



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

IA PER LA MODA

Data Visualization

Alessia Angeli

Studente di dottorato in Data Science and Computation
Dipartimento di Informatica – Scienza e Ingegneria



CV (in breve)

- Studente di dottorato in Data Science and Computation
- Laurea Magistrale in Matematica, curriculum generale/applicativo
- Laurea Triennale in Matematica



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

VARLAB: VIRTUAL AND AUGMENTED REALITY LAB

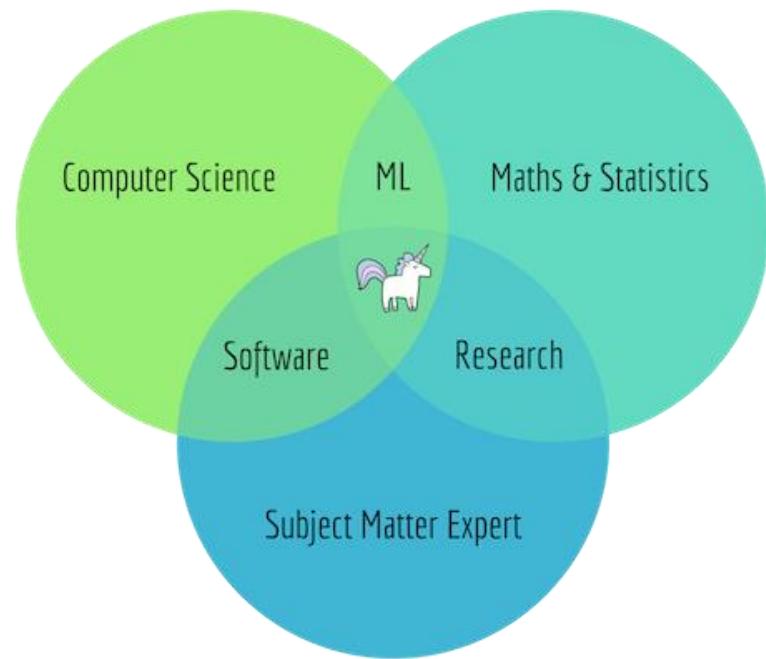


Contatti

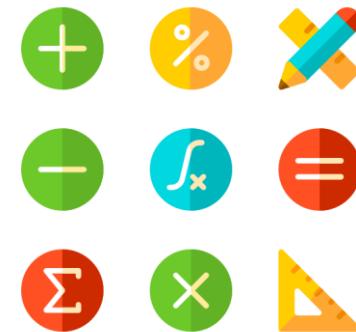


ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

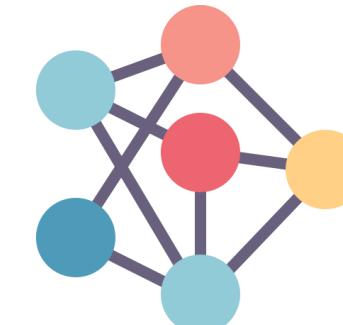
DATA SCIENCE



Maths



Machine and Deep Learning



Data Visualization



Augmented Reality





ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA



Data Visualization – Introduzione

Alessia Angeli

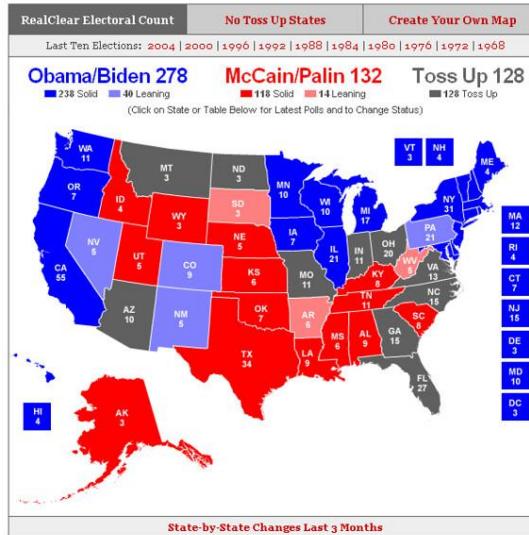
Studente di dottorato in Data Science and Computation
Dipartimento di Informatica – Scienza e Ingegneria

Introduzione

La visualizzazione dati è:

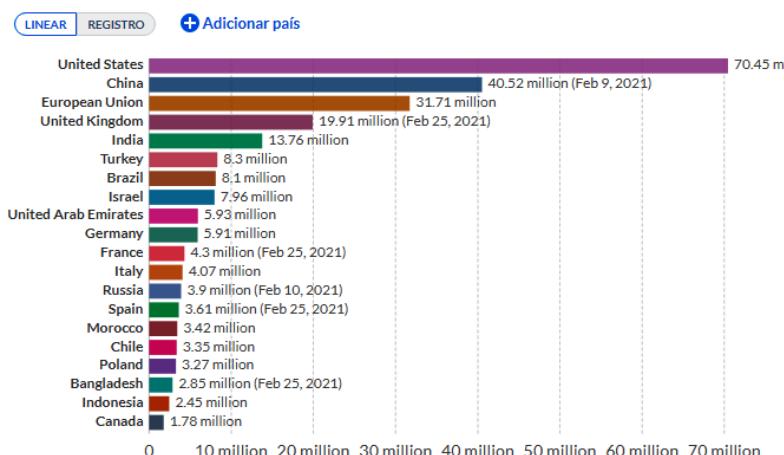
- difficile da valutare ed è facile commettere errori;
- altamente dipendente dal problema/dominio considerato.

RealClearPolitics Electoral College
(270 Electoral Votes Needed To Win)



Doses de vacina COVID-19 administradas, 26 de fevereiro de 2021

Número total de doses de vacinação administradas. Isso é contado como uma dose única e pode não ser igual ao número total de pessoas vacinadas, dependendo do regime de dose específico (por exemplo, as pessoas recebem doses múltiplas).

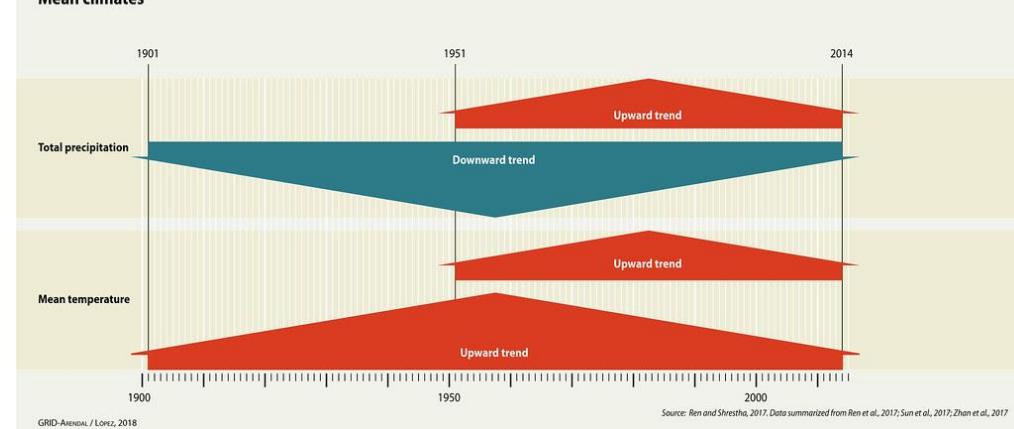


Fonte: dados oficiais coletados por Our World in Data - última atualização em 27 de fevereiro, 10:40 (horário de Londres)

CC BY

Observed climate changes over the HKH

Mean climates



<https://www.flickr.com/>

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Facciamo un passo indietro... Perché visualizzare?

- La visualizzazione è un canale molto efficace per il cervello umano.

ESEMPIO

- **Gioco del 15** ([EuroVis'09 keynote](#))

ci sono 2 giocatori

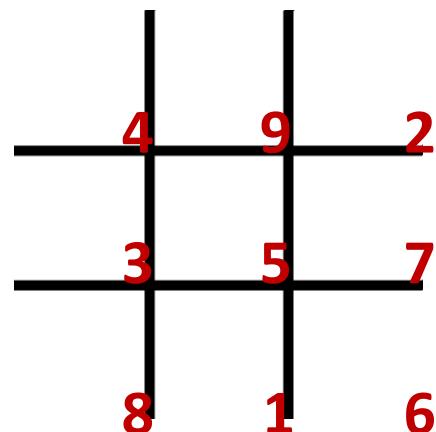
ogni giocatore sceglie a turno una cifra da 1 a 9

una volta che una cifra è stata scelta questa non può più essere usata da nessun giocatore

si continua a giocare

il primo giocatore che arriva ad avere tre cifre la cui somma è 15 vince

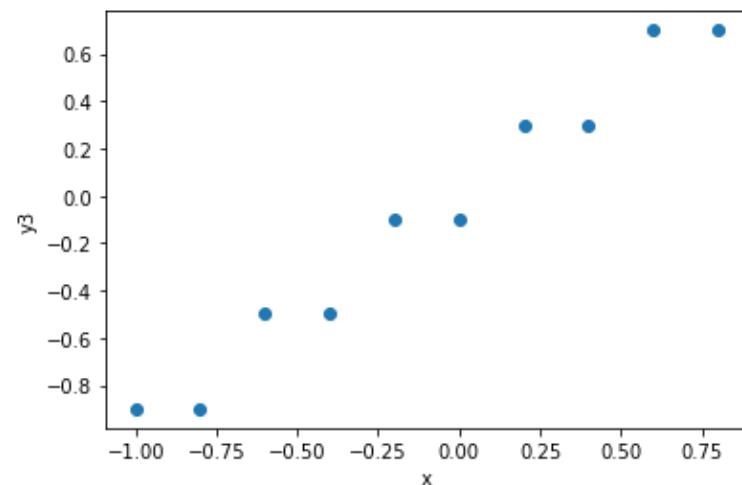
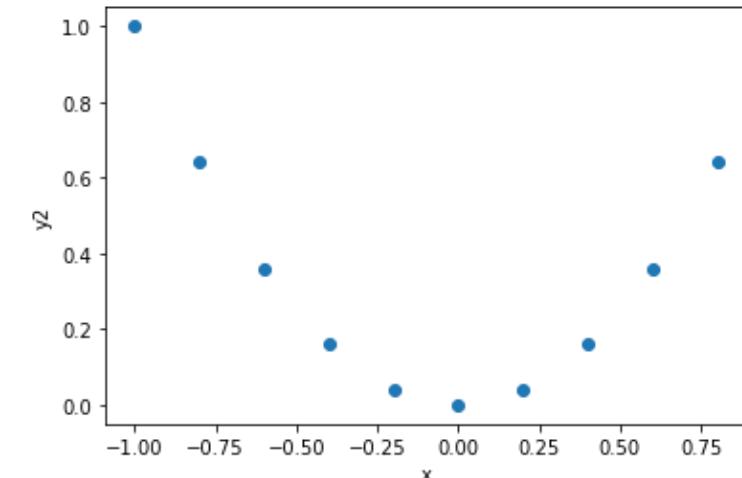
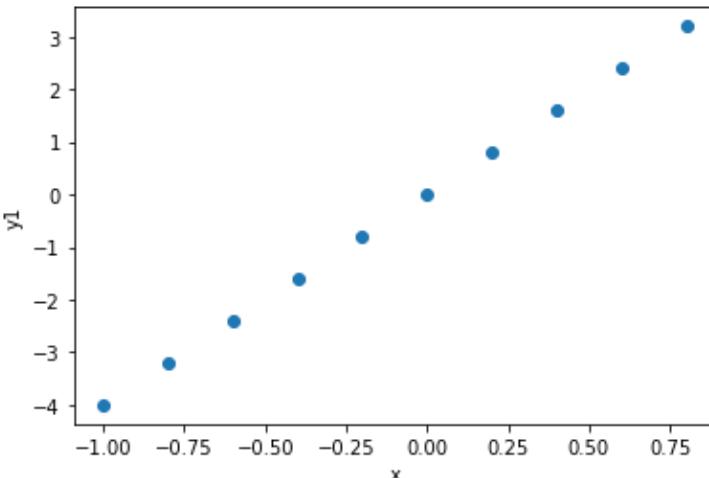
- Vi è familiare questo gioco?



Perché visualizzare?

	x	y1	y2	y3
0	-1.00	-4.00	1.00	-0.90
1	-0.80	-3.20	0.64	-0.90
2	-0.60	-2.40	0.36	-0.50
3	-0.40	-1.60	0.16	-0.50
4	-0.20	-0.80	0.04	-0.10
5	-0.00	-0.00	0.00	-0.10
6	0.20	0.80	0.04	0.30
7	0.40	1.60	0.16	0.30
8	0.60	2.40	0.36	0.70
9	0.80	3.20	0.64	0.70

	x	y1	y2	y3
count	10.00	10.00	10.00	10.00
mean	-0.10	-0.40	0.34	-0.10
std	0.61	2.12	0.33	0.00
min	-1.00	-4.00	0.00	-0.90
25%	-0.55	-2.20	0.07	0.50
50%	-0.10	0.40	0.20	0.10
75%	0.25	1.40	0.57	0.20
max	0.80	3.20	1.00	0.70



Come creare una visualizzazione efficace?

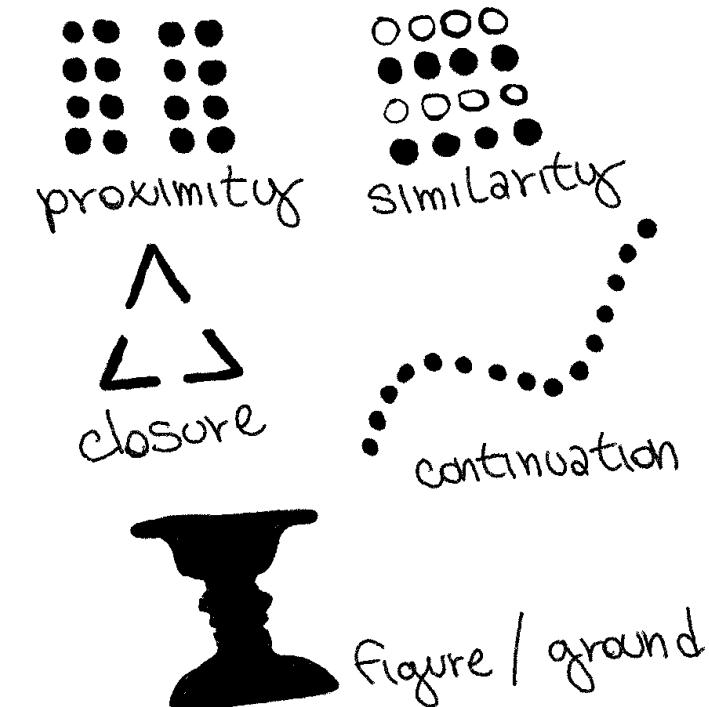
- Comprendere (almeno un po') la **percezione umana**;
- Definire gli **obiettivi**;
- Esplorare lo **spazio delle possibili visualizzazioni**.



L'efficacia e la percezione – I principi della Gestalt

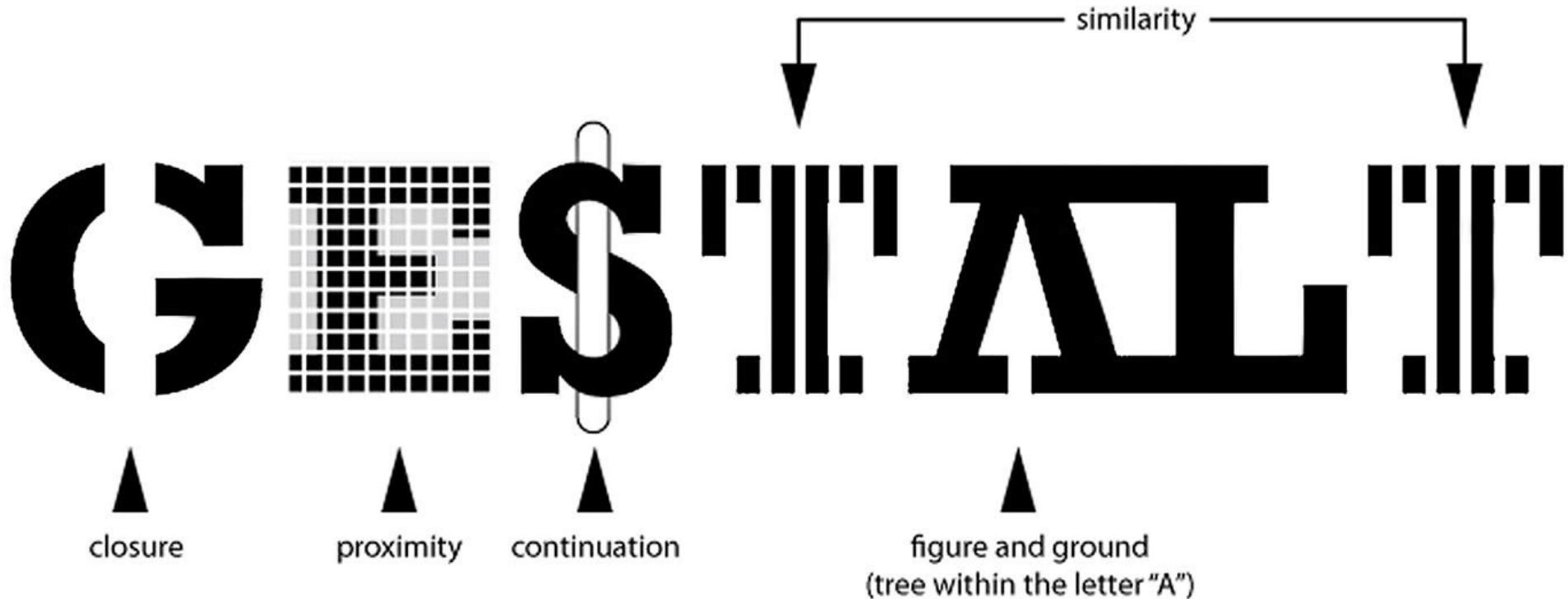
I principi della Gestalt:

1. Principio di **prossimità** o vicinanza;
2. Principio di **similarità** o somiglianza;
3. Principio di **chiusura** o completamento;
4. Principio di **continuità** di direzione;
5. Principio di **contrasto** (figura/sfondo).

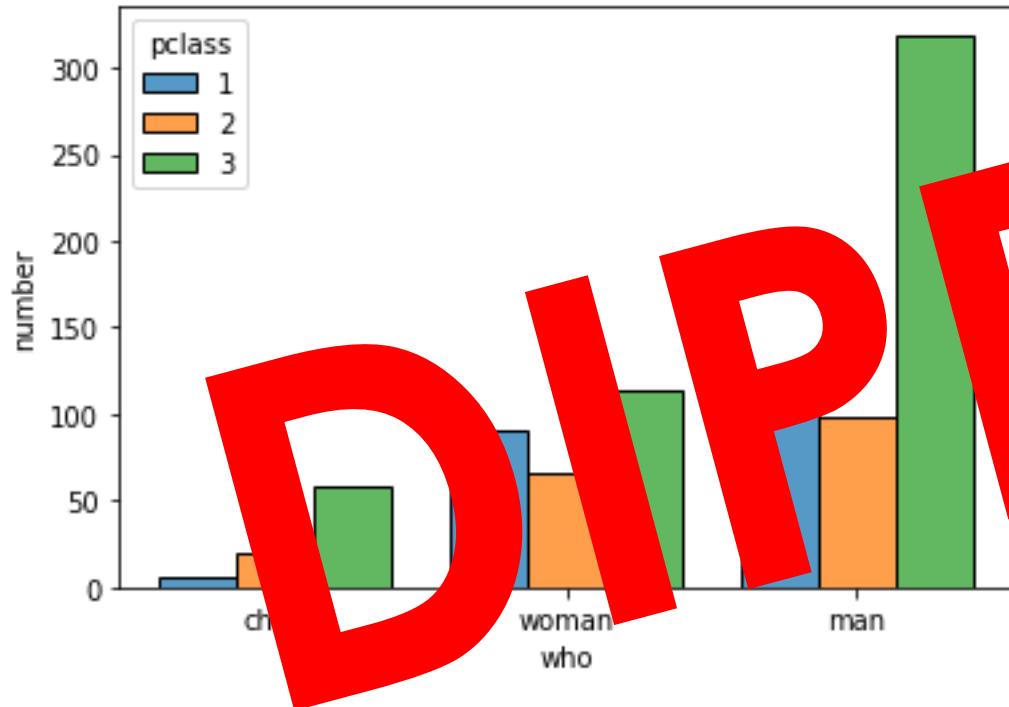


- Gestalt = «pattern» dal tedesco.
- I principi della Gestalt descrivono come la nostra mente individua/percepisce gli elementi all'interno di un gruppo.
- Utilizzare i principi della Gestalt in visualizzazione può portare miglioramenti nell'efficacia del risultato.

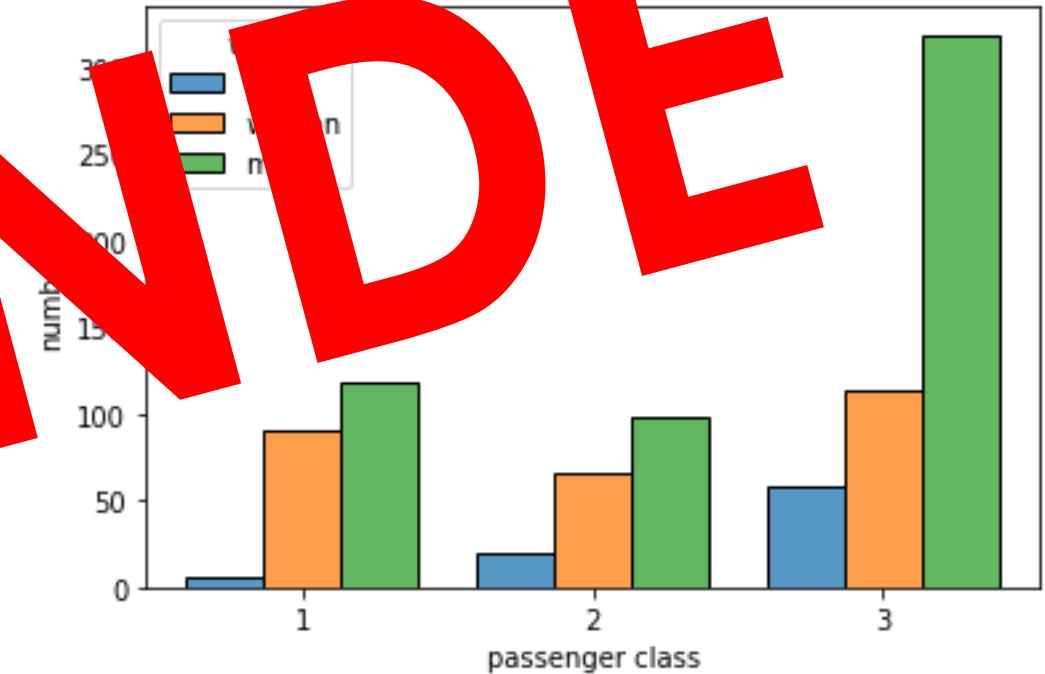




L'efficacia e gli obiettivi – Quale visualizzazione è più efficace?



DIPENDE



L'efficacia di una visualizzazione può essere misurata
solo in relazione agli obiettivi e al problema considerato.

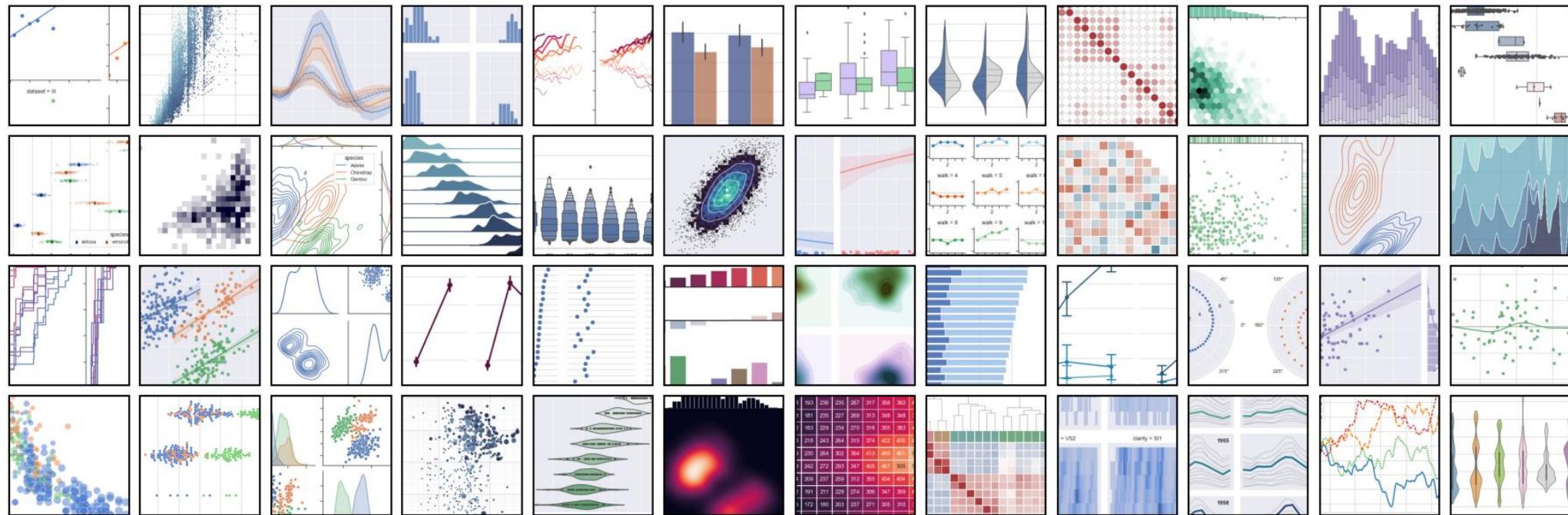


L'efficacia e lo spazio delle possibili visualizzazioni

Lo spazio delle possibili visualizzazioni è molto esteso.

Infatti, se si vuole creare una visualizzazione, subito ci si trova a dovere rispondere a domande come:

- Quale tipo di visualizzazione scegliere?
- Combinare questa con altre visualizzazioni? Quali? Come?
- Creare una visualizzazione interattiva? In che modo?



Per una visualizzazione efficace – Domande importanti

QUALI dati devono essere visualizzati?

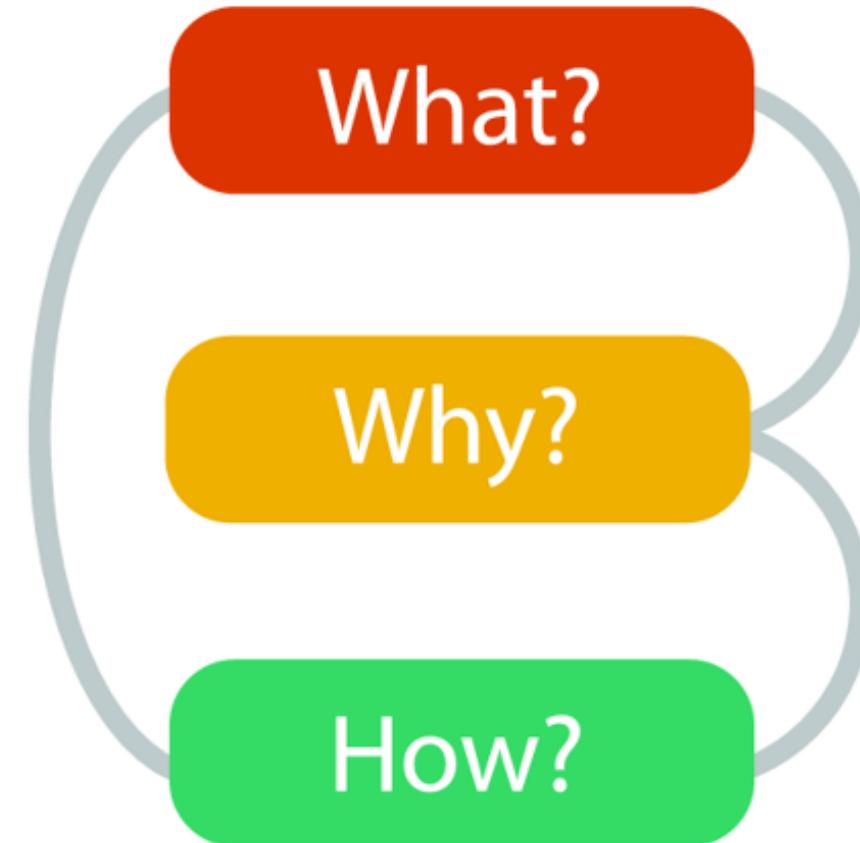
- Tipo di dati.

PERCHE' questi dati devono essere visualizzati?

- Chi saranno gli utenti;
- Qual è il problema studiato;
- Qual è l'obiettivo da raggiungere.

COME questi dati devono essere visualizzati?

- Che tipo di grafico utilizzare.



Il processo di progettazione è solitamente
un **perfezionamento iterativo**.



WHAT? – Tipo di DATI

Possiamo identificare 5 principali tipi di dati:

- **Items:** dati con valori discreti;
(e.g., colori occhi -> marroni, verdi, azzurri, ...)
- **Valori:** dati con valori continui;
(e.g., lunghezza capelli -> 10cm, 14.5cm, ...)
- **Links:** relazioni tra items;
(e.g., altezza -> meno alto, più alto, alto uguale)
- **Attributi:** valori che misurano proprietà di items, links o insiemi di questi;
(e.g., esempi precedenti per persona)
- **Posizioni:** valori spaziali.
(e.g., latitudine e longitudine)



WHAT? – Tipo di ATTRIBUTI

➔ Attribute Types

→ Categorical



(e.g., colore occhi, colore capelli)

→ Ordered

→ Ordinal



(e.g., giudizi scolastici)

→ Quantitative



(e.g., voti scolastici)

➔ Ordering Direction

→ Sequential



(e.g., peso)

→ Diverging



(e.g., temperatura in °C)

→ Cyclic



(e.g., giorni della settimana)



WHAT? – Tipo di ATTRIBUTI

survived	pclass	sex	age	sibsp	parch	fare	embarked	class	who	adult_male	deck	embark_town	alive	alone
0	0	3	male	22.0	1	0	7.2500	Third	man	True	NaN	Southampton	no	False
1	1	1	female	38.0	1	0	71.2833	First	woman	False	C	Cherbourg	yes	False
2	1	3	female	26.0	0	0	7.9250	Third	woman	False	NaN	Southampton	yes	True
3	1	1	female	35.0	1	0	53.1000	First	woman	False	C	Southampton	yes	False
4	0	3	male	35.0	0	0	8.0500	Third	man	True	NaN	Southampton	no	True
5	0	3	male	NaN	0	0	8.4583	Third	man	True	NaN	Queenstown	no	True
6	0	1	male	54.0	0	0	51.8625	First	man	True	E	Southampton	no	True
7	0	3	male	2.0	3	1	21.0750	Third	child	False	NaN	Southampton	no	False
8	1	3	female	27.0	0	2	11.1333	Third	woman	False	NaN	Southampton	yes	False
9	1	2	female	14.0	1	0	30.0708	Second	child	False	NaN	Cherbourg	yes	False

Ordered - Ordinal

Ordered - Quantitative

Categorical



WHY? – L'astrazione del problema: dai dati alla visualizzazione

Individuare ed analizzare:

area di interesse



pubblico
(utenti)



problema

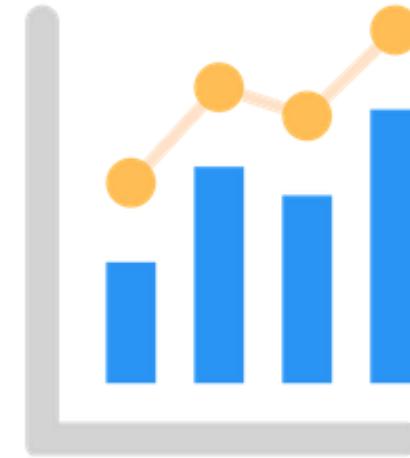
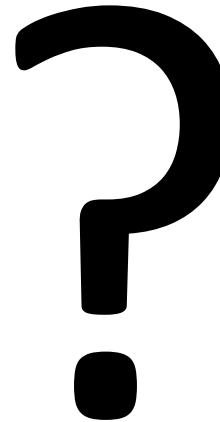
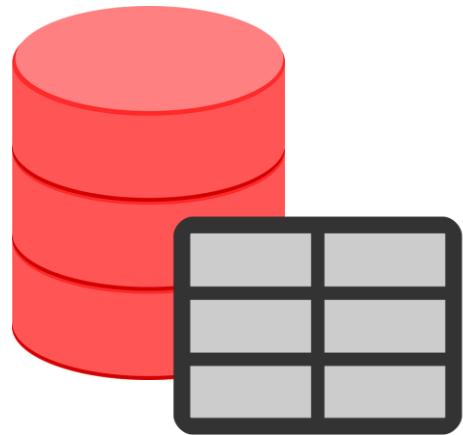


obiettivi da
raggiungere



HOW? – La visualizzazione dati

Come utilizzare questi strumenti per visualizzare al meglio dati di interesse.





ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA



Data Visualization – Elementi e Principi

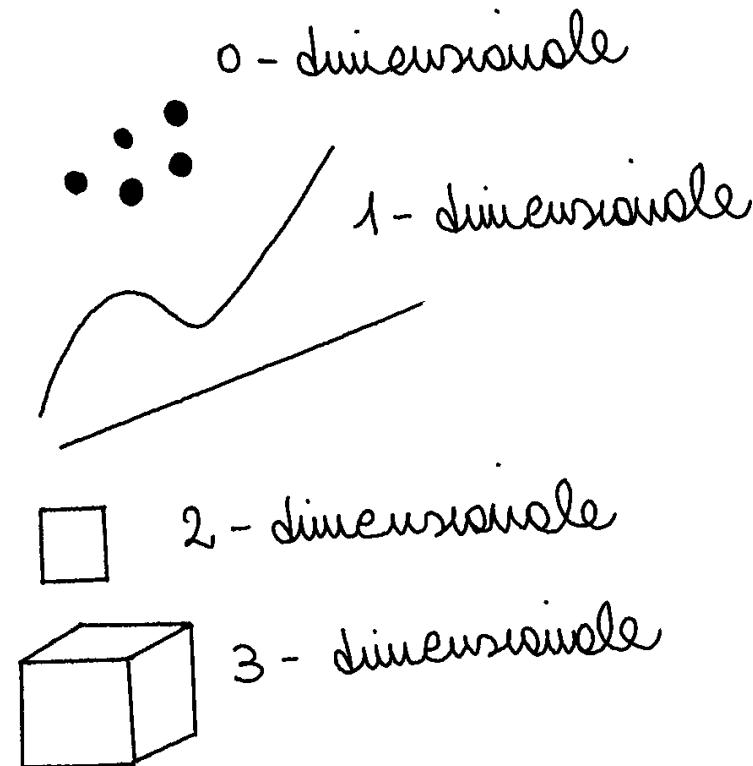
Alessia Angeli

Studente di dottorato in Data Science and Computation
Dipartimento di Informatica – Scienza e Ingegneria

Elementi di visualizzazione - MARKS

Con **marks** (segni) ci si riferisce agli oggetti primitivi che ci permettono di creare un grafico:

- Punti
- Linee
- Aree
- Volumi



Elementi di visualizzazione - CHANNELS

In visualizzazione i **channels** (canali) sono variabili che permettono di controllare l'apparenza dei marks.

④ Position

→ Horizontal



→ Vertical



→ Both



④ Color



④ Shape



④ Tilt



④ Size

→ Length



→ Area

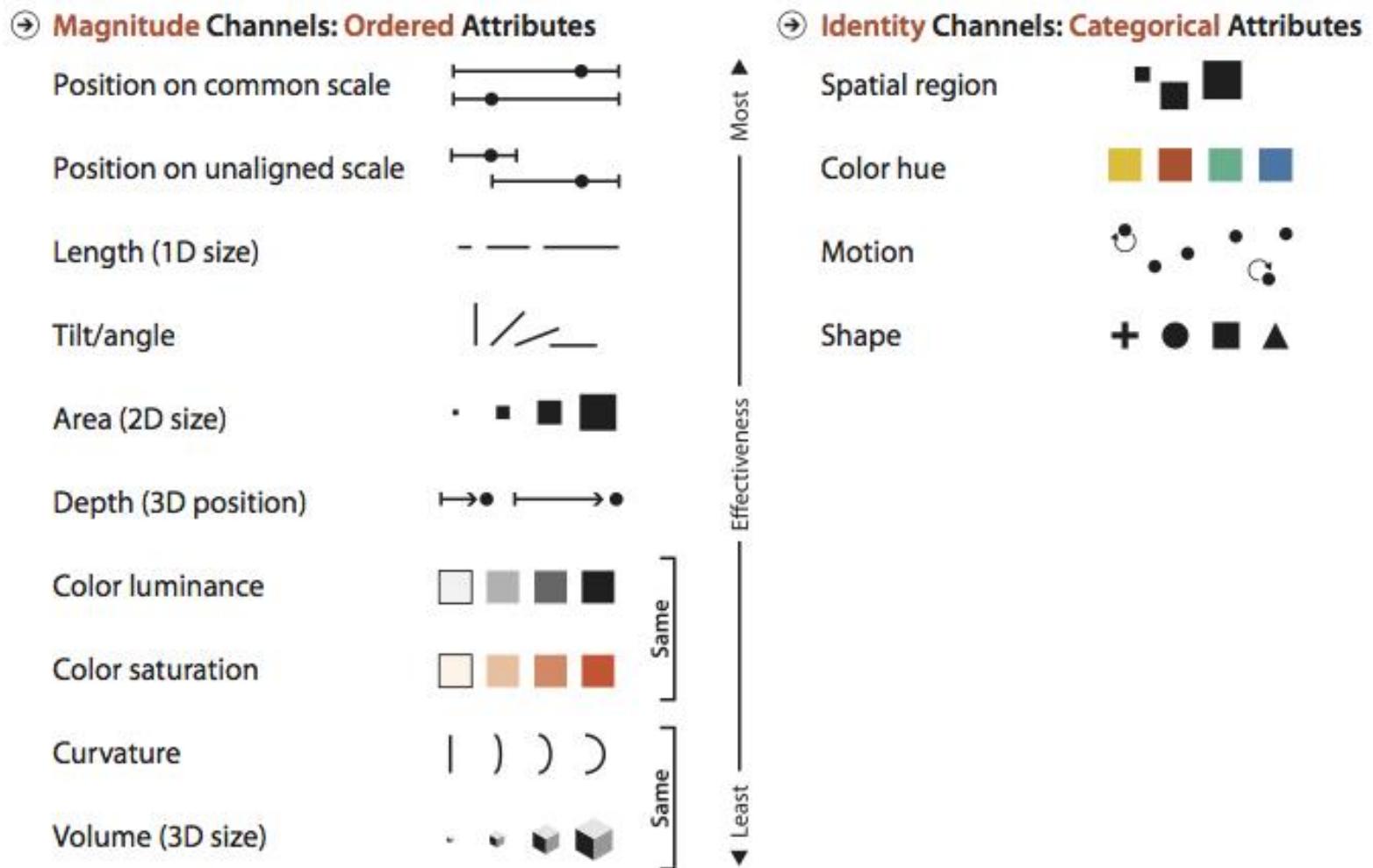


→ Volume



Elementi di visualizzazione - CHANNELS

Possiamo poi distinguere due tipi di channels: **channels di intensità** e **channels di identità**.



Elementi di visualizzazione – MARKS e CHANNELS

I **marks** vengono utilizzati per rappresentare:

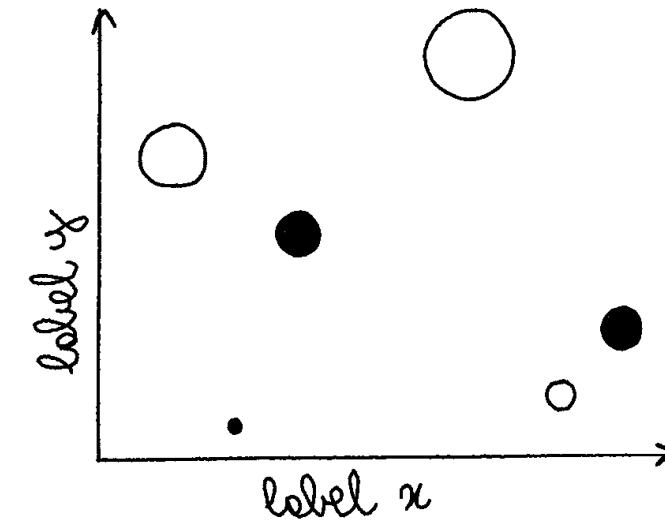
- Items
- Valori
- Links

I channels vengono usati per rappresentare:

- Attributi
- Posizioni

In particolare:

- I **channels di identità** possono rappresentare attributi categorical e posizioni
- I **channels di intensità** possono rappresentare attributi ordered

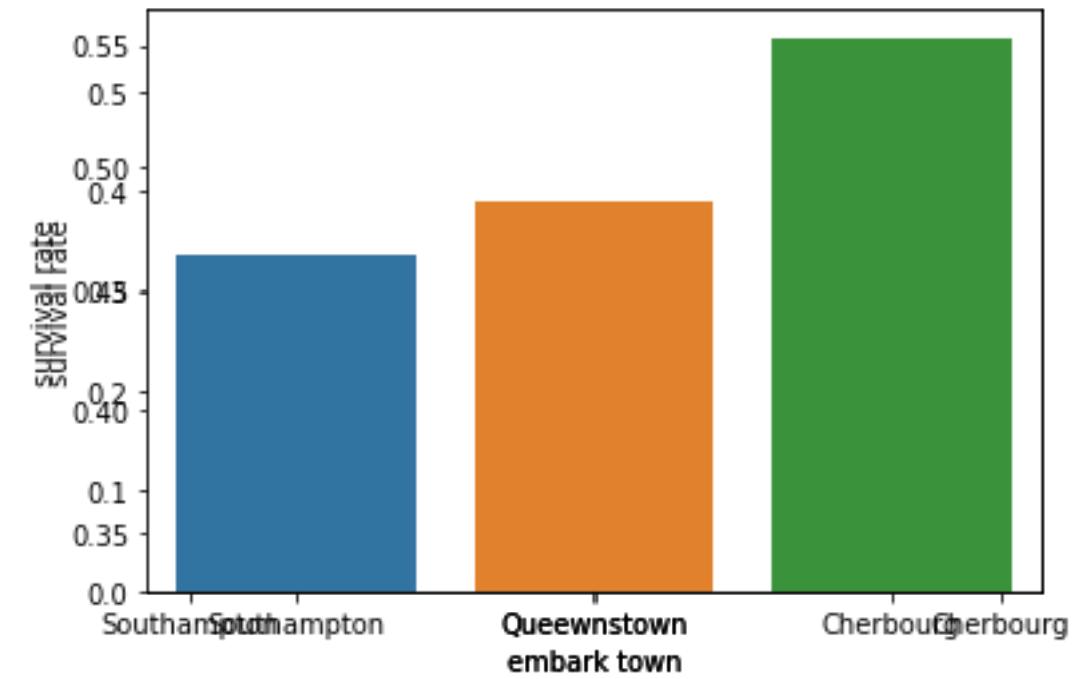
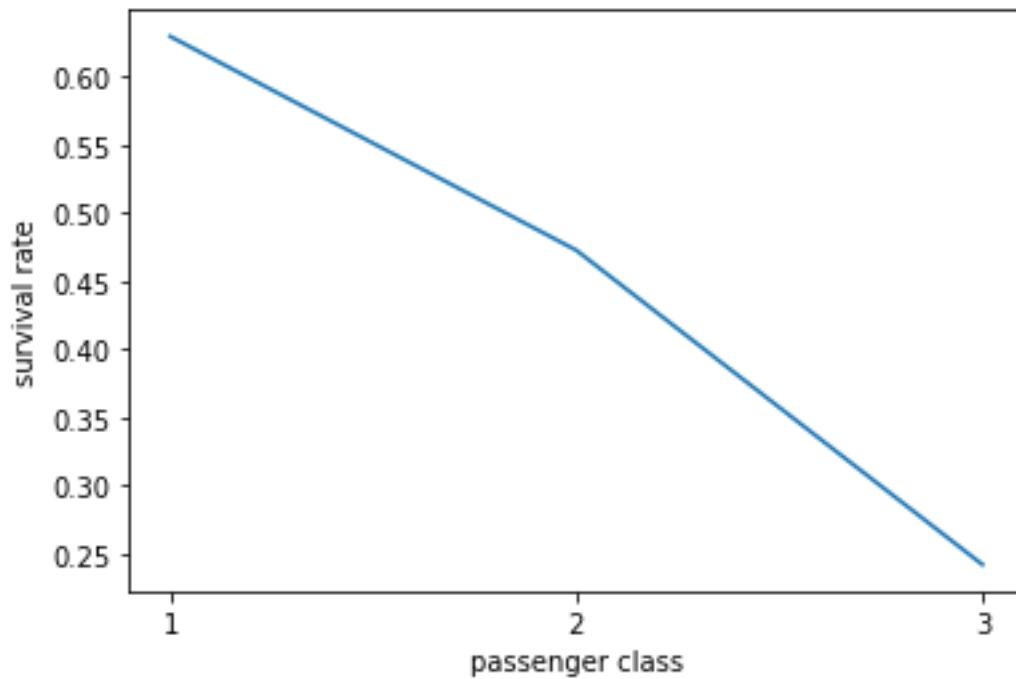


punti
direzione orizzontale
direzione verticale
colore
dimensione



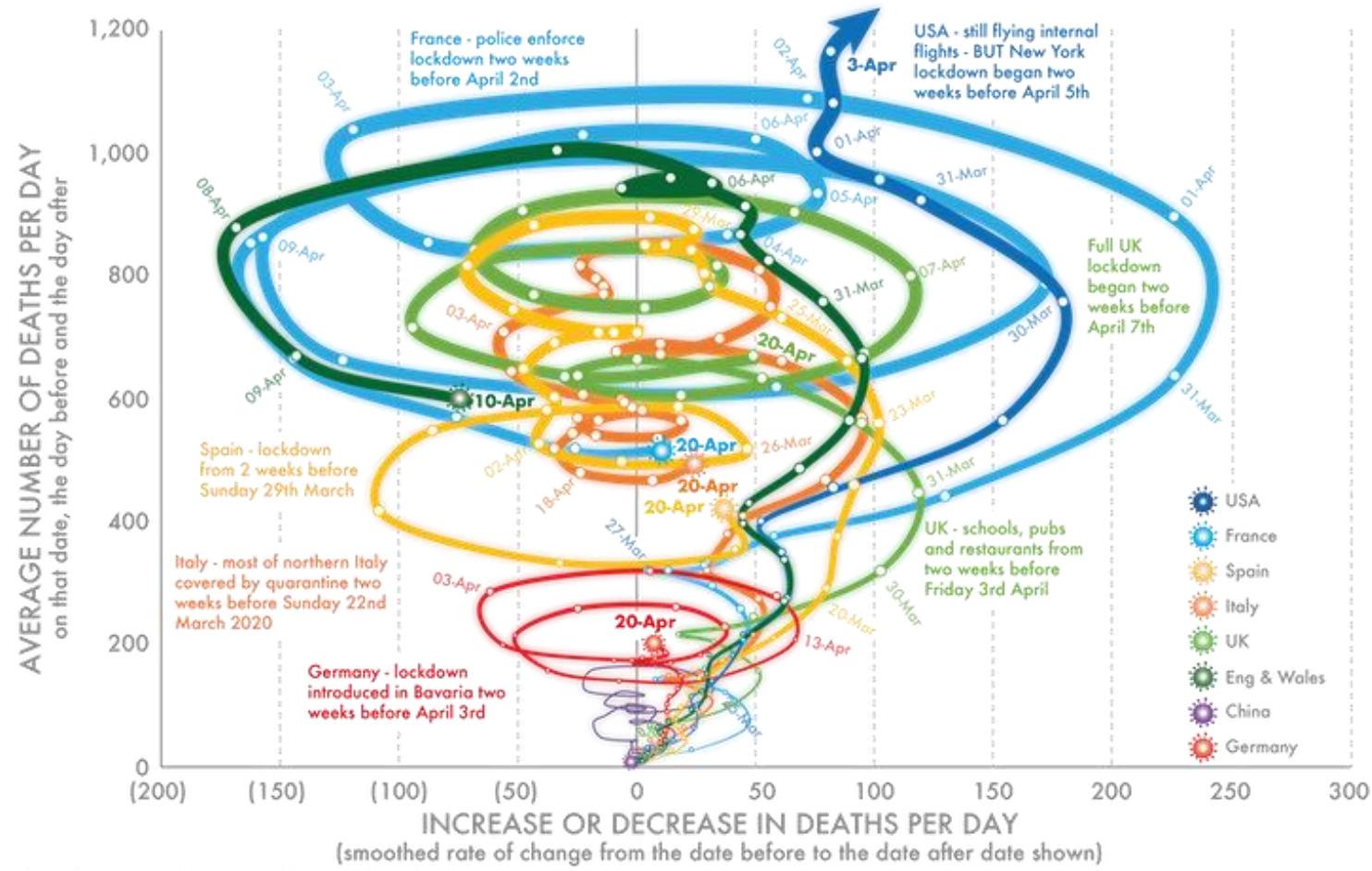
I principi della visualizzazione – Espressività

La rappresentazione visiva dovrebbe mostrare **solo** le informazioni di interesse.



I principi della visualizzazione – Efficacia

L'importanza degli attributi dovrebbe corrispondere alla rilevanza dei channels utilizzati.



A mio avviso grafico
non efficace.

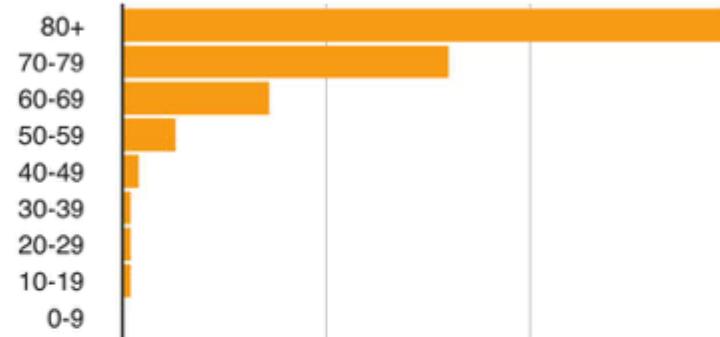


I principi della visualizzazione – Efficacia

Death rate varies by age, health and sex

Case fatality ratio

Age



Health condition



Sex



Source: Chinese Centre for Disease Control and Prevention

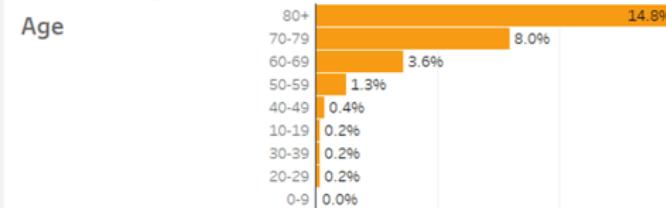


Osservazioni/commenti?

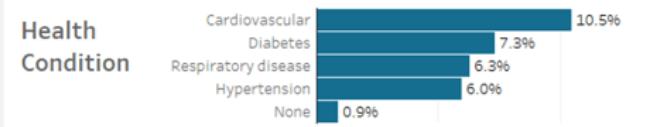
Death rate varies by age, health and sex

Case fatality ratio

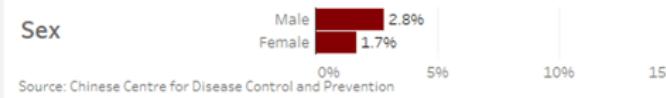
Age



Health Condition



Sex



Source: Chinese Centre for Disease Control and Prevention

NOTE: this is a redesign of a BBC graphic by Andy Cotgreave, exploring axis lengths. The redesign is in response to a tweet from Alice Casey (@cased)

What percent of people who contract coronavirus die (estimated)?

Case fatality ratio

Age

Health Condition

Sex

Source: Chinese Centre for Disease Control and Prevention

NOTE: this is a redesign of a BBC graphic by Andy Cotgreave, exploring axis lengths. The redesign is in response to a tweet from Alice Casey (@cased)



I principi della visualizzazione – Come misurare l'efficacia?

Come misurare l'efficacia?

- **Accuratezza:** quanto i valori possono essere stimati accuratamente dal grafico;
- **Discernibilità:** quanti valori possono essere percepiti;
- **Separabilità:** quanto si riescono a distinguere gli elementi usati nella visualizzazione;
- **Popout:** quanto è facile distinguere un item di interesse dagli altri;
- **Grouping:** quanto un channel usato fa percepire un gruppo di dati “simili”.



Tufte's principles – Chart Lies

Il Chart Lies può essere quantificato attraverso il **Lie Factor (LF)**:

$$LF = \frac{\text{rapporto misure dei valori nel grafico}}{\text{rapporto misure dei valori nei dati}}$$

Questo numero rappresenta il rapporto tra la rappresentazione del dato e il valore del dato stesso.

In visualizzazione, per ottenere un risultato efficace, i valori dovrebbero essere rappresentati in diretta proporzionalità con le quantità che rappresentano.

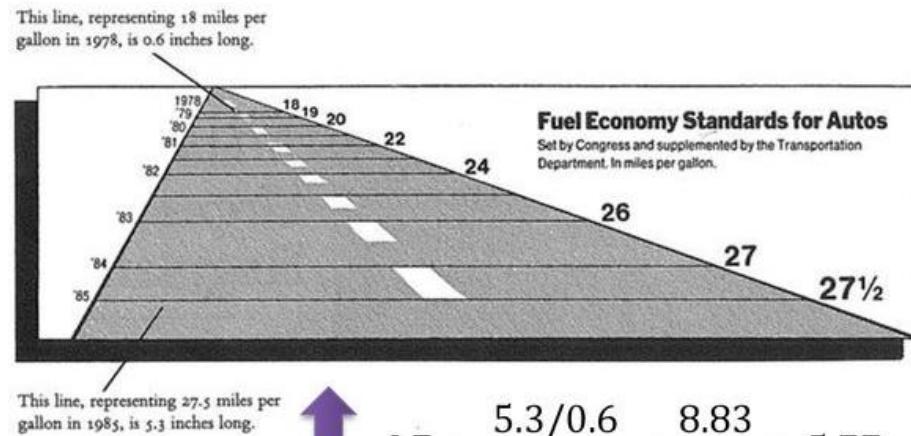
Quindi... Se questo numero è **minore di 0.95 o maggiore di 1.05** allora è indice di una sostanziale distorsione nella visualizzazione.



Tufte's principles – Chart Lies

In queste visualizzazioni la rappresentazione dei valori, misurata sul grafico, non è direttamente proporzionale al valore numerico che questi rappresentano.

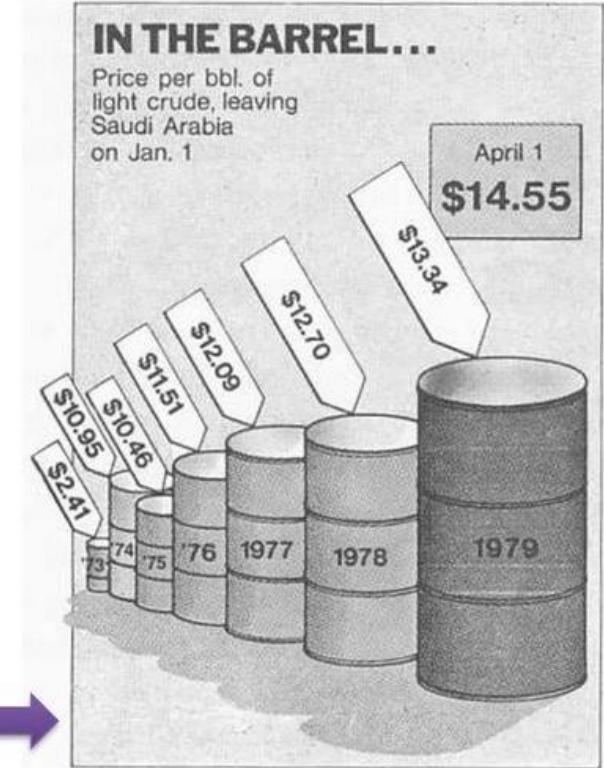
Il **Lie Factor** (LF) risulta infatti, in entrambi i casi, maggiore di 1.05.



$$LF = \frac{5.3/0.6}{27.5/18} = \frac{8.83}{1.53} = 5.77$$

Reprinted from
Tufte (2009), p.
57 & p. 62

$$LF = \frac{4280\% \text{ (change in volume)}}{454\% \text{ (change in price)}} = 9.4$$



Questo è solo uno dei modi di come si può “barare” in visualizzazione dati.

Tufte's principles – Chartjunk

Con Chartjunk ci si riferisce a tutto ciò che è presente nel grafico ma che non rappresenta direttamente i dati.

Un modo per misurare il Chartjunk è calcolare il **Data Ink Ratio (DIR)**:

$$\text{DIR} = \frac{\text{quantità inchiostro dati}}{\text{quantità inchiostro totale grafico}}$$

Molti ricercatori sostengono che qualsiasi informazione non necessaria presente in un grafico distoglia il lettore dall'obiettivo principale e che questa sia "dannosa" per l'efficacia della visualizzazione, quindi sostengono che sia buona pratica cercare di massimizzare il DIR.

Tuttavia, un'altra corrente di pensiero, sostiene che aggiungere determinati oggetti conosciuti ad una visualizzazione dati porti il lettore a ricordare meglio il grafico (e le informazioni) che osserva.

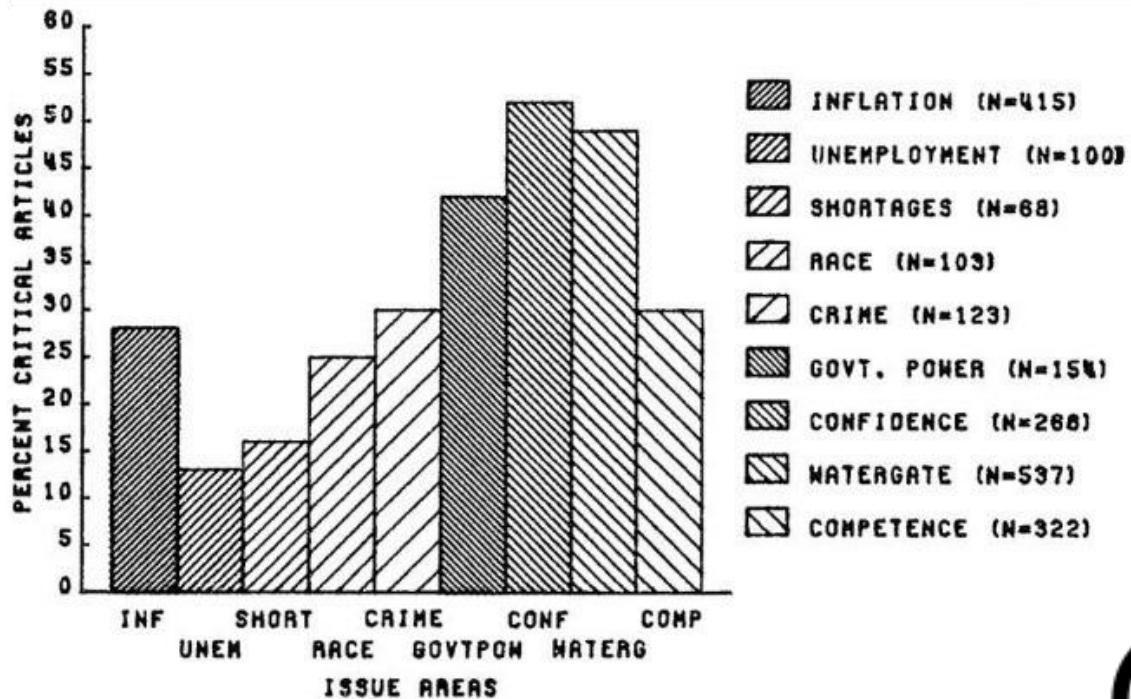


"The Chartjunk debate"

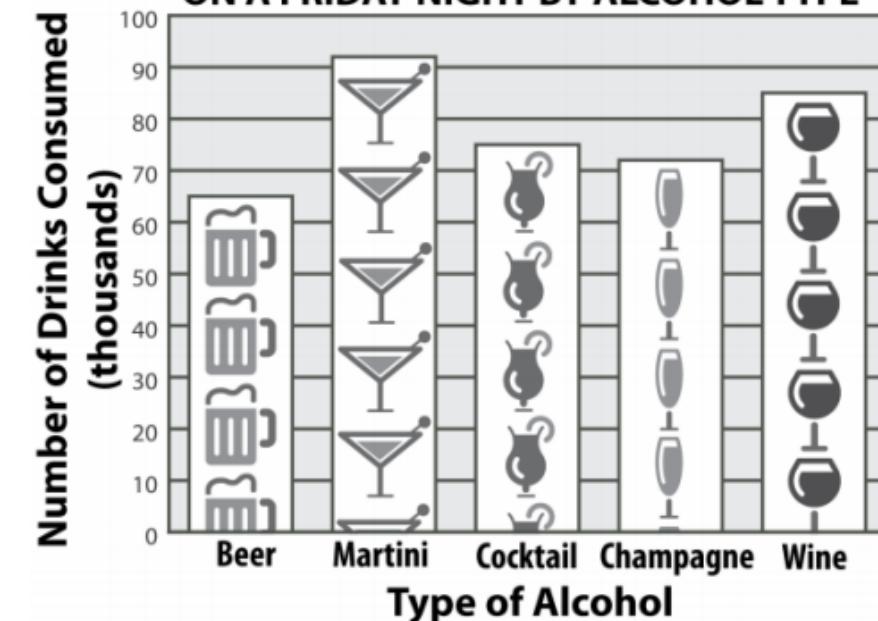
<https://dl.acm.org/doi/abs/10.1145/1753326.1753716>



Tufte's principles – Chartjunk



NUMBER OF DRINKS CONSUMED IN MANHATTAN
ON A FRIDAY NIGHT BY ALCOHOL TYPE



«Perfection is achieved not when there is nothing more to add, but when there is nothing left to take away.»

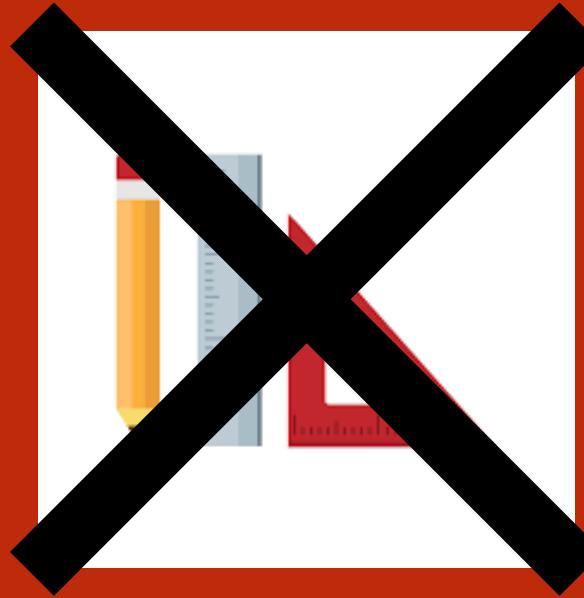
— Antoine de Saint-Exupéry



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA



non trattato in
questa lezione

Data Visualization – Visualization Tools

Alessia Angeli

Studente di dottorato in Data Science and Computation
Dipartimento di Informatica – Scienza e Ingegneria



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA



Data Visualization – I Colori

Alessia Angeli

Studente di dottorato in Data Science and Computation
Dipartimento di Informatica – Scienza e Ingegneria

I COLORI – Lo scopo del colore nella visualizzazione dati

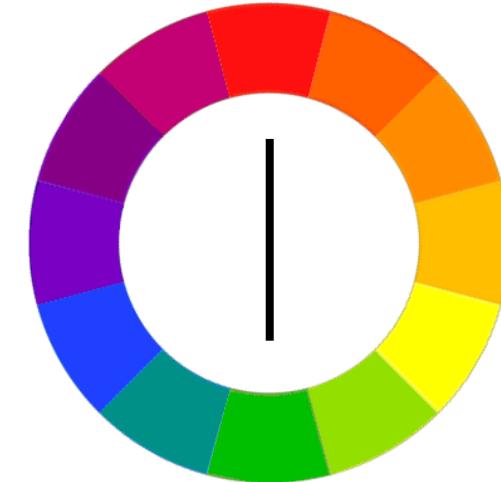
- Per **etichettare** (colore come nome);
- Per **misurare** (colore come quantità);
- Per **rappresentare** (colore come simbolo);
- Per **decorare** (colore come bellezza).



I COLORI – Colori complementari e colori analoghi

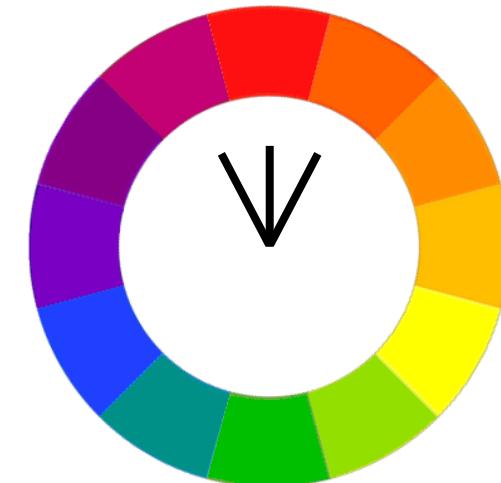
Colori complementari:

- Si trovano alle estremità opposte della scala dei colori, creano un forte contrasto se usati fianco a fianco.
- Da riservare alle occasioni in cui si vuole che qualcosa si distingua.



Colori analoghi:

- Si trovano uno accanto all'altro sulla ruota dei colori.
- E' necessario assicurarsi che ci sia abbastanza contrasto tra colori analoghi per usarli in visualizzazione, per fornire chiarezza nell'interpretazione.



I COLORI – La percezione dei colori

Gli uomini percepiscono i colori in tre dimensioni:

- **Hue** (tinta/colore):

è il colore primario. Dal rosso all'arancione, giallo, verde, azzurro, indaco e viola.



Hue

- **Lightness** (luminosità):

descrive lo spettro di una tinta dal chiaro allo scuro, quanto una tinta è mischiata con il nero.



Lightness

- **Saturation** (saturazione):

è l'intensità di un colore, quanto un colore è mischiato con il bianco. I colori intensi sono colori saturi mentre i toni della terra, insieme al grigio, nero e bianco sono colori desaturati.



Saturation



I COLORI – Possiamo ordinare i colori?

- **Hue** – NO

Questa caratteristica è utile per dati categorici in quanto non possiede un ordine interno proprio.



Hue



Lightness



Saturation



I COLORI – Come utilizzarli in visualizzazione dati

Color maps e **color palettes** sono mappe utilizzate per codificare uno (o più) attributi con i colori.

La differenza tra color maps e color palettes è che le prime codificano i colori in modo continuo, mentre le seconde in modo discreto.

Color map – continuo



Color palette – discreto



NOTA: per maggiore chiarezza nelle prossime slides si farà riferimento solo alle color palettes, però tutto quanto visto si può essere riportato anche per le color maps.



I COLORI – Come utilizzarli in visualizzazione dati

I tre tipi principali di color palettes sono:

- **Sequential** (sequenziale):

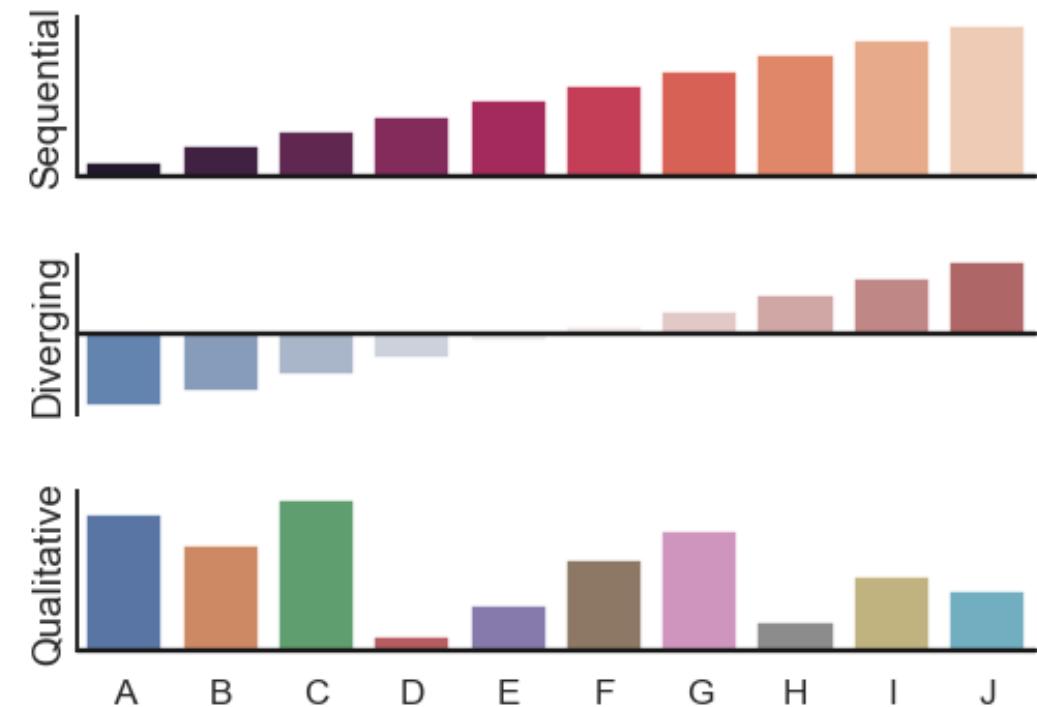
composta da colori che possono essere ordinati da un «basso» ad un «alto».

- **Diverging** (divergente):

si può pensare composta da due palette sequenziali che si estendono in direzioni opposte rispetto ad un punto centrale.

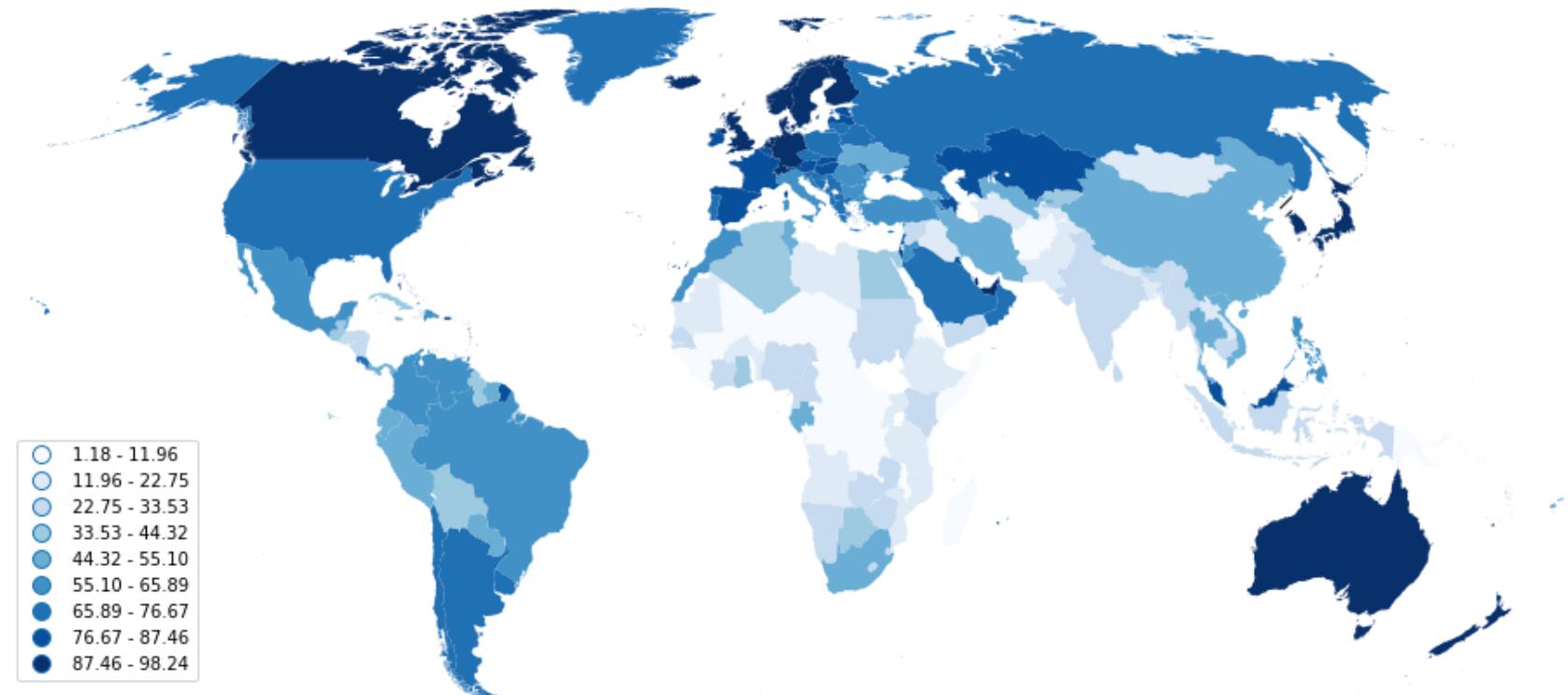
- **Qualitative/categorical** (qualitativa/categorica):

composta da colori con alto contrasto (soprattutto tra colori adiacenti) senza un ordine interno proprio.



I COLORI – Sequential color palette

Individuals using the Internet (% of population) in 2016



Data types:

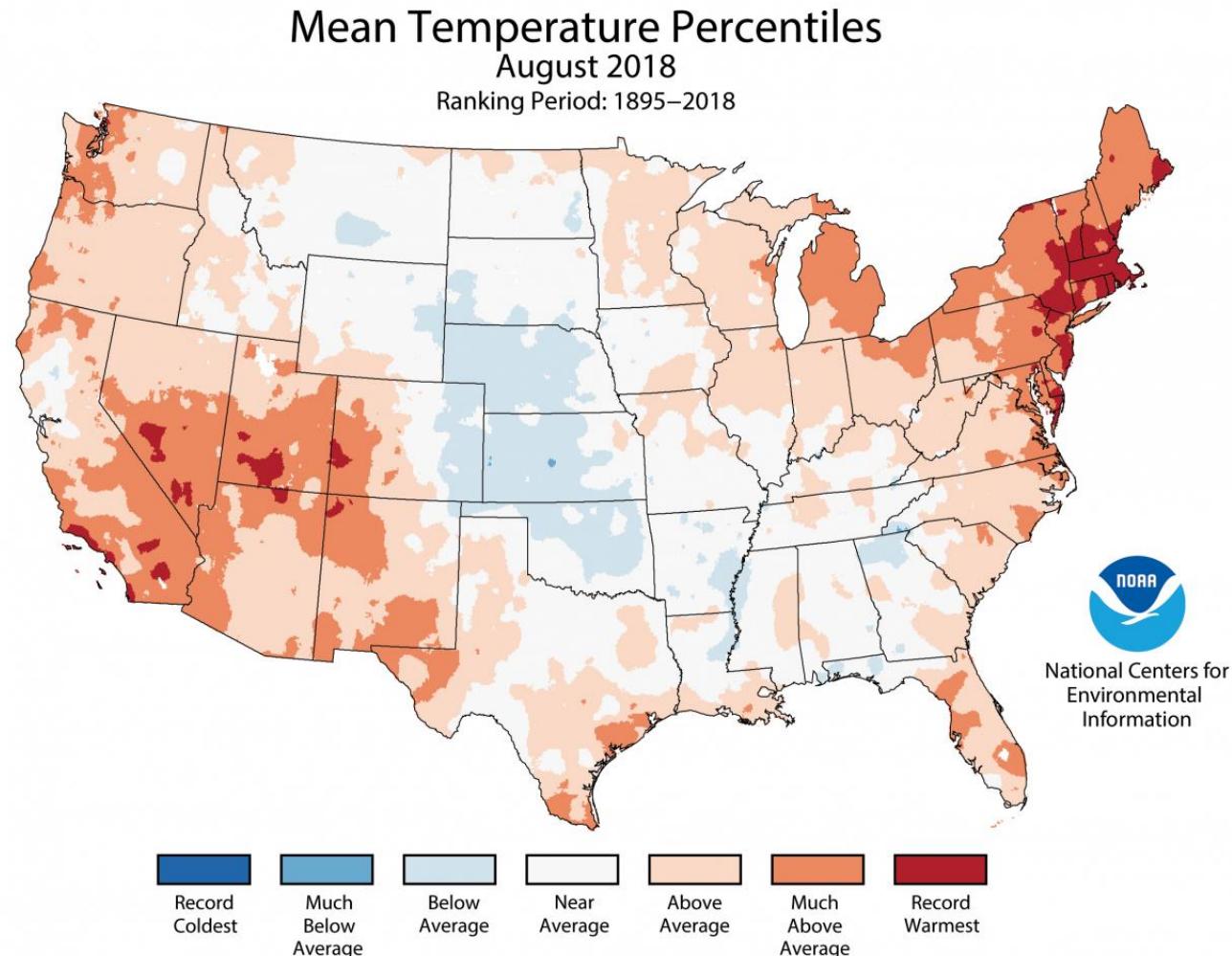
- **Ordinal** data
- **Interval** data
- **Ratio** data

Individuals who have used the Internet from any location in the last 3 months via any device based on the International Telecommunication Union, World Telecommunication/ICT Development Report and database. Data: World Bank - worldbank.org • Author: Ramiro Gómez - ramiro.org



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

I COLORI – Diverging color palette



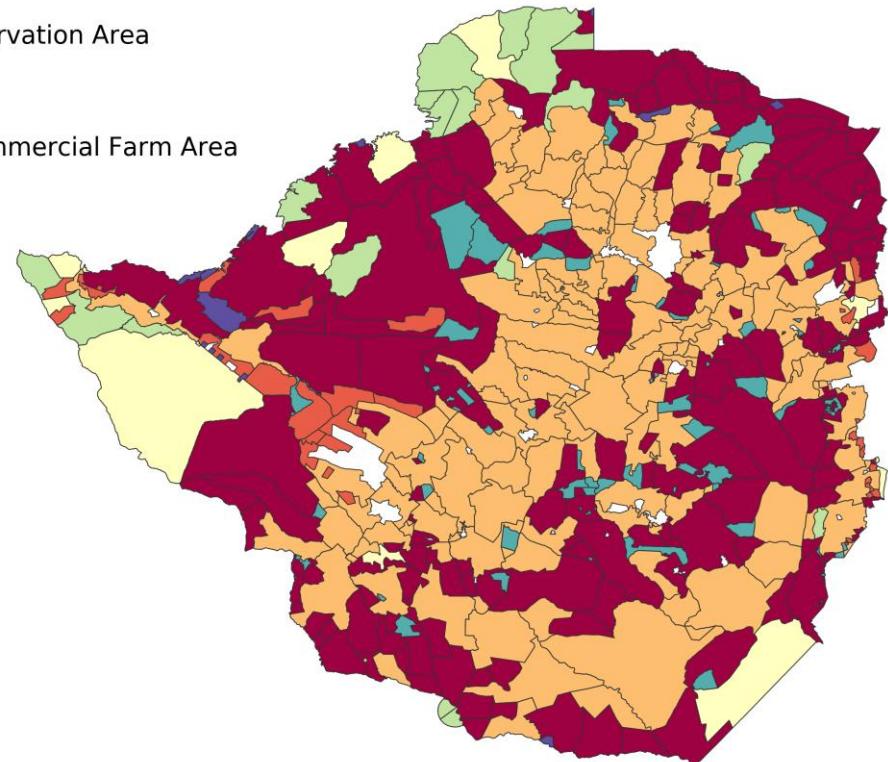
Data types:
Interval/ratio data con un punto di **average/zero** rilevante.



I COLORI – Qualitative/categorical color palette

- Communal Land
- Forest Land
- Intensive Conservation Area
- National Park
- Safari Area
- Small-Scale Commercial Farm Area
- State Land

Land Usage in Rhodesia



Data types:

Categorical data.

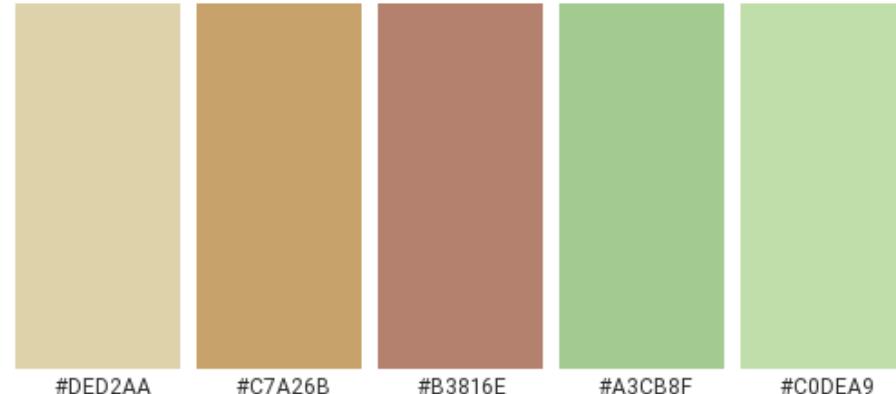
NOTA: limitare il numero di tonalità usate ad un massimo di 7 per una buona distinzione.



I COLORI – I colori trasmettono emozioni

Un'idea generale:

- I **toni della terra** sono una miscela o tonalità di marroni, che possono includere anche colori più ricchi contenenti un po' di marrone, come l'arancione, il rosso, il verde. Questi sono considerati rassicuranti e accomodanti.
- I **colori non naturali**, invece, risaltano e risultano ideali per mostrare un avviso.



Pastel Earthy Tones Color Scheme - by SchemeColor.com



Happy Colors Color Scheme - by SchemeColor.com



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

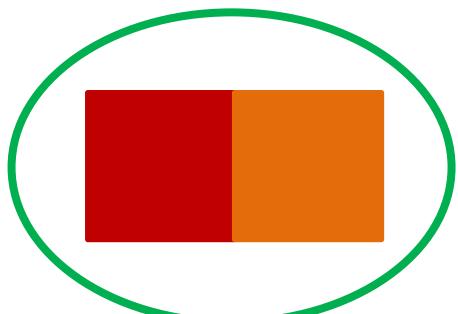
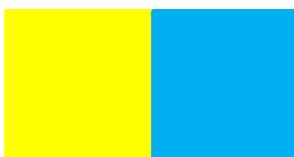
I COLORI – Controlla gli accostamenti di colore

Per cercare di scegliere accostamenti di colore efficaci controlla in bianco/nero!

Abbinamenti di colori contrastanti potrebbero stancare gli occhi degli utenti.

Soltanamente i contrasti ottimali sono quelli che rimangono tali anche in bianco/nero.

Quindi... Tra questi abbinamenti quale scegliereste?



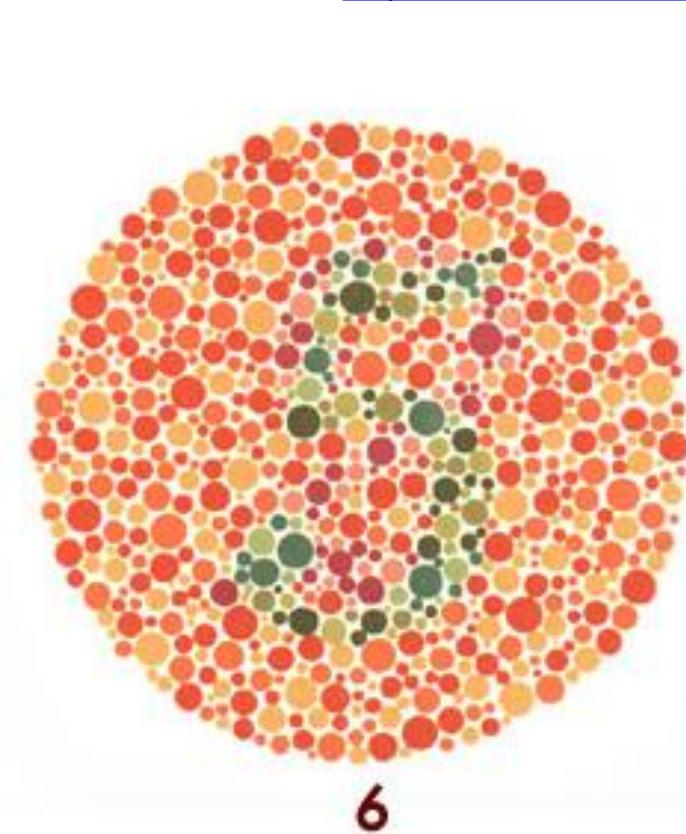
Hue as Value
with Grayscale
for comparison



I COLORI – Il daltonismo

Daltonismo – in medicina, cecità ereditaria per i colori, che si trasmette secondo lo schema della cosiddetta eredità legata al sesso: può interessare tutti i colori (acromatopsia) o riguardarne soltanto uno o pochi, per lo più il rosso (protanopia) e il verde (deutanopia).

<https://www.treccani.it/vocabolario/daltonismo/>



NOTA: per chi fosse interessato esistono diversi siti internet per costruire color palettes come:



<https://colorbrewer2.org/#type=sequential&scheme=BuGn&n=3>



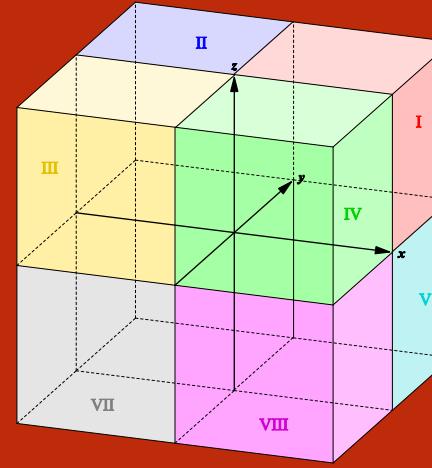
<https://www.palettable.io/FFBF00>



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA



Data Visualization – Space & Attributes

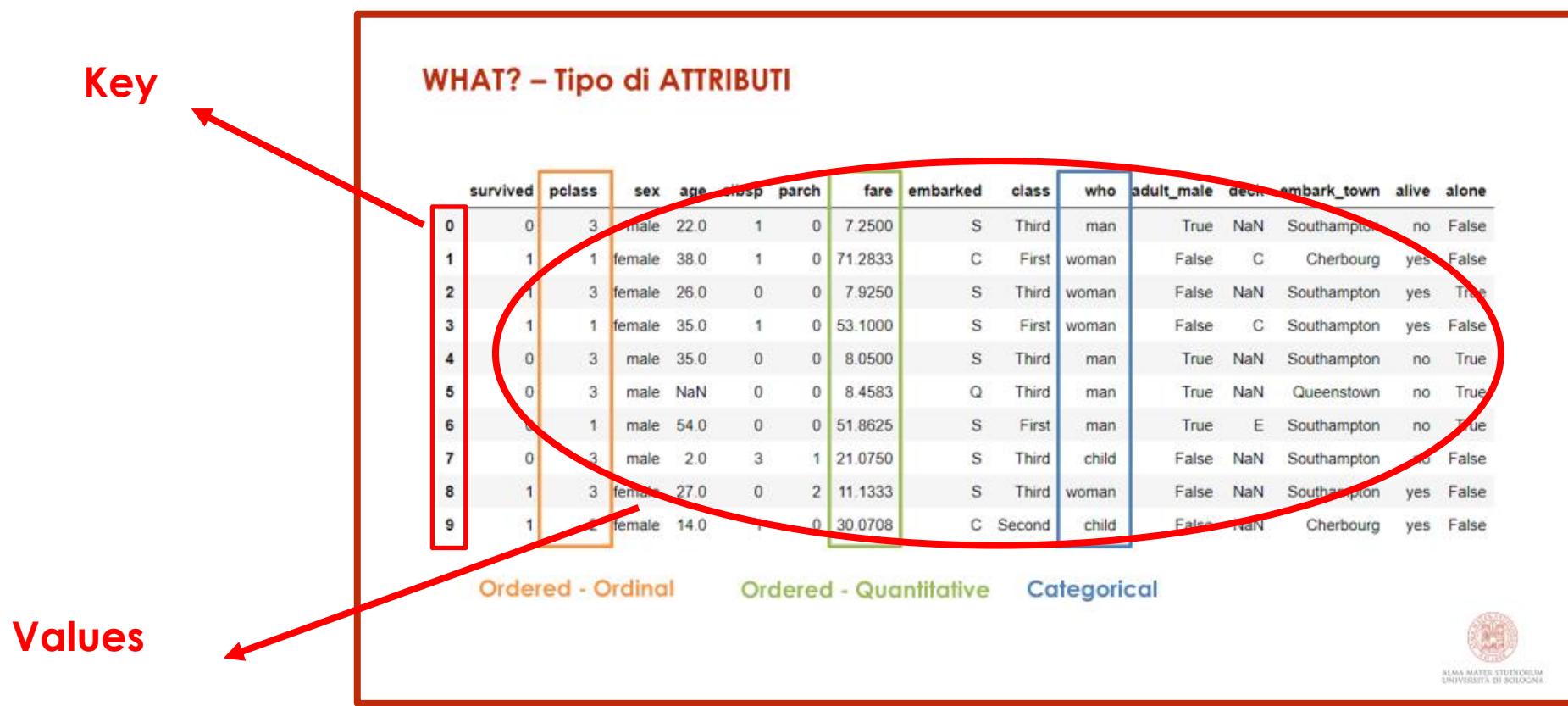
Alessia Angeli

Studente di dottorato in Data Science and Computation
Dipartimento di Informatica – Scienza e Ingegneria

Attributes: Keys and Values

- **Attribute – Key:** un attributo viene definito chiave quando identifica univocamente (in modo unico) un item (elemento);
- **Attribute – Value:** un attributo se non è una chiave viene definito valore.

WHAT? – Tipo di ATTRIBUTI



	survived	pclass	sex	age	sibsp	parch	fare	embarked	class	who	adult_male	deck	embark_town	alive	alone
0	0	3	male	22.0	1	0	7.2500	S	Third	man	True	NaN	Southampton	no	False
1	1	1	female	38.0	1	0	71.2833	C	First	woman	False	C	Cherbourg	yes	False
2	1	3	female	26.0	0	0	7.9250	S	Third	woman	False	NaN	Southampton	yes	True
3	1	1	female	35.0	1	0	53.1000	S	First	woman	False	C	Southampton	yes	False
4	0	3	male	35.0	0	0	8.0500	S	Third	man	True	NaN	Southampton	no	True
5	0	3	male	NaN	0	0	8.4583	Q	Third	man	True	NaN	Queenstown	no	True
6	0	1	male	54.0	0	0	51.8625	S	First	man	True	E	Southampton	no	True
7	0	3	male	2.0	3	1	21.0750	S	Third	child	False	NaN	Southampton	no	False
8	1	3	female	27.0	0	2	11.1333	S	Third	woman	False	NaN	Southampton	yes	False
9	1	2	female	14.0	1	0	30.0708	C	Second	child	False	NaN	Cherbourg	yes	False

Ordered - Ordinal Ordered - Quantitative Categorical

Key

Values



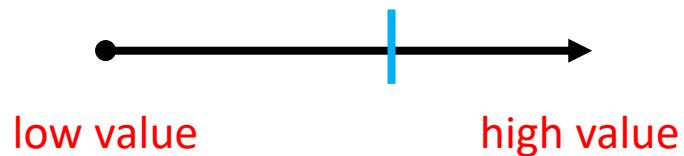
ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA



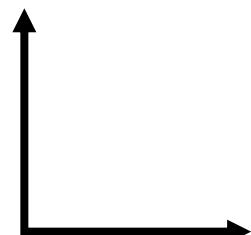
ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Space & Quantitative Attributes

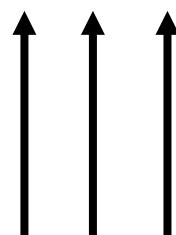
- Esprimere il valore con la posizione:



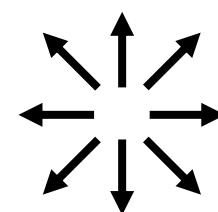
- Scegliere il modo più efficace per disegnare gli assi.



ORTOGONALI



PARALLELI

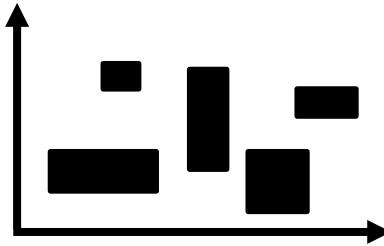


RADIALI

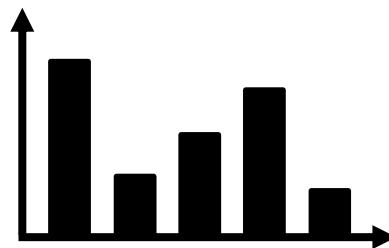


Space & Qualitative Attributes

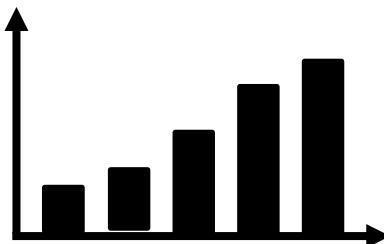
- Codificare categorical key(s) in regioni separate;



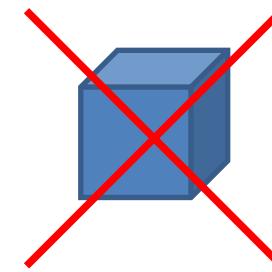
- Scegliere un allineamento appropriato;



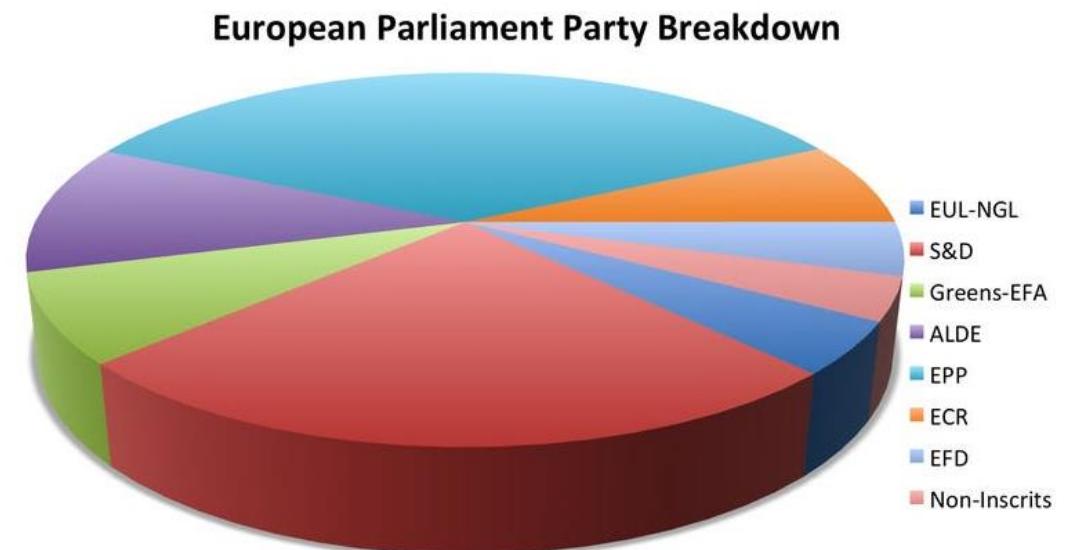
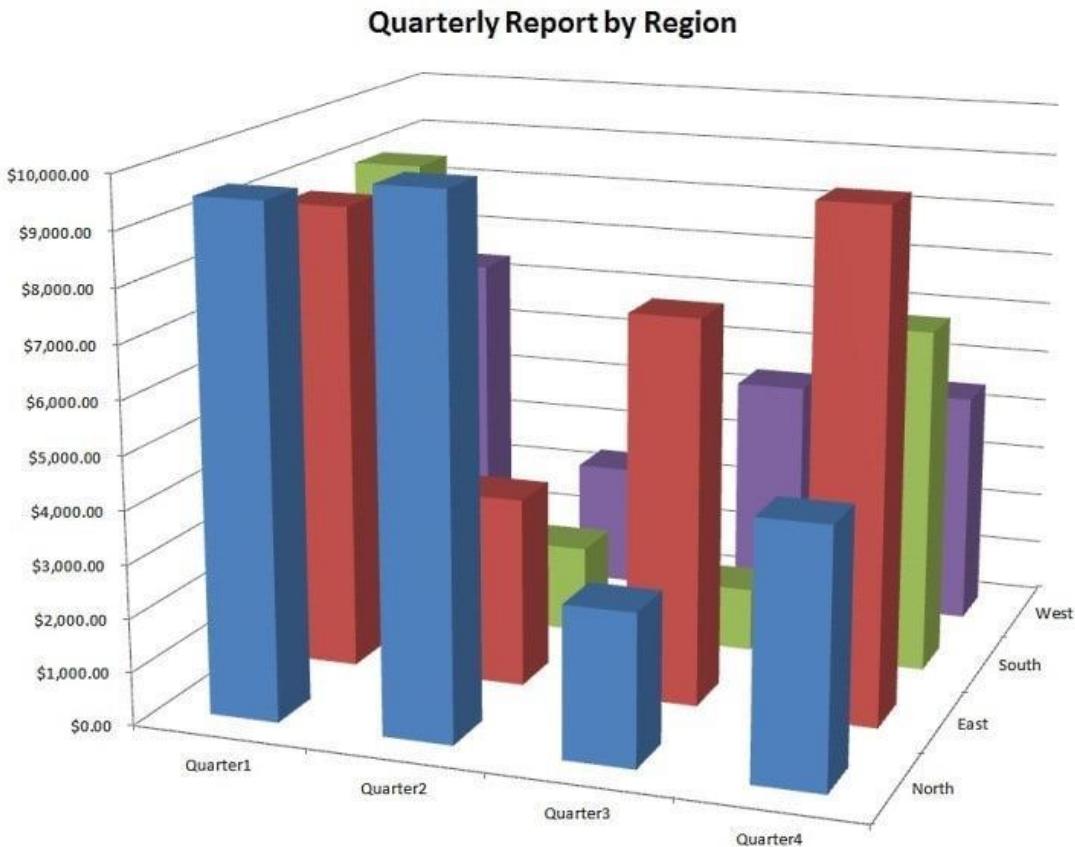
- Scegliere un ordine appropriato.



Attenzione alle visualizzazioni 3D!

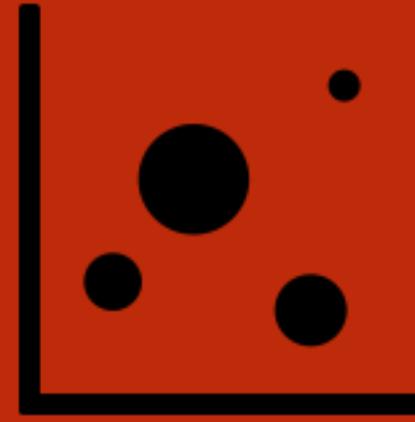


3D... Senza valide ragioni? Anche no.





ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA



Data Visualization – Plots/Graphs

Alessia Angeli

Studente di dottorato in Data Science and Computation
Dipartimento di Informatica – Scienza e Ingegneria

Quantitative Attributes

- Scatter plot
- Histogram
- Scatter plot matrix
- Box plot
- Violin plot
- Radar chart

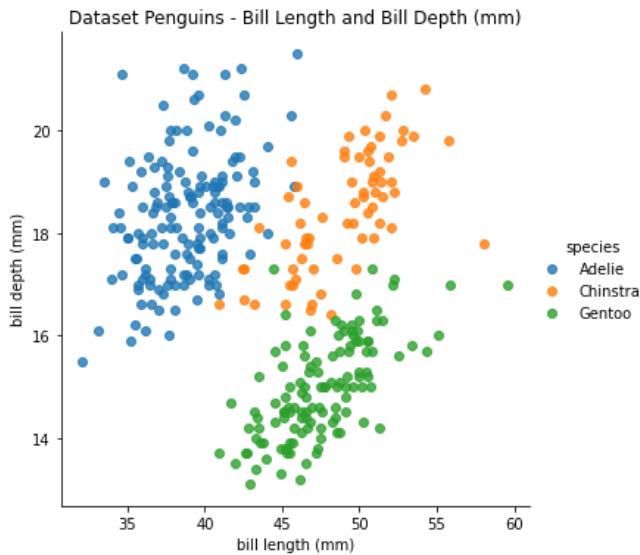
...



Scatter plot

What?

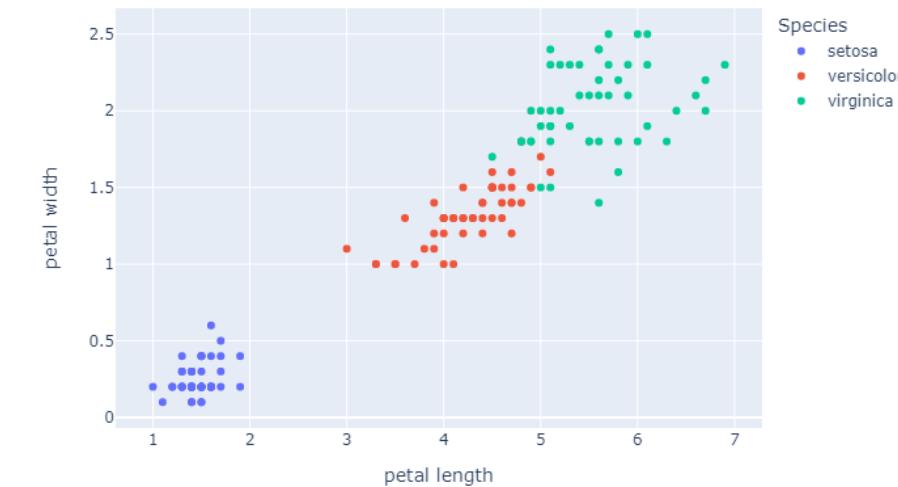
- 2 quantitative attributes;



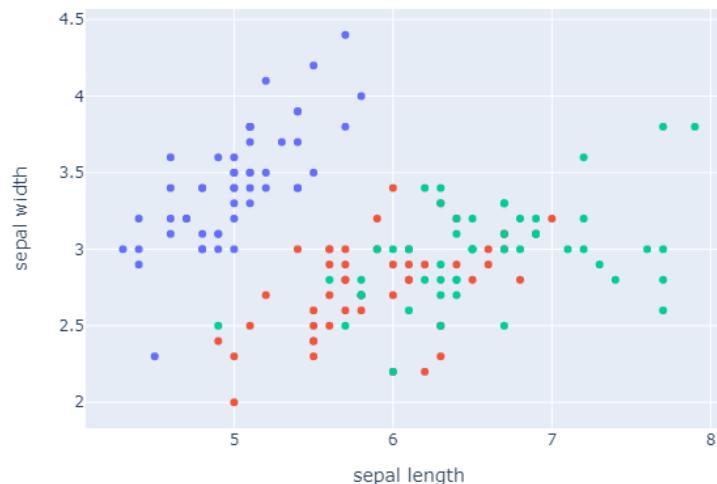
Why?

- Visualizzare correlazioni e distribuzioni;
- Identificare outliers, patterns e clusters;

Petal Length and Petal Width of different Iris Species



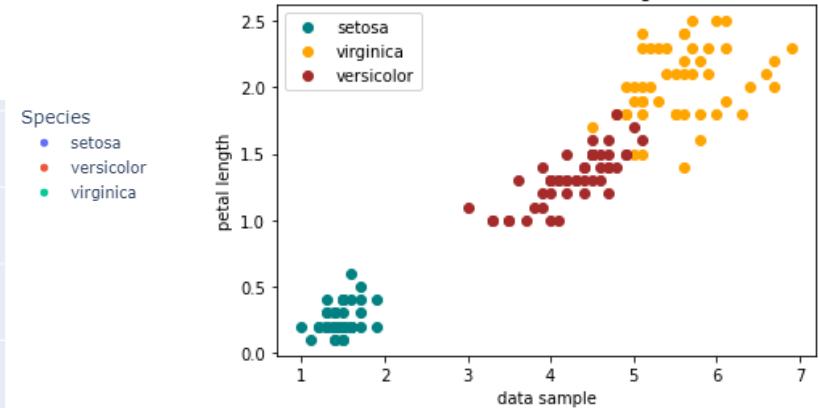
Sepal Length and Sepal Width of different Iris Species



Remarks

- Fino a ~100 items;
- Colore e dimensione possono essere usati per codificare categorical attributes aggiuntivi (bubble plot).

Dataset Iris - Petal Length

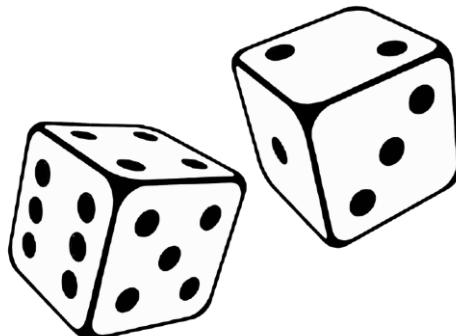


Definizione

DISTRIBUZIONE DI PROBABILITÀ: Una distribuzione di probabilità è un modello matematico che associa ai valori (possibili) di una variabile aleatoria (continua o discreta) le probabilità che tali valori possano essere assunti da tale variabile. Formalmente le distribuzioni vengono espresse da funzioni matematiche, **funzione densità di probabilità** e **funzione di probabilità**, rispettivamente per variabili aleatorie continue e discrete.

ESEMPIO

Si lanciano 2 dadi e si considera come variabile aleatoria la somma risultante.



Somma	# Combinazioni	Probabilità
2	1	0.03
3	2	0.06
4	3	0.08
5	4	0.11
6	5	0.14
7	6	0.17
8	5	0.14
9	4	0.11
10	3	0.08
11	2	0.06
12	1	0.03



Histogram

What?

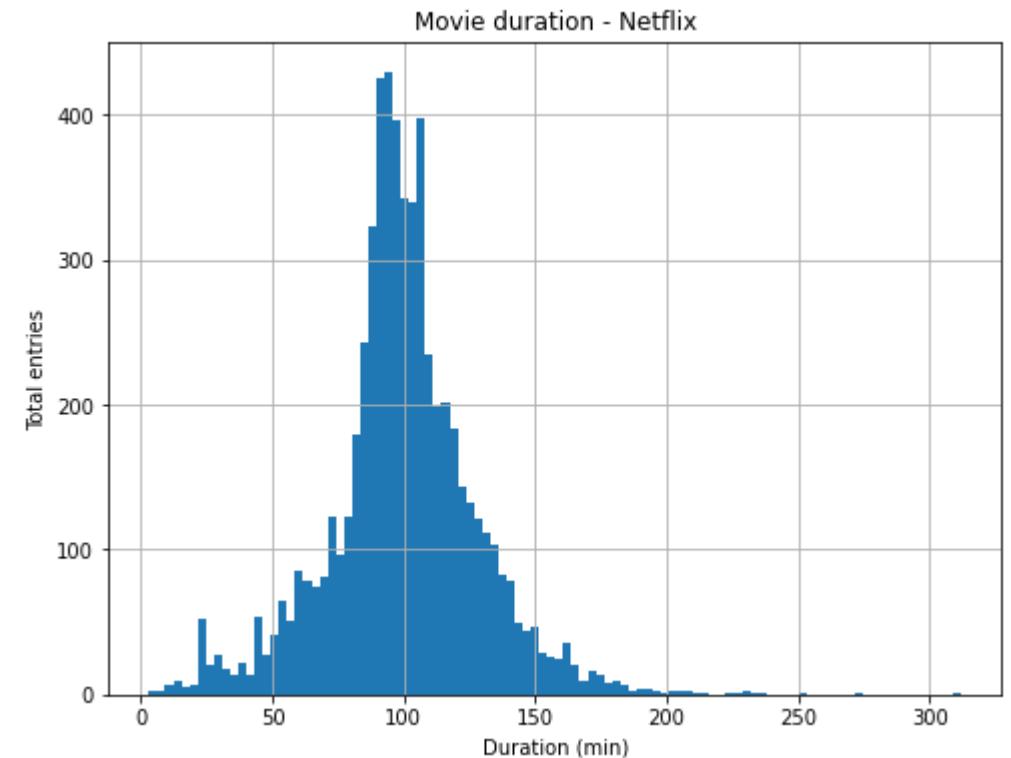
- 1 quantitative attribute;

Why?

- Visualizzare distribuzioni;
- Identificare patterns e range;

Remarks

- Una linea (o un'area) può essere visualizzata per mostrare la funzione di densità calcolata;
- Gli items possono essere visualizzati con dei punti.



Definizione

MATRICE: una matrice è una tabella ordinata. Le righe orizzontali vengono chiamate *righe* della matrice e le righe verticali *colonne* della matrice.

Generalmente una matrice si indica con una lettera maiuscola e viene scritta nel modo seguente:

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

dove i pedici di ogni elemento della matrice indicano, rispettivamente, la riga e la colonna in cui l'elemento è posizionato.

Quindi a_{ij} è l'elemento della matrice A che si trova nella riga i -esima e nella colonna j -esima.



Scatter plot matrix

What?

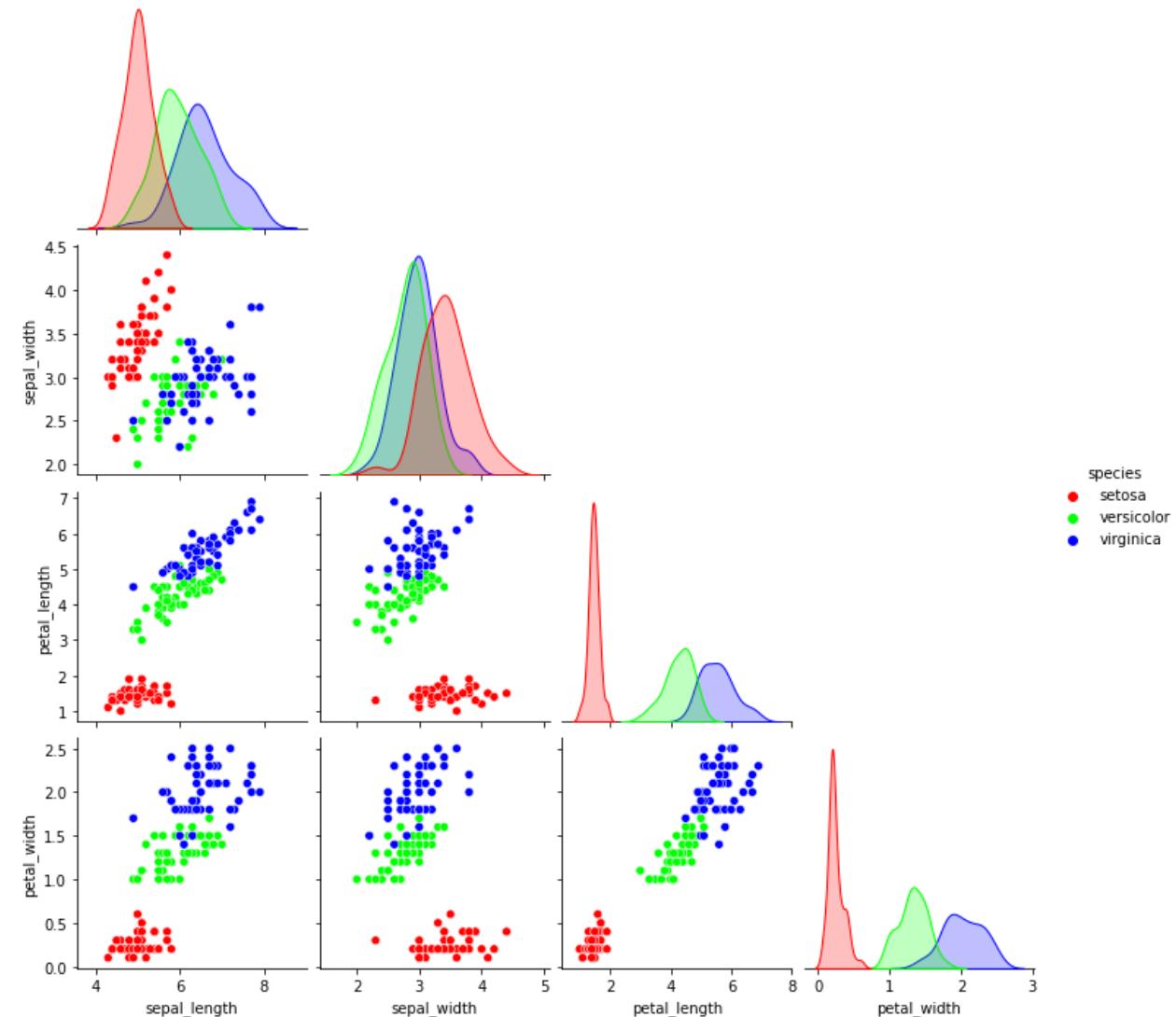
- N quantitative attributes;

Why?

- Visualizzare correlazioni e distribuzioni;
- Identificare outliers, patterns e clusters;

Remarks

- Fino a ~12 attributi e ~100 items;
- E' possibile visualizzare solo la parte triangolare inferiore della matrice.



Statistica descrittiva – alcune definizioni

Considerando un insieme di dati numerici si definiscono:

5 7 4 6 5

MEDIA (MEDIA ARITMETICA): rapporto tra la somma dei dati e il numero dei dati.

$$(5+7+4+6+5)/5=5.4$$

MODA: il valore del dato che si presenta con maggiore frequenza (possono essere presenti più valori di moda).

5 – dato con massima frequenza (2)

MEDIANA: è il valore centrale tra i dati ordinati in modo crescente o decrescente. Se l'insieme contiene un numero di dati dispari c'è un unico valore centrale e questo è la mediana. Se l'insieme contiene un numero di dati pari, invece, ci sono due valori centrali e di solito come mediana viene considerata la media aritmetica di questi.

5 – è il valore centrale in 4 5 5 6 7



Statistica descrittiva – alcune definizioni

Oltre alla mediana, che divide a metà un insieme di dati ordinati, vengono usati anche altri indici che dividono tale insieme in determinate percentuali detti **quantili**, **quartili** e **percentili**.

PERCENTILI: sono un caso particolare dei quantili e, come si intuisce dal nome, dividono l'insieme di dati ordinati in 100 parti.

- il 1° percentile lascia alla sua sinistra un centesimo (1%) degli elementi dell'insieme ordinato;
- il 10° percentile lascia alla sua sinistra il 10% degli elementi;
- il 50° percentile (che coincide con la mediana) lascia alla sua sinistra il 50% degli elementi;
- ...

QUARTILI: questi si ottengono dividendo l'insieme di dati ordinati in 4 parti uguali.

- il **primo quartile** (che coincide con il 25-esimo percentile) è il valore che lascia alla sua sinistra il 25% degli elementi;
- il **secondo quartile** (che coincide con la mediana e con il 50-esimo percentile) è il valore che lascia alla sua sinistra il 50% dei dati;
- il **terzo quartile** (che coincide con il 75-esimo percentile) è il valore che lascia il 75% degli elementi a sinistra e il 25% a destra.



Box plot

What?

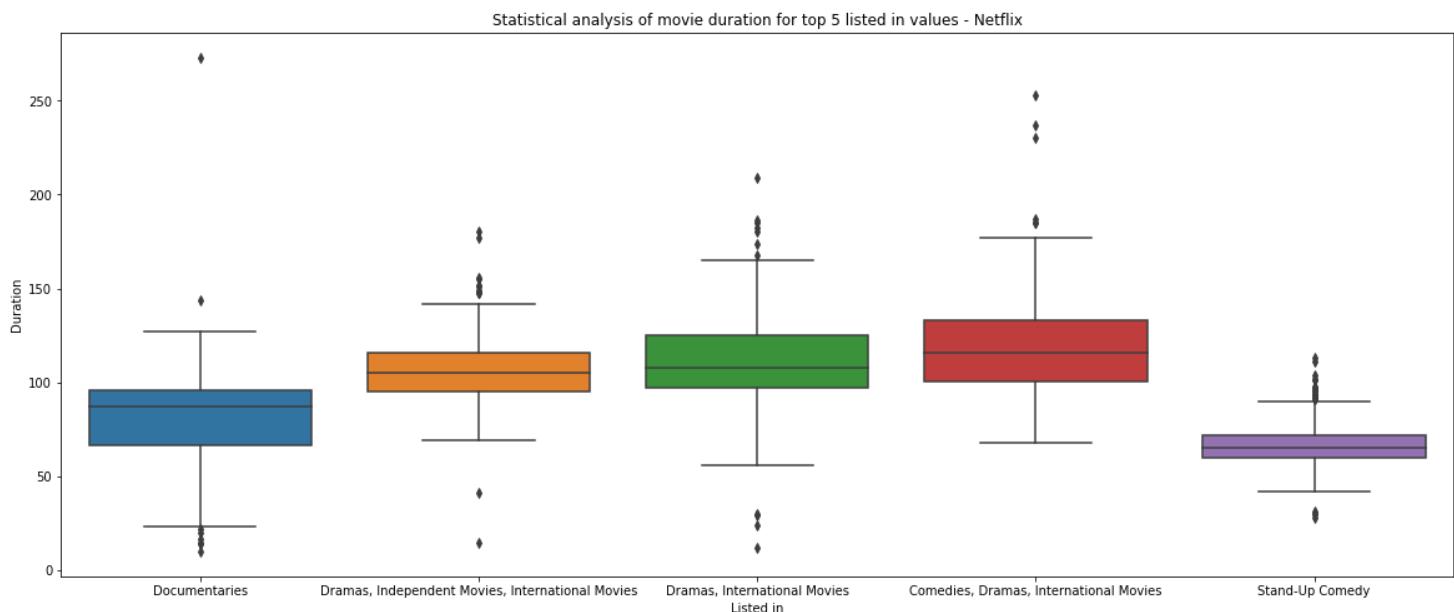
- N quantitative attributes
(oppure 1 quantitative attribute ed 1 categorical key);

Why?

- Visualizzare distribuzioni;
- Identificare outliers, valori estremi, range etc.:

Remarks

- Il colore può codificare un categorical attribute aggiuntivo;
- Possibile effettuare raggruppamenti.



Violin plot

What?

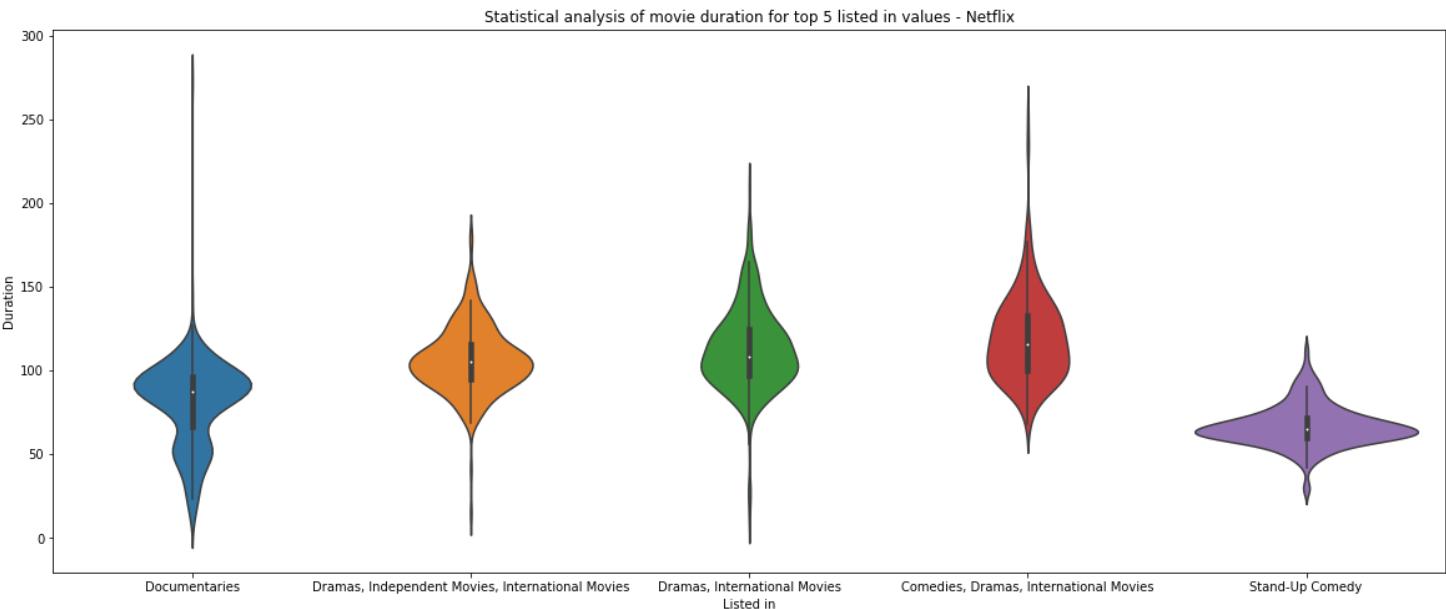
- N quantitative attributes (oppure 1 quantitative attribute ed 1 categorical key);

Why?

- Visualizzare distribuzioni;
- Identificare range;

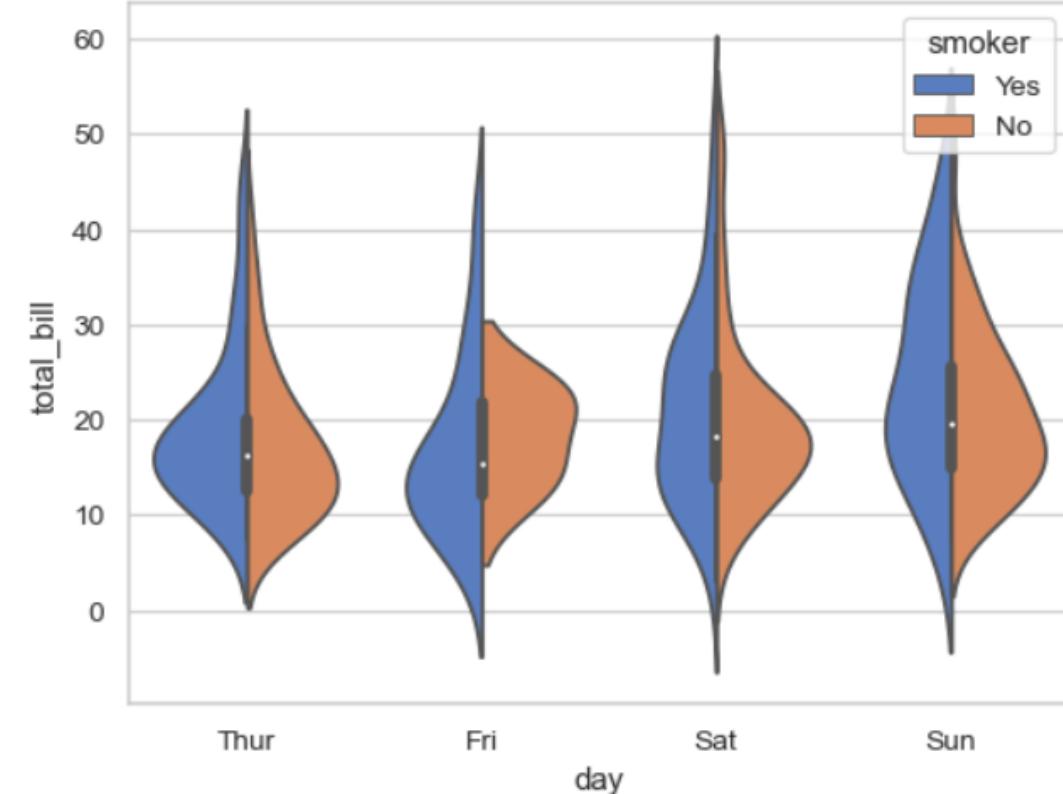
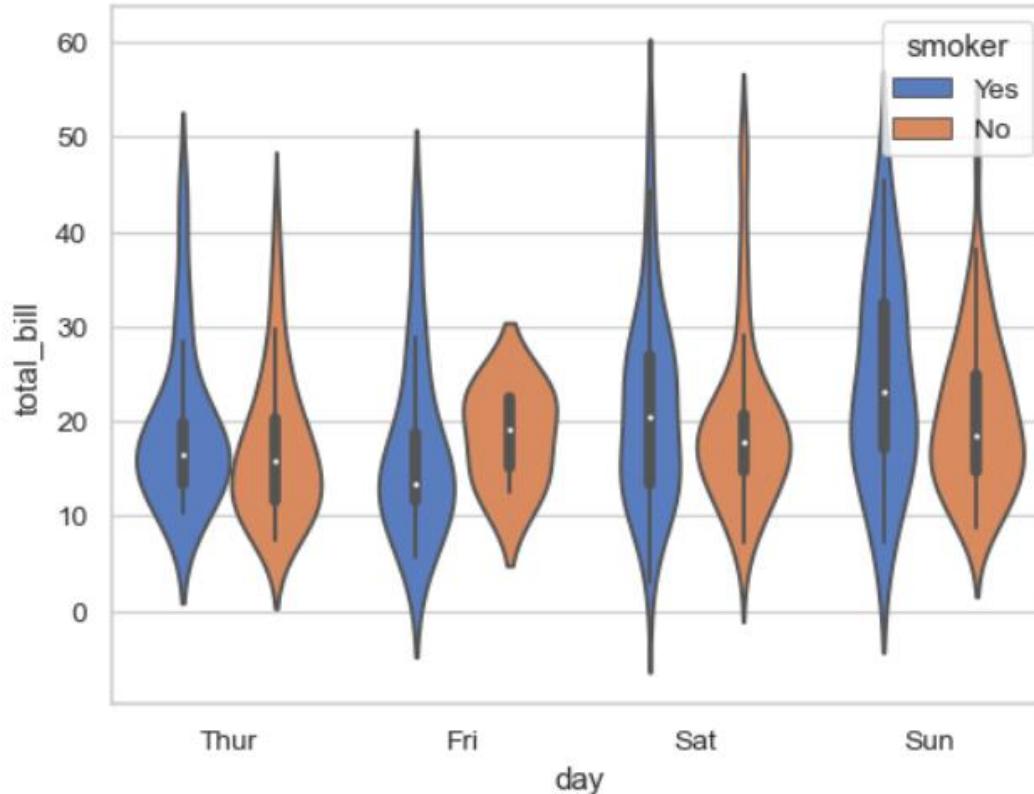
Remarks

- Il colore può codificare un categorical attribute aggiuntivo (è possibile effettuare anche uno split se i dati lo consentono).



Esempi – SOLO immagine

Violin plot



Radar chart

Characteristics of Iris Species

What?

- N quantitative attributes;

Why?

- Identificare patterns;
- Confrontare valori;

Remarks

- Fino a ~12 attributi;
- Il colore può codificare una categorical key aggiuntiva (fino a 3-4 valori).



Qualitative Attributes

- Bar plot
- Multi-set bar plot
- Pie chart
- Word Cloud

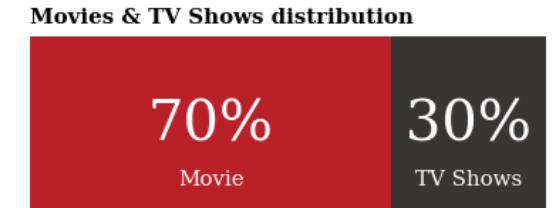
...



Bar plot

What?

- 1 quantitative attribute;
- 1 categorical key;

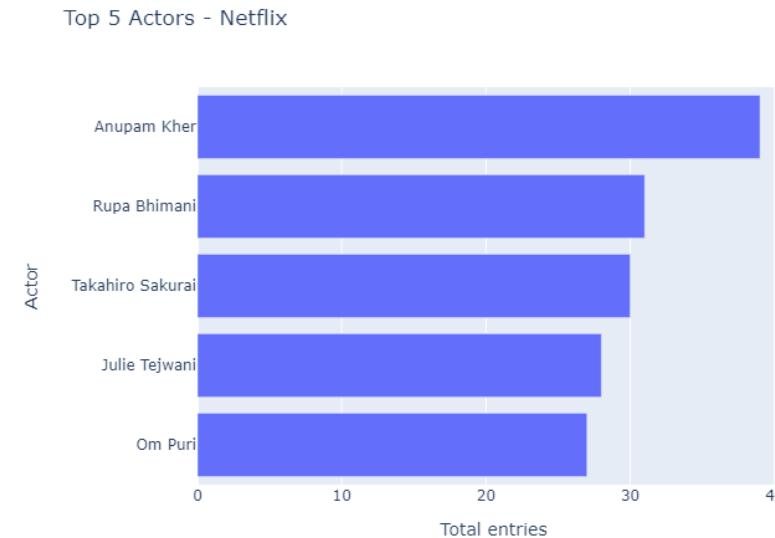


Why?

- Confrontare/evidenziare valori;
- Identificare valori estremi;

Remarks

- Fino a ~100 barre;
- Keys vs valori ordinati;
- Non adatto per visualizzare trends.



Multi-set bar plot

What?

- 1 quantitative attribute;
- 2 categorical keys;

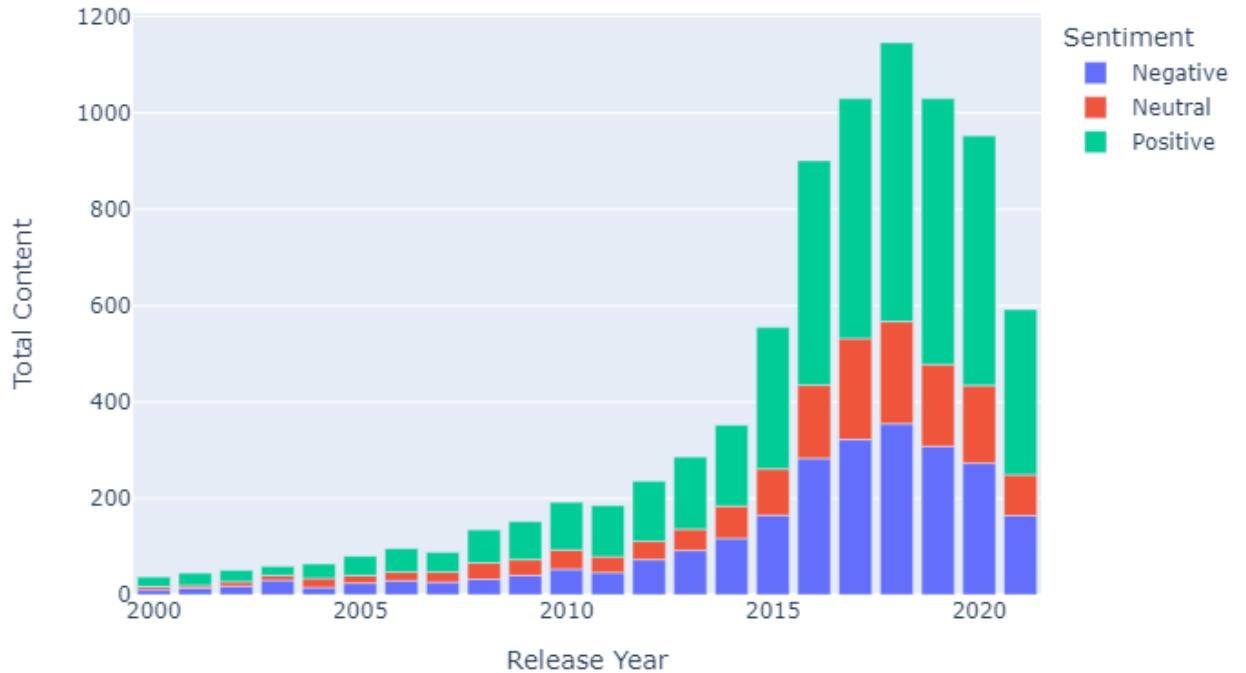
Why?

- Confrontare valori;
- Identificare patterns;

Remarks

- Visualizzare fino a ~100 barre;
- Riuscire a raggruppare/confrontare items, patterns.

Sentiment of contents - Netflix



Distribution of Content Ratings - Netflix

Pie chart

What?

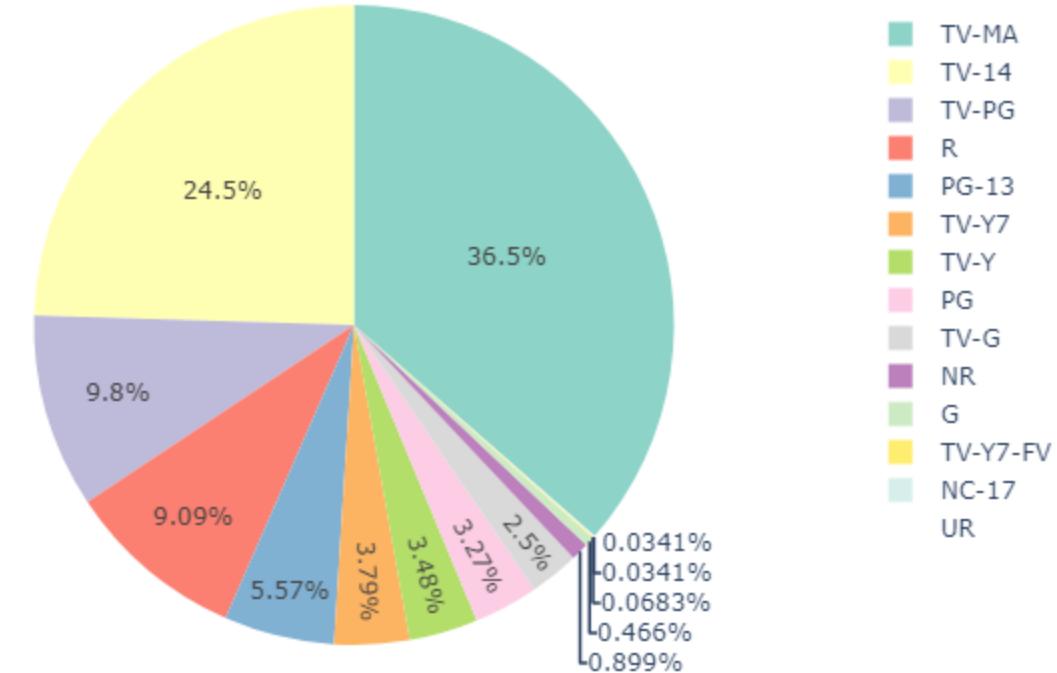
- 1 quantitative attribute;
- 1 categorical key;

Why?

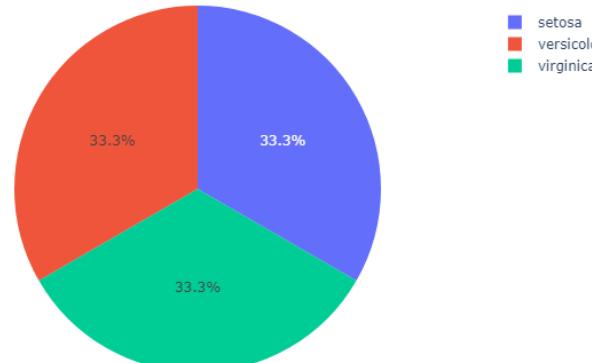
- Evidenziare una parte rispetto al tutto;

Remarks

- Meno dettagliato ed accurato rispetto al layout lineare;
- L'area centrale può essere rimossa (donut chart).



Iris Species



Word cloud

Word cloud - Netflix

What?

- 1 categorical attribute -> testo;
- 1 quantitative attribute -> frequenza;

Why?

- Visualizzare distribuzione di parole;
- Visualizzare un sommario;

Remarks

- Scarsa accuratezza;
- Bias dovuto alla lunghezza e alla struttura delle parole.



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

2 Categorical Keys

Heatmap

...



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Definizione

MATRICE DI CONFUSIONE: è un metodo per visualizzare le performance di un algoritmo rispetto ad un problema di classificazione dove gli outputs possono essere due o più classi.

Nel caso di problema di classificazione binario (due classi in outputs) la matrice di confusione sarà composta da quattro elementi: **True Positive (TP)**, **False Positive (FP)**, **False Negative (FN)**, **True Negative (TN)**.

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

Inoltre, la matrice di confusione è estremamente comoda per calcolare *Precision*, *Recall*, *Accuratezza*, ... (se ne parlerà nelle prossime lezioni).



Heatmap

What?

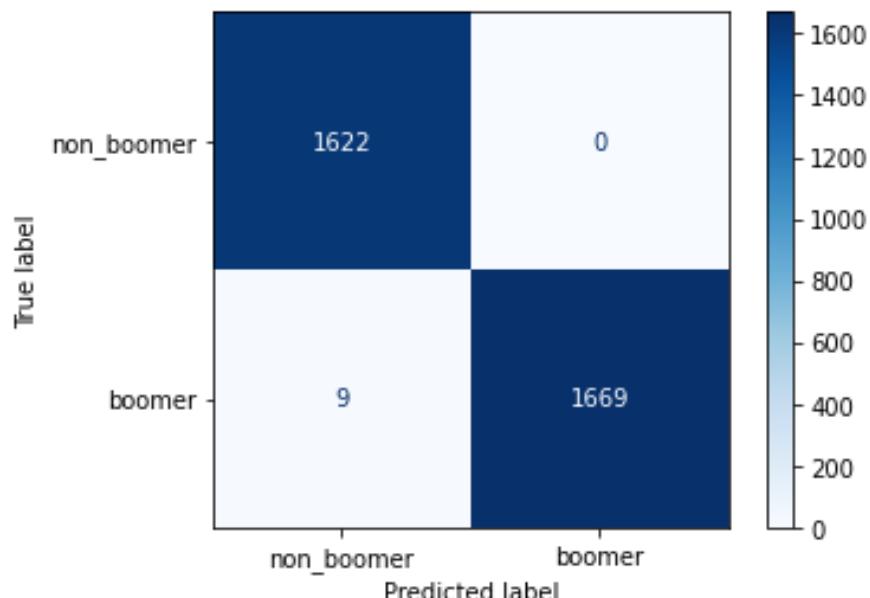
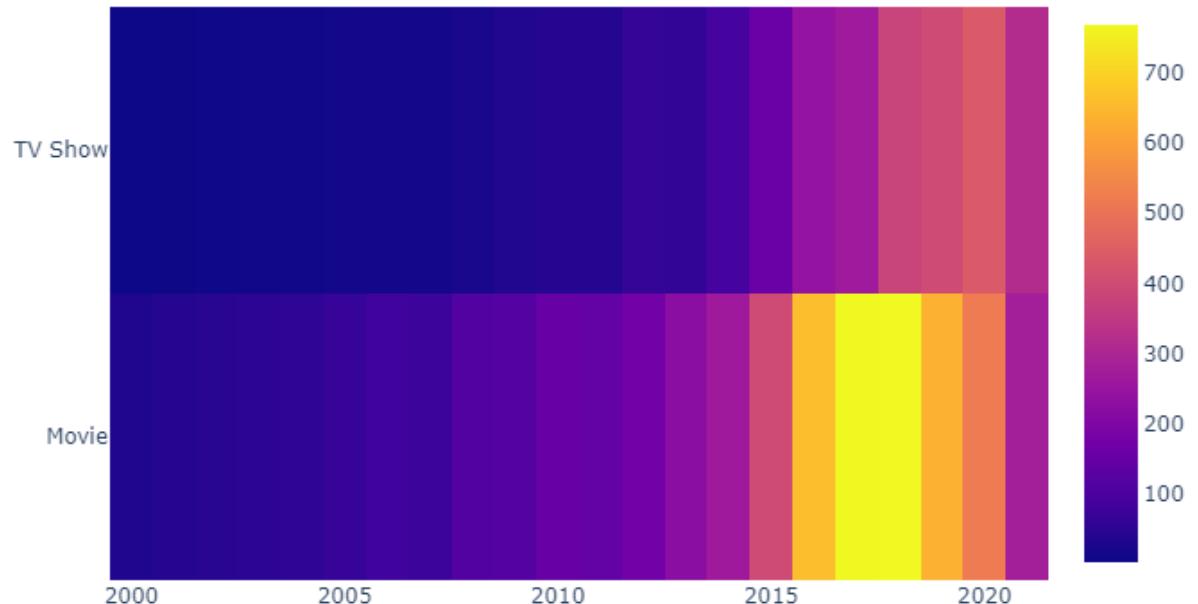
- 2 categorical key;
- 1 quantitative attribute;

Why?

- Visualizzare correlazioni;
- Identificare patterns, outliers;
- Confusion matrix for classification result visualization;

Remarks

- Fino a ~1M di items;
- L'ordine delle keys influisce la visibilità dei patterns.



For dealing with time

Line graph
Stacked area graph
...



Line graph

What?

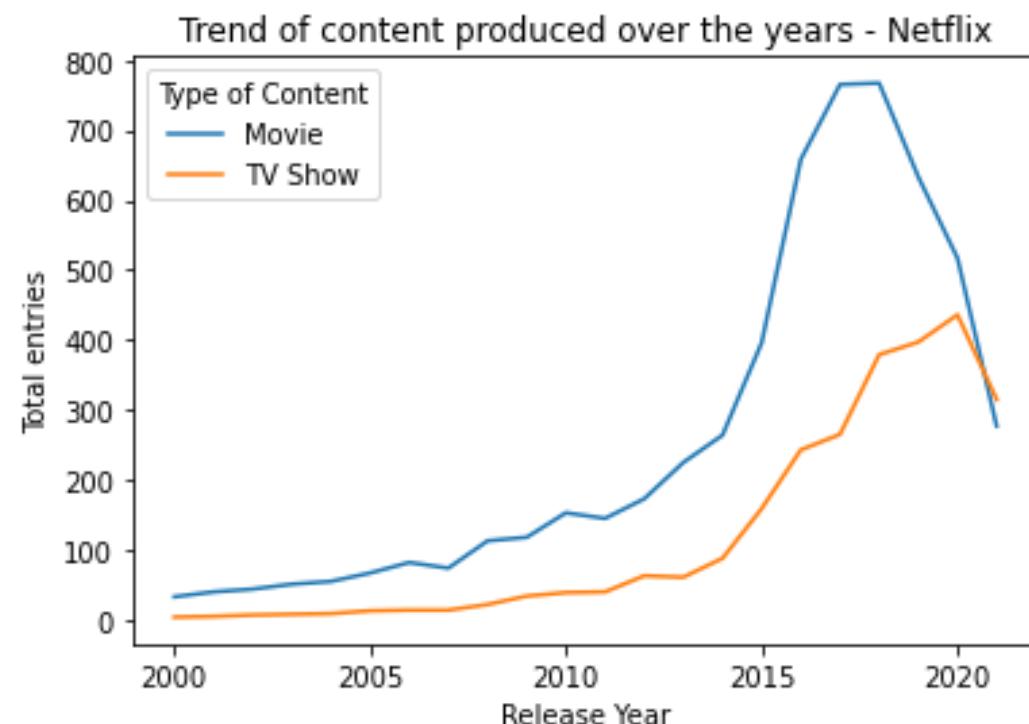
- 1 ordered key -> time;
- 1 quantitative attribute;

Why?

- Identificare e confrontare trends;

Remarks

- Fino a 10-20 linee;
- Il colore può codificare un categorical attribute additivo.

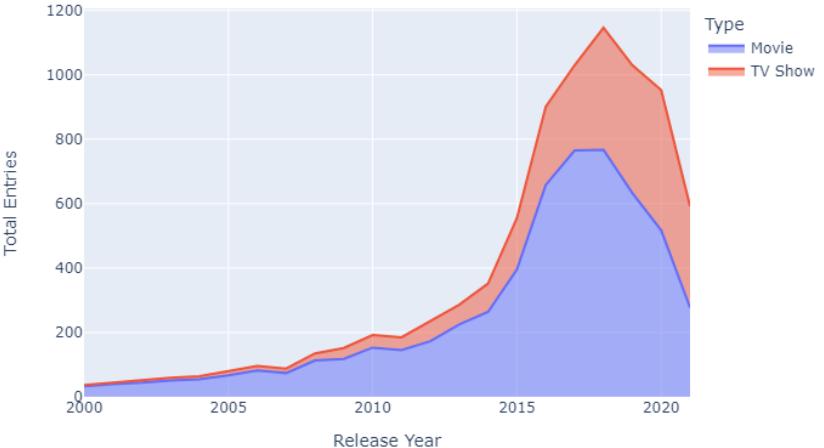


Stacked area graph

What?

- 1 ordered key -> time;
- 1 categorical attribute;

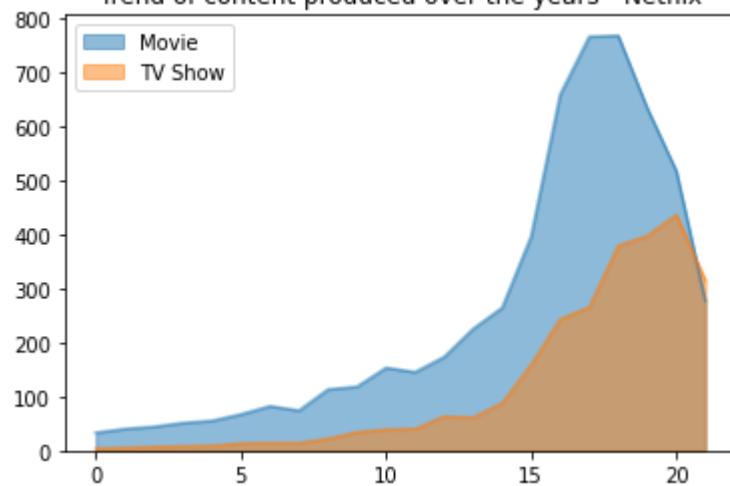
Trend of content produced over the years - Netflix



Why?

- Visualizzare trends;
- Evidenziare una parte rispetto al tutto;
- Confrontare valori;

Trend of content produced over the years - Netflix



Remarks

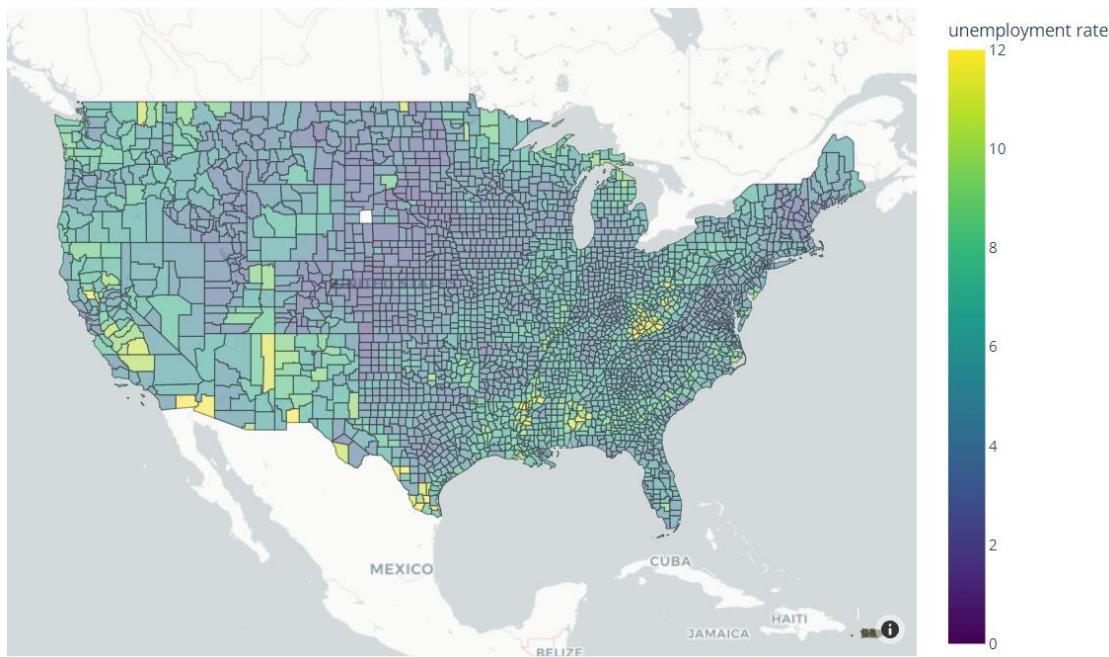
- Fino a pochi valori.

ATTENZIONE ALLA BASE DI RIFERIMENTO DELLE AREE COLORATE



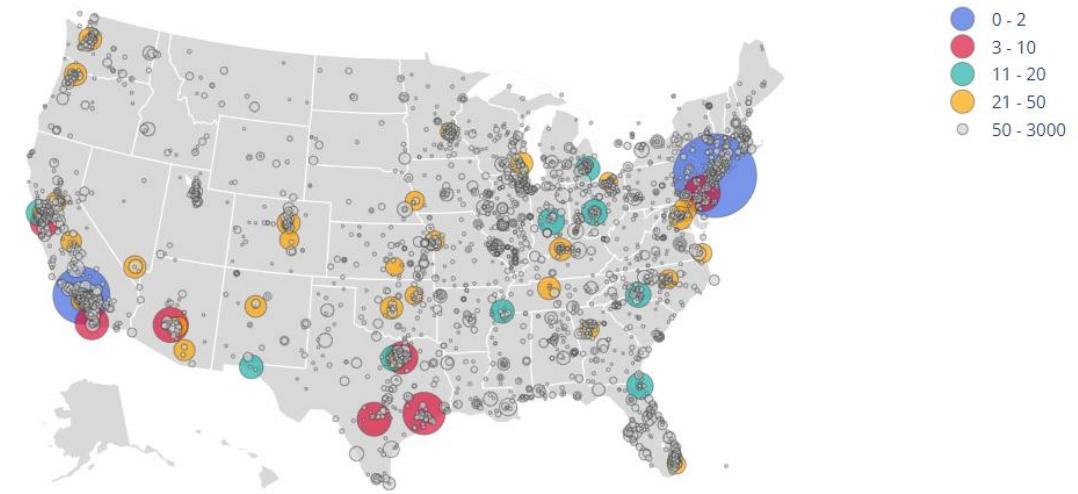
Esempi – SOLO immagine

Choropeth map e Bubble map



unemployment rate
12
10
8
6
4
2
0

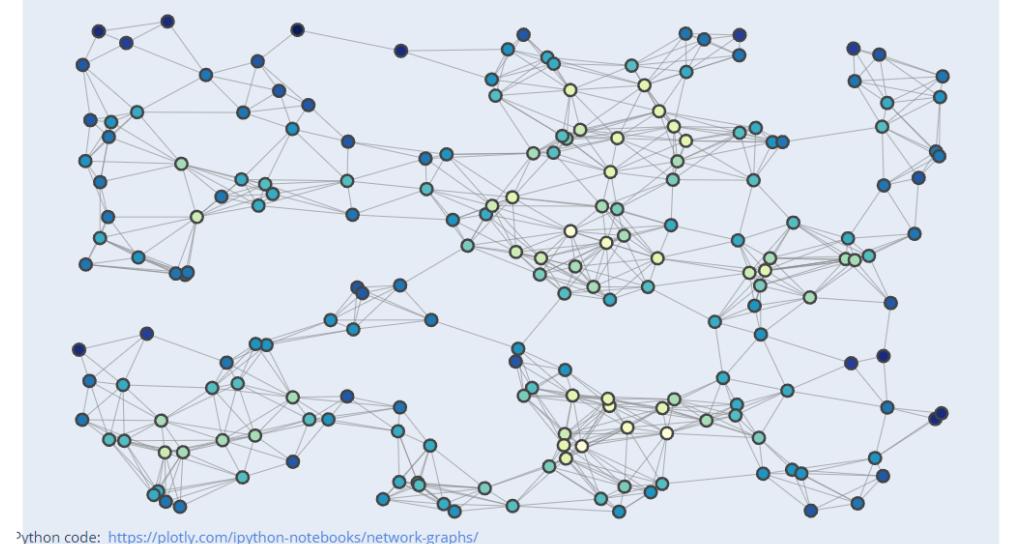
2014 US city populations
(Click legend to toggle traces)



Esempi – SOLO immagine

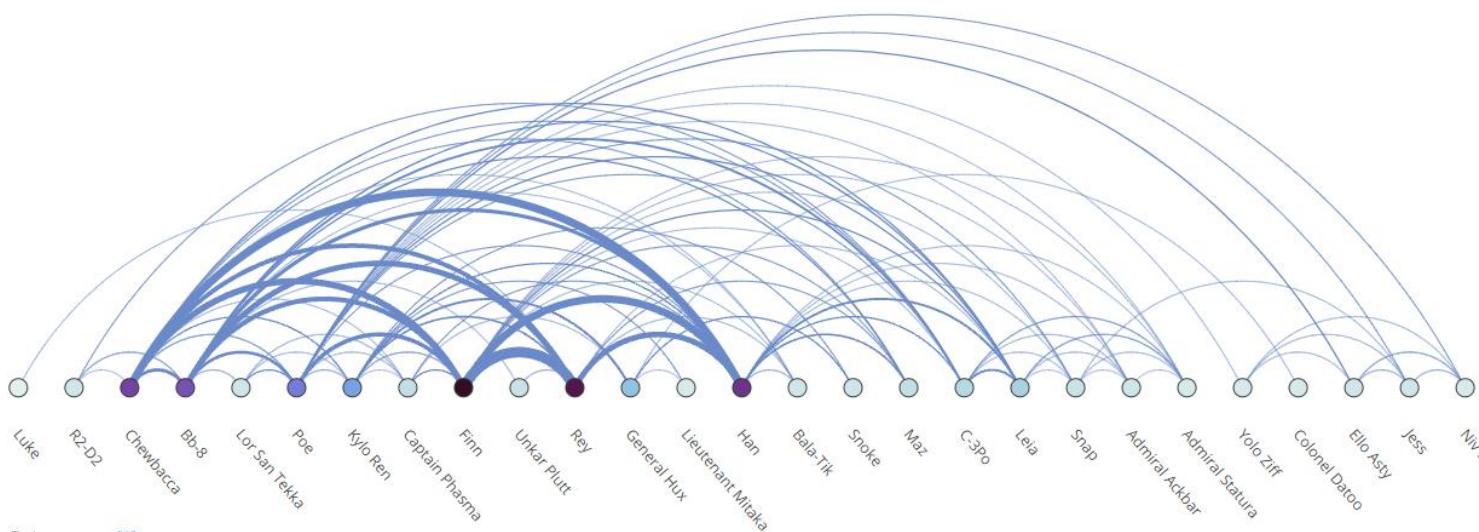
Node-link diagram e Arc diagram

Network graph made with Python



Python code: <https://plotly.com/ipython-notebooks/network-graphs/>

Arc Diagram of Star Wars Characters that Interacted in The Force Awakens



Data source: [1]

<https://plotly.com/python/network-graphs/>

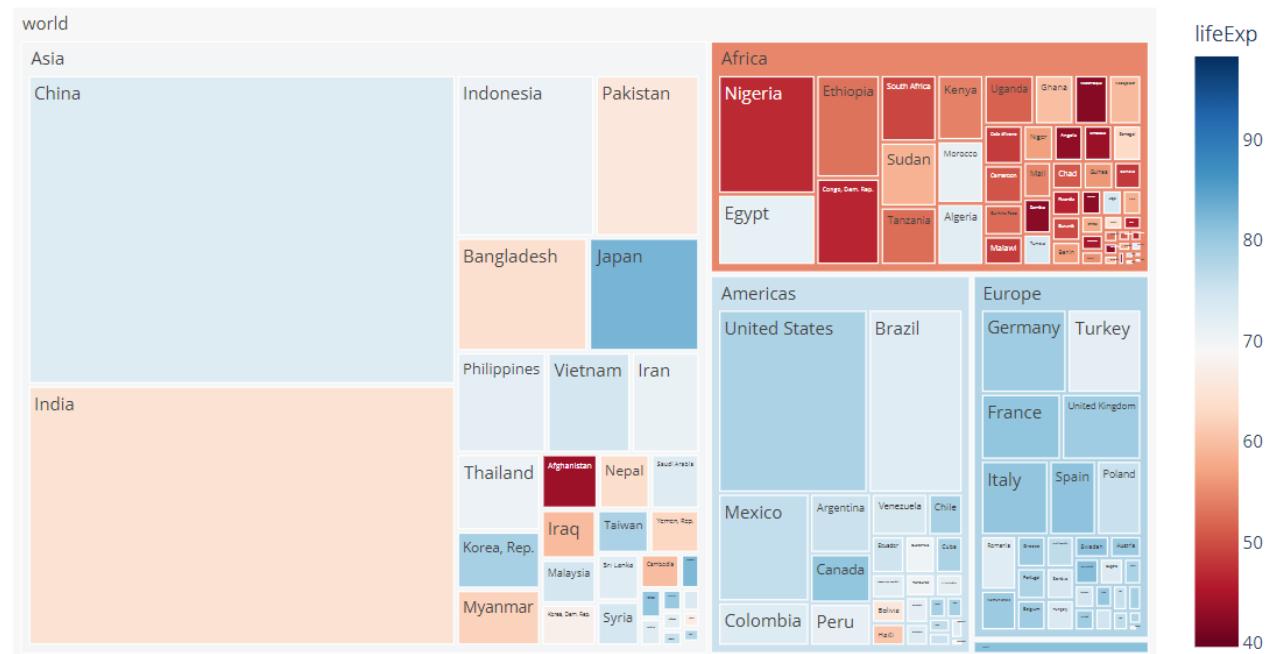
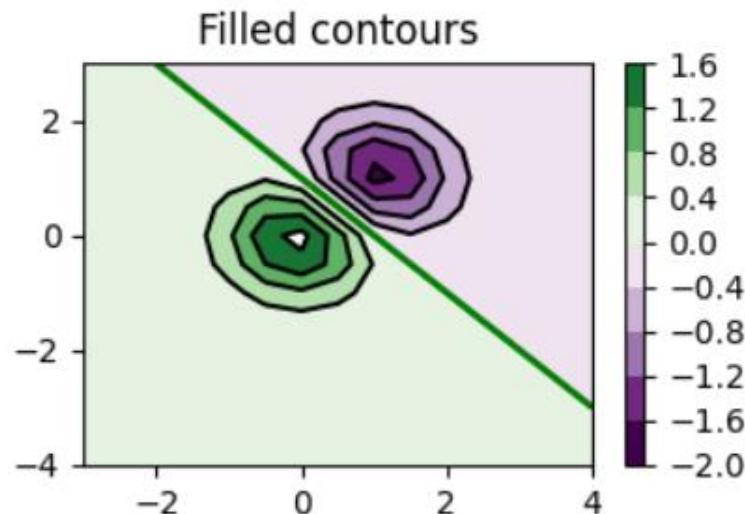
<https://chart-studio.plotly.com/~empet/13574/arc-diagram-of-star-wars-characters-that-interacted-in-the-force-awakens/#plot>



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Esempi – SOLO immagine

Isocontour plot e Tree map



https://matplotlib.org/stable/gallery/images_contours_and_fields/contour_image.html#sphx-glr-gallery-images-contours-and-fields-contour-image-py

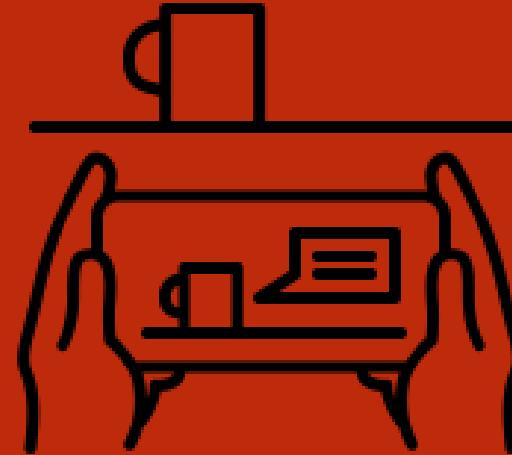
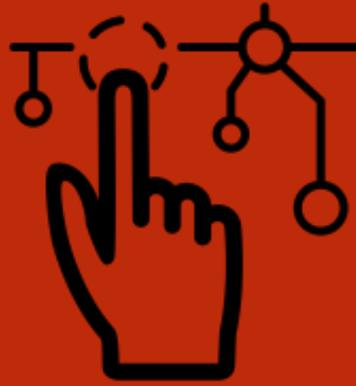
<https://plotly.com/python/treemaps/>



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA



Data Visualization – Visualizzazione Interattiva & AR

Alessia Angeli

Studente di dottorato in Data Science and Computation
Dipartimento di Informatica – Scienza e Ingegneria

Visualizzazione interattiva

L'utente ha la possibilità di:

Cambiare

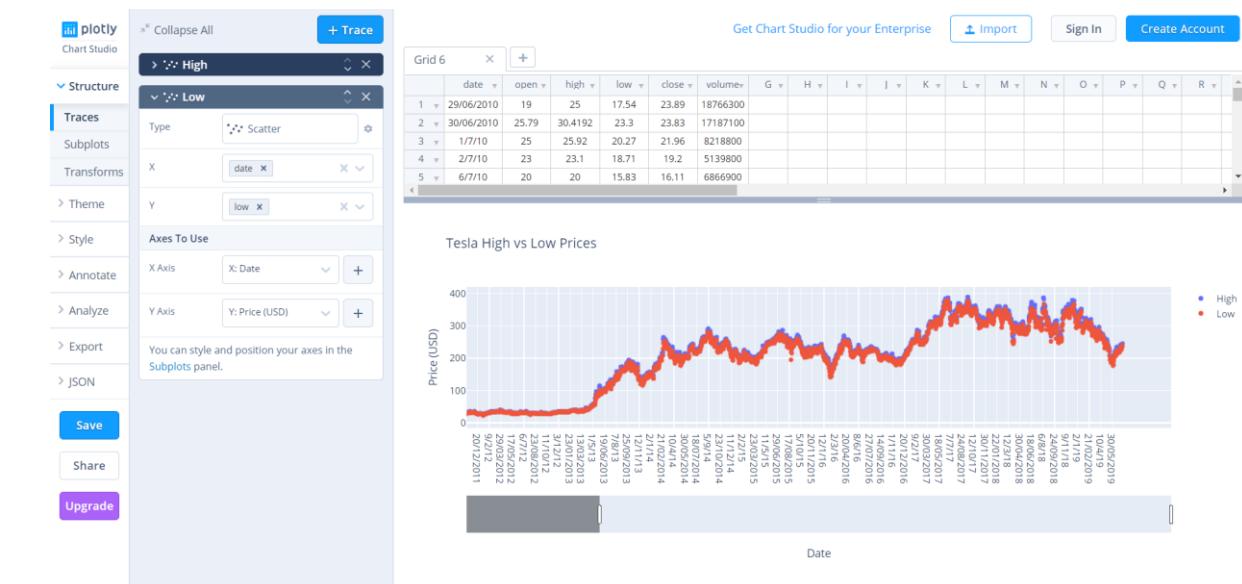
- Codifica (tipo di grafico)
- Parametri (colore, dimensioni, aggiungere elementi al grafico)
- Organizzazione (allineamento colonne/righe)
- Ordine
- Tipo di aggregazione

Selezionare

- Items
- Tooltips per ulteriori informazioni

Navigare

- Tra items
- Tra attributi



2017 eclipse Sooner Later
Drag the globe to explore the eclipse paths.



<https://towardsdatascience.com/interactive-data-visualization-167ae26016e8>

<https://www.tableau.com/learn/articles/interactive-map-and-data-visualization-examples>



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Visualizzazione interattiva - VANTAGGI

- Tipo di visualizzazione flessibile, potente, intuitivo.
- L'analisi esplorativa dei dati può cambiare durante lo stesso processo di analisi.
- Possibile cambio di attività fluido attraverso codifiche visive diverse a supporto di attività diverse.
- Le transizioni animate possono fornire un supporto eccellente alla visualizzazione.
- C'è un'evidenza empirica che le transizioni animate aiutino le persone a rimanere concentrate.

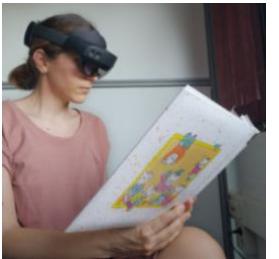


Visualizzazione interattiva - SVANTAGGI

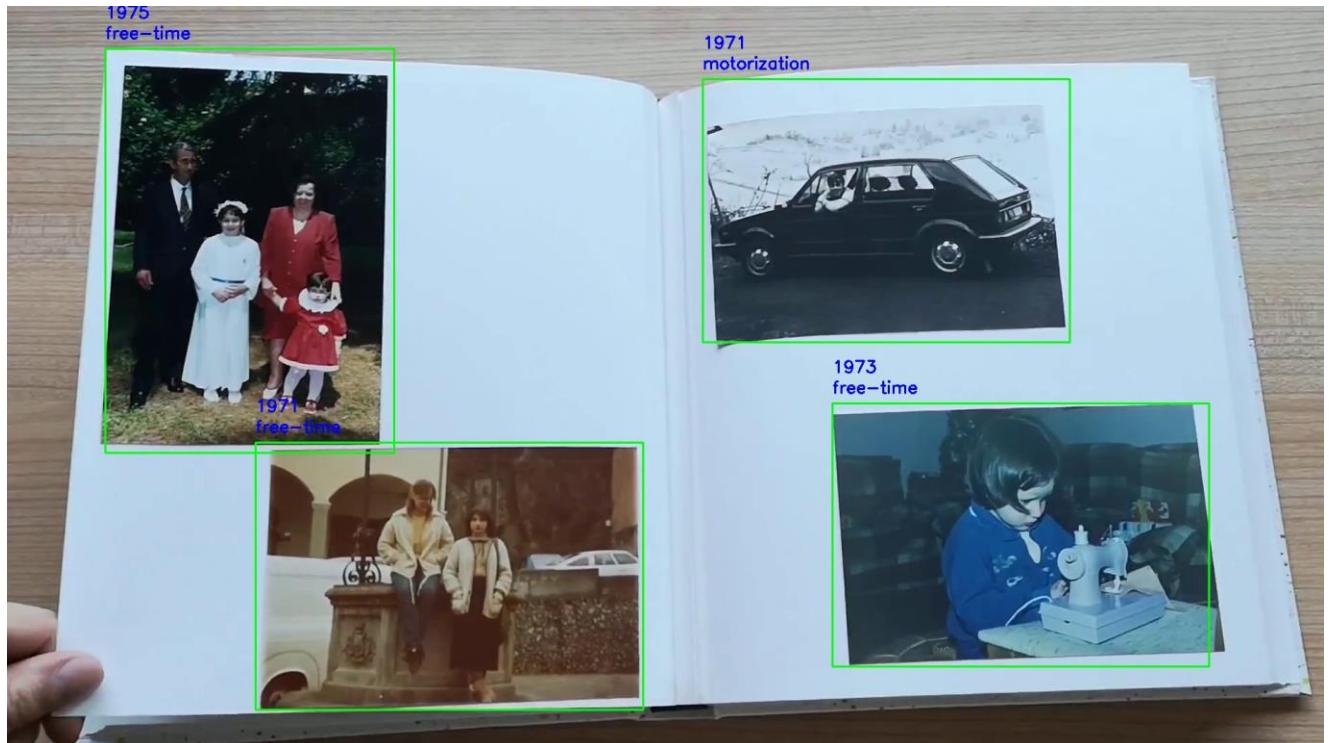
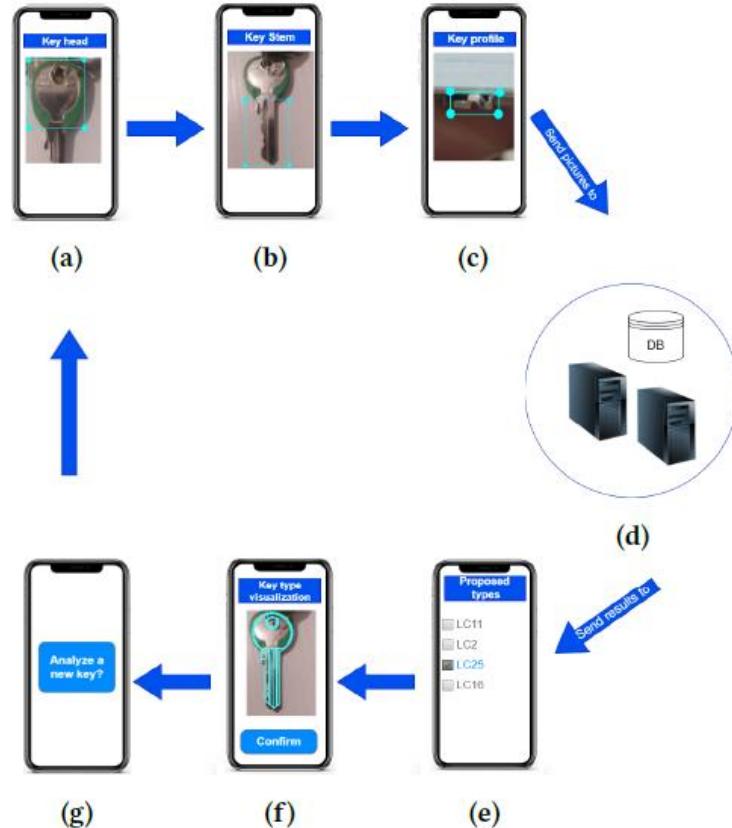
- L'interazione ha un costo in termini di tempo (variabile, a volte significativo)
- L'interazione impone un carico cognitivo.
- I controlli per l'interazione potrebbero richiedere molto spazio sullo schermo.
- Gli utenti potrebbero non interagire come pianificato dal designer (e.g., i registri del NYTimes mostrano che circa il 90% degli utenti non interagisce oltre lo scrollytelling (Aisch, 2016)).



Visualizzazione e Augmented Reality (AR)



Visualizzazione e Augmenter Reality (AR) – alcuni nostri progetti



[Stacchio, L., Angeli, A., Hajahmadi, S., Marfia, G. \(2021\). Revive Family Photo Albums through a Collaborative Environment Exploiting the HoloLens 2. In Proceedings of the 2021 IEEE International Symposium on Mixed and Augmented Reality Adjunct \(ISMAR-Adjunct\), to appear.](#)

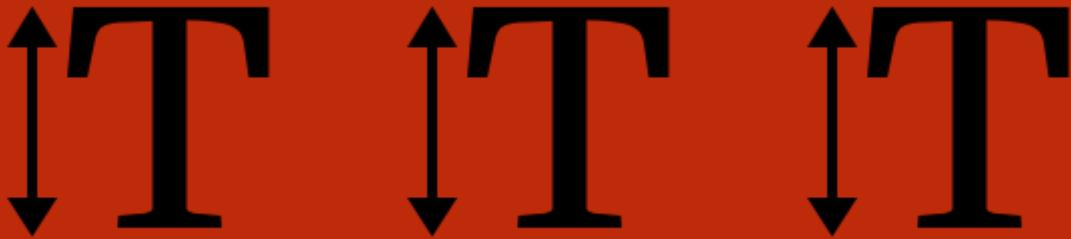
[Stacchio, L., Hajahmadi, S., & Marfia, G. \(2021, March\). Preserving Family Album Photos with the HoloLens 2. In *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops \(VRW\)* \(pp. 643-644\). IEEE.](#)

[Stacchio, L., Angeli, A., & Marfia, G. \(2021, September\). Empowering Locksmith Crafts via Mobile Augmented Reality. In *Proceedings of the Conference on Information Technology for Social Good* \(pp. 305-308\).](#)





ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA



Data Visualization – Principi della Gestalt & Tipografia

Alessia Angeli

Studente di dottorato in Data Science and Computation
Dipartimento di Informatica – Scienza e Ingegneria

7 ERESIE tipografiche



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Usa meno fonts,
meno dimensioni, e meno stili
possibili.



pt.2

Le linee più lunghe sono difficili da leggere, si dovrebbe ridurre il numero di caratteri per linea tra 50 e 70, in modo che gli occhi umani possano facilmente gestire la riga successiva. Lo stesso vale per i caratteri piccoli, sono anche difficili da leggere. Le persone devono avvicinarsi.



pt.3

*Usate le dimensioni dei caratteri per creare delle gerarchie.
Altrimenti, tutto ha la stessa priorità.*



pt.4

Quando usate font di dimensioni diverse,
date loro il giusto
contrasto.



Evita i riquadri

Se sono
veramente
necessari

Evitali
comunque



Non abbiate paura degli spazi vuoti: sono vostri amici. Hanno la stessa importanza degli spazi pieni. I vostri lettori vi ameranno.



Evitare il posizionamento centrale.

Altrimenti tutti gli altri elementi



Ruoteranno intorno al centro, in un layout molto confuso.



7 DOGMI tipografici



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

pt.1

**Usate meno font,
meno pesi e meno
stili possibile.**



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Le linee più lunghe sono difficili da leggere, si dovrebbe ridurre il numero di caratteri per linea tra 50 e 70, in modo che gli occhi umani possano facilmente gestire la riga successiva.

Lo stesso vale per i caratteri piccoli, che sono anche difficili da leggere. Le persone devono avvicinarsi.



pt.3

Usare le dimensioni del carattere per creare gerarchie.

Altrimenti, tutto avrà la stessa priorità.



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

pt.4

Quando si usano
diverse dimensioni
di caratteri,
date loro il giusto contrasto.



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Evita i riquadri

Se sono
veramente
necessari

Usa lo spazio e
contrasto per
definire blocchi



pt.6

Non fatevi spaventare
dagli spazi vuoti.
Sono vostri amici.

Hanno la stessa
importanza degli
spazi pieni.

I vostri lettori
vi ameranno.



pt.7

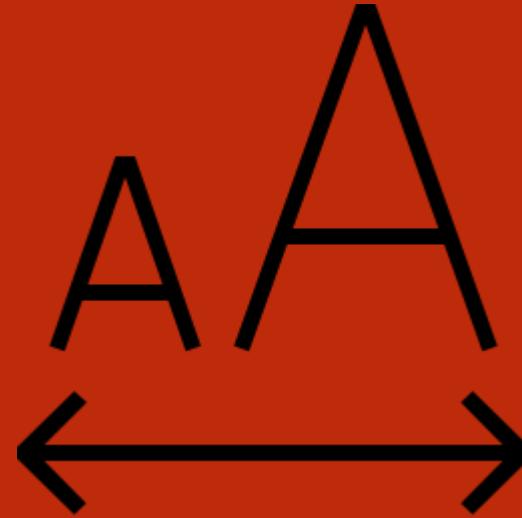


**Evita il posizionamento centrale.
Altrimenti tutti gli altri elementi ruoteranno
intorno a loro, in un layout molto confuso.**





ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA



Data Visualization – Fonts

Alessia Angeli

Studente di dottorato in Data Science and Computation
Dipartimento di Informatica – Scienza e Ingegneria

Come riconoscere un “buon” font

Un font ben disegnato:

1. Possiede tutti glifi, compresi anche quelli di alfabeti non latini.
2. Dispone di diversi pesi, dal Light al Black aumenta le possibilità espressive.
3. Alcuni font hanno anche la versione monospaziata (mono), consigliata per la visualizzazione di dati numerici.

Google fonts dispone di alcuni ottimi candidati:

Inter, IBM Plex, Source, Lato, Roboto, Fira, PT, Karla, Work Sans, etc.



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Glyph Overview



Family Overview

Styles

Regular
Italic
Semi Bold
Semi Bold Italic
Bold
Bold Italic
Black
Black Italic
Heavy
Heavy Italic

Gradient

About

Radial is a emotionally charged geometric-turned-humanist sans serif. Very informal geometry and tonal variety of 5 weights, makes it an absolute piece of contemporary workhorse with dramatic personality range, that performs flawlessly on both muscular display and small font sizes.

Formats

Open Type OTF, True Type TTF, WOFF, EOT

Encoding

Radial
Latin Extended, Cyrillic Extended, Greek

Details

Designer:
Milos Mitrovic

Released:
2018

Foundry:
Gradient

Contact

Milos Mitrovic
Nyhaugla 11
5230 Paradis,
Bergen, Norway

Info@wearegradient.net
www.wearegradient.net



LMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Font Features

Case Sensitive Forms	Avant-Garde «Hystory» {Hello} ¿Cómo estás? (Brackets)	AVANT-GARDE «HISTORY» {HELLO} ¿CÓMO ESTÁS? (BRACKETS)	Superscript Subscript	BERGEN1 OSLO2 CO ₂ H ₂ O	BERGEN ¹ OSLO ² CO ₂ H ₂ O
Discretionary Ligatures	official	official	Stylistic Set 1	Malaga Sangene Bang Балкан	Malaga Sangene Bang Балкан
Tabular Figures	08.06.1980 21.11.2011	08.06.1980 21.11.2011	Bulgarian Alternates	вгджзийклп тшщъю	үвгджзийкл нмщъю
Fractions	32 4/24 12 7/8	32 ¼ 12 ⅞	Circled Numbers	(1)(2)(3)(4)(5)(6)(7) (8)(9)	①②③④⑤⑥⑦⑧⑨
Ordinals Feminine, Masculine	No1 Sr.a	Nº1 Sr. ^a			



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA



Data Visualization – In parte...

Alessia Angeli

Studente di dottorato in Data Science and Computation
Dipartimento di Informatica – Scienza e Ingegneria

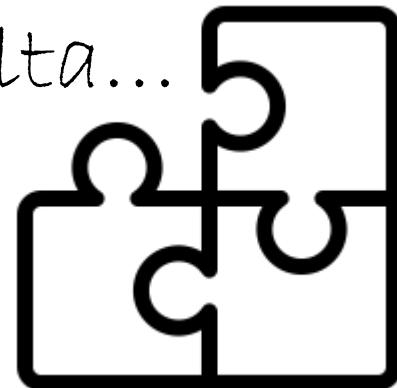
Solo una parte...

La visualizzazione dati non è solo quanto visto, è molto di più...

Una parte fondamentale è analizzare/selezionare i dati/risultati da mostrare. Spesso questa parte è una delle più complesse sia per analizzare i dati con l'obiettivo di costruire un modello (**feature selection** e **feature extraction**) sia con l'obiettivo di visualizzare dati e/o risultati (**analisi esplorativa**).

Ricordando poi, come già detto in precedenza, che come prima operazione i dati vanno controllati: dati mancanti, dati oggettivamente non corretti (e.g., età di uomini con valori negativi, prezzi di articoli con valori negativi), ...

un pezzo alla volta...





ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA



Data Visualization – Quindi...

Alessia Angeli

Studente di dottorato in Data Science and Computation
Dipartimento di Informatica – Scienza e Ingegneria

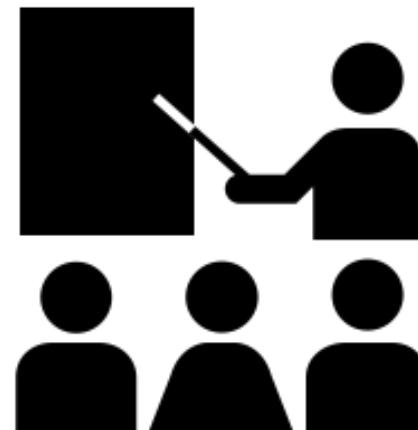
Visualizzazione efficace?

Come abbiamo visto, ci sono molti modi per definire l'efficacia ma... quello che conta di più è il fatto che i vostri utenti o lettori siano in grado di **estrarre informazioni dai dati visualizzati**:

- In modo accurato;
- Con uno sforzo ragionevole;
- Con grande sicurezza.

Il test è l'**acquisizione di conoscenza**:

- Se i vostri utenti non imparano niente di nuovo... c'è qualcosa che non va!
- L'acquisizione di conoscenza deve essere il vostro metro di valutazione.



ATTENZIONE

Il problema

- Capire al meglio di che cosa hai/hanno bisogno tu/gli utenti.

Dal problema alla visualizzazione

- Mostrare quanto richiesto.

Il tipo di visualizzazione

- Scegliere la strada per visualizzare i dati ottimale rispetto alla richiesta.



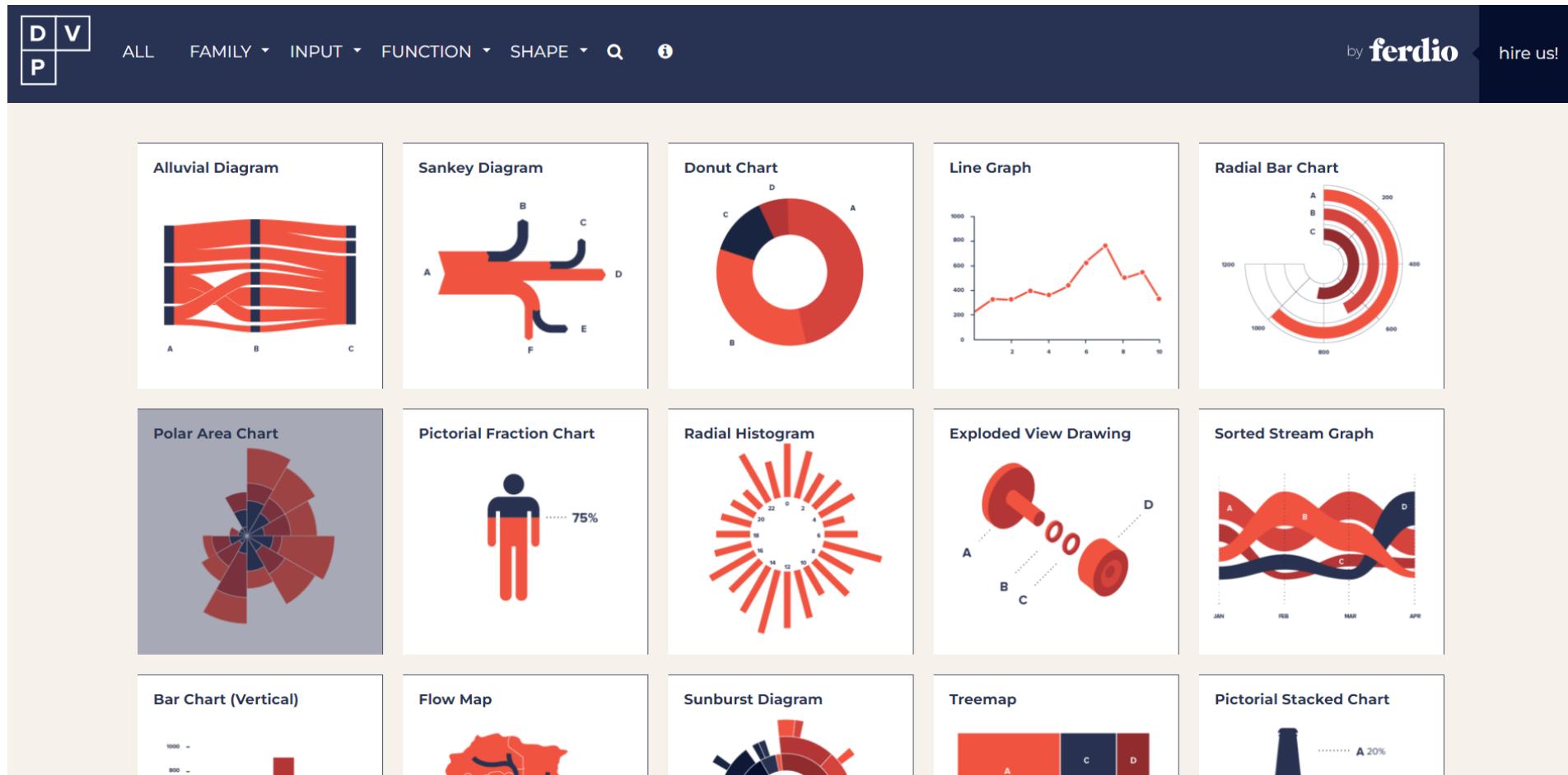
Un consiglio

«Sbagliando sì impara...» (?) – in Data Visualization sì!

La maggior parte delle visualizzazioni non è da subito ottimale e si continua ad imparare a crearne di migliori fallendo. Più fallisci, più la tua visualizzazione migliora. Non smettere di ripetere. Non accontentarti della prima visualizzazione che ti viene in mente! Genera e confronta alternative.



ALCUNE RISORSE – pt.1



<https://datavizproject.com/>



pt.2

The Data Visualisation Catalogue

About • Blog • Shop • Resources

CN 中文 ES Español RU Русский TR Türkçe

echo-analytics.io

Echo Analytics POIs

OPEN

Search by Function View by List



Arc Diagram



Area Graph



Bar Chart



Box & Whisker Plot



Brainstorm



Bubble Chart



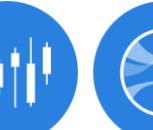
Bubble Map



Bullet Graph



Calendar



Candlestick Chart



Chord Diagram



Choropleth Map



<https://datavizcatalogue.com/>



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

pt.3

The image shows the homepage of the Flourish website. The background is a world map with numerous small, colorful dots scattered across it, representing data points. In the center, there is a white rectangular callout containing the following text:
**Beautiful and easy
data visualization
and storytelling**
Below this text is a white button with the text "Get started for free". At the bottom of the callout, there is a link "Find out more ↓". At the very bottom of the page, there is a blue horizontal bar with the URL <https://flourish.studio/>.

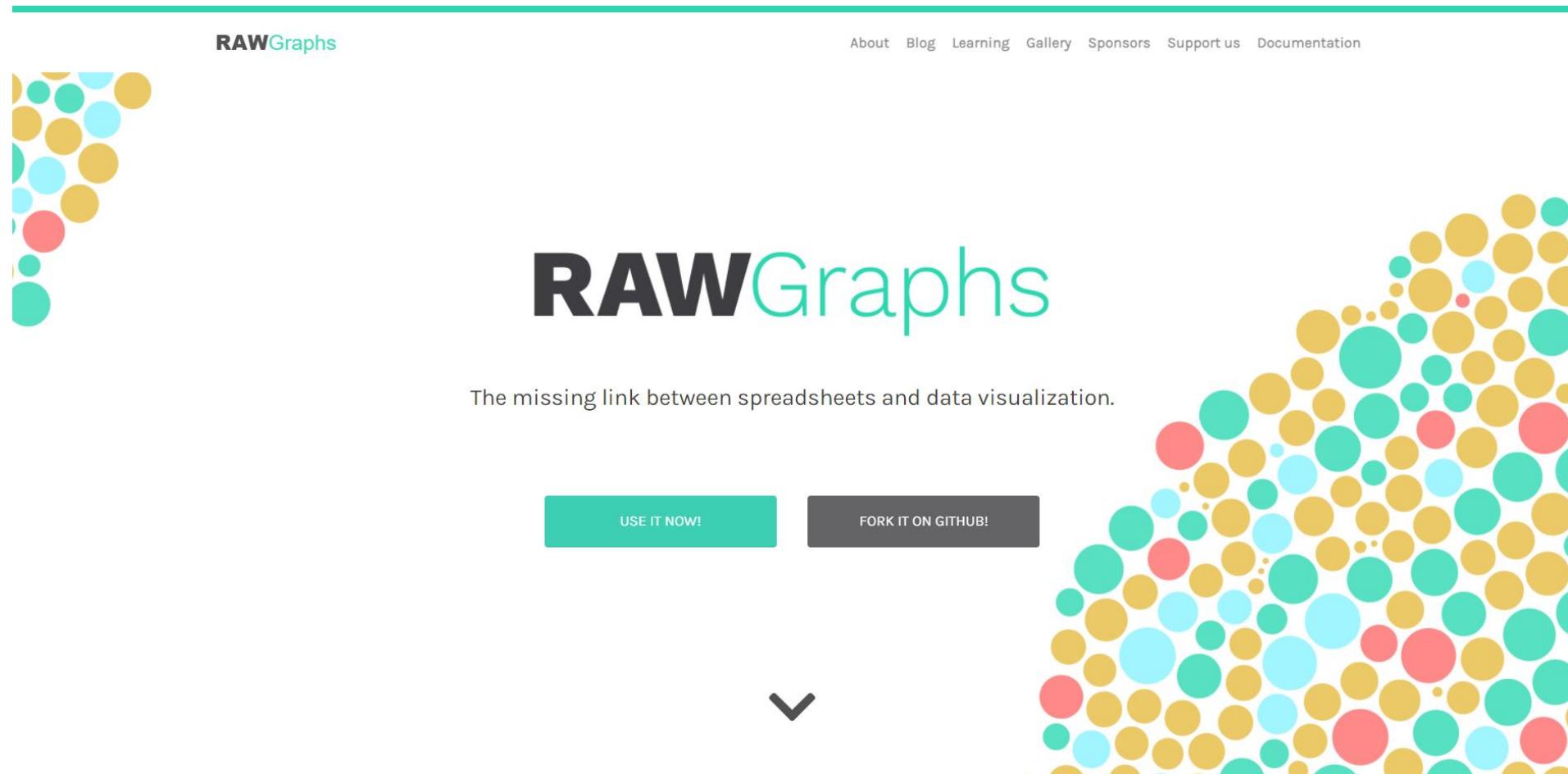


ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

pt.4

RAWGraphs

About Blog Learning Gallery Sponsors Support us Documentation



The missing link between spreadsheets and data visualization.

USE IT NOW! FORK IT ON GITHUB!

▼

<https://rawgraphs.io/>



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

pt.5



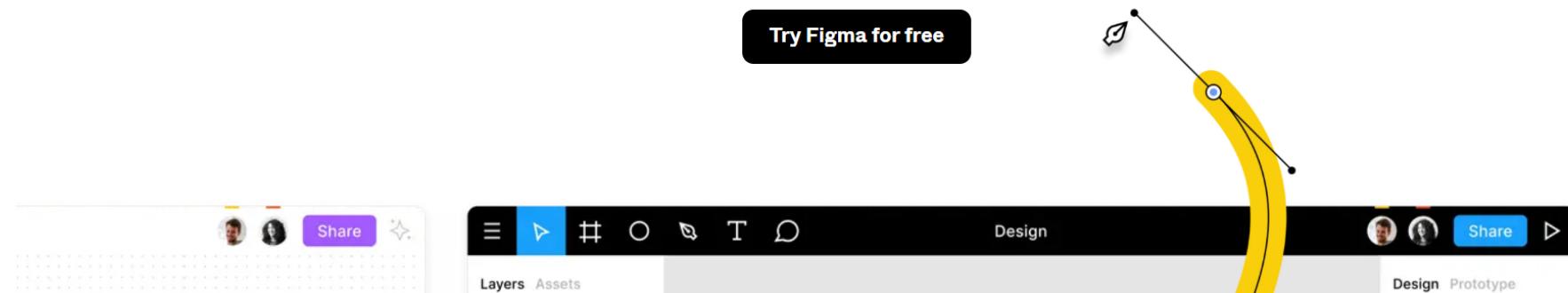
Products ▾ Enterprise ▾ Pricing Community ▾ Company ▾ Log in

Sign up

Nothing great is made alone.

Figma connects everyone in the design process
so teams can deliver better products, faster.

Try Figma for free



<https://www.figma.com/>



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Riferimenti

Corso «Data and Results Visualization», Daniele Loiacono, Politecnico di Milano (2019).

Corso Coursera (online) “Visualization for Data Journalism”, Margaret Yee Man Ng, University of Illinois at Urbana-Champaign (2021).

Workshop ART-ER, SCIENZA IBRIDA 2021, Data Visualization and Story Telling, Matteo Moretti, Sheldon.studio (2021).

Sitografia presente slide per slide.





ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Alessia Angeli

Dipartimento di Informatica – Scienza e Ingegneria

alessia.angeli2@unibo.it

www.unibo.it