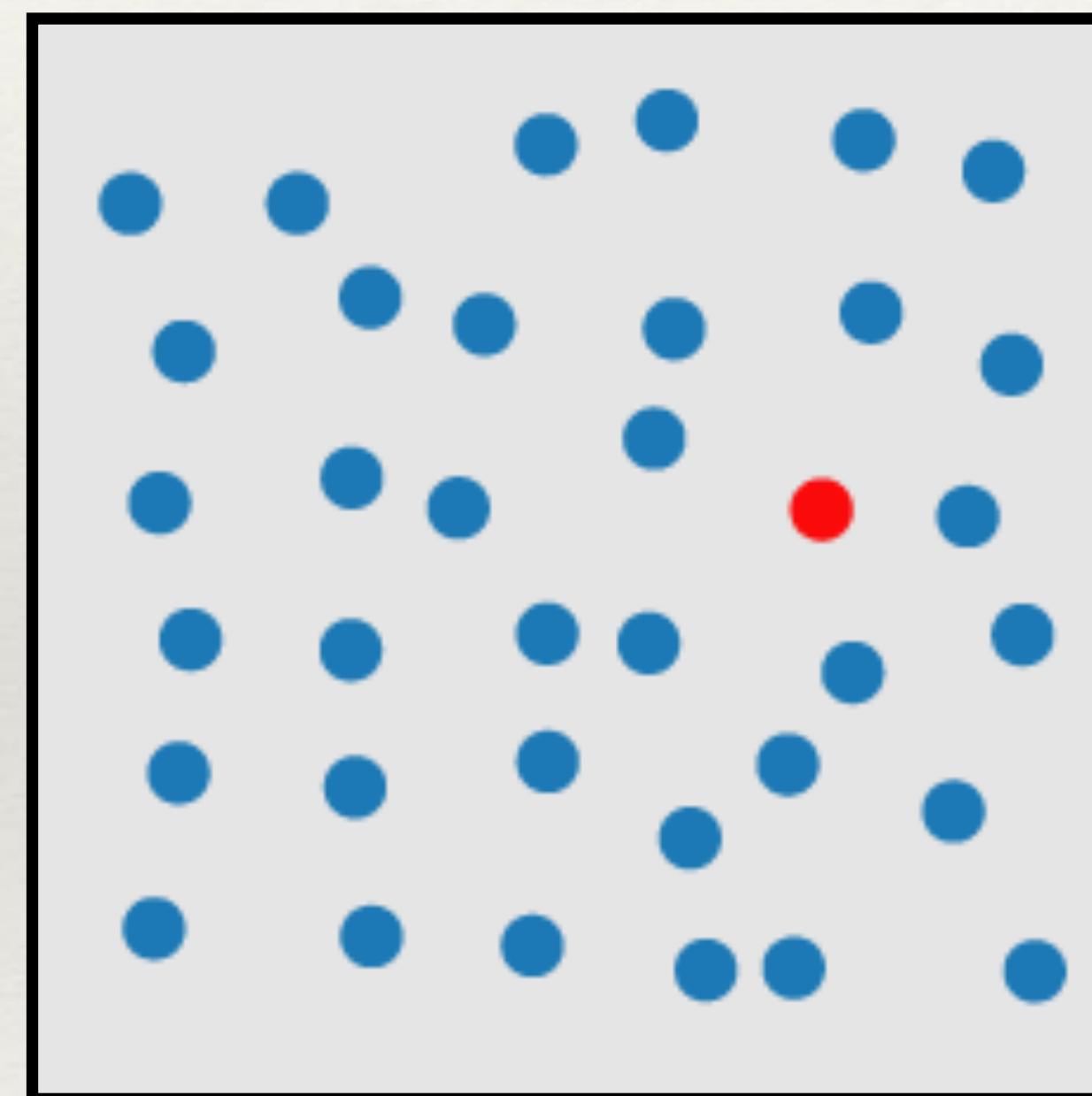


Spatial Position

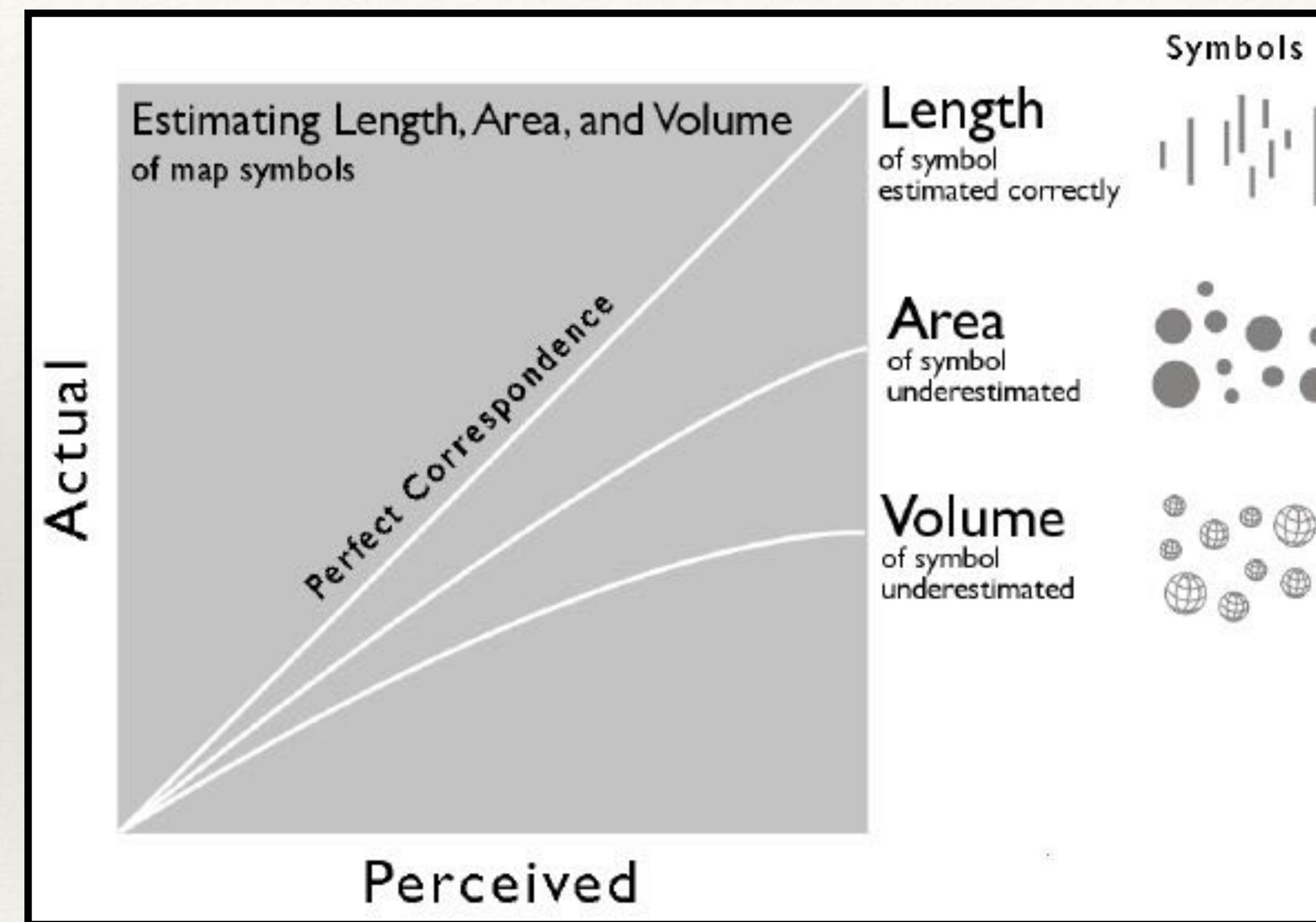
2-D Position



Esempio



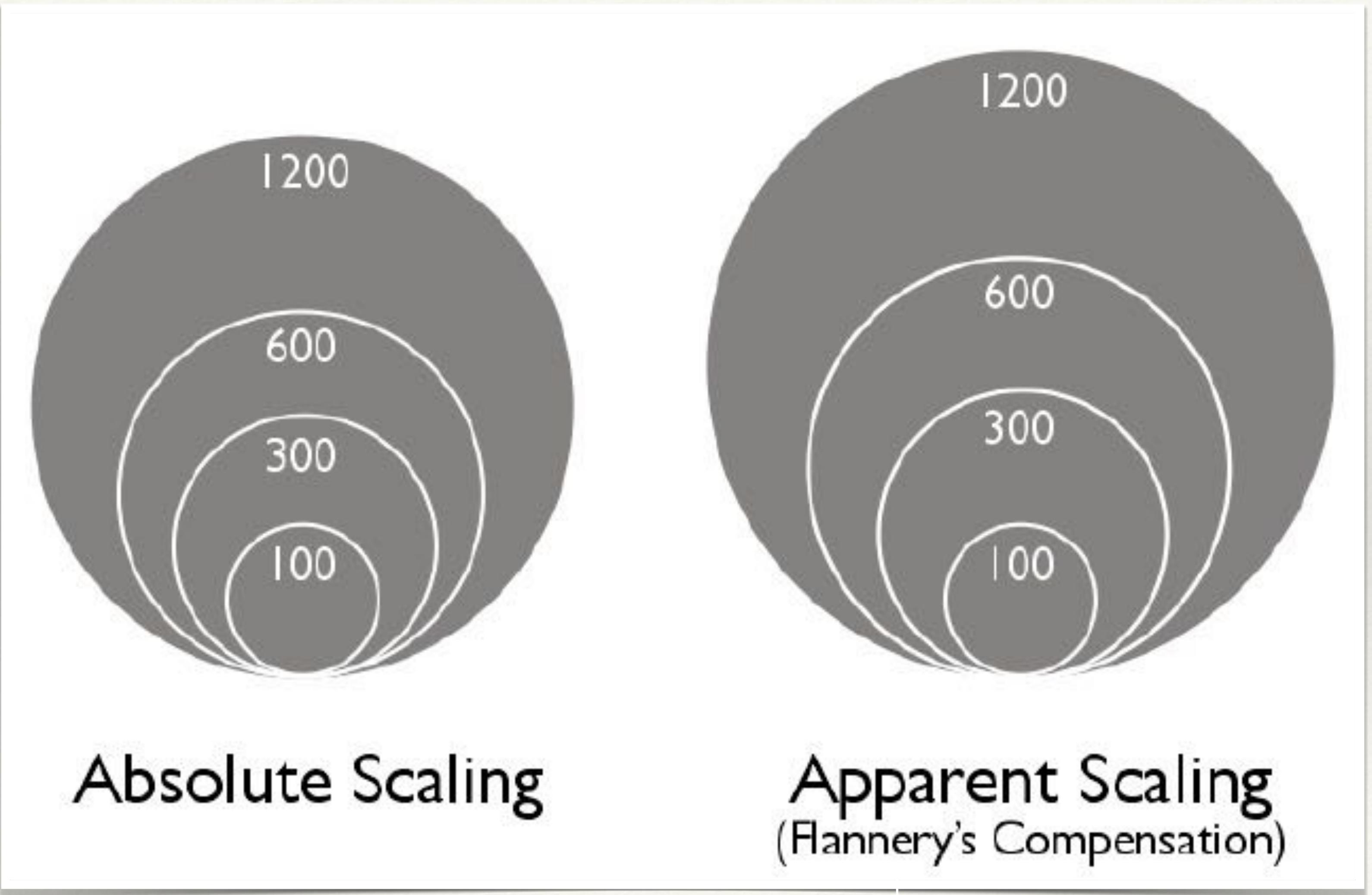
Forme



<https://makingmaps.net/2007/08/28/perceptual-scaling-of-map-symbols/>

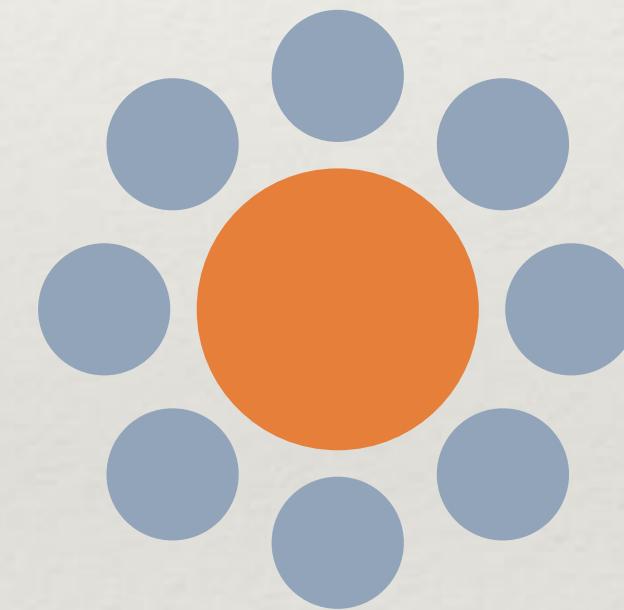
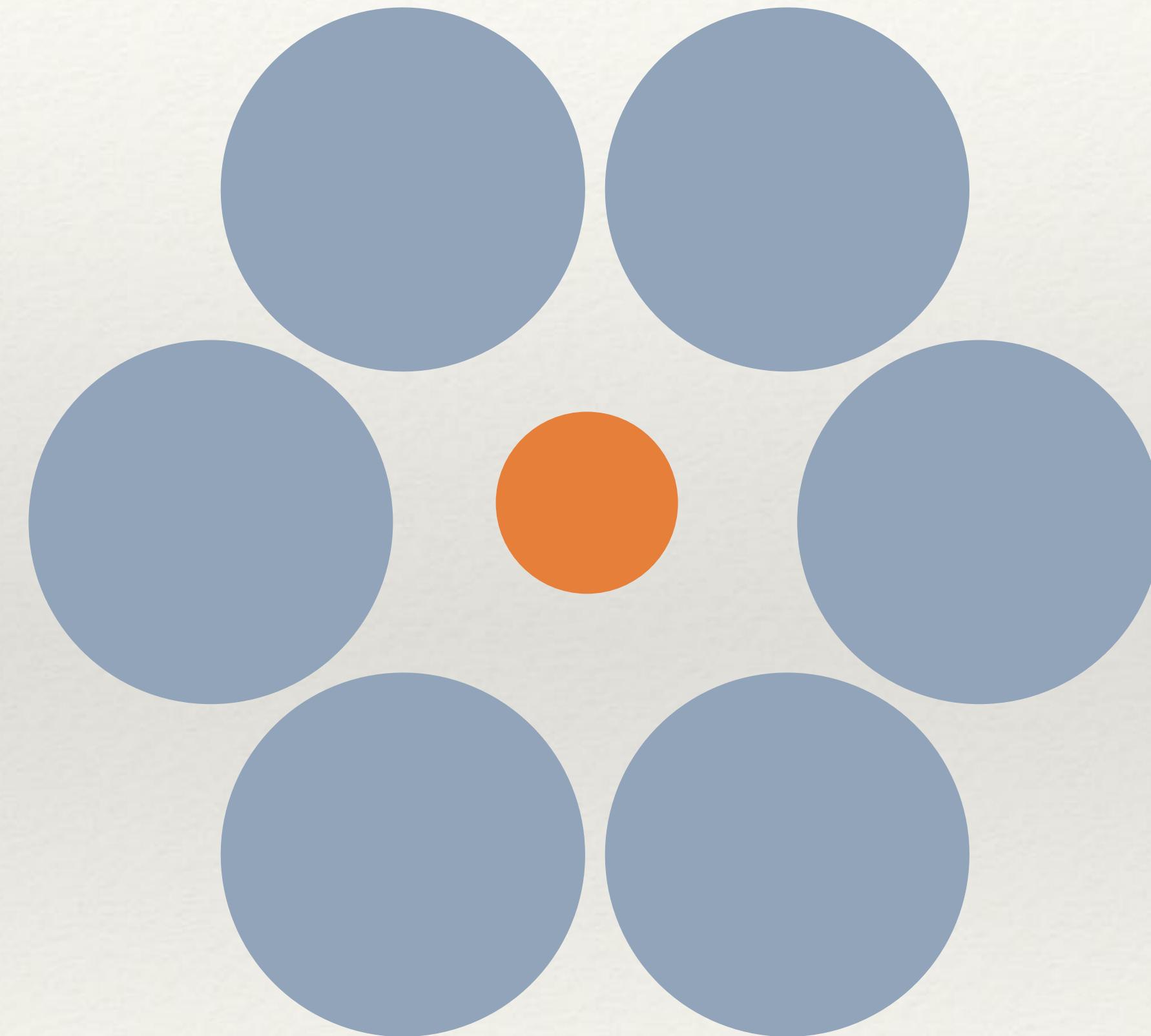
Forme

- ❖ Dimensioni effettive e percepite nei cerchi



Forme

- ❖ Dimensioni effettive e percepite nei cerchi



Animazioni

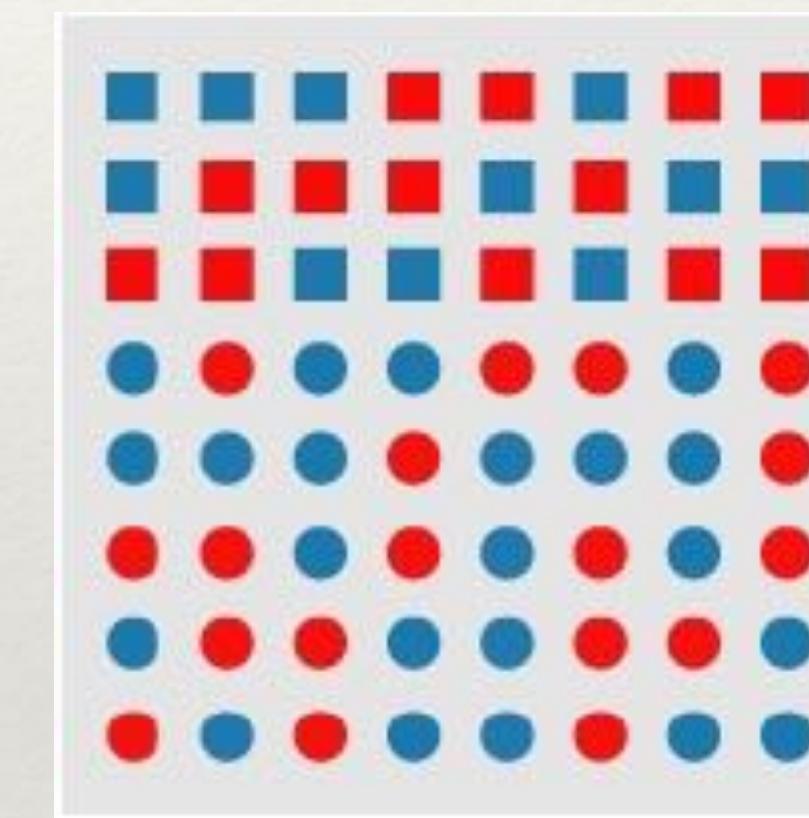
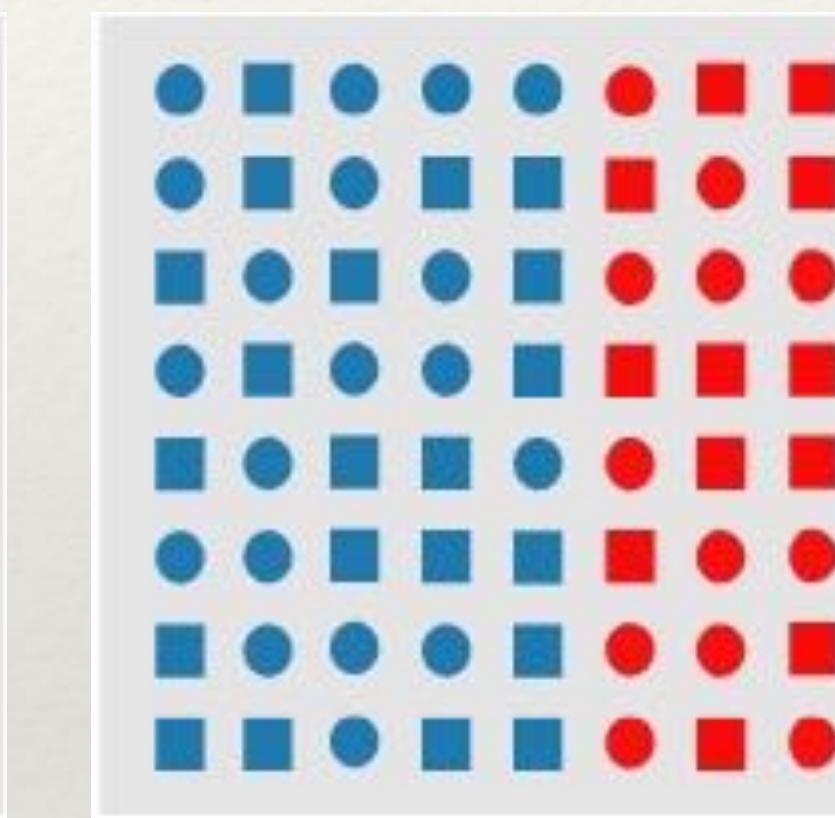
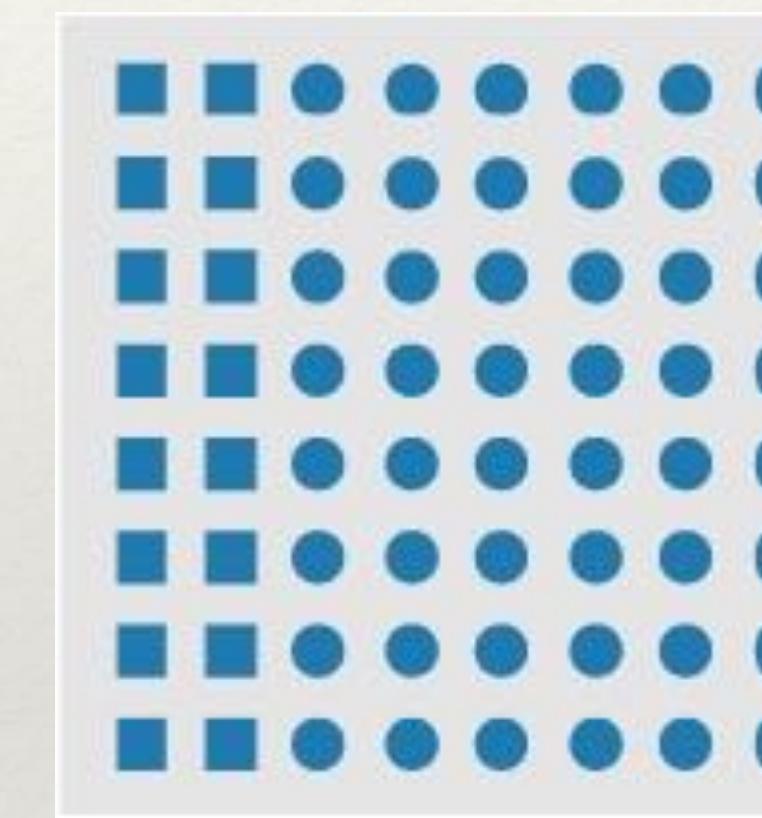
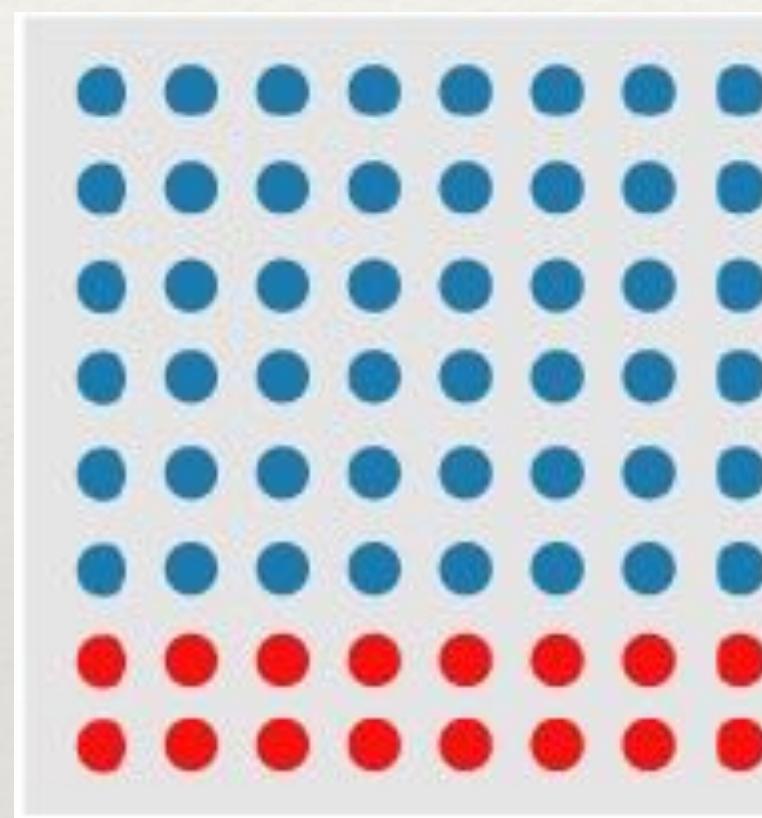
- ❖ Due attributi
 - ❖ Intermittenza (compare / scompare)
 - ❖ Movimento (da una posizione all'altra)
- ❖ Uno dei modi più efficaci per attirare l'attenzione
- ❖ Molto spesso abusato nel marketing

Attenzione e movimento

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

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

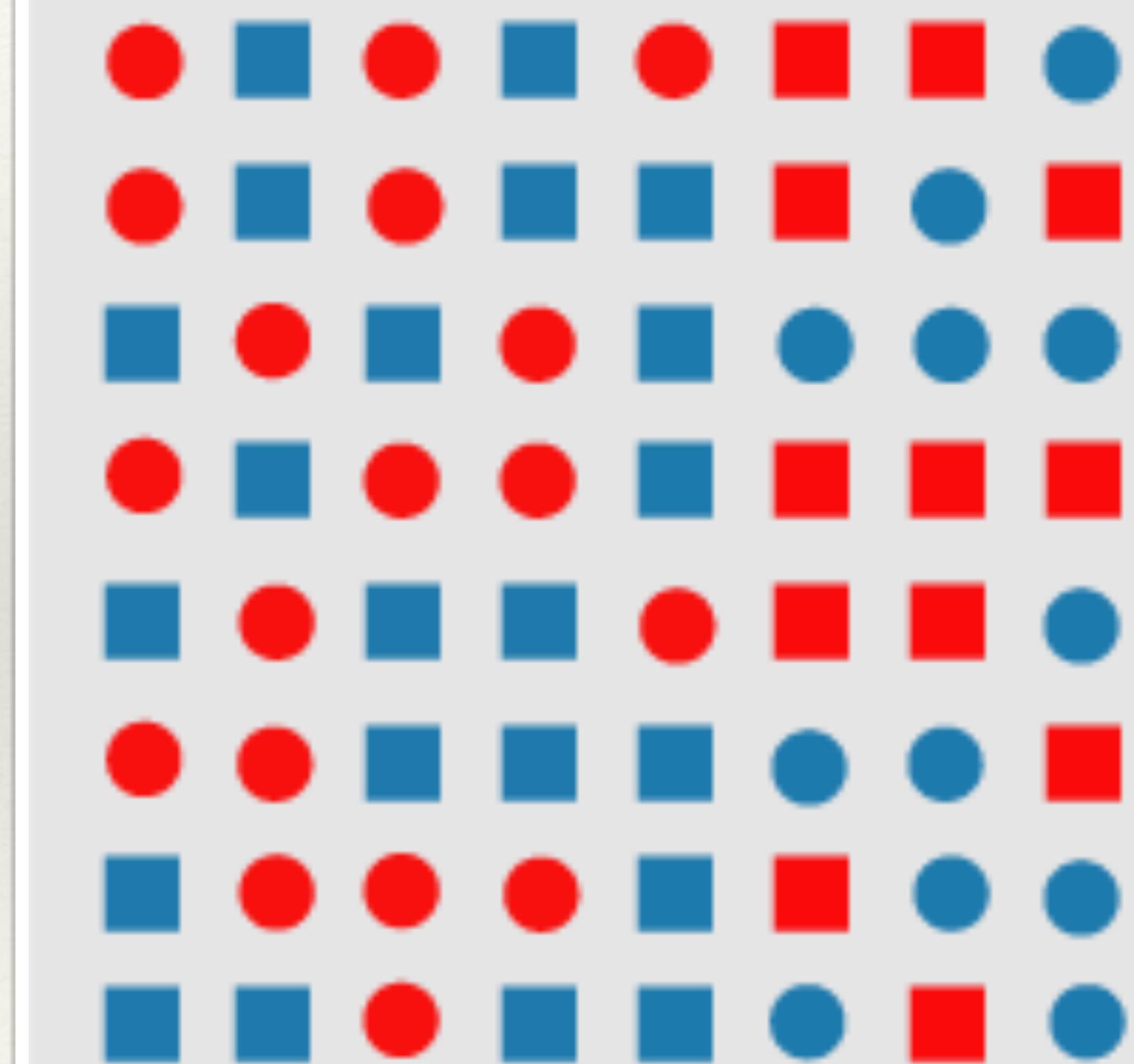
Contorni e margini



<https://www.csc2.ncsu.edu/faculty/healey/PP/>

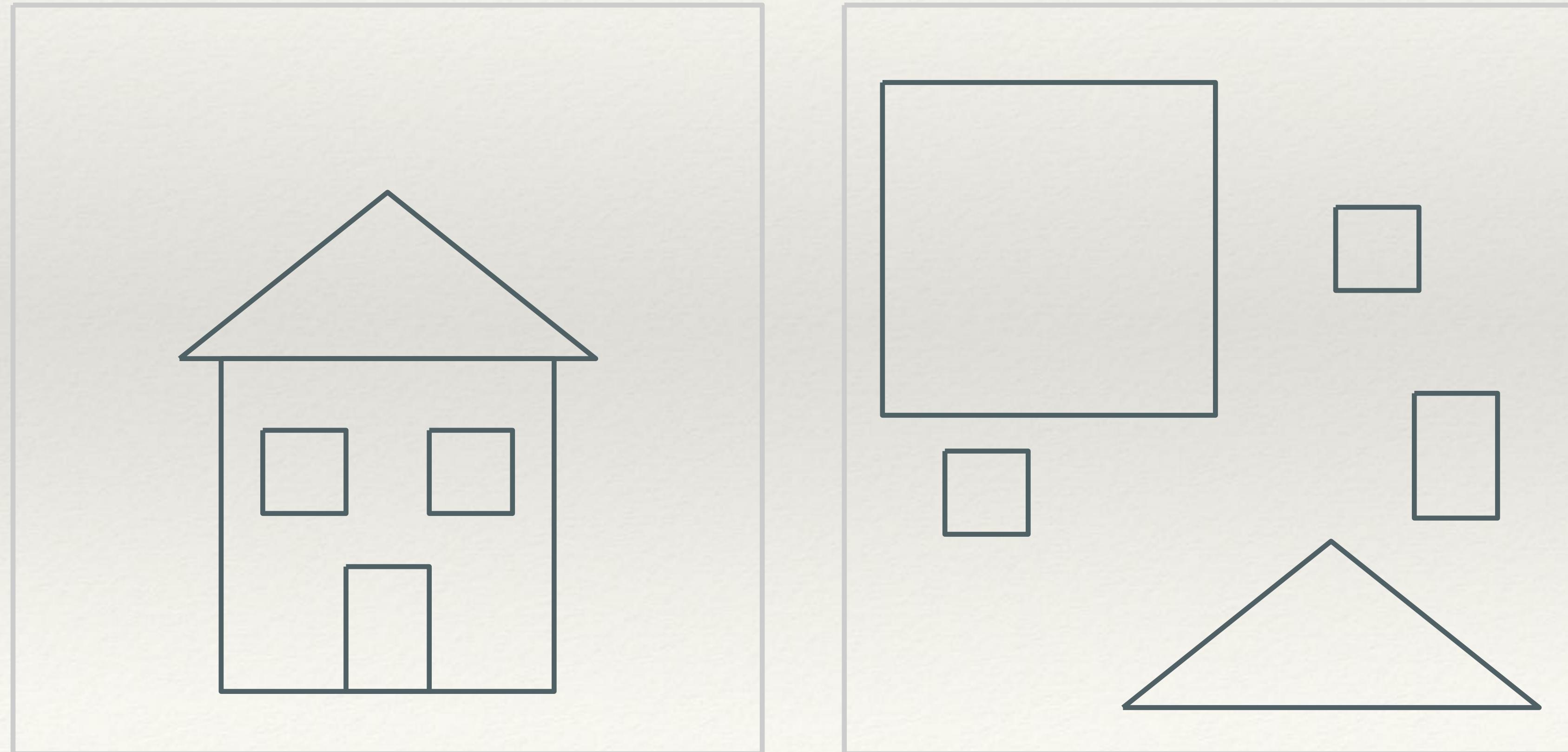
Contorni e margini (2)

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



Gestalt

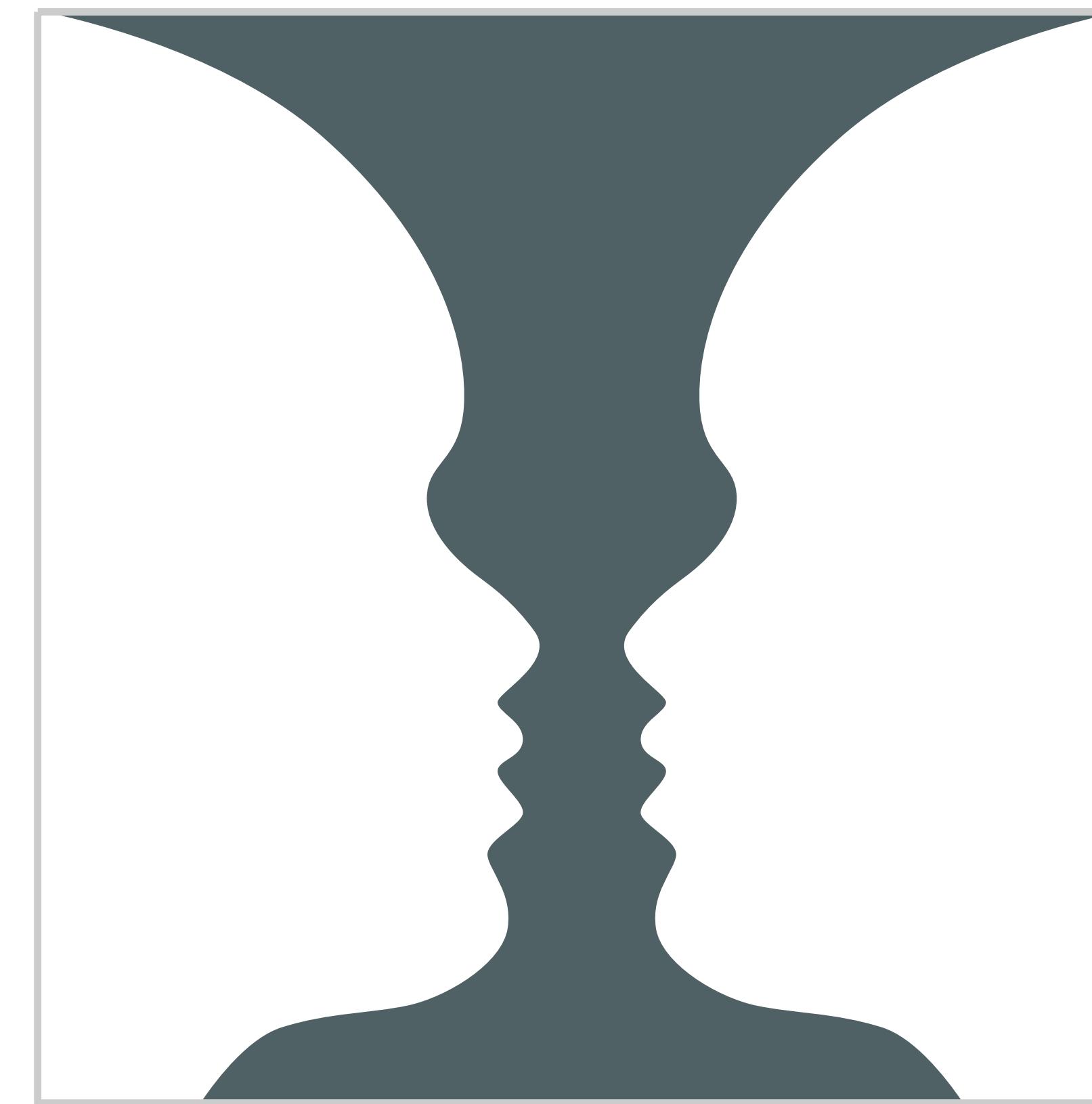
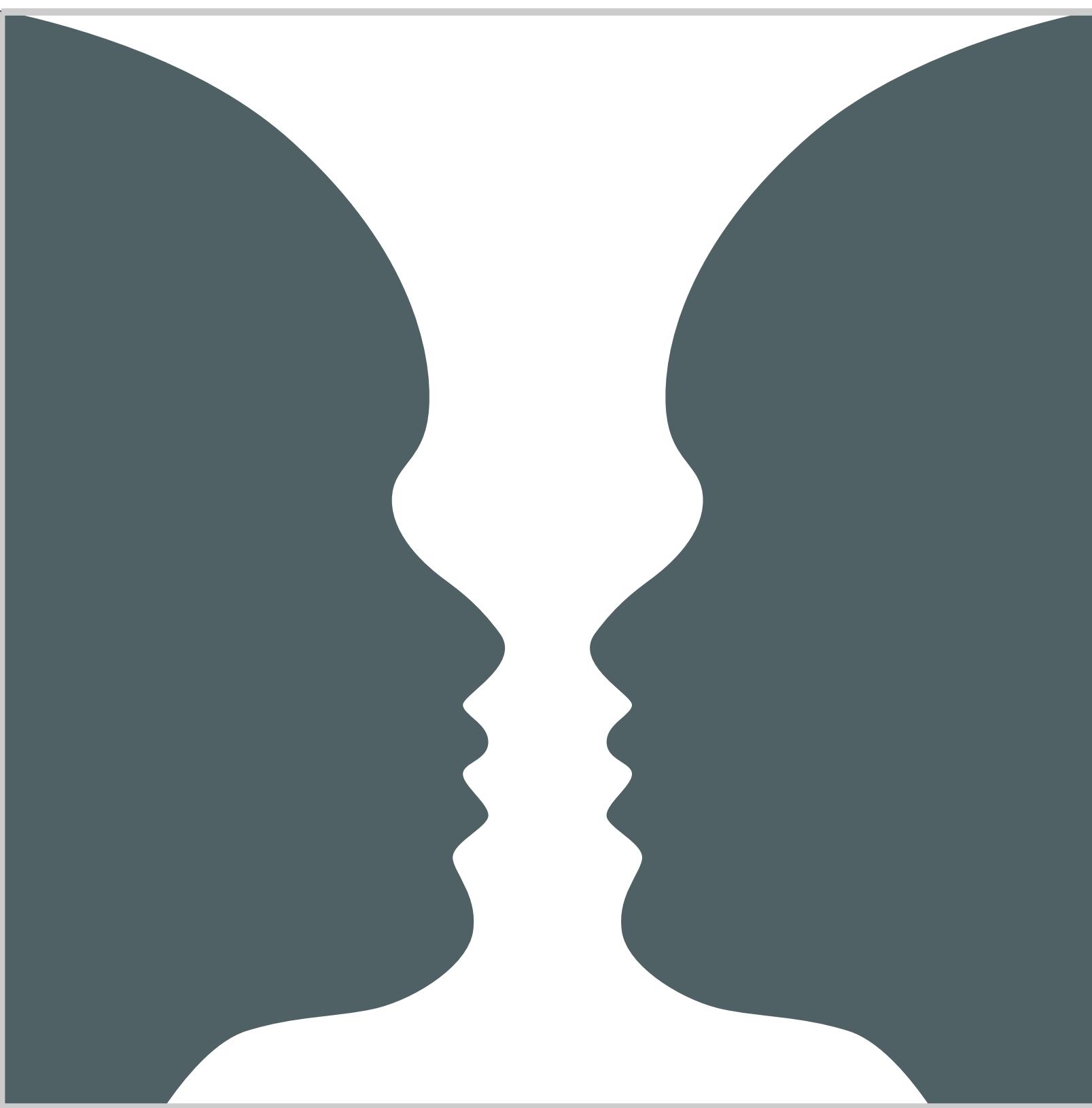
Tendenza a percepire oggetti come parti di un tutto



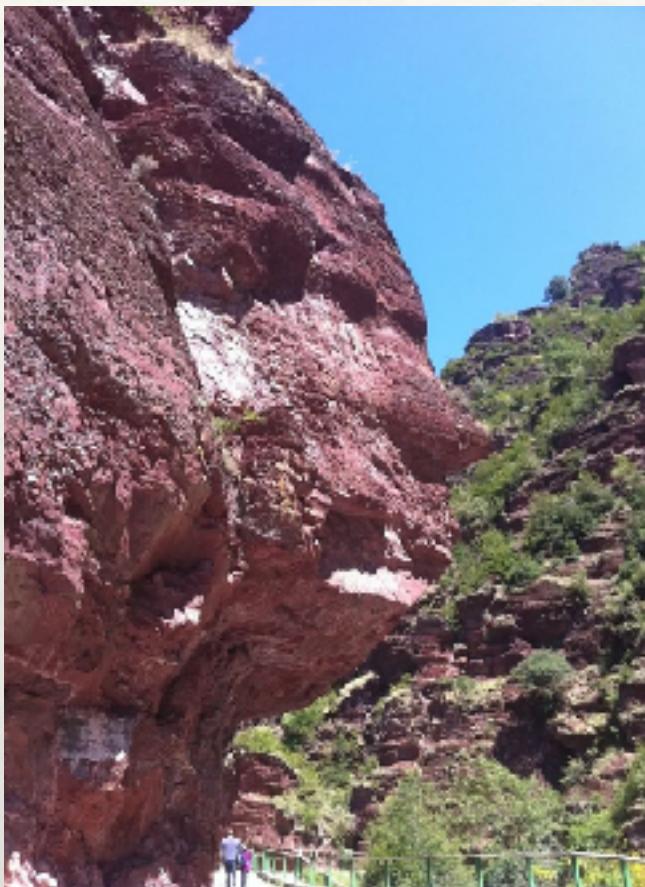
Chiusura

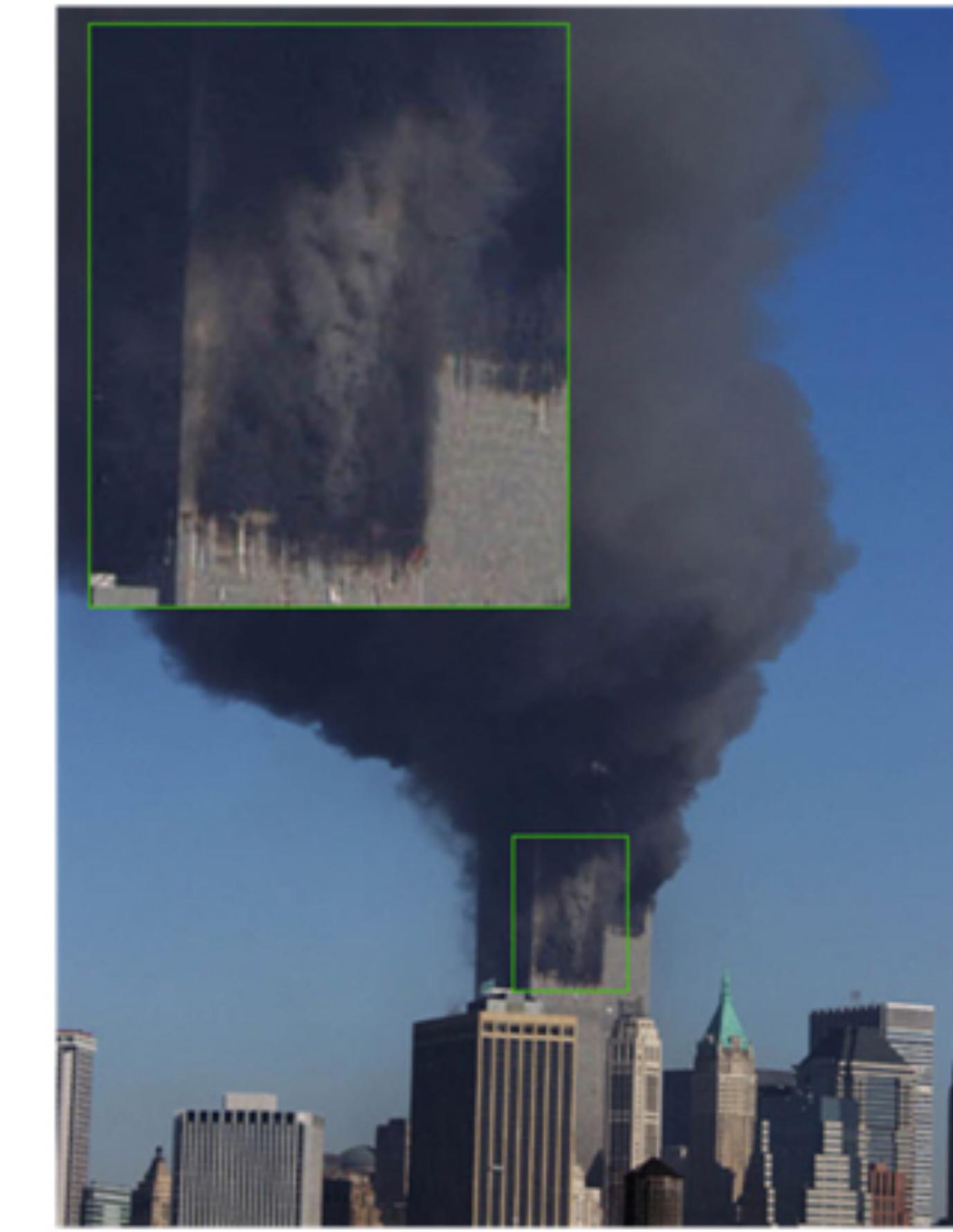
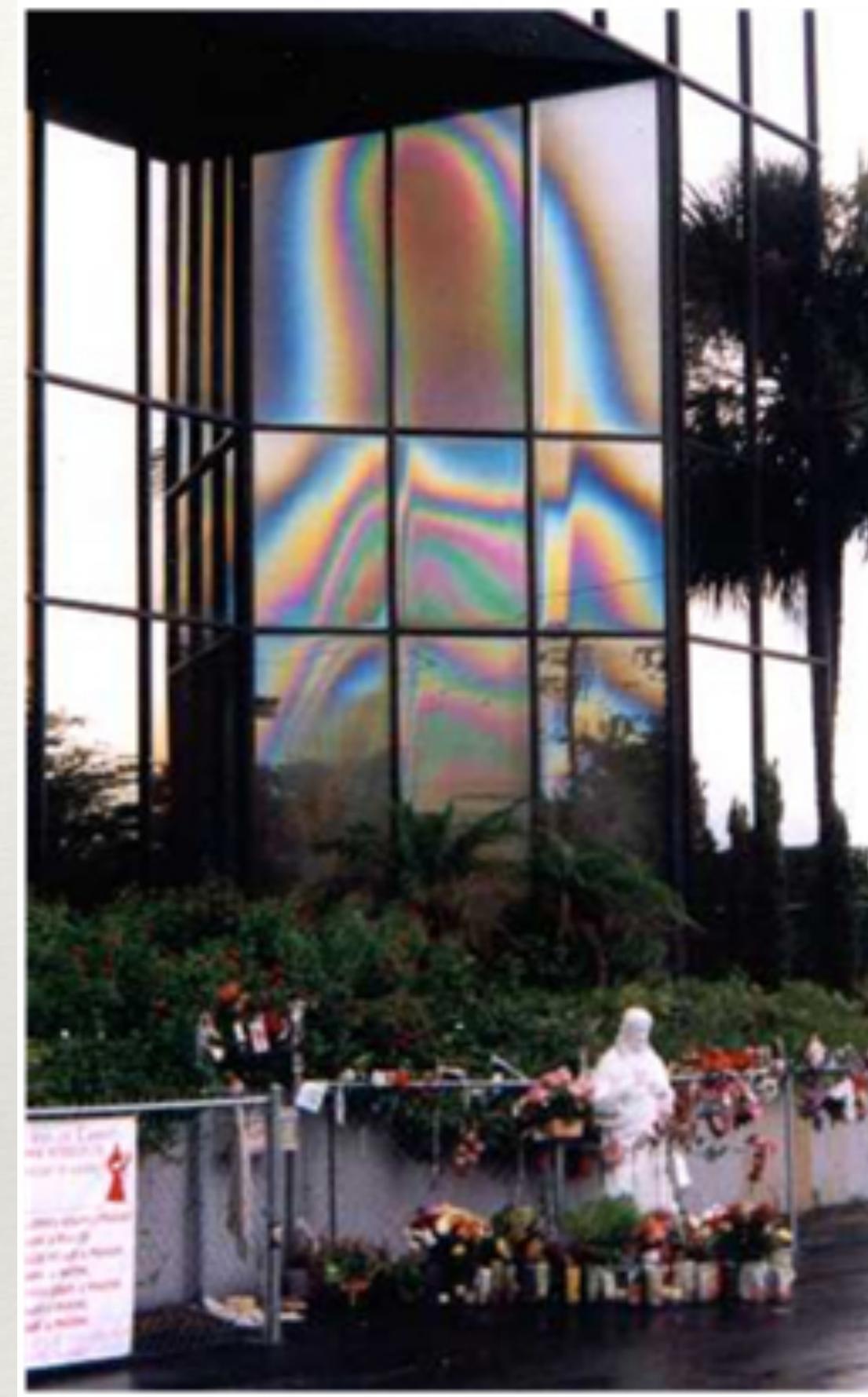


Figure e sfondo



Pareidolia





Effetto spesso usato per supportare credenze, superstizioni o teorie del complotto.

Riferimenti

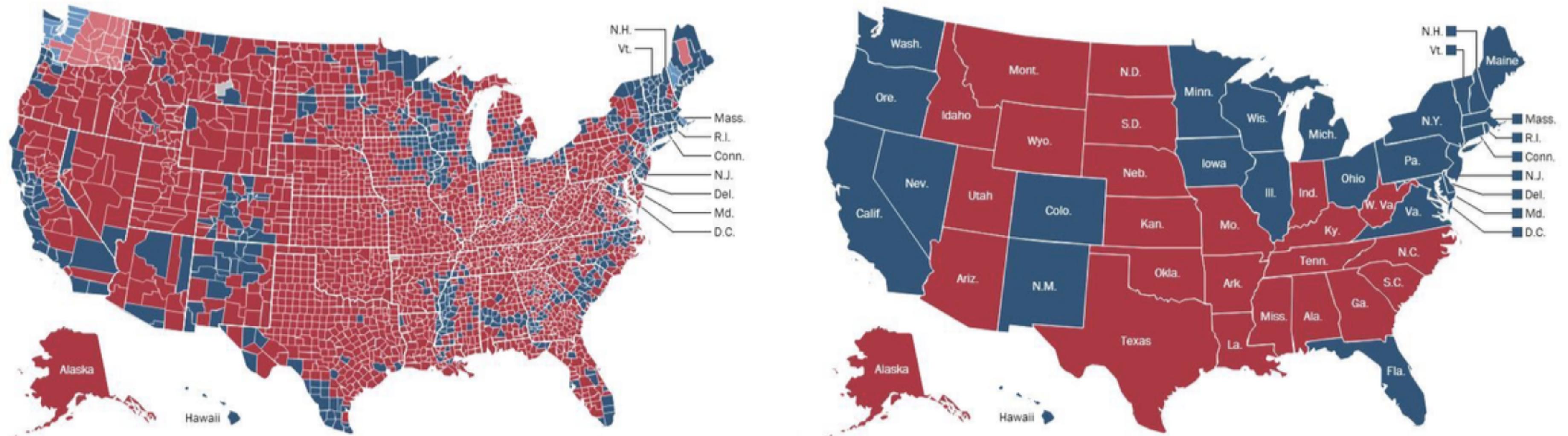
<https://en.wikipedia.org/wiki/Pareidolia>

<http://gestaltrevision.be/en/master-index/66-what-we-do/overview/research-areas/mid-level/multistability/71-pareidolia>

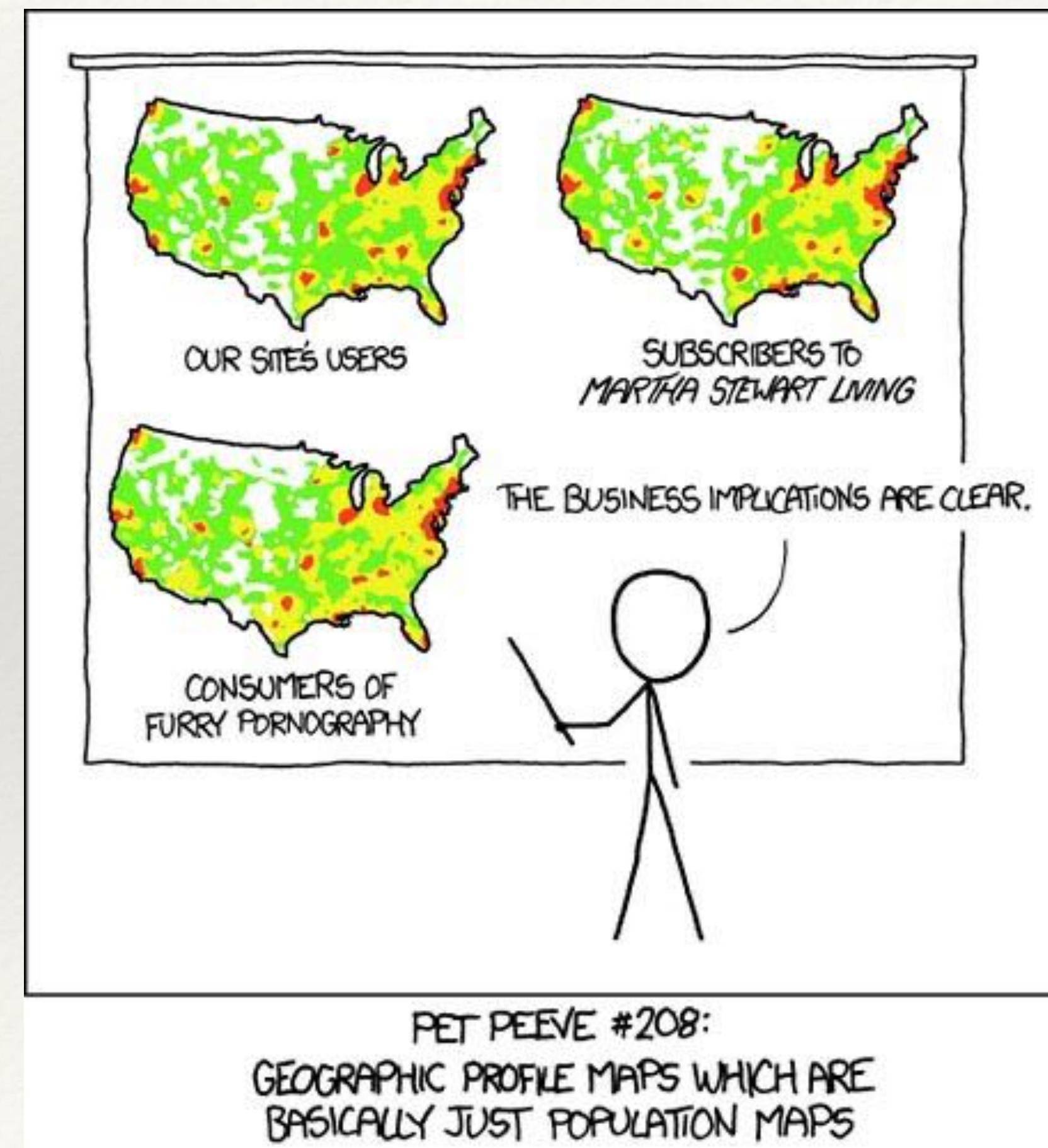
Lezione #5

“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”

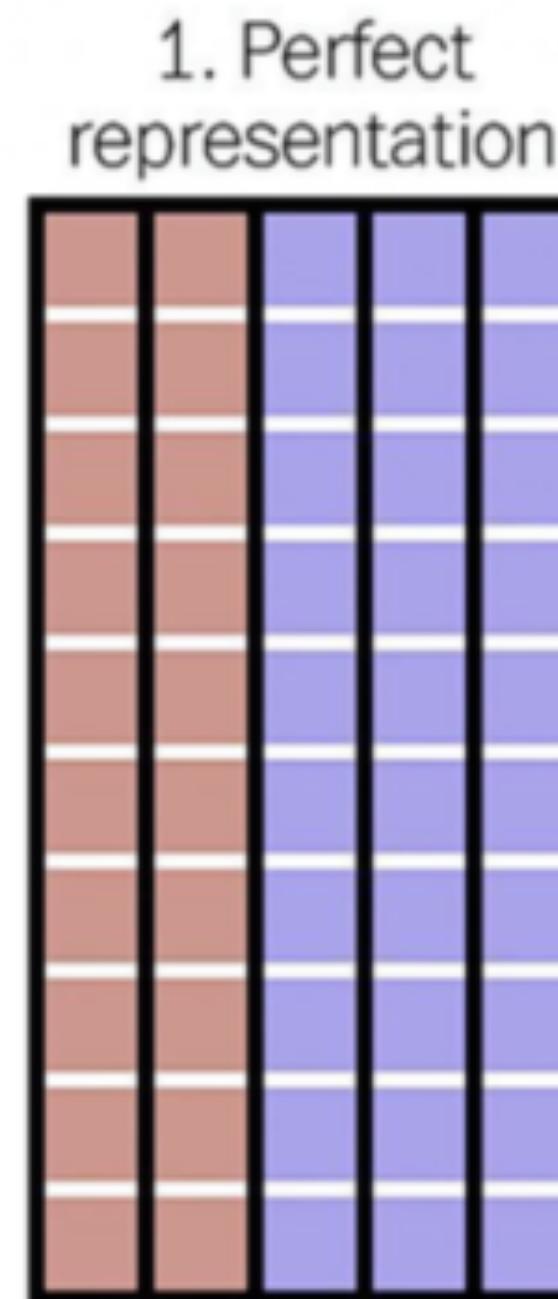
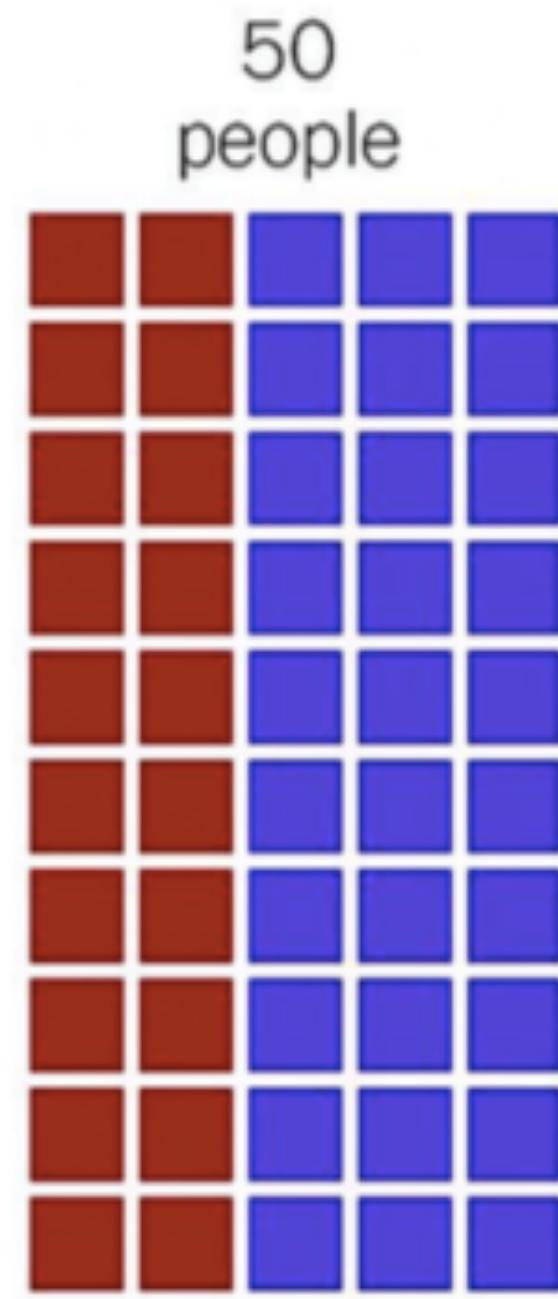
Raccontare storie diverse con stessi dati



anche "mappe di calore" e dotmap hanno problemi



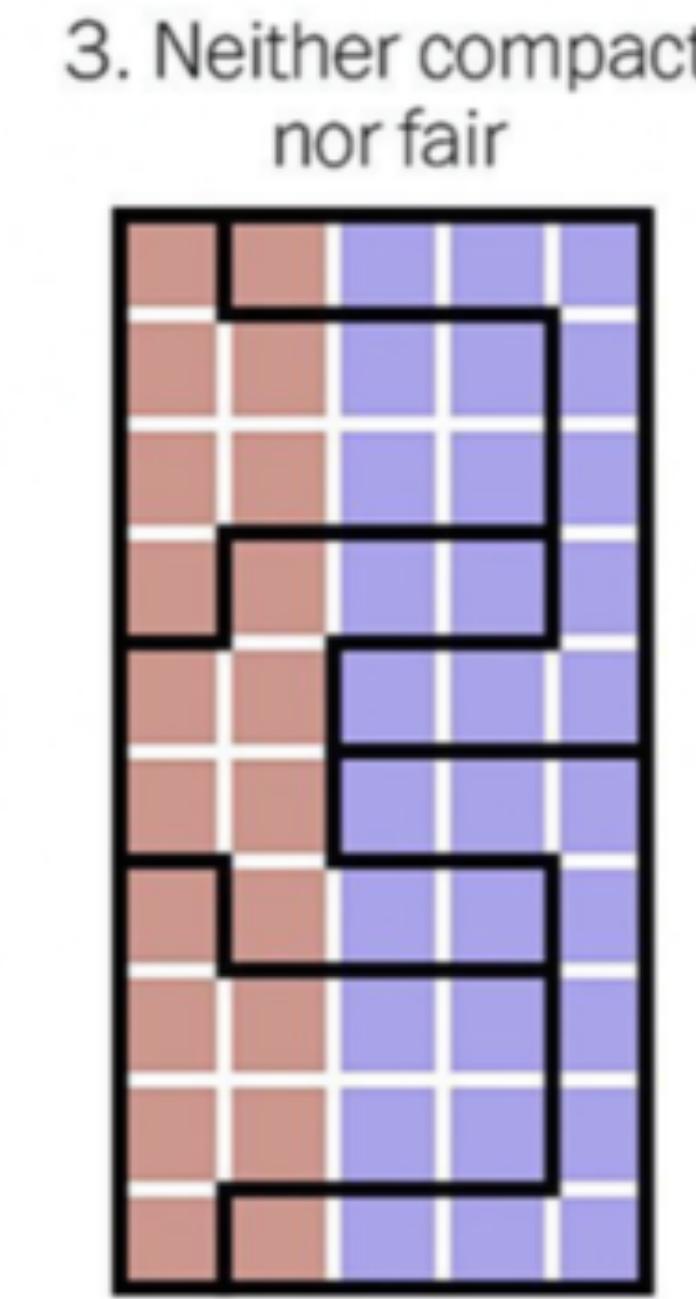
gerrymandering



BLUE WINS



BLUE WINS



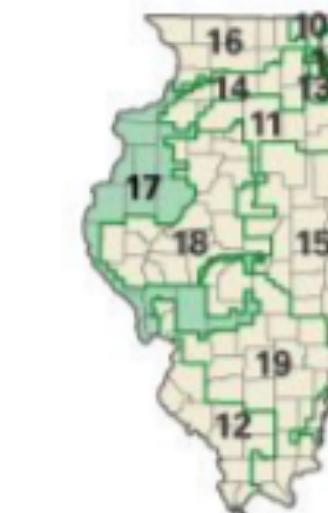
RED WINS

Congressional District 17



nationalatlas.gov™

17 Congressional District
Fulton County



Illinois (19 Districts)

Congressional District 2



nationalatlas.gov™

2 Congressional District
Grand County



Utah (3 Districts)

Congressional District 22



nationalatlas.gov™

22 Congressional District
Harris County



Texas (32 Districts)

Congressional District 12



nationalatlas.gov™

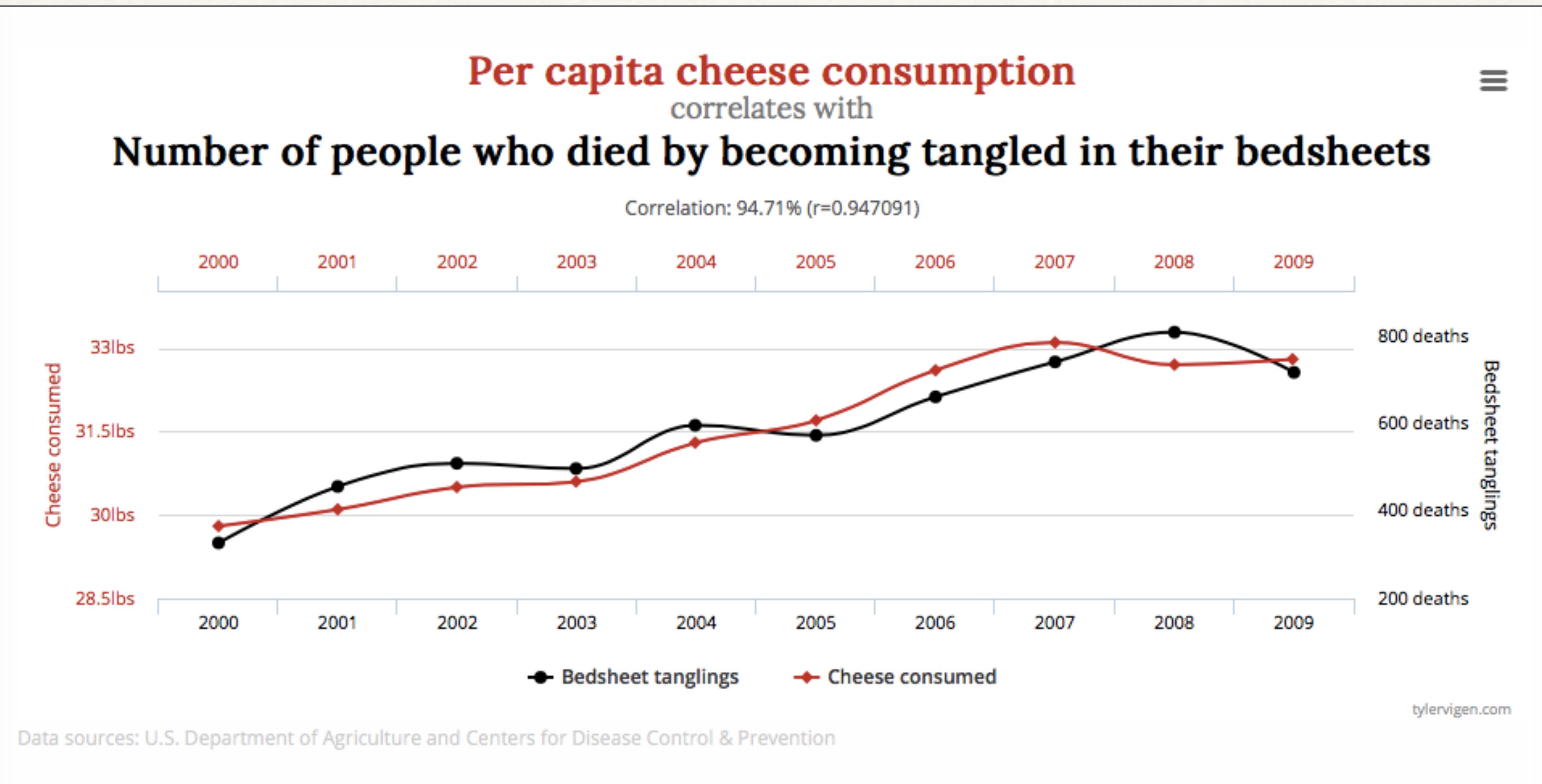
12 Congressional District
Rowan County



North Carolina (13 Districts)

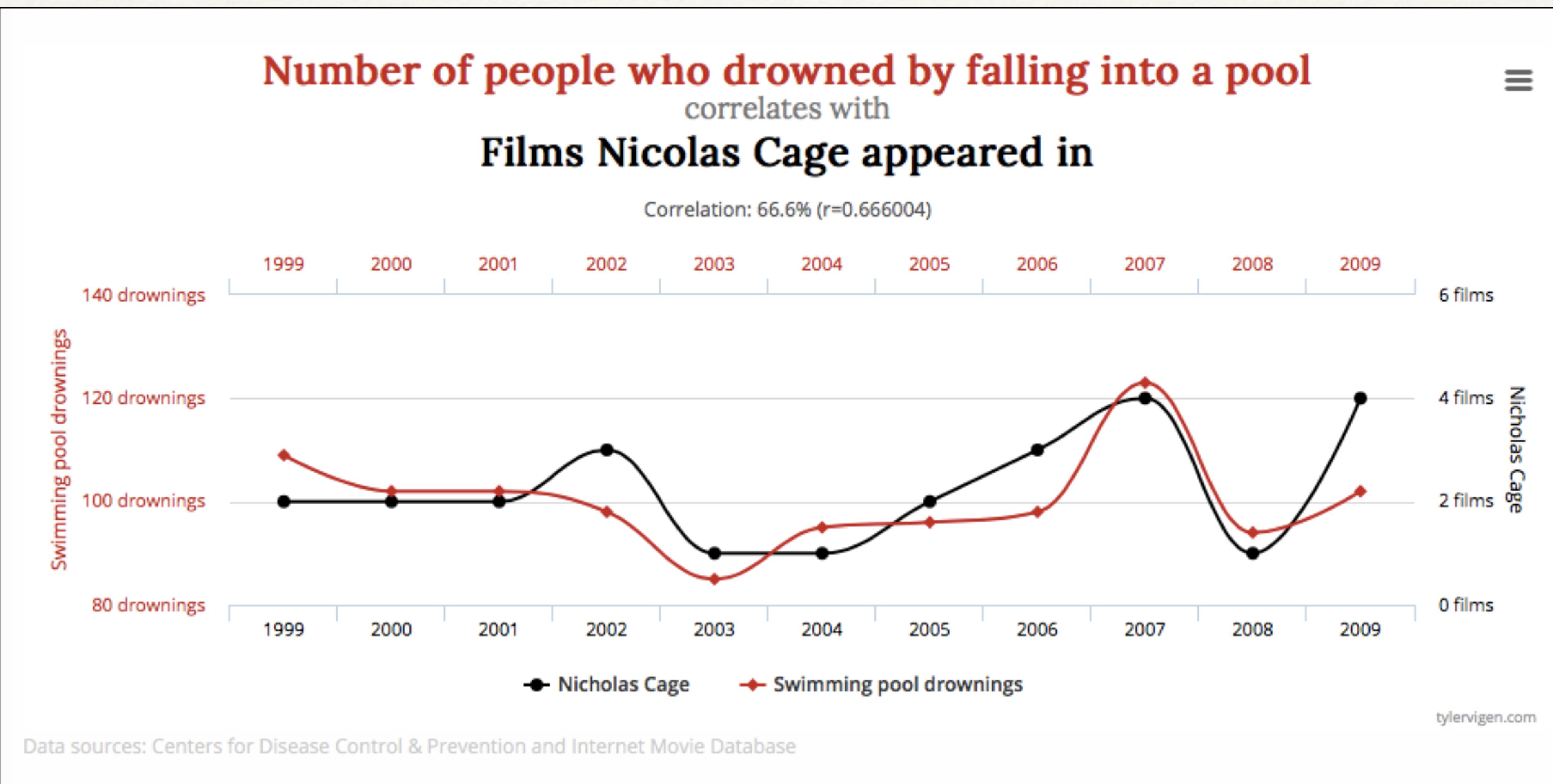
Correlazioni spurie

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



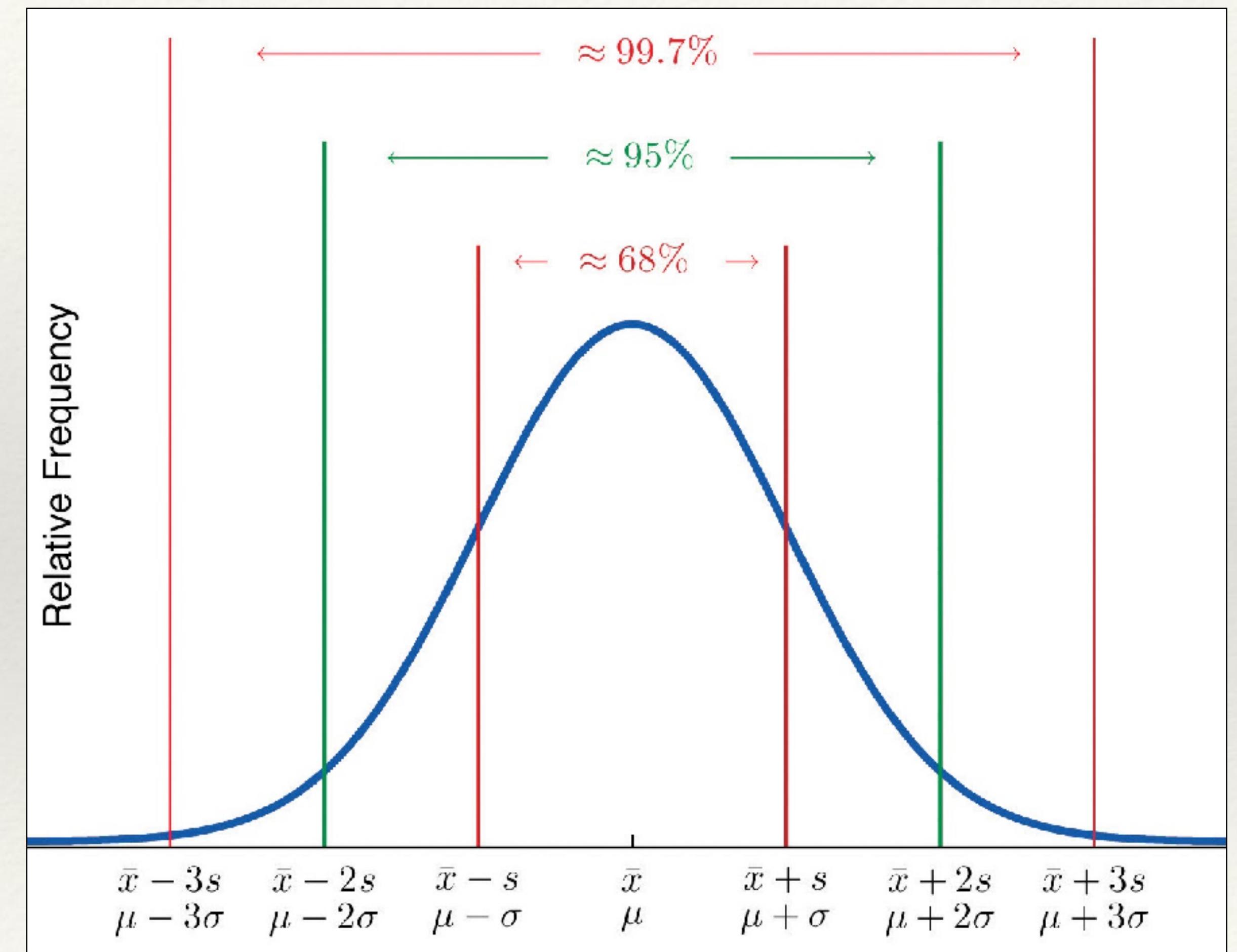
Correlazioni spurie

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

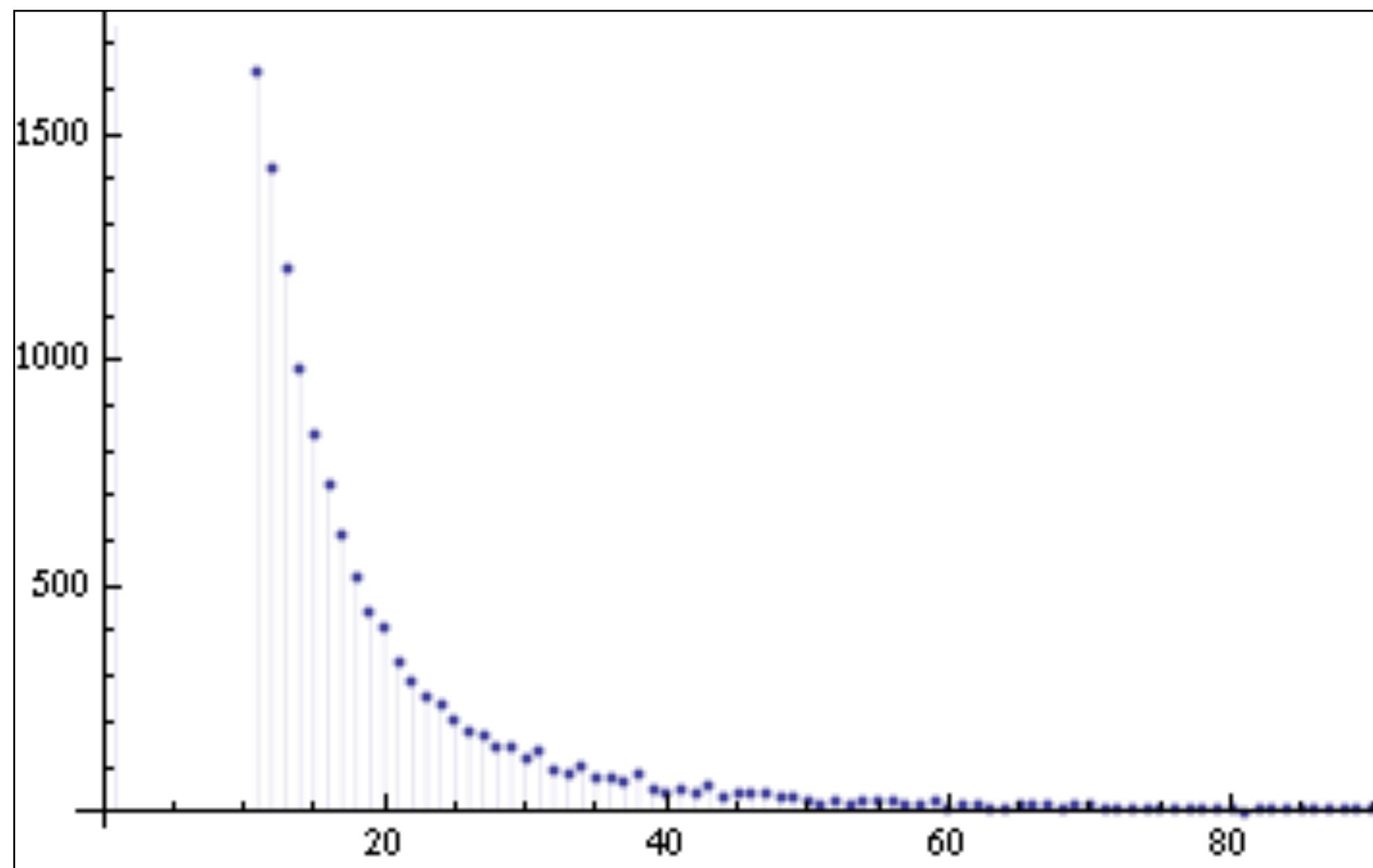


La media statistica

- ❖ Distribuzione delle altezze negli individui:
la media ci aiuta a fare predizioni
- ❖ Ma attenzione, anche rimanendo alla
statistica elementare, non c'è solo la media!



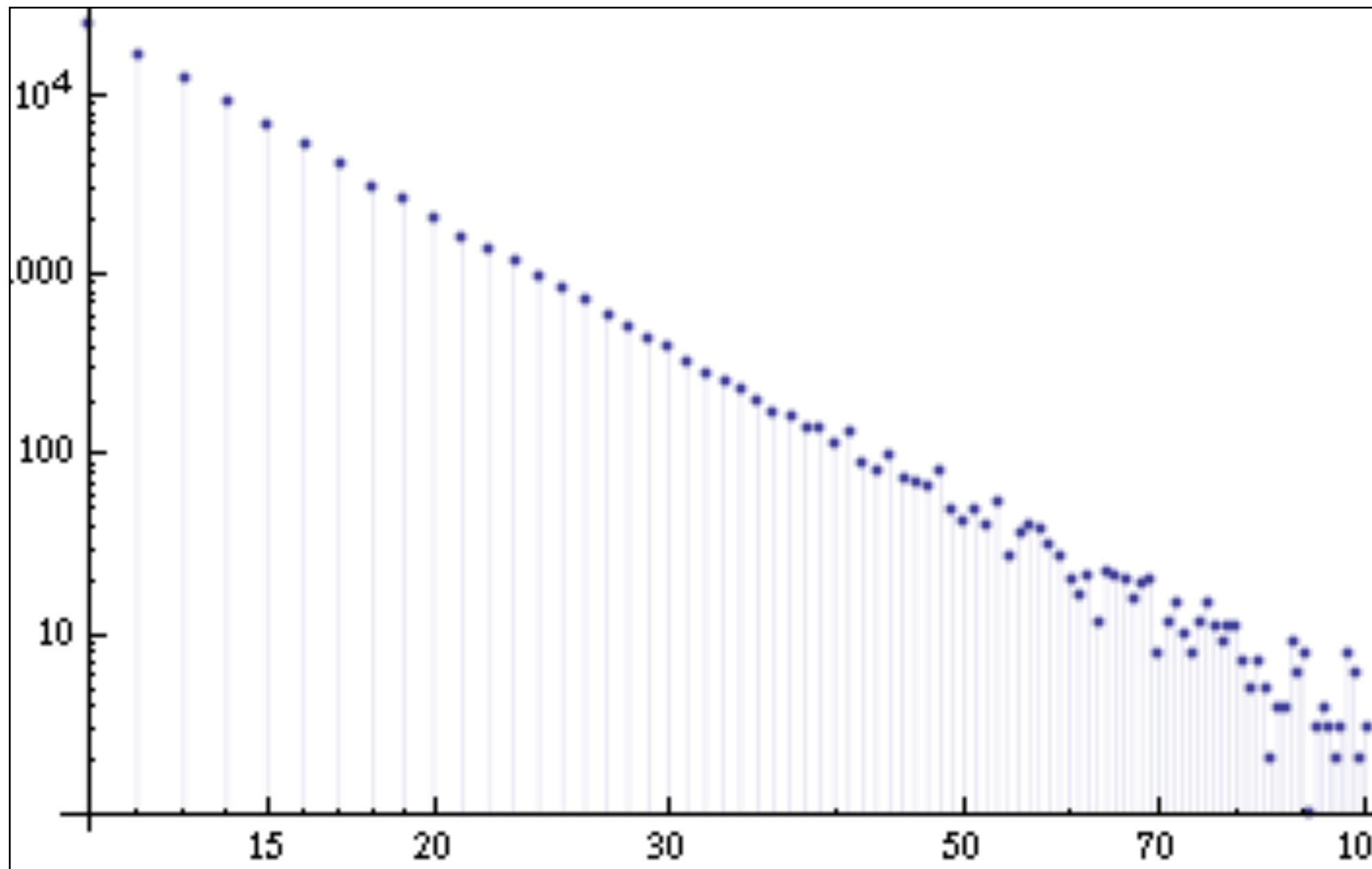
La media statistica



esempio: Distribuzione della ricchezza

La media statistica

La media in questo caso non ha alcun valore statistico!



In questo caso gli eventi estremi sono poco probabili, ma comunque MOLTO più probabili che con una curva a "campana"

esempio: Distribuzione della ricchezza

Esempio: climate change

- ❖ "Quest'anno anno abbiamo avuto le minime più basse degli ultimi anni: con questo freddo, come si fa a parlare di *riscaldamento globale*?"

 **Donald J. Trump** 
@realDonaldTrump

In the beautiful Midwest, windchill temperatures are reaching minus 60 degrees, the coldest ever recorded. In coming days, expected to get even colder. People can't last outside even for minutes. What the hell is going on with Global Waming? Please come back fast, we need you!

199.000 03:28 - 29 gen 2019 

177.000 utenti ne stanno parlando 

Capire un fenomeno complesso da pochi eventi locali è sbagliato

Abbiamo bisogno di molto più contesto e di tante tante analisi

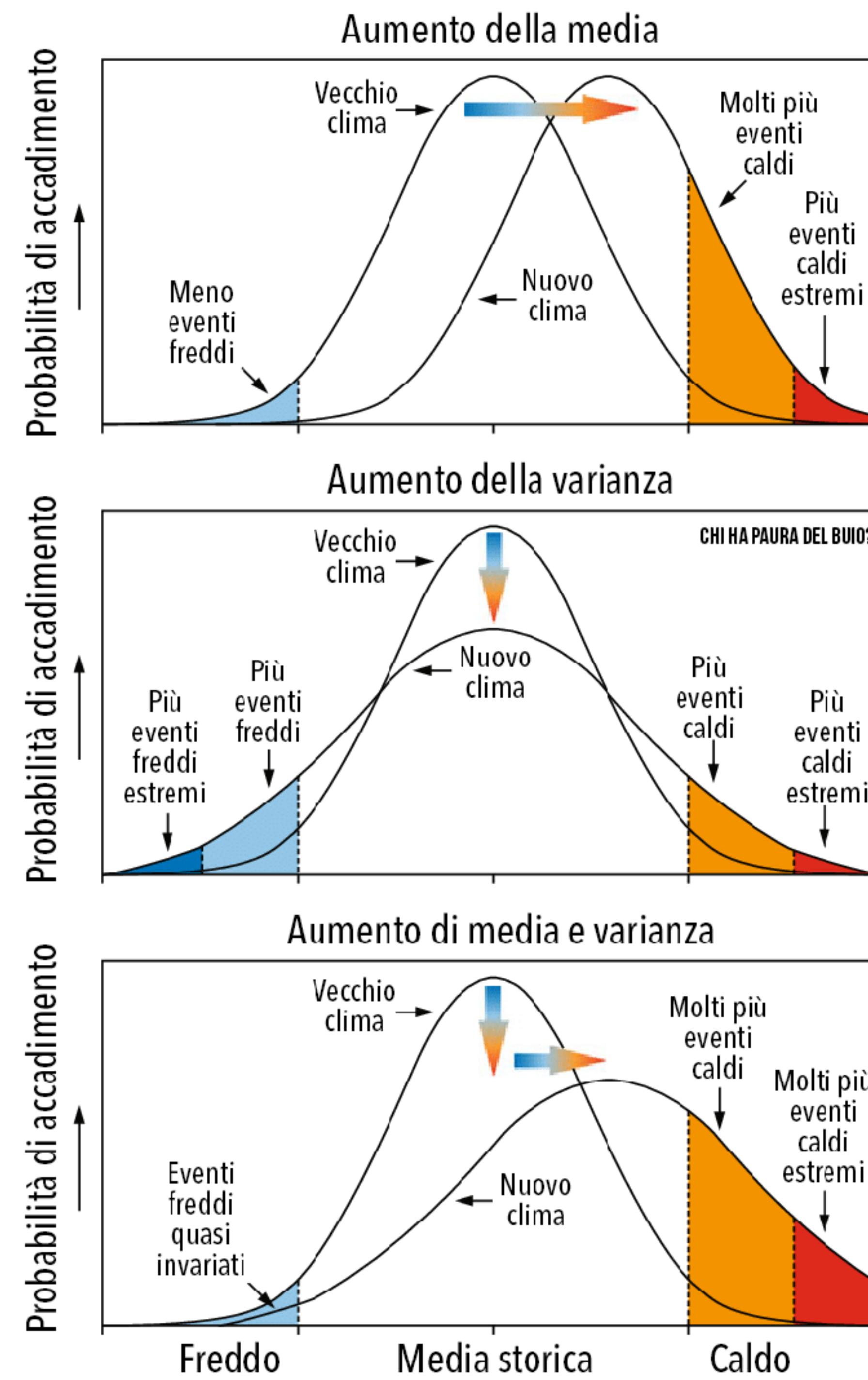
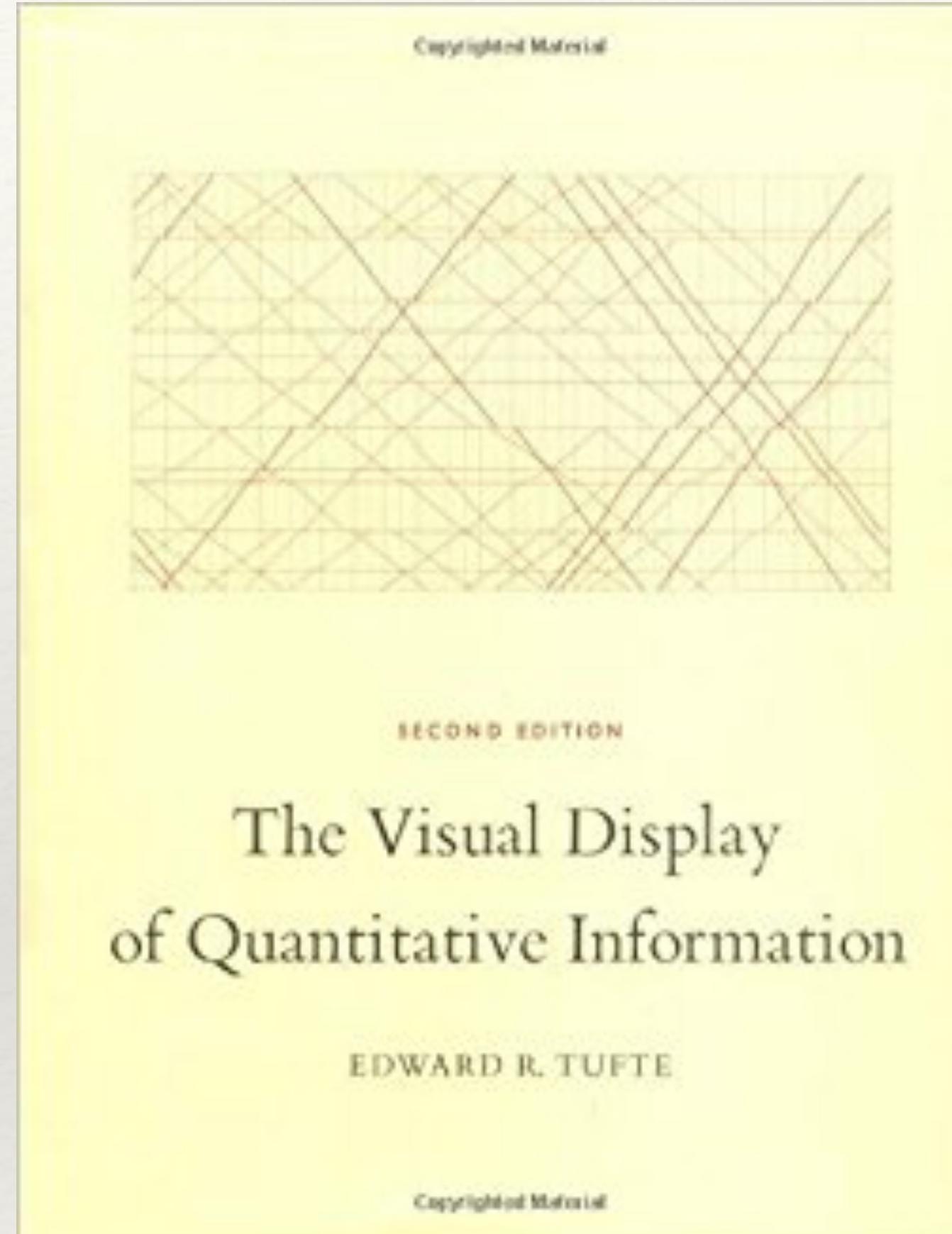


Immagine tratta da Olaf Jensen -
"Chi ha paura del buio?"

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”

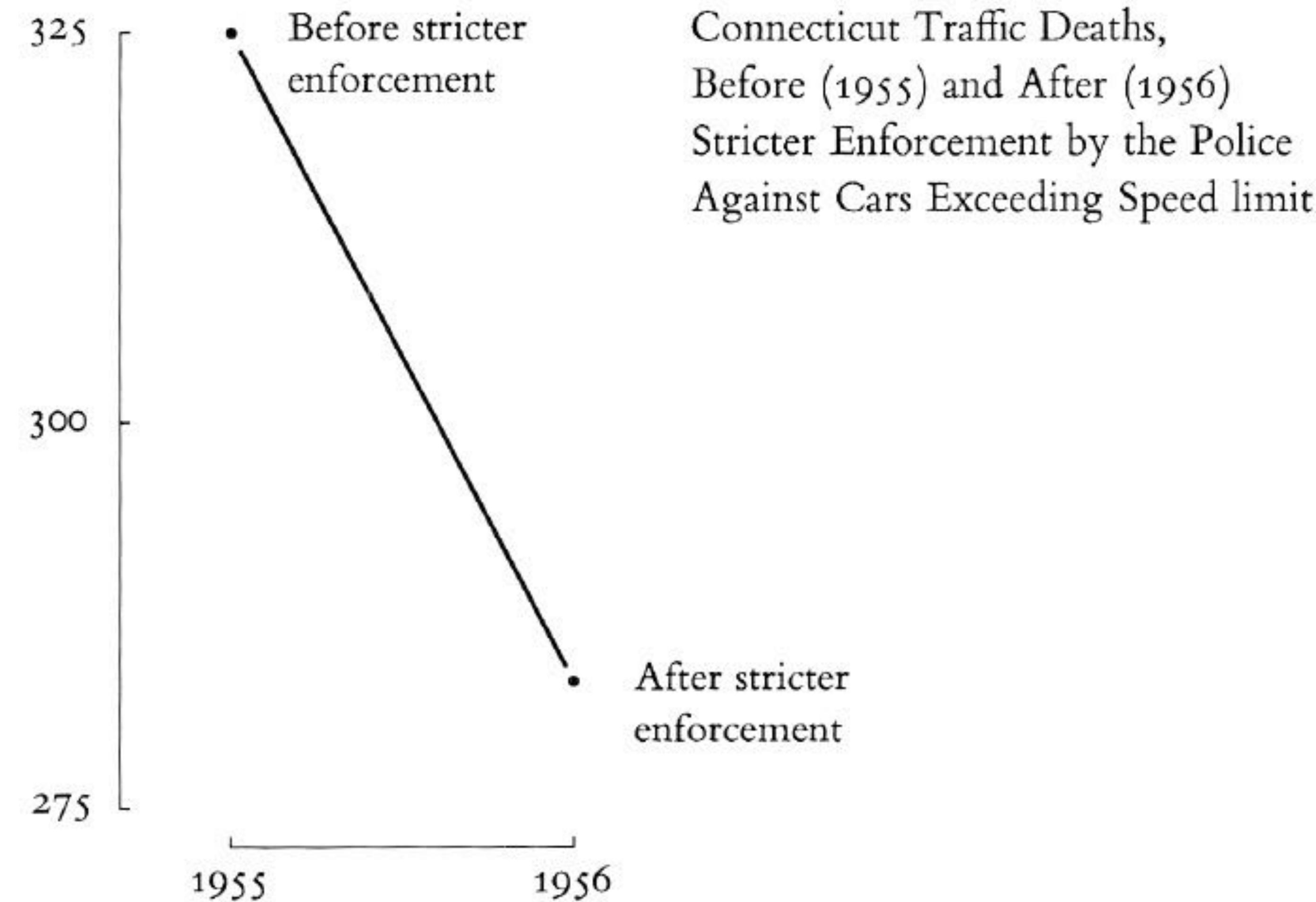
Misurare le bugie nei grafici



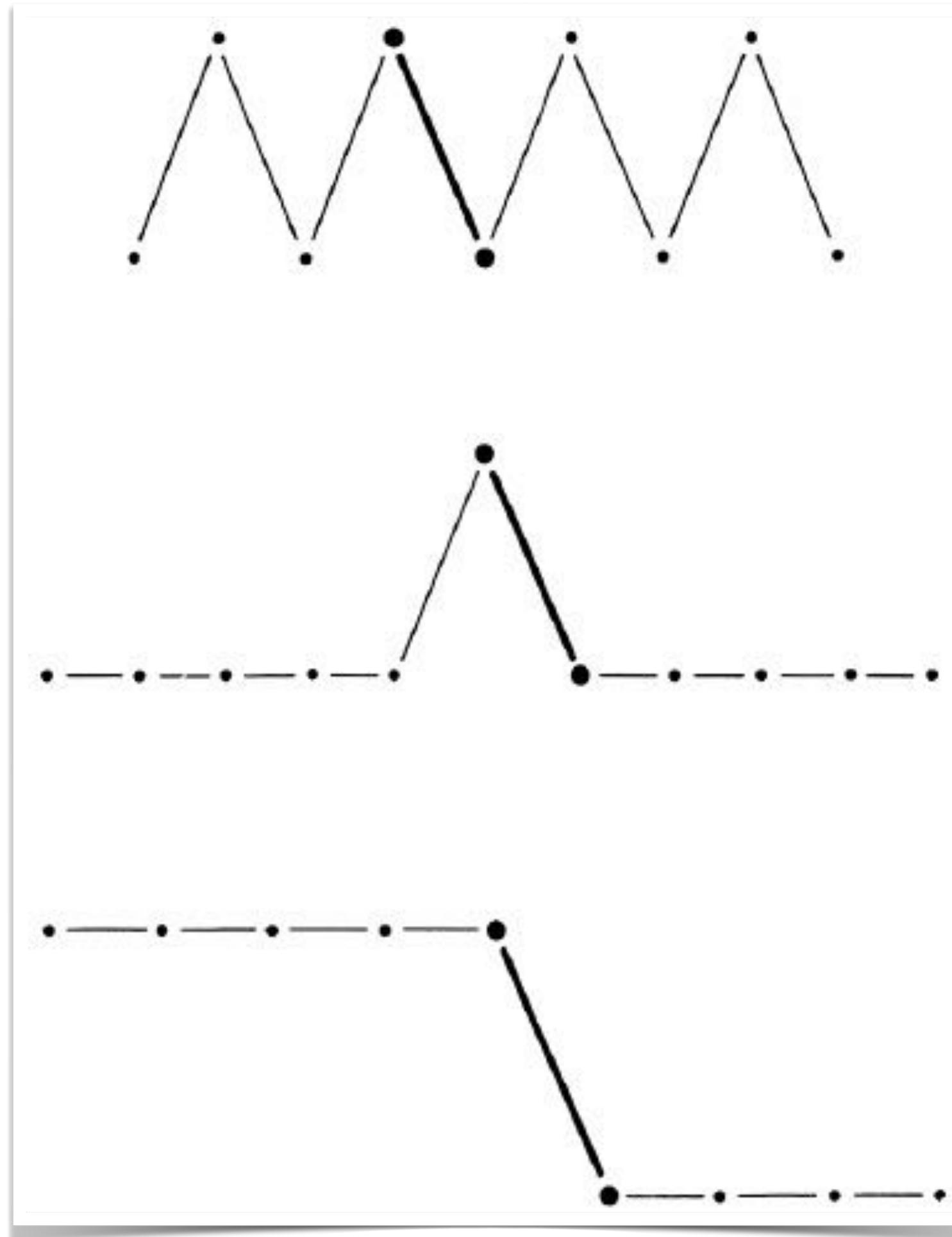
The Visual Display of Quantitative Information Edward R. Tufte

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

Integrità grafica

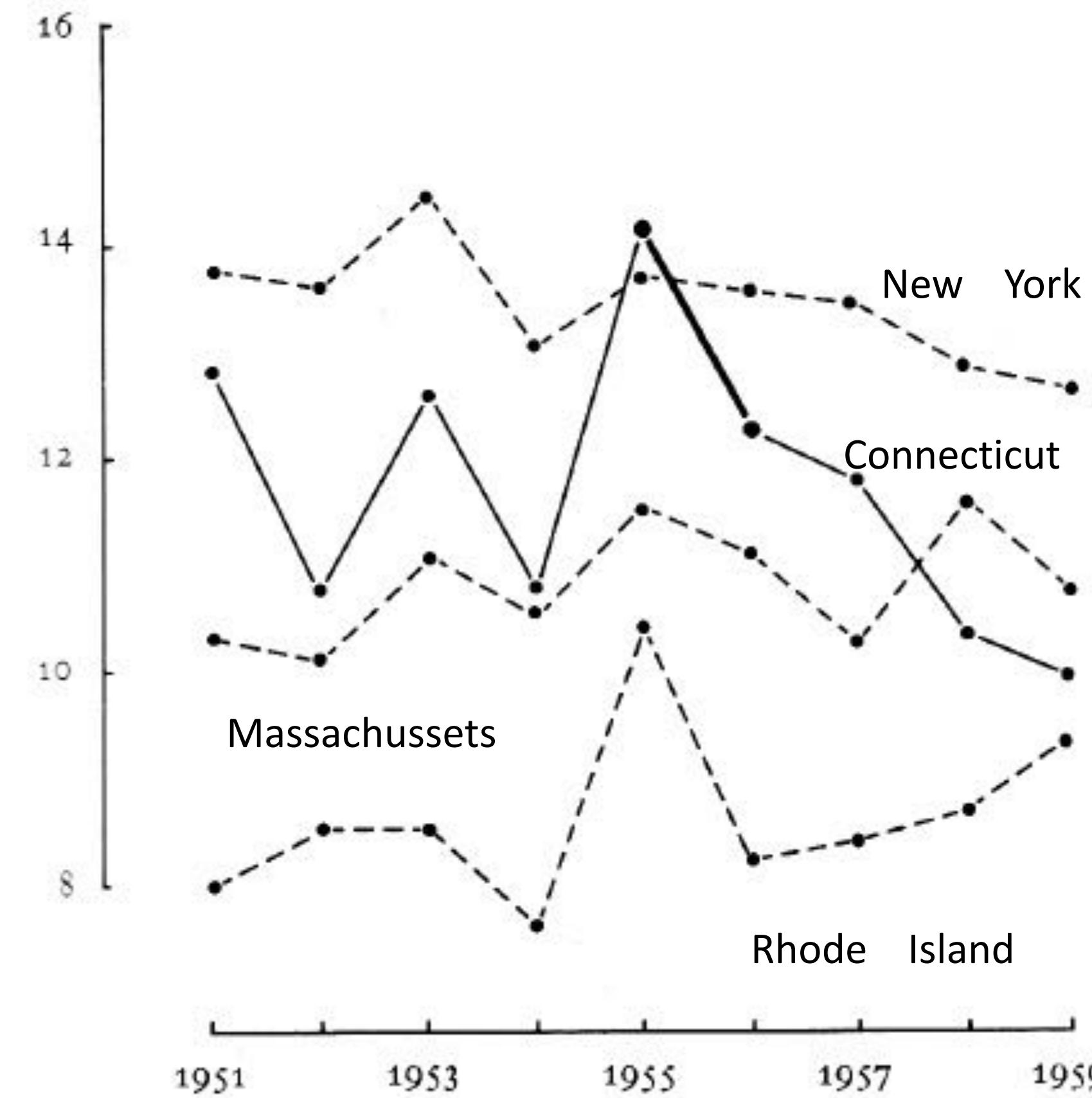
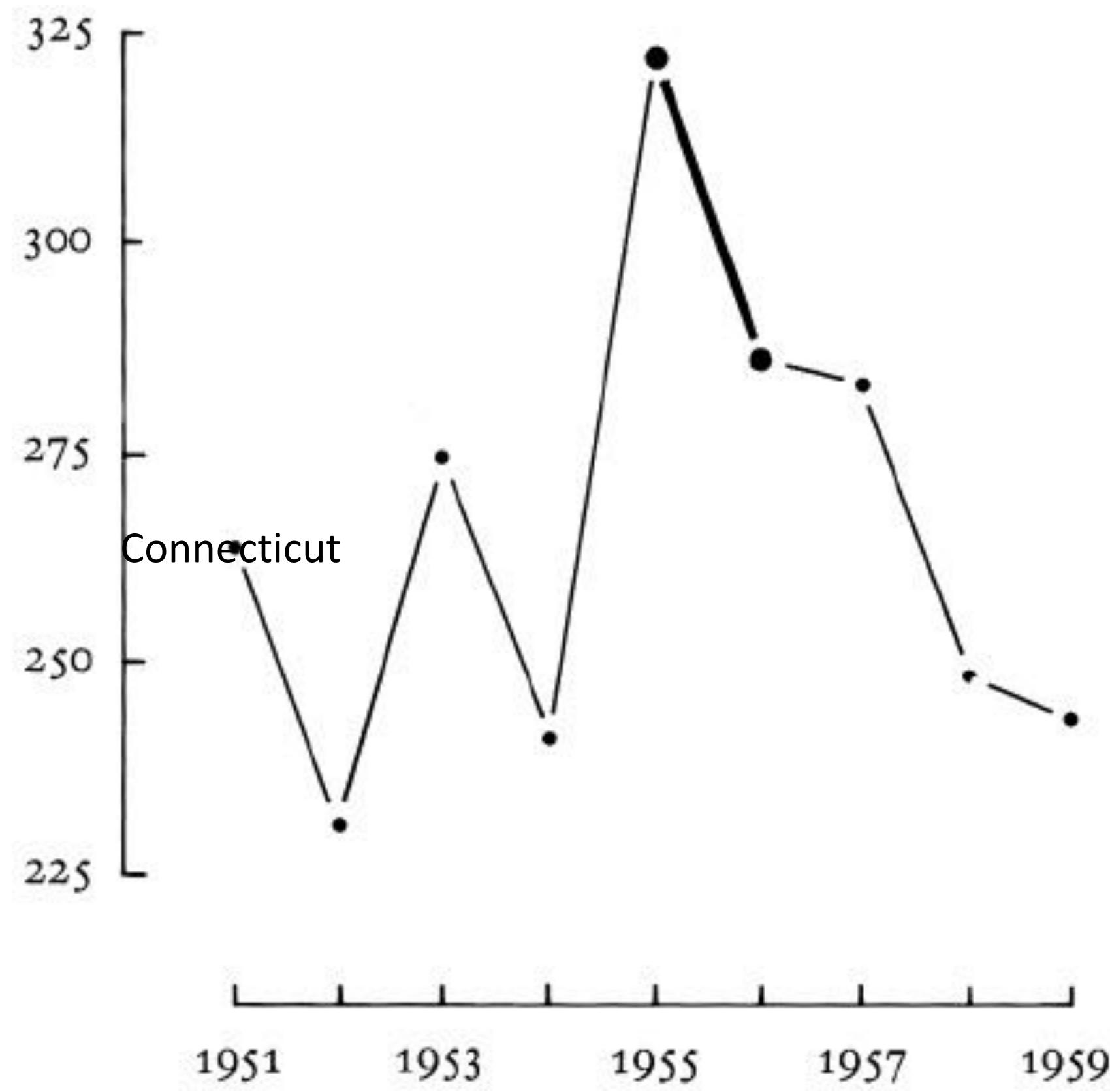


Integrità grafica



- ❖ Niente contesto
- ❖ Da quale pattern viene il grafico precedente?

Integrità grafica



Lezione #7

“Non introdurre bugie nascondendosi dietro l’idea di migliorare l’estetica della infografica o di semplificarne la lettura. Gli errori sono misurabili, prima o poi vengono a galla e la reputazione si perde in un baleno”

Basta così?

- ❖ Per nulla! Abbiamo parlato solo di una parte del processo dell'analisi dei dati, ovvero la loro visualizzazione
- ❖ Se siete degli scienziati e avete elaborato un **modello fisico** (deterministico o stocastico): non innamoratevene mai! il modello deve essere sempre **falsificabile** (Aristotele, Popper, etc.) e deve essere validato. È vero fino a quando non sarà falsificato e **prima o poi accadrà**.
- ❖ Se siete degli esperti di **machine learning** e avete addestrato il vostro modello: attenzione al bias nel data set iniziale, all'overfitting. Questi modelli sono una scatola nera: anche se le predizioni sono ottime, non capiamo davvero cosa succede dentro. Usare con cautela e competenza
- ❖ Attenzione ai **vostri pregiudizi**. Ne abbiamo già parlato, ma repetita iuvant...

Scoraggiati?

- ❖ "Questo secolo oramai alla fine
Saturo di parassiti senza dignità
Mi spinge solo ad essere migliore
Con più volontà" (F. Battiato, "E ti vengo a cercare", 1988)
- ❖ Mai scoraggiarsi! C'è tanto buon giornalismo e gli amici sui social possono fare ottime rassegne stampa, ma bisogna imparare l'arte dello **scetticismo**.
- ❖ Internet può essere il vostro nemico, ma anche il vostro migliore amico:
https://www.generazioniconnesse.it/site/_file/documenti/Comunicazione/Fake_news/Dispensa_Docenti_decalogo_bastabufale.pdf

Come diventare un detective anti-bufala di Paolo Attivissimo

1. Non condividere notizie che non hai verificato
2. Usa gli strumenti di Internet per verificare le notizie
3. Chiedi le fonti e le prove
4. Chiedi aiuto a una persona esperta o a un ente davvero competente
5. Ricorda che anche Internet e i social network sono manipolabili
6. Riconosci i vari tipi e gli stili delle notizie false
7. Hai un potere enorme: usalo bene
8. Dai il buon esempio: non lamentarti del buio, ma accendi una luce

Inoltre (mie aggiunte):

9. Se la notizia è riportata in un articolo scientifico, controlla se non sia stato ritratto o se non sia stato smentito da una ricerca successiva: nelle Scienze Dure non esistono Verità assolute ed eterne come in matematica!
10. Se la notizia è in linea con quella che credi sia la verità: applica le regole di sopra DUE volte

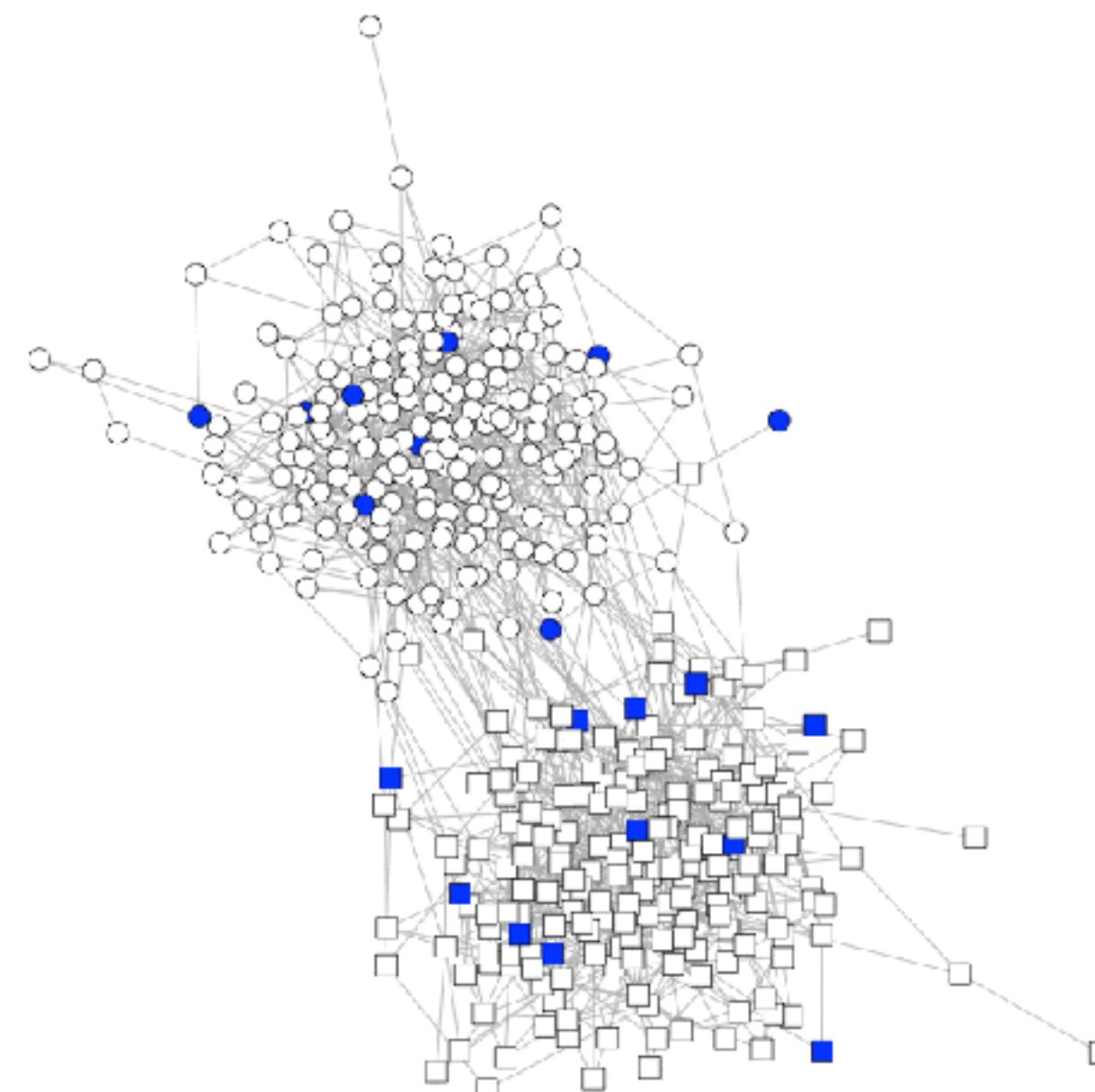
Lavori recenti

<http://arcs.di.unito.it>

“dimenticare” (bufale vs teorie del complotto)

LOW Forgetting Rate

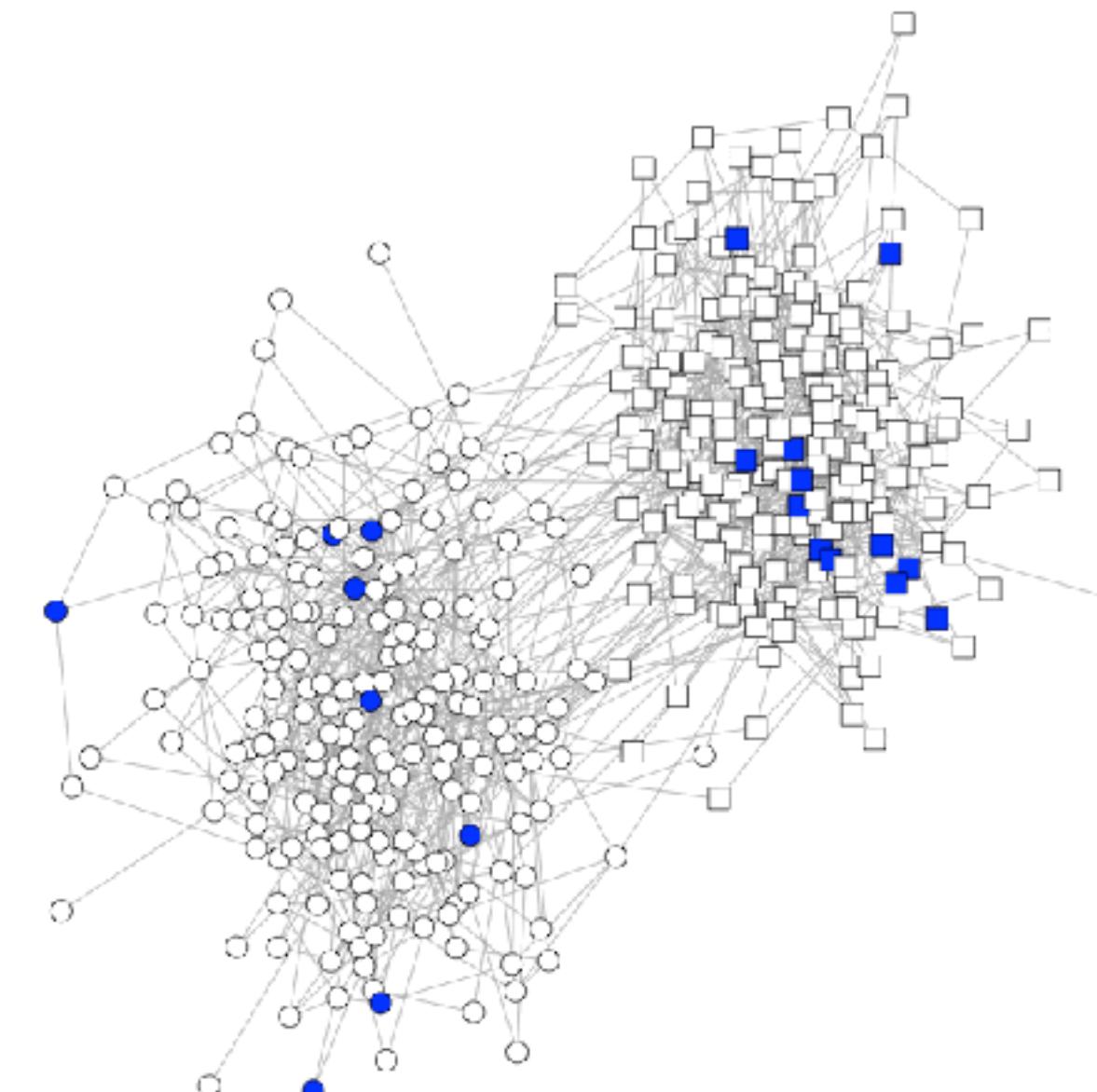
$$p_f = 0.1$$



Time = 1

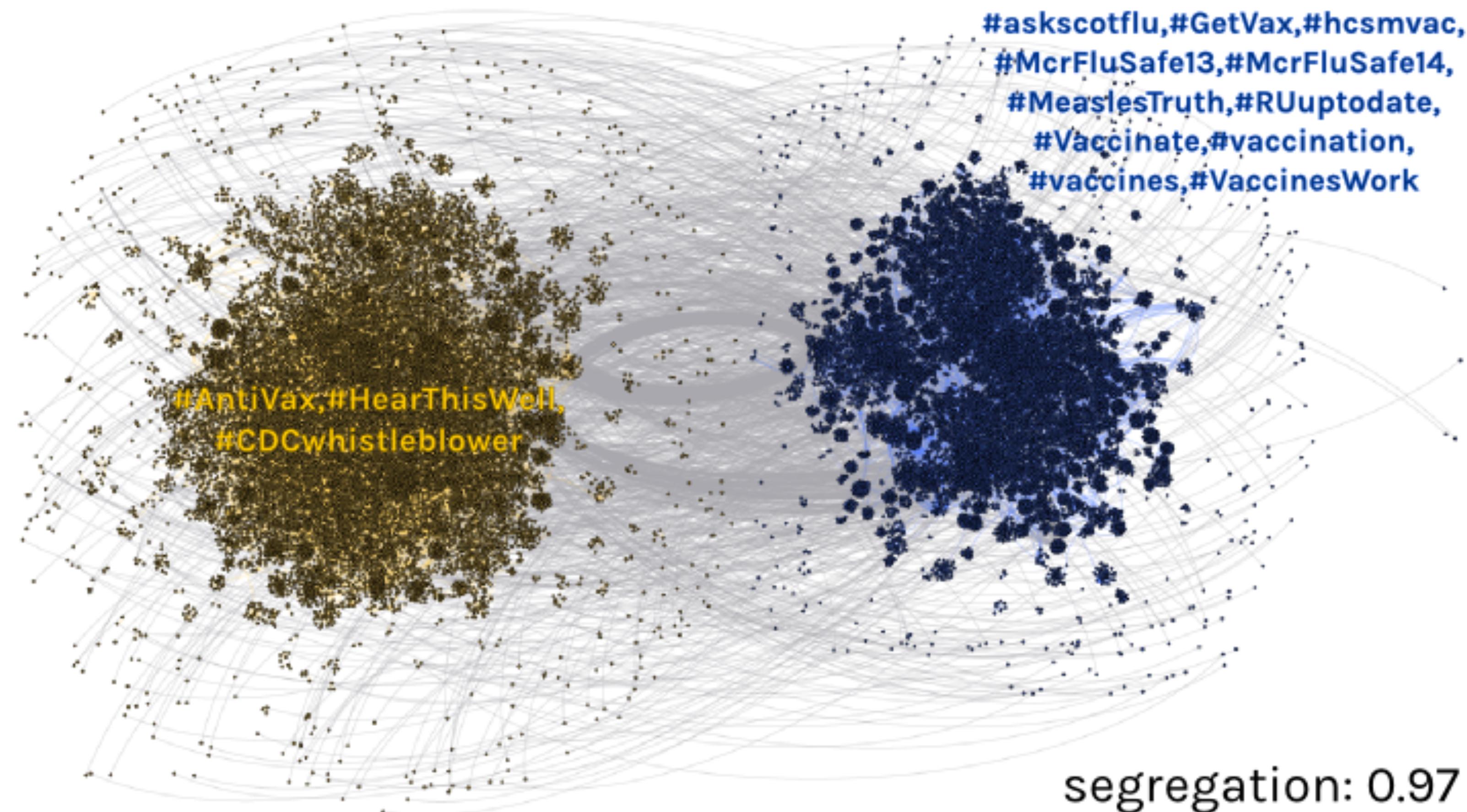
HIGH Forgetting Rate

$$p_f = 0.8$$



Time = 1

dati twitter: dibattito su vaccini

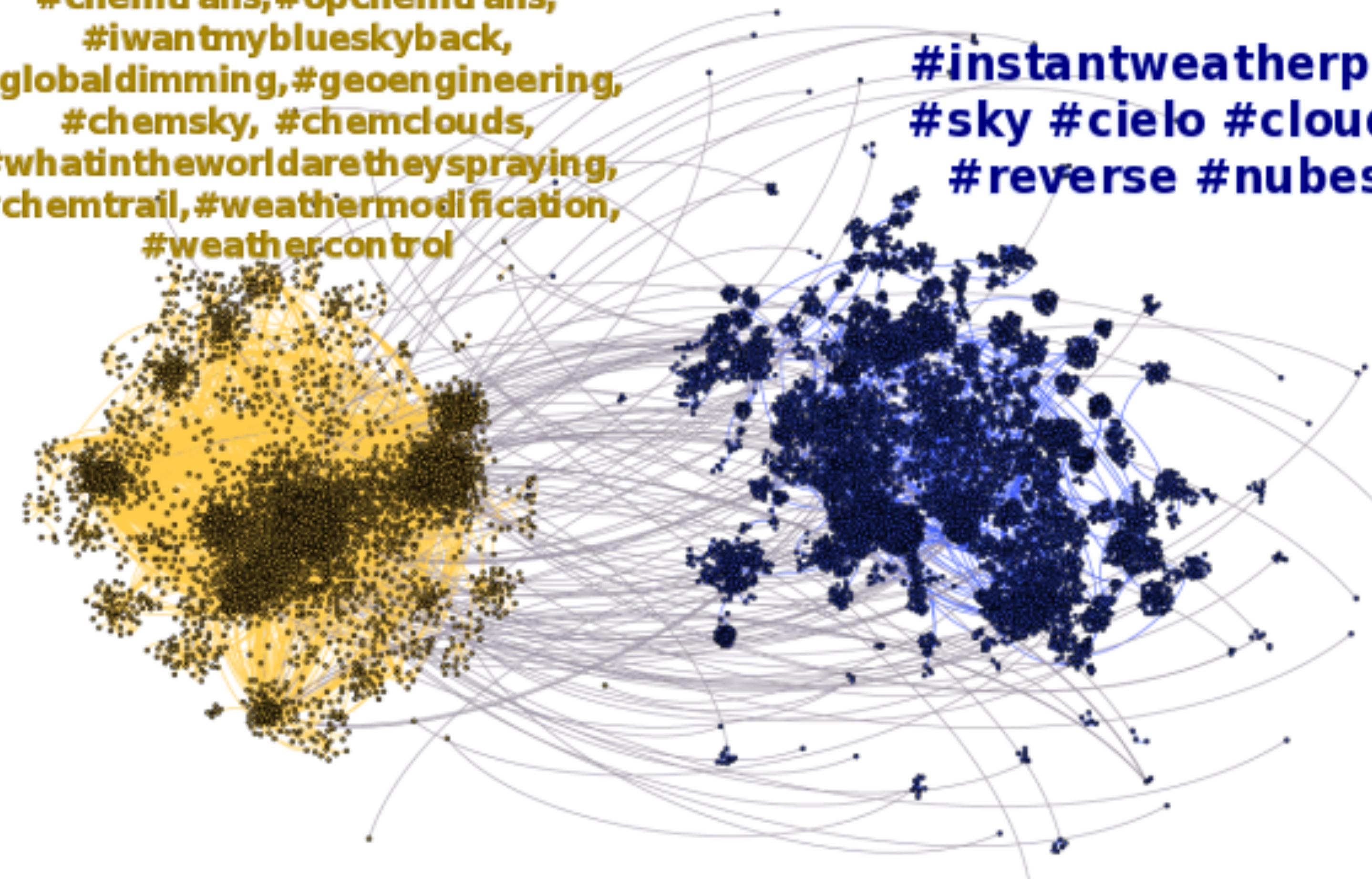


twitter data from IU <https://osome.iuni.iu.edu>

dati twitter: dibattito su scie chimiche

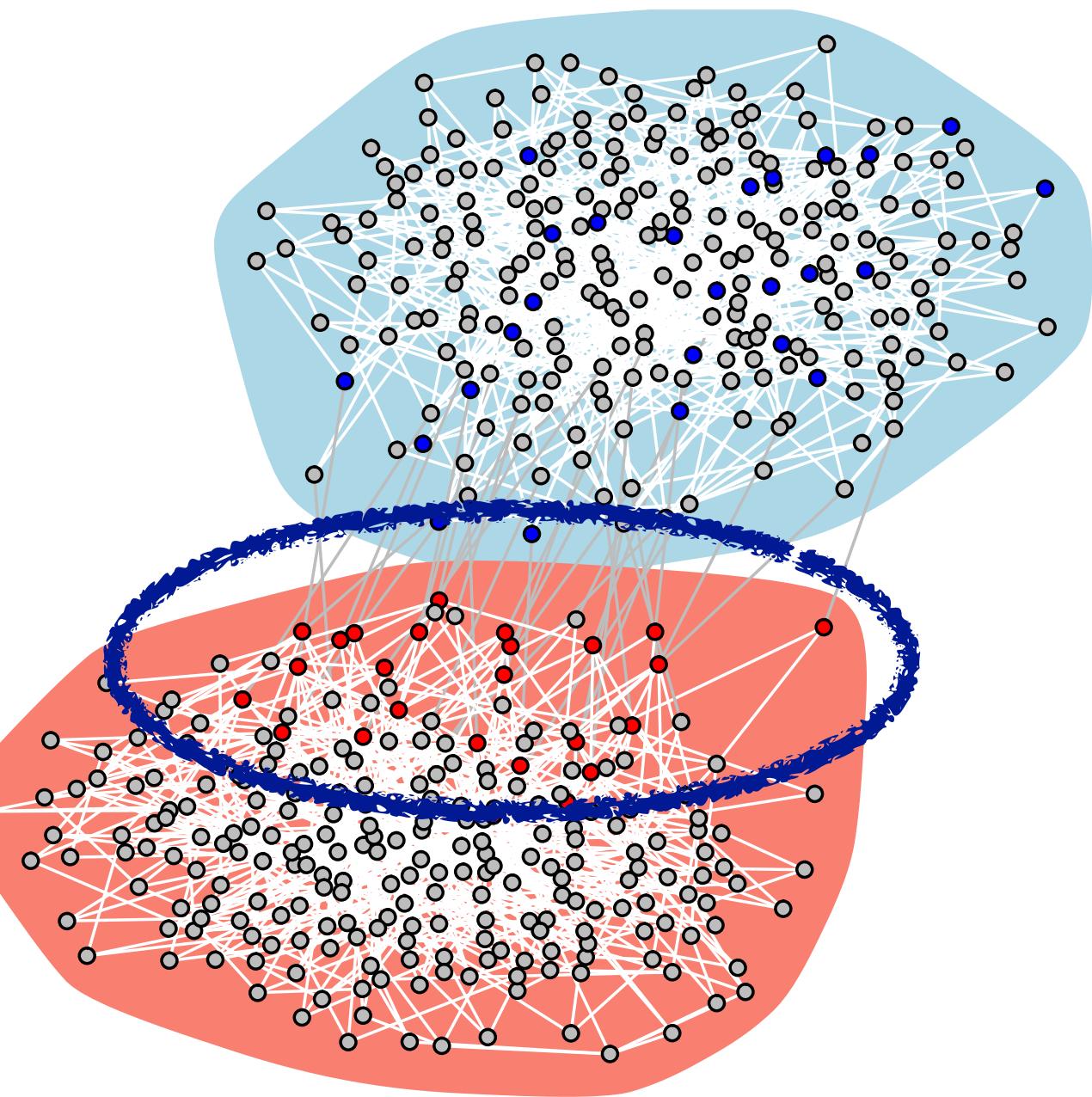
#chemtrails, #opchemtrails,
#iwantmyblueskyback,
#globaldimming, #geoengineering,
#chemsky, #chemclouds,
#whatintheworldaretheyspraying,
#chemtrail, #weathermodification,
#weathercontrol

#instantweatherpro
#sky #cielo #clouds
#reverse #nubes

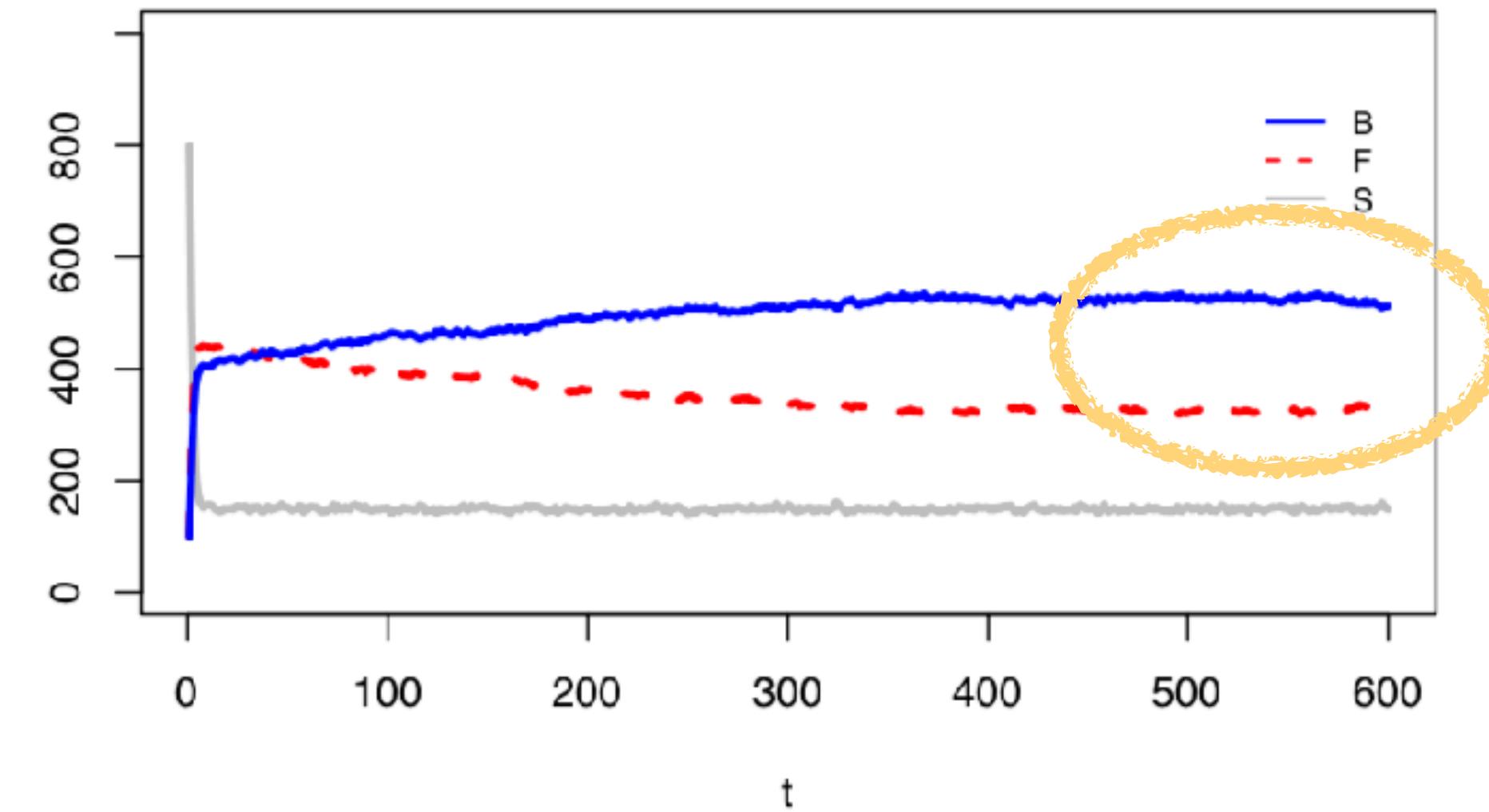


twitter data from IU <https://osome.iuni.iu.edu>

La posizione migliore per il fact checker



risultati della simulazione



SPECIAL REPORT

Finland is winning the war on fake news. What it's learned may be crucial to Western democracy

By Eliza Mackintosh, CNN

Video by Edward Kiernan, CNN

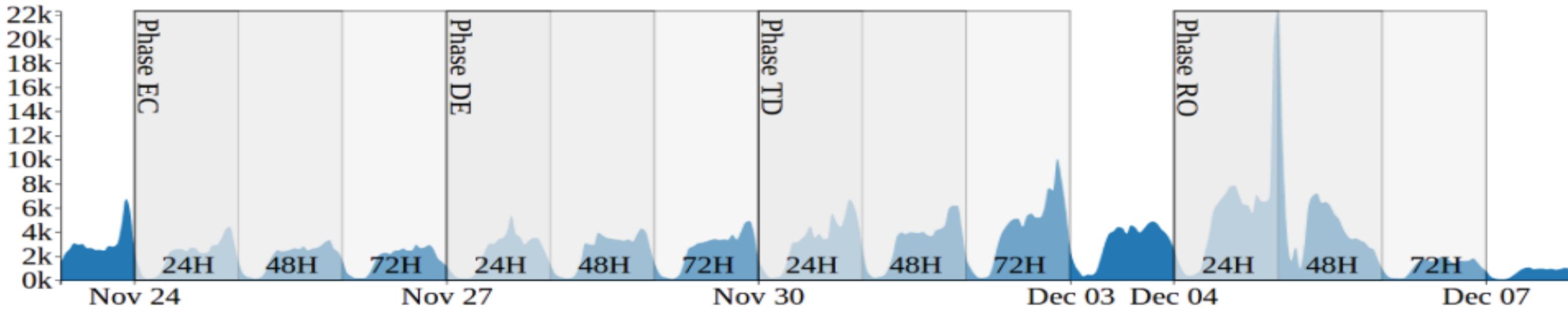


Helsinki, Finland (CNN) – On a recent afternoon in Helsinki, a group of students gathered to hear a lecture on a subject that is far from a staple in most community college curriculums.

Standing in front of the classroom at Espoo Adult Education Centre, Jussi Toivanen worked his way through his PowerPoint presentation. A slide titled “Have you been hit by the Russian troll army?” included a checklist of methods used to deceive readers on social media: image and video manipulations, half-truths, intimidation and false profiles.

Referendum Italiano 2016

Tweet collezionati

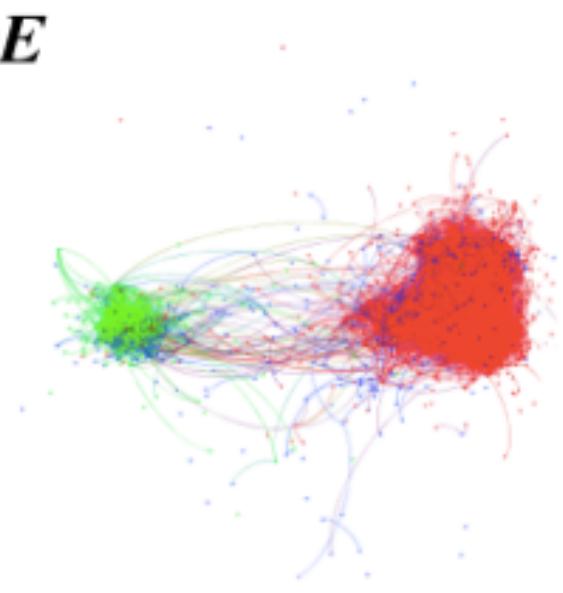


EC



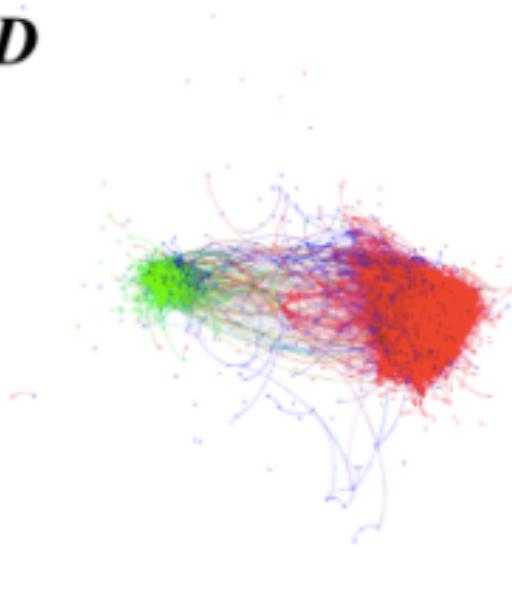
opinione rilevata **contraria**

DE



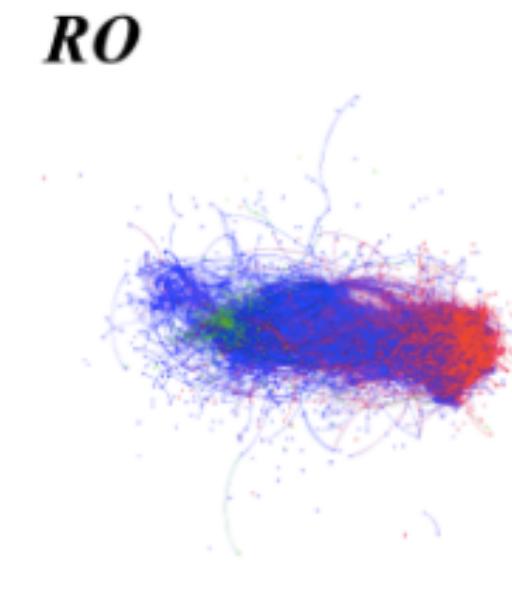
opinione rilevata **a favore**

TD



opinione rilevata **nulla**

RO

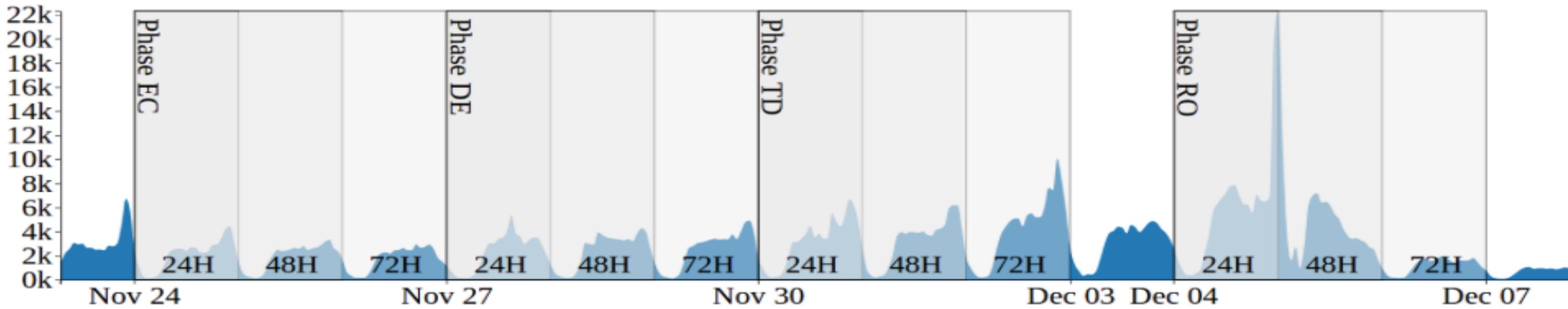


Retweet Network

forte segnale di
omofilia

Referendum Italiano 2016

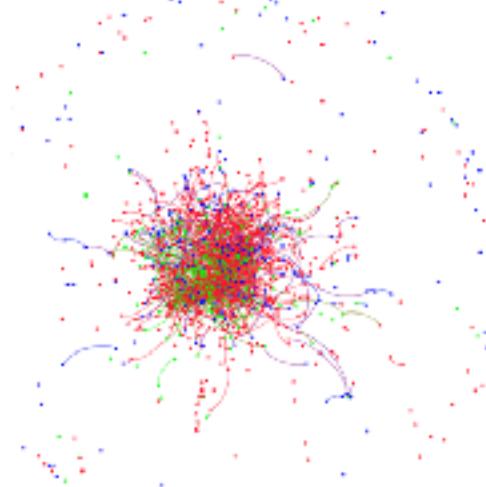
Tweet collezionati



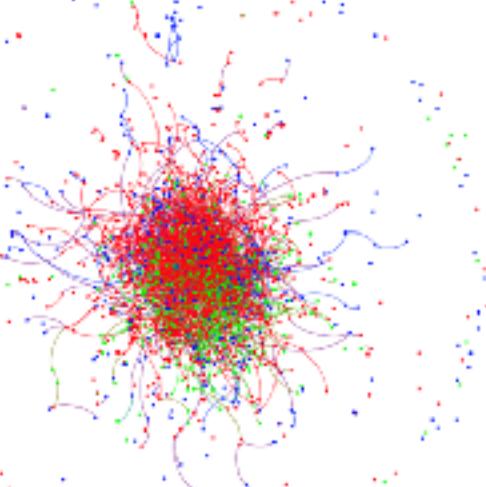
EC



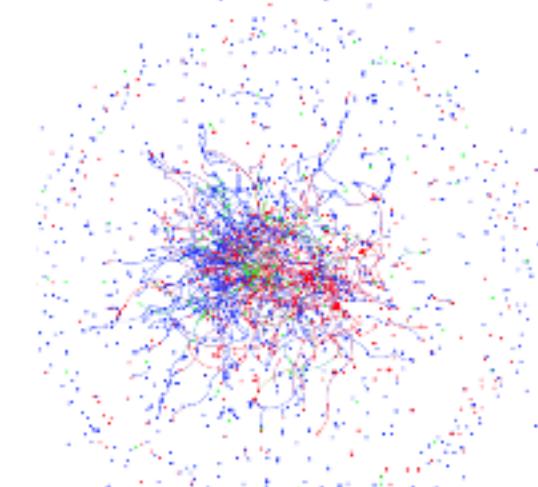
DE



TD

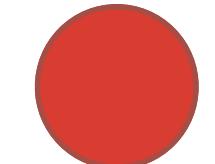


RO

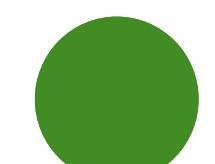


Reply-to Network

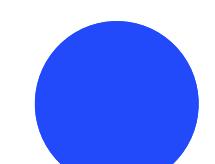
segna di omofilia
inversa o eterofilia



opinione rilevata **contraria**

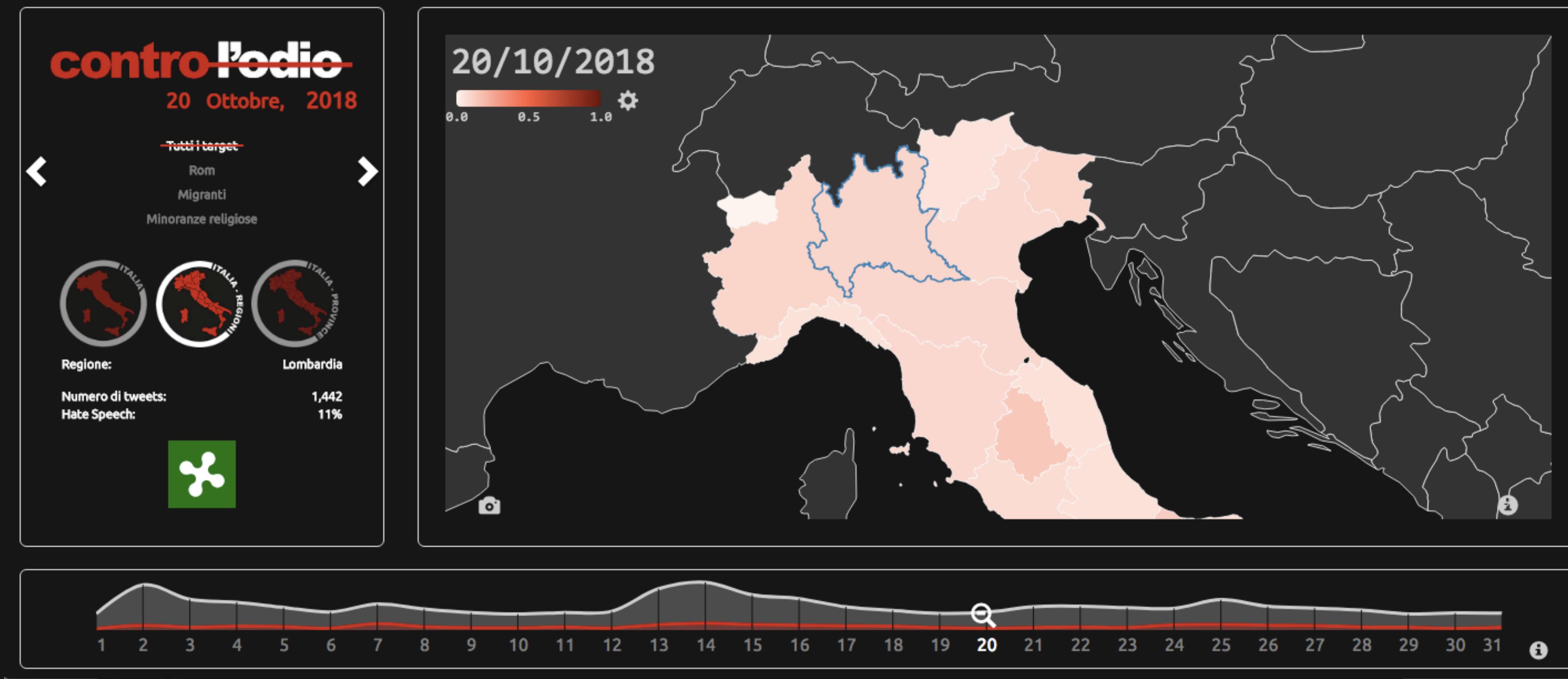


opinione rilevata **a favore**

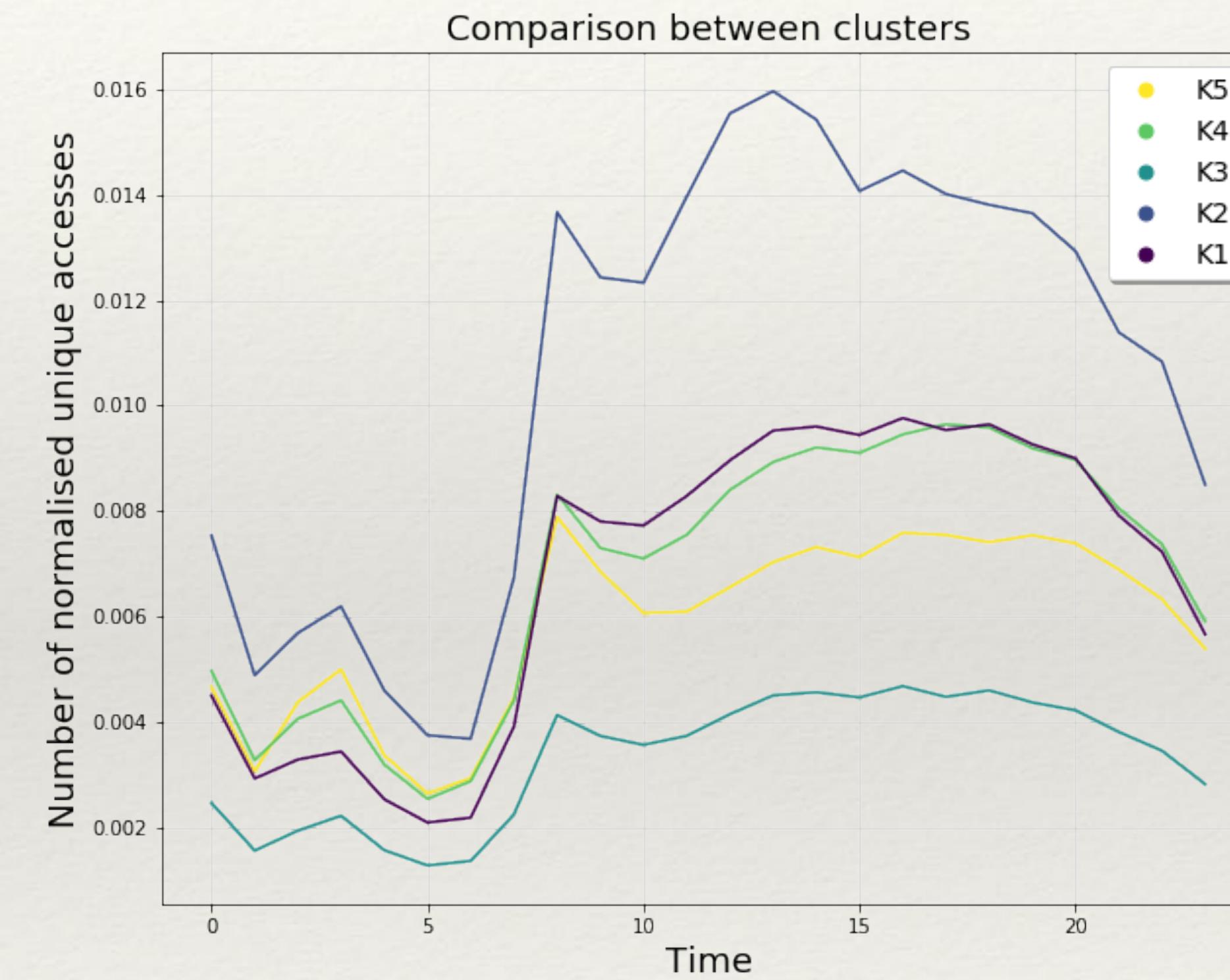
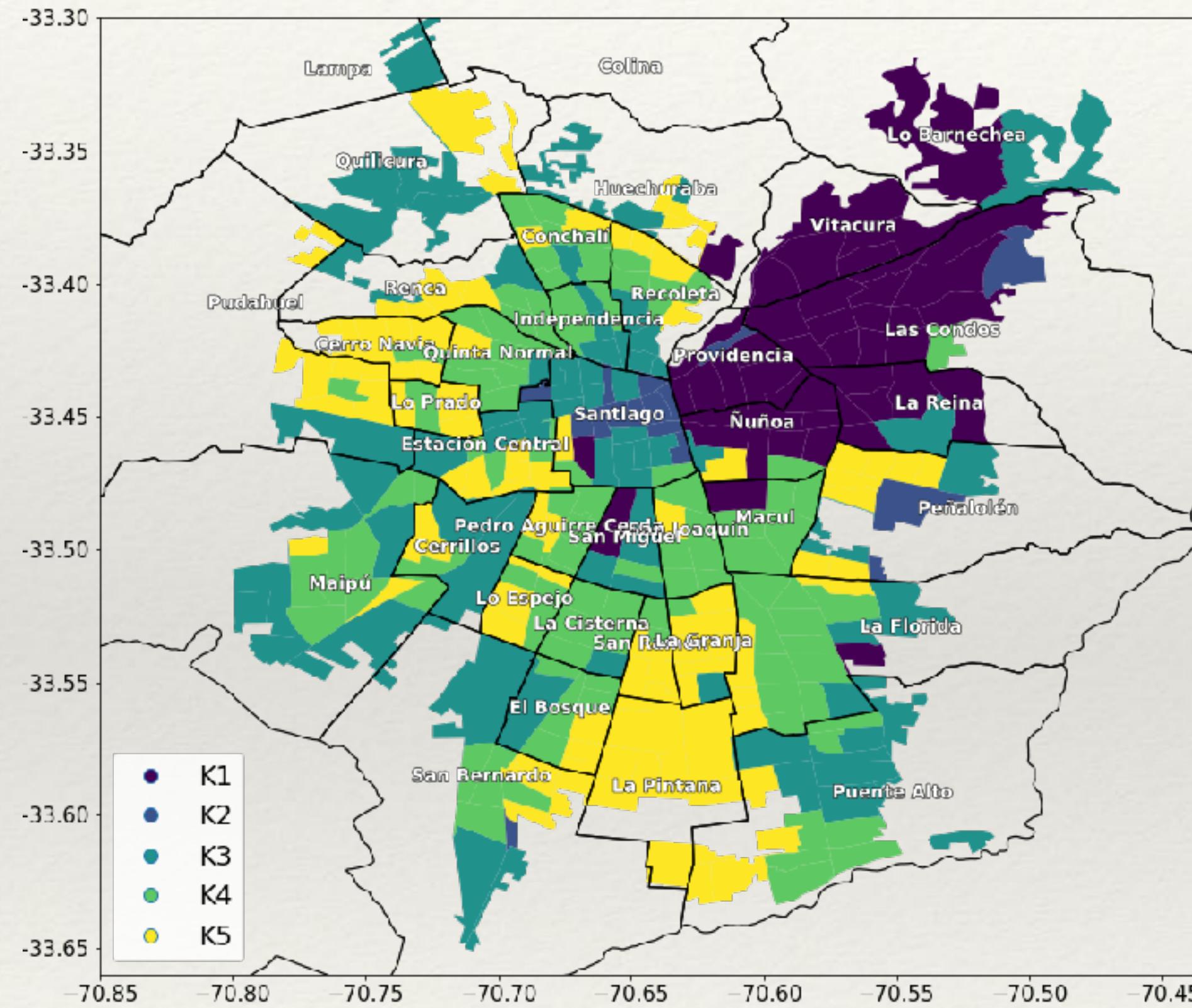


opinione rilevata **nulla**

Hate speech monitoring (Contro l'Odio)



Accesso alle notizie (Santiago del Cile)



Riferimenti

- LM Aiello, A Barrat, C Cattuto, G Ruffo, R Schifanella, 2010. [Link creation and profile alignment in the aNobii social network](#), 2010 IEEE 2nd Int.. Conf. on Social Computing, 249-256
- LM Aiello, A Barrat, C Cattuto, G Ruffo, R Schifanella, [Link creation and information spreading over social and communication ties in interest based online social network](#), EPJ Data Science 1 (1), 12
- LM Aiello, M. Deplano, R Schifanella, G Ruffo. 2012, [People are Strange when you're a Stranger: Impact and Influence of Bots on Social Networks](#), in Proc. of the 6th Intern. AAAI Conf. on Weblogs and Social Media (ICWSM'12), Dublin, Ireland
- M Tambuscio, G Ruffo, A Flammini, and F Menczer. 2015. [Fact-checking Effect on Viral Hoaxes: A Model of Misinformation Spread in Social Networks](#). In Proc. of the 24th Int. Conf. on World Wide Web (WWW '15 Companion)
- M Tambuscio, D F M Oliveira, G L Ciampaglia, G Ruffo, 2019. [Network segregation in a model of misinformation and fact-checking](#), Journal of Computational Social Science (2018) 1: 261.
- M Tambuscio, G. Ruffo. 2019, [Fact-checking strategies to limit urban legends spreading in a segregated society](#), Applied Network Science Journal, Springer
- M Lai, M Tambuscio, V Patti, P Rosso, G. Ruffo, [Stance Polarity in Political Debates: a Diachronic Perspective of Network Homophily and Conversations on Twitter](#), Data & Knowledge Engineering Journal, online: September 2019
- A T E Capozzi, V Patti, G Ruffo, and C Bosco. 2018. [A Data Viz Platform as a Support to Study, Analyze and Understand the Hate Speech Phenomenon](#). In Proceedings of the 2nd International Conference on Web Studies (WS.2 2018), ACM
- A T. E. Capozzi, M Lai, V Basile, C Musto, M Polignano, F Poletto, M Sanguinetti, C Bosco, V Patti, G Ruffo, G Semeraro and M Stranisci. [Computational Linguistics Against Hate: Hate Speech Detection and Visualization on Social Media in the "Contro L'Odio" Project](#), in Proc. of CLIC 2019
- S Vilella, D Paolotti, G Ruffo, L Ferres, [News and the city: understanding online press consumption patterns through mobile data](#), EPJ Data Sci. 9, 10 (2020). <https://doi.org/10.1140/epjds/s13688-020-00228-9>



di.unito.it



GIANCARLO
RUFFO



MARCELLA
TAMBUSCIO



MIRKO
LAI



ROSSANO
SCHIFANELLA



ANDRÉ
PANISSON



LUCA
AIELLO



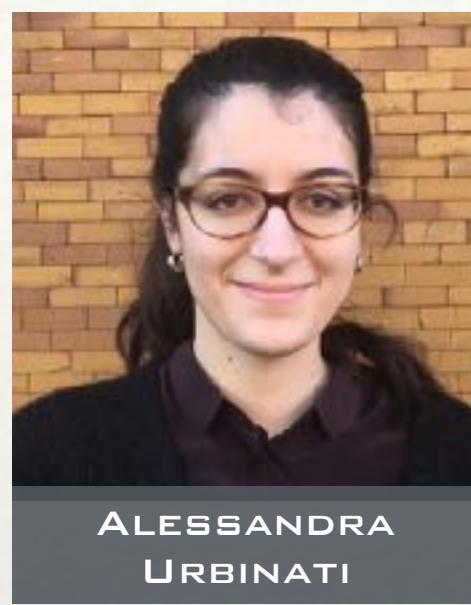
VIVIANA
PATTI



ARTHUR
CAPOZZI



ALFONSO
SEMERARO



ALESSANDRA
URBINATI



SALVATORE
VILELLA



EDOARDO
GALIMBERTI



EMILIO
SULIS



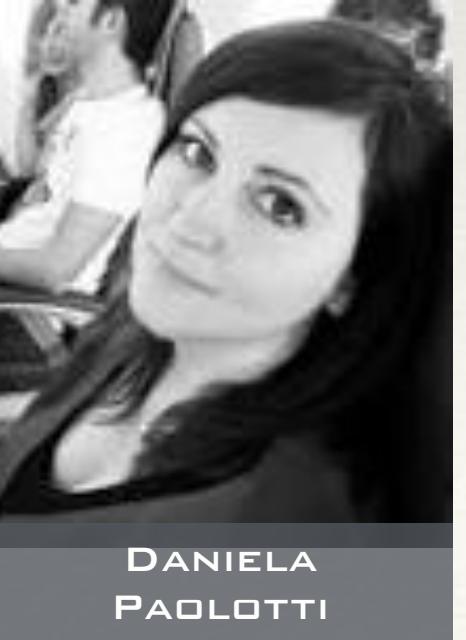
MARTINA
DEPLANO



CRISTINA
BOSCO

ARC²S: Applied Research on Computational Complex Systems

Thanks!



DANIELA
PAOLOTTI



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



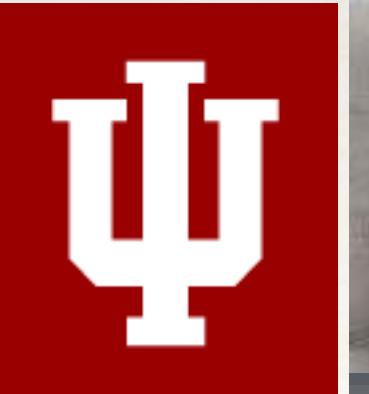
PAOLO
ROSSO



Universidad del Desarrollo



LEO
FERRES



IU
Institut Universitaire de France



CHENGCHENG
SHAO



GIOVANNI LUIGI
CIAMPAGLIA



ALESSANDRO
FLAMMINI



FIL
MENCZER