

# **Mapping e Big Data**

## **Lezione 6**

Mario Verdicchio

Università degli Studi di Bergamo

Anno Accademico 2022-2023

# Modello

- Un **modello** è un insieme di segni e di relazioni tra loro che descrive, con grado di accuratezza variabile, un aspetto della realtà.
- Su tale descrizione, gli osservatori posso costruire spiegazioni o previsioni sull'aspetto della realtà descritto.
- Tipicamente, la descrizione di un modello è una semplificazione della realtà che omette elementi inutili per le successive spiegazioni e previsioni e che invece mette in risalto gli elementi considerati importanti per tali processi

# Le visualizzazioni sono modelli

- Tutte le visualizzazioni sono modelli
- La mappe, ad esempio, sono modelli
- Come tali, omettono dettagli che non servono allo scopo per cui sono state create
- Una mappa su un navigatore, ad esempio, non dice nulla sullo stato dei cespugli sul ciglio delle strade



In questa zona della mappa non c'è nulla, ma questo non vuol dire che nella realtà ci sia un vuoto. In quella parte della città non ci sono elementi utili per la navigazione in automobile, secondo l'opinione di chi ha realizzato questa visualizzazione.

# L'omissione come vantaggio

- L'omissione di elementi permette agli osservatori della visualizzazione di focalizzarsi su ciò che è importante
- Un modello è tanto migliore quanto più si approssima a ciò che rappresenta senza essere inutilmente complesso e quanto più facile è per i suoi osservatori interpretarlo correttamente

# L'omissione come svantaggio

- Attenzione: sono i disegnatori della visualizzazione a decidere quali elementi includere e quali omettere
- A causa di incompetenza o, peggio, di intenzioni malevoli, i disegnatori possono creare visualizzazioni fuorvianti, che danno agli osservatori una rappresentazione della realtà che non è approssimata, bensì falsa

# La solita NCTA

- La National Cable & Telecommunications Association fornisce un altro esempio negativo

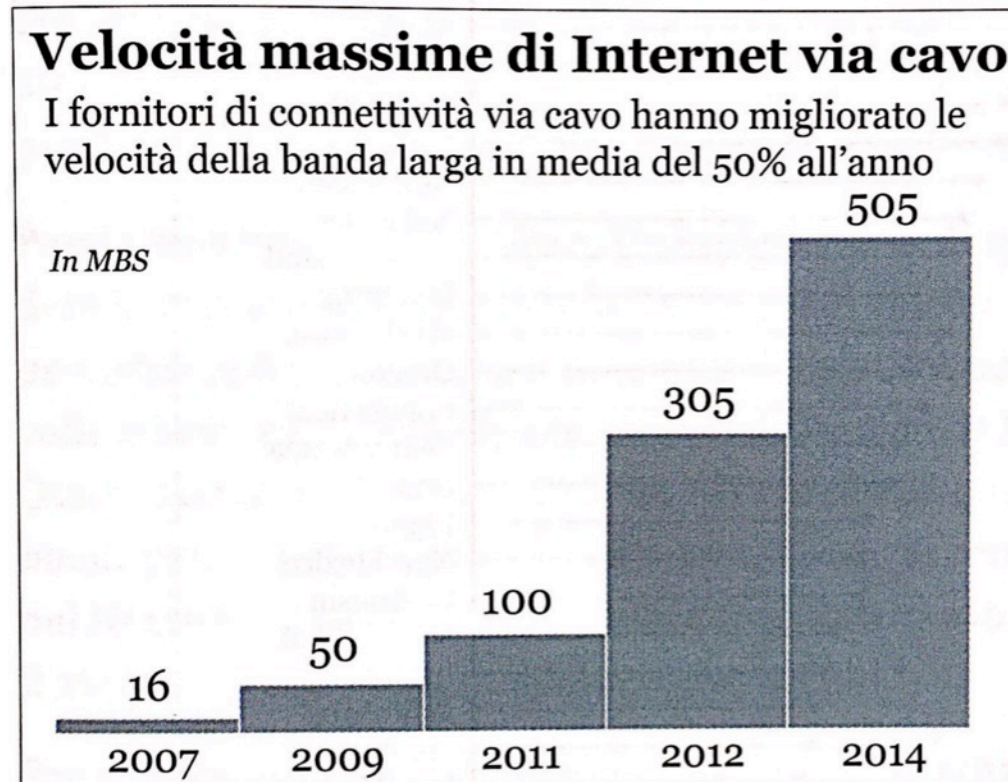
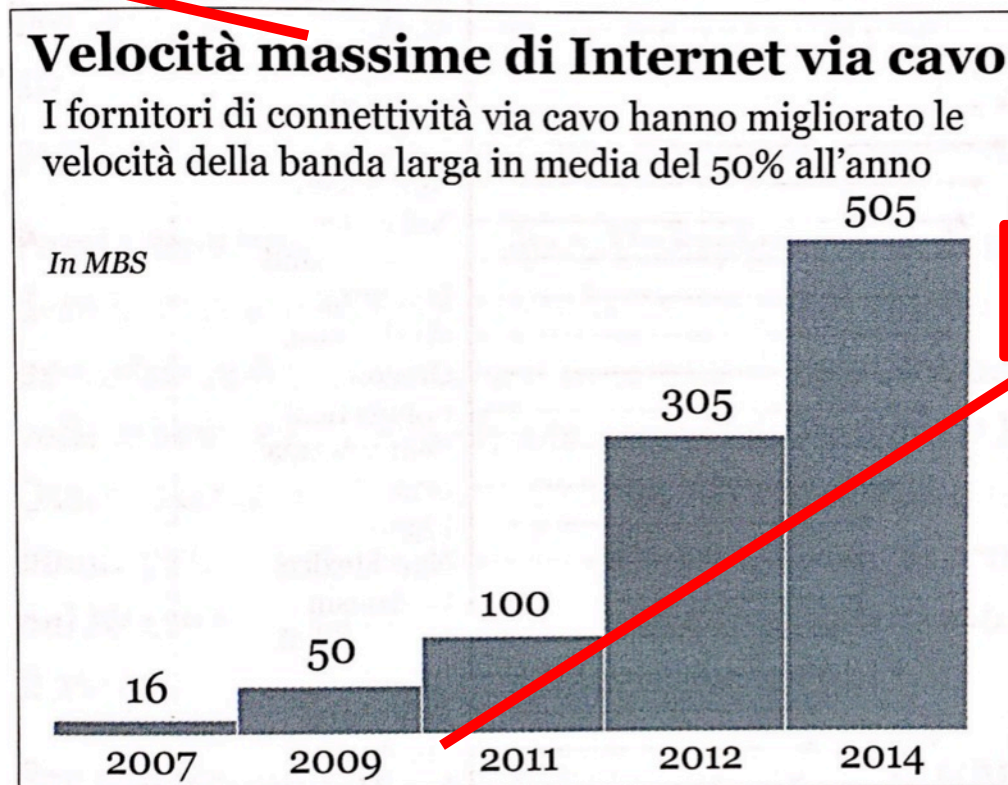


Figura tratta da  
“L’arte del vero” di A. Cairo

Queste sono le velocità massime. Molto bene. Però:

- quali sono le velocità minime?
- quali sono le velocità medie?
- quante persone hanno accesso alla velocità massima?
- come è distribuita geograficamente la velocità



Perché mancano certi anni?



# Anche chi è ben intenzionato...

- ...può fornire visualizzazioni fuorvianti
- Il rischio è che si elaborino politiche utili, ma non nelle zone dove c'è più bisogno
- Ecco un esempio dalla NHTSA (National Highway Traffic Safety Administration) sulle vittime degli incidenti stradali tra il 1975 e il 2012 nei vari stati degli USA

# Variazione percentuale: vittime della strada 1975-2012

## VITTIME TOTALI

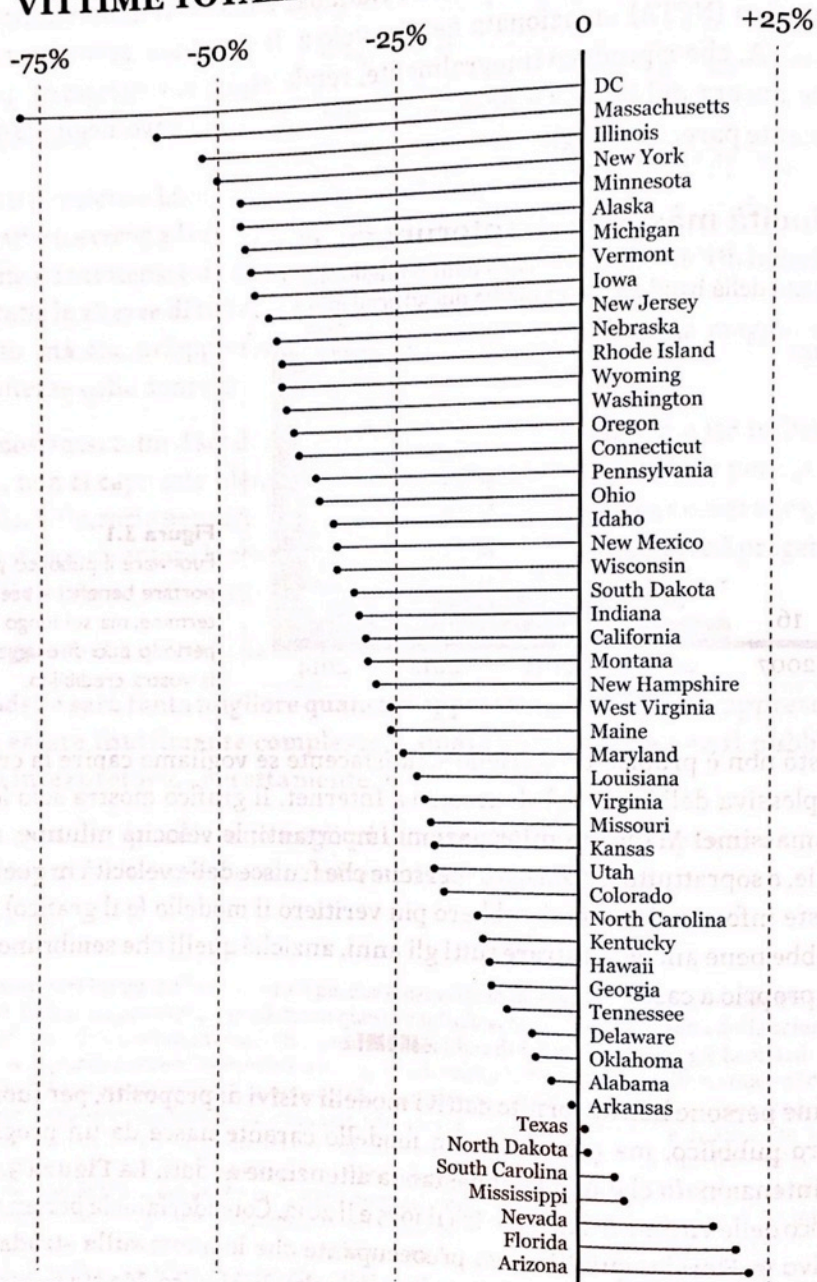
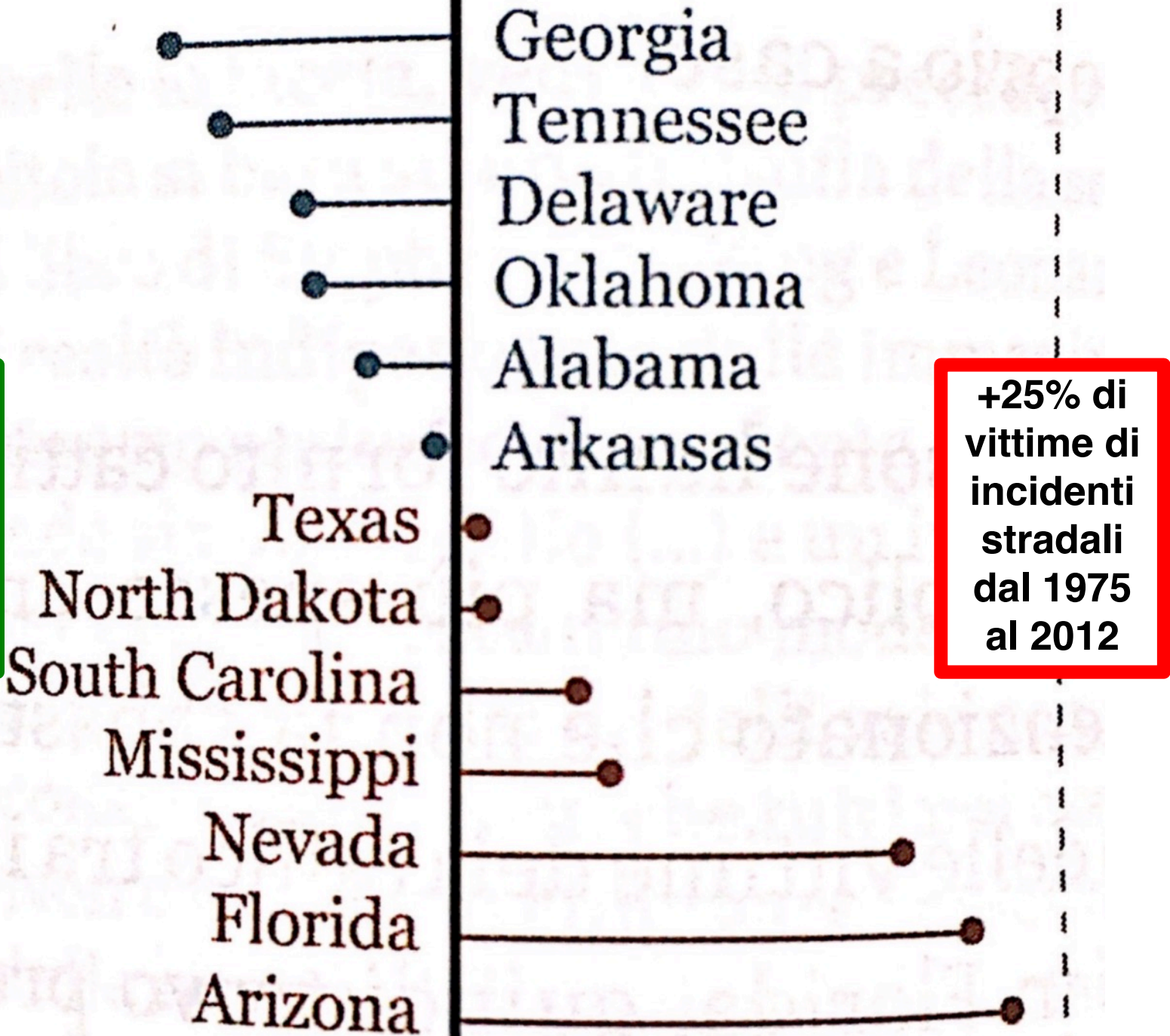


Figura tratta da "L'arte del vero" di A. Cairo

**-25% di  
vittime di  
incidenti  
stradali  
dal 1975  
al 2012**



**+25% di  
vittime di  
incidenti  
stradali  
dal 1975  
al 2012**

# Questa visualizzazione è veritiera?

- Non riporta dati falsi (anche se si dovrebbero fare accertamenti sulle fonti di dati a riguardo)
- Induce, però, l'osservatore a ragionamenti errati, come ad esempio pensare che la situazione delle morti su strada in Florida sia notevolmente peggiorata
- Perché il ragionamento è errato? Perché mancano dati indispensabili per una corretta interpretazione dei dati sulle morti su strada:
  - la popolazione dello stato è nel frattempo aumentata?
  - il traffico nello stato è nel frattempo aumentato?
- I valori assoluti hanno sicuramente un valore, ma se confrontiamo due periodi diversi, allora occorrono controlli di altre variabili che rendano tali valori relativi alle circostanze a cui si riferiscono
- Altrimenti sarebbe come equiparare il potere di acquisto di un dollaro nel 2012 con quello di un dollaro nel 1975



# What Food a Dollar Could Buy the Year

**1975: 1 pound of Chock full o'Nuts coffee**





# CHOCK FULL O'NUTS® COFFEE HEAVENLY ORIGINAL, MEDIUM ROAST, GROUND

\$5.00

★★★★★ (54) Reviews | Write a Review

## Size and Packaging

11.3 oz Can

Qty

-

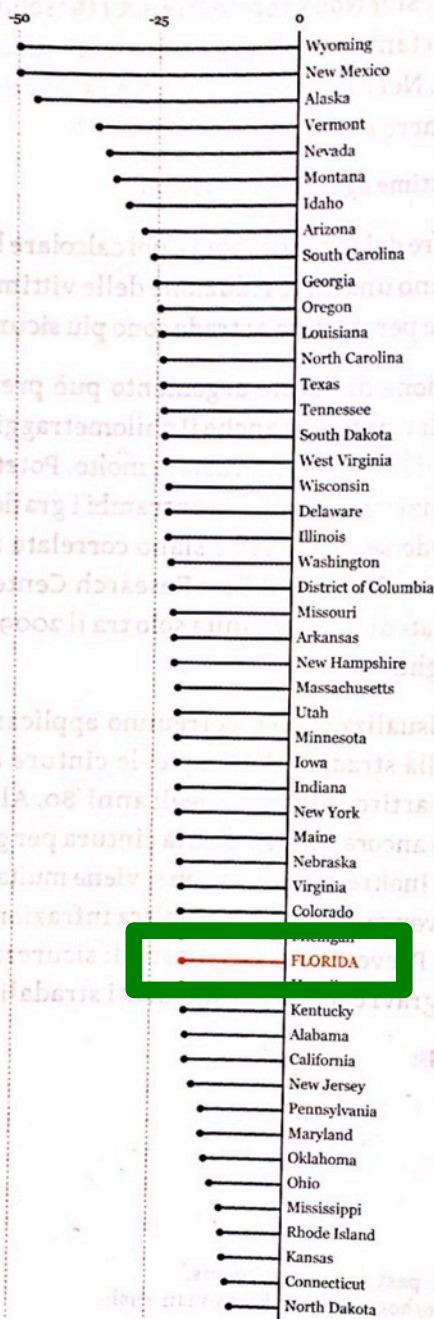
1

+

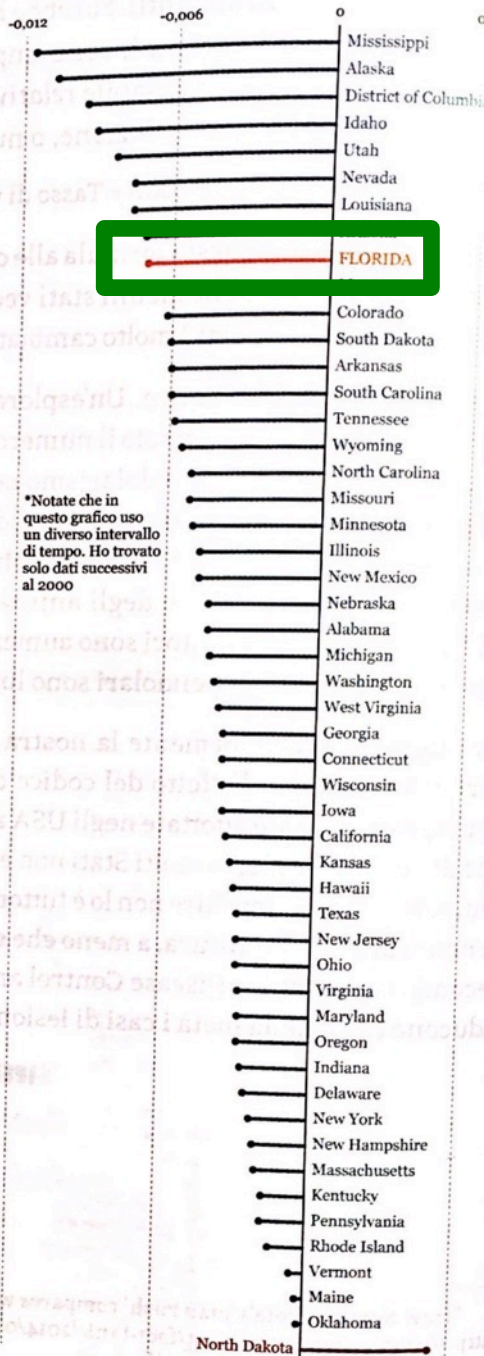
11,3 oz sono 0,7 pound

Se 0,7 pound costano 5 dollari, vuol dire che 1 pound costa più di 7 dollari

Variazione vittime del traffico 1975-2012  
PER 100.000 VEICOLI



Variazione vittime del traffico 2000-2012  
PER MILIONE DI MIGLIA-AUTO\*



\*Notate che in questo grafico uso un diverso intervallo di tempo. Ho trovato solo dati successivi al 2000

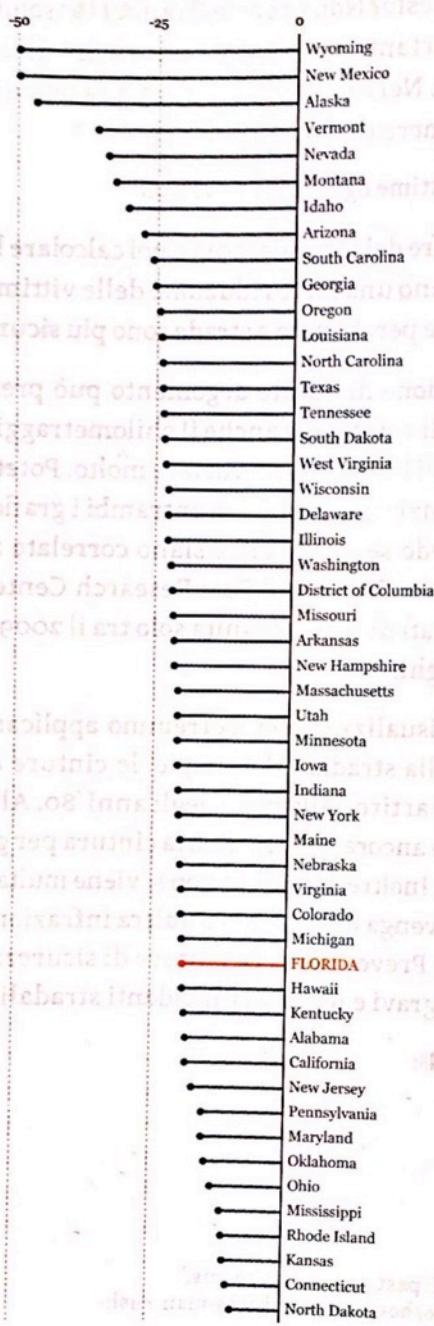
## Sorpresa!

Se prendiamo in considerazione dati sulla popolazione, in particolare dati sulla quantità di automobili, non solo tutti gli stati mostrano un miglioramento (decine di vittime in meno ogni 100mila veicoli), ma la Florida sale anche di qualche posizione in classifica.

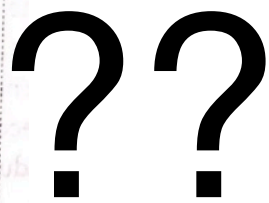
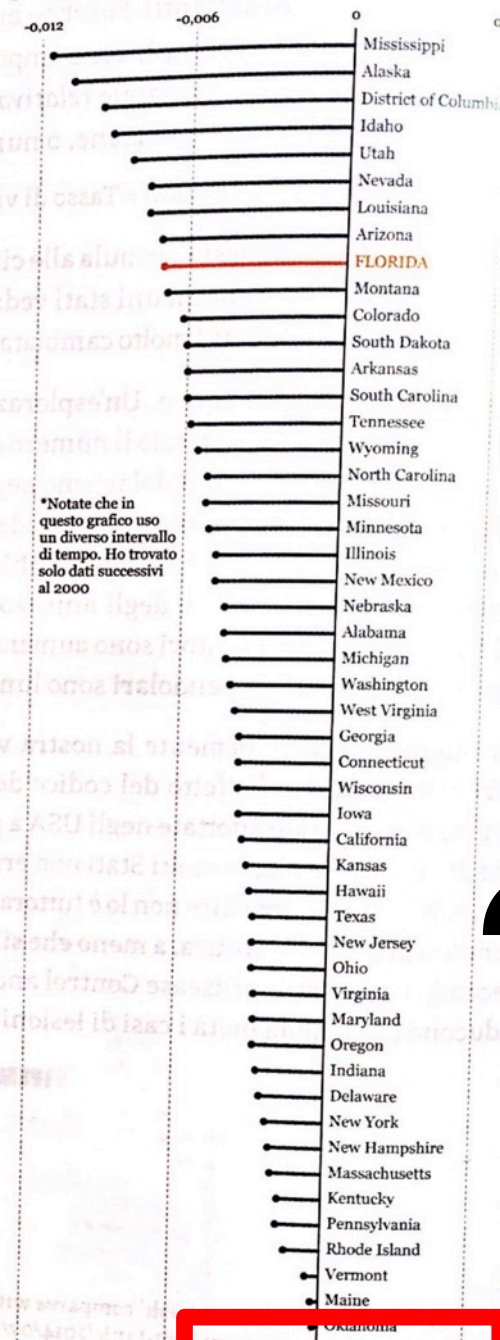
Attenzione: la quantità di auto è una variabile significativa, ma non dice nulla su quanto le auto vengano usate dalla popolazione (potrebbero essere auto usate solo per 100km al mese da persone che non devono fare pendolarismo). Ancor più significativo è il dato delle morti rispetto alle distanze percorse in auto. Per variazione di numero di vittime per 1 milione di miglia percorse, la Florida si piazza ancora meglio e arriva nella top 10 per miglioramento della sicurezza.



**Variazione vittime del traffico 1975-2012  
PER 100.000 VEICOLI**



**Variazione vittime del traffico 2000-2012  
PER MILIONE DI MIGLIA-AUTO\***



North Dakota



# Il misterioso caso del Nord Dakota



Immagine di TUBS, tratta da Wikipedia.org

# Perché?

- Si possono solo fare ipotesi.
- Un'ipotesi è basata sull'esplosione dell'industria petrolifera verso la fine degli anni 2000 nello stato, che si è rivelato avere un suolo ottimale per il fracking, il che ha portato a:
  - un incremento di 100mila lavoratori (dal 2009 al 2014)
  - con grandi tempi di pendolarismo (perché nessuno vive vicino ai campi petroliferi dove lavora)
- Questa ipotesi è corretta? È sbagliata? Solo una cosa è certa...



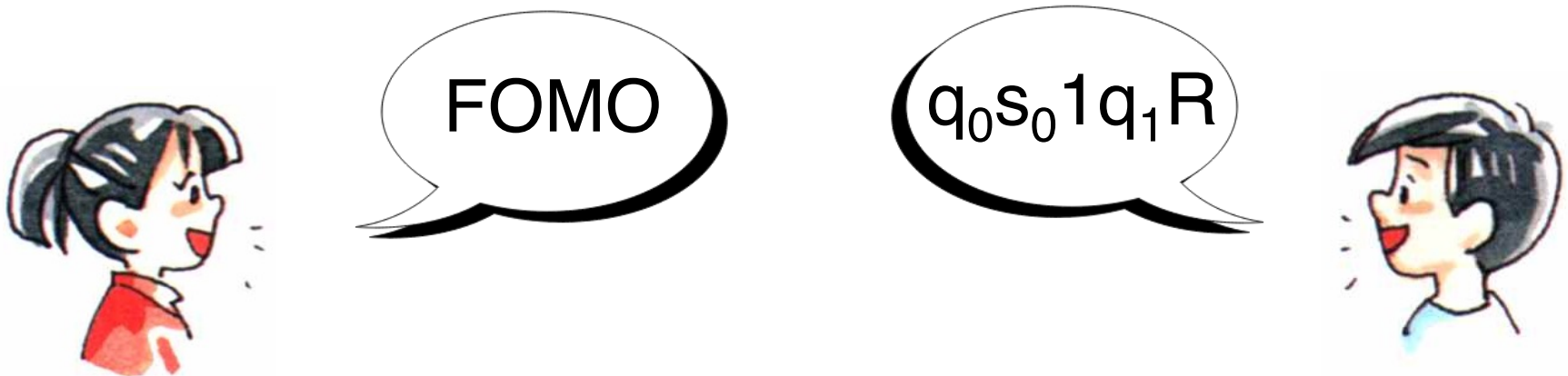
Foto di una strada del Nord Dakota. Fonte: [www.usnews.com](http://www.usnews.com)

Non saranno i dati sugli incidenti mortali che ci daranno una spiegazione. È vero che i dati possono aiutare l'osservatore a formulare spiegazioni, ma le cause di un fenomeno sono spesso esterne al fenomeno stesso.

Quindi, per verificare le ipotesi sull'incremento delle morti su strada in Nord Dakota, avremo bisogno di altri dati: un'indagine più vasta si rende necessaria.

# Visualizzazioni e mente umana

- In comunicazione, viene definito **contesto** l'insieme di informazioni che il mittente e il destinatario devono avere in comune perché la trasmissione di un messaggio abbia successo





# Visualizzazioni e mente umana

- Un discorso molto simile avviene per le visualizzazioni: sono un modello che funge da tramite tra la mente del disegnatore e la mente dell'osservatore

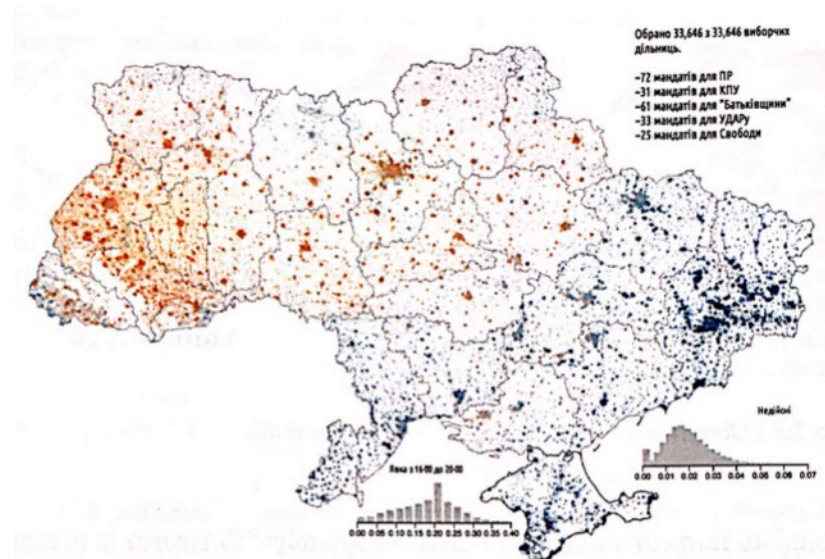
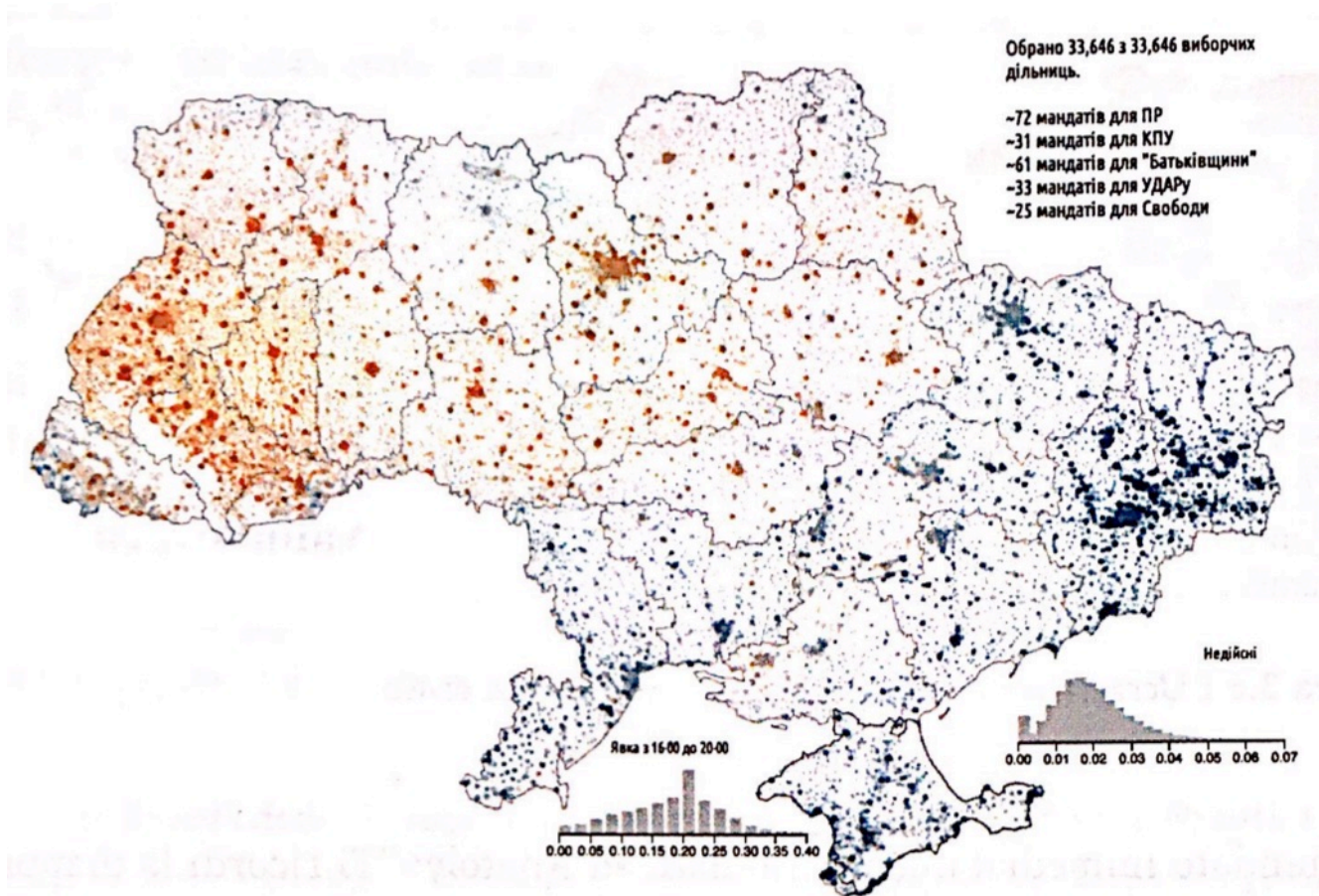


Figura tratta da "L'arte del vero" di A. Cairo

# Il caso Ucraina: 2012



Questa visualizzazione mostra i risultati delle elezioni ucraine del 2012, che hanno confermato la maggioranza parlamentare a sostegno del presidente filo-russo Viktor Yanukovich. I cerchi sulle mappe rappresentano i distretti elettorali, e hanno dimensioni in proporzione al numero di votanti. Il loro colore è determinato dal partito che ha vinto: arancione per il partito di opposizione filo-occidentale, blu per il partito di maggioranza filo-russo.

# Il caso Ucraina: 2013



Figura tratta da "L'arte del vero" di A. Cairo

Un anno dopo, a seguito del rifiuto del presidente di fare accordi con l'Unione Europea e delle sue proposte di un ravvicinamento alla Russia, sono scoppiate delle proteste, culminate con la fuga di Yanukovich in Russia e la sua sostituzione con un presidente ad interim in attesa di nuove elezioni.

Nel riportare le notizie sulle rivolte in Ucraina, molti giornali hanno pubblicato visualizzazioni come quelle sopra, mostrando un paese nettamente diviso in due per orientamento politico e lingua.



# Il caso Ucraina: it's complicated

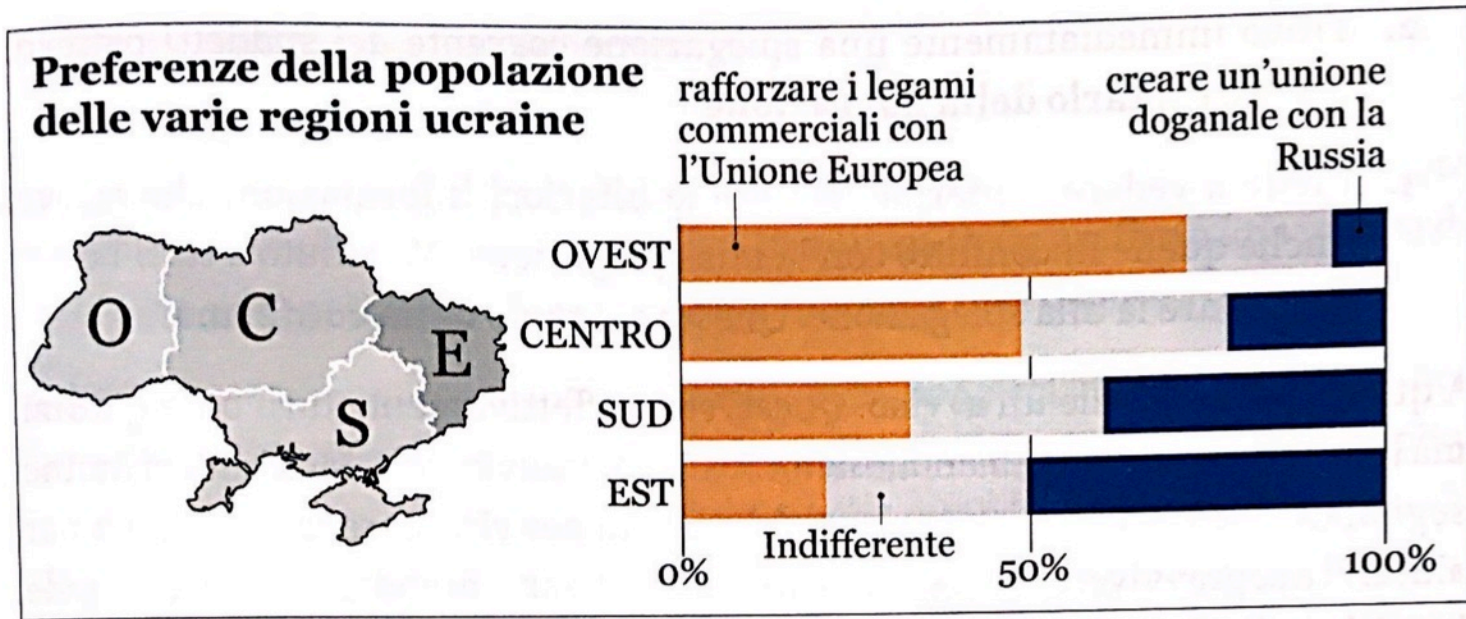


Figura tratta da "L'arte del vero" di A. Cairo

In realtà, raccogliendo dati più dettagliati, il quadro si fa più complicato. Non si ha una netta divisione tra Ovest e Est, tra pro-Europa e pro-Russia. Questi orientamenti sono netti solo in regioni ben delimitate del paese. Nella regione centrale e nella regione meridionale c'è una significativa fetta della popolazione che risulta indifferente alle due opzioni, ma anche decisiva nello stabilire una maggioranza all'interno della regione.



# Visualizzazioni e mente umana

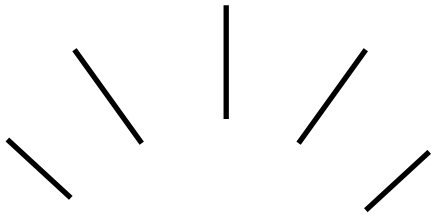
- Le visualizzazioni aiutano molto con la selezione degli elementi significativi di una realtà da rappresentare, e l'eliminazione di altri elementi non importanti
- Questa semplificazione, però, portata all'eccesso può anche semplificare troppo e fornire una rappresentazione non veritiera, o che induce nell'osservatore ragionamenti errati (come la divisione dell'Ucraina in due)
- In molti casi, per chiarire un argomento bisogna aumentare le informazioni, non ridurle

# Visualizzazioni e mente umana

- Ci sarà sempre una discrepanza, seppur piccola, tra quanto concepito dal disegnatore e quanto compreso dall'osservatore
- È essenziale ridurre al minimo la possibilità di fraintendimenti, anche perché la mente umana ha una naturale tendenza a compiere errori

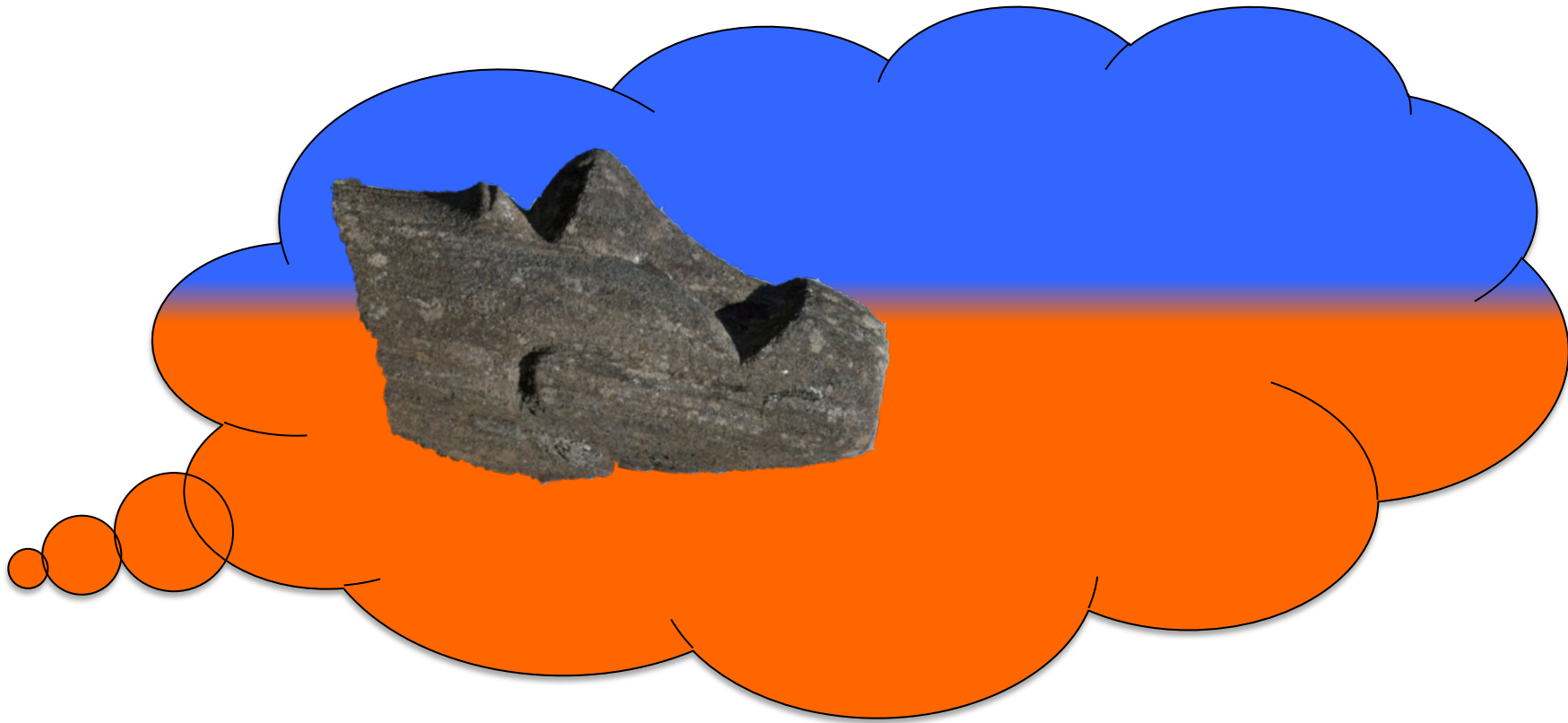
# Apofenia

Superficie di Marte (Viking-1, 1976). Fonte: NASA.



Si definisce **apofenia** la tendenza a trovare pattern indipendentemente dal fatto che siano reali o meno. In inglese si chiama **patternicity**.

# Narrazione



Esiste anche la tendenza alla **narrazione**, ossia a trovare subito una spiegazione coerente dei pattern trovati. (Attenzione: coerente non vuol dire necessariamente né vera né realistica.)

# Conferma



Il **bias di conferma** è la tendenza, una volta che si è formata una narrazione, a interpretare ogni ulteriore nuovo dato come conferma della narrazione, anche quelli che sono in conflitto con essa.

# Perché facciamo così

- Non ci sono teorie solide sulle cause di queste tendenze
- Un'ipotesi che ha un certo successo è di tipo evolucionistico:
  - sono sopravvissuti quegli esseri viventi che hanno “visto” un predatore pericoloso e sono scappati subito
  - non sono sopravvissuti quegli altri che hanno voluto aspettare per avere una conferma
- Arrivare a conclusioni affrettate che sovrastimano il rischio è un vantaggio nella natura selvaggia
- Arrivare a conclusioni affrettate, però, vuol dire spesso arrivare a conclusioni errate



# Vedete dei pattern?

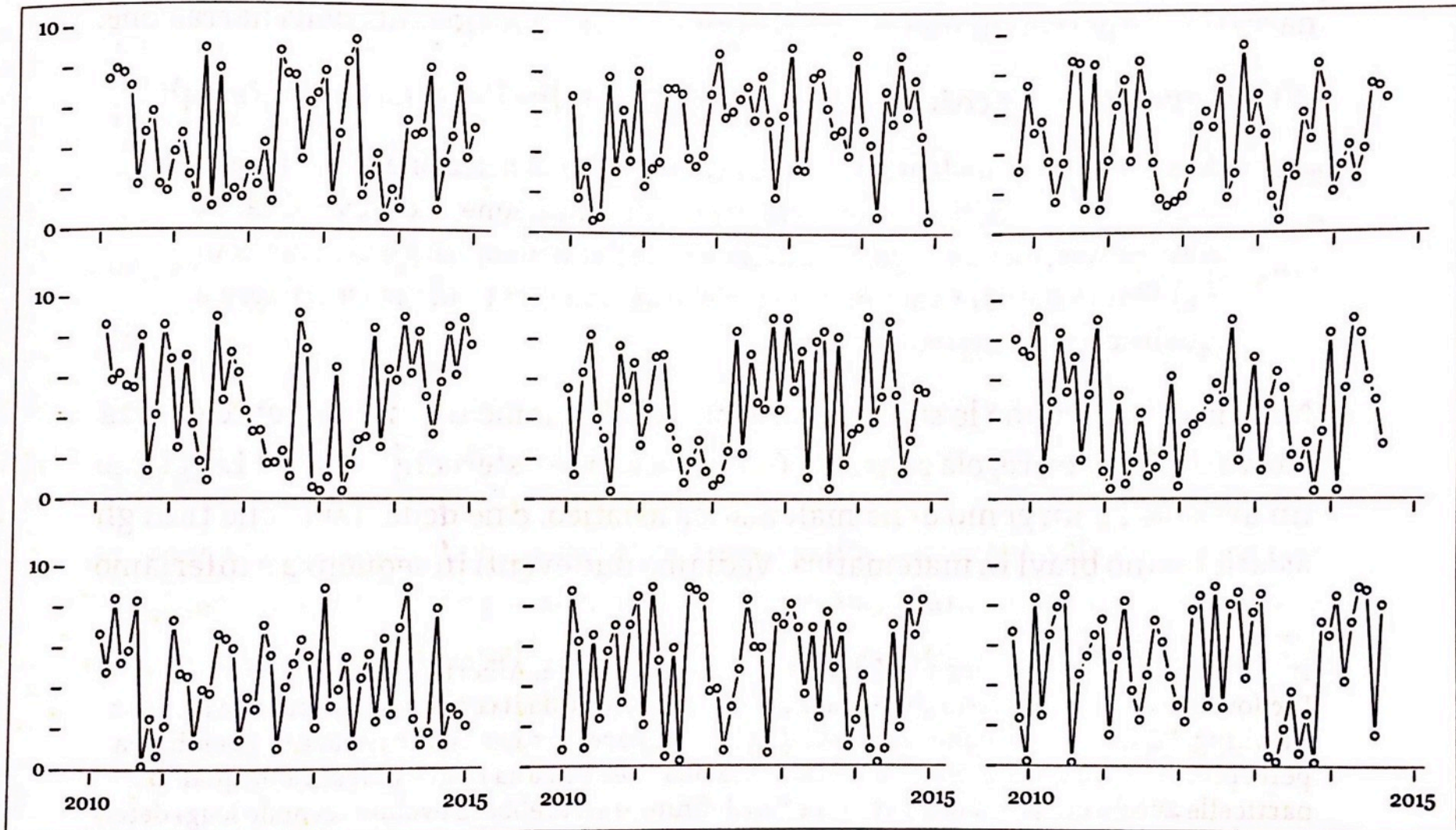


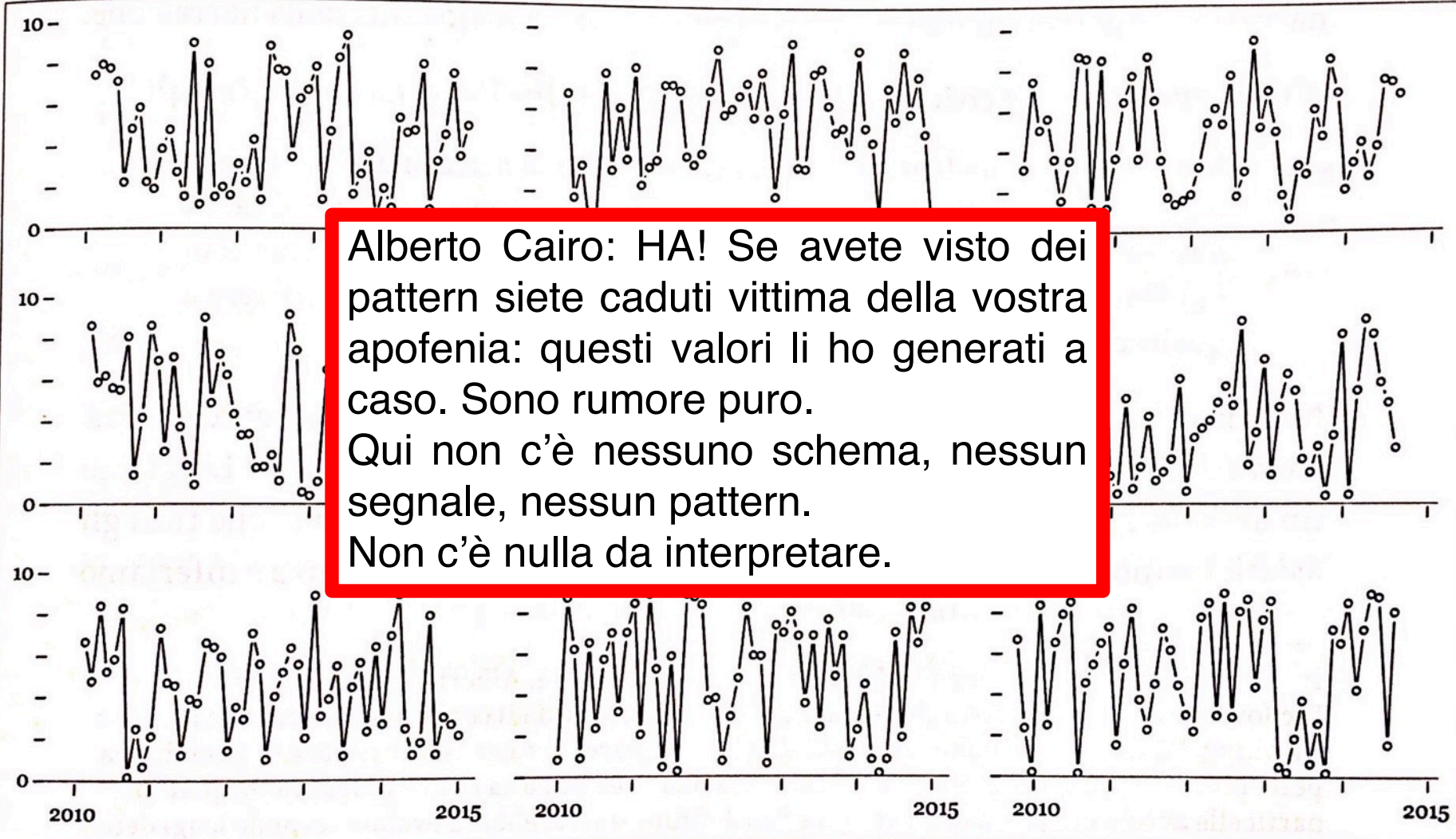
Figura tratta da "L'arte del vero" di A. Cairo

# Pattern vs noise

- Definiamo **noise** (o **rumore**) fluttuazioni casuali in un segnale
- Più in generale, possiamo identificare il rumore con valori casuali
- I pattern, invece, sono definiti da segnali ben precisi, da schemi ben definiti
- L'apofenia è la tendenza a scambiare rumore per pattern

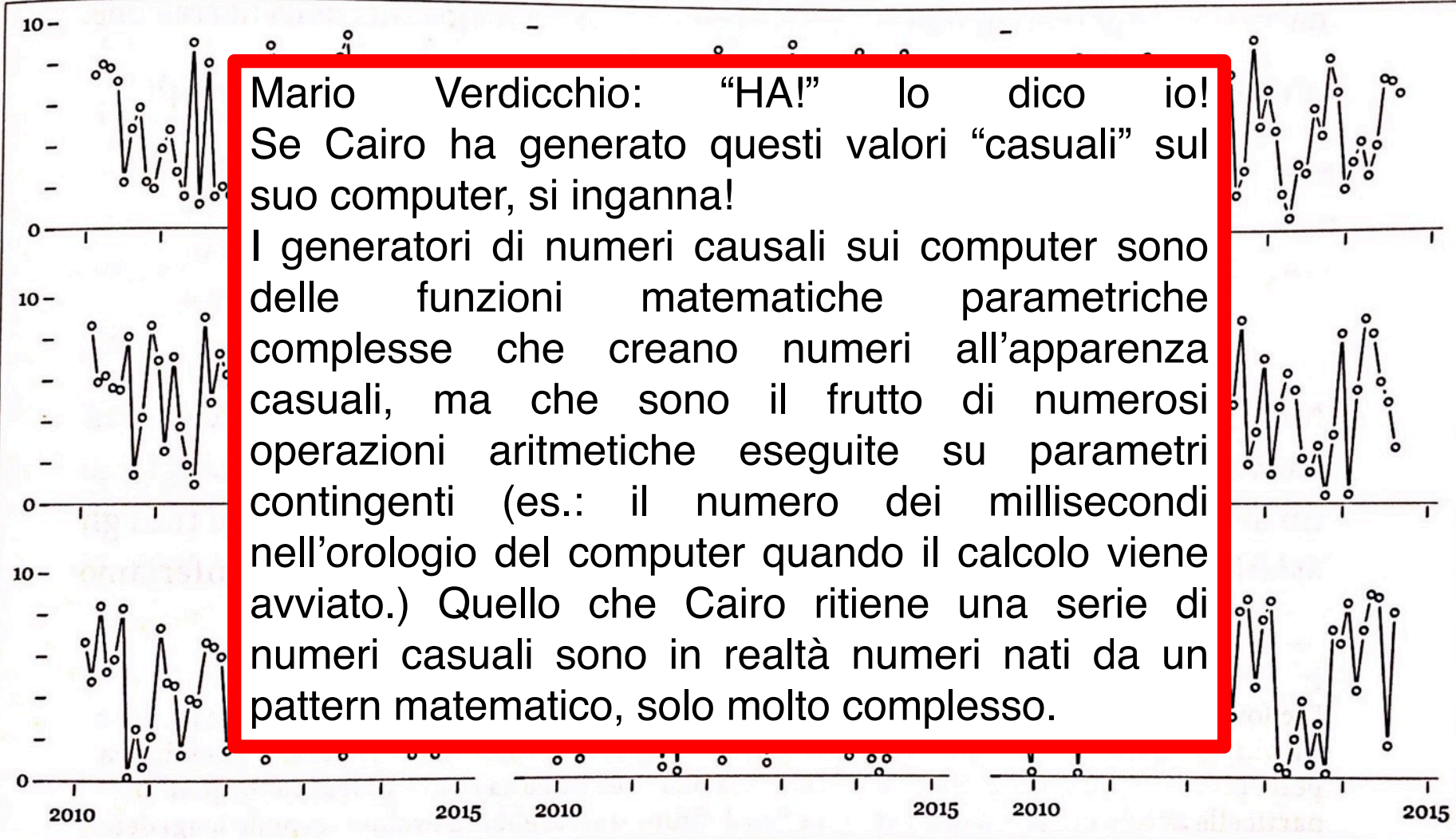


# Tornando all'esempio...



Alberto Cairo: HA! Se avete visto dei pattern siete caduti vittima della vostra apofenia: questi valori li ho generati a caso. Sono rumore puro. Qui non c'è nessuno schema, nessun segnale, nessun pattern. Non c'è nulla da interpretare.

# Tornando all'esempio...



Mario Verdicchio: “HA!” lo dico io!  
Se Cairo ha generato questi valori “casuali” sul suo computer, si inganna!  
I generatori di numeri casuali sui computer sono delle funzioni matematiche parametriche complesse che creano numeri all'apparenza casuali, ma che sono il frutto di numerosi operazioni aritmetiche eseguite su parametri contingenti (es.: il numero dei millisecondi nell'orologio del computer quando il calcolo viene avviato.) Quello che Cairo ritiene una serie di numeri casuali sono in realtà numeri nati da un pattern matematico, solo molto complesso.

# Plot twist!



Vedo pattern inesistenti...

Immagine tratta dal film "Il sesto senso" di M. Night Shyamalan

Un osservatore delle visualizzazioni di Cairo, a causa della sua apofenia, vede pattern dove non ci sono.

Cairo genera la visualizzazione per mezzo di un computer, e non vede alcun pattern nei dati, perché li genera con una funzione "random".

In realtà, se si tratta di una funzione che gira su un computer, i suoi risultati non possono essere davvero casuali, perché il computer è un sistema deterministico che funziona solo con regole matematiche precise.

Quindi, Cairo, che accusa gli altri di vedere pattern, può essere accusato di non vederli!

# Facciamo chiarezza 1/3

- Cairo, in realtà, non ha tutti i torti, e probabilmente sa benissimo (da data scientist qual è) che i dati generati “casualmente” da un computer hanno solo l'apparenza casuale
- Vengono chiamati in maniera più corretta dati “pseudo-casuali”

# Facciamo chiarezza 2/3

- Le funzioni “random” che i programmatori ci mettono a disposizione sono disegnate apposta per rendere i pattern matematici delle funzioni sottostanti completamente invisibili all’occhio umano: dovremmo essere in grado di calcolare l’ora esatta (al millisecondo) in cui la funzione viene lanciata, conoscere perfettamente il suo codice, e eseguire miliardi di calcoli al secondo per poter riconoscere tali pattern una volta che i risultati vengono stampati su schermo

# Facciamo chiarezza 3/3

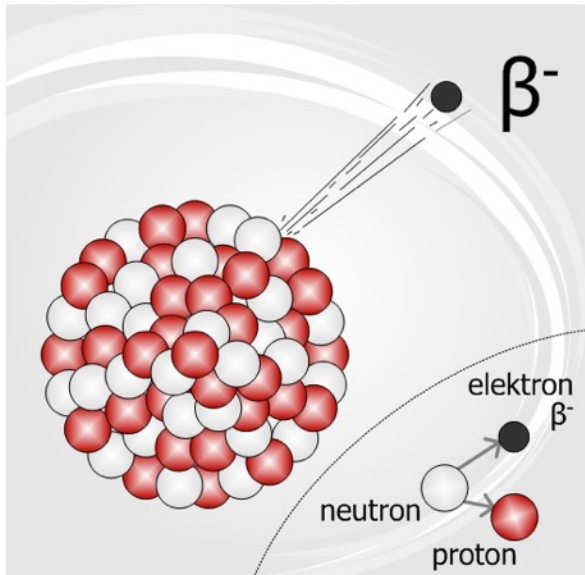
- Con le funzioni “random” dei computer, i pattern sono effettivamente impossibili da vedere, quindi possiamo benissimo accettare l’ipotesi che non esistano, anche perché se anche in qualche modo riuscissimo a calcolarli, darebbero solo indicazioni su processi in atto all’interno del nostro computer, e non rifletterebbero alcun aspetto interessante della realtà che ci circonda

# Pattern: da vedere o da ignorare?

- L'apofenia è problematica perché fa vedere pattern laddove non esistono
- Le visualizzazioni in generale, ma soprattutto in geourbanistica, devono puntare ad essere veritiere, quindi l'apofenia va evitata
- Ignorare pattern che non ci interessano (come le funzioni matematiche all'interno di un computer) va bene
- Ignorare pattern che potrebbero invece essere utili ai nostri scopi può inficiare l'utilità di una visualizzazione



# Vera casualità?



Esistono dispositivi per creare numeri casuali veri, e sono basati sul decadimento di atomi radioattivi. Nel nucleo di questi atomi, un neutrone ogni tanto decade e diventa un protone e rilascia un elettrone e un antineutrino elettronico. L'elettrone colpisce un sensore e il decadimento è percepito.

Considerati 4 atomi (1,2,3,4), sia  $T_{12}$  l'intervallo di tempo tra il decadimento di 1 e il decadimento di 2; sia  $T_{34}$  l'intervallo di tempo tra il decadimento di 3 e il decadimento di 4.

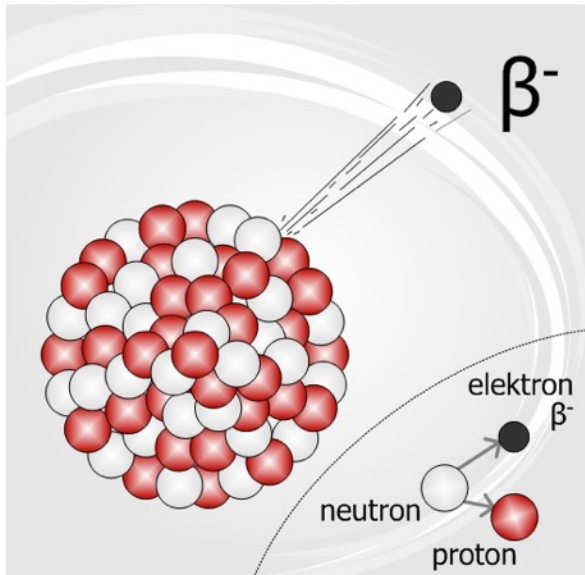
Se  $T_{12} < T_{34}$  scriviamo uno 0.

Se  $T_{12} \geq T_{34}$  scriviamo un 1.

Questa cifra è veramente casuale, perché basata su fenomeni fisici che non hanno alcun pattern.



# Chi lo dice?



Esistono dispositivi per creare numeri casuali veri, e sono basati sul decadimento di atomi radioattivi. Nel nucleo di questi atomi, un neutrone ogni tanto decade e diventa un protone e rilascia un elettrone e un antineutrino elettronico. L'elettrone colpisce un sensore e il decadimento è percepito.

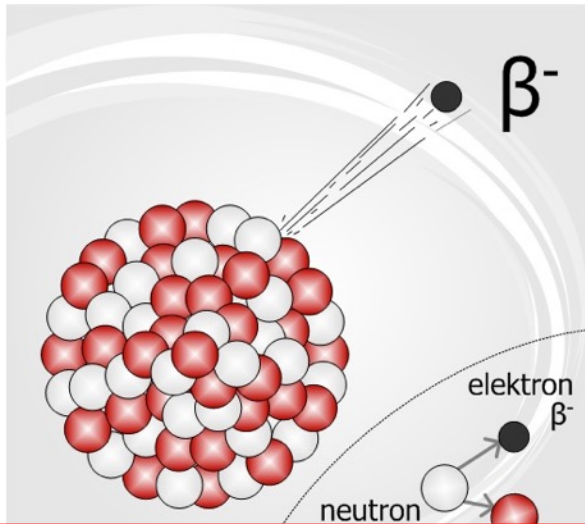
Considerati 4 atomi (1,2,3,4), sia  $T_{12}$  l'intervallo di tempo tra il decadimento di 1 e il decadimento di 2; sia  $T_{34}$  l'intervallo di tempo tra il decadimento di 3 e il decadimento di 4.

Se  $T_{12} < T_{34}$  scriviamo uno 0.

Se  $T_{12} \geq T_{34}$  scriviamo un 1.

Questa cifra è veramente casuale, perché basata su fenomeni fisici che non hanno alcun pattern.

# Chi lo sa?



Non è che si tratta di un pattern che esiste ma che è per noi invisibile? Non un pattern computazionale dentro al computer, ma un pattern di una legge fisica universale che non abbiamo ancora scoperto o che non riusciremo mai a scoprire?

Esistono dispositivi per creare numeri casuali veri, e sono basati sul decadimento di atomi radioattivi. Nel nucleo di questi atomi, un neutrone ogni tanto decade e diventa un protone e rilascia un elettrone e un antineutrino elettronico. L'elettrone colpisce un sensore e il decadimento è percepito.

Considerati 4 atomi (1,2,3,4), sia  $T_{12}$  l'intervallo di tempo tra il decadimento di 1 e il decadimento di 2; sia  $T_{34}$  l'intervallo di tempo tra il decadimento di 3 e il decadimento di 4.

Se  $T_{12} < T_{34}$  scriviamo uno 0.

Se  $T_{12} \geq T_{34}$  scriviamo un 1.

Questa cifra è veramente casuale, perché basata su fenomeni fisici che non hanno alcun pattern.

# Da “la Repubblica” del 7 aprile 2021

## L'anomalia del muone: l'esperimento che suggerisce l'esistenza di nuove forze della natura

di Matteo Marini

*L'acceleratore di particelle del Fermilab, a Chicago, ha misurato un'anomalia nel valore del momento magnetico del muone. Sembrerebbe un dettaglio riservato agli appassionati di fisica. Invece è una notizia che apre la porta alla presenza di nuove particelle. Perfino di un secondo bosone di Higgs*



# La creazione di storie

- Qui entra in gioco la **narrazione**, ovvero letteralmente la creazione di **storie**
- La parola “storia” sembra subito indicare finzione, quindi sarebbe più opportuno chiamarle **ipotesi**
- Tornando all’esempio degli atomi radioattivi, abbiamo diverse ipotesi:
  - il loro decadimento è davvero casuale
  - il loro decadimento è regolato da leggi non ancora scoperte
  - il loro decadimento è regolato da leggi che sono impossibili da scoprire

# Le storie piacciono di più

*“Al suo livello più elementare, una storia è una descrizione di un accadimento che contiene una qualche forma di sensazione, o di teatralità. In altre parole, è una spiegazione di causa ed effetto inzuppata nelle emozioni (...) Siamo nati per raccontare storie, e abbiamo la tendenza a credere a quelle che inventiamo.”*

Will Storr

“The Unpersuadables: Adventures with the Enemies of Science”

(2014)



# “Non prendere freddo!”

- Questa è una delle storie più dure a morire, perché ancora molte persone ci credono
- La scienza ha più volte sperimentalmente confermato che il freddo non fa venire il “raffreddore” (il cui nome stesso è una conseguenza della storia):
  - alcuni virus si mantengono attivi più a lungo se le temperature sono basse
  - quando fa freddo la gente tende a stare più tempo all’interno, dove ha più probabilità di entrare in contatto con altre persone che hanno un virus
- Sono i virus a causare il raffreddore (**e peggio**), non il freddo

# Il potere della narrazione

- La narrazione può essere uno strumento potente per comunicare in modo efficace
- È però anche pericolosa perché rende ciechi alle eventuali prove che dovrebbero spingere a rivedere o scartare i modelli adottati finora
- L'attivismo radicale, la difesa acritica delle cause, il giornalismo "d'opinione" sono tutte attività che si basano sulla narrazione
- A rafforzare il potere della narrazione, c'è il confirmation bias

# Confirmation bias



# Confirmation bias


- Attenzione: il conforto delle bugie nella vignetta **non** deriva dal fatto che quelle bugie dicono cose piacevoli (es.: “sono ancora in ottima forma”, “i capelli grigi aumentano il mio fascino”, “queste slide sono davvero interessanti”)
- Tale conforto deriva dalla convergenza tra le bugie e il modello nella mente di chi le sente: sembrano **confermare** quanto noi crediamo, quindi tendiamo a volere che siano vere
- Allo stesso modo, le verità scomode non sono scomode perché sono negative in sé, ma perché causano **dissonanza cognitiva**

# Dissonanza cognitiva



In "A Theory of Cognitive Dissonance" (1957), Leon Festinger ha per la prima volta proposto l'ipotesi che gli esseri umani avessero il desiderio di mantenere una coerenza psicologica per poter vivere bene. Una persona che percepisce una inconsistenza tra le sue credenze, o tra le sue credenze il suo comportamento, prova un disagio psicologico ed è spinto a ridurre tale dissonanza.






I medici  
esagerano  
sui danni del  
fumo.

Un aumento statistico  
della probabilità di un  
tumore non vuol dire  
che mi verrà di sicuro.

Se smettessi  
ingrasserei, e  
questo è altrettanto  
pericoloso.

Fumare è  
davvero  
piacevole e vale  
la pena rischiare.



bla bla bla  
bla bla bla  
bla bla bla

bla bla bla bla bla bla  
bla bla bla bla bla bla  
bla bla bla bla bla bla

bla bla bla bla bla  
bla bla bla bla bla  
bla bla bla bla bla

Ognuno di questi pensieri viene formulato per ridurre la dissonanza, ossia mitigare il modello secondo cui il fumo fa male (l'unico scientifico) con altri modelli alternativi (basati sulle emozioni e sulla retorica). Il desiderio di ridurre la dissonanza rende la persona meglio predisposta verso storie che confermino i modelli alternativi. Questo è il cognitive bias.



# Torniamo al climate change

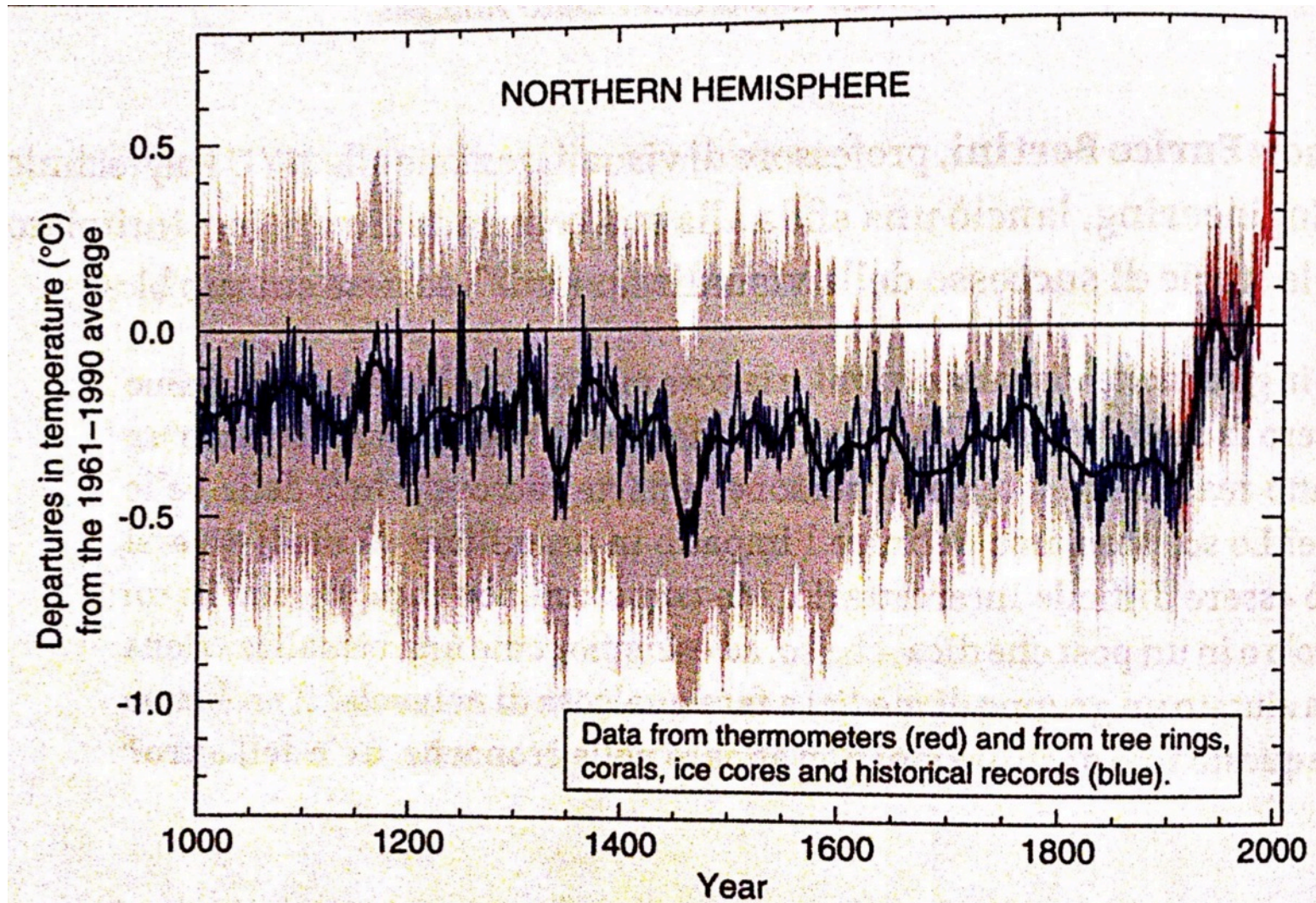


Grafico di Mann, Bradley e Hughes. Figura tratta da "L'arte del vero" di A. Cairo

“Il motivo principale per cui la gente è in disaccordo sul cambiamento climatico non è che è stato comunicato in modo poco comprensibile, ma che le posizioni sul tema trasmettono valori (preoccupazioni per la comunità contro indipendenza del singolo, abnegazione prudente contro eroica ricerca di gratificazione, umiltà contro astuzia, armonia con la natura contro dominio della stessa) che segnano linee di demarcazione culturali.

“La semplice amplificazione o il miglioramento della chiarezza dell’informazione in materia non genererà un consenso da parte del pubblico se chi comunica il rischio non tiene conto degli indizi che determinano quanto le percezioni dei rischi del cambiamento climatico siano espressioni dell’impegno culturale di chi le forma. Una disattenzione del genere rischia di aumentare il divario tra gli schieramenti.



“I cittadini che credono in valori di gerarchia e individualismo snobbano le informazioni scientifiche sui cambiamenti climatici in parte perché associano il problema all’antagonismo contro commercio e industria.

“Le persone tendono a interpretare le sfide alle credenze predominanti nella loro comunità culturale come attacchi alla competenza di coloro in cui credono e che considerano delle guide.

“Questa implicazione, che naturalmente provoca una resistenza, può essere rafforzata quando comunicatori con un’identità culturale riconoscibile accusano aspramente di scarsa intelligenza o integrità chi li contrasta.”

Testo tratto da:

Dan Kahan et al. “The tragedy of the risk-perception commons: culture conflict, rationality conflict, and climate change.”

*Temple University legal studies research paper* (2011).

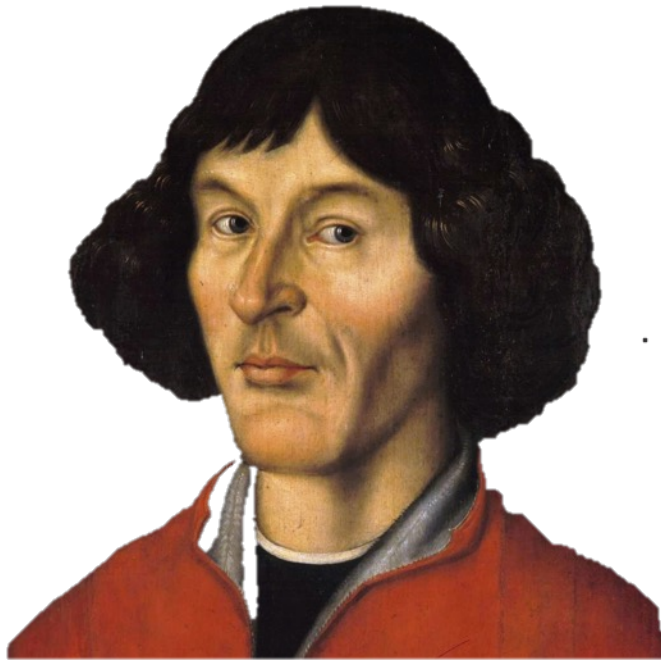
- Il modo in cui presentiamo le informazioni è importante quanto la solidità delle informazioni stesse.

# Modelli

- Tornando al concetto di modello, abbiamo visto che si può riferire a visualizzazioni, ma anche a credenze personali, o teorie scientifiche
- Lo scopo dei designer, dei ricercatori e degli scienziati onesti è quello di trovare modelli il più possibile vicini alla realtà
- Vista l'affinità con la scienza, un'idea potrebbe essere quella di prendere in prestito caratteristiche del metodo scientifico

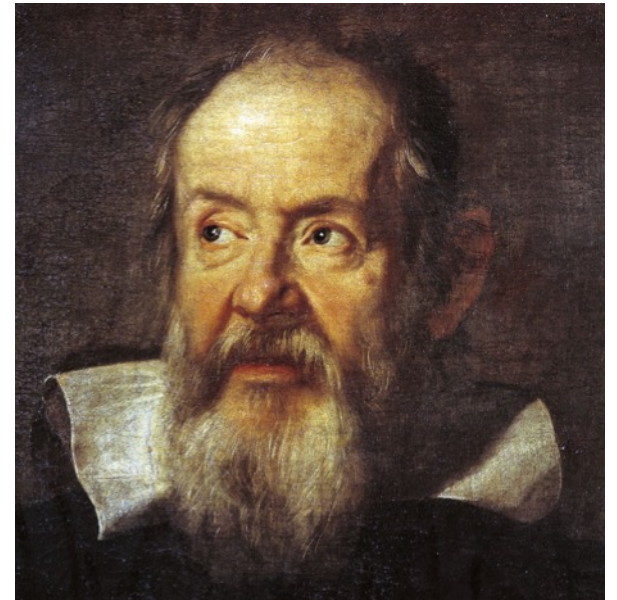
# Metodo scientifico

- Generalmente si riconosce a Francis Bacon (1561–1626) il merito di aver formalizzato per primo il metodo scientifico.
- Avvocato di successo e eminente filosofo londinese, propose un nuovo approccio all'indagine scientifica, che pubblicò nel 1621 nel come “Novum Organum Scientiarum”.
- Propose il ragionamento induttivo (dal particolare al generale) come base del pensiero scientifico.
- Era convinto che solo un sistema chiaro e preciso di indagine scientifica avrebbe assicurato il dominio dell'uomo sul mondo.



**Nicolaus Copernicus  
(1473–1543)**

**I lavori di Copernico e di  
Galileo hanno  
avuto un'influenza  
notevole su Bacon.**



**Galileo Galilei  
(1564 – 1642)**





# Dal metodo scientifico alle visualizzazioni

- Per rendere le visualizzazioni più veritiere possibili, applichiamo gli strumenti di pensiero rigorosi:
  - logica
  - statistica
  - sperimentazione
- Informazioni più abbondanti e migliori portano a modelli migliori
- 400 anni di scienza empirica e sperimentale hanno dimostrato che questi strumenti funzionano
- Se anche la comprensione perfetta della realtà è irraggiungibile (vedi il decadimento degli atomi), come teorie scientifiche sempre migliori soppiantano le precedenti e ci danno una migliore descrizione della realtà, così possiamo aspirare ad avere visualizzazioni che approssimano la realtà sempre meglio

# Esempio: giornalisti in Georgia



Immagine di TUBS, tratta da Wikipedia.org

# Esempio: giornalisti in Georgia

- Titolo di un giornale del 15/08/2013:  
**“Uno studio rivela che oltre un quarto dei laureati in giornalismo vorrebbe aver scelto un'altra carriera”.**
- L'articolo inizia con un'indicazione della fonte di dati:  
“Circa il 28% dei laureati in giornalismo vorrebbe aver scelto un altro settore, dice l'annuale indagine sui laureati del Grady College dell'Università della Georgia.”
- È vero? È falso? O meglio, i dati di questa indagine sono una approssimazione buona della realtà?

# Possibile miglioramento 1

- Il dato citato dall'articolo si riferiva all'anno accademico appena concluso
- Un possibile miglioramento viene dall'osservare dati analoghi da anni precedenti
- Cairo lo ha fatto e ha scoperto quanto mostrato in questo grafico

# Laureati (triennali) in giornalismo e comunicazione di massa che vorrebbero aver scelto un'altra carriera

**Percentuale**

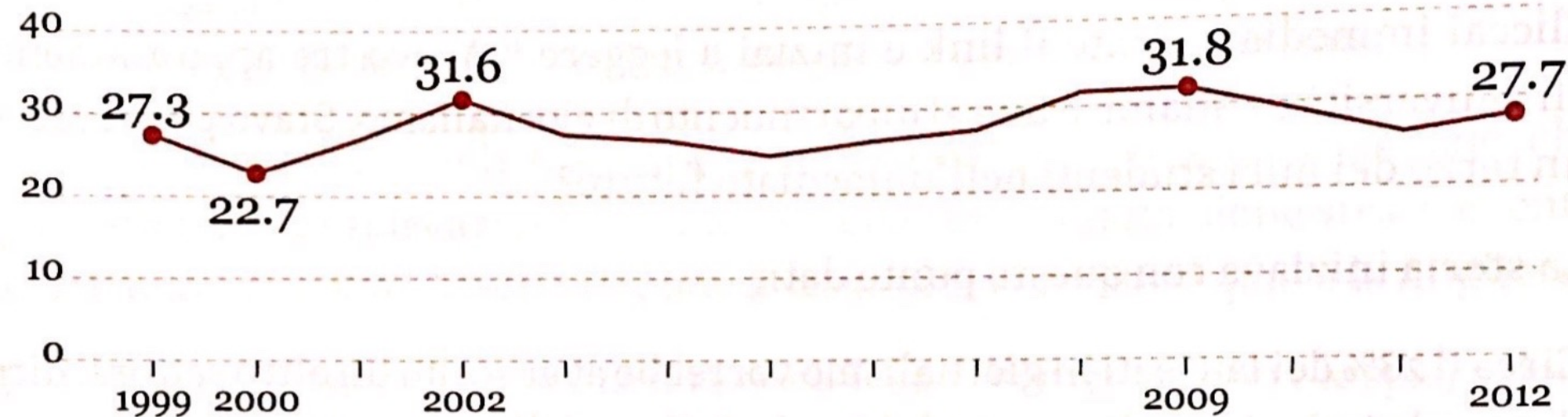


Figura tratta da "L'arte del vero" di A. Cairo

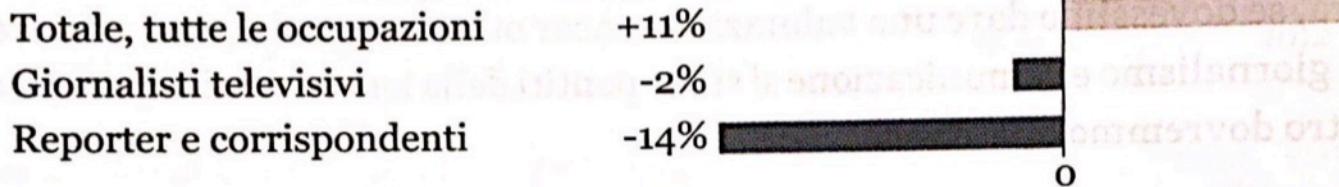
- Scopriamo che tale percentuale è rimasta pressoché invariata (a parte una valle nel 2000 che merita ulteriori indagini) dal 1999
- Quindi già un titolo quasi sensazionalista che sembra alludere a un fenomeno nuovo, o a un crollo, sembra inadeguato e non rappresentativo della realtà



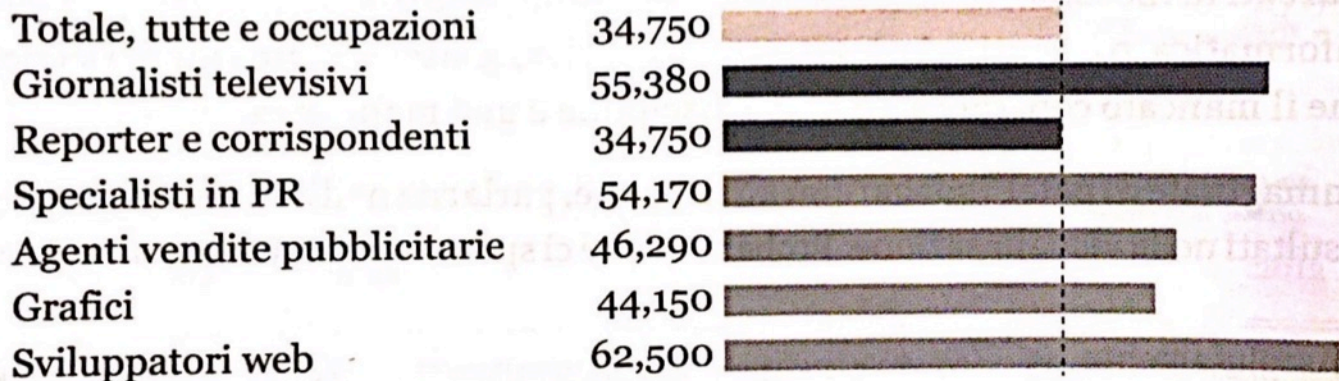
# Possibile miglioramento 2

- La percentuale non è cambiata significativamente dal 1999, ma molte cose sono cambiate dal 1999:
  - c'è stato il boom di Internet
  - c'è stato 9/11
  - c'è stata la grande crisi dell'editoria
  - c'è stata la crisi economica del 2008
- Un possibile miglioramento del modello è quello di ampliare il contesto prendendo in considerazione altri dati che sono connessi con quelli iniziali
- Esempio: quanto si guadagna nel giornalismo?

### **Variazione nell'occupazione, proiezione 2012-2022**



### **Stipendi medi annuali, maggio 2012. In \$ USA**



FONTE: U.S. Bureau of Labor Statistics, Occupational Employment Statistics

Figura tratta da "L'arte del vero" di A. Cairo

- Con dati ulteriori dal Bureau di statistica del lavoro (quello della mazza da hockey delle richieste di sussidio legate al covid-19), abbiamo conferma che il mercato del lavoro per i giornalisti è in crisi. Non solo...
- Scopriamo che la professione più comune per chi si laurea in giornalismo, ossia reporter, è pagata non tanto bene: esattamente lo stipendio medio negli USA, con media calcolata su tutti i lavori, inclusi i peggio pagati.

# Possibile miglioramento 3

- Non si hanno dati a riguardo, ma come se la cavano le altre facoltà?
- Chi ci dice che il 28% di laureati in giornalismo insoddisfatti non sia una quantità molto più piccola confrontata con il 40% dei laureati in legge, o il 50% dei laureati in lettere? (Ipotesi tutte da verificare)
- Allargare il contesto con ulteriori dati sicuramente aiuta a interpretare meglio i dati che abbiamo e a costruire visualizzazioni più veritiere

# Un titolo migliore

- Con tutte le informazioni aggiuntive che abbiamo scoperto con altri dati, collegati al data set iniziale, il solo menzionato nell'articolo, siamo in grado di proporre un titolo migliore.
- Non più:  
    **“Uno studio rivela che oltre un quarto dei laureati in giornalismo vorrebbe aver scelto un'altra carriera”**
- bensì:  
    **“Uno studio rivela che gli scossoni al mercato del lavoro degli ultimi dieci non hanno intaccato la fermezza dei laureati in giornalismo”.**
- Tutta un'altra **storia**.

# Lezione imparata

- Non affrettatevi a scrivere un titolo o una storia, né a progettare una visualizzazione immediatamente dopo aver trovato un pattern, un punto dati o un fatto interessante.
- Fermatevi e pensate.
- Cercate altre fonti e persone che possano aiutarvi a togliere i paraocchi e a eliminare il confirmation bias, cercando fattori esterni che possano aiutarvi a spiegare quanto scoperto.
- Solo a quel punto potete decidere:
  - che cosa dire
  - come dirlo
  - quanti dettagli dare