

# *Data visualization*

## Presentazione delle informazioni

Modalità di rappresentazione grafica

Obiettivi

Illustrare l'approccio alla rappresentazione grafica delle informazioni

Presentare alcune regole per costruire grafici efficaci

Sperimentare alcuni processi di rappresentazione

Fulvio Sbroiavacca



# Joseph Minard 1861

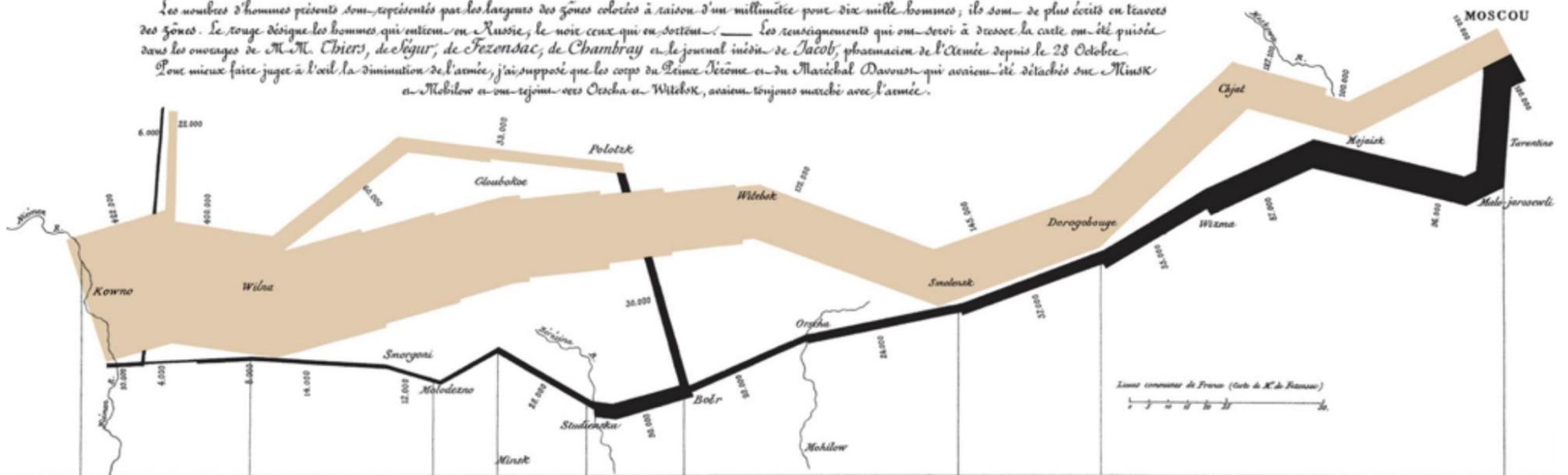
Una delle visualizzazioni più famose mai realizzate è stata di Joseph Minard e raffigura il viaggio di Napoleone mentre marciava verso la Russia per la sua campagna di Russia del 1812

## Carte Figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813.

Desinée par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite. Paris, le 20 Novembre 1869.

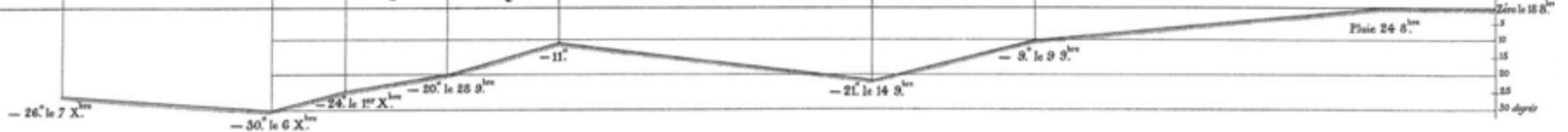
Les nombres d'hommes présents sont représentés par les largeurs des zones colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en lettres des zones. Le rouge désigne les hommes qui entrent en Russie, le noir ceux qui en sortent. — Les renseignements qui ont servi à dresser la carte ont été puisés dans les ouvrages de M. M. Chiers, de Legur, de Fezondac, de Chambray et le journal inédit de Jacob, pharmacien de l'Armée depuis le 28 Octobre.

Pour mieux faire juger à l'œil la diminution de l'armée, j'ai supposé que les corps du Prince Jérôme et du Maréchal Davout, qui avoient été détachés sur Minsk et Mohilow et qui rejoindrent Orscha et Witebsk, avoient toujours marché avec l'armée.



## TABLEAU GRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.

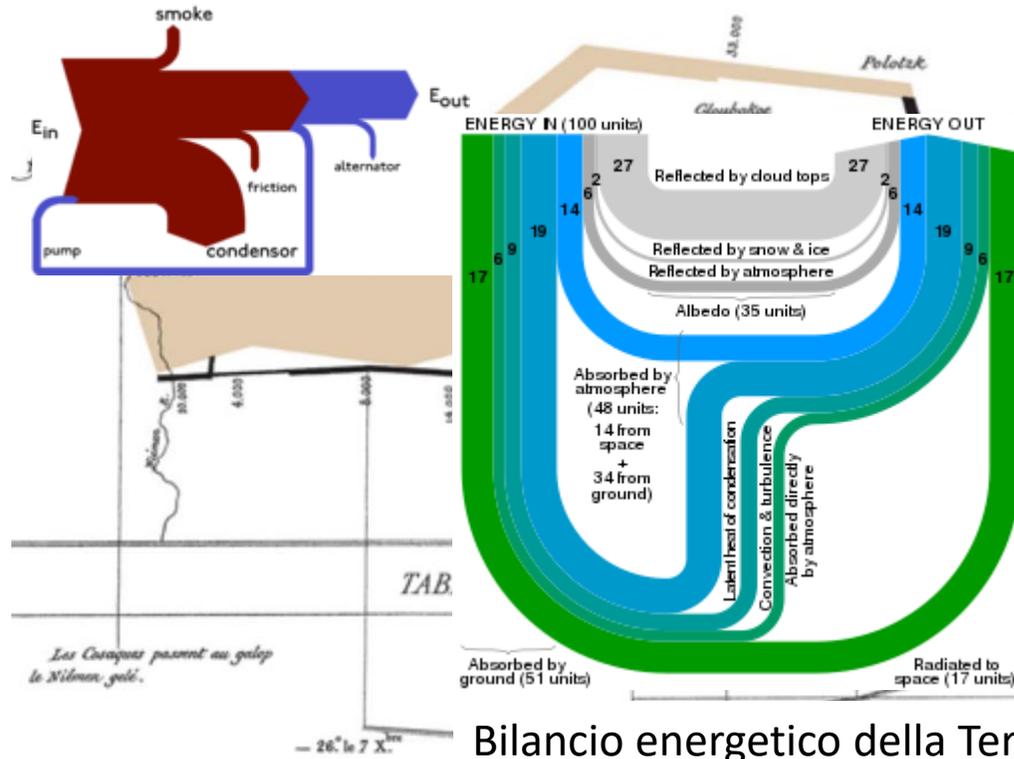
Les Cosaques passent au galop le Nilman gelé.



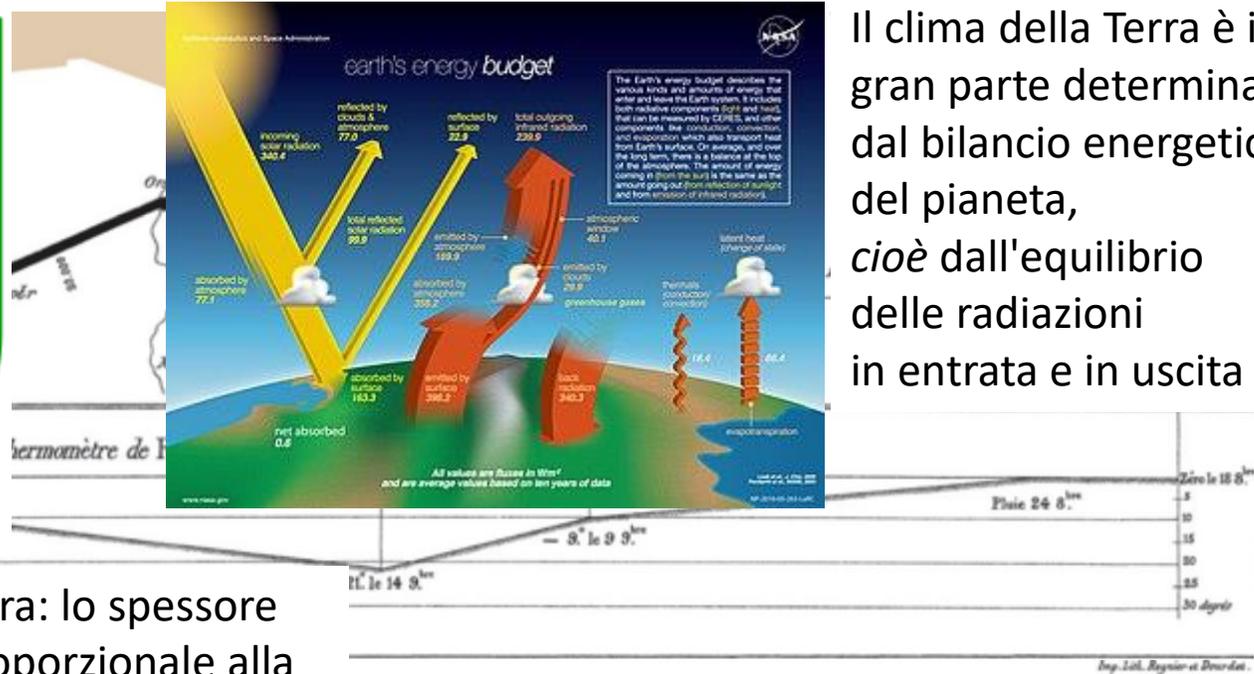
# Joseph Minard 1861

- I diagrammi di Sankey sono un tipo di diagramma di flusso in cui la larghezza delle frecce è proporzionale alla portata

successives en hommes de l'Armée de  
M. Minard, Inspecteur Général  
is par les largeurs des zones colorées à raison  
ou en Russie, le noir ceux qui en sortent  
de Fezonsac, de Chambrey et le jo  
de l'armée, j'ai supposé que les corps du Prince  
Nobilow et un régiment vers Ostcha et Witebsk)



L'illustrazione mostra l'esercito di Napoleone in partenza dal confine polacco-russo. Una banda spessa illustra le dimensioni del suo esercito in punti geografici specifici durante la loro avanzata e ritirata. Visualizza sei tipi di dati in due dimensioni: il numero delle truppe napoleoniche; la distanza percorsa; temperatura; latitudine e longitudine; la direzione del viaggio; e posizione relativa a date specifiche senza menzionare Napoleone. L'interesse di Minard risiedeva nei travagli e nei sacrifici dei soldati. Questo tipo di grafico a bande per l'illustrazione dei flussi fu successivamente chiamato **diagramma di Sankey**.



Il clima della Terra è in gran parte determinato dal bilancio energetico del pianeta, cioè dall'equilibrio delle radiazioni in entrata e in uscita

Bilancio energetico della Terra: lo spessore della linea è linearmente proporzionale alla quantità relativa di energia

# Il potere della visualizzazione nella scienza dei dati

## Un'immagine dice davvero più di mille parole

*"La visualizzazione riguarda davvero la cognizione esterna, ovvero il modo in cui le risorse esterne alla mente possono essere utilizzate per potenziare le capacità cognitive della mente." - **Stuart Card***

La quantità e la complessità delle informazioni prodotte nel campo della scienza, dell'ingegneria, degli affari e dell'attività umana quotidiana stanno aumentando a ritmi sbalorditivi

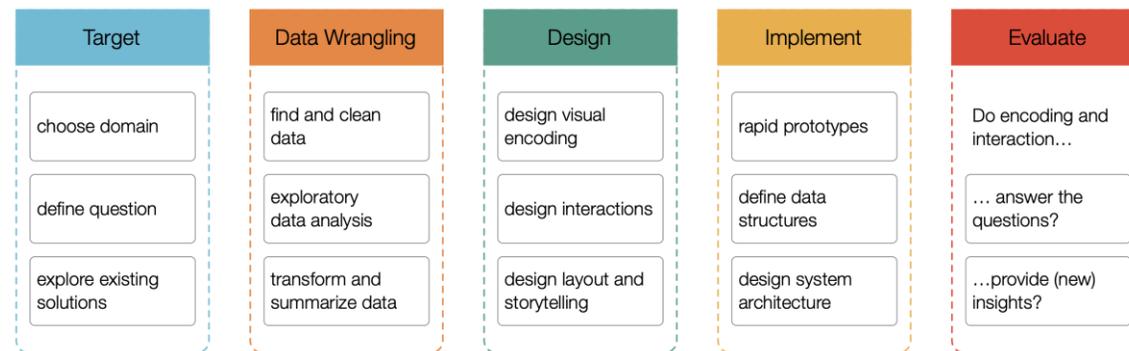
Una buona visualizzazione non solo presenta un'interpretazione visiva dei dati, ma lo fa migliorando la **comprensione**, la **comunicazione** e il **processo decisionale**

# Obiettivi di visualizzazione

**Esplorazione dei dati** > *trova l'ignoto*

**Analisi dei dati** > *verifica delle ipotesi*

**Presentazione** > *comunicare e diffondere*



## Problemi comuni

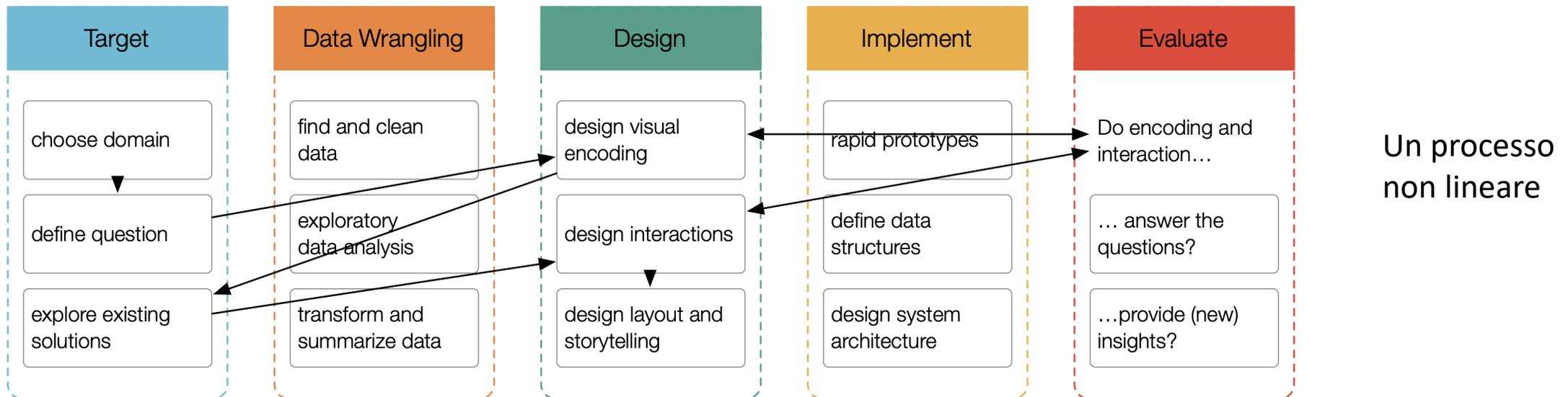
- Situazione del dominio: sono state comprese correttamente le esigenze degli utenti? Forse si sta affrontando il problema sbagliato (fase iniziale)
- Astrazione di dati/attività: si sta mostrando la cosa giusta ai fruitori? Forse si sta utilizzando l'astrazione sbagliata (fase iniziale)
- Codifica/interazione visiva: il modo in cui si mostrano i dati funziona? Forse si sta utilizzando la codifica sbagliata (fase di design)
- Algoritmo: si sta utilizzando l'algoritmo giusto? Forse viene utilizzato l'algoritmo sbagliato (fase di realizzazione)

Alcuni metodi qualitativi per acquisire informazioni su come le persone interpreteranno e risponderanno alle loro idee o progetti

- Studi osservazionali ("Pensa ad alta voce")
- Interviste agli esperti (note anche come critiche al design)
- Focus group

# Un modello in cinque fasi

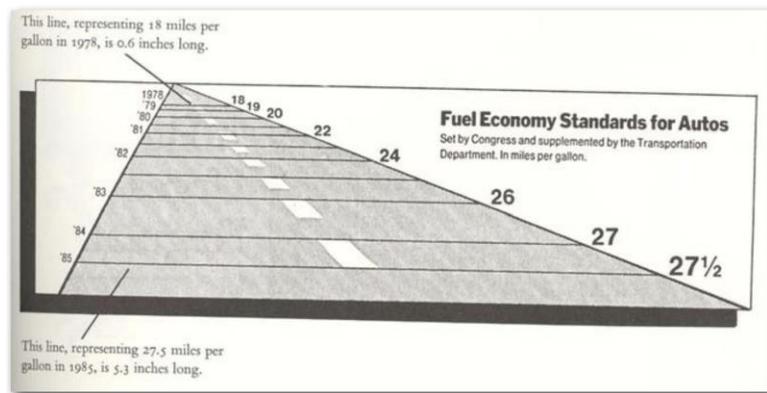
- Isolare un obiettivo o una domanda specifica
- Inserire i dati in un formato utilizzabile, eseguire analisi esplorative (90% lavoro Data Scientist)
- Sviluppare una storia che si vuole raccontare con i dati (strettamente correlato all'obiettivo definito e al messaggio che si vuole comunicare)
- Realizzare la visualizzazione
- Valutare se inviare il messaggio che si desidera comunicare e se risponde alla domanda a cui si è deciso di rispondere



# Tre regole per una visualizzazione (Edward Tufte)

## Integrità grafica

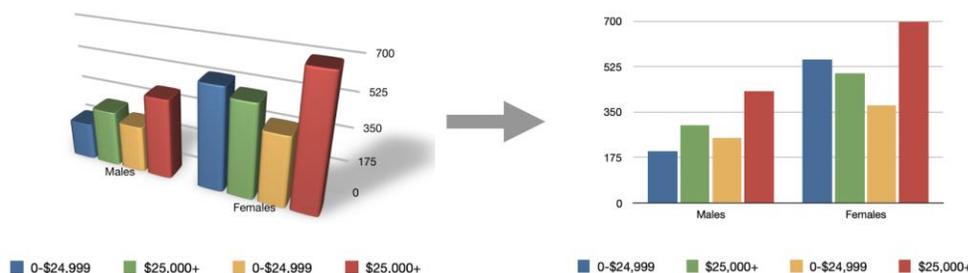
$$\frac{\text{Size of effect shown in graphic}}{\text{Size of effect in data}}$$



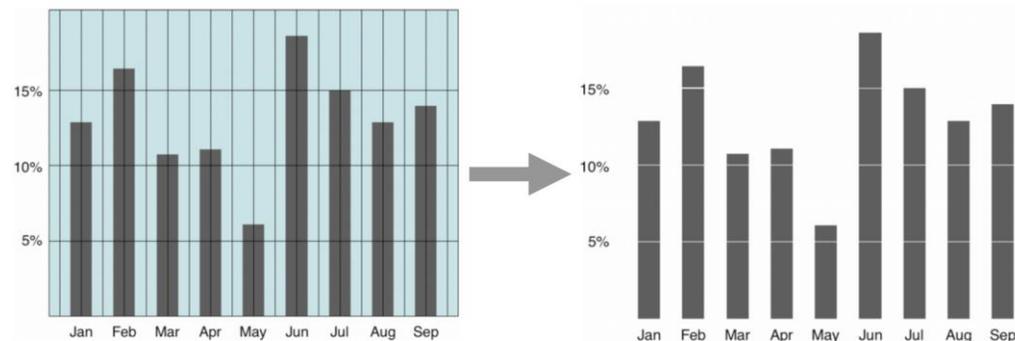
Tufte, VDQI

## Chiarezza e minimalismo

$$\text{Data-Ink Ratio} = \frac{\text{Data ink}}{\text{Total ink used in graphic}}$$



No  
Elementi  
visivi  
estranei



# Data visualization

Per interpretare i risultati ottenuti con l'**analisi dei dati** e della **scienza dei dati** (enormi quantità di dati) e prendere decisioni *data oriented* è necessario produrre una **chiara visualizzazione delle informazioni** prodotte

Per trasformare i dati in conoscenza

- è essenziale presentarli nel modo giusto
- utilizzando elementi visivi come grafici, diagrammi, tabelle e mappe, per vedere e comprendere tendenze, valori particolari o anomali (tendenze e correlazioni potrebbero non essere rilevati osservando dati testuali)

*Una visualizzazione efficace consente di esporre i contenuti eliminando dai dati il superfluo e portando in primo piano le informazioni utili*

# La scelta della forma rappresentativa

La scelta di una forma di grafico dipende dal concetto che si vuole esprimere

Prima di scegliere il tipo di rappresentazione:

- Determinare **cosa** si vuole esprimere
- **Quali relazioni** si vogliono evidenziare

*Un grafico non è giusto o sbagliato in assoluto:  
deve essere finalizzato allo scopo*

# Quale rappresentazione grafica utilizzare?

La scelta di una forma di grafico dipende dal concetto che si vuole esprimere

Procedere per passi:

- scegliere un **titolo**
- produrre una **frase** che riassume cosa si vuole evidenziare
- scegliere un **tipo di grafico** che realizzi la frase

*Un grafico non è giusto o sbagliato in assoluto:  
deve essere finalizzato allo scopo*

# Tipi di rappresentazione grafica

Componenti di un intero

Rapporto tra due variabili

Confronto tra fenomeni diversi

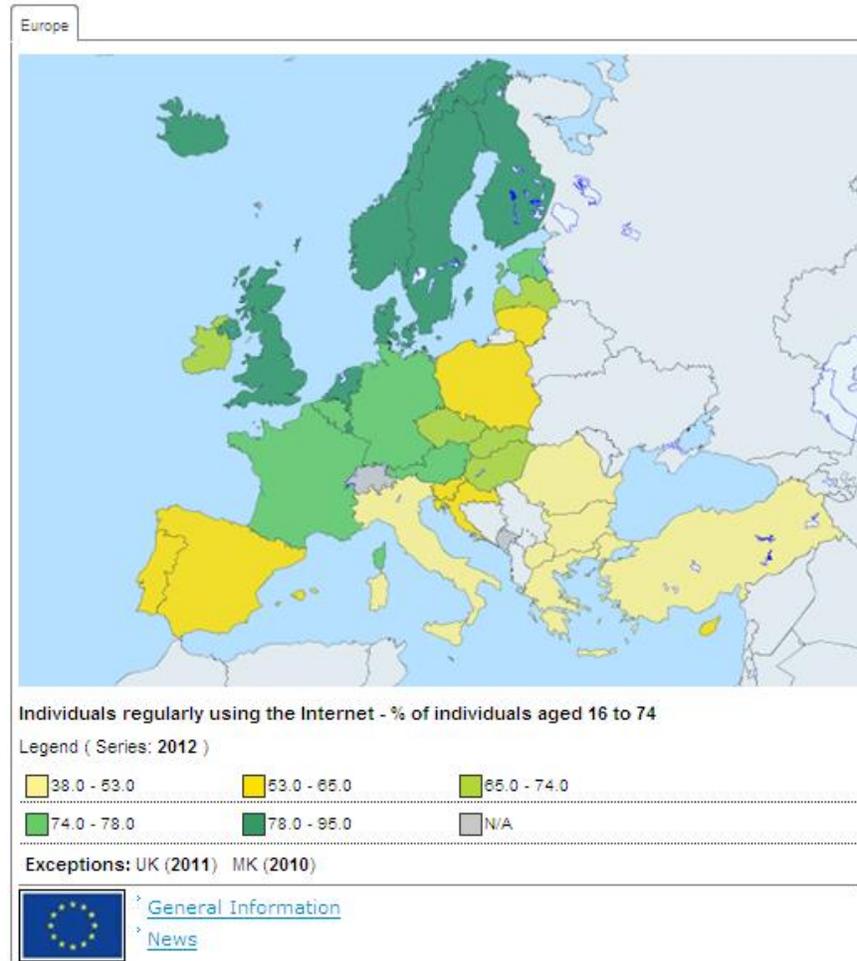
Rappresentazioni temporali

Un caso particolare: Grafici azionari

Rappresentazioni di tipo territoriale

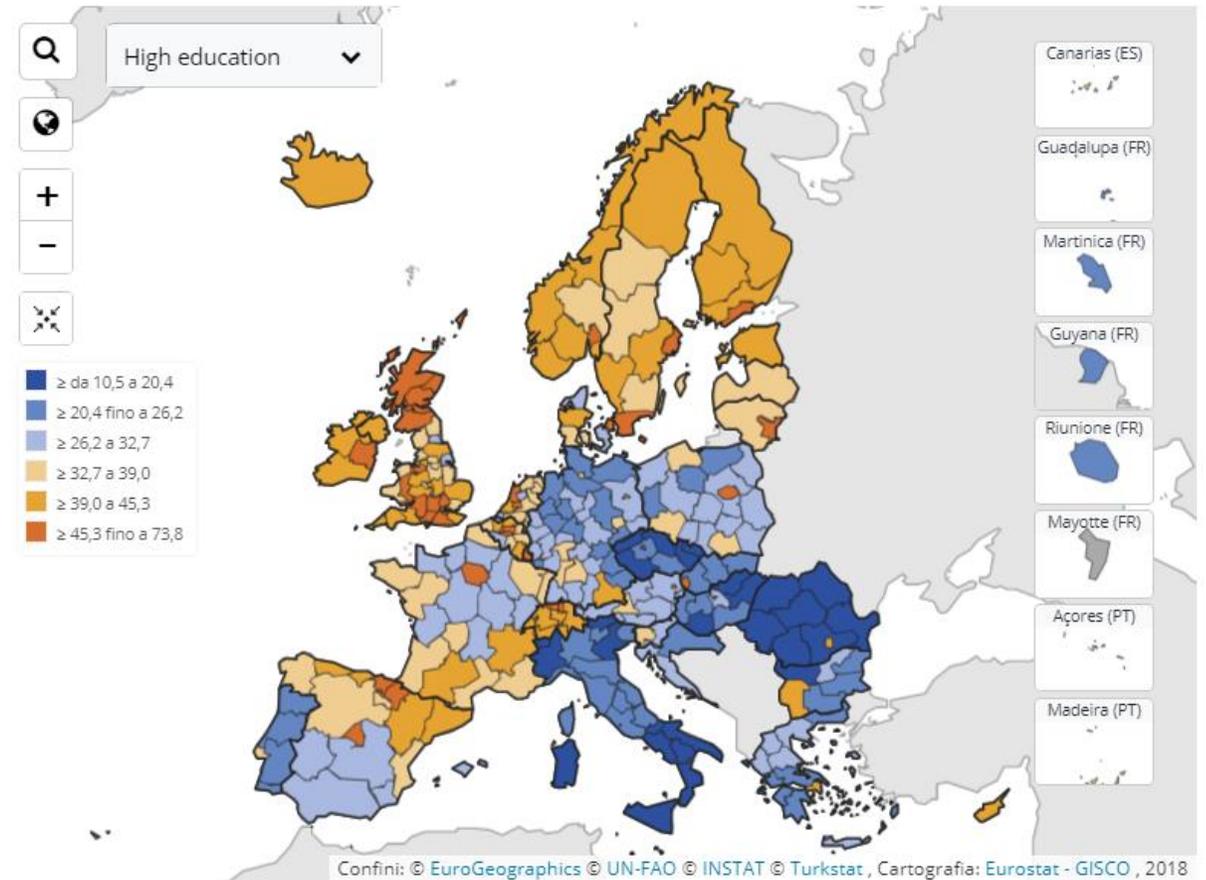
*Un grafico non è giusto o sbagliato in assoluto:  
deve essere finalizzato allo scopo*

# Rappresentazioni di tipo territoriale



[ec.europa.eu/eurostat](https://ec.europa.eu/eurostat)

**Livello di istruzione, 2019**  
(come% della popolazione di età compresa tra 25 e 64 anni)



Livello di istruzione basso: non più di un'istruzione primaria o secondaria inferiore; livello di istruzione medio: istruzione secondaria superiore o post-secondaria non terziaria; alto livello di istruzione: istruzione terziaria.

Fonte: Eurostat - [accesso al set di dati](#)

<https://ec.europa.eu/eurostat/cache/digpub/regions/>

# Rappresentazioni di tipo territoriale – scarico di un set di dati

The image shows a composite of three browser screenshots from the website [www.datiopen.it](http://www.datiopen.it). The top screenshot shows the homepage with the logo "DatiOpen.it Il portale Italiano dell'Open Data" and a search bar. The middle screenshot shows a blog post titled "Data Smart People: statistiche su formazione ed occupazione" with a description of the smart community and a list of data indicators. The bottom screenshot shows a data visualization interface for "LAUREATI IN SCIENZA E TECNOLOGIA PER REGIONE" with various filters and a download menu.

**Blog Post:** Data Smart People: statistiche su formazione ed occupazione. La smart community è caratterizzata da elevati livelli di qualifica dei suoi componenti, che contribuiscono alla vita pubblica come sorgenti che alimentano una rete pacifica e consapevole con molteplici interessi, frutto della continua formazione ed affinamento delle conoscenze. Con Data Smart People il Ministero dello sviluppo economico, c

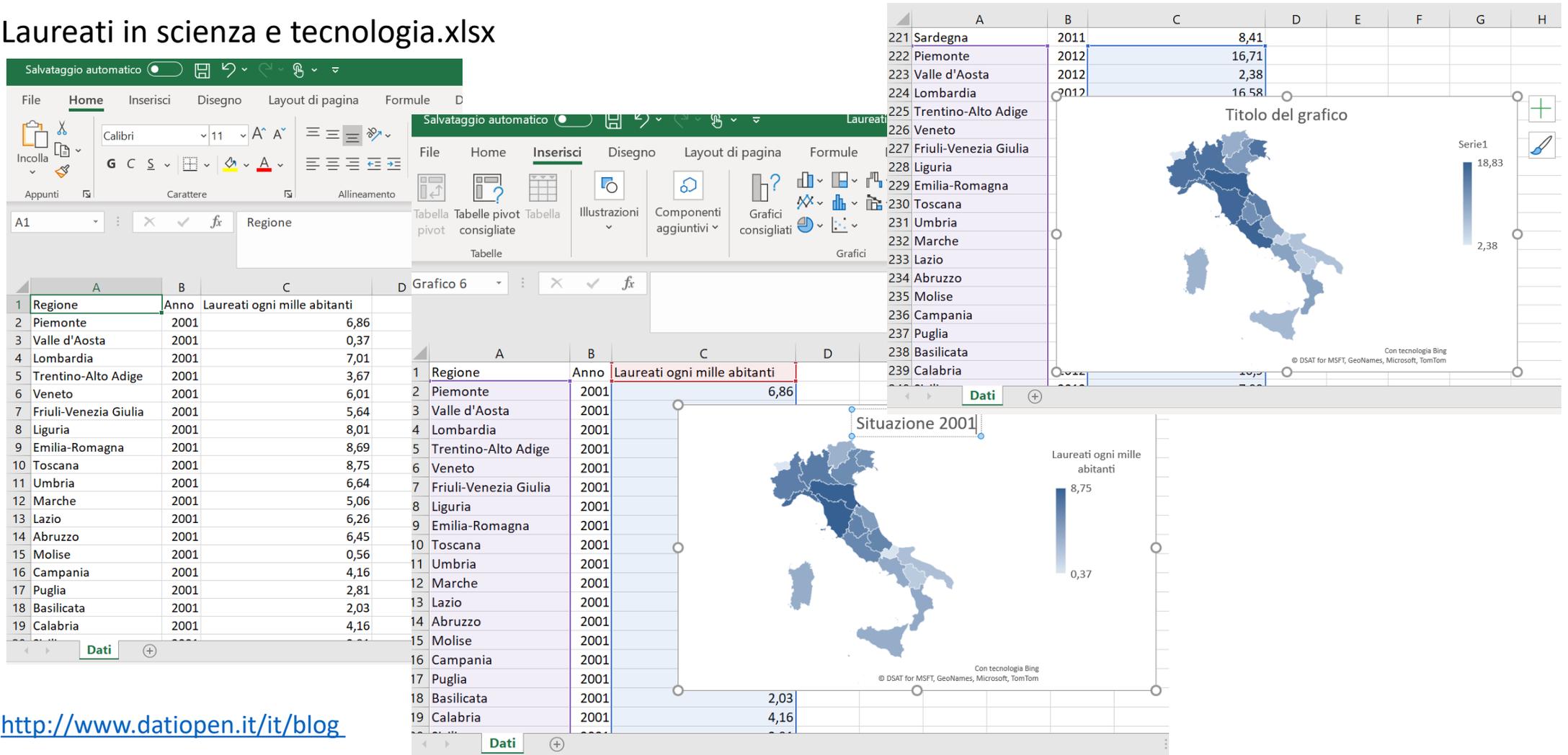
**Data Download Options:**

- File originale: DOWNLOAD (LAUREATI IN SCIENZA E TECNOLOGIA.XLSX)
- Esporta in CSV: Dato Filtrato, Dato Completo
- Esporta in altri formati: XML/Atom, JSON
- Linked Open Data: RDF/TTL

<http://www.datiopen.it/it/blog>

# Rappresentazioni di tipo territoriale – utilizzo di un set di dati

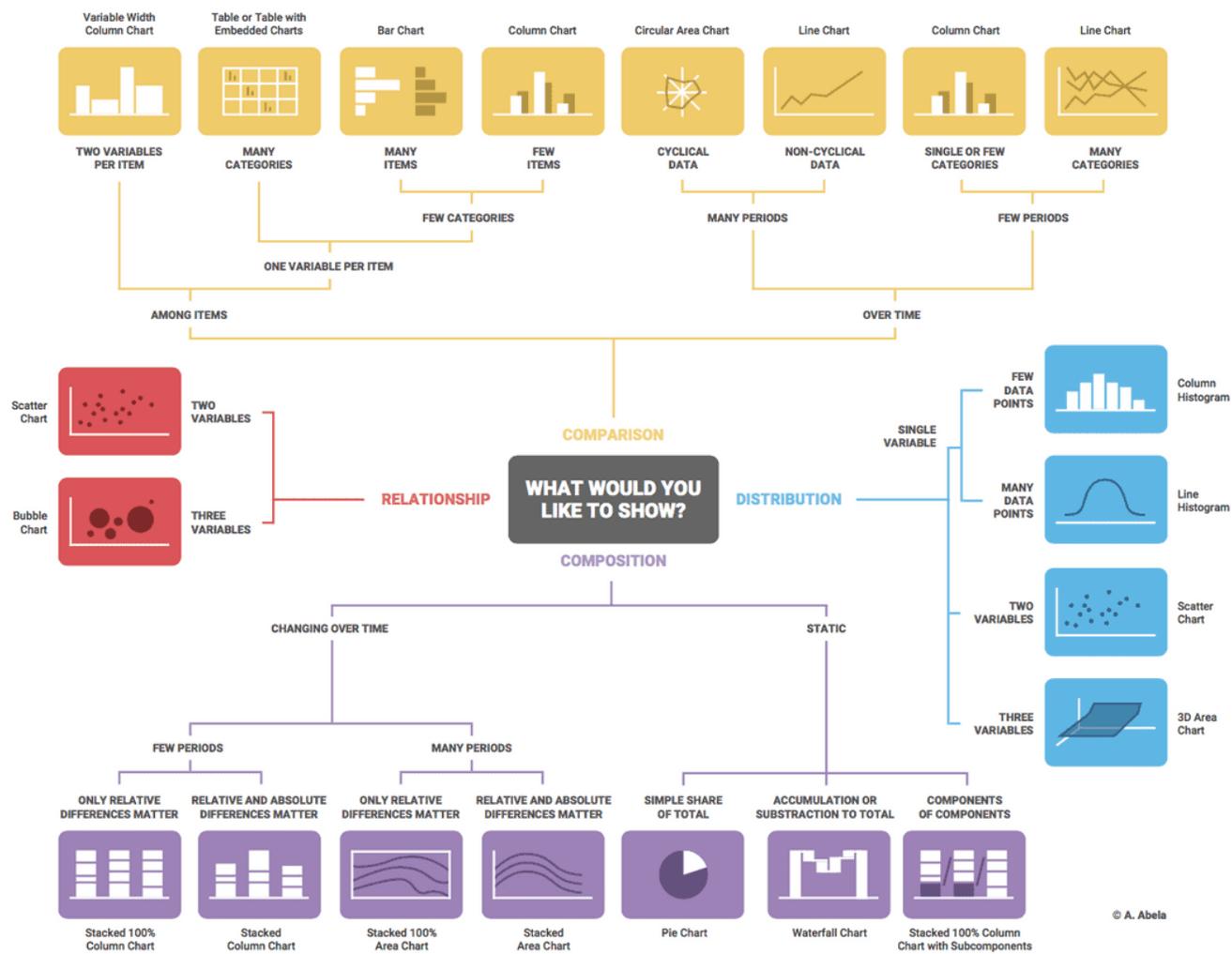
## Laureati in scienza e tecnologia.xlsx



# Tipi di rappresentazione grafica e possibili scelte

	Torte	Linee	Istogrammi	Mappe
Rappresentazioni temporali	Red	Green	Green	Red
Componenti di un intero	Green	Red	Green	Red
Rapporto tra due variabili	Red	Green	Green	Red
Confronto tra fenomeni diversi	Red	Red	Green	Red
Rappresentazioni di tipo territoriale	Red	Red	Red	Green

# Risorsa per la scelta del grafico



Cosa si vuole rappresentare?

**Confronto** : quando vogliamo confrontare tra loro i diversi valori o attributi all'interno dei dati. Esistono alcune varianti, a seconda dei dati. Ad esempio: i dati includono la variabile temporale? Quanti periodi di tempo? Quante variabili e categorie presentano i dati ?

**Composizione** : quando vogliamo sapere come sono composti i dati, cioè quali caratteristiche generali sono presenti nel data set. Esistono alcune varianti, a seconda dei dati. Ad esempio: ci sono cambiamenti nel tempo? Con quanti periodi di tempo? Nei dati statici, hai valori che si accumulano ?

**Distribuzione** : quando vogliamo capire come sono distribuiti i singoli punti dati all'interno dell'intero set di dati. A seconda del numero di variabili in cui vogliamo analizzare la distribuzione, possiamo scegliere grafici a barre, grafici a linee o grafici a dispersione.

**Relazione** : in questo caso ci interessa sapere come i valori e gli attributi sono correlati tra loro. Per rappresentare i valori, i grafici a dispersione vengono solitamente utilizzati quando sono coinvolte due variabili e i grafici a bolle vengono utilizzati quando sono coinvolte tre variabili.

# Esempio di ausilio



## Microsoft Power BI

Confronto per confrontare l'entità delle misure

Modifica nel tempo - Per visualizzare la tendenza al cambiamento delle misure

Part-to-whole per identificare le parti che compongono un totale di misura

Flusso per visualizzare un flusso o relazioni dinamiche

Classifica per classificare le misure in un ordine

Spaziale per visualizzare le misure su mappe spaziali

Distribuzione per visualizzare la distribuzione dei valori

Correlazione per mostrare le correlazioni tra le misure

Singolo per presentare valori singoli

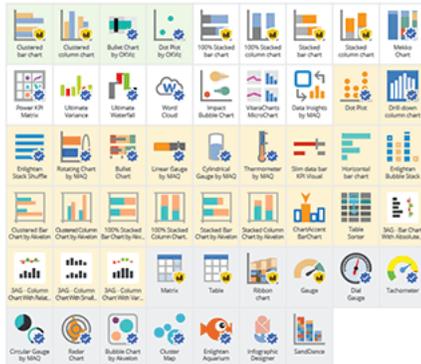
Narrativa raccontare una storia con i dati

Filtro per controllare i filtri dei rapporti

Miscellanea

### COMPARISON

Display measures compared by their magnitude



### CHANGE OVER TIME

Display the changing trend of measures



### RANKING

Display measures by their rank order



### SPATIAL

Display measures over spatial maps



### FLOW

Display a flow or dynamic relations



### PART-TO-WHOLE

Display the parts of a measure



### DISTRIBUTION

Display the distribution of a measure



### CORRELATION

Display relations between measures



### SINGLE

Display single values



### FILTER

Control report filters



### NARRATIVE

Tell a story with data



### MISCELLANEOUS



Recommended (green), There is a better alternative (yellow), Don't use in the category (grey), Built-in visual (blue), Certified visual (purple), R required (red)

2018 © SQLBI AND OKVIZ ARE TRADEMARK OF SQLBI CORP. / POWER BI IS A TRADEMARK OF MICROSOFT CORP.

Riferimento per scegliere gli elementi visivi giusti in Microsoft Power BI

<https://www.sqlbi.com/ref/power-bi-visuals-reference/#>

# Un ausilio interattivo

Il progetto Data Viz - <https://datavizproject.com/>

The screenshot displays the website [datavizproject.com](https://datavizproject.com/) with a dark blue header. The header includes a logo with 'D', 'V', and 'P' in a grid, navigation menus for 'ALL', 'FAMILY', 'INPUT', 'FUNCTION', and 'SHAPE', a search icon, and a 'by ferdio hire us!' button. The main content area features a grid of ten interactive chart examples:

- Alluvial Diagram:** A flow diagram with three vertical columns labeled A, B, and C, connected by red and dark blue flows.
- Sankey Diagram:** A flow diagram starting from a large red block labeled A, branching into several smaller blocks labeled B, C, D, E, and F.
- Donut Chart:** A donut chart divided into four segments labeled A, B, C, and D.
- Line Graph:** A line graph with a red line showing data points over 10 units on the x-axis and values from 0 to 1000 on the y-axis.
- Radial Bar Chart:** A radial bar chart with three concentric rings labeled A, B, and C, and numerical values (200, 400, 600, 800, 1000) around the perimeter.
- Polar Area Chart:** A polar area chart with multiple segments of varying lengths radiating from a central point.
- Pictorial fraction chart:** A pictorial chart showing a human figure with a red lower half and a dark blue upper half, labeled with '75%'.
- Radial Histogram:** A radial histogram with a central circle and numerous thin red bars radiating outwards, labeled with numbers 0 through 22.
- Exploded View Drawing:** An exploded view drawing of a mechanical part with segments labeled A, B, C, and D.
- Sorted Stream Graph:** A stream graph with four vertical columns labeled JAN, FEB, MAR, and APR, showing overlapping red and dark blue flows.

# Potenziare la rappresentazione

## Raccontare una storia

un grafico = un messaggio

La visualizzazione non deve limitarsi al grafico in quanto tale

Per facilitare la comprensione bisogna raccontare una storia in cui tutte le informazioni sono presenti (data storytelling)

Commenti, leggenda, colori adeguati, ... tutto deve consentire a chi fruisce della rappresentazione di comprendere al meglio le informazioni da trasmettere

## Rendere azionabile la rappresentazione

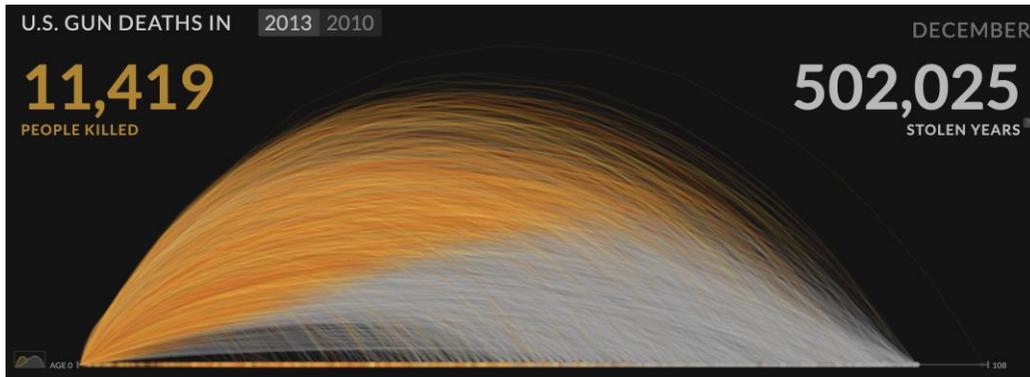
La presentazione delle informazioni e degli indicatori di performance può permettere al fruitore di compiere delle azioni

Ad esempio una curva rappresenta una certa tendenza nel tempo, nel prossimo periodo come si comporterà? Modificando alcuni parametri tra un livello minimo ed un massimo quale comportamento assumerà? Quale sarà il campo di variabilità?

Più rappresentazioni possono essere raggruppate in una dashboard con possibili azioni predefinite

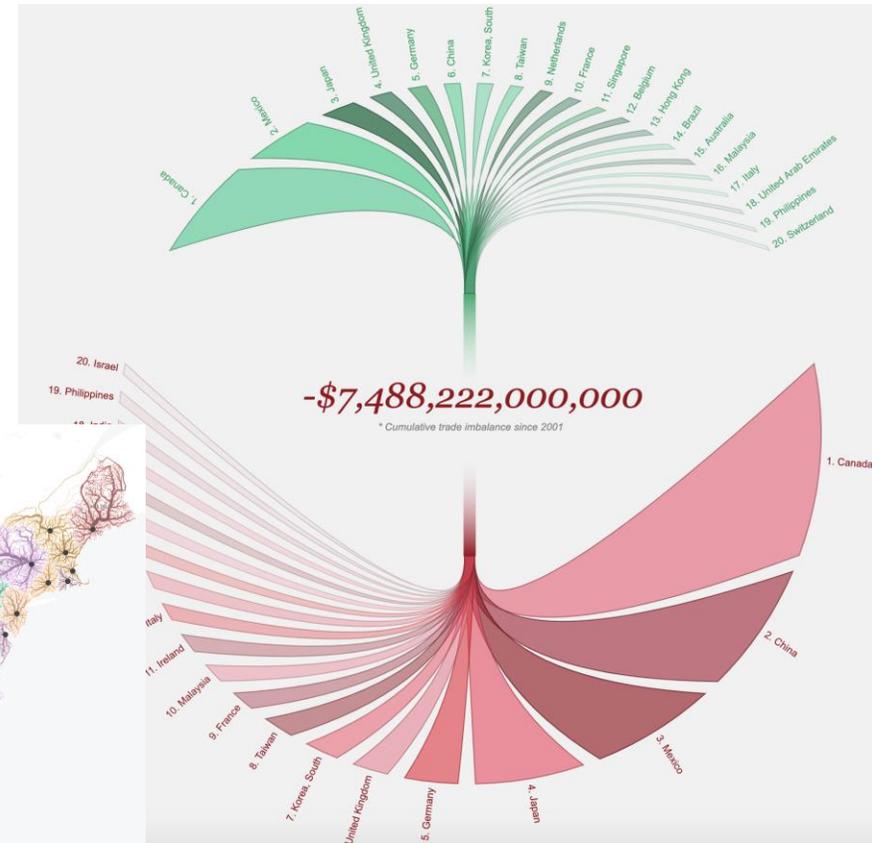
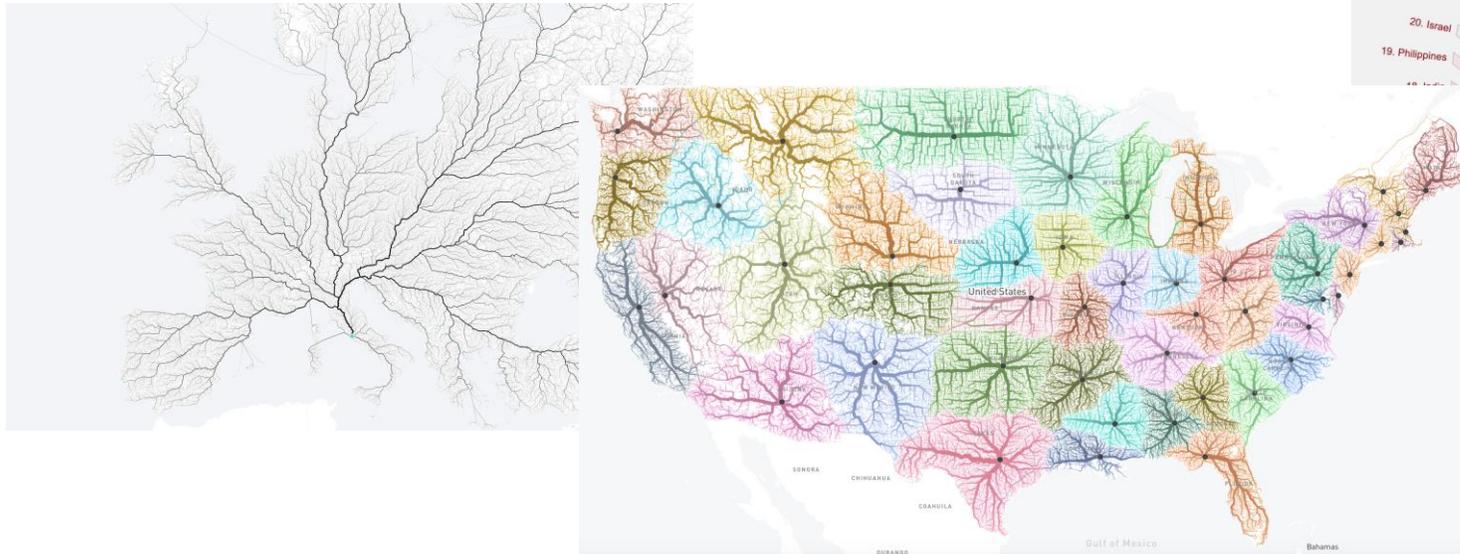
# Esempi visualizzazioni interattive

<https://guns.periscopic.com/?year=2013>

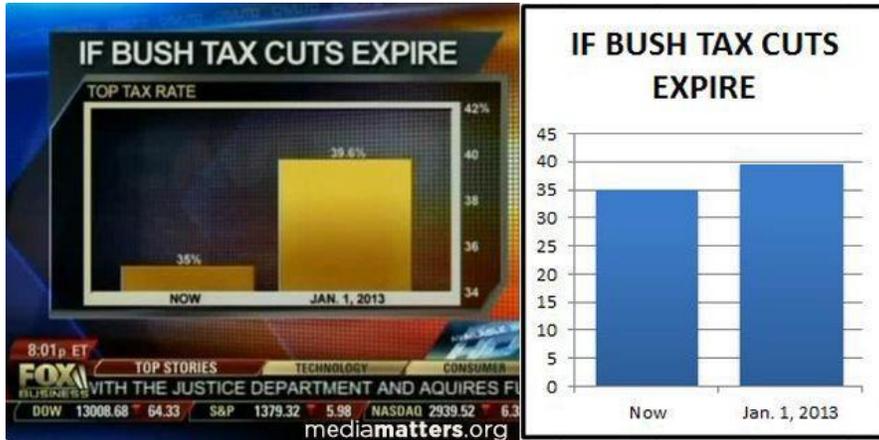


Galleria esempi su GitHub:

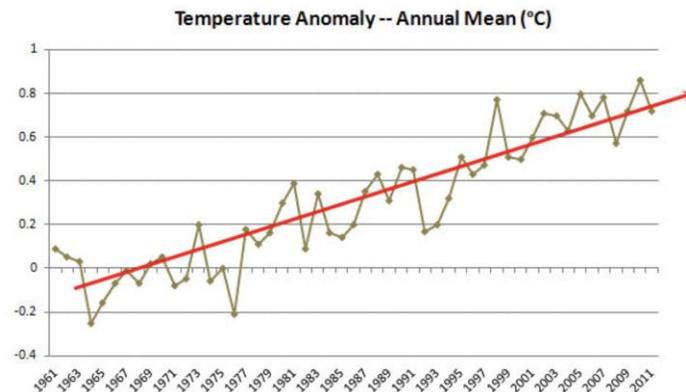
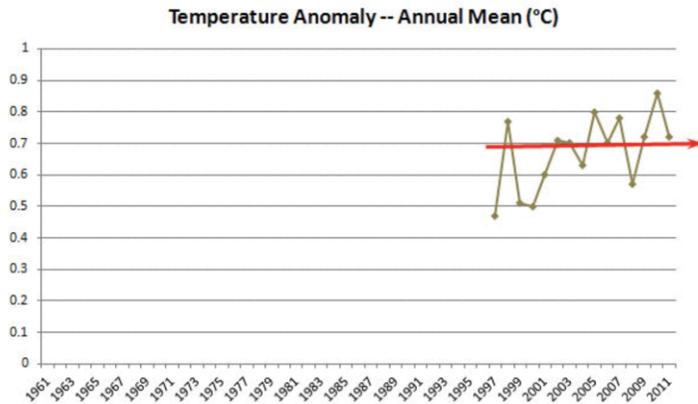
<https://github.com/d3/d3/wiki/Gallery>



# Esempi visualizzazioni ingannevoli



*Dishonest Fox Chart: Bush Tax Cut Edition*



Esempi di rappresentazioni di cattiva qualità: <https://viz.wtf/>  
<https://medium.economist.com/mistakes-weve-drawn-a-few-8cdd8a42d368>

Quest'opera è stata rilasciata con licenza Creative Commons Attribuzione - Non commerciale - Condividi allo stesso modo 4.0 Internazionale.  
Per leggere una copia della licenza visita il sito web <http://creativecommons.org/licenses/by-nc-sa/4.0/>.