

La retorica del dato sempre in agguato

“Acqua, acqua dovunque e neppure una goccia da bere!”

(Samuel Taylor Coleridge, La ballata del vecchio marinaio)

“Se tortureremo a sufficienza i dati, confesseranno tutto”

(Ronald Coase, premio Nobel per l'economia nel 1991)

“La percezione è una fantasia che coincide con la realtà”

(Chris Firth, Making Up the Mind: How the Brain Creates Our Mental World)

“È più difficile notare l'assenza di qualche cosa che la sua presenza”

(Martin Lindstrom, Small data. I piccoli indizi che svelano i grandi trend, 2016)

Articolo di

Andrea Granelli

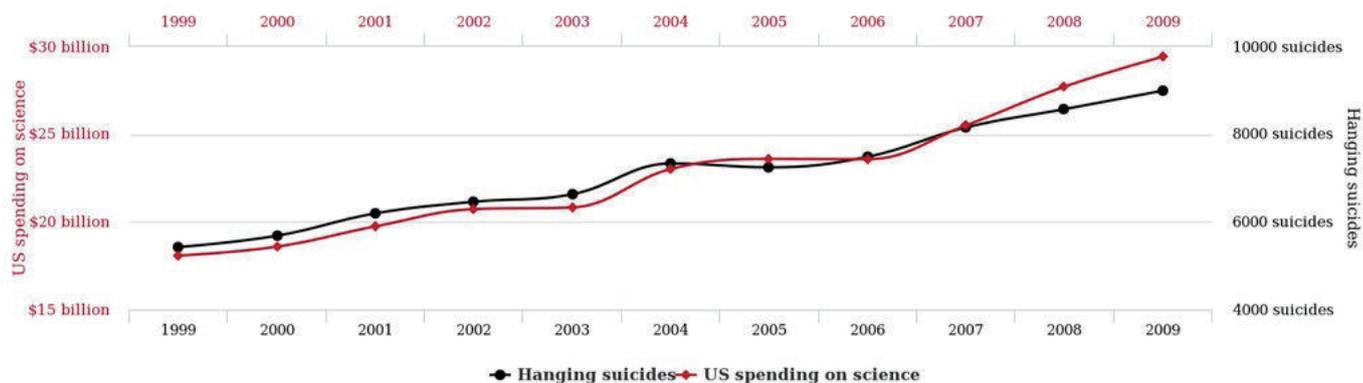
L'analfabetismo matematico e statistico dilaga e - sempre più frequentemente - prendiamo decisioni sbagliate o ci spaventiamo per dei rischi che non hanno quella rilevanza statistica che tendiamo ad attribuire loro. Il disastro dei mutui spazzatura, la trasformazione delle nostre abitudini alimentari a valle della “sindrome” della mucca pazza, ma anche l'ossessione per una continua e sempre più invasiva diagnosi preventiva o i comportamenti irrazionali a fronte del rischio-terrorismo sono il tipico prodotto di questa carenza educativa.

A complicare il quadro si aggiungono alcuni “meccanismi” del nostro cervello che privilegiano specifiche interpreta-

zioni dei dati e un intero armamentario di strumenti retorici, applicati non tanto alle parole quanto ai numeri, particolarmente subdoli in applicati a entità (i numeri appunto) considerati per loro natura oggettivi, rappresentazione veritiera di fenomeni. **Gerd Gigerenzer**, psicologo esperto di processi decisionali e direttore del Center for Adaptive Behavior and Cognition al Max Planck di Berlino, sostiene nel suo libro *“Imparare a rischiare. Come prendere decisioni giuste”* che spesso gli esperti sbagliano a illustrare la probabilità di un fenomeno, e i comportamenti irrazionali della gente (come nel caso di alcuni farmaci contraccettivi, o durante il picco mediatico della “mucca pazza”) dipendono da errori comunicativi (talvolta, ahimé, voluti),

dove si nasconde la classe di riferimento o si tende a comunicare l'aumento del rischio relativo anche quando l'aumento del rischio assoluto è irrilevante: passare dalla probabilità di 1 su un milione a quella di 2 su un milione è irrilevante, ma raddoppia il rischio relativo. Nel caso del terrorismo, invece, sono proprio i terroristi che decidono di comunicare per suggestionarci. Facendo leva sul funzionamento del nostro cervello o, meglio, della sua componente più antica. Tutti ricordiamo l'11 settembre 2001. Subito dopo si scatenò la paura di volare e l'aumento dei viaggi in auto ebbe conseguenze gravi. Prima dell'attacco il numero degli incidenti di macchina mortali era assai vicino alla media dei cinque anni precedenti; viceversa, que-

US spending on science, space, and technology correlates with Suicides by hanging, strangulation and suffocation



tylervigen.com

sto numero restò sopra la media in ciascuno dei dodici mesi successivi all'11 settembre e superò anche, quasi sempre, tutti i dati dei cinque anni precedenti. Si stima che complessivamente circa 1.600 americani abbiano perso la vita sulle strade per avere deciso di evitare il rischio di volare. **Osama bin Laden** commentò soddisfatto l'attentato alle torri gemelle spiegando di avere usato pochissimo denaro per produrre un altro enorme danno: "Al-Qaeda ci ha speso 500.000 dollari, mentre l'America ne ha persi, secondo la stima più bassa, più di 500 miliardi, fra l'incidente e le sue conseguenze; questo significa che ogni dollaro di Al-Qaeda ne ha sconfitto un milione" (*Daved Gartenstein Ross, "Bin Laden's War of a thousand cuts will live on", The Atlantic, 3 maggio 2011*). Perché questo? Ciò dipende dal funzionamento del nostro cervello, in particolare della sua parte più antica, formatasi quando l'uomo viveva in piccole bande di cacciatori-raccoglitori. Quando sono in parecchi a morire in modo spettacolare e tutti insieme, come l'11 settembre, il nostro cervello reagisce con l'ansia poiché è a rischio la sopravvivenza stessa della comunità: ma quando sono altrettanti, o ancora di più, a morire in un modo distribuito nel tempo, come negli incidenti di macchina o di moto, è meno probabile che ci venga paura. Nei soli Stati Uniti muoiono sulle strade circa 35.000 persone all'anno; eppure, pochi hanno paura di prendere l'auto.

I terroristi, dunque, tendono a sfruttare gli eventi a bassa probabilità in cui improvvisamente restano uccise molte persone - i cosiddetti "rischi terrificanti" - poiché essi fanno scattare un principio psicologico inconscio: "se muoiono tutte insieme molte persone, reagisci con la paura ed evita quella situazione". Non si tratta di paura della morte in quanto tale, ma di morire in un determinato modo: tutti insieme nello stesso istante, o a breve tempo. Ma i rischi reali sono radicalmente più bassi.

Disegnare l'informazione

Nancy Duarte cita, nel suo *"HBR Guide to Persuasive presentation"*, due comportamenti del nostro cervello da tenere presente. Innanzitutto l'uso di grafici 3D senza la presenza di un terzo asse dei dati rende i numeri più grandi (e quindi più rilevanti di quanto non lo siano). Nelle pie chart 3D, ad esempio, la "fetta" in primo piano sembra ingannevolmente più grande di una fetta di analoghe proporzioni in secondo piano. In secondo luogo la tecnica di alterare le proporzioni degli assi può rendere le variazioni rappresentate più o meno significative: ad esempio i picchi di un grafico perdono incisività se si dilata l'asse orizzontale del grafico. L'information designer **Alberto Cairo** ci ricorda, inoltre, che il cervello umano non è bravo a calcolare le aree delle superfici; è molto più bravo a confrontare dimensioni singole come la lunghezza o

l'altezza. Oltretutto il cervello è anche una macchina estremamente pigra. Quando si trova a doversi chiedere se l'orso che corre verso di noi è grande abbastanza da rappresentare una minaccia, non perde tempo o energia ad analizzare se è alto e grosso. Vedere soltanto se è alto è già sufficiente. Pertanto i diagrammi a torta possono essere fuorvianti; fanno infatti sottovalutare le differenze. Anche i colori possono giocare brutti scherzi. In particolare i colori puri non sono comuni in natura, e quindi il cervello viene attratto da loro. Il modo migliore per disorientare il lettore di un grafico è quindi di farcirlo di oggetti colorati con tonalità pure. Ma le tecniche più insidiose per ingannare la nostra lettura dei numeri sono di tipo retorico: esiste infatti un insieme di fallacie che spesso vengono usate per ingannare "torturando a sufficienza di dati", come ci ricorda il premio Nobel **Ronald Coase**. Una "fallacia" è un argomento che sembra corretto ma non lo è realmente. È un errore argomentativo nascosto, di solito costruito ad arte per convincere un interlocutore. La fallacia non è dunque propriamente un ragionamento, dal momento che si basa su inferenze invalide; ma non per questo è poco efficace nel convincere. La sua forza sta nell'essere un meccanismo "quasi-corretto", una mezza verità che inganna perché vi ritroviamo frammenti di verità. La tradizione ha tentato ripetutamente di identificare le forme

generali di questi errori ingannevoli: errori che non sembrano errori. Un recente libro sull'argomento (*Franca D'Agostini, "Verità avvelenata. Buoni e cattivi argomenti nel dibattito pubblico"*) ne ha individuate 112 tipologie!

Le fallacie non si limitano a operare nell'ambito della parola ma sono molto efficaci e presenti anche nel mondo dei numeri, dove sono particolarmente insidiose per la pretesa di realtà (e quindi di verità) che il numero può vantare rispetto alla semplice parola pronunciata. Cinque sono le fallacie classiche, fallacie ahinoi usate con grande frequenza, che contribuiscono a suggerire una specifica (e spesso fuorviante) interpretazione di una serie di dati:

1) Ignorare quei dati che disturbano il significato:

spesso vengono etichettati come errori di misura e tolti dalla serie. Questa fallacia viene spesso usata per risolvere il cosiddetto "problema dei fatti": la presenza di alcuni dati scomodi che rischiano di invalidare una specifica teoria.

2) Manipolare i dati aggiungendo alla serie considerata ulteriori dati costruiti ad hoc in grado di confermare quanto si vuole dimostrare. I fautori di questa "regolazione" dell'osservazione (che viene motivata dal fatto che la stessa rilevazione dei dati può essere fallace e incorrere in errori di misura) citano spesso una frase di Albert Einstein: "Se i fatti e la teoria non concordano, cambia i fatti". Peccato che l'attribuzione non sia affatto dimostrata, come si accorse a sue spese Ivanka Trump quando, nel giugno 2013, la citò in un suo tweet...

3) Definire in modo vago il fenomeno che si vuole misurare. È il classico ambito del discorso politico che promette azioni future dando indicatori di impatto ingannevoli. Un bel articolo di Tim Harford sul Financial Times ("Where the Truth Lies with Statistics", 12 maggio 2017) ne analizza diversi: "aumenteremo le



tasse dei ricchi", "accetteremo solo gli immigrati con elevate professionalità", "taglieremo le spese inutili".

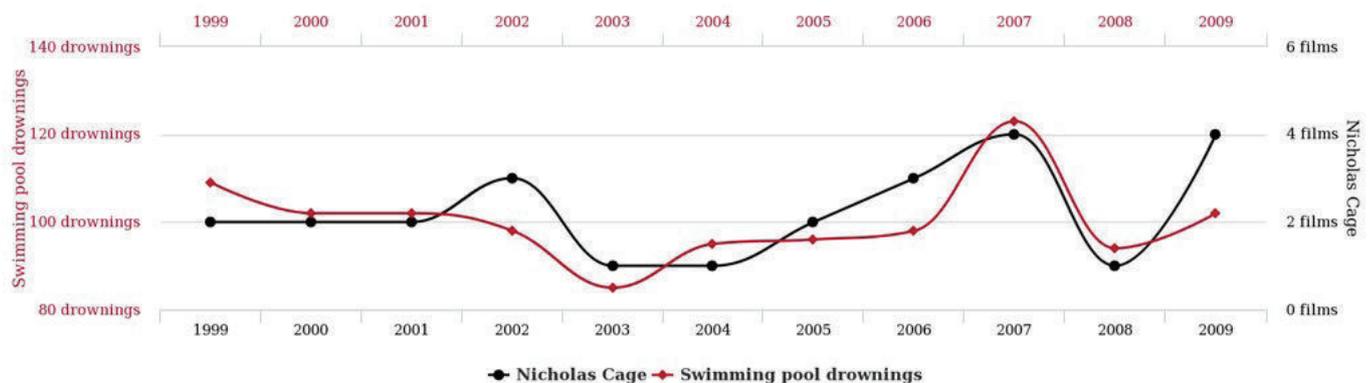
4) Estrapolare il significato dei dati a un contesto su cui non si applicano. Nella retorica classica viene chiamata "fallacia induttiva". La serie di numeri mette in luce una proprietà che vale per un contesto ma il cui risultato viene applicato a un altro contesto, simile ma non uguale (magari con diversi fattori di contor-

no, differenti condizioni iniziali, ...). È il rischio di passare da una condizione particolare (es. vale per la città di Milano) a una generale (es. vale per tutte le città italiane).

5) Usare i dati come "fumo negli occhi": i dati si presentano per giustificare un'asserzione non sono realmente rilevanti, ossia il legame concettuale e logico tra i dati in premessa e la conclusione è solo apparente. Nella retorica classica viene chiamata

DARE I NUMERI

Number of people who drowned by falling into a pool correlates with Films Nicolas Cage appeared in



tylervigen.com

“fallacia di rilevanza”. Spesso questa tecnica viene anche usata come falsa pista, come “distrazione”, per deviare l’attenzione su altri temi (magari lontani dall’oggetto della discussione). È la classica situazione delle performance review dove un manager, nonostante i suoi risultati siano al di sotto delle attese, presenta numeri e tabelle per dimostrare che non è poi andata così male, visto il contesto peggiore di quanto si pensasse, e che egli ha fatto il massimo possibile e quindi i cattivi risultati (ma sono poi così cattivi vista la situazione? ...) non sono proprio colpa sua.

Fallacia, correlazione e causalità

Nel suo “Il cigno nero” **Nassim Nicholas Taleb** fa un passaggio interessante su una tendenza naturale dell’uomo che chiama “fallacia narrativa” (e che viene anche detta correlazione illusoria): la tendenza cioè a costruire senza motivo storie intorno a fatti. Una volta che entriamo nella storia e che ci cattura, perdiamo il rapporto con il fatto scatenante che diventa secondario e può anche “cadere in disgrazia” ed essere smentito dalla stessa storia. Il processo alla base di questo fenomeno è detto di categorizzazione ed è la modalità che gli individui adottano per ordinare e semplificare la realtà, raggruppando persone, oggetti ed eventi in categorie ed elementi di senso, in base alla loro somiglianza rispetto

alle loro azioni, intenzioni e atteggiamenti. Questo meccanismo è lo stesso che crea gli stereotipi sociali. Un (oramai ex) studente della Harvard Law School, **Tyler Vigen**, ha addirittura creato un sito su questo tema dal nome “*Spurious Correlations*” (www.tylervigen.com). Tra le correlazioni spurie, cioè dati correlati senza nessun legame causale, la più famosa è rappresentata dal grafico che rappresenta l’andamento dell’“US spending on science, space, and technology” con “Suicides by hanging, strangulation and suffocation”. Molto interessante è anche la correlazione spuria tra il tasso di divorzi nel Maine e il consumo di margarina. Vigen ha creato il suo sito per mettere in luce “a fun way to look at correlations and to think about data”. Il problema è però che molti invece ci credono davvero e tendono a considerare la correlazione come indicatore assoluto di causalità. La base teorica di questo assunto è un famoso articolo del 2008 di **Chris Anderson**, allora direttore di rivista *Wired* (“*The End of Theory: The Data Deluge Makes the Scientific Model Obsolete*”). Ascoltiamo le sue parole: “Di fronte alla disponibilità di enormi quantità di dati questo approccio, ipotesi, modello teorico e test, diventa obsoleto. [...] La correlazione soppianta la causalità e le scienze possono avanzare addirittura senza modelli teorici coerenti, teorie unificate o una qualche tipo di spiegazione meccanicistica”.

Collegato alla rappresentazione dei numeri si apre anche il capitolo delle immagini, spesso usate per rappresentare i numeri stessi. L’argomento richiede un trattamento a sé ma forse è utile ricordare - per completare questa breve riflessione sulla retorica dei numeri - che le immagini possono essere sia ingannevoli (pensiamo agli innumerevoli esempi prodotti dalla teoria della Gestalt oppure alle anamorfosi create dallo street artist Julian Beever) oppure ambigue per la loro strutturale duplicità. La più famosa è “*My wife and my mother-in-law*” di W.E. Hill, dove coesistono due immagini: quella di una giovane ballerina e quella di una vecchia befana.

Paradossi e logica “fuzzy”

In questa veloce disamina sugli inganni dei numeri non possiamo infine non citare uno dei più famosi paradossi - il “sorite” - che ha da sempre affascinato i filosofi fin dall’antichità. Inventato dal megaride **Eubulide** (lo stesso del paradosso del mentitore) è una fallacia induttiva basata sul concetto di “non più ma non ancora” che, in certi contesti, può diventare un vero e proprio ragionamento paradossale. Un suo famoso enunciato è quello di **Cyrano de Bergerac**: “Parigi è la più bella città del mondo; la mia è la più bella strada di Parigi; la mia casa è la più bella casa della strada; la mia stanza la più bella stanza della casa; io sono l’uomo più bello della

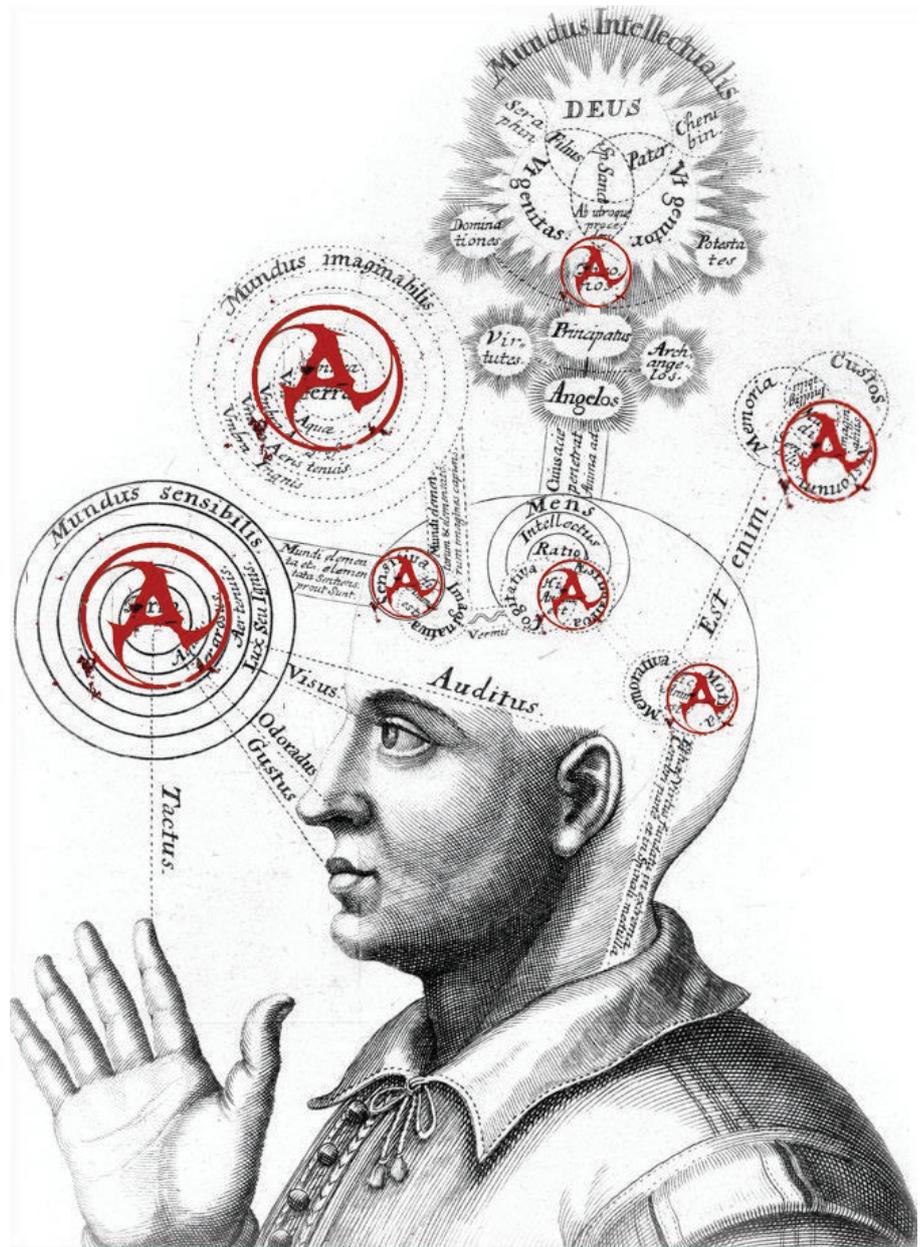
stanza: dunque io sono l'uomo più bello del mondo”.

Una versione più “numerica” è la seguente: “Una persona che ha cinque anni di età è, biologicamente parlando, un bambino. Se qualcuno è un bambino, resterà tale dopo che è trascorso un secondo. In tal caso, trascorso un altro secondo resterà ancora un bambino, come pure un altro secondo dopo, e un altro secondo dopo ancora. Quindi, dopo 630.720.000 secondi resterà ancora un bambino, anche se avrà ormai venticinque anni!”. La versione originale di Eubulide si basava sul ragionamento che aggiungendo un granello di sabbia alla volta, non si riesce mai a formare un mucchio (soros in greco significa infatti mucchio). La domanda sottesa al ragionamento è: “quando un gruppo di granelli di sabbia diventa effettivamente mucchio?”. I contesti in cui il paradosso ha la massima efficacia sono quelli dove vengono usati concetti vaghi, quelli che è difficile definire con esattezza: giovinezza, bellezza, mucchio, ecc.

Una “soluzione” a questo paradosso ci viene dalla logica fuzzy: essere un bambino (o un gruppo di granelli di sabbia) sembra una proprietà che gradualmente si dissolve, lasciando il posto alla comparsa, altrettanto graduale, di un'altra proprietà (essere un adulto o un mucchio di sabbia). Sembra quindi naturale supporre che il valore di verità di “Andrea è un bambino” si trasformi in modo graduale dal vero al falso. La verità, quindi, si presenta per gradi e una sua descrizione puramente numerica rischia di essere fallace.

Che fare?

Il suggerimento più naturale è ovviamente studiare il funzionamento del nostro cervello e imparare a usare (e riconoscere) le tecniche della retorica applicate ai numeri. Più conosciamo le scorciatoie che usa il nostro cervello per capire il mondo, più possiamo prevederle e soprattutto sfruttarle per i nostri scopi. Come osserva **Robert Spence** (“*Information Visualization. Design for Interaction*”), la visualizzazione non è

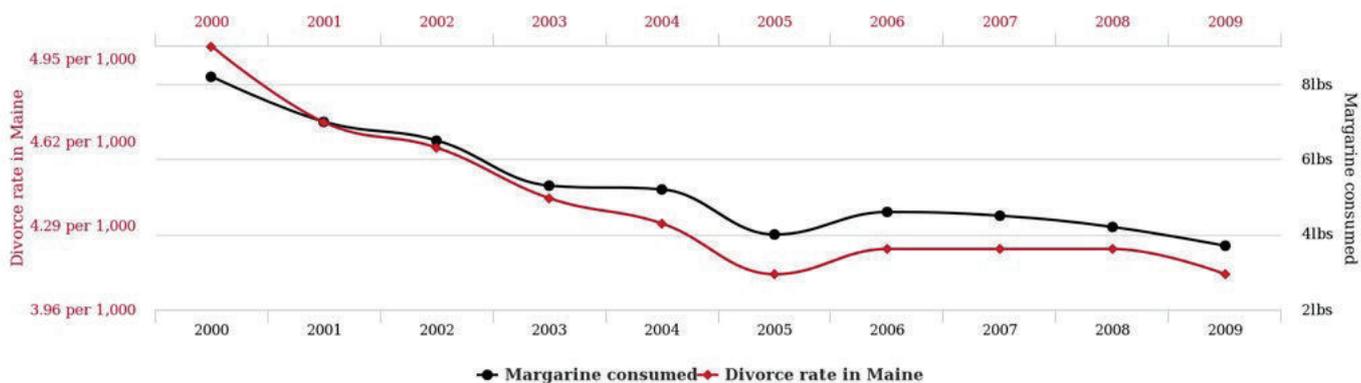


qualcosa che ha luogo su una pagina o su uno schermo: avviene nella mente. La rappresentazione sulla pagina o sullo schermo non è che un sussidio per facilitare questa intuizione. Volendo, però, andare più sul pratico, ci sono diverse cose che si possono fare con maggiore facilità. Innanzitutto incominciare a usare i Big data, e soprattutto i sistemi di business analytics, con grande cautela e nel modo appropriato. Non possono essere l'unica fonte per comprendere i fenomeni. Come ci ricorda **Martin Lindstrom** nel suo “*Small data. I piccoli indizi che svelano i grandi trend*”: “I Big data spiegano le correlazioni ma non

le causalità”. L’abilità dei lettori della realtà è unire diverse fonti di dati (sia qualitative sia quantitative) e soprattutto non staccare mai il senso critico e chiedersi sempre “perché”. In secondo luogo prestare molta attenzione a come comunichiamo i numeri. Come gli antichi retori ci insegnano, l’abilità comunicativa (*l’ars rethorica*) è uno strumento potente, ma va usato con estrema attenzione. Ricordiamo le meravigliose parole di **Gorgia da Lentini**, siciliano del X secolo e considerato il padre della retorica: “La parola è un potente sovrano, poiché con un corpo piccolissimo e del tutto invisibile conduce a com-

DARE I NUMERI

Divorce rate in Maine correlates with Per capita consumption of margarine



tylervigen.com

pimento opere profondamente divine. Infatti essa ha la virtù di troncare la paura, di rimuovere il dolore, d'infondere gioia, d'intensificare la compassione". Per questo motivo il comunicatore deve avere un profondo senso etico. Ha in mano una pistola che, se mal utilizzata, può fare del male. Deve quindi evitare le fallacie e le tecniche manipolative sopra identificate. Ma deve anche usare con abilità ciò di cui dispone. Di seguito alcune indicazioni pratiche: **Edward Tufte**, uno dei più riconosciuti teorici della visualizzazione e dell'information design, ha definito in una famosa intervista sul *Communication Quarterly* (*Mark Zachary, Charlotte Thiralls, An "Interview with Edward E. Tufte"*) il rapporto tra forma e funzione: "Le validità di un grafico analitico presuppone la trasformazione di principi intellettivi in principi visivi. Perciò, se il compito intellettuale è capire la causalità, il principio grafico richiesto sarà: "Mostra la causalità. Se il compito intellettuale è rispondere a una domanda e confrontarla con delle alternative, il principio grafico sarà: "Mostra i paragoni". Il punto è che i grafici analitici non vanno definiti in base alla loro facilità d'uso per l'utente, o necessariamente alla loro leggibilità o a quello che ne pensano gli psicologi o gli arredatori: piuttosto, le loro strutture devono essere definite basandosi su come la struttura coadiuva la riflessione analitica sull'evidenza".

Efficienza nel disegno dei dati

Sempre Tufte suggerisce un vero e proprio principio di efficienza nell'illustrare le informazioni, usando solo ciò che è strettamente necessario. Lo chiama "rapporto dati-inchiostro" poiché rappresenta una sorta di misura della quantità di inchiostro utilizzata per rappresentare i dati in un grafico. Tufte definisce infatti come elementi dati-inchiostro quelli che non possono essere eliminati senza distruggere l'integrità della presentazione. Tutto il resto può essere eliminato in quanto mera decorazione che tende a distogliere il lettore da ciò che è veramente importante. Infatti, come ci ricorda **Alberto Cairo** nel suo *"L'arte funzionale"*, una buona rappresentazione grafica soddisfa due obiettivi fondamentali. Innanzitutto presenta le informazioni: le infografiche non devono necessariamente semplificare il messaggio, devono soprattutto chiarirlo. Poi permette agli utenti di esplorare le informazioni, "entrando" nei numeri e scoprendone i nessi meno apparenti. Infine non eccedere mai nei dati da mostrare. Troppi dati confondono. Questo vale sia per il numero di dati in una tavola sia soprattutto per il numero di tavole numeriche di una presentazione. Spesso la quantità di tavole numeriche è proporzionale al grado di insicurezza di chi presenta. Se il fenomeno è articolato e complesso da descrivere, si possono costruire allegati pieni di slides

da utilizzare solo se la discussione lo richiede. L'approccio "esaustivo" che vuole dimostrare il controllo del fenomeno che si sta discutendo presentando moltissime tavole su ciascuna dimensione o aspetto del fenomeno è controproducente. Su questo tema del "troppo stroppia" (che un'altra dizione popolare declina in "il troppo è troppo") **Antonio Spadaro** aggiunge un'ulteriore prospettiva ricordando, in un articolo di *Wired* (11 marzo 2015), una riflessione di Papa Francesco sui rischi dell'informazione eccessiva. Rispondendo a uno studente che gli chiedeva, domanda strana da fare a un Papa, "Come faccio a vivere bene al tempo dell'information overload?", **Papa Francesco** risponde evocando un'immagine, quella dell'uomo-museo, che accumula informazioni senza sapere che farsene. E gli contrappone quella di uomo-sapiente, di homo sapiens. Chi accumula dati senza sapere cosa farsene "musealizza" la sua vita e il rischio è sempre più concreto, visto come la rete ci bombarda di risposte. Anzi, la capacità predittiva dei motori di ricerca è talmente straordinaria che otteniamo una risposta prima ancora di formulare la domanda. E allora rischiamo di perdere l'abitudine a fare domande, il gusto per le domande, la capacità di fare domande. Avendo infatti a disposizione tutte le risposte possibili, non ci poniamo più il problema di farci delle domande. E ci resta poco spazio per la sorpresa.