

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Data Visualization

Rimini – 21/10/2021

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

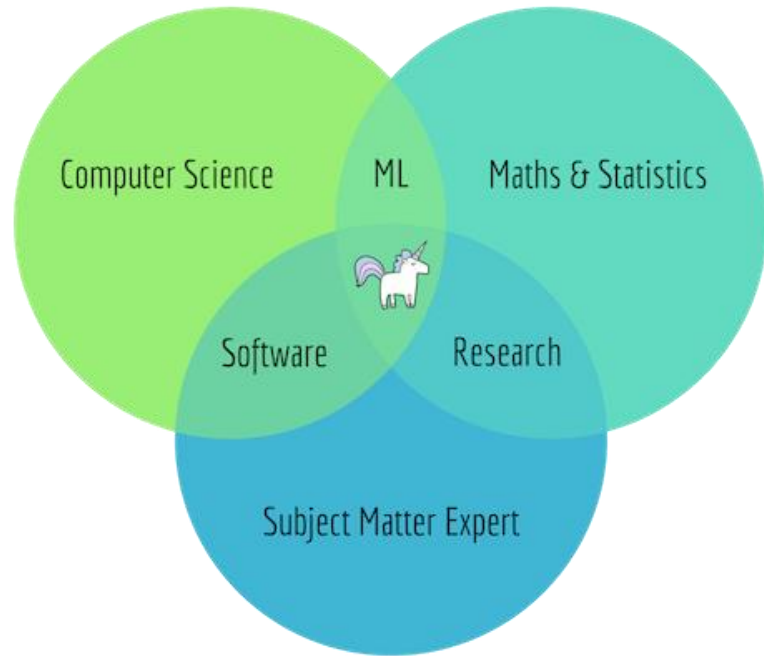


Contatti

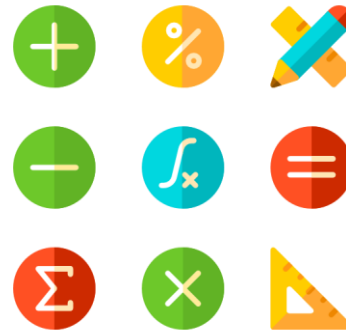


ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

DATA SCIENCE



Maths



Machine and Deep Learning



Data Visualization



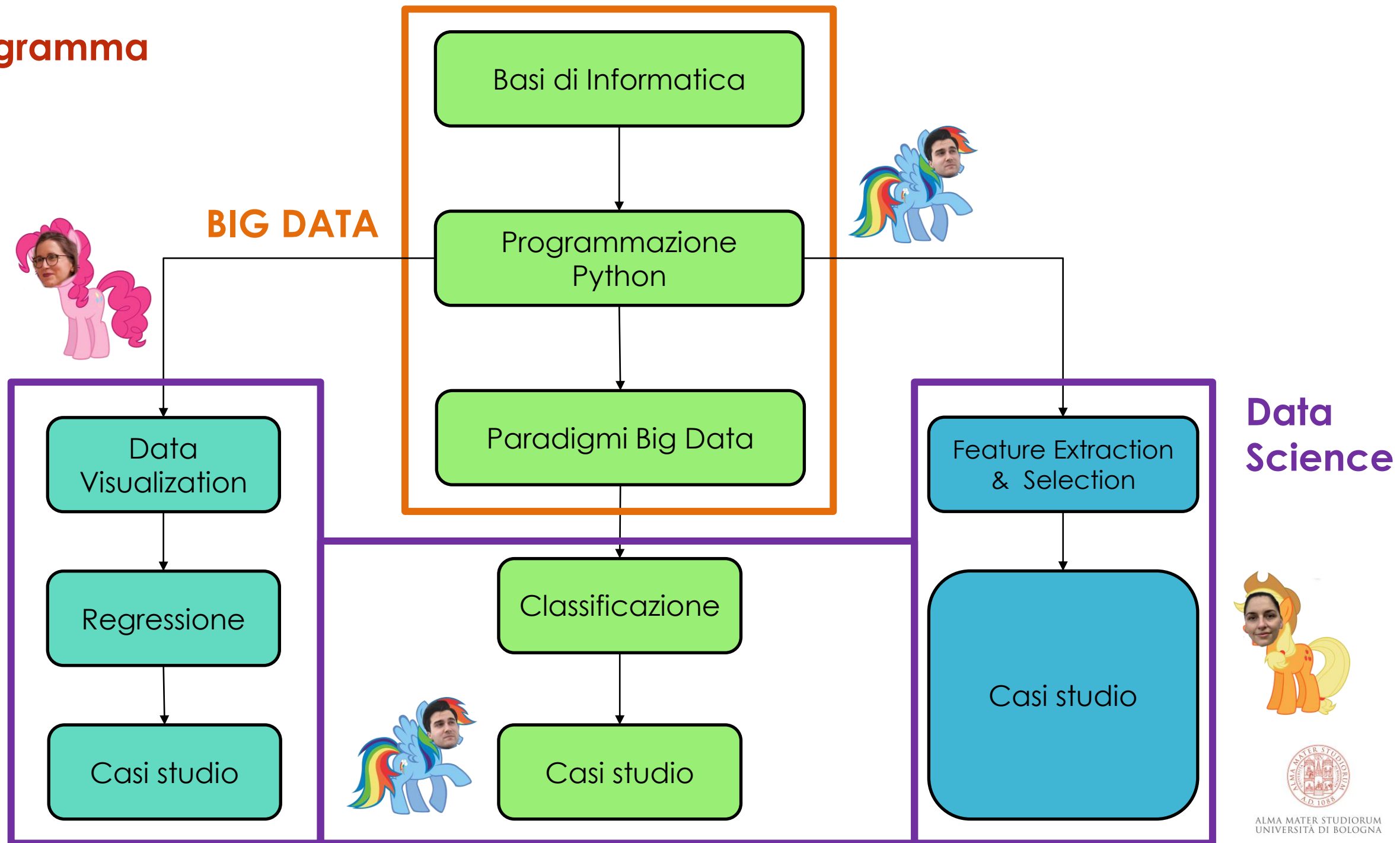
Augmented Reality

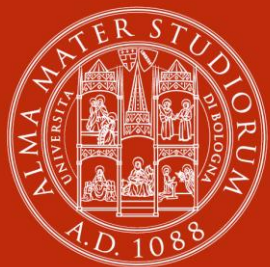


Unicorni fantastici e dove trovarli (data science)



Programma





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:

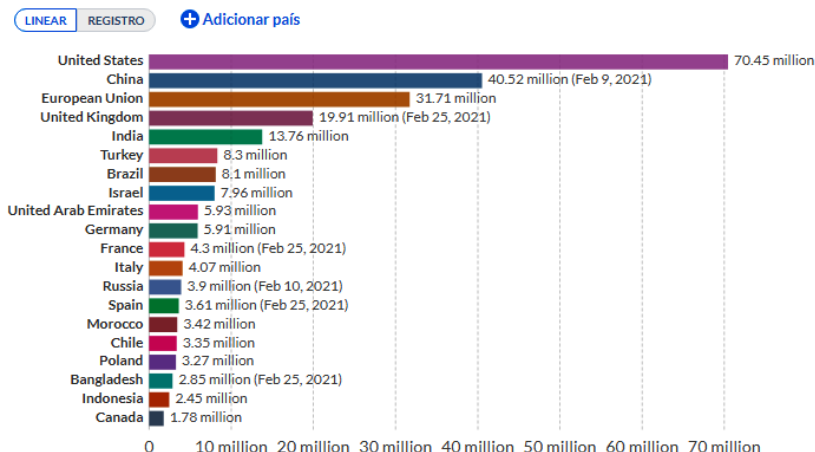
- difficile da valutare ed è facile commettere errori;
- altamente dipendente dal problema/dominio considerato.



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

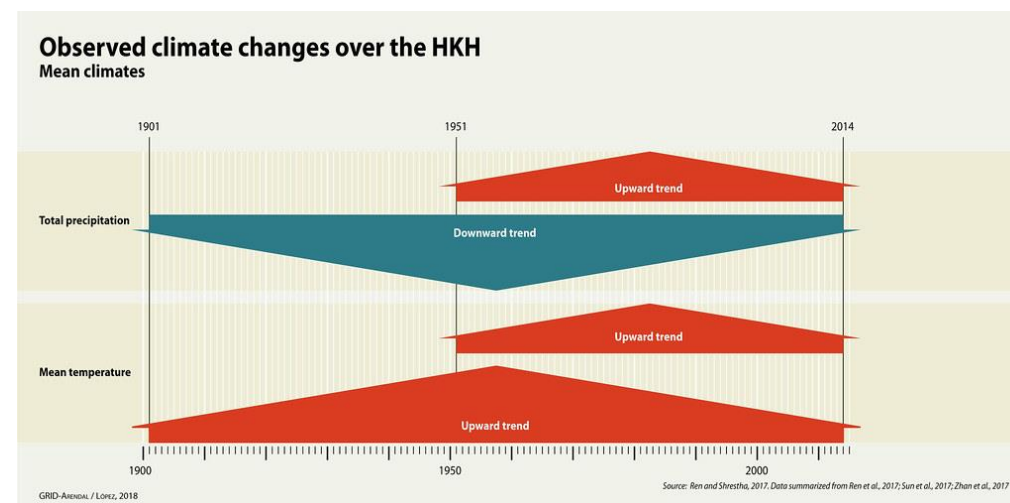
Número total de doses de vacinação administradas. Isso é contado como uma dose única e pode não ser igual ao número total de pessoas vacinadas, dependendo do regime de dose específico (por exemplo, as pessoas recebem doses múltiplas).

Our World
in Data



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

CC BY



<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](#))

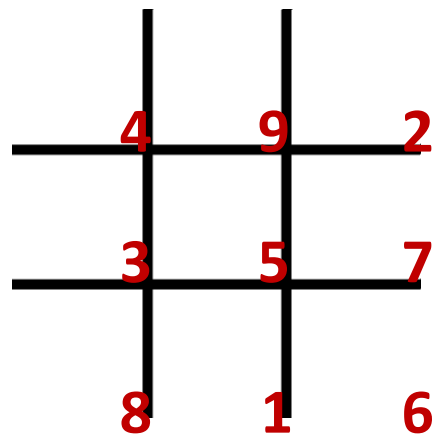
ci sono 2 giocatori

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

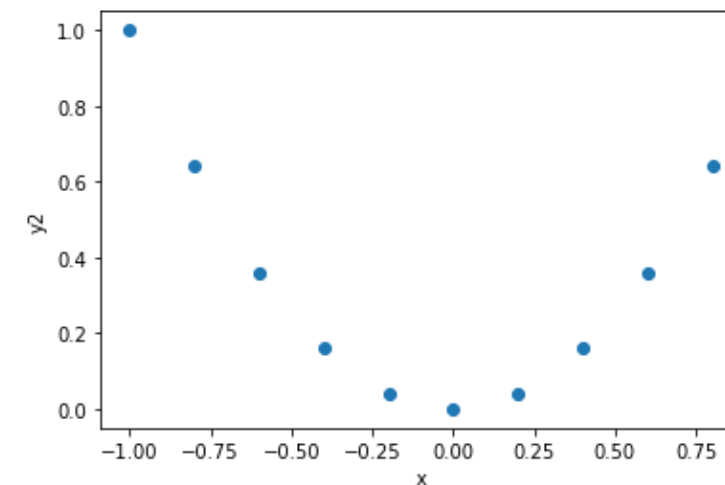
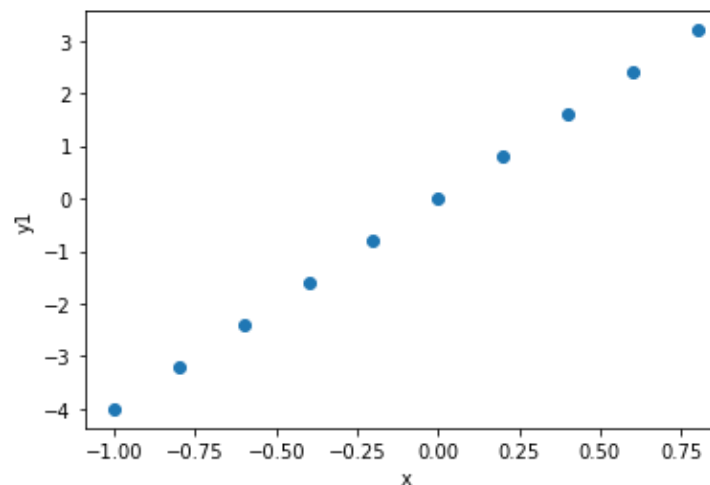
il primo giocatore che arriva ad avere tre cifre la cui somma è 15 vince

- Vi è familiare questo gioco?

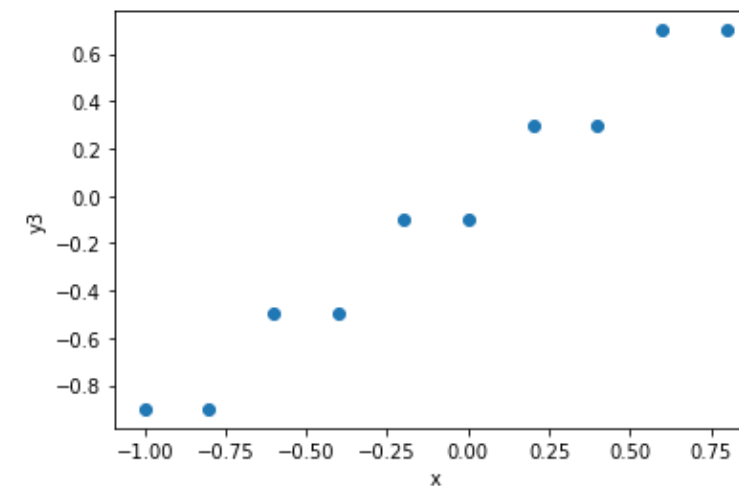


Perché visualizzare?

	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	0.60
min	-1.00	-4.00	0.00	-0.90
25%	-0.55	-2.20	0.07	-0.50
50%	-0.10	-0.40	0.26	-0.10
75%	0.25	1.40	0.57	0.30
max	0.80	3.20	1.00	0.70



Come creare una visualizzazione efficace?

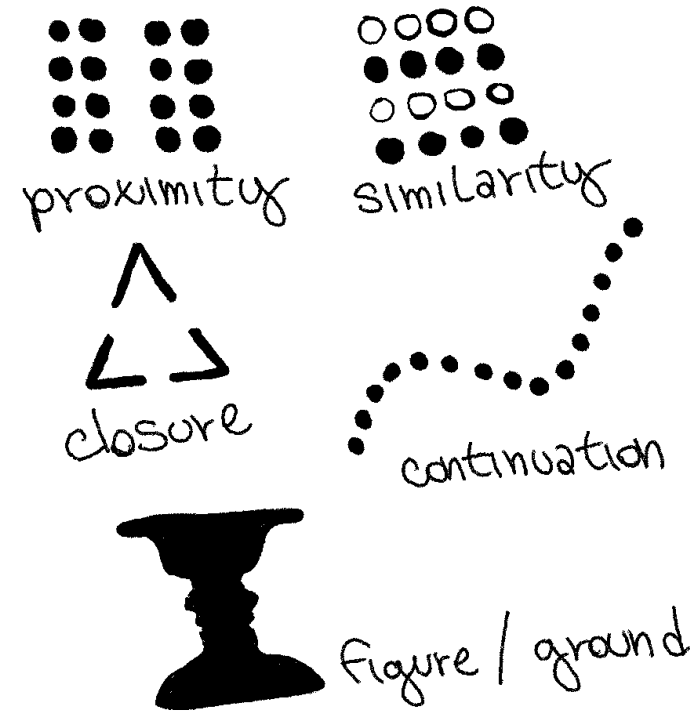
- Comprendere (almeno un po') la **percezione umana**;
- Definire gli **obiettivi**;
- Esplorare lo **spazio delle possibili visualizzazioni**.



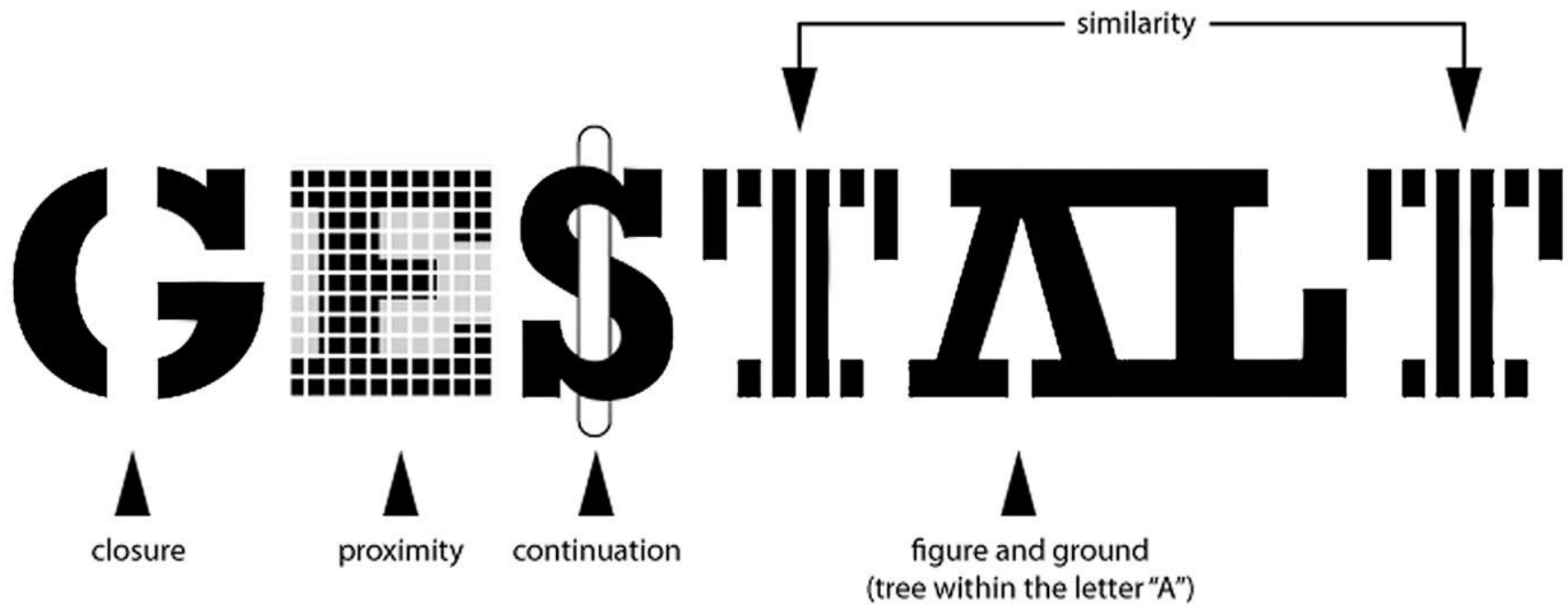
L'efficacia e la percezione – I principi della Gestalt

I principi della Gestalt:

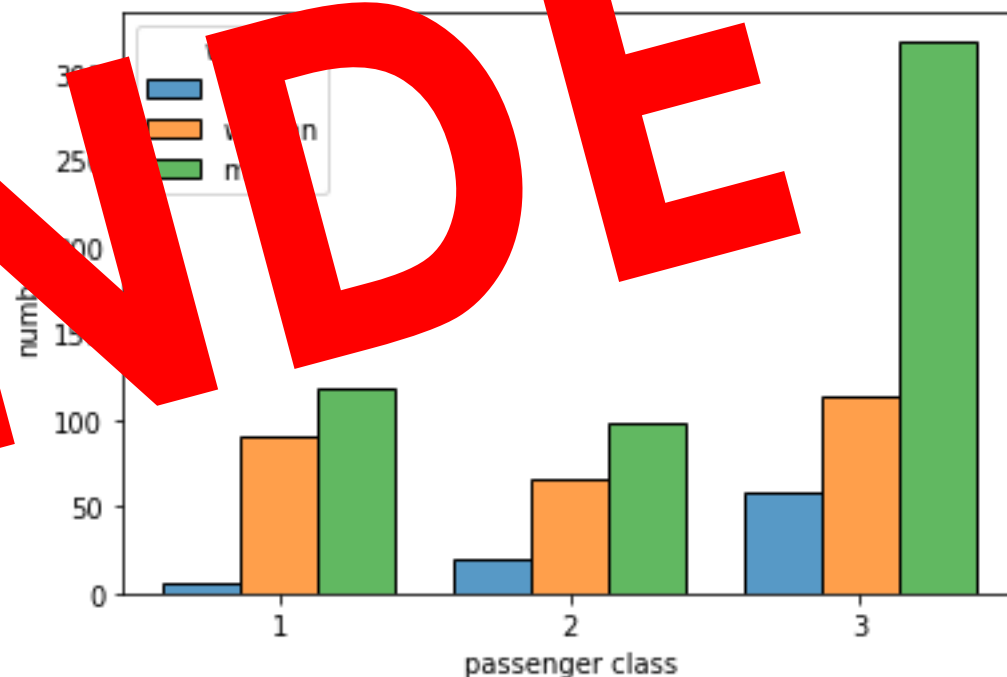
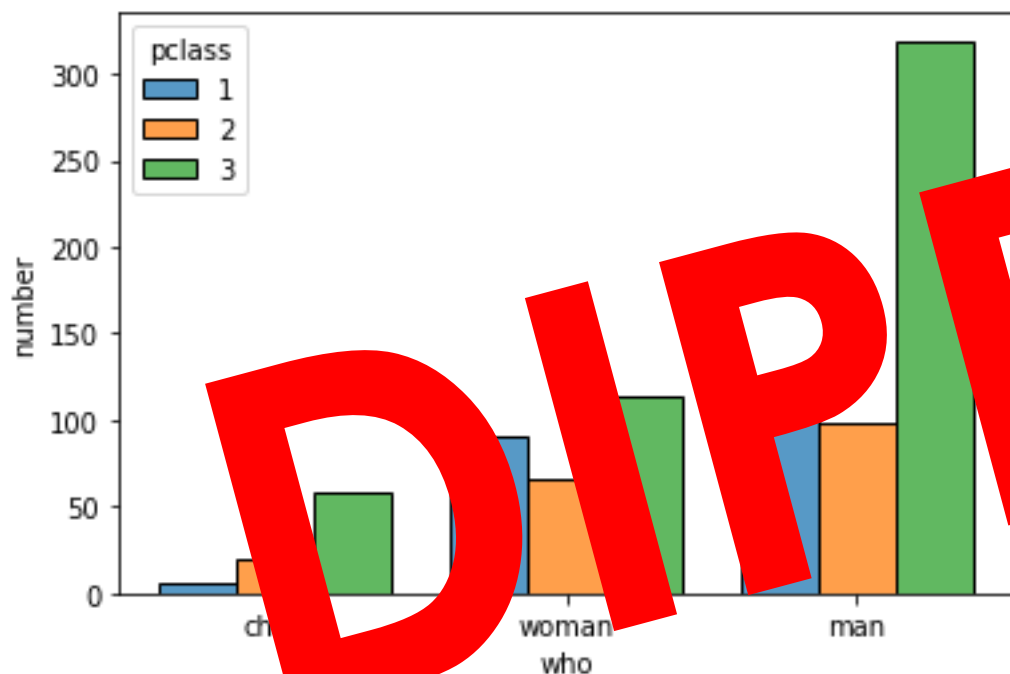
1. Principio di **prossimità** o vicinanza;
2. Principio di **similarità** o somiglianza;
3. Principio di **chiusura** o completamento;
4. Principio di **continuità** di direzione;
5. Principio di **contrasto** (figura/sfondo).



- 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?



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

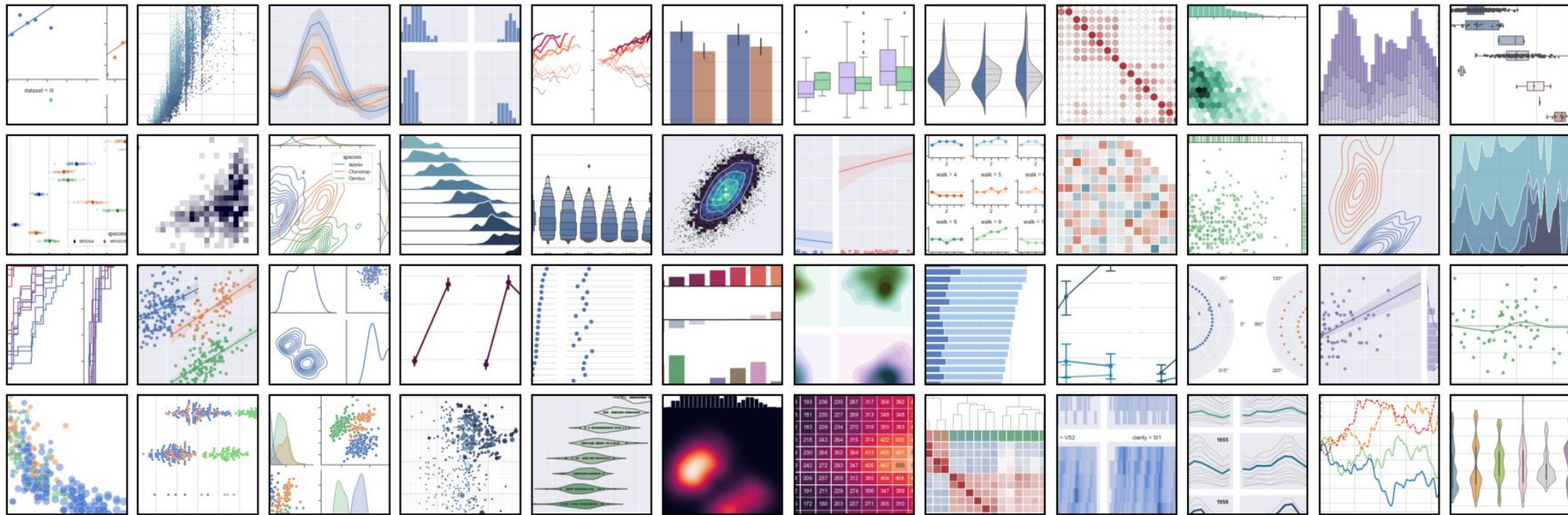


L'efficacia e lo spazio delle possibili visualizzazioni

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?

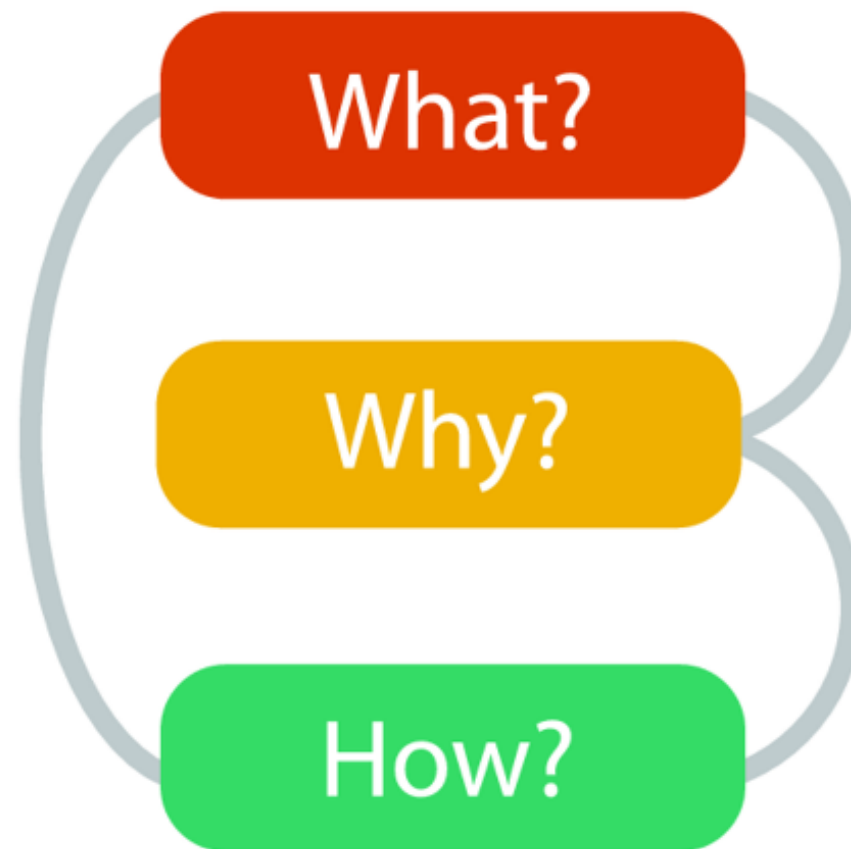
- Tipo di dati.

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	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



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

Individuare ed analizzare:

area di interesse



pubblico
(utenti)



problema



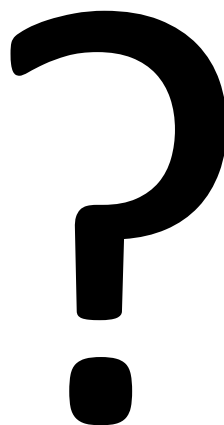
obiettivi da
raggiungere

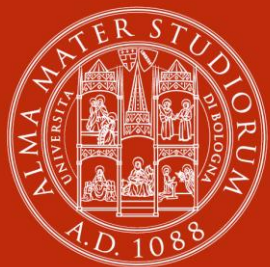


HOW? – La visualizzazione dati

In particolare, in queste lezioni vedremo:

- Le principali tipologie di grafici (e.g., istogramma, line plot, diagramma a torta, box plot);
- Come costruire i grafici utilizzando alcune librerie di Python (i.e., matplotlib, seaborn, plotly);
- Come utilizzare questi strumenti per visualizzare al meglio dati di interesse.





ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA



Data Visualization – Elementi e Principi

Alessia Angeli

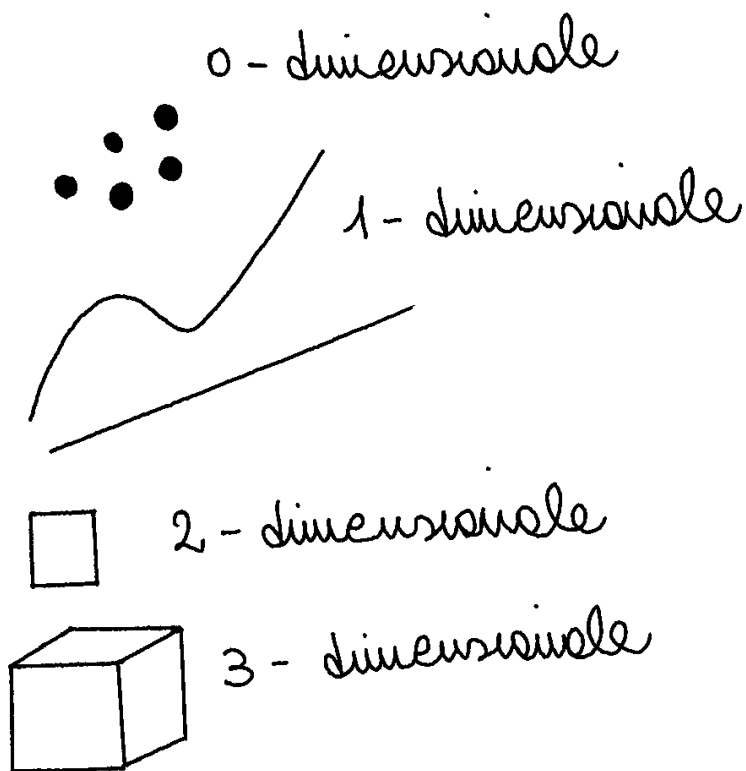
Studente di dottorato in Data Science and Computation

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.

➞ Position

➞ Horizontal



➞ Vertical



➞ Both



➞ Color



➞ Shape



➞ Tilt

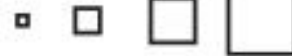


➞ Size

➞ Length



➞ Area

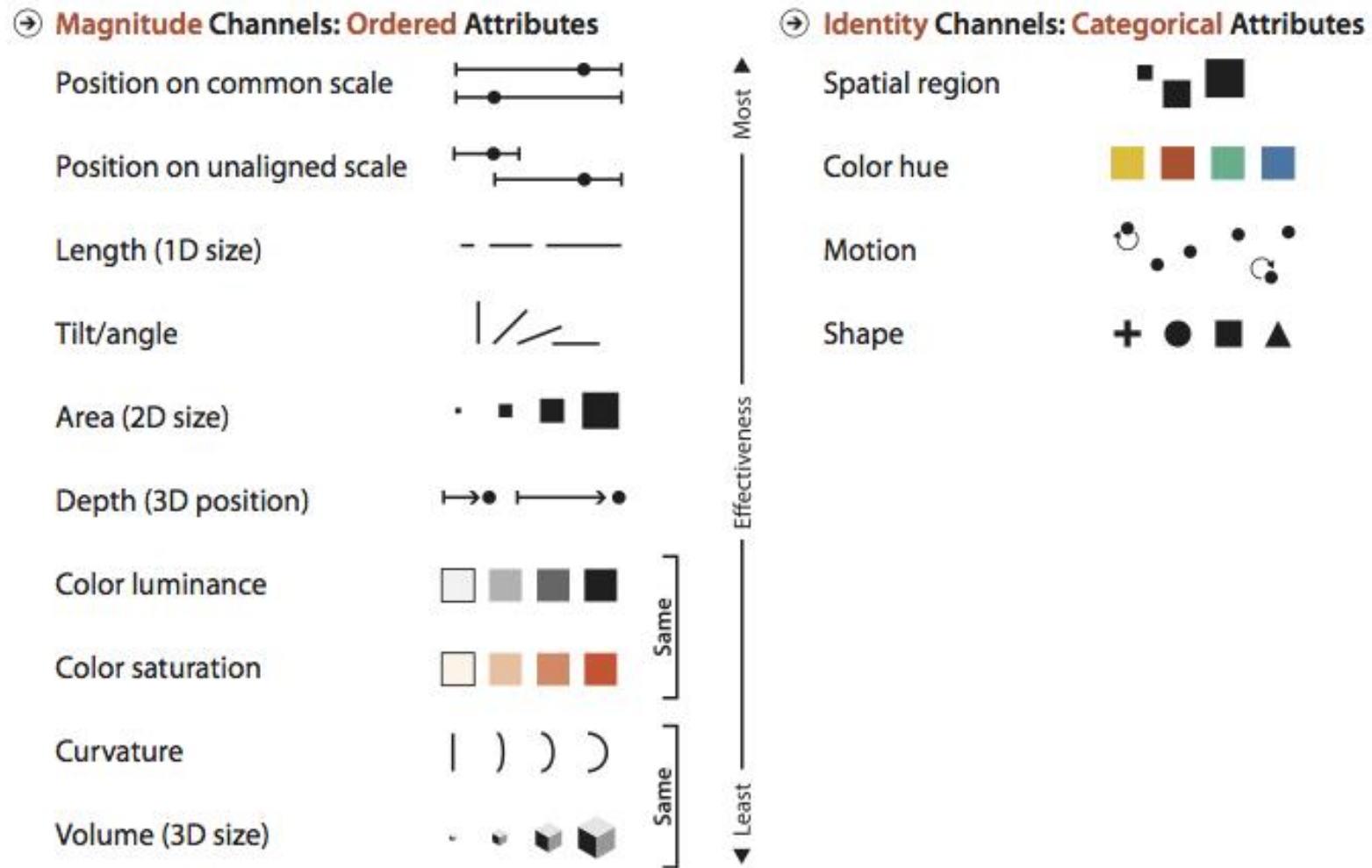


➞ Volume



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:

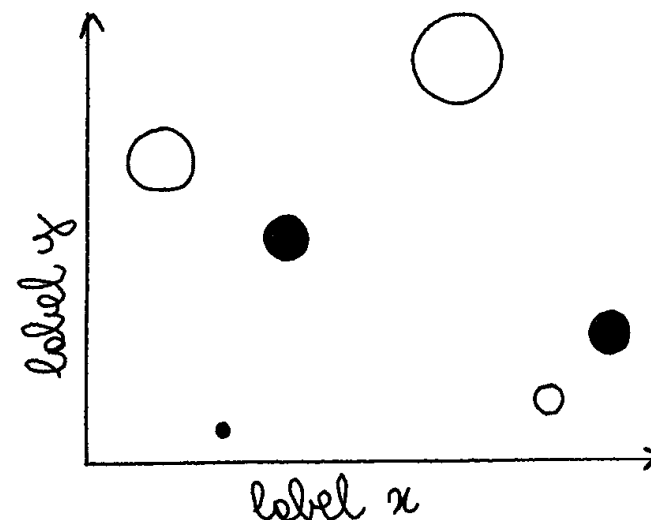
- Items
- Valori
- Links

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

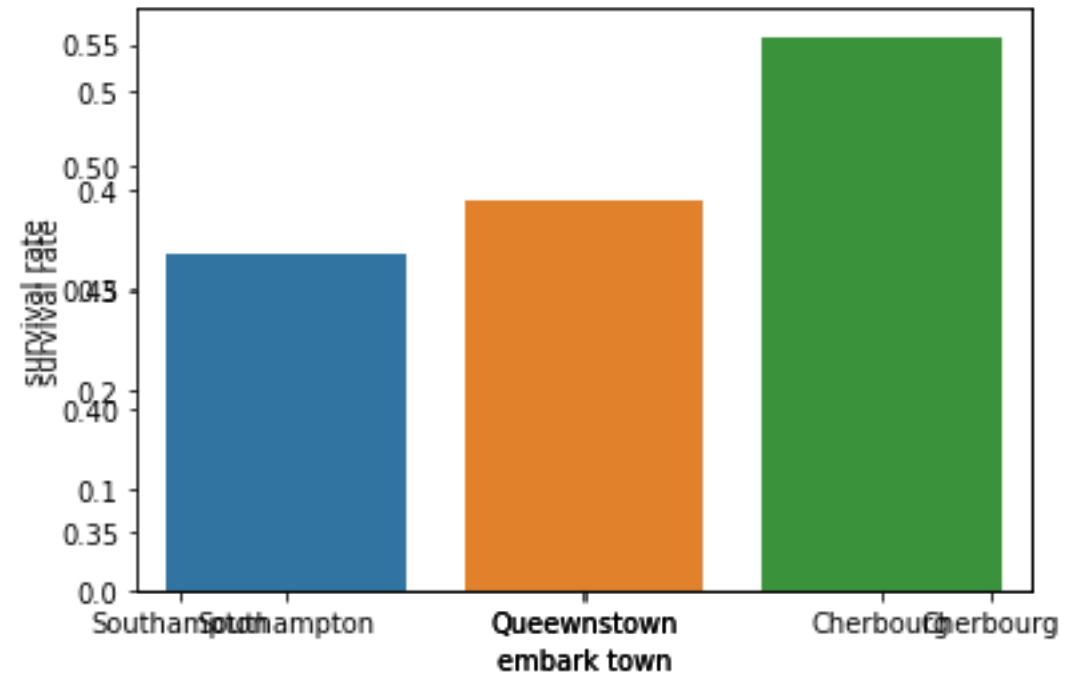
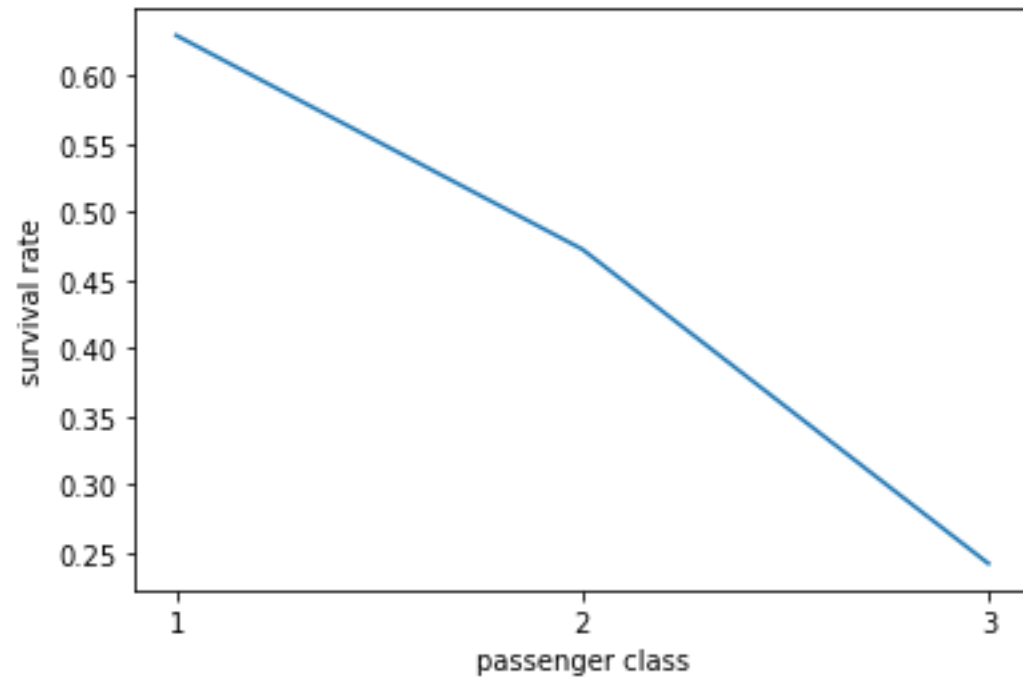


punti
direzione orizzontale
direzione verticale
colore
dimensione



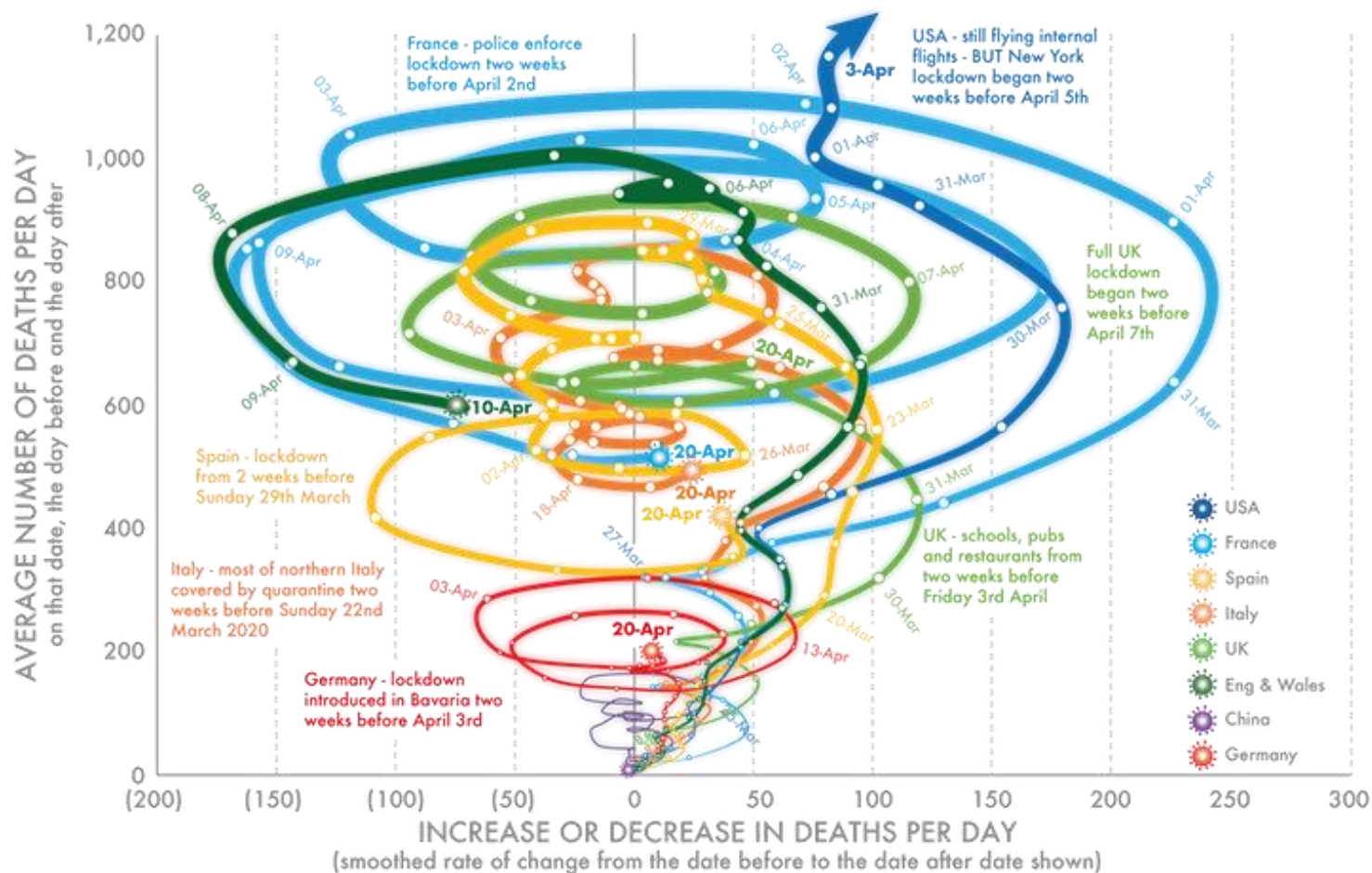
I principi della visualizzazione – Espressività

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.

<https://theconversation.com/three-charts-that-show-where-the-coronavirus-death-rate-is-heading-137103>

<https://www.oldstreetsolutions.com/good-and-bad-data-visualization>



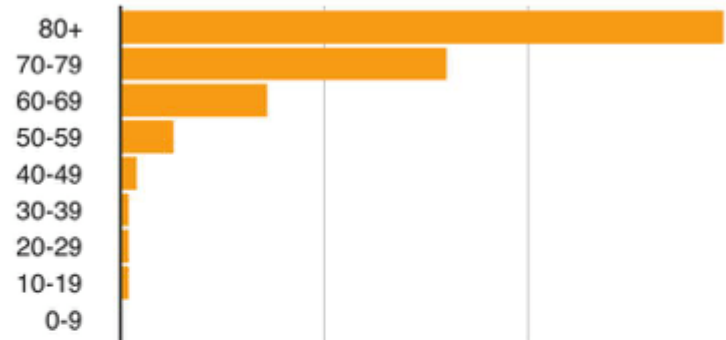
ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

I principi della visualizzazione – Efficacia

Death rate varies by age, health and sex

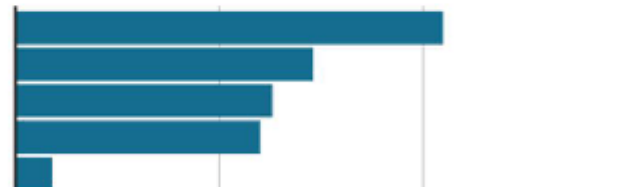
Case fatality ratio

Age



Health condition

Cardiovascular
Diabetes
Respiratory disease
Hypertension
None



Sex

Male
Female



Source: Chinese Centre for Disease Control and Prevention

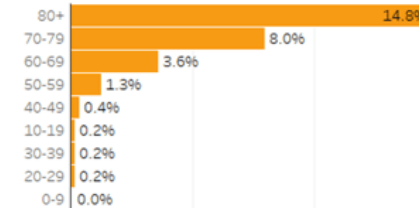
BBC

Osservazioni/commenti?

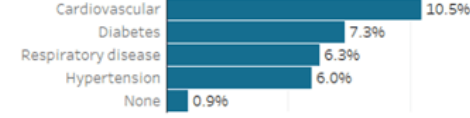
Death rate varies by age, health and sex

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)



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

I principi della visualizzazione – Come misurare l'efficacia?

Come misurare l'efficacia?

- **Accuratezza:** quanto i valori possono essere stimati accuratamente dal grafico;
- **Discriminabilità:** 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.

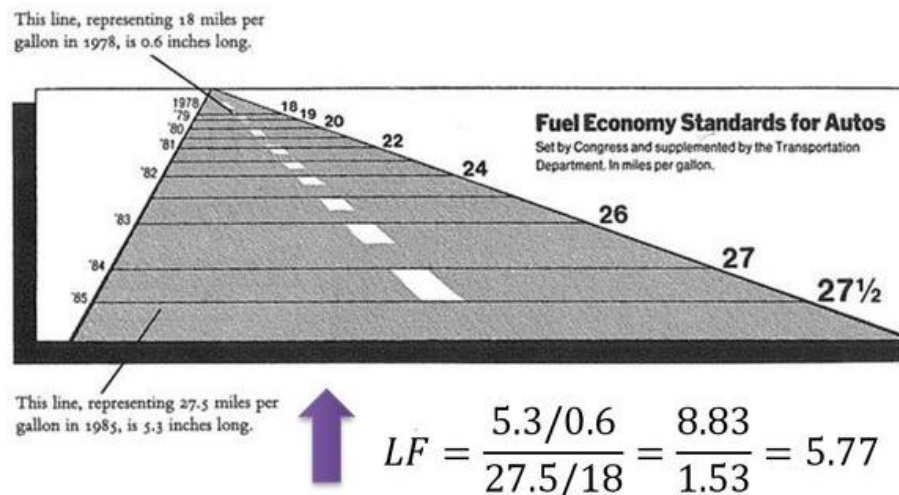
Quindi... Se questo numero è **minore di 0.95** o **maggiore di 1.05** allora è indice di una sostanziale distorsione nella visualizzazione.



Tufte's principles – Chart Lies

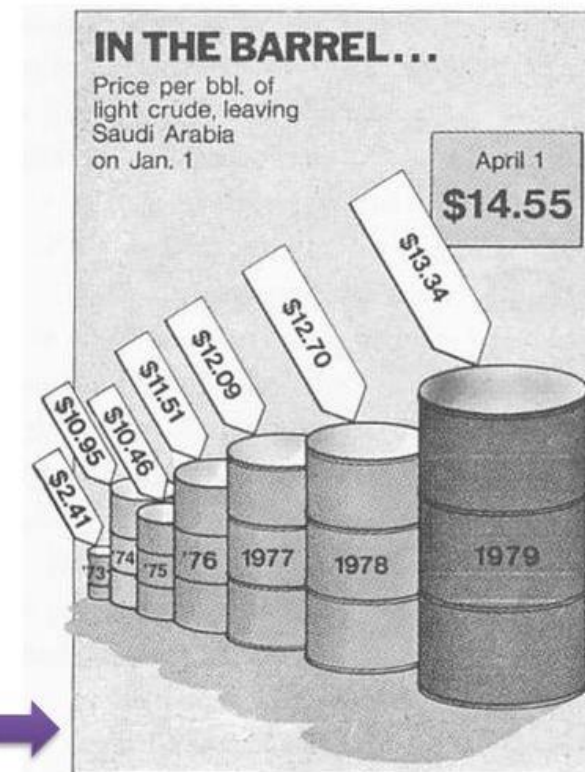
In queste visualizzazioni la rappresentazione dei valori, misurata sul grafico, **non** è direttamente proporzionale al valore numerico che questi rappresentano.

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



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

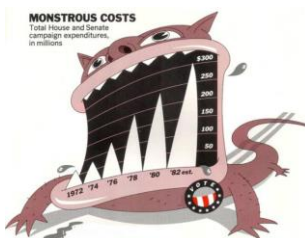
Con Chartjunk ci si riferisce a tutto ciò che è presente nel grafico ma che non rappresenta direttamente i dati.

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 distolga 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.

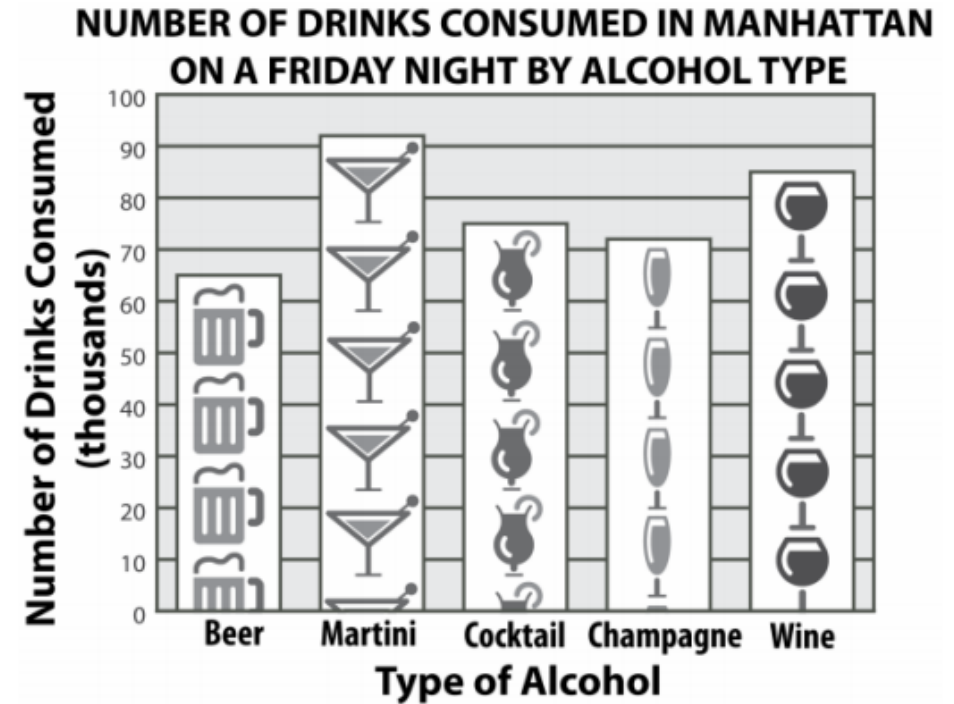
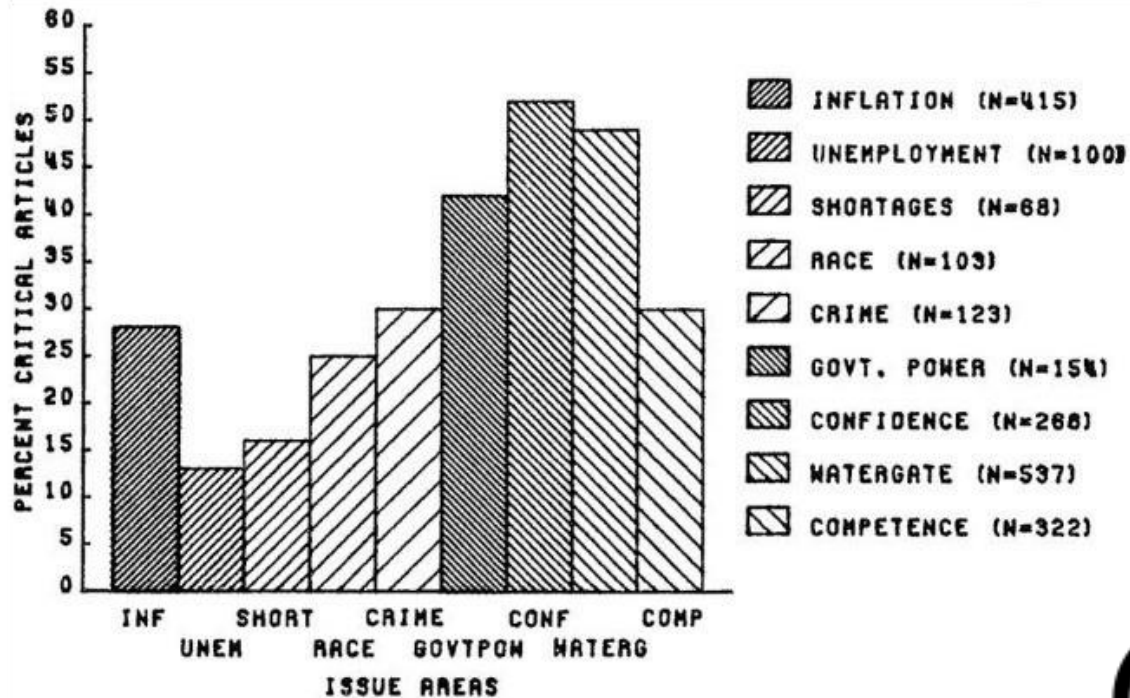


“The Chartjunk debate”

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



Tufte's principles – Chartjunk



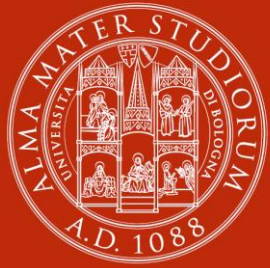
trade-off



«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



Data Visualization – Visualization Tools

Alessia Angeli

Studente di dottorato in Data Science and Computation

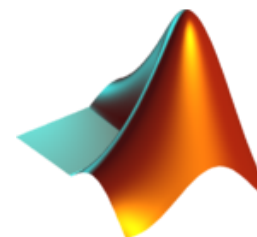
Dipartimento di Informatica – Scienza e Ingegneria

Visualization tools - overview

In queste lezioni utilizzeremo **Python** per costruire grafici.

Esistono però molti altri strumenti per poterlo fare:

- Excel;
- Google Fogli;
- R;
- Matlab;
- Tableau;
- ...



Visualization tools - overview

Le librerie di Python che utilizzeremo principalmente saranno:

- **Pandas** -> costruzione e gestione dataset (dataframe)
- **Matplotlib** -> visualizzazione
- **Seaborn** -> visualizzazione
- **Plotly** -> visualizzazione



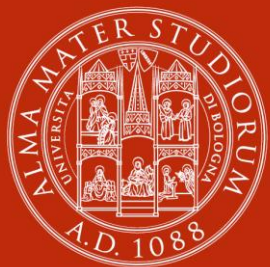
Hands-On

In queste lezioni utilizzeremo [Python](#) come linguaggio di programmazione.

Le esercitazioni verranno fatte utilizzando [Google Colab](#) per aprire notebook di [Jupyter](#).

- [Notebook 1 – Introduzione a Pandas](#)
- [Notebook 2 – Introduzione a Matplotlib](#)
- [Notebook 3 – Introduzione a Seaborn](#)
- [Notebook 4 – Introduzione a Plotly](#)





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

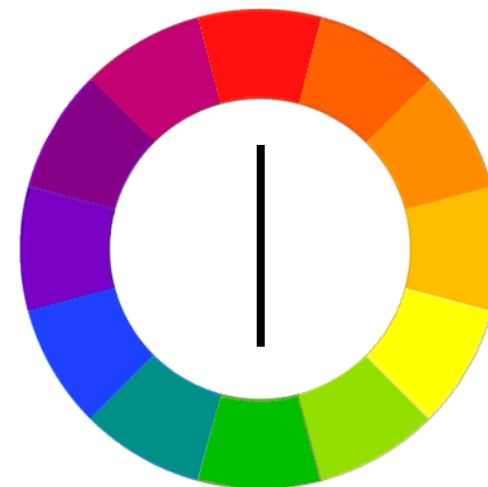
- Per **etichettare** (colore come nome);
- Per **misurare** (colore come quantità);
- Per **rappresentare** (colore come simbolo);
- Per **decorare** (colore come bellezza).



I COLORI – Colori complementari e colori analoghi

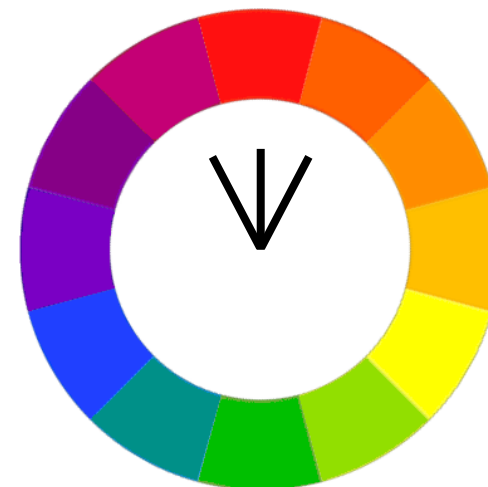
Colori complementari:

- 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.

- **Lightness** (luminosità):

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

- **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.



Hue



Lightness



Saturation

I COLORI – Possiamo ordinare i colori?

- **Hue** – NO

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

- **Lightness** – SI'

- **Saturation** – SI'



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

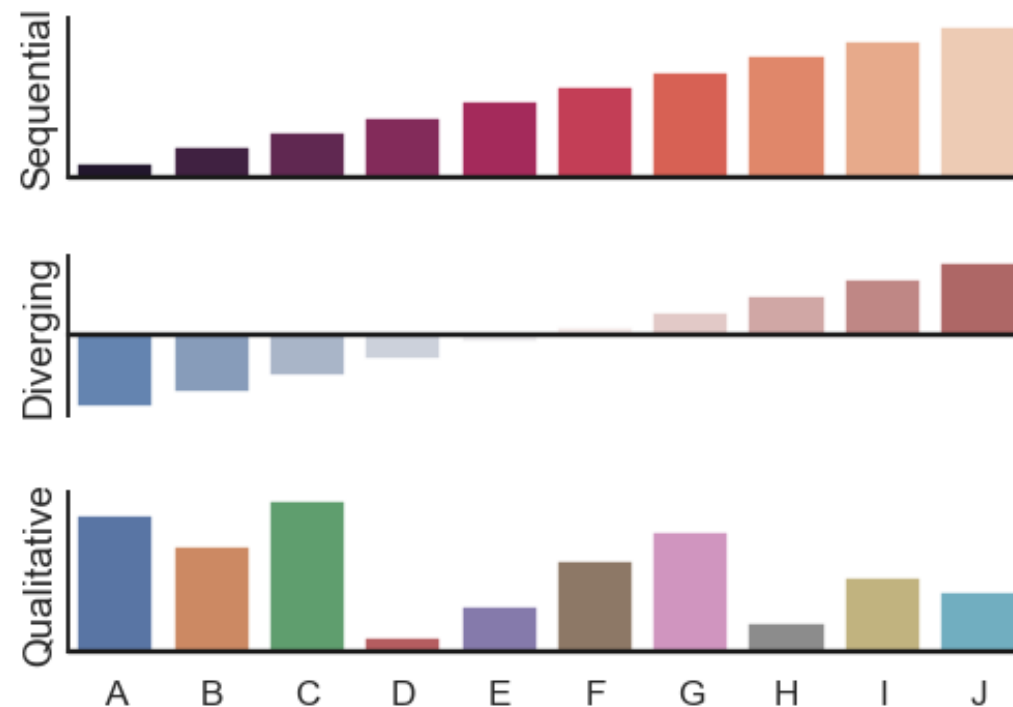


NOTA: per maggiore chiarezza nelle prossime slides si farà riferimento solo alle color palettes, però tutto quello visto si può essere riportato anche per le color maps.

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.

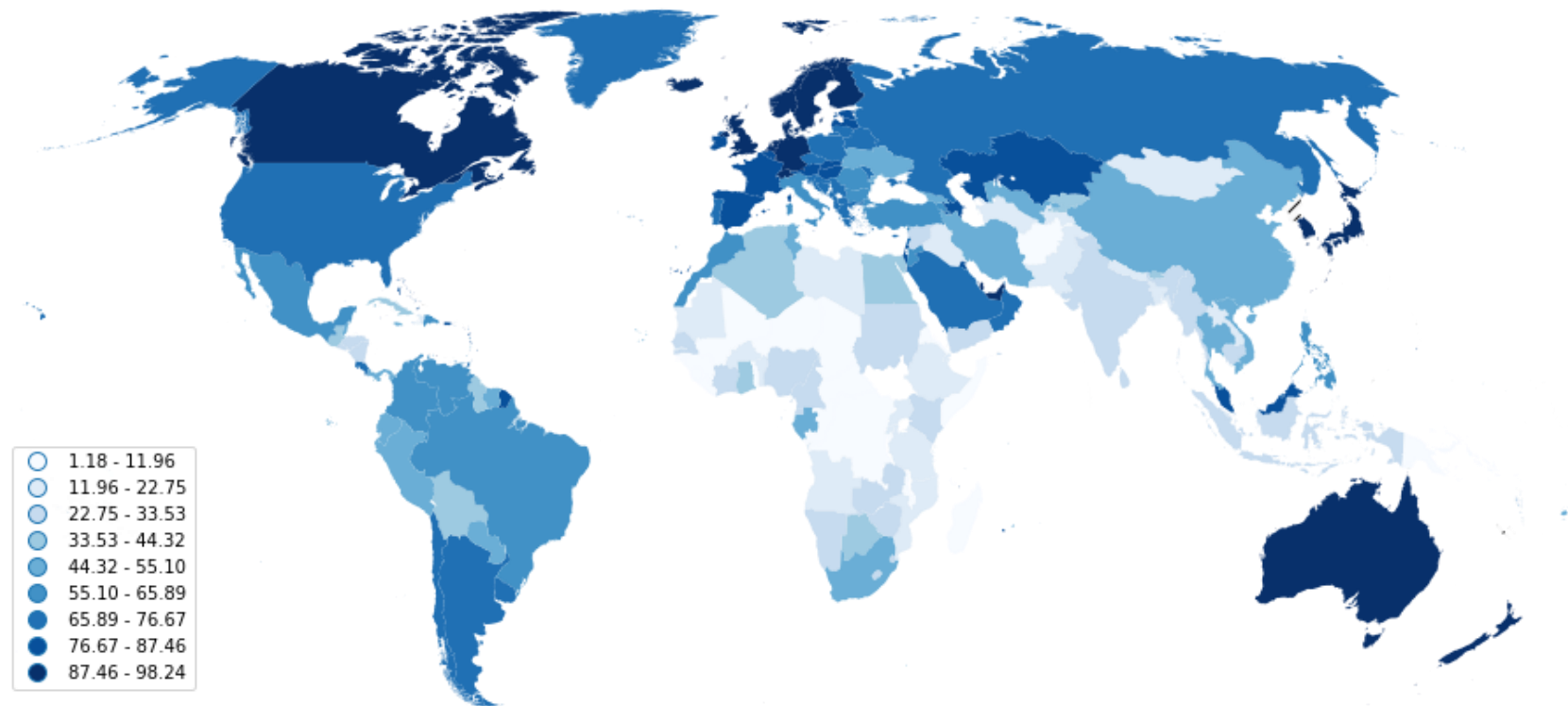


I COLORI – Sequential color palette

Individuals using the Internet (% of population) in 2016

Data types:

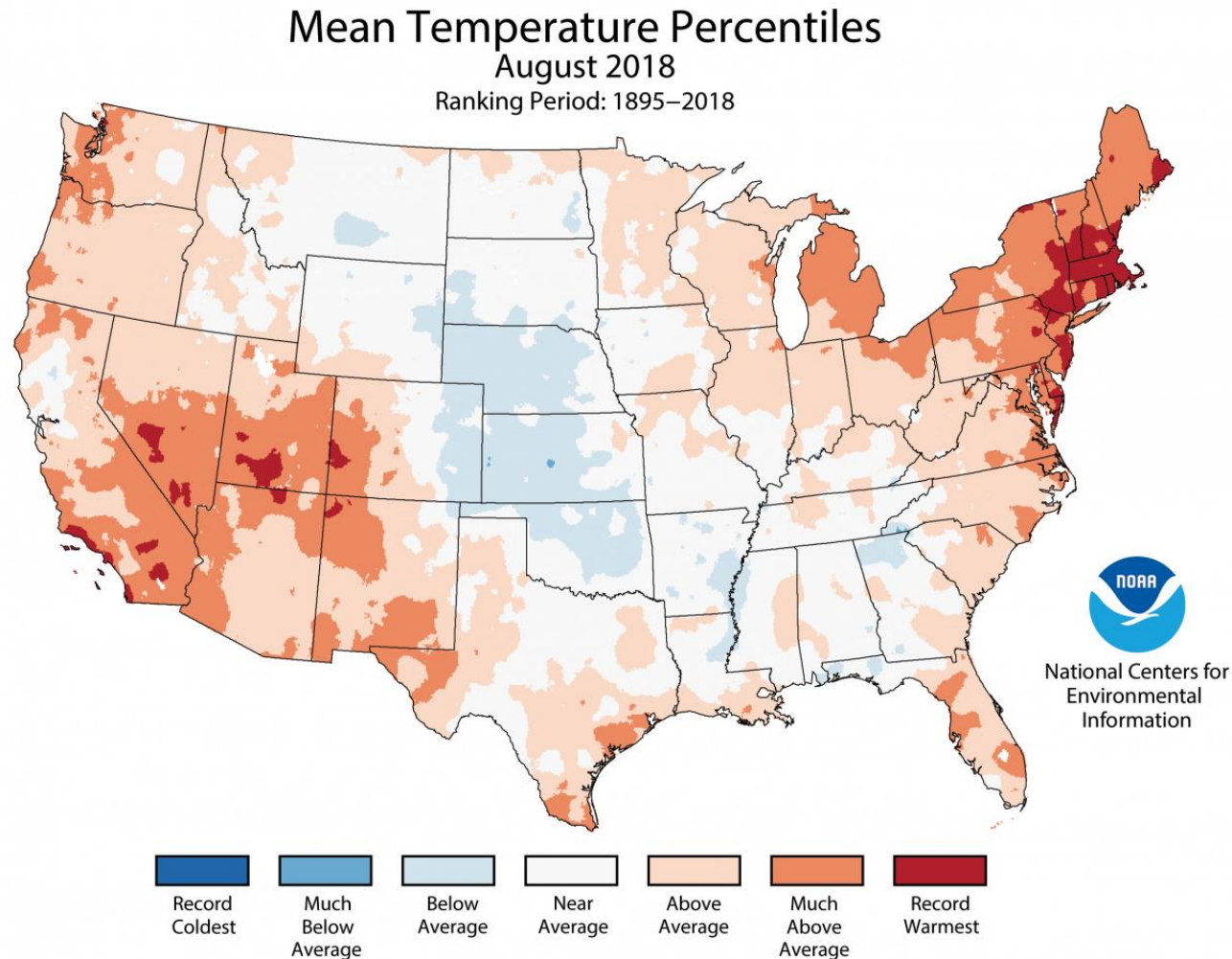
- **Ordinal** data
- **Interval** data
- **Ratio** data



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



I COLORI – Diverging color palette



Created: Wed Sep 05 2018

Data Source: 5km Gridded Dataset (nClimGrid)

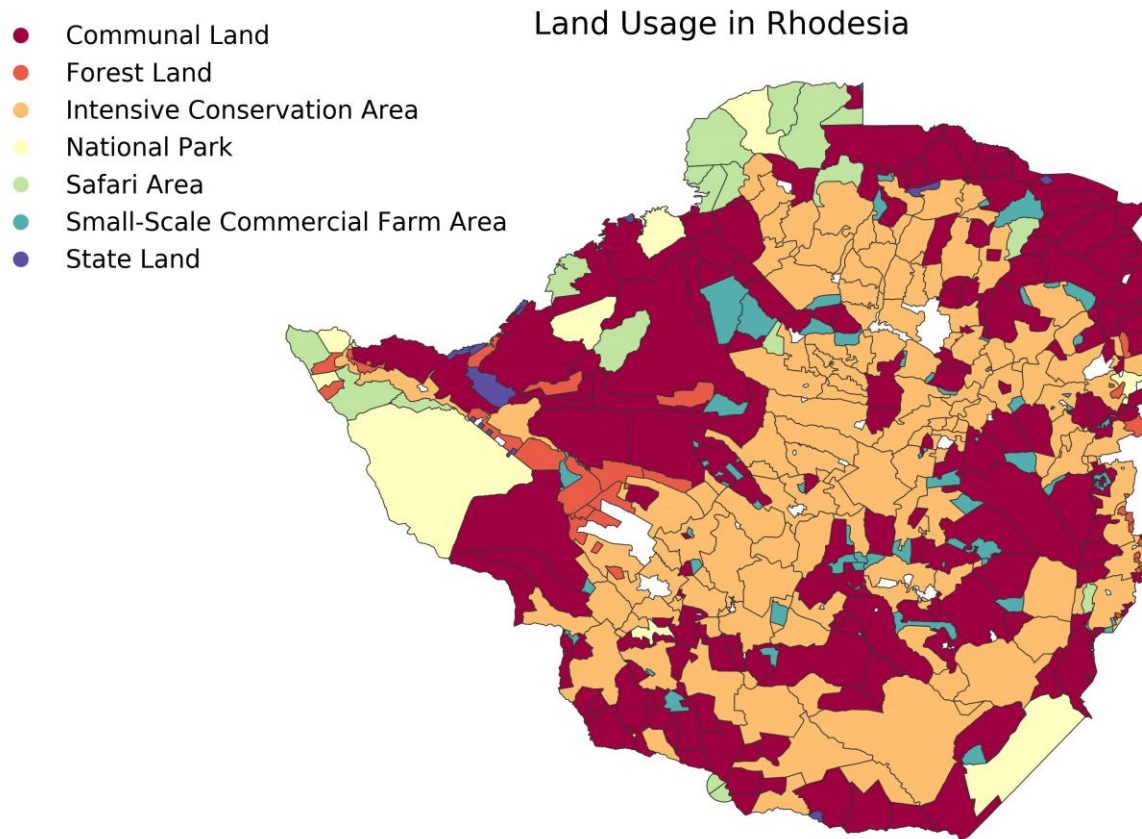
Data types:

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



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

I COLORI – Qualitative/categorical color palette



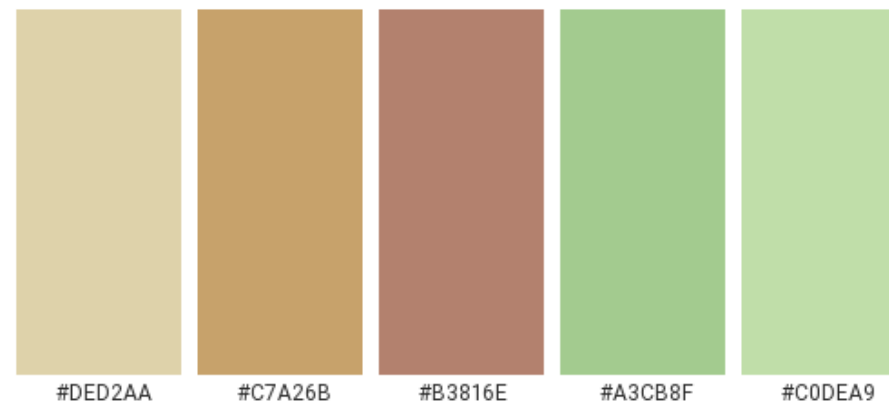
Data types:
Categorical data.

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

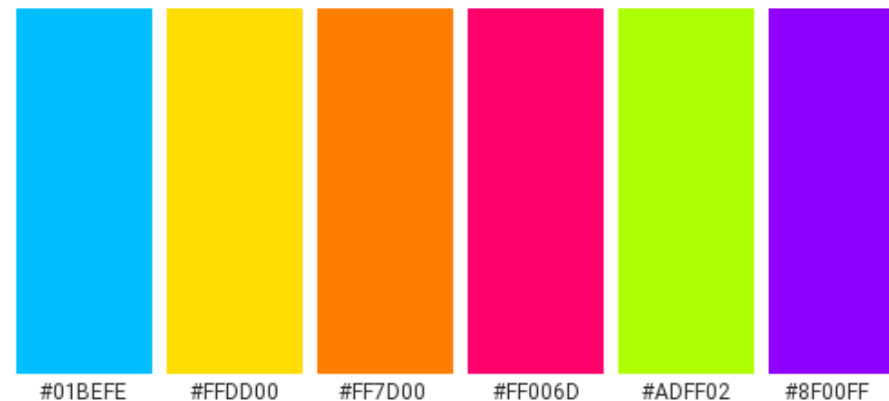
I COLORI – I colori trasmettono emozioni

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

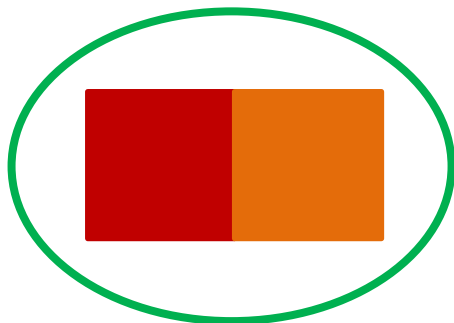
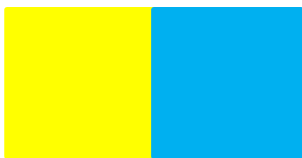
I COLORI – Controlla gli accostamenti in bianco e nero!

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

Abbinamenti di colori contrastanti potrebbero stancare gli occhi degli utenti.

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

Quindi... Tra questi abbinamenti quale scegliereste?



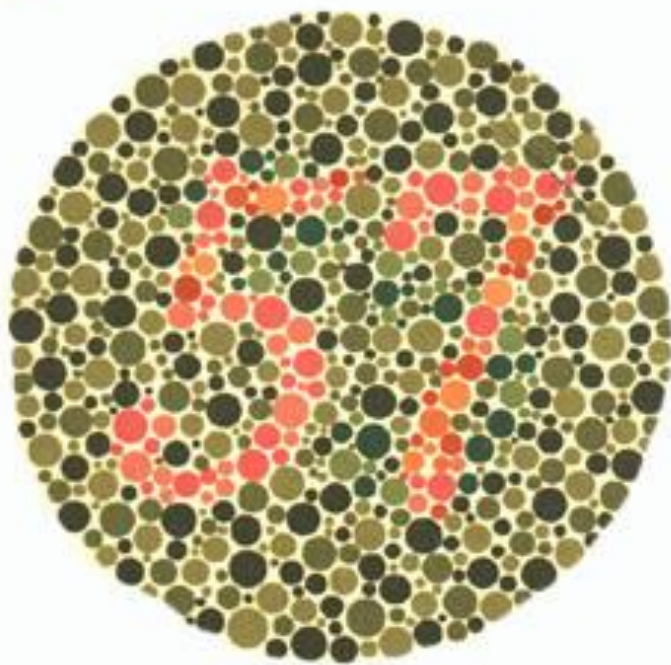
Hue as Value
with Grayscale
for comparison



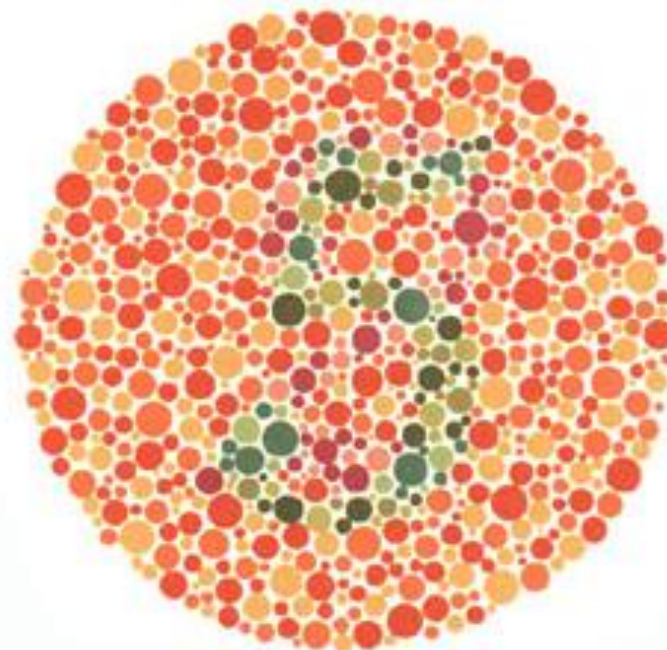
I COLORI – Il daltonismo

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 (*deuteranopia*).

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



5



6



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Hands-On

- [Notebook 5 – Introduzione ai Colori](#)

NOTA: per chi fosse interessato esistono diversi siti internet per costruire color palettes, che si possono successivamente esportate ed utilizzate in Python, come:



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



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



Riferimenti

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

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