

Master Degree in Computer Science

Scientific and Large Data Visualization

*Francesco Caprari:580154
Academic Year: 2023-2024*

Contents

1	Intro	3
1.1	Information Visualization	3
1.1.1	Perchè la visualizzazione?	4
1.2	Fundamentals of 3D Computer Graphics	5
1.3	Scientific Visualization	5
2	Information Visualization	7
2.1	Data	7
2.2	Tipi di attributi	7
2.3	Semantica degli attributi	8
2.4	Grafici Fondamentali (Bivariate data)	8
2.4.1	Bar Charts	8
2.4.2	Histograms	9
2.4.3	Pie Charts	9
2.4.4	Donut Charts	10
2.4.5	Scatter plots	11
2.4.6	Slope Charts	11
2.4.7	Line Charts	12
2.4.8	Area Charts	12
2.4.9	Chropleth maps	12
2.4.10	Symbol maps	13
2.5	Grafici Fondamentali (Attributi Multipli)	13
2.5.1	Stacked Bar Charts	13
2.5.2	Grouped Bar Charts	14
2.5.3	Stacked Line Charts	14
2.5.4	Line Charts Series	15
2.5.5	Bubble Charts	15
2.5.6	Small multiples	16
2.5.7	Symbol Maps	16
2.6	Grafici Fondamentali (Attributi Gerarchici)	17
2.6.1	Sunburst Charts	17
2.6.2	Treempas	17
3	Information Visualization II	19
3.1	Visual Encoding	19
3.2	Graphical Elements	19
3.3	Visual decoding	20
3.3.1	Quality Evaluation	20
3.3.2	Percentuale di Accuracy	21
3.4	Altri tipi di grafici	21
3.4.1	WordClouds	21
3.4.2	Calligrams	22
3.4.3	Flow Maps	22
3.4.4	Chernoff faces	23
3.4.5	Multidimensional Icons	23

3.4.6 Petal as a gliph	24
----------------------------------	----

1 Intro

Information Visualization

- Quando la rappresentazione visiva non è ovvia...
- Percezione visiva, migliori pratiche, dati multidimensionali, grafi e reti

3D Computer Graphics

- Rappresentazione in 3D (mesh poligonali)
- Trasformazione dei dati 3D in immagini generate al computer: rendering, illuminazione, texturizzazione

Scientific Visualization

- Illustrazione grafica dei dati scientifici per estrarre informazioni sui fenomeni

1.1 Information Visualization

Consiste nella Trasformazione dei dati per estrarre informazioni utili, ci sono diverse definizioni:

- La visualizzazione delle informazioni è l'uso di rappresentazioni visive, computerizzate, interattive, di dati astratti per amplificare la cognizione [S. T. Card, 1999].
- Visualizzare significa rendere visibili e comprensibili certi fenomeni e parti della realtà; molti di questi fenomeni non sono naturalmente accessibili all'occhio nudo, e molti di essi non sono neanche di natura visiva [J. Costa, 1998].
- Visualizzazione delle informazioni vs visualizzazione scientifica: Nella visualizzazione scientifica esiste una relazione naturale tra ciò che viene rappresentato e la sua rappresentazione, mentre nella visualizzazione delle informazioni la relazione è convenzionale.

Ci sono due prospettive diverse nell'interpretare l'Information Visualization:

- **Visualizzazione come scienza applicata:** Il valore di una buona visualizzazione consiste nel permetterci di individuare pattern nei dati e quindi la scienza della percezione dei pattern può fornire una base per decisioni di design [C. Ware, Information Visualization. Perception for design, 2013].
- **Visualizzazione come tecnologia:** Gli infografici sono uno strumento visivo per la comunicazione, la comprensione e l'analisi [A. Cairo, The functional art, 2013]. Le limitazioni funzionali formano: come il design di un oggetto tecnologico deve dipendere dal compito che dovrebbe aiutare la forma grafica dovrebbe essere vincolata dalle funzioni della tua presentazione.



Figure 1: Visualizzazione dati di Netflix

1.1.1 Perchè la visualizzazione?

Le visualizzazioni esplicative sono strumenti per presentare informazioni, comunicare dati e messaggi, spiegare qualcosa a qualcun altro, ha tre scopi principali: Explanation, confirmation e exploration.

Siamo una specie visiva. Il sistema visivo umano è molto bravo nell'identificare e analizzare i pattern. Le visualizzazioni sono come artefatti cognitivi (strumenti che gli esseri umani hanno costruito per aiutare a pensare meglio, ad esempio l'abaco). La visualizzazione ha a che fare con la cognizione distribuita (il nostro sistema cognitivo non è esclusivamente composto dal nostro cervello, mente e sensori, ma anche dall'ambiente intorno a noi, che utilizziamo per memorizzare e manipolare informazioni). Le visualizzazioni permettono il trattamento parallelo delle informazioni, anziché sequenziale. Le visualizzazioni esplorative sono strumenti per i lettori per analizzare ciò che viene loro presentato. Molto spesso, l'esito dell'analisi esplorativa non è solo la risposta alle domande originali, ma la generazione di nuove domande.

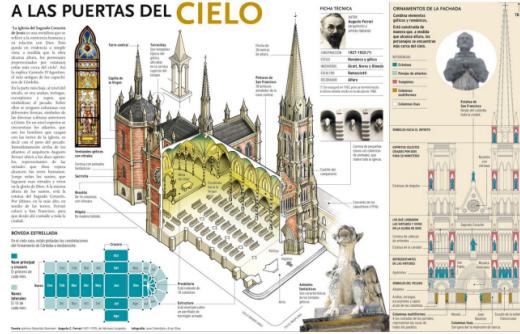


Figure 2: Juan Colombata & Enzo Oliva La Voz del Interior

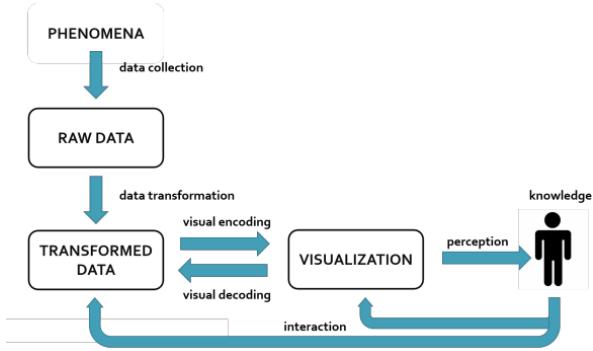


Figure 3: Visualization pipeline

1.2 Fundamentals of 3D Computer Graphics

Animazioni e serious gaming, Progettazione assistita da computer (CAD) e modellazione di prodotti, Patrimonio culturale, Archeologia, Architettura, Fabbricazione digitale e stampa 3D, Biologia e monitoraggio ambientale, Medicina e sistemi di eHealth, Moda del futuro, Sicurezza e difesa...



Figure 4: Esempi di utlizzo dei modelli 3D

1.3 Scientific Visualization

Tecniche informatiche per la generazione di rappresentazioni visuali interattive di dati spazio-temporali acquisiti o simulati (collegamento naturale con il mondo 3D+tempo in cui viviamo). Le illustrazioni a la comunicazione visuale della conoscenza è parte della nostra storia, anche oggi l'incremento della quantità dei dait da analizzare è necessario un approccio grafico per la visualizzazione. La Visualizzazione Scientifica si applica in diversi campi come l'Ingegneria, la medicina E la Scienza. La Visualizzazione Scientifica riguarda anche la definizione di algoritmi efficienti per la manipolazione interattiva ed esplorazione dei dati e delle loro caratteristiche.

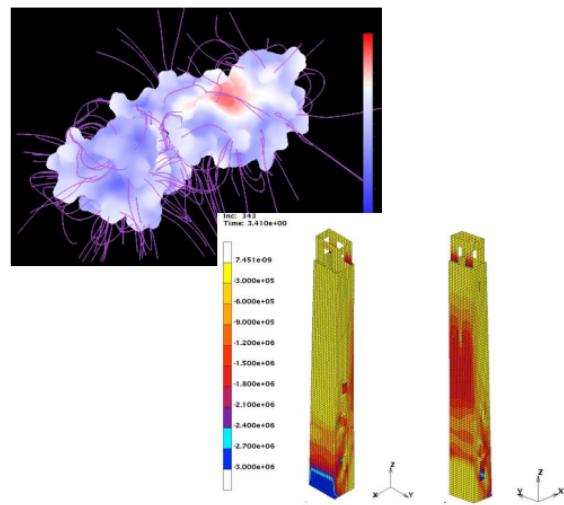


Figure 5: Esempi di Visualizzazione Scientifica

2 Information Visualization

La visualizzazione delle informazioni è un modo per comunicare significativamente i dati e aiutare le persone a dare un senso a grandi quantità di informazioni. Le visualizzazioni come artefatti cognitivi che consentono il trattamento parallelo delle informazioni. Tre scopi principali della visualizzazione:

- **Explanatory:** Le visualizzazioni esplicative sono strumenti per presentare informazioni, comunicare dati e messaggi, spiegare qualcosa a qualcun altro.
- **Exploratory:** Le visualizzazioni esplorative sono strumenti per i lettori per analizzare ciò che viene loro presentato. Molto spesso, l'esito dell'analisi esplorativa non è solo la risposta alle domande originali, ma la generazione di nuove domande.
- **Confirmatory:** Nell'analisi confermativa, le visualizzazioni sono destinate a testare ipotesi.

Bisogna decidere cosa visualizzare

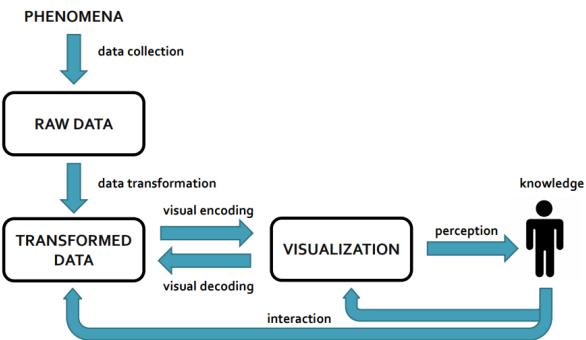


Figure 6: Visualization pipeline

2.1 Data

Informazioni fattuali come misurazioni o statistiche, utilizzate come base per il ragionamento, la discussione o il calcolo. I dati sono collezioni di *items* e *attributes* degli items. Gli *Items* sono gli oggetti/entità che vogliamo visualizzare. Gli *Attributi* sono le proprietà degli oggetti/entità.

Una *Tabella* è una griglia di colonne e righe, dove le righe rappresentano gli item e le colonne gli attributi. Un *network* è una collezione di nodi che rappresentano gli item, connessi tramite dei link; entrambi nodi e link possono avere attributi.

2.2 Tipi di attributi

- *Quantitativi*: sono gli attributi dove i valori rappresentano quantità misurate, questi valori possono essere ordinati, ma può essere calcolata anche la distanza tra i valori.
- *Categorici*: sono gli attributi dove i valori descrivono le categorie, possono essere di tre tipi, **Nominali** se non hanno un ordine particolare, **Ordinali** se possono essere ordinati, **Binary** se hanno solo due stati.

2.3 Semantica degli attributi

- **Spaziali e temporali:** Esempio la location, latitudine e longitudine o la data di assunzione come attributo temporale.
- **Sequnziali, ciclici divergenti:** Esempio: i mesi dell'anno sono ciclici, la temperatura è un attributo divergente.
- **Gerarchici:** Tipi di prodotto con sottocategorie, un esempio sono i vestiti.

Si deve selezionare la visualizzazione appropriata a seconda del tipo e dalla semantica dell'attributo. Esempio:

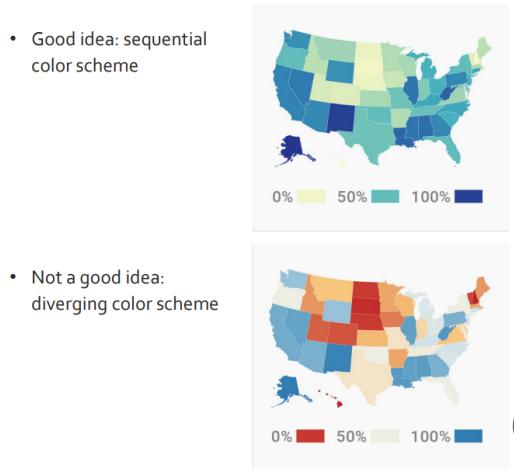


Figure 7: Differenza di Visualizzazione

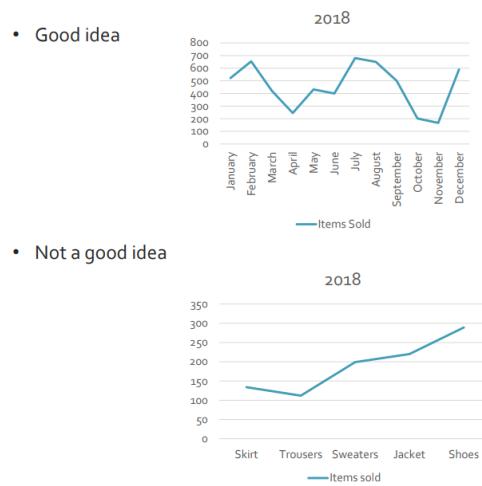


Figure 8: Differenza di Visualizzazione

2.4 Grafici Fondamentali (Bivariate data)

2.4.1 Bar Charts

Visualizza come una quantità misurata si distribuisce tra categorie. Ogni barra rappresenta una categoria e la lunghezza della barra è una quantità misurata in quella categoria. Dati bivariati: nominale/ordinale e quantitativo. Non confondere con gli istogrammi

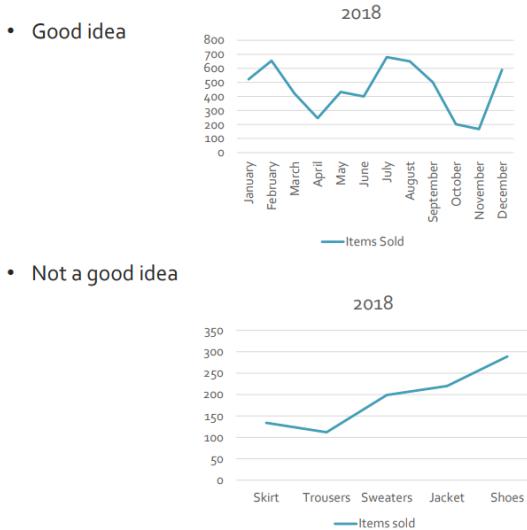


Figure 9: Bar Charts

2.4.2 Histograms

Frequenza degli elementi. Dati bivariati: una variabile indipendente quantizzata in intervalli (blocchi) e una variabile dipendente.

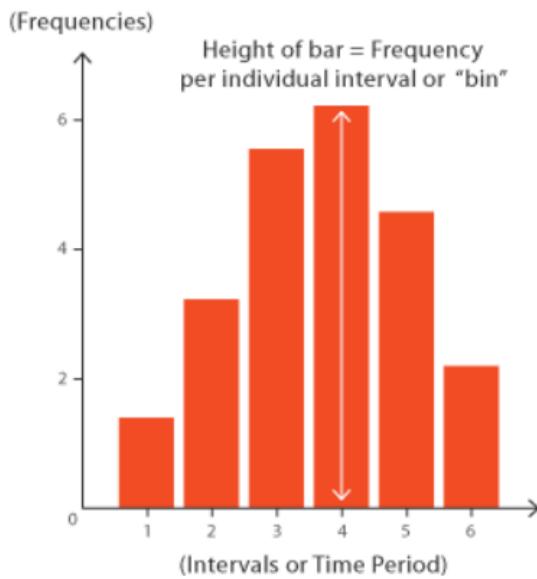
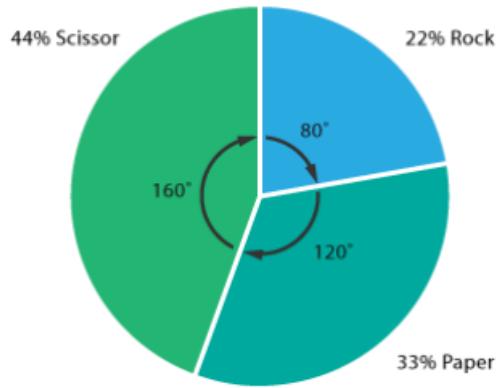


Figure 10: Histograms

2.4.3 Pie Charts

Mostrare proporzioni e percentuali tra le categorie. Utile per dare un'idea rapida e confrontare una fetta rispetto al totale, non adatto per confronti accurati. Altri svantaggi: numero limitato di valori, occupazione dello spazio.



Data			
Rock	Paper	Scissor	TOTAL
2	3	4	9
To calculate percentages			
$2/9=22\%$	$3/9=33\%$	$4/9=44\%$	100%
Degrees for each "pie slice"			
$(2/9) \times 360 = 80^\circ$	$(3/9) \times 360 = 120^\circ$	$(4/9) \times 360 = 160^\circ$	360°

Figure 11: Pie Charts

2.4.4 Donut Charts

Grafici a torta con l'area centrale tagliata. Maggiore enfasi sulla lunghezza dell'arco rispetto all'area. Più efficienti in termini di spazio.

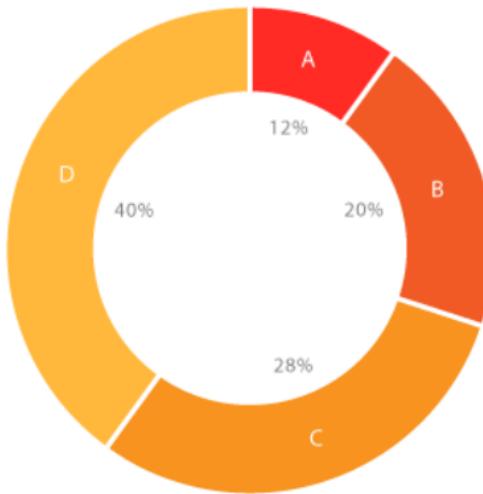


Figure 12: Donut Charts

2.4.5 Scatter plots

Relazione tra attributi: visualizzazione di come una quantità è correlata a un'altra (e analisi di cluster, valori anomali, ecc.). Dati bivariati: due attributi quantitativi indipendenti.

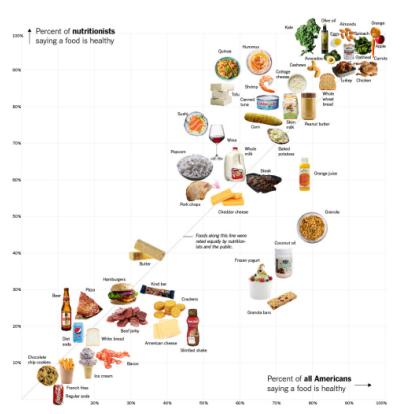


Figure 13: Scatter plots

2.4.6 Slope Charts

Alternativa ai Scatter plots: assi paralleli, ogni elemento è una linea che collega due quantità.

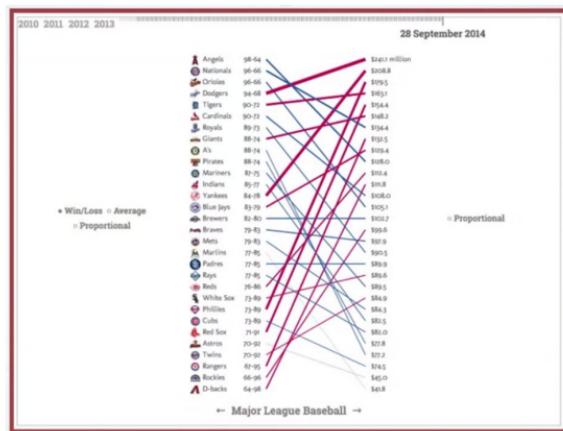


Figure 14: Slope Charts

2.4.7 Line Charts

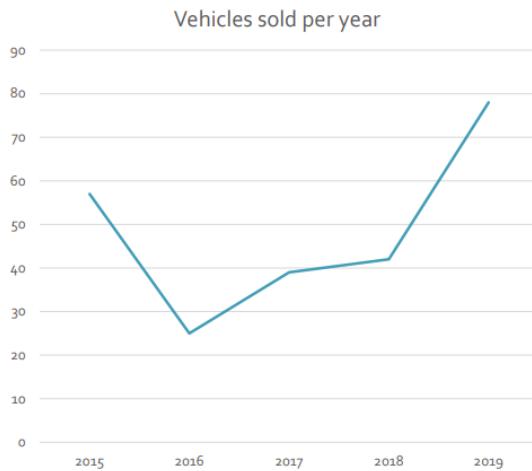


Figure 15: Line Charts

2.4.8 Area Charts

Line Charts dove l'area sotto alla linea è riempita.

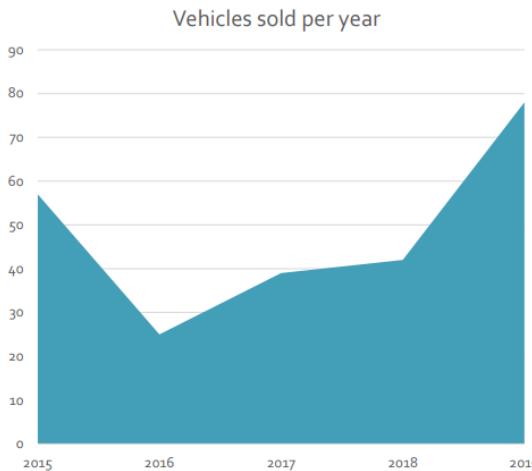


Figure 16: Area Charts

2.4.9 Chropleth maps

Come si distribuisce una quantità nelle diverse aree/geografiche/regioni. Colori, sfumature, pattern sono utilizzati per rappresentare la quantità associata alle aree/regioni. Buona panoramica (ma non confronto accurato). Rischio: confondere l'area geografica con i valori dei dati (prestare attenzione alla normalizzazione).

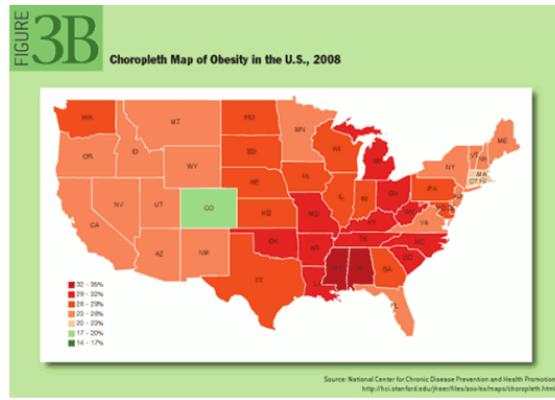


Figure 17: Choropleth maps

2.4.10 Symbol maps

Come si distribuisce una quantità lungo due coordinate spaziali. Un simbolo (spesso un disco o un quadrato) è posizionato in un punto e dimensionato in modo che la sua area sia proporzionale alla quantità associata al punto.

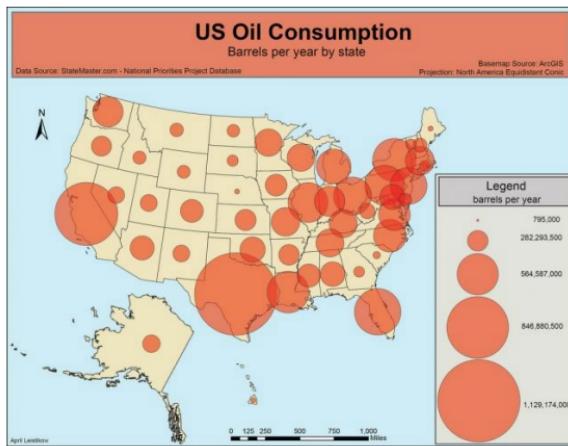


Figure 18: Symbol maps

2.5 Grafici Fondamentali (Attributi Multipli)

2.5.1 Stacked Bar Charts

Rappresentazione di attributi aggiuntivi. Tante barre quanti sono le categorie del primo attributo, e altrettanti segmenti all'interno di ciascuna barra quanti sono i valori dell'altro attributo categorico. Utile per comunicare la proporzione

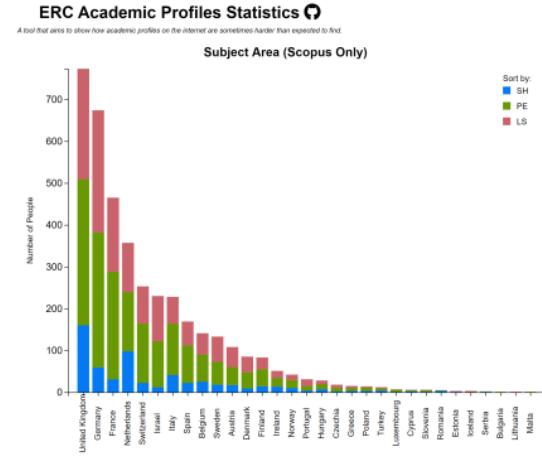


Figure 19: Stacked Bar Charts

2.5.2 Grouped Bar Charts

Rappresentazione di attributi aggiuntivi. Lo stesso grafico a barre viene ripetuto più volte per il numero di categorie nell'attributo aggiuntivo. Utile per confrontare i valori individuali

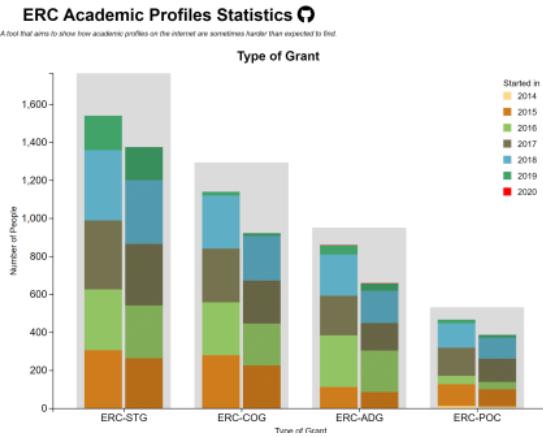
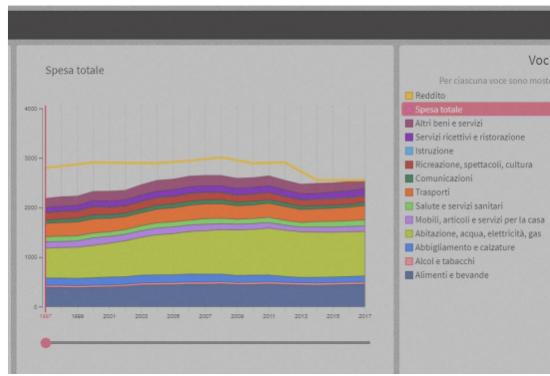


Figure 20: Grouped Bar Charts

2.5.3 Stacked Line Charts

I valori tra le categorie sono impilati uno sopra l'altro. Non è ideale per leggere e confrontare i valori individuali nel tempo: la forma delle linee è influenzata dalla forma delle linee sottostanti (interazione/filtraggio per compensare). Non è significativo per i numeri negativi e dati che non dovrebbero essere sommati (ad esempio, temperatura).



2.5.4 Line Charts Series

Raccolta di linee nello stesso grafico. Facile confrontare i valori, ma le curve sovrapposte possono rendere il grafico difficile da leggere

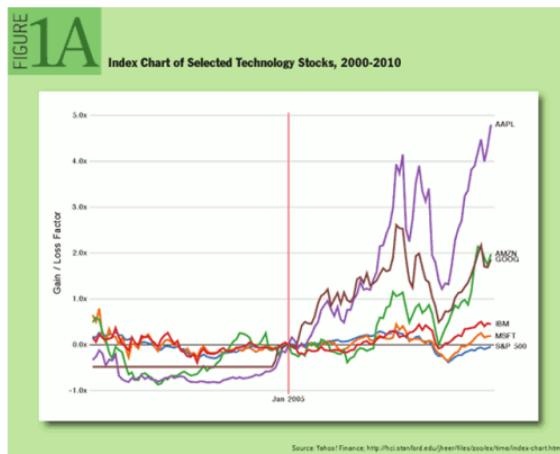


Figure 21: Line Charts Series

2.5.5 Bubble Charts

Variante degli scatterplot, adattata per rappresentare attributi aggiuntivi. Le aree dei cerchi rappresentano un attributo aggiuntivo (avviso: area, non raggio). I colori possono distinguere le categorie o rappresentare attributi aggiuntivi. Capacità limitata di dimensioni dei dati (interazione per compensare).

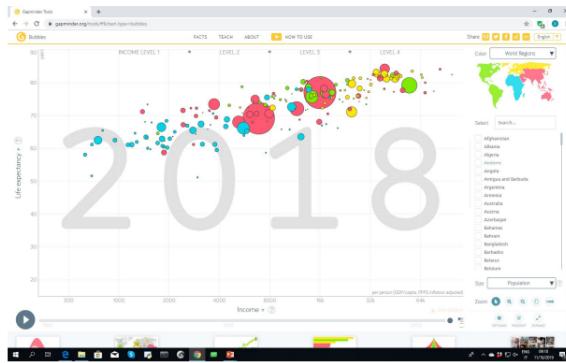


Figure 22: Bubble Charts

2.5.6 Small multiples

Valido per tutti i tipi di grafici.

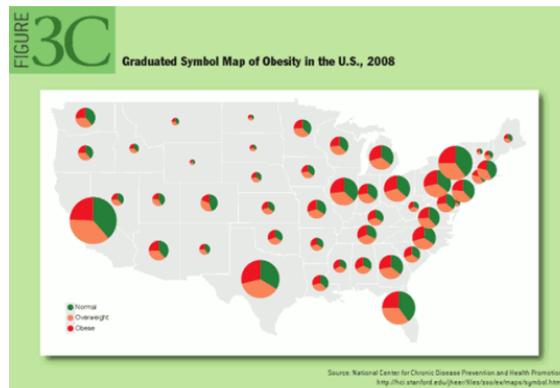


Figure 23: Small multiples

2.5.7 Symbol Maps

Questo tipo di mappa utilizza simboli, come icone o forme, per visualizzare informazioni diverse relative a diverse caratteristiche o attributi su una mappa geografica.

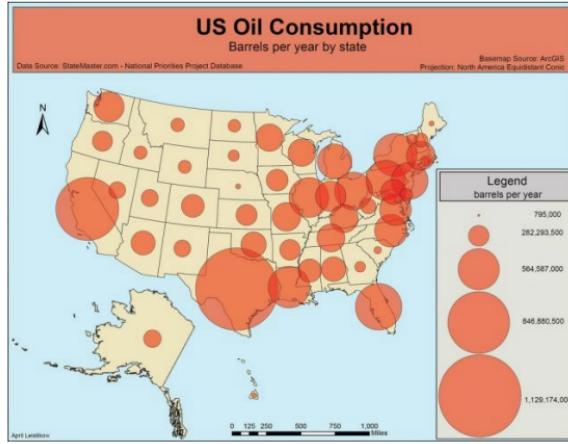


Figure 24: Symbol Maps

2.6 Grafici Fondamentali (Attributi Gerarchici)

2.6.1 Sunburst Charts

Per mostrare gerarchie, ogni anello rappresenta un livello nella gerarchia, muovendosi verso l'esterno.

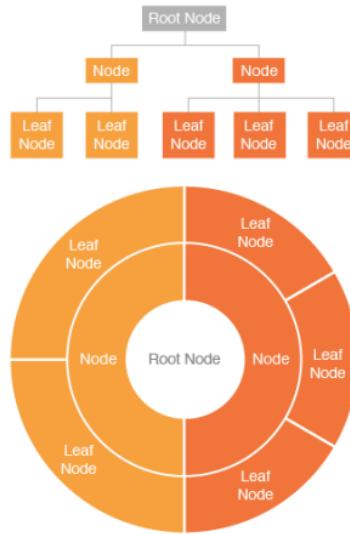


Figure 25: Sunburst Charts

2.6.2 Treempas

Un modo alternativo per visualizzare le gerarchie, utilizzando la dimensione dell'area. Ogni rettangolo rappresenta una categoria, con rettangoli sottocategorie nidificati. Quando una quantità è assegnata a una categoria, la sua dimensione dell'area viene visualizzata in proporzione a quella quantità e alle altre quantità all'interno della stessa categoria padre in una relazione parte-tutto. Diverse strategie di impaginazione (ad esempio, algoritmo squarified) Compatto e efficiente nello spazio, ottimo per avere una panoramica.

delle strutture, perfetto per confrontare proporzioni tramite la dimensione dell'area, ma meno efficace nel mostrare i livelli nella gerarchia rispetto, ad esempio, ai grafici a raggiera (sunburst charts).

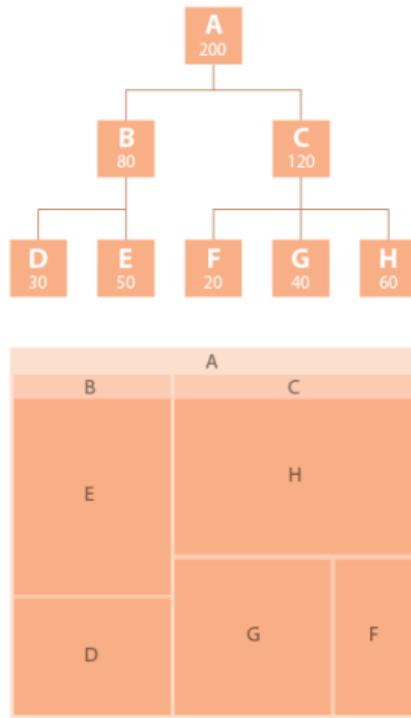


Figure 26: Treempas

3 Information Visualization II

3.1 Visual Encoding

Le visualizzazioni sono composte da **marks** (segnali) e **channels** (canali). I marks sono gli oggetti grafici che rappresentano gli elementi di dati (ad esempio, punti, linee, barre). I channels sono le proprietà grafiche che rappresentano gli attributi dei dati (ad esempio, colore, posizione, forma, dimensione).

La codifica visiva significa passare dai dati alle rappresentazioni visive. La codifica richiede non solo la scelta di un grafico appropriato per i dati in questione (tipo e semantica), ma anche la selezione degli elementi grafici individuali e delle loro proprietà.

3.2 Graphical Elements

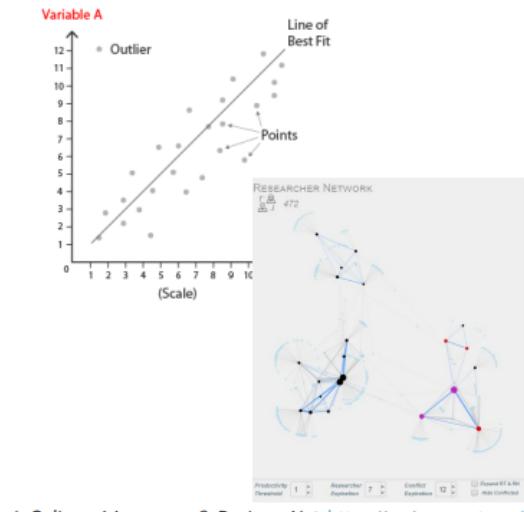


Figure 27: Esempio di Marks

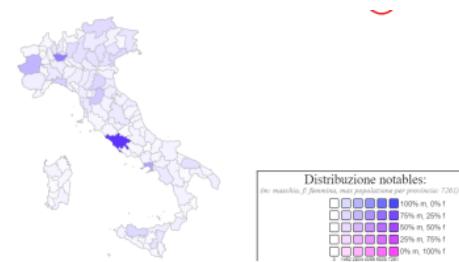


Figure 28: Esempio di Channels

Le componenti contestuali sono elementi che rendono più semplice interpretare le visualizzazioni. Un esempio sono: le labesls, annotazioni, legende, griglie, assi cartesiani.

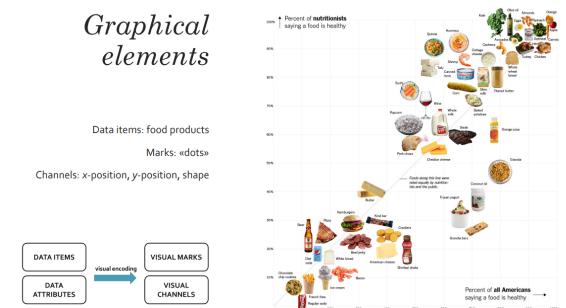


Figure 29: Esempio completo di Channels e Marks di un grafico

3.3 Visual decoding

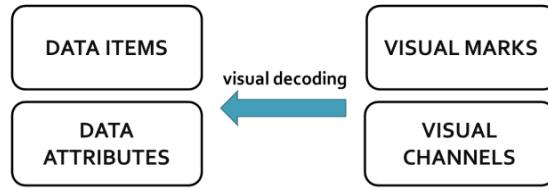


Figure 30: Visual Decoding

Per **Visual Decoding** si intende destrutturare una rappresentazione visiva nei suoi principali elementi e identificare:

- gli elementi grafici, quali sono i segnali visivi? Quali sono i canali visivi?
- le regole di mappatura (ossia, le informazioni che i singoli elementi grafici rappresentano) quali elementi di dati rappresentano i segnali? Quali attributi rappresentano i canali?

È utile per valutare e ridisegnare le visualizzazioni.

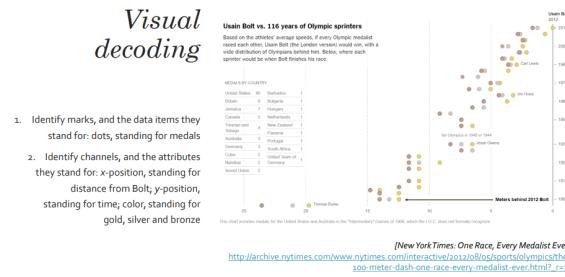


Figure 31: Esempio completo di Decoding

3.3.1 Quality Evaluation

La valutazione di una visualizzazione può basarsi su due principi guida principali: **Expressiveness** ed **Effectiveness**. **Expressiveness**: la Visual representation dovrebbe rappresentare tutte e solo le relazioni che esistono nei dati. Le informazioni rilevanti dovrebbero essere prioritarie e quindi codificate con i canali più efficaci/accurati.

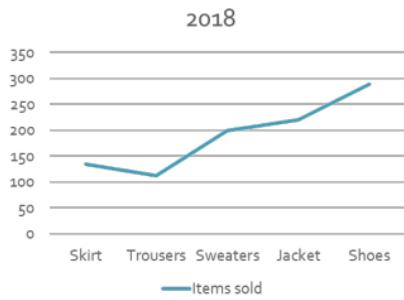


Figure 32: Dati non ordinati che sembrano ordinati (nell'esempio seguente, stiamo mostrando informazioni su un trend che non è nei dati: la forma della linea non ha significato).

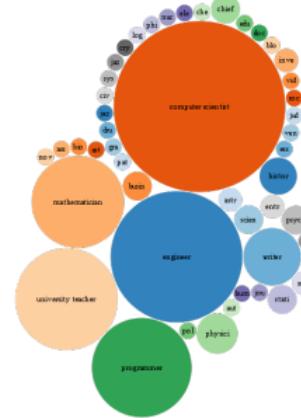


Figure 33: utilizzo dei colori quando non restituiscono alcuna informazione

3.3.2 Percentuale di Accuracy

Quanto sono efficaci i canali nel trasmettere diversi tipi di attributi? Uno dei (possibili) riassunti per attributi quantitativi.

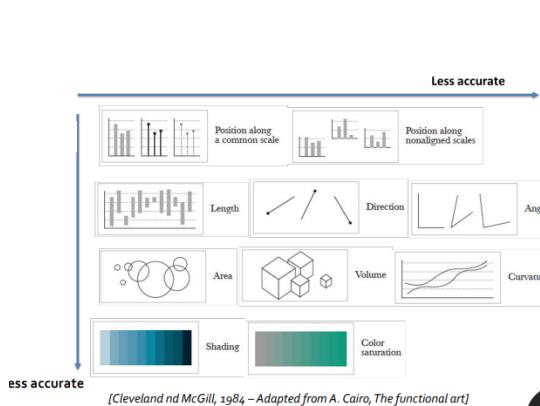


Figure 34: Riassunto dell'accuracy di ogni grafico.

QUANTITATIVE	ORDINAL	CATEGORICAL
Position	Position	Position
Length	Density	Hue
Angle	Saturation	Texture
Slope	Hue	Connection
Area	Texture	Containment
Volume	Connection	Density
Density	Containment	Saturation
Saturation	Length	Shape
Hue	Angle	Length
	Slope	Angle
	Area	Slope
	Volume	Area
		Volume

[Data Visualization Course by John C. Hart, for Coursera, 2015]

Figure 35: Riassunto dell'accuracy di ogni grafico

3.4 Altri tipi di grafici

3.4.1 WordClouds

Rappresentazione di quanto frequentemente le parole appaiono in un determinato corpo di testo attraverso la dimensione della parola. Variazioni nell'arrangiamento e nel colore. Principalmente utilizzato per motivi estetici.



Figure 36: WordClouds

3.4.2 Calligrams

Testi disposti in modo tale da formare un'immagine tematicamente correlata. L'immagine creata dalle parole illustra il testo esprimendo visivamente qualcosa associato (o in contrasto) a ciò che il testo dice.



Figure 37: Calligrams

3.4.3 Flow Maps

Raffigurare il movimento delle entità posizionando linee tracciate sopra delle mappe (spazio e tempo) Lo spessore, il colore, ecc. possono codificare informazioni aggiuntive In questa mappa: dimensioni delle truppe, distanza percorsa, temperatura, latitudine e longitudine, direzione del viaggio, tempo.

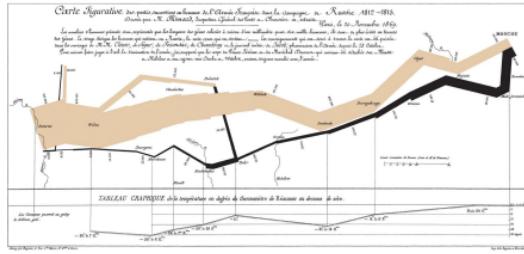


Figure 38: Flow Maps

3.4.4 Chernoff faces

Ragionamento: siamo molto bravi nel riconoscere i volti. Introdotto da Herman Chernoff nel 1973 Variabili sono mappate su tratti del viso (larghezza/curvatura della bocca, dimensione verticale del viso, dimensione/inclinazione/separazione degli occhi, dimensione delle sopracciglia, posizione verticale delle sopracciglia...)

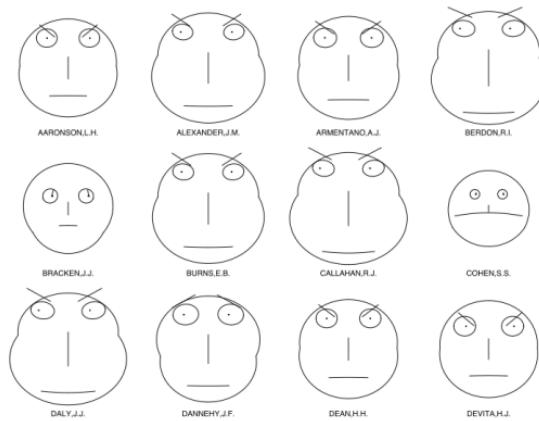


Figure 39: Chernoff faces

3.4.5 Multidimensional Icons

Spence e Parr (1991) proposero di codificare le proprietà di un oggetto in una semplice rappresentazione tramite icone. Hanno applicato questo approccio per verificare le offerte di permanenza.

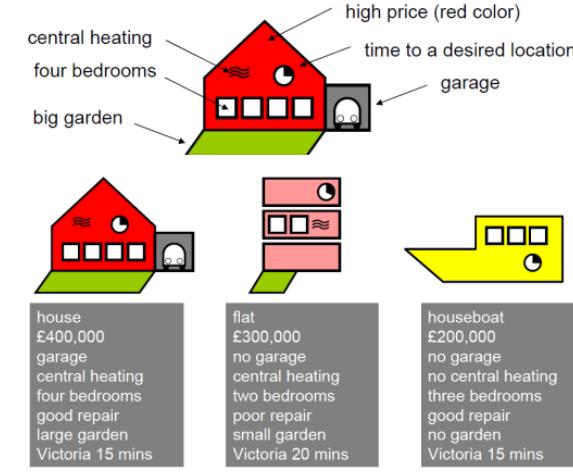


Figure 40: Multidimensional Icons

3.4.6 Petal as a gliph

L'idea di Moritz Stefaner per visualizzare un indice di vita consiste nel mappare diverse variabili (relative alle condizioni di vita materiale e alla qualità della vita) in petali di diverse dimensioni, per confrontare il benessere tra paesi.

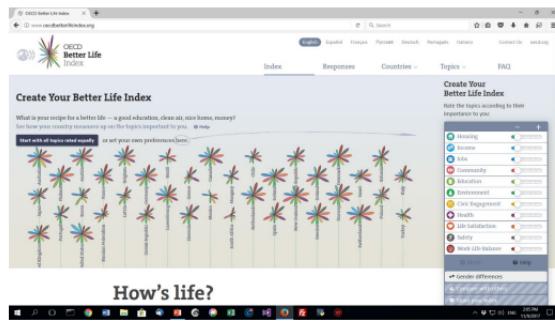


Figure 41: Petal as a gliph