

Mapping e Big Data

Lezione 5

Mario Verdicchio

Università degli Studi di Bergamo

Anno Accademico 2022-2023

Visualizzazione

- Una **visualizzazione** è una rappresentazione visiva dei dati con l'obiettivo di consentire:
 - comunicazione
 - analisi
 - scoperta
 - esplorazione
 - e altre attività legate a dati e informazioni

Tipi di visualizzazione

- Grafici
- Mappe di dati
- Infografica
- Visualizzazione di dati
- Applicazione giornalistica

Grafico

- Un **grafico** è una visualizzazione in cui i dati sono codificati con simboli che hanno forme, colori e proporzioni diversi
- Noto anche come **diagramma**
- Spesso i simboli di un grafico sono inseriti all'interno di un piano cartesiano
- Tipi di grafico senza piano cartesiano sono i grafici a torta

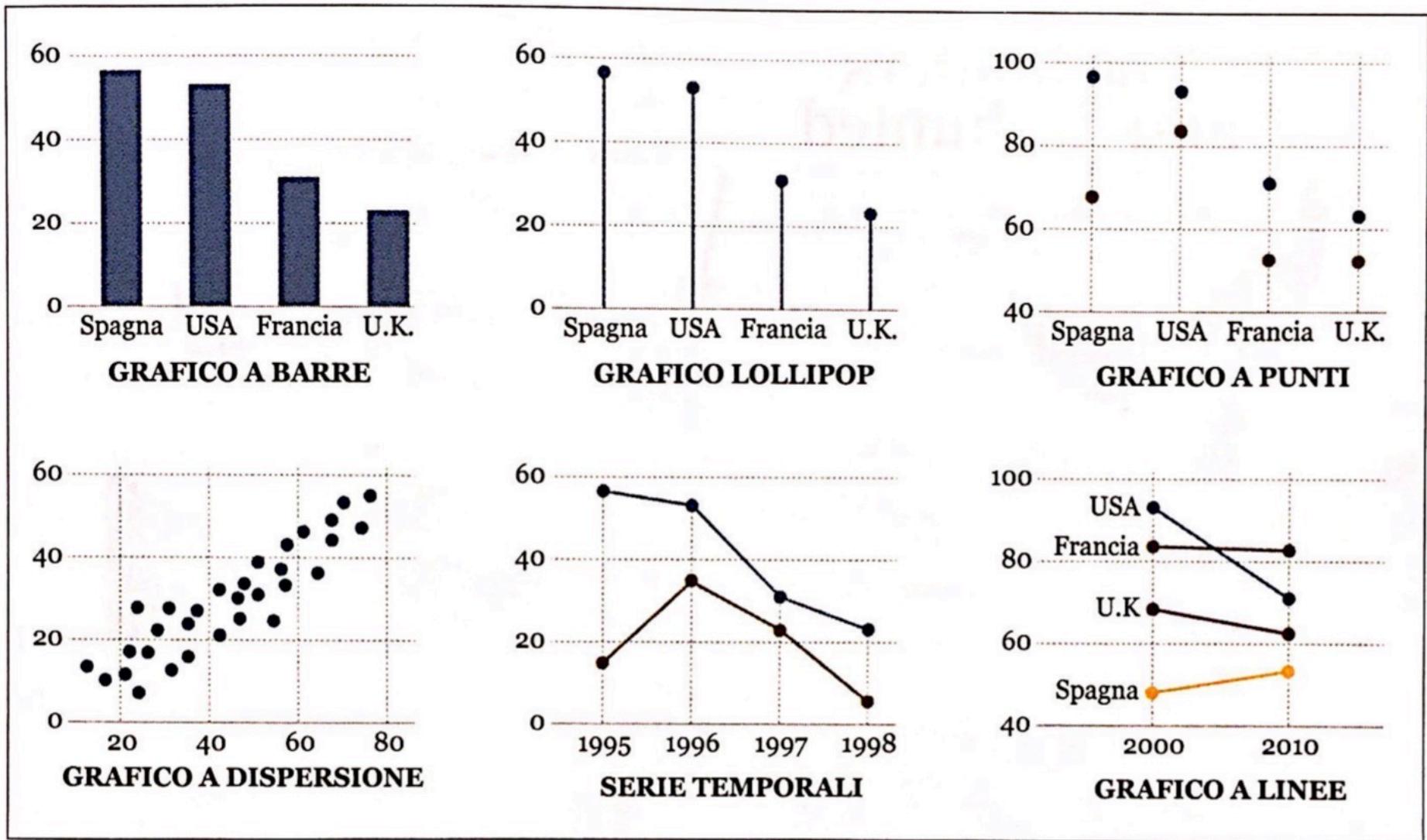


Figura tratta da "L'arte del vero" di A. Cairo

Mappa di dati

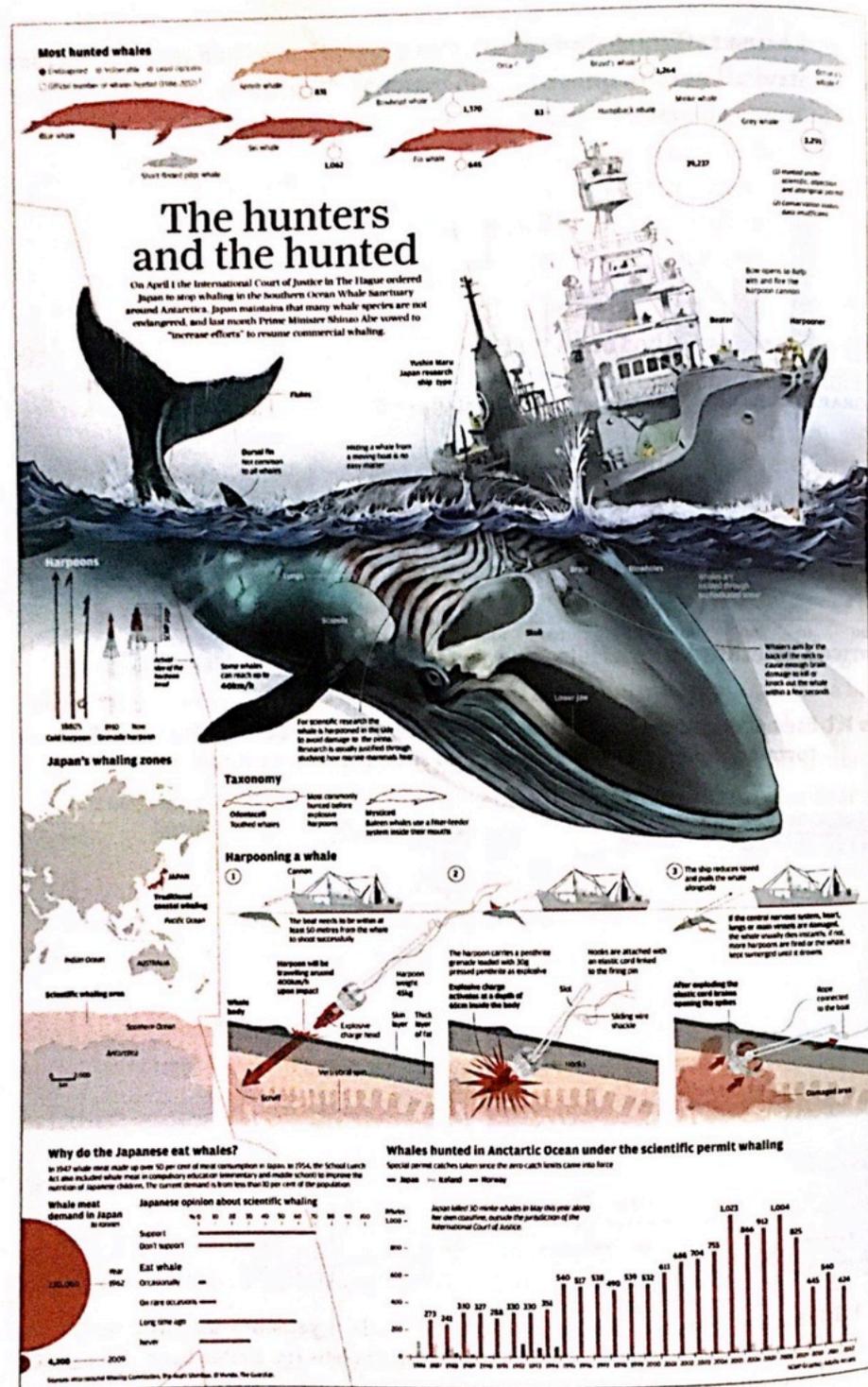
- Una **mappa** è una raffigurazione di un'area geografica
- Una **mappa di dati** è la rappresentazione di dati relativi a un'area geografica
- Una mappa di dati tipicamente include una mappa dell'area a cui si riferisce



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

Infografica

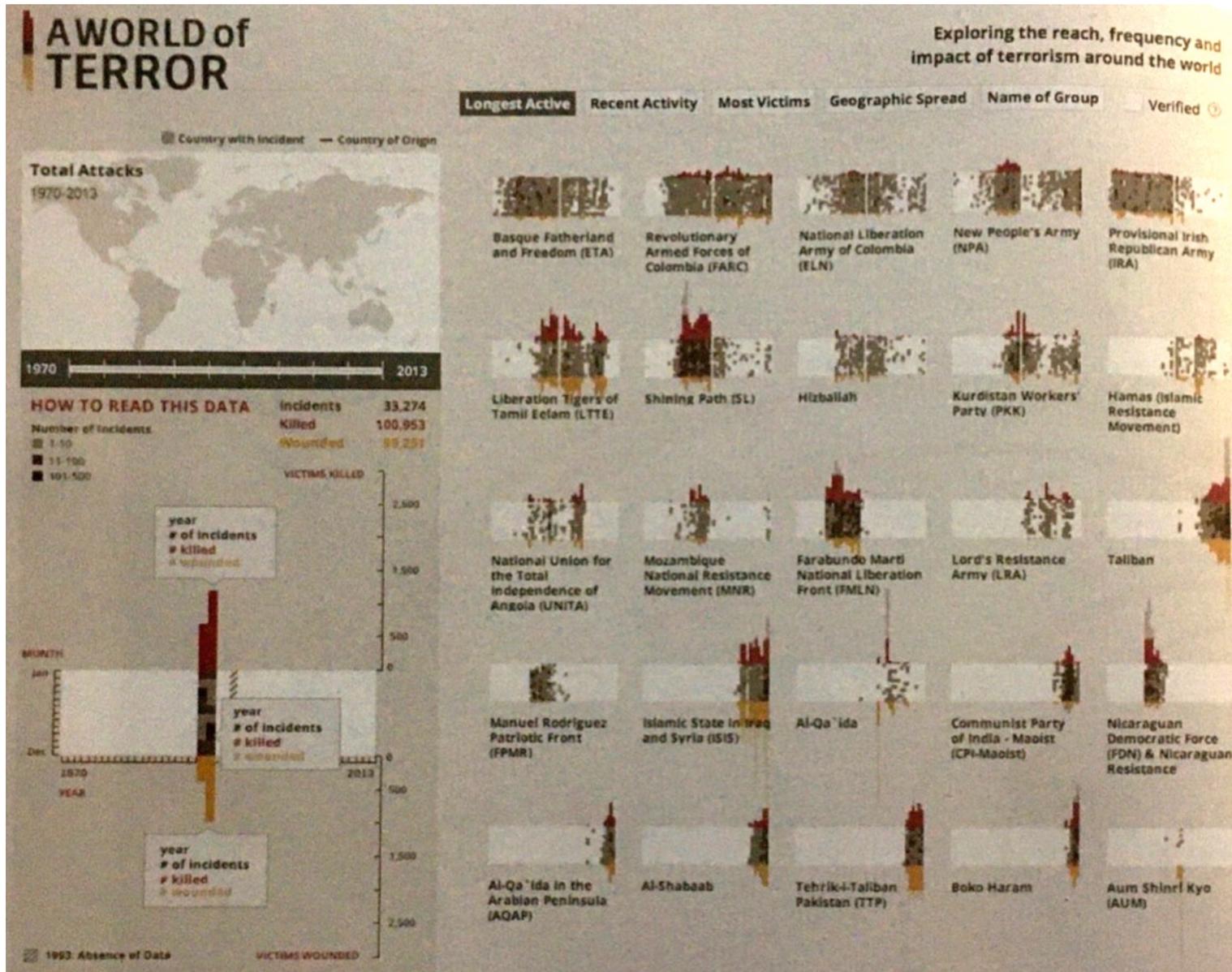
- Un'**infografica** è una rappresentazione visiva in più parti concepite per comunicare uno o più messaggi
- Le parti di un'infografica possono essere:
 - grafici
 - mappe
 - illustrazioni e testi
 - audio
- Queste parti forniscono spiegazioni e contesto
- Il progettista dell'infografica non mostra tutte le informazioni che ha raccolto, ma solo quelle importanti per il concetto (o concetti) che sta cercando di trasmettere
- Le infografiche sono tipicamente colorate e divertenti, ma l'obiettivo fondamentale è informare il pubblico, quindi chiarezza e approfondimento sono più importanti degli abbellimenti grafici



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

Visualizzazione dei dati

- Una **visualizzazione dei dati** è un modo di esporre dati alle persone, per consentire loro di analizzare i dati, esplorarli, e fare delle scoperte
- Tipicamente una visualizzazione dei dati **non** sono intese per trasmettere messaggi dei suoi progettisti, bensì di presentare nella maniera più neutra possibile i dati
- Starà al pubblico trarre le proprie conclusioni



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

Applicazione giornalistica

- “**Applicazione giornalistica**” è un termine che è stato introdotto per la prima volta da ProPublica, un’organizzazione no-profit per il giornalismo investigativo
- Per applicazione giornalistica si intende una visualizzazione che consente alle persone di correlare i dati presentati alle loro vite
- Una caratteristica dell’applicazione giornalistica è la personalizzazione in funzione dei bisogni di ciascun osservatore
- Questa personalizzazione può essere realizzata tramite simulatori, calcolatrici, database online interattivi

PRO PUBLICA Journalism in the Public Interest

Receive our top stories daily
 Email address **SUBSCRIBE**

Home Our Investigations Tools & Data MuckReads Get Involved About Us

Treatment Tracker

The Doctors and Services in Medicare Part B

By Lena Groeger, Charles Ornstein, and Ryann Grochowski Jones, ProPublica, May 15, 2014

Medicare recently released, for the first time, details on 2012 payments to individual doctors and other health professionals serving the 46 million seniors and disabled in its Part B program. Part B covers services as varied as office visits, ambulance mileage, lab tests, and the doctor's fee for open-heart surgery. Use this tool to find and compare providers. | [Related story >](#)

Search for a Provider, City or Zip Code

All States

For example: Rebecca Torres, Chicago, 63141

Mental Health Providers Score Highest in Intensity of Office Visits

Office visits are reimbursed on a scale of 1-5, with 5 being the most intensive and costly. Mental health professionals bill Medicare for level 5 visits most frequently compared to other specialties, with cancer doctors following close behind. Here's the breakdown of office visits by specialty (for patients seen at least once before).

Most intensive & costly

Least intensive & costly

BY LEVEL 5 VISITS
BY TOP SPECIALTIES

The bigger the slice, the higher percentage of office visits at that level. Hover over each slice to see more detail. Arranged by specialties with highest % of level 5 visits.

Neuropsychiatry

Psychiatry

Geriatric Psychiatry

Nuclear Medicine

Hospice and Palliative Care

Neurology

Gynecological/Oncology

Hematology

Medical Oncology

Hematology/Oncology

Cardiac Electrophysiology

Thoracic Surgery

Cardiac Surgery

Infectious Disease

Geriatric Medicine

Cardiac Electrophysiology

Thoracic Surgery

Cardiac Surgery

Infectious Disease

Geriatric Medicine

Medicare Payments by State

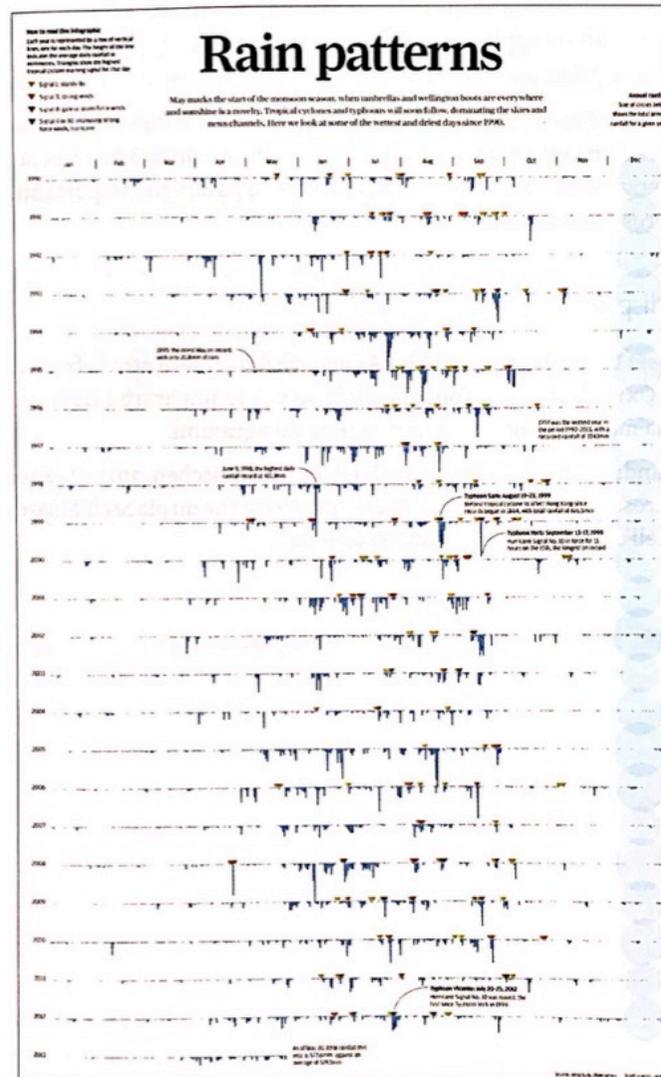
| State | Payments |
|----------------|----------|
| NATIONAL | \$77.3B |
| CALIFORNIA | \$7.86B |
| FLORIDA | \$7.62B |
| TEXAS | \$6.66B |
| NEW YORK | \$6.45B |
| NEW JERSEY | \$3.56B |
| ILLINOIS | \$3.39B |
| PENNSYLVANIA | \$3.13B |
| MICHIGAN | \$2.79B |
| NORTH CAROLINA | \$2.78B |
| OHIO | \$2.47B |
| GEORGIA | \$2.21B |
| VIRGINIA | \$2.03B |
| TENNESSEE | \$1.85B |
| MASSACHUSETTS | \$1.81B |
| MARYLAND | \$1.78B |
| ARIZONA | \$1.59B |
| INDIANA | \$1.53B |
| ALABAMA | \$1.49B |
| MISSOURI | \$1.43B |
| SOUTH CAROLINA | \$1.34B |
| WASHINGTON | \$1.34B |
| KENTUCKY | \$1.14B |
| CONNECTICUT | \$1.02B |
| LOUISIANA | \$1.02B |
| WISCONSIN | \$924M |
| OKLAHOMA | \$810M |
| COLORADO | \$776M |
| ARKANSAS | \$754M |
| KANSAS | \$734M |

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

Visualizzazioni ibride

- Data la complessa e ricca nature delle visualizzazioni, questa tassonomia non può essere rigorosa per tutti gli esempi di visualizzazioni
- Esistono visualizzazioni che hanno caratteristiche di una categoria di visualizzazioni unite a quelle di un'altra categoria

Esempio: infografica + visualizzazione di dati



Visualizzazione di
The South China Morning Post
Figura tratta da
“L’arte del vero” di A. Cairo

Ciò che importa davvero...

- non è una tassonomia rigorosa e definitiva.
- Si può sempre descrivere una visualizzazione difficile da classificare con le caratteristiche tipiche delle diverse categorie a cui si avvicina
- L'importante è che una visualizzazione sia illuminante per chi la osserva
- A questo scopo, i suoi progettisti devono fare riferimento a importanti principi e parametri

Le 5 qualità delle buone visualizzazioni

La mazza da hockey

- Cairo ritiene che una visualizzazione in particolare, resa pubblica in un rapporto del 2001 dall'IPCC (Intergovernmental Panel on Climate Change) riassume in sé le 5 qualità delle visualizzazioni di successo
- Si tratta di un grafico della differenza tra la temperatura media di ciascun anno e la temperatura media del periodo 1961-1990, realizzato da Mann e Bradley (University of Massachusetts) e Hughes (University of Arizona)
- Si chiama “grafico a mazza da hockey” per via della sua forma



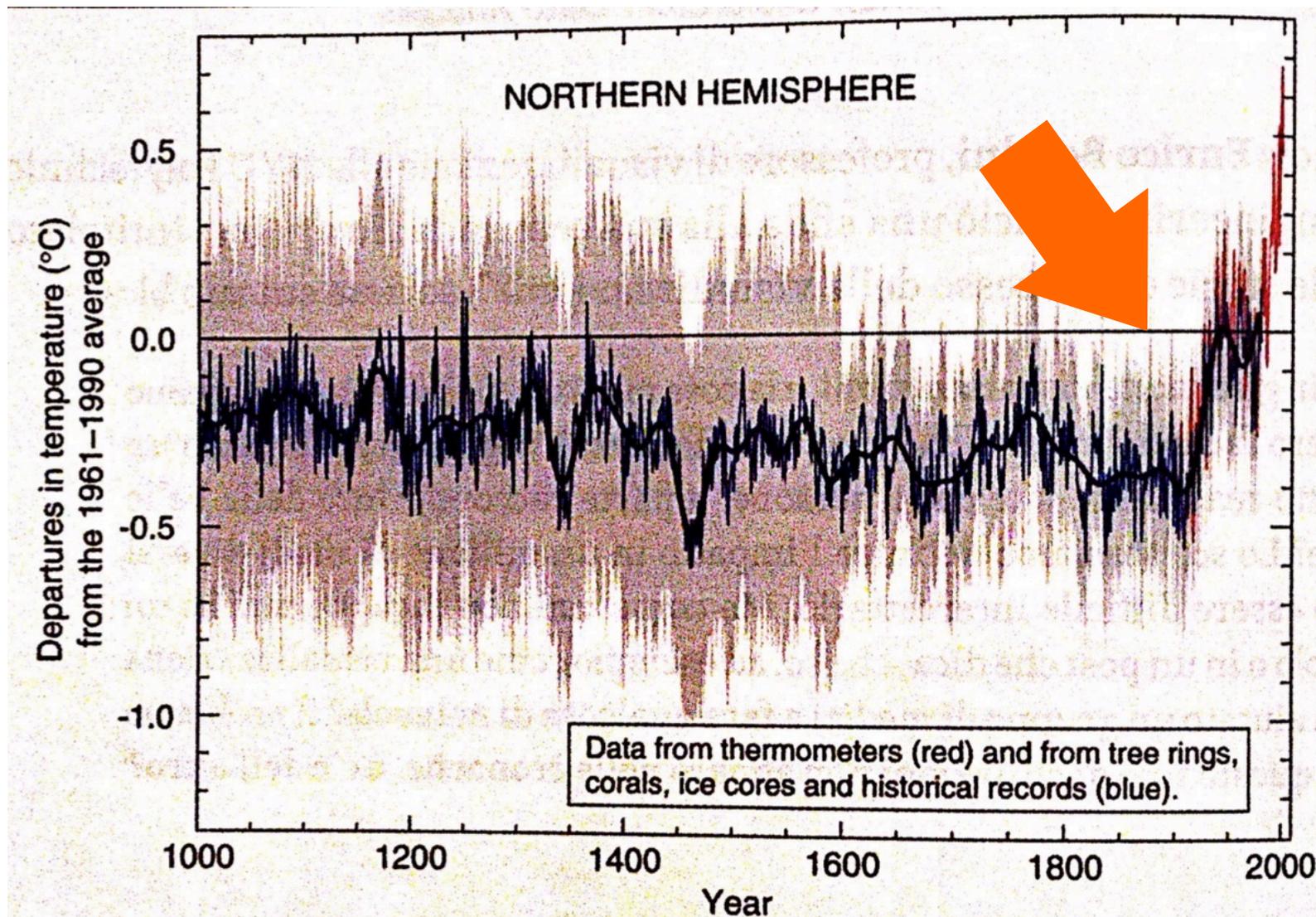


Grafico di Mann, Bradley e Hughes. Figura tratta da "L'arte del vero" di A. Cairo

La mazza da hockey

- I valori delle temperature sono stati misurati con termometri quotidianamente solo dal XVII secolo
- Per avere dati sulle temperature nei periodi precedenti, gli scienziati hanno misurato variabili “proxy” (che sono legate alle variabili a cui siamo interessati con relazioni e funzioni a noi note)
- Le stime statistiche sul passato non sono mai completamente accurate, e l’incertezza è mostrata visivamente con la fascia grigia, all’interno della quale gli scienziati sono sicuri al 100% che il valore reale ricada
- Più ci si avvicina al presente, minore è l’incertezza grazie alle misurazioni sempre più precise e ai dati sempre più disponibili, e quindi la fascia grigia si restringe

Le 5 qualità della mazza da hockey

- Il grafico a mazza da hockey non ha lasciato indifferenti, anzi, è tuttora considerato una delle visualizzazioni più emblematiche e convincenti, grazie alle sue qualità
 1. è **veritiero**, perché basato su ricerche oneste e approfondite
 2. è **funzionale**, perché è una raffigurazione accurata dei dati e consente all'osservatore di eseguire operazioni significative (es. vedere il cambiamento nel corso del tempo)
 3. è **bello**, nel senso di intrigante e esteticamente piacevole
 4. è **profondo**, perché rivela prove che altrimenti sarebbero difficili da vedere
 5. è **illuminante**, perché se accettiamo le prove che raffigura cambierà il nostro modo di pensare

Visualizzazioni **veritiere**

Un caso nelle telecomunicazioni USA

- La NCTA (National Cable & Telecommunications Association) è un'organizzazione commerciale che gestisce le PR e il lobbying per conto delle aziende USA della comunicazione via cavo
- Nel 1992, negli USA è stata introdotta una regolamentazione delle trasmissioni via cavo (Cable Television Consumer Protection and Competition Act)
- Nel 1996, la regolamentazione viene rilassata con il Telecommunication Act
- Sulla home page della NCTA viene pubblicato il seguente grafico, a dimostrare che la deregolamentazione fa bene, perché porta a maggiori investimenti nel settore

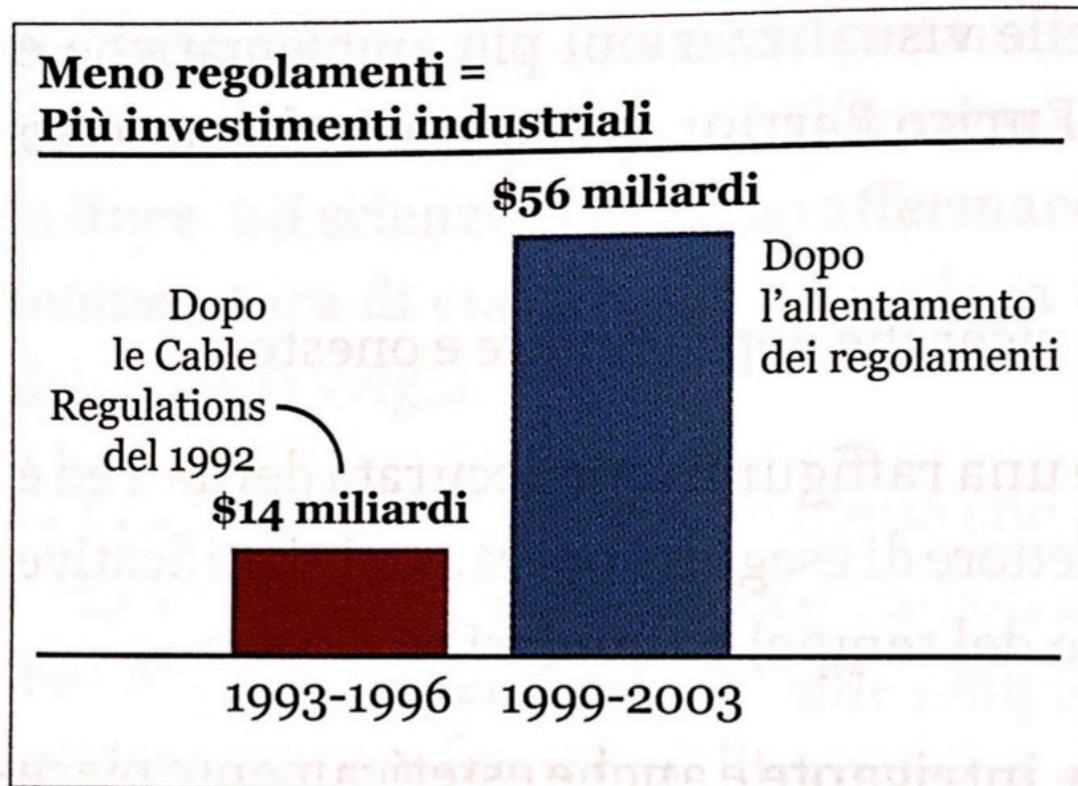
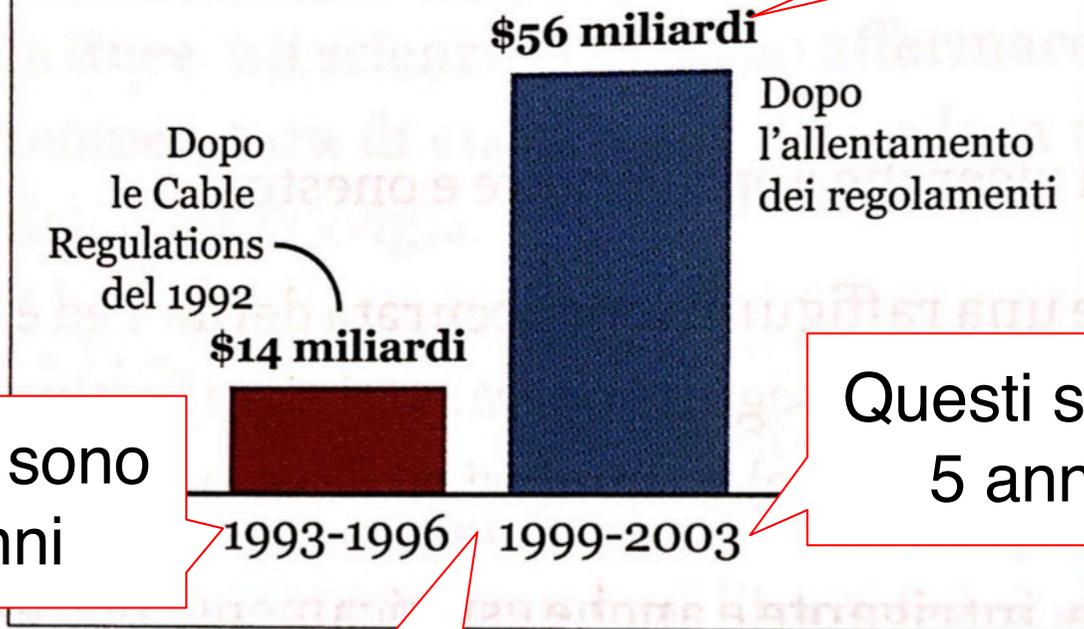


Figura tratta da "L'arte del vero" di A. Cairo

**Meno regolamenti =
Più investimenti industriali**



Tenuto conto dell'inflazione?

Questi sono 4 anni

Questi sono 5 anni

Dove sono 1997 e 1998?

Indagando meglio sul sito della NCTA...

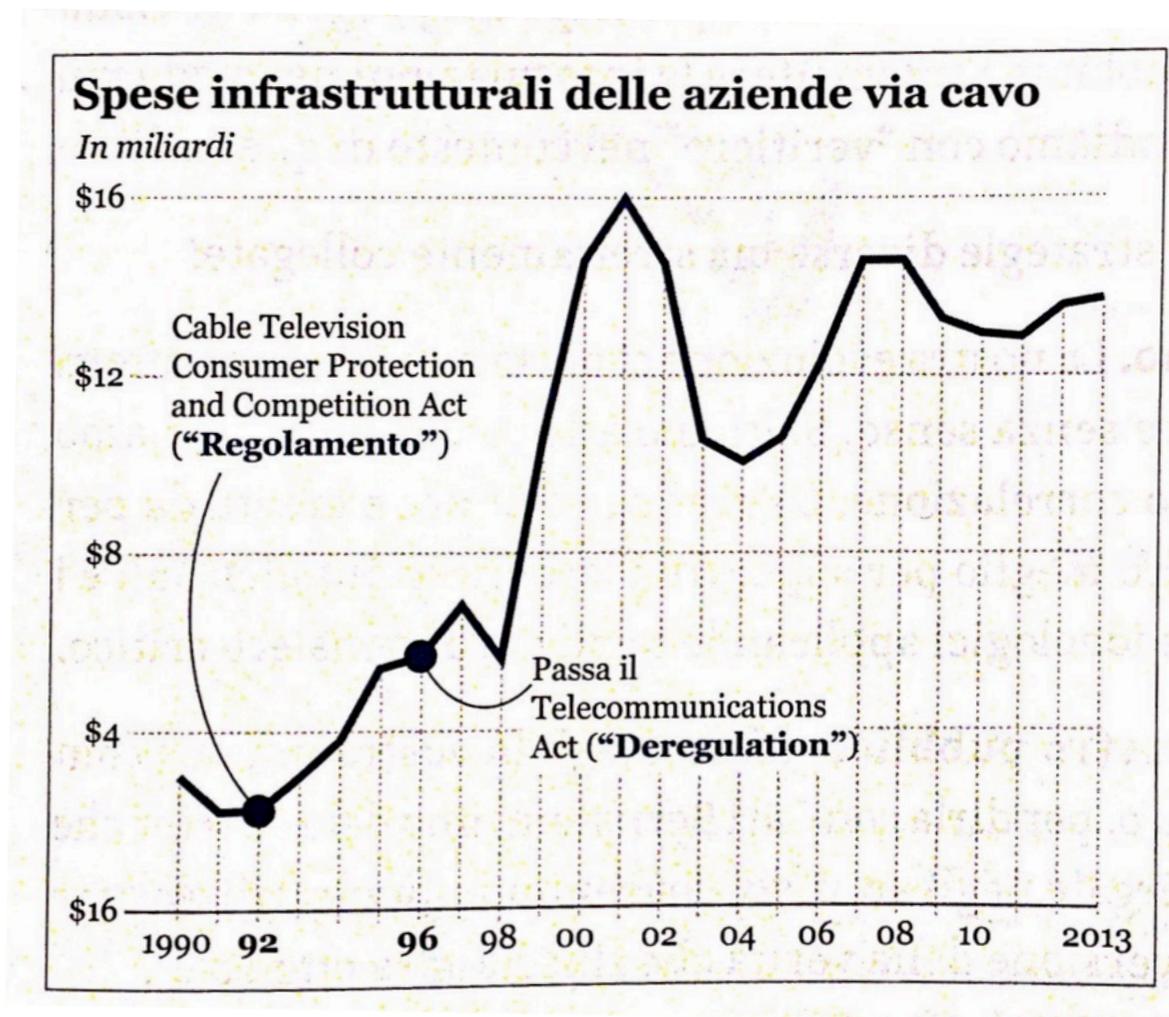
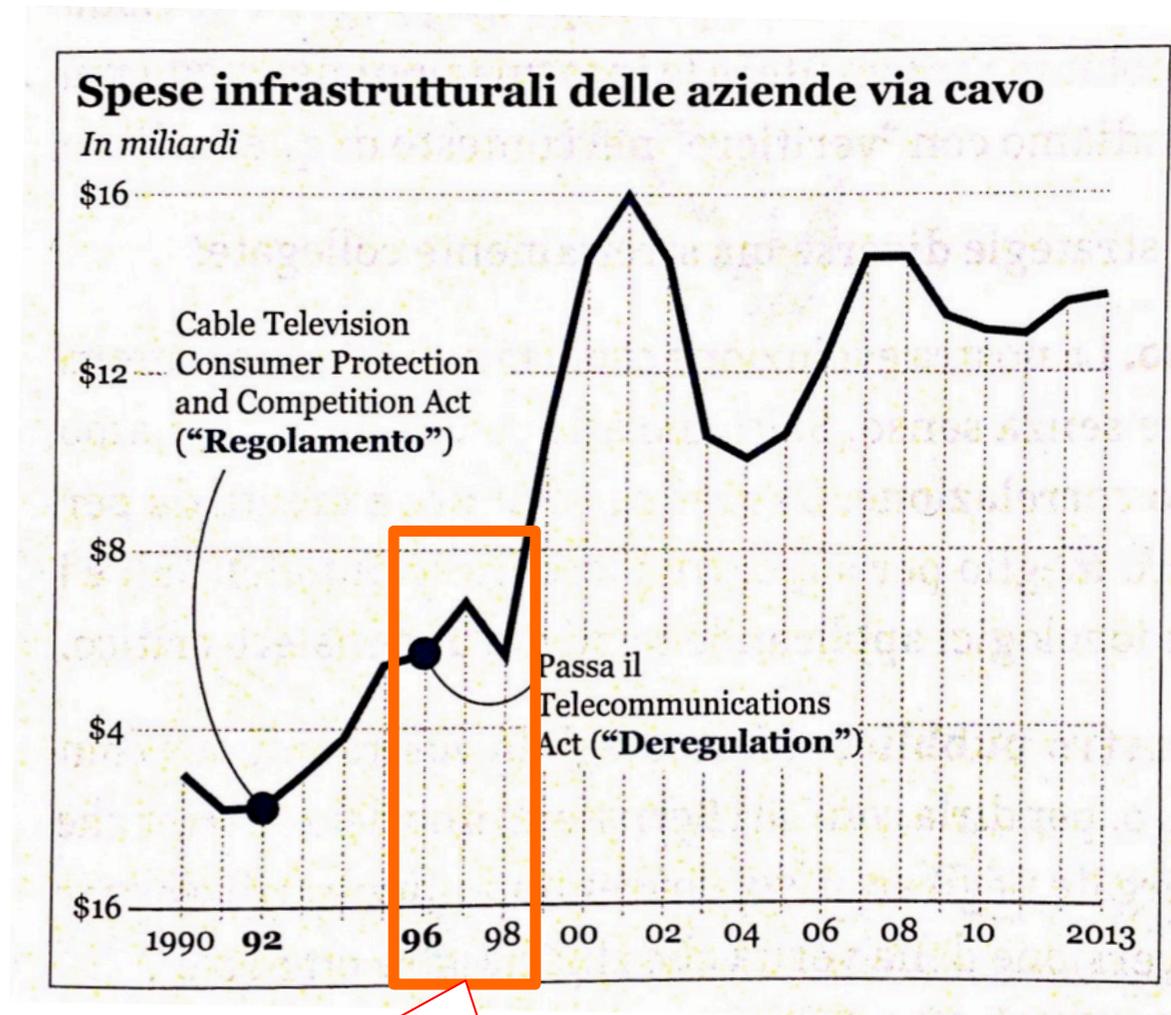


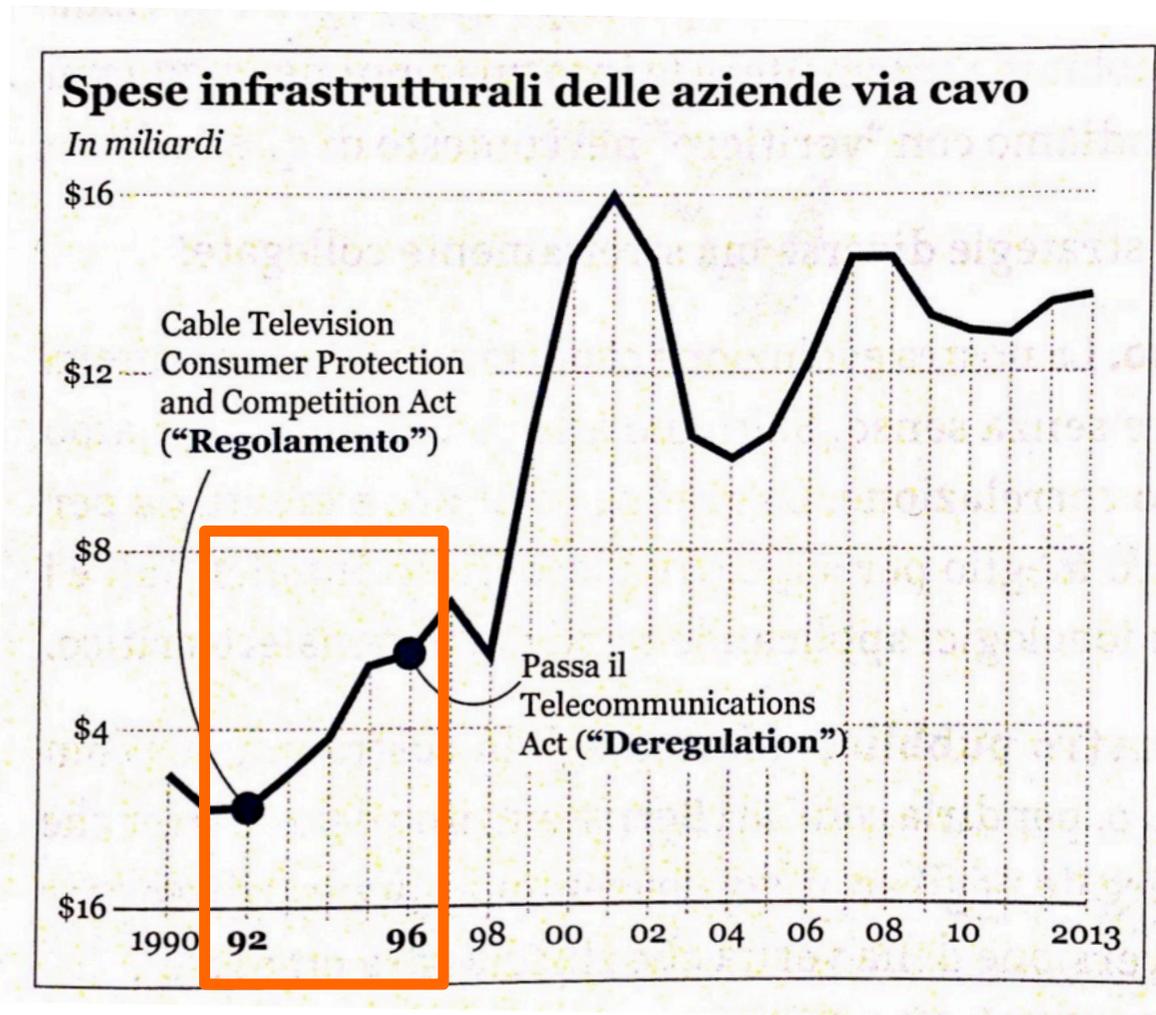
Figura tratta da "L'arte del vero" di A. Cairo

Indagando meglio sul sito di NCTA...



Dove sono 1997 e 1998?

Indagando meglio sul sito della NCTA...



I fatti

- Dopo l'introduzione della regolamentazione, gli investimenti sono continuati a salire
- Dopo il rilassamento della regolamentazione, invece, gli investimenti hanno visto una caduta brusca, poi compensata da un'altra grande crescita

Analisi dei fatti

- A che cosa è dovuta la caduta degli investimenti dopo la deregolamentazione?
 - alla deregolamentazione stessa?
 - alla crisi economica del 1997-98?
- A che cosa è dovuta la crescita dopo tale caduta?
 - alla deregolamentazione?
 - al boom delle dot com e della new economy?
- Non abbiamo una risposta certa, ma siamo certi del fatto che il messaggio del primo grafico non ha prove a suo sostegno

- I fatti scoperti tramite il secondo grafico meritano un'analisi più approfondita
- Non è detto che questa analisi porti a dei risultati che smentiscano del tutto il primo grafico
- Come risultato, però, la fiducia nei confronti della comunicazione di NCTA ne esce sicuramente danneggiata
- Visualizzare dati fuorvianti non solo è moralmente inaccettabile, ma mina anche la credibilità
- Due raccomandazioni:
 - essere onesti con il pubblico
 - evitare l'autoinganno, non saltare alle conclusioni

Altro esempio



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

“I costi dei due prodotti sono fortemente correlati.”

Scala dei valori
del prodotto 1

Attenzione!

Scala dei valori
del prodotto 2



Che cosa succede se mettiamo i costi
dei due prodotti sulla stessa scala?

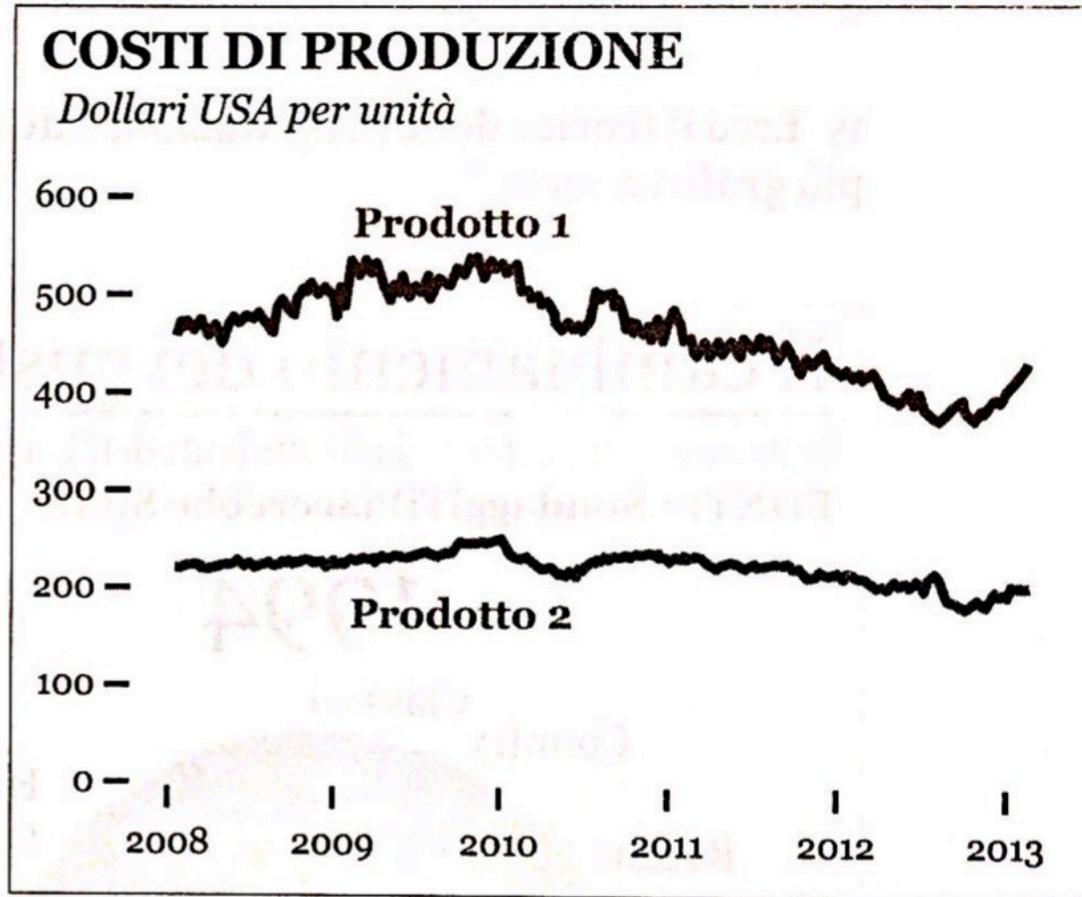


Figura tratta
da "L'arte del
vero" di A.
Cairo

La correlazione scompare, o comunque
diventa molto più debole

Visualizzazioni **funzionali**

Funzionalità

- Il fattore principale nella scelta della forma grafica di una visualizzazione non è il gusto personale, ma il pensiero razionale
- Il pensiero che guida ogni progettazione di un oggetto che deve essere usabile e utile
- Si pensa allo scopo per cui l'oggetto viene creato e si plasma la materia prima per consentire alle persone (gli utenti dell'oggetto) di raggiungere tale scopo

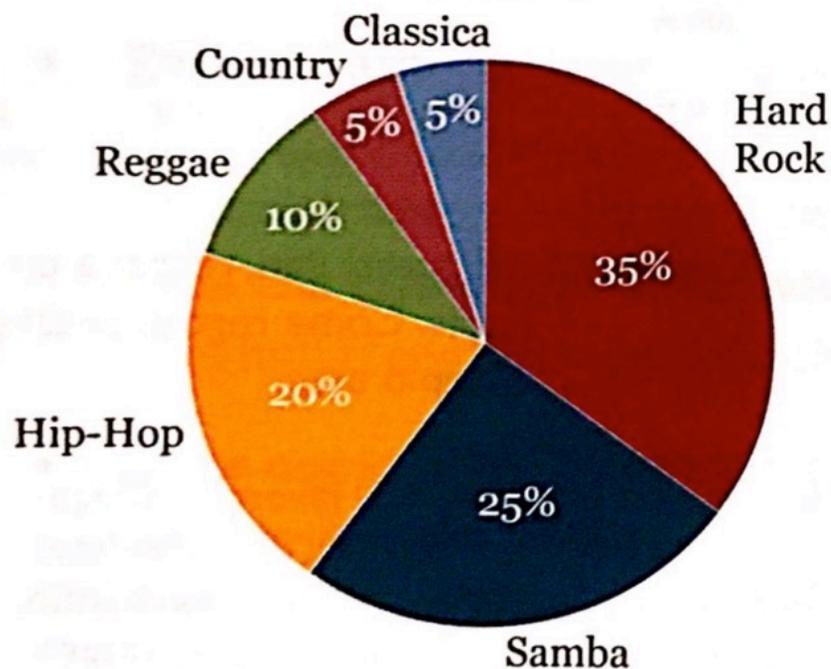
Un esempio: i grafici a torta

Il cambiamento dei gusti musicali in vent'anni

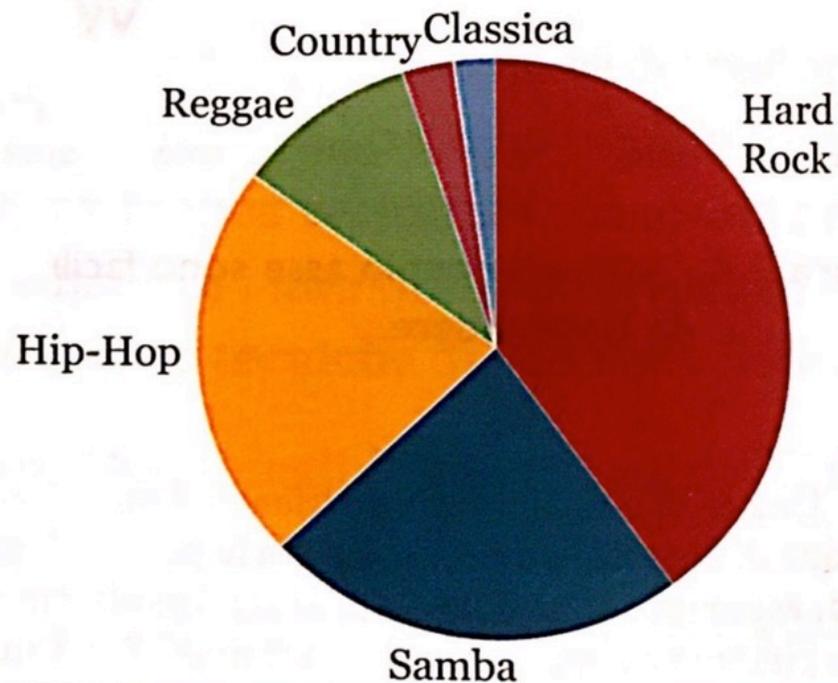
Stili musicali preferiti dagli studenti dell'Università di Miami. Sondaggio condotto su 1.000 studenti.

FONTE: SondaggiTiPiacerebbe SpA.

1994



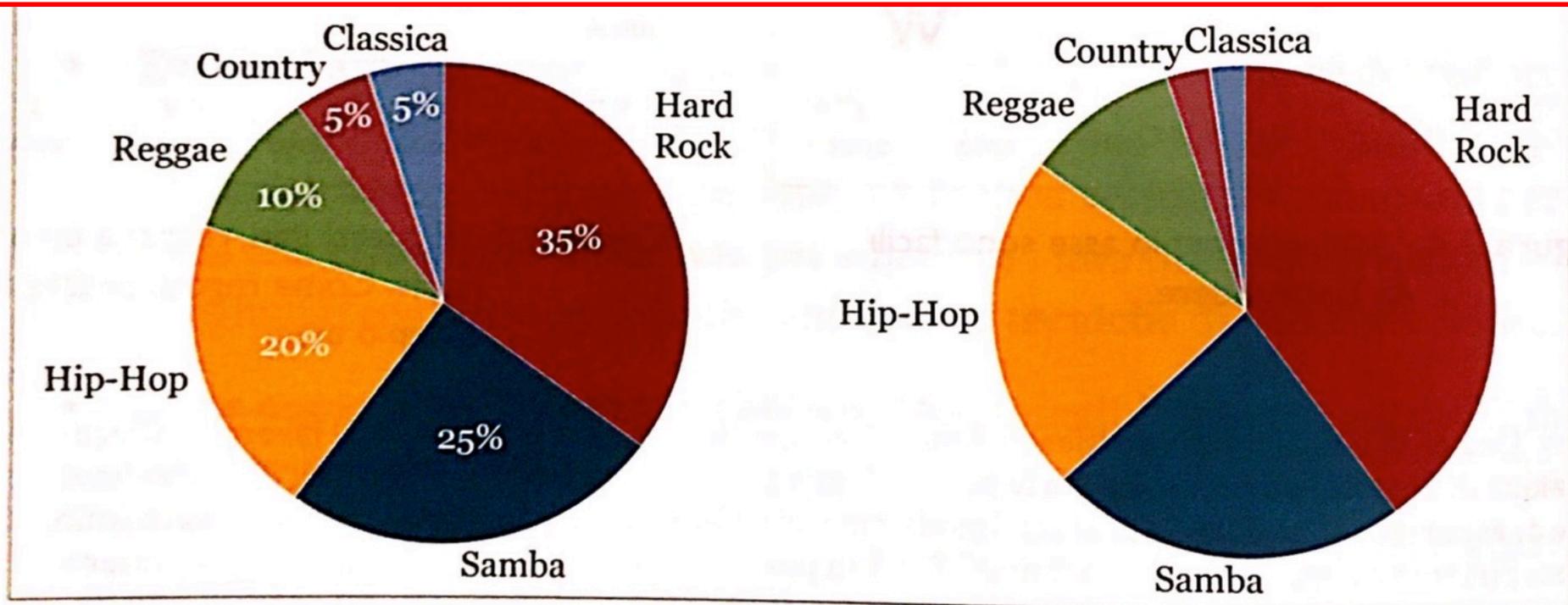
2014



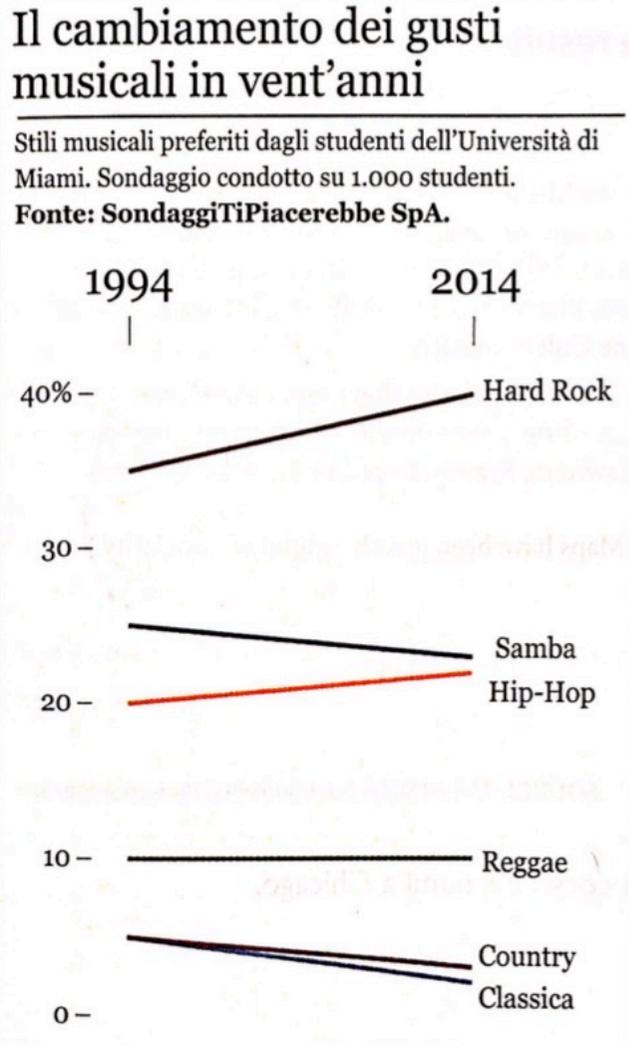
Un esempio negativo

Riesco a vedere che hard rock è cresciuto, perché la fetta inizia a “mezzogiorno” in entrambe le torte, e si nota che finisce più “tardi” nel 2014. Ma la samba? La sua fetta si è spostata e non riesco a capire se è rimasta uguale, se è lievemente cresciuta o lievemente diminuita. Senza le cifre percentuali nel 2014 sono perso.

Domanda: ma se mi serve leggere le cifre, a che cosa serve la visualizzazione?
Questa è una visualizzazione non è per niente **funzionale**.



Un'alternativa funzionale



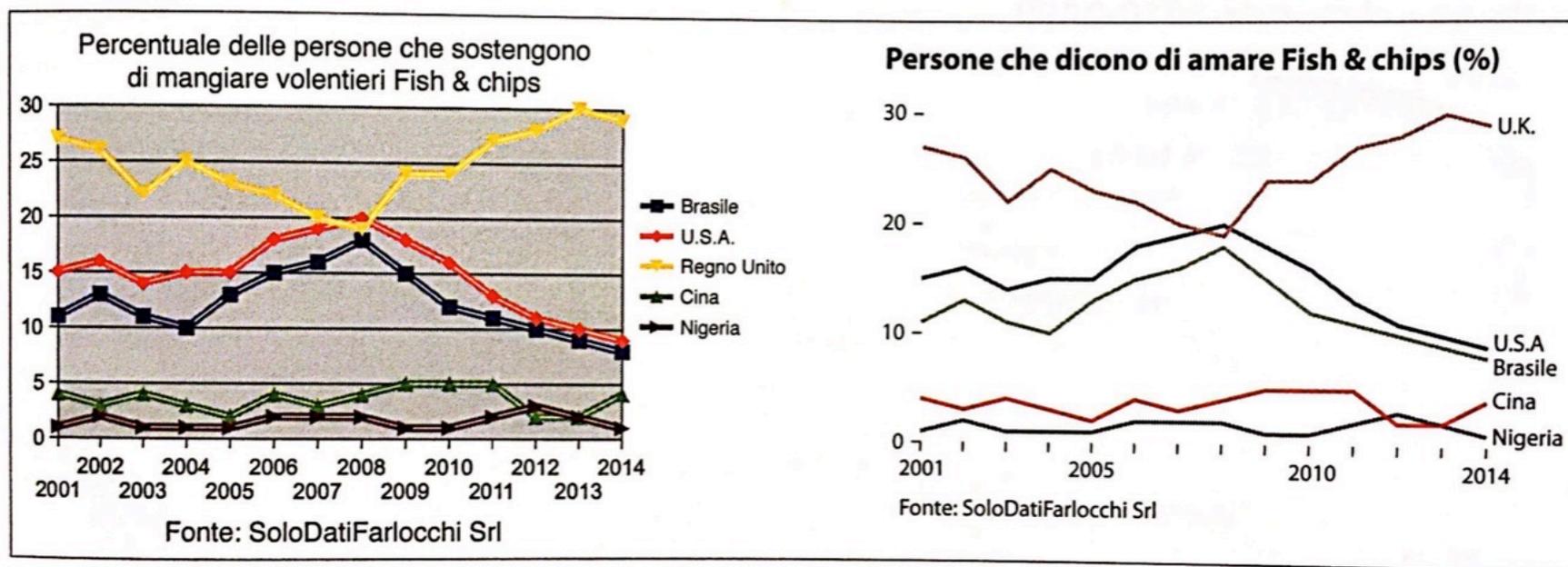
Questo è un **grafico angolare**, o **“slope chart”** in inglese. Il nome in inglese rende meglio l'idea di base del grafico: “slope” vuol dire pendenza, pendio.

Proprio questi pendii ci aiutano a vedere istantaneamente gli andamenti dei generi musicali e le differenze tra 1994 e 2014. Le percentuali quasi non ci interessano, perché gli andamenti sono già visualizzati benissimo senza bisogno di parlare di cifre.

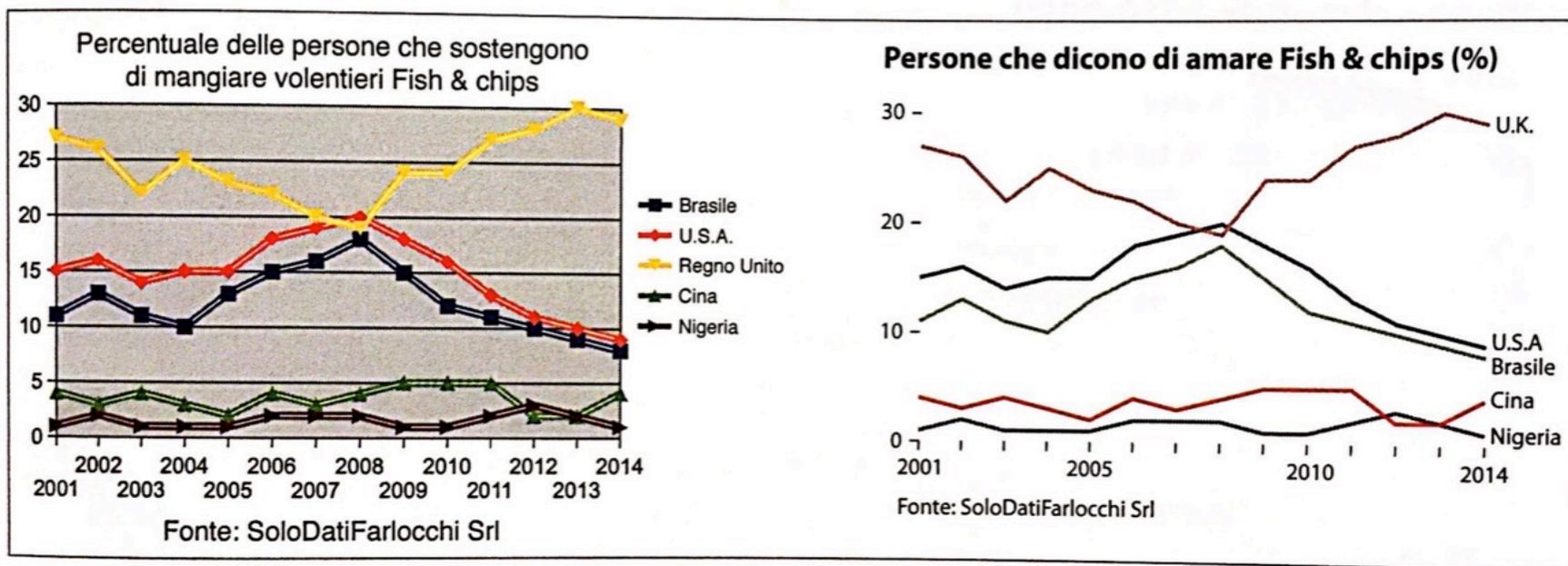
Lo scopo della visualizzazione deve guidare la decisione sulla forma da dare ai dati.

Visualizzazioni
belle

Quale dei due grafici è più bello?



Quale dei due grafici vi piace di più?



Le solite banalità

- Non è bello ciò che è bello, ma è bello ciò che piace
- La bellezza è soggettiva
- La bellezza è negli occhi di chi guarda

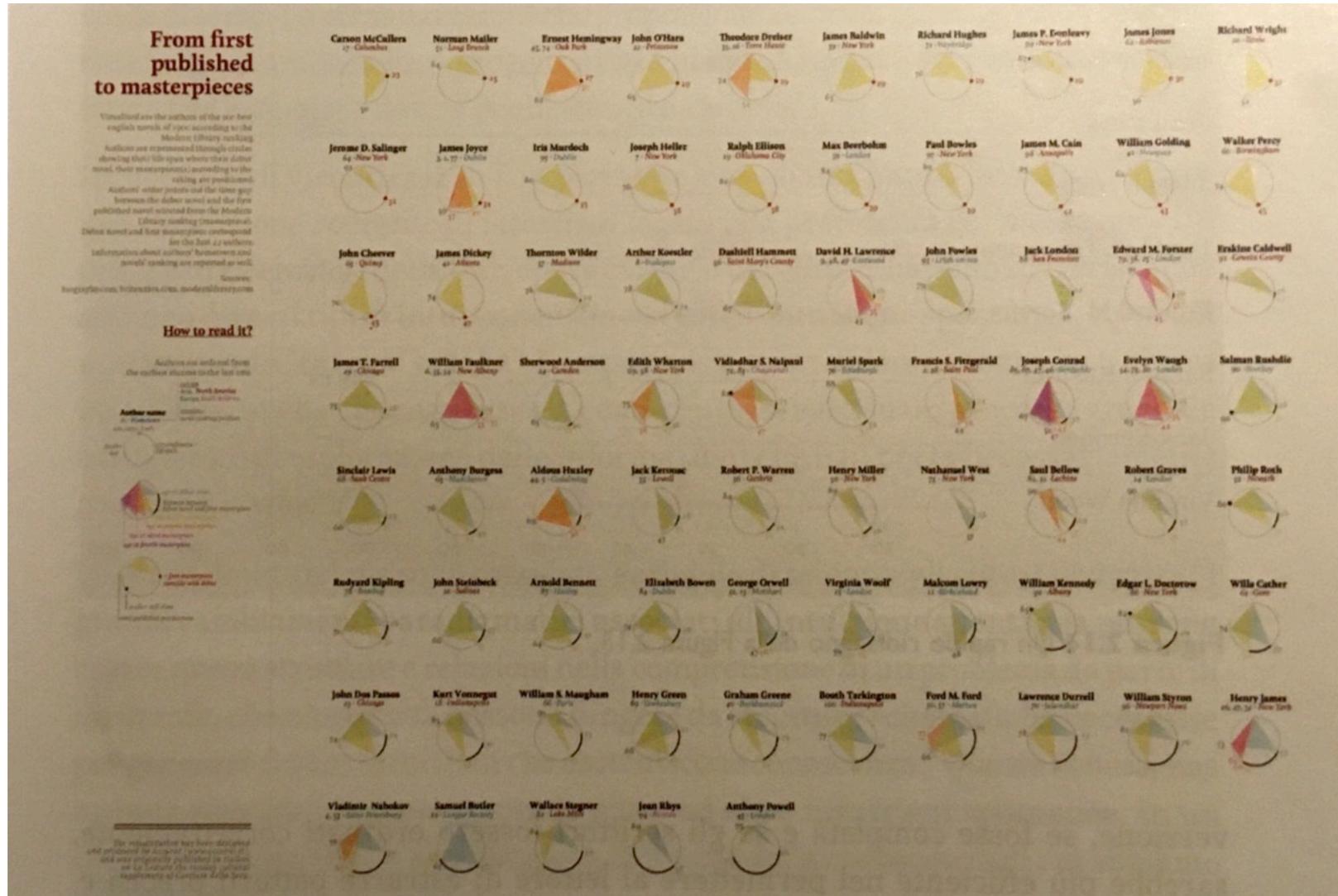
Saranno pure proverbi e modi di dire triti, ma hanno tutti in comune il fatto di porre l'accento non sulle qualità dell'oggetto osservato, ma sull'esperienza dell'osservatore



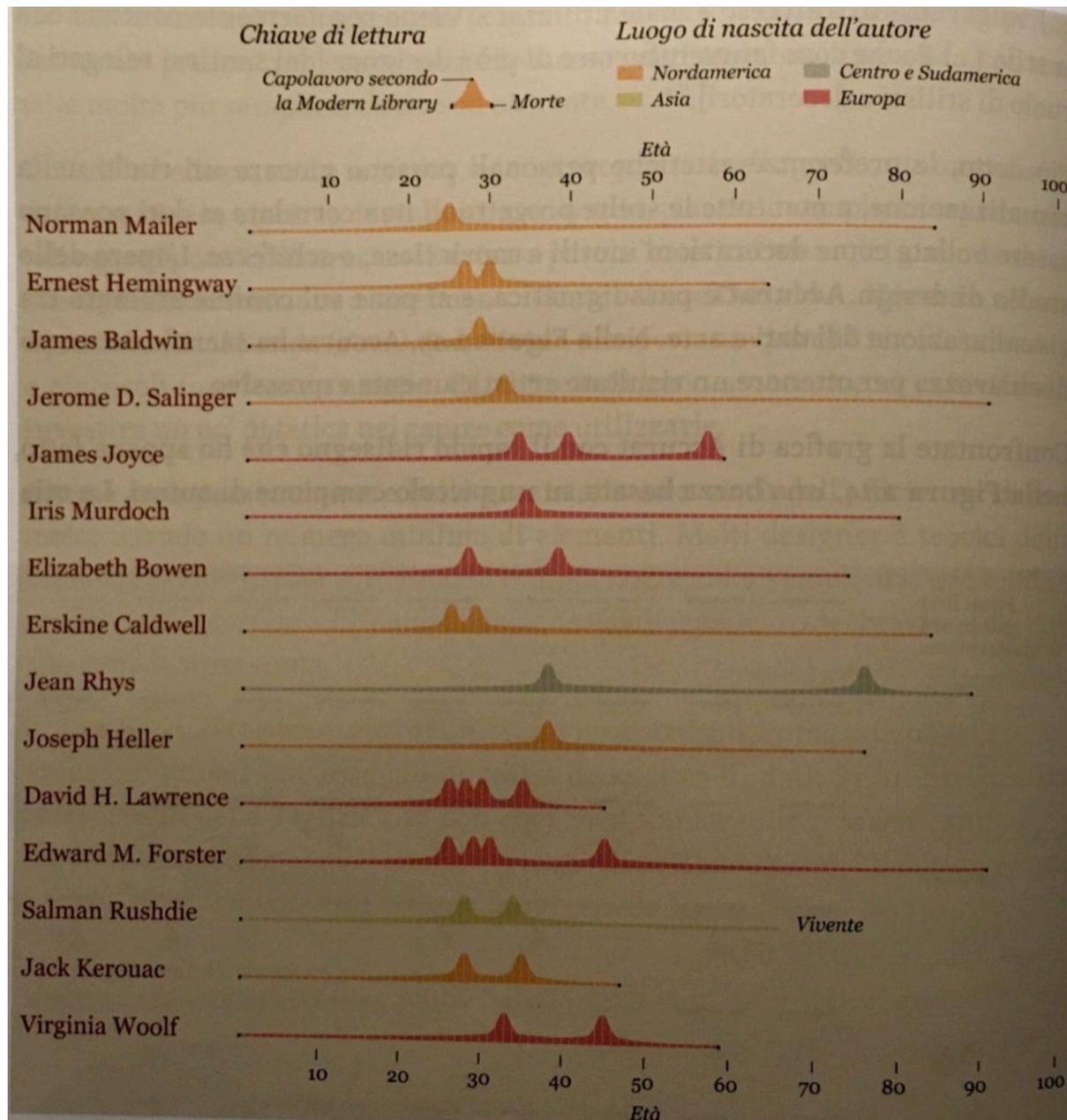
L'attenzione per gli osservatori

- Da questo punto di vista, il discorso della bellezza si collega al discorso sulla funzionalità
- “Si pensa allo scopo per cui l'oggetto viene creato e si plasma la materia prima per consentire alle persone (gli utenti dell'oggetto) di raggiungere tale scopo”
- Lo scopo, in questo caso, è di dare agli utenti un'esperienza estetica significativa
- A scapito della leggibilità della visualizzazione? Dipende, appunto, dallo scopo

Visualizzazione di Accurat.it

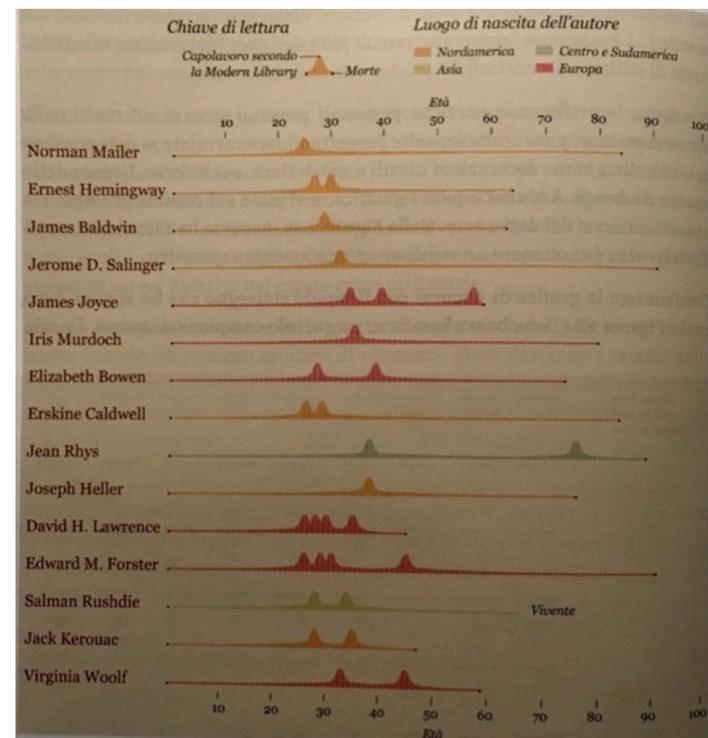
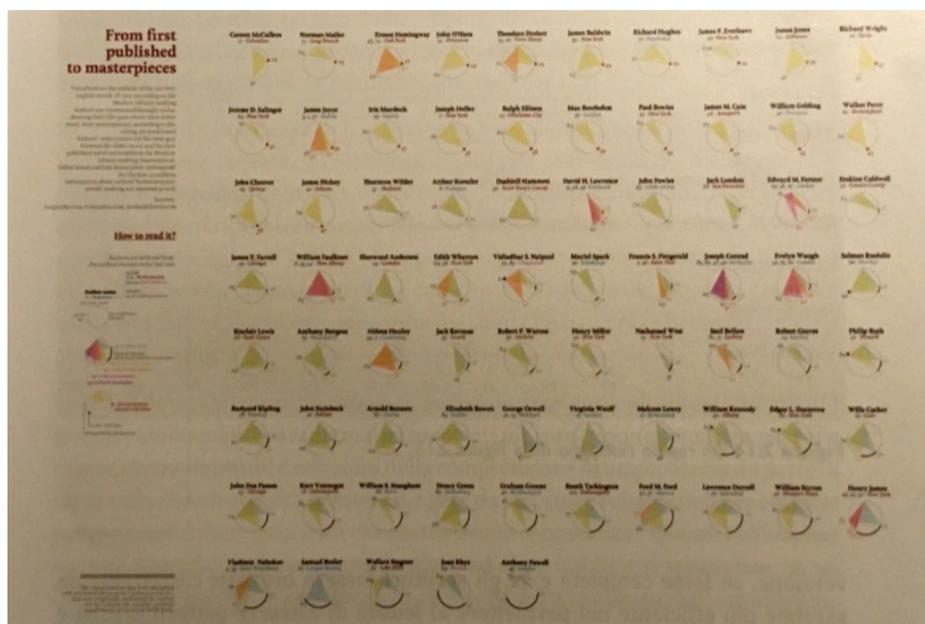


Visualizzazione di A.Cairo



Estetica vs chiarezza

- Le visualizzazioni mostrano a che età scrittori famosi hanno scritto lavori che sono considerati capolavori
- La versione originale è quella di Accurat.it
- Cairo ha creato una visualizzazione basata sulla prima per dimostrare che:
 - la sua visualizzazione presenta i dati in maniera che l'utente li capisce più velocemente
 - la visualizzazione di Accurat.it ha un valore estetico maggiore, in altre parole, è più bella



- Dimenticandoci della soggettività della bellezza e prendendo come corretta la valutazione di Cairo, essendo la sua visualizzazione più chiara, è anche la migliore?
- Lo stesso Cairo dice di no: se lo scopo è quello di affascinare gli osservatori con un'esperienza estetica, allora, seppur richiede un maggiore sforzo di lettura e comprensione, la visualizzazione di Accurat.it raggiunge il suo obiettivo.

Visualizzazioni **profonde**

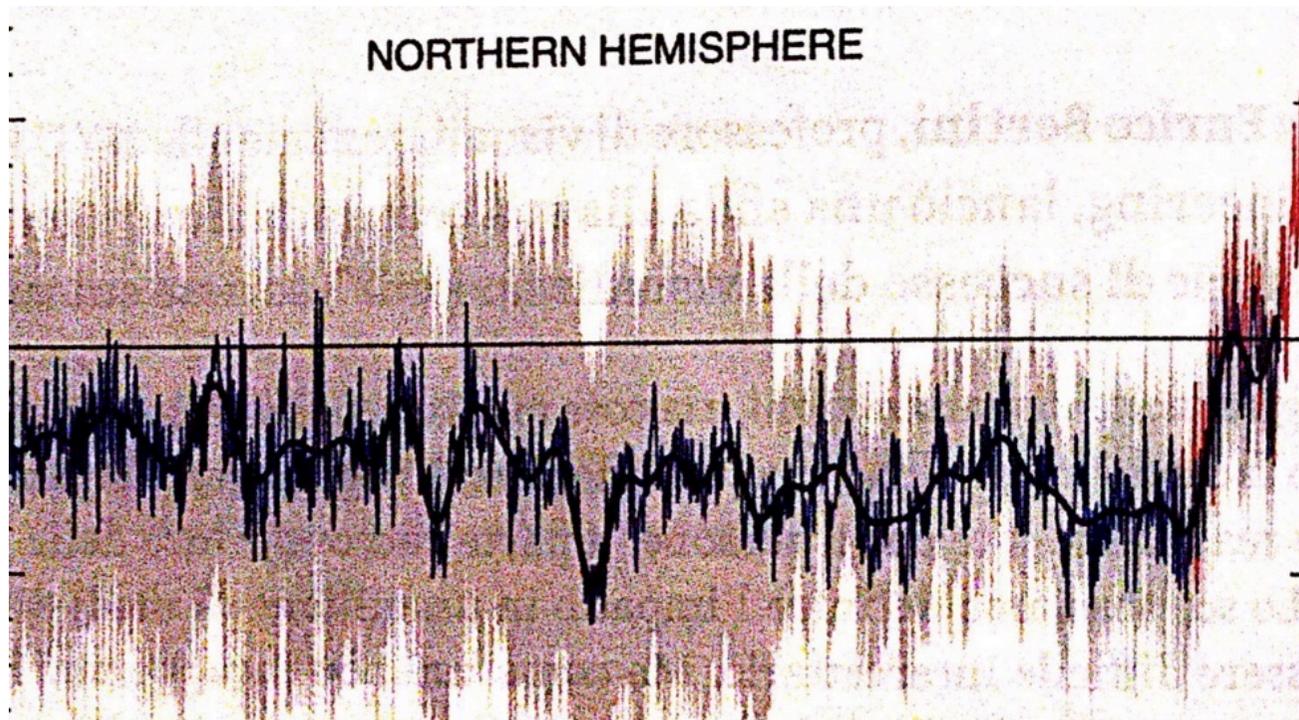
“Profondità”

- La parola “profondità” in questo contesto si riferisce alla capacità delle visualizzazioni di permettere all’osservatore di fare delle scoperte che non sarebbero possibili se i dati fossero presentati in modo diverso
- “Lo scopo della visualizzazione è l’intuizione, non le immagini”
da *Readings in Information Visualization*
di Kard, Mackinlay e Schneiderman

Intuizione

- L'idea di base è che la scoperta scaturisca da un'intuizione che è stimolata nell'osservatore dalla visualizzazione
- Definire il concetto di intuizione non è facile, ma si può pensare di distinguere due tipi di intuizione:
 - intuizione spontanea
 - intuizione graduale

Intuizione spontanea



“Hai visto che impennata?
Ma che cosa sta succedendo??”

Intuizione graduale

- È un processo meno immediato, non ha momenti di “eureka” o “wow”, ma che porta alla costruzione di conoscenza
- L'intuizione che porta alla conoscenza è più comune nelle visualizzazioni interattive, dove l'utente ha la possibilità di esaminare i diversi aspetti di un problema complesso

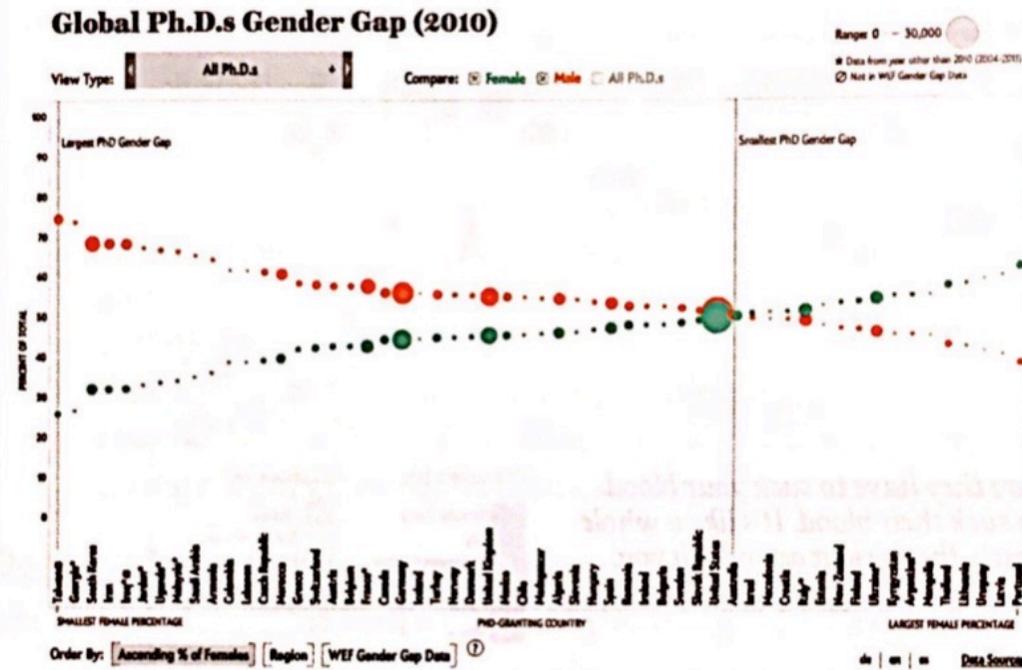


How Nations Fare in PhDs by Sex [Interactive]

How women and men fare in doctoral studies around the world

Sep 16, 2014 | [View in español](#)

In the U.S., women are going to college and majoring in science and engineering fields in increasing numbers, yet here and around the world they remain underrepresented in the workforce. Comparative figures are hard to come by, but a disparity shows up in the number of Ph.D.s awarded to women and men. The chart here, assembled from data collected by the National Science Foundation, traces the gender gap at the doctoral level for 56 nations. The situation in individual countries varies widely, but as the numbers make clear, there are interesting exceptions to the global trend.



Un esempio di visualizzazione interattiva è questo grafico di Periscopic per la rivista Scientific American sul genere degli studenti di dottorato nelle varie discipline.
www.scientificamerican.com/article/how-nations-fare-in-phds-by-sex-interactive/

Visualizzazioni **illuminanti**

Illuminazione nel senso di persone illuminate

- Il termine “illuminazione” in un contesto come questo dove si parla di intuizione, acquisizione di conoscenza, può essere equivoco
- Non stiamo parlando di utenti che si illuminano perché hanno fatto una scoperta (quella è la profondità delle visualizzazioni)
- Stiamo parlando del risultato di una buona visualizzazione: persone illuminate, ovvero che hanno cambiato modo di pensare rispetto a una questione e sono diventate persone migliori

Buffy vs Katrina

- Buffy è un personaggio di invenzione, protagonista di una serie TV fantasy andata in onda dal 1997 al 2003
- Katrina è il nome dell'uragano che nell'agosto del 2005 ha distrutto la costa della Louisiana e New Orleans, causando più di 1800 morti
- Esistono visualizzazioni molto veritiere, funzionali, belle e profonde su entrambi gli argomenti

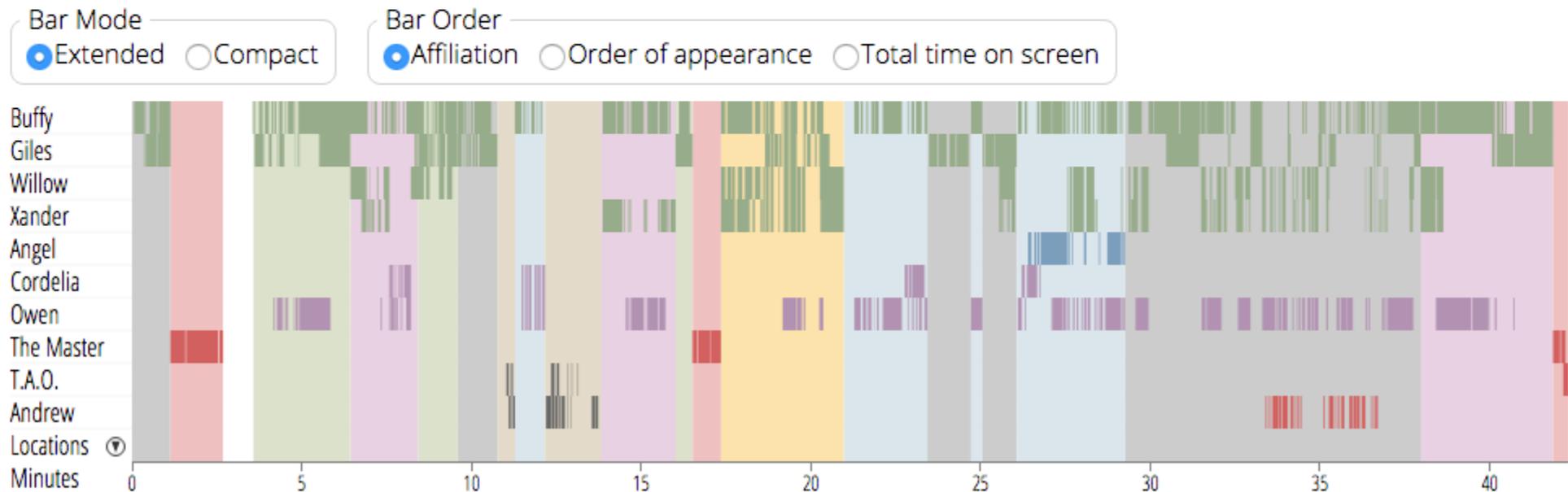
visualizingbuffy.com

1x05 - Never Kill a Boy on the First Date

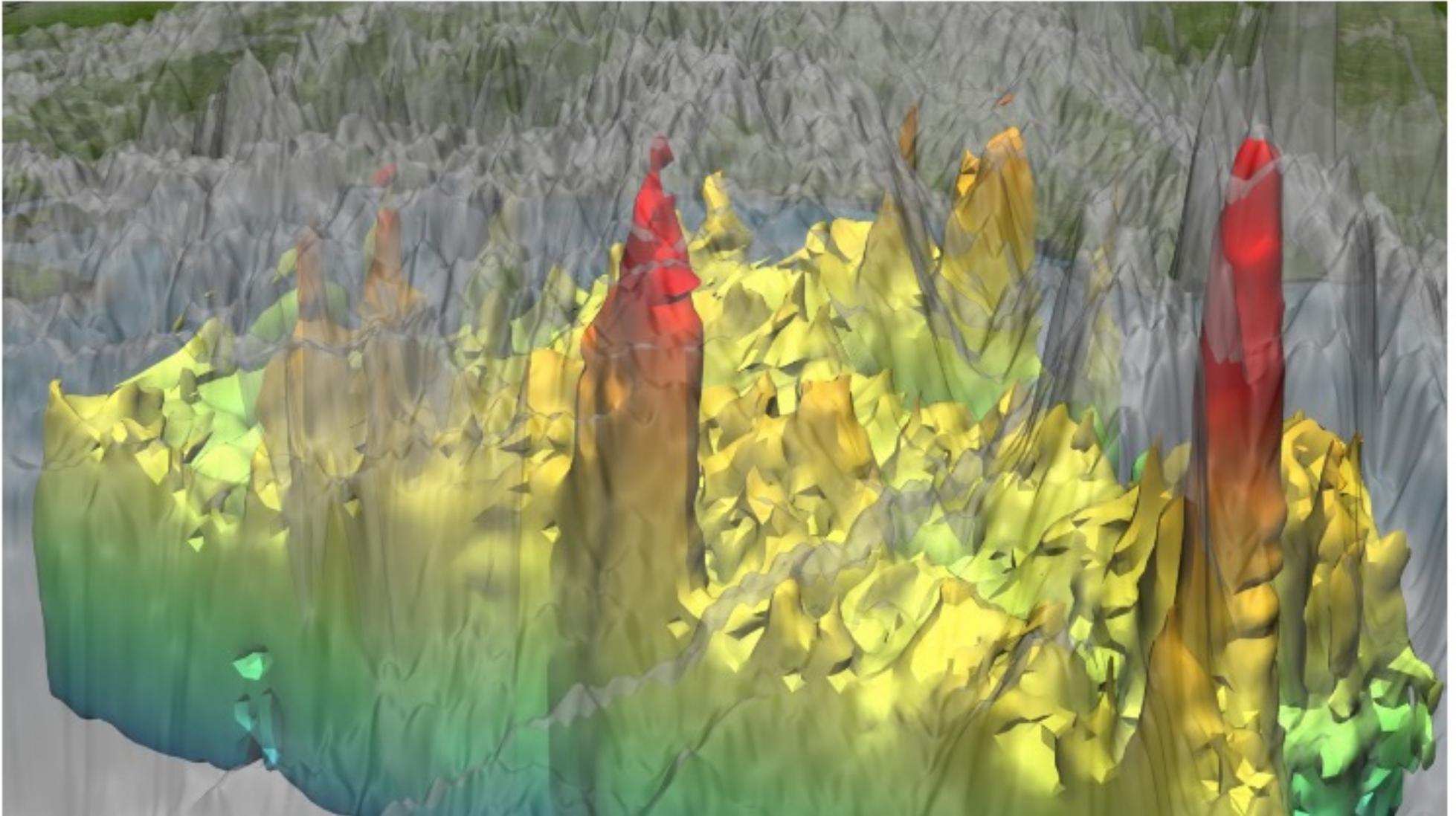
Synopsis:

Buffy sets her sights on the mysterious and brooding Owen, opting to go on a date rather than help Giles prevent a prophecy about an unstoppable vampire rising to help the Master.

Character appearances



svs.gsfc.nasa.gov/Gallery/Katrina2015.html

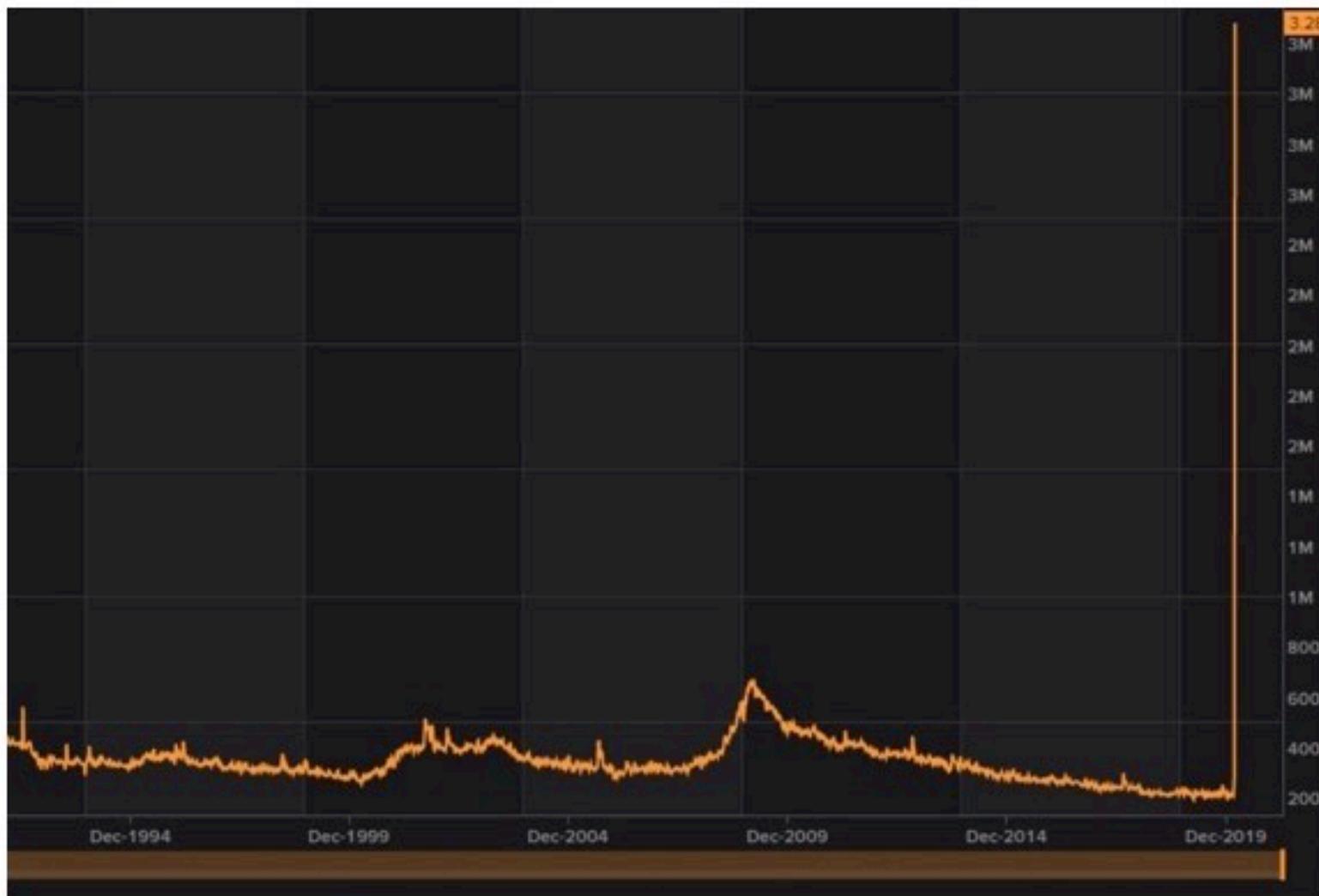


Relativismo vs illuminazione

- Tra Buffy e Katrina, qual è l'argomento più importante?
- “Dipende dal pubblico di riferimento” è una risposta corretta dal punto di vista della funzionalità, ma moralmente relativistica
- L'illuminazione, in questo contesto, punta a rendere le persone migliori
- Da questo punto di vista, alcuni argomenti sono più importanti di altri perché sono più critici per il benessere degli altri
- “Do good with data” è il motto dello studio di visualizzazione di Periscope

A proposito di benessere...





Il grafico delle richieste di disoccupazione per settimana negli Stati Uniti dagli anni Novanta a oggi

21 marzo 2020, Fonte: Bureau of Labor Statistics