

Università degli Studi di Padova

DIPARTIMENTO DI MATEMATICA “TULLIO LEVI-CIVITA”

CORSO DI LAUREA IN INFORMATICA



## Data visualization e infografiche: studio e applicazione attraverso la libreria D3.js

*Tesi di laurea triennale*

*Relatore*

Prof. Lamberto Ballan

*Laureando*

Jessica Carretta

---

ANNO ACCADEMICO 2023-2024



# Sommario

Il presente documento esibisce un resoconto dell'attività di stage condotta dalla laureanda Jessica Carretta presso l'azienda Zucchetti S.p.A. (sede di Padova) per una durata di circa trecentosei ore.

Esso si propone di esaminare il campo della [Data Visualization](#) e delle infografiche, riportandone anche un'applicazione pratica effettuata tramite la libreria [D3.js](#).

Nello specifico, il documento è strutturato come segue.

La parte introduttiva delinea le aspettative iniziali e i piani per lo stage, inquadrando il tutto nel contesto aziendale.

Successivamente, vengono esplorati i fondamenti della [Data Visualization](#) e analizzate le caratteristiche delle infografiche. Inoltre, una particolare attenzione viene rivolta anche all'individuazione di meccanismi che ottimizzino la scelta di tali raffigurazioni. In seguito, si approfondisce, attraverso un caso pratico, la creazione di infografiche interattive realizzate mediante [D3.js](#).

Infine, il documento si conclude con una riflessione sull'esperienza effettuata, inclusi i risultati ottenuti e le considerazioni personali sul percorso compiuto.

Per quanto riguarda il documento in sé, si riportano di seguito le convenzioni tipografiche adottate:

- gli acronimi, le abbreviazioni e i termini ambigui o di uso non comune menzionati vengono definiti nel glossario, situato alla fine del presente documento;
- i termini in lingua straniera o facenti parti del gergo tecnico sono evidenziati con il carattere *corsivo*.

*“I was born to be anything I wanted to be”*

— Palaye Royale, “Anxiety”

# Ringraziamenti

*In primo luogo, vorrei ringraziare il Prof. Lamberto Ballan, relatore della mia tesi, per l'aiuto fornitomi durante la stesura del lavoro.*

*In secondo luogo, desidero ringraziare i miei genitori per avermi dato l'opportunità di raggiungere questo traguardo.*

*Infine, un grazie speciale va anche ai miei amici per le risate e i momenti condivisi insieme durante questi anni e al mio ragazzo, Filippo, per avermi supportato e sopportato anche nei momenti più difficili.*

*Longare, Settembre 2024*

Jessica Carretta

# Indice

<b>1</b>	<b>Introduzione</b>	<b>1</b>
1.1	Il contesto di riferimento . . . . .	1
1.1.1	L'azienda ospitante . . . . .	1
1.2	Aspettative iniziali . . . . .	2
1.2.1	La proposta di stage . . . . .	2
1.2.2	Pianificazione delle attività . . . . .	2
1.2.3	Obiettivi fissati . . . . .	4
1.2.4	Vincoli . . . . .	4
<b>2</b>	<b>Data visualization</b>	<b>5</b>
2.1	Definizione e applicazioni . . . . .	5
2.1.1	La necessità di rappresentare i dati . . . . .	5
2.1.2	Obiettivi della data visualization . . . . .	6
2.1.3	Applicazioni . . . . .	6
2.2	Principi e linee guida . . . . .	7
2.2.1	Creare visualizzazioni efficaci . . . . .	7
2.2.2	Linee guida . . . . .	9
2.3	Classificare i grafici . . . . .	10
2.3.1	Metodo di classificazione . . . . .	10
2.3.2	Chart-chooser . . . . .	12
<b>3</b>	<b>Infografiche</b>	<b>18</b>
3.1	Definizione e applicazioni . . . . .	18
3.1.1	Cos'è un infografica e in cosa si distingue dalla data visualization	18
3.1.2	Applicazioni e vantaggi . . . . .	19
3.2	Classificare le infografiche . . . . .	19
3.2.1	Metodi di classificazione . . . . .	19
3.2.2	Infographic-helper . . . . .	20
3.3	Design delle infografiche . . . . .	23
3.3.1	I fattori che influenzano il design . . . . .	23
3.3.2	Interattività . . . . .	32
3.3.3	Principi di Gestalt . . . . .	33
3.4	Come costruire un'infografica . . . . .	34
3.4.1	L'Infomodel . . . . .	34
3.4.2	Principi e linee guida . . . . .	37
<b>4</b>	<b>Applicazione pratica di infografica web con D3.js</b>	<b>40</b>
4.1	Progettazione dell'infografica . . . . .	40

4.1.1	Identificazione degli interlocutori . . . . .	40
4.1.2	Identificazione della storia e dei suoi effetti . . . . .	42
4.1.3	Identificazione del dataset . . . . .	44
4.1.4	Analisi dei dati e rappresentazione della storia . . . . .	44
4.2	Implementazione e uso dell'infografica . . . . .	48
4.2.1	Strumenti e tecnologie utilizzati . . . . .	48
4.2.2	Implementazione . . . . .	51
4.3	Validazione . . . . .	68
<b>5</b>	<b>Conclusioni</b> . . . . .	<b>70</b>
5.1	Rendiconto dei risultati . . . . .	70
5.1.1	Vincoli e obiettivi generali . . . . .	70
5.1.2	Attività svolte e obiettivi specifici . . . . .	70
5.2	Considerazioni finali . . . . .	71
<b>A</b>	<b>Appendice A</b> . . . . .	<b>73</b>
A.1	Grafici risultanti da Chart-chooser . . . . .	73
	<b>Acronimi e abbreviazioni</b> . . . . .	<b>80</b>
	<b>Glossario</b> . . . . .	<b>81</b>
	<b>Bibliografia</b> . . . . .	<b>85</b>

# Elenco delle figure

1.1	Logo Zucchetti S.p.A. . . . .	1
2.1	Diagramma di Venn: intersezione tra data science, comunicazione e design nella visualizzazione dei dati . . . . .	7
2.2	Gerarchia DIKV (Data-Information-Knowledge-Wisdom) . . . . .	8
2.3	Sistema gerarchico di priorità per le regole di Chart-chooser - priorità 20 e 15 . . . . .	15
2.4	Sistema gerarchico di priorità per le regole di Chart-chooser - priorità dalla 10 all'1 . . . . .	16
3.1	Gerarchia dei fattori che influenzano il design delle infografiche . . . . .	24
4.1	Esempio di cambio sfondo tra sezioni . . . . .	53
4.2	Esempio di tooltip su testo . . . . .	54
4.3	Esempio di tooltip su un grafico . . . . .	54
4.4	Esempio di contenitore “collassato” . . . . .	54
4.5	Pulsante per l'apertura del chatbot. . . . .	55
4.6	Sezione introduttiva dell'infografica . . . . .	55
4.7	Sezione dell'infografica sui requisiti del corso . . . . .	56
4.8	Sezione dell'infografica sul profilo degli studenti . . . . .	57
4.9	Diagramma di Sankey sul profilo degli studenti animato . . . . .	58
4.10	Sezione dell'infografica sui luoghi d'interesse . . . . .	59
4.11	Mappa dei luoghi dell'Università - visualizzazione individuale delle puntine e pop-up . . . . .	60
4.12	Sezione dell'infografica sulle materie di studio . . . . .	61
4.13	Sezione dell'infografica sul percorso universitario per durata e voto . . . . .	62
4.14	Sezione dell'infografica riguardante i percorsi intrapresi post-laurea . . . . .	63
4.15	Diagramma di Venn sulla proporzione di studenti e lavoratori dopo l'animazione . . . . .	64
4.16	Sezione dell'infografica sull'opinione finale dei laureati . . . . .	64
4.17	Sezione dell'infografica sulle fonti . . . . .	65
4.18	Esempio di finestra della chat . . . . .	65

# Elenco delle tabelle

1.1	Obiettivi da raggiungere . . . . .	4
3.1	Backbone shape dei pattern di Visual Information Flow (VIF) . . . . .	27
5.1	Stato di soddisfacimento degli obiettivi a fine stage . . . . .	71
A.1	Grafici risultanti da Chart-chooser . . . . .	79



# Capitolo 1

## Introduzione

*In questo capitolo verranno delineate le aspettative iniziali e i piani per lo stage, inquadrando il tutto nel contesto aziendale.*

### 1.1 Il contesto di riferimento

#### 1.1.1 L'azienda ospitante

Zucchetti S.p.A. nasce nel 1978 da un'intuizione di Domenico "Mino" Zucchetti, che per primo in Italia realizzò un *software* capace di realizzare la dichiarazione dei redditi in maniera automatizzata. Questo prodotto ebbe un grande successo e l'azienda iniziò quindi a produrre *software* per automatizzare altre procedure contabili aziendali.



**Figura 1.1:** Logo Zucchetti S.p.A.

A oggi, Zucchetti è una delle principali aziende in Italia nel settore dell'[Information and Communication Technology \(ICT\)](#) e un attore sempre più rilevante sul mercato internazionale. Il gruppo può infatti contare più di 9.000 dipendenti distribuiti in 15 paesi e oltre 700.000 clienti presenti in più di 50 nazioni.

Oltre al consolidamento della sua presenza sul territorio, Zucchetti ha anche ampliato la sua offerta di prodotti. Essa fornisce gestionali aziendali e soluzioni di [Enterprise Resource Planning \(ERP\)](#) per la gestione del personale, la comunicazione aziendali, la rilevazione delle presenze e degli accessi, etc. Zucchetti opera, inoltre, nel campo dell'[e-business](#) e della [business intelligence](#) e offre servizi per robotica, automazione, sicurezza informatica e [server farm](#).

#### Servizi offerti

La *mission* di Zucchetti è quella di supportare le aziende nel migliorare la loro competitività ed efficienza operativa attraverso soluzioni innovative e di qualità. A tal fine, Zucchetti S.p.A. fornisce ai suoi clienti:

- **Un supporto pre-vendita**, analizzando e studiando le soluzioni che meglio si adattano al problema;
- **Un supporto post-vendita**, installando il prodotto e offrendo supporto tecnico;
- **Una formazione del personale**, per consentire al cliente di sfruttare al meglio il prodotto acquistato;
- **Un aggiornamento costante**, per restare al passo con le nuove tecnologie e le nuove normative in materia fiscale, contabile e amministrativa.

### Clientela target

Zucchetti S.p.A. si rivolge a una vasta gamma di clientela, comprendente imprese di ogni dimensione e settore, ma anche professionisti e associazioni di categoria, oltre che CAF e Pubblica Amministrazione. La sua offerta diversificata di soluzioni *software* e servizi tecnologici è pensata infatti per soddisfare le esigenze specifiche di ciascun segmento di mercato.

## 1.2 Aspettative iniziali

### 1.2.1 La proposta di stage

Lo scopo del progetto di stage è quello di sviluppare un sistema di visualizzazione interattiva dei dati che fornisca agli utenti una comprensione non convenzionale, ma pur sempre intuitiva, delle analisi effettuate. Per realizzare tale visualizzazione viene proposto l'utilizzo della libreria [D3.js](#), nota per la sua flessibilità e capacità di creare visualizzazioni grafiche dinamiche e coinvolgenti.

A tal fine, è necessario studiare e analizzare in profondità le varie modalità di rappresentazione grafica dei dati, focalizzandosi sull'identificazione delle tecniche più idonee per rendere chiari e informativi i vari tipi di dati possibili. A partire da questo studio, si potrà passare alla creazione di un'infografica web, che combinerà grafici realizzati tramite [D3.js](#) con altri elementi aggiuntivi di chiarimento. Una volta completata l'infografica, si prevede di sviluppare un *template* riutilizzabile per future rappresentazioni simili e verificarne l'efficacia e l'utilità mediante adeguati metodi di valutazione.

Questo progetto di stage offre dunque l'opportunità di approfondire i seguenti campi:

- **Data Visualization** e **infografiche**;
- **Utilizzo della libreria [D3.js](#)**, attraverso la creazione di grafici dinamici e interattivi inseriti all'interno dell'infografica;
- **Sviluppo Web con [Hypertext Markup Language \(HTML\)](#), [Cascading Style Sheets \(CSS\)](#) e [JavaScript \(JS\)](#)**, attraverso la realizzazione di un'interfaccia web che implementi l'infografica.

### 1.2.2 Pianificazione delle attività

Si riporta di seguito la pianificazione delle attività di stage suddivise nelle otto settimane di lavoro previste.

- **Prima Settimana - Formazione**
  - Introduzione e approfondimento sui principi chiave e sulle metodologie della [Data Visualization](#);
  - Formazione sulle tecnologie da adottare, in particolare sulla libreria [D3.js](#).
- **Seconda Settimana - Selezione e sviluppo di grafici**
  - Identificazione dei criteri per la scelta del tipo di grafico più adatto, considerando la struttura dei dati e gli obiettivi della visualizzazione;
  - Sviluppo dei grafici individuati utilizzando la libreria [D3.js](#).
- **Terza Settimana - Integrazione di algoritmi a grafici**
  - Sviluppo di grafici avanzati, che integrano algoritmi, attraverso l'uso di [D3.js](#).
- **Quarta Settimana - Test e Documentazione**
  - Realizzazione di test che verifichino la bontà del codice prodotto nelle settimane precedenti;
  - Produzione di documentazione riguardante il prodotto realizzato.
- **Quinta Settimana - Approfondimento sulle infografiche**
  - Ricerca approfondita di informazioni e guide riguardanti il concetto di infografica;
  - Raccolta di elementi grafici e di design da accompagnare ai grafici per migliorarne la comprensibilità e l'attrattività visiva;
  - Realizzazione preliminare dell'infografica utilizzando [HTML](#), [CSS](#) e [JS](#).
- **Sesta Settimana - Completamento dell'infografica e creazione del template**
  - Ultimazione dell'infografica in [HTML](#), [CSS](#) e [JS](#);
  - Creazione di un *template* dell'infografica, riutilizzabile per future analisi simili.
- **Settima Settimana - Metodi di validazione dell'infografica**
  - Individuazione di criteri che permettano di valutare la qualità dell'infografica (e.g. *chatbot* con [Large Language Model \(LLM\)](#) che interagisce con l'utente sull'infografica ottenendo, indirettamente, il suo feedback per miglioramenti futuri);
  - Implementazione dei metodi di validazione individuati.
- **Ottava Settimana - Test e Documentazione**
  - Realizzazione di test che verifichino la bontà del codice prodotto nelle settimane precedenti;
  - Produzione di documentazione riguardante il prodotto realizzato.

### 1.2.3 Obiettivi fissati

Sono fissati i seguenti obiettivi da raggiungere attraverso le attività sopraelencate.

Identificativo	Tipo	Descrizione dell'obiettivo
O01	Obbligatorio	Comprendere i principi fondamentali della <a href="#">Data Visualization</a> .
O02	Obbligatorio	Selezionare il grafico più adatto per tipo di struttura dati e obiettivo della visualizzazione.
O03	Obbligatorio	Realizzare grafici informativi e interattivi utilizzando <a href="#">D3.js</a> .
O04	Obbligatorio	Realizzare un'infografica web funzionante in <a href="#">HTML</a> , <a href="#">CSS</a> e <a href="#">JS</a> che visualizzi i risultati degli algoritmi di Intelligenza Artificiale e <a href="#">Machine Learning</a> .
O05	Obbligatorio	Creare un <i>template</i> dell'infografica.
O06	Obbligatorio	Individuare dei metodi di validazione dell'infografica.
D01	Desiderabile	Sviluppare infografiche alternative.
D02	Desiderabile	Implementare un <i>chatbot</i> con <a href="#">LLM</a> che interagisca con l'utente sull'infografica.

**Tabella 1.1:** Obiettivi da raggiungere

### 1.2.4 Vincoli

Si riportano di seguito i vincoli imposti per lo svolgimento dell'attività di stage.

#### Vincoli temporali

La durata dello stage è fissata a otto settimane, per un totale complessivo compreso tra le 300 e le 320 ore in base alla necessità. Gli orari di lavoro sono dal lunedì al venerdì, dalle ore 9.00 alle ore 18:00, con pausa pranzo dalle 13:00 alle 14:00.

#### Vincoli tecnologici

La proposta di stage prevede l'utilizzo della libreria [D3.js](#) per lo sviluppo di grafici interattivi. Questi andranno poi a inserirsi in infografiche web da sviluppare attraverso [HTML](#), [CSS](#) e [JS](#). Non sono stati imposti limiti ulteriori su eventuali tecnologie aggiuntive.

## Capitolo 2

# Data visualization

*In questo capitolo verrà presentata una panoramica della [Data Visualization](#), spiegandone caratteristiche, usi e principi. Viene inoltre proposto un metodo di classificazione dei grafici per ottimizzare le visualizzazioni.*

### 2.1 Definizione e applicazioni

#### 2.1.1 La necessità di rappresentare i dati

La *visualizzazione dei dati* ha una tradizione millenaria. Sin dagli albori della civiltà umana, le persone hanno sviluppato metodi per rappresentare visivamente i dati al fine di renderli più facili da comprendere, memorizzare e trasmettere. Negli ultimi decenni, tuttavia, questa esigenza ha conosciuto un incremento senza precedenti.

Viviamo infatti nell'epoca dell'[information overload](#) e dei [big data](#), dove le informazioni arrivano con una velocità, un volume e una varietà così travolgenti che non siamo in grado di comprenderli se non attraverso un ulteriore strato di astrazione. Tale condizione si deve soprattutto allo sviluppo delle tecnologie digitali, grazie alle quali siamo arrivati a disporre di una quantità apparentemente infinita di informazioni, tra sapere sociale, motori di ricerca e volontà delle persone di esprimersi.

In questo contesto, la [Data Visualization](#) è diventata fondamentale per interpretare e trarre intelligenza da questa grande mole di dati. Nello specifico, essa consente di fissare l'attenzione in segni maneggevoli - le rappresentazioni grafiche - che consentono una lettura dell'informazione semplice e intuitiva. Tramite tali segni, è possibile cogliere funzioni strutturali e relazioni impreviste tra fenomeni, oltre che scoprire regolarità tra di essi che altrimenti sarebbero rimaste nascoste, facilitando così eventuali risoluzioni e valutazioni.

Possiamo dunque definire la [Data Visualization](#) come l'insieme dei meccanismi di traduzione da dati a rappresentazione grafica mediante i quali l'informazione viene resa più chiara ed efficace. Più precisamente, potremmo dire che è un processo atto ad amplificare la cognizione dei dati attraverso una traslazione di questi in un contesto visivo.

### 2.1.2 Obiettivi della data visualization

Gli obiettivi alla base della [Data Visualization](#) sono:

- **Identificare**, ovvero trovare un *dataset* significativo, né troppo banale né troppo complesso, da cui sia possibile ricavare delle informazioni utili;
- **Manipolare**, ossia analizzare e combinare i dati in modo da chiarirne il significato;
- **Formattare**, ovvero standardizzare il modo con cui si accede ai dati per una consumazione più efficiente;
- **Presentare**, ossia rappresentare i dati formattati in modo da esprimerne il significato nascosto.

Si desidera dunque massimizzare l'efficacia ed efficienza della visualizzazione. Ciò comporta:

- Per quanto riguarda l'efficienza: ridurre la complessità e il rumore nella visualizzazione, eliminando così tutti gli elementi superflui che potrebbero portare addirittura a correlazioni incorrette.
- Per quanto riguarda l'efficacia: fornire informazioni comprensibili e utili in modo che sia possibile prendere decisioni informate in base a esse.

Il fine ultimo di questi obiettivi è quello di:

- **Rendere i dati comprensibili e memorabili**,
- Facilitare nuove scoperte e **individuare *trend* e valori anomali**,
- **Visualizzare** velocemente **relazioni e regolarità** nei dati,
- **Rendere più consapevole la presa di decisioni** e stimolare la formulazione di ulteriori domande più specifiche.

### 2.1.3 Applicazioni

Nell'età contemporanea, il processo di visualizzazione dei dati è ormai diventato essenziale per la gestione quotidiana di qualsiasi impresa, governo e organo informativo.

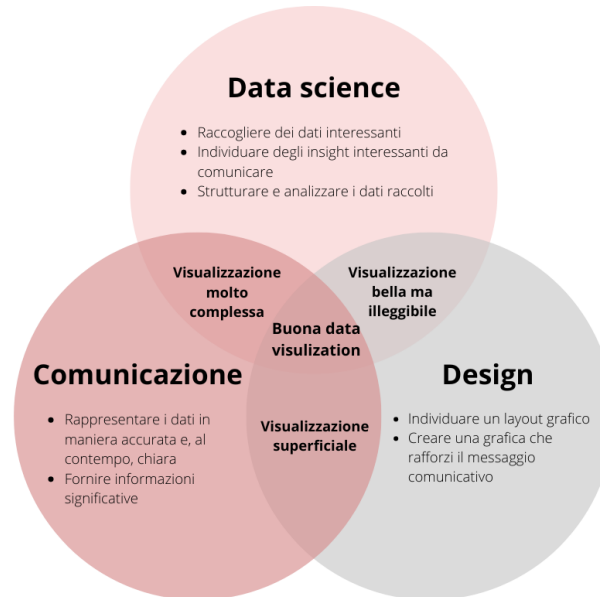
Imprese e pubbliche amministrazioni sfruttano la [Data Visualization](#) per comprendere meglio la propria organizzazione e prendere decisioni più informate basate sui dati, ad esempio per quanto riguarda la gestione delle risorse e lo sviluppo di strategie future. Questo approccio, inoltre, è cruciale anche per la percezione esterna dell'organizzazione e, dunque, per contraddistinguerla da enti a essa analoghi.

Alla stessa maniera, anche per gli organi informativi (e.g. i giornali) la [Data Visualization](#) è diventata indispensabile. Infatti, solo grazie a essa è possibile ricavare e comprendere a fondo determinate informazioni, oltre che renderle fruibili in maniera chiara e intuitiva. Ciò risulta particolarmente importante per tali organi in quanto la loro *mission* è proprio diffondere e comunicare informazioni.

## 2.2 Principi e linee guida

### 2.2.1 Creare visualizzazioni efficaci

Una buona visualizzazione nasce dall'incontro di comunicazione, [Data Science](#) e design.

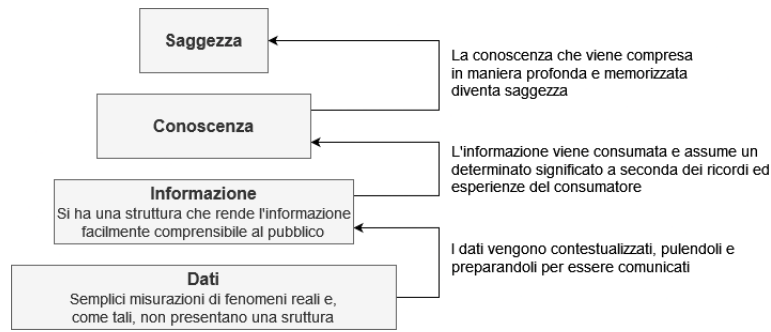


**Figura 2.1:** Diagramma di Venn: intersezione tra data science, comunicazione e design nella visualizzazione dei dati

Se realizzata correttamente, infatti, essa è capace di ricavare da *dataset* complessi delle informazioni significative e trasmetterle in maniera intuitiva. Usando le parole di Edward R. Tufte, statistico e luminare nel campo della [Data Visualization](#), una *visualizzazione dei dati* eccellente consiste in “idee complesse comunicate con chiarezza, precisione ed efficienza”<sup>1</sup>.

Da ciò ne consegue che l’obiettivo della [Data Visualization](#) è quello di rendere il più efficiente possibile il processo che il cervello umano compie nell’acquisire conoscenza e *saggezza* a partire dall’osservazione di fenomeni. Tale processo può essere descritto, in maniera semplificata, attraverso la cosiddetta gerarchia DIKV (*Data-Information-Knowledge-Wisdom*) di seguito riportata.

<sup>1</sup>Edward R. Tufte. *The Visual Display of Quantitative Information, Second Edition*. Graphics Press, LLC, 2006.



**Figura 2.2:** Gerarchia DIKV (Data-Information-Knowledge-Wisdom)

Come si nota subito dall'immagine 2.2, è essenziale ottenere dati puliti, completi e rappresentativi (responsabilità della [Data Science](#)). Solo successivamente sarà possibile rappresentare tali dati e ottenere una conoscenza valida (responsabilità di comunicazione e design).

È importante notare che nella *visualizzazione dei dati*, la grafica deve aspirare a oggettività, precisione e, soprattutto, funzionalità. Infatti, sebbene il processo di creazione possa somigliare a quello di un'applicazione artistica, essa non è e non deve essere solo un'espressione estetica. La grafica deve, in realtà, coadiuvare il fruitore in attività intellettive. Per descrivere tale principio, possiamo dunque prendere in prestito le parole di Louis Sullivan sull'architettura modernista del XX secolo: "La forma segue la funzione". Tale massima può infatti essere applicata anche all'architettura dell'informazione nel contesto della [Data Visualization](#), giacché un complesso di dati può assumere più forme, ma non tutte le forme sono sempre adatte, usabili e funzionali allo scopo della visualizzazione. Infatti, la loro scelta deve dipendere dal tipo di dati e obiettivo della rappresentazione.

Il processo di decisione della rappresentazione grafica viene meglio approfondito nella sezione 2.3.

Una volta decisa la forma migliore, bisogna comunque adattarla alla visualizzazione del caso specifico attraverso delle scelte di design. In tal senso, si può adottare un approccio più minimalistico, promosso dal sopracitato Edward Tufte, o un approccio più "artistico".

Per Tufte è essenziale combinare semplicità di design a complessità dei dati. Per lui, il fruitore della visualizzazione deve poter ricavare il maggior numero di idee e informazioni dal minor uso possibile di spazio e inchiostro. A tal proposito, egli sviluppa le seguenti formule:

$$\text{Rapporto dati-inchiostro} = \frac{\text{inchiostro usato per la codifica dei dati}}{\text{inchiostro totale usato per la stampa del grafico}}$$

$$\text{Densità dei dati} = \frac{\text{numero di voci in una matrice di dati}}{\text{area occupata dal grafico}}$$

evidenziando che quanto più elevati sono tali rapporti, migliore sarà la rappresentazione, a condizione però di non sovrappopolare lo spazio. Ciò implica che sia necessario evitare di utilizzare tutti gli elementi decorativi o, più in generale, tutte le parti che non comunicano dati, ovvero tutto ciò che Tufte chiama [chartjunk](#) (e.g. la maggior parte delle griglie e sfondi posti sotto i grafici).



D'altra parte, si potrebbe pensare che finché la funzione principale rimane comunicare dati ci si può divertire con la forma con cui essi vengono presentati. Infatti, ciò che Tufte potrebbe considerare *chartjunk*, e dunque superfluo, potrebbe invece rivelarsi utile al processo di memorizzazione e dunque portare più facilmente alla *saggezza*.

Qualsiasi approccio si decida di intraprendere, la rappresentazione dovrebbe sempre permettere di suscitare riflessioni non superficiali ed evidenziare tendenze, *pattern* o fenomeni altrimenti invisibili. Tale profondità si concretizza quasi sempre tramite la rappresentazione di dati multivariati, che consentono l'analisi di un fenomeno in maniera più accurata. Solo in un secondo momento ci si può dedicare all'estetica della presentazione, utilizzando gli spazi rimanenti. Tale precauzione è da prendere affinché i dati ricevano subito tutto lo spazio necessario ed eventuali effetti speciali o decorazioni siano inseriti solo se ne presenta l'opportunità.

### 2.2.2 Linee guida

Nel visualizzare dati è necessario garantire che le informazioni siano rappresentate accuratamente e senza distorsioni. A tal scopo, è dunque necessario perseguire un'**integrità grafica**. Ciò, secondo il sopracitato Tufte, implica l'uso di:

- proporzioni corrette tra la rappresentazione dei numeri sul grafico e le quantità numeriche fornite, come pure tra il numero di dimensioni rappresentate graficamente e il numero delle variabili nei dati;
- scale appropriate che mostrino la vera variazione dei dati (e.g. scale con intervalli regolari);
- etichette dettagliate e chiare che non fuorviino l'utente;
- elementi che contestualizzino il grafico.

Come accennato precedentemente, altrettanto importante è la **scelta del design**. Ciò include:

- La scelta di un formato e *layout* appropriato, scegliendo una combinazione di elementi che meglio mostri la riflessione che si vuole comunicare. Ciò comporta:
  - per l'occidente, un orientamento (specie per quanto riguarda il testo) che va dall'alto a sinistra verso il basso a destra;
  - l'uso di grafici corredati da piccoli messaggi di chiarimento, piuttosto che spiegazioni sparse.
- L'uso coerente e complementare di parole, numeri e figure.
- Visualizzare i dati in maniera accessibile, anche e soprattutto per quanto riguarda dettagli di dati complessi. Ciò comporta:
  - l'uso di codifiche non troppo elaborate;
  - tenere in considerazione anche le persone daltoniche e *color-deficient* nella scelta dei colori del grafico;
  - l'uso di caratteri leggibili.

## 2.3 Classificare i grafici

Lo sviluppo e disponibilità al pubblico di nuovi strumenti avanzati ha consentito anche a utenti non esperti di visualizzare dati ed estrarne informazioni dettagliate, permettendo loro di creare grafici e diagrammi. Tuttavia, persiste ancora un ampio divario di conoscenza tra questi utenti e i modelli visivi esistenti. Ne consegue che spesso le forme visive scelte per la rappresentazione non sono le più efficaci essendo limitate da un ristretto numero di opzioni conosciute. In alcuni casi, addirittura, la rappresentazione scelta risulta essere anche poco comprensibile o fuorviante per il caso d'uso specifico.

Senza una classificazione chiara, diventa dunque difficile per questo tipo di utenti selezionare la tecnica di visualizzazione più adeguata. Nelle seguenti sezioni, si esaminano dunque i criteri di classificazione identificati, assieme a uno strumento prototipale che mira ad automatizzare questo processo di classificazione.

### 2.3.1 Metodo di classificazione

I criteri principali da tenere in considerazione nella scelta del grafico sono:

- l'obiettivo della visualizzazione,
- i tipi di dati disponibili.

#### Obiettivi della visualizzazione

Quando parliamo di “obiettivo della visualizzazione” intendiamo il tipo di relazione tra i dati che vogliamo mettere in evidenza attraverso la rappresentazione grafica. I principali obiettivi individuati sono:

1. **Divergenza** (*deviation*), quando si vogliono mettere in luce le differenze di valori a partire da un punto fisso (solitamente lo 0, ma può essere ad esempio anche una media).
  - Esempio: confrontare la produttività del turno di giorno con quello del turno di notte.
2. **Correlazione** (*correlation*), quando si vuole enfatizzare la relazione tra due o più variabili. In questi casi, il fruitore della visualizzazione assumerà che le variabili coinvolte siano in relazione causale, pertanto se così non fosse è necessario specificarlo.
  - Esempio: esplorare come la qualità del sonno influenzi le prestazioni accademiche.
3. **Ranking**, quando si vuole mostrare la posizione degli elementi rispetto agli altri e il loro valore effettivo passa in secondo piano.
  - Esempio: mostrare i dieci paesi con il tasso di inquinamento più alto.
4. **Distribuzione** (*distribution*), quando si vogliono mostrare eventuali *pattern* nei dati, evidenziando valori ed eventi di un *dataset*, e per quali casi essi si verificano.
  - Esempio: mostrare la frequenza di incidenti stradali per tipo di veicolo.

5. **Cambiamento nel tempo** (*time*), quando si vogliono enfatizzare dei *trend* che variano nel tempo.
  - Esempio: analizzare come l'uso dei social media è cambiato negli ultimi cinque anni.
6. **Composizione** (*composition*), quando si vuole mostrare come una singola entità viene suddivisa nelle sue componenti.
  - Esempio: mostrare la composizione delle emissioni di gas serra per settore industriale.
7. **Grandezze** (*magnitude*), quando si vogliono comparare grandezze, che siano esse assolute (per un maggior livello di dettaglio) o relative.
  - Esempio: mostrare la dimensione del mercato degli smartphone in termini di volumi di vendita.
8. **Spazi** (*spatial*), quando si vogliono mostrare *pattern* geografici o posizioni precise.
  - Esempio: mostrare i paesi con la maggiore concentrazione di risorse naturali.
9. **Flussi** (*flow*), quando si vogliono mostrare volumi o intensità del movimento/-cambiamento tra due o più stati o condizioni.
  - Esempio: esplorare la rete di interconnessioni finanziarie tra le banche.

### Tipi di dati

Una prima distinzione nel tipo di dati riguarda:

- **dati numerici**,
- **dati categorici**.

Il *dataset* fornito per la visualizzazione potrà dunque essere composto da dati tutti numerici, tutti categorici oppure da un mix dei due.

Un'altra distinzione si ha tra:

- **dati univariati**, quando ho un *dataset* molto piccolo composto da una sola variabile;
- **dati multivariati**, quando ho un *dataset* composto da più variabili.

In base alle suddette classificazioni, si può specificare ulteriormente il tipo di dato:

- In caso di *dataset* con multiple variabili numeriche:
  - **dati ordinati**, dove è presente almeno una variabile numerica che è ordinata;
  - **dati non ordinati**, dove l'ordine non è rilevante.
- In caso di *dataset* con multiple variabili categoriche:
  - **organizzazione in sottogruppi**, una o più variabili categoriche specificano e rappresentano una sottopopolazione di un'altra variabile categorica;

- **organizzazione indipendente**, le variabili categoriche sono indipendenti tra di loro;
- **organizzazione gerarchica**, una o più variabili categoriche presentano un rapporto gerarchico rispetto ad un'altra variabile categorica padre.
- In caso di almeno una variabile numerica e una categorica, con multiple variabili categoriche solo se organizzate in sottogruppi o in relazione gerarchica:
  - **dati con singole occorrenze**, quando a ogni individuo della popolazione, individuato dalle variabili categoriche, corrisponde un solo dato numerico;
  - **dati con multiple occorrenze**, quando a ogni individuo della popolazione, individuato dalle variabili categoriche, corrispondono multipli valori numerici.

### 2.3.2 Chart-chooser

*Chart-chooser* è uno strumento prototipale sviluppato nel corso dello stage che si serve del metodo di classificazione descritto nelle pagine precedenti per consigliare a un utente non esperto uno o più grafici che potrebbero rappresentare al meglio i dati disponibili. Tale compito di classificazione viene gestito e realizzato principalmente tramite un [sistema di regole](#), il cui funzionamento viene approfondito al paragrafo [Implementazione del sistema di regole](#). Per quanto riguarda il linguaggio di programmazione utilizzato, l'intero strumento è implementato usando [JS](#).

Lo strumento è disponibile a [questo link](#).

#### Analisi dei dati

In primis, è necessario individuare il numero di variabili numeriche e categoriche presenti nel *dataset*. *Chart-chooser* effettua tale processo in maniera semplificata, controllando se il primo valore per ogni variabile è un numero (variabile numerica) o una stringa (variabile categorica).

Nel caso sia necessario, viene inoltre richiesto all'utente di rispondere ad alcune domande per chiarire ulteriori aspetti riguardanti il tipo di dato e individuare le ulteriori sotto-distinzioni (come descritte nella sezione precedente). Ad esempio, nel caso si abbiano dati multivariati numerici verrà chiesto se i dati sono ordinati o meno, mentre se si hanno dati univariati numerici non saranno necessarie ulteriori domande.

#### Individuazione dell'obiettivo

Una volta definito il tipo di dato è necessario individuare l'obiettivo della visualizzazione. A tal fine, *Chart-chooser* richiede all'utente di fornire una breve descrizione di ciò che vuole rappresentare e comunicare. Il testo immesso viene poi elaborato da un [LLM](#), il quale interpreta la descrizione e ne estrae l'obiettivo più pertinente tra i nove possibili (definiti e descritti nella sottosezione [Obiettivi della visualizzazione](#)). Come [LLM](#) viene usato un modello a 7 miliardi di parametri, utilizzato attraverso [llama.cpp](#) e reso disponibile dall'azienda ospitante. I parametri sono perlopiù quelli di default, tranne la temperatura che viene impostata a 0 per garantire che l'output del modello sia il più predicibile possibile e restituisca solamente il numero dell'obiettivo.

Per quanto riguarda invece il metodo di [Prompt Engineering](#) utilizzato, si adopera una tecnica di [Few-Shot Learning \(FSL\)](#) implementata nello scenario dell'[In-Context](#)

**Learning (ICL).** Nello specifico, per ogni obiettivo vengono fornite al LLM alcune descrizioni d'esempio in modo tale da facilitare al modello il compito di classificazione e migliorarne le prestazioni. Questo approccio permette al LLM di comprendere meglio il contesto e rispondere con una maggiore precisione.

Il *prompt* utilizzato è il seguente:

```

1 Questa è una conversazione tra User e Llama, un chatbot amichevole.
  Llama è disponibile, onesto, bravo a scrivere e non manca mai di
  rispondere immediatamente e con precisione a qualsiasi
  richiesta.
2
3 User:
4 Dati i seguenti obiettivi:
5 Obiettivo 1: Mostrami la differenza tra i dati; Confronta due serie
  di dati divergenti; Mostrami la variazione rispetto a un valore.
6 Obiettivo 2: Mostrami un dato in relazione ad un altro; Mostrami
  come sono associati i due dati; Mostrami la correlazione tra un
  dato e i suoi due fattori.
7 Obiettivo 3: Mostrami la classifica; Mostrami i top; Mostrami i dati
  più importanti; Mostrami i più comuni; Mostrami i principali.
8 Obiettivo 4: Mostrami quante volte si verifica un evento; Mostrami
  la distribuzione del dato; Mostrami la media; Mostrami la
  frequenza.
9 Obiettivo 5: Mostrami la variazione nel tempo del dato; Mostrami la
  tendenza del dato nel tempo; Mostrami le differenze tra i dati
  nel tempo; Mostrami il cambiamento dei dati nel tempo; Mostrami
  le previsioni future.
10 Obiettivo 6: Mostrami i componenti del dato; Mostrami il dato diviso
  nei sottogruppi; Mostrami la struttura del dato; Mostrami la
  ripartizione del dato.
11 Obiettivo 7: Mostrami la dimensione effettiva del dato; Mostrami le
  unità; Mostrami la misura del dato; Mostra la rilevanza del dato
  all'interno di un contesto; Mostrami la grandezza dei dati.
12 Obiettivo 8: Mostrami la posizione; Mostrami i flussi nel mondo;
  Mostrami la mappa di distribuzione; Mostrami i paesi in cui si
  verifica l'evento; Mostrami la variazione nello spazio del dato.
13 Obiettivo 9: Mostrami i flussi tra i vari punti; Mostrami la rete;
  Mostrami le relazioni e la loro intensità; Mostrami cosa
  influisce nel dato; Mostrami come sono relazionati all'interno
  di un sistema.
14
15 Classifica la seguente frase, dimmi solamente il numero dell'
  obiettivo senza altre parole:
16 ${descrizione_utente}
17 Llama:

```

### Implementazione del sistema di regole

Le informazioni raccolte sono utilizzate da un [sistema di regole](#) per combinare le analisi e le classificazioni effettuate e restituire il nome dei possibili grafici applicabili al caso d'uso, se presenti. Si precisa, infatti, che non per tutti gli obiettivi sono presenti tutti i tipi di dato. Pertanto, la combinazione potrebbe portare a un risultato nullo.

La tecnologia per il [sistema di regole](#) scelta è [json-rules-engine](#). Tale decisione è motivata dalla sua leggerezza e velocità, nonché dal fatto che le regole siano semplici strutture [JSON](#), il che le rende più leggibili dall'uomo e facili da mantenere.

Gli elementi principali che caratterizzano un sistema di questo tipo sono:

- I *fatti*, ovvero dei valori costanti o funzioni pure.

- Le *regole*, ovvero delle strutture composte da:
  - *Condizioni*, ovvero i “requisiti” della regola, presentati nella forma:  
`[fact: nome_fatto, operator: tipo_operatore, value: valore]`.  
 Tali “requisiti” sono soddisfatti se la relazione definita dall’operatore tra il valore del *fatto* e il valore prestabilito indicato risulta vera. Si precisa, inoltre, che possono essere creati insieme di queste *condizioni* attraverso l’uso di operatori booleani.
  - *Eventi*, ovvero delle operazioni che il *motore* esegue in base al soddisfacimento o meno delle *condizioni*.
- Il *motore di inferenza* in sé, a cui vengono dati in input i *fatti* e aggiunte le *regole*. Esso è il cuore del sistema in quanto si occupa di valutare le *regole* basandosi sui *fatti* forniti, innescando *eventi* a seguito del soddisfacimento o meno delle *condizioni*.

Per quanto riguarda il funzionamento specifico di *Chart-chooser*, si hanno come *fatti* iniziali solamente il numero di variabili numeriche e categoriche. Per quanto riguarda gli altri *fatti*, che richiedono un input da parte dell’utente (e.g. l’obiettivo o il fatto di essere dati numerici ordinati o meno), questi sono forniti al *motore* durante la sua esecuzione, all’innescarsi di determinati *eventi*. Ciò permette di poter valutare subito le regole sui *fatti* già noti e nel frattempo poter ottenere dall’utente il valore dei rimanenti tramite **Promise**, fermando eventualmente l’esecuzione del *motore* solo prima di regole che necessitano dei *fatti* ancora sconosciuti. In tale maniera, inoltre, vengono richiesti all’utente e forniti al *motore* solamente i *fatti* che concernono la situazione specifica. Ad esempio, nel caso in cui si abbiano solamente dati numerici, non verrà chiesta all’utente e non sarà fornita inutilmente al *motore* l’informazione riguardante l’organizzazione delle variabili categoriche.

Il funzionamento descritto è reso possibile grazie all’adozione di un [sistema di regole](#) gerarchico. Nello specifico, a ciascuna *regola* viene attribuita una *priorità* che ne determina l’ordine di valutazione: una *priorità* maggiore comporta una valutazione anticipata; mentre, in caso di più *regole* a uguale *priorità*, queste verranno valutate in parallelo. Questo meccanismo viene sfruttato per realizzare una classificazione stratificata, in cui ogni livello di *priorità* rappresenta una fase di classificazione più dettagliata rispetto al livello precedente. Ne consegue che le *regole* appartenenti allo stesso livello siano mutualmente esclusive, poiché ciascuna di esse definisce un gruppo specifico in cui possono essere catalogati i dati forniti. Per consentire ai livelli inferiori di conoscere il gruppo identificato precedentemente e procedere con ulteriori sottoclassificazioni dello stesso, viene aggiunto ai *fatti* un valore booleano per ogni *regola*. Tale valore rappresenta il soddisfacimento delle *condizioni* della *regola*, assumendo il valore di **true** in caso positivo o **false** altrimenti.

Si ha dunque il seguente [sistema di regole](#):

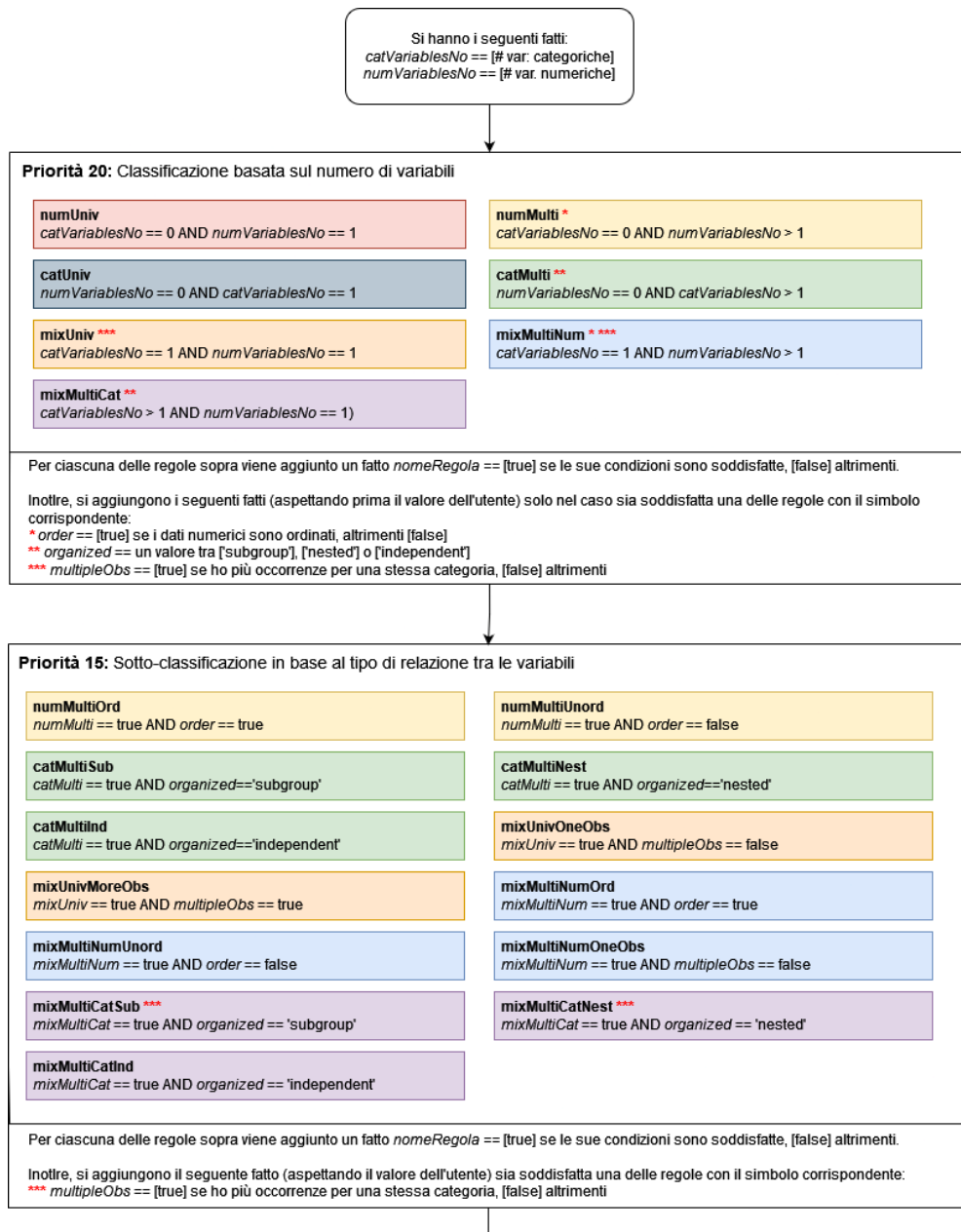
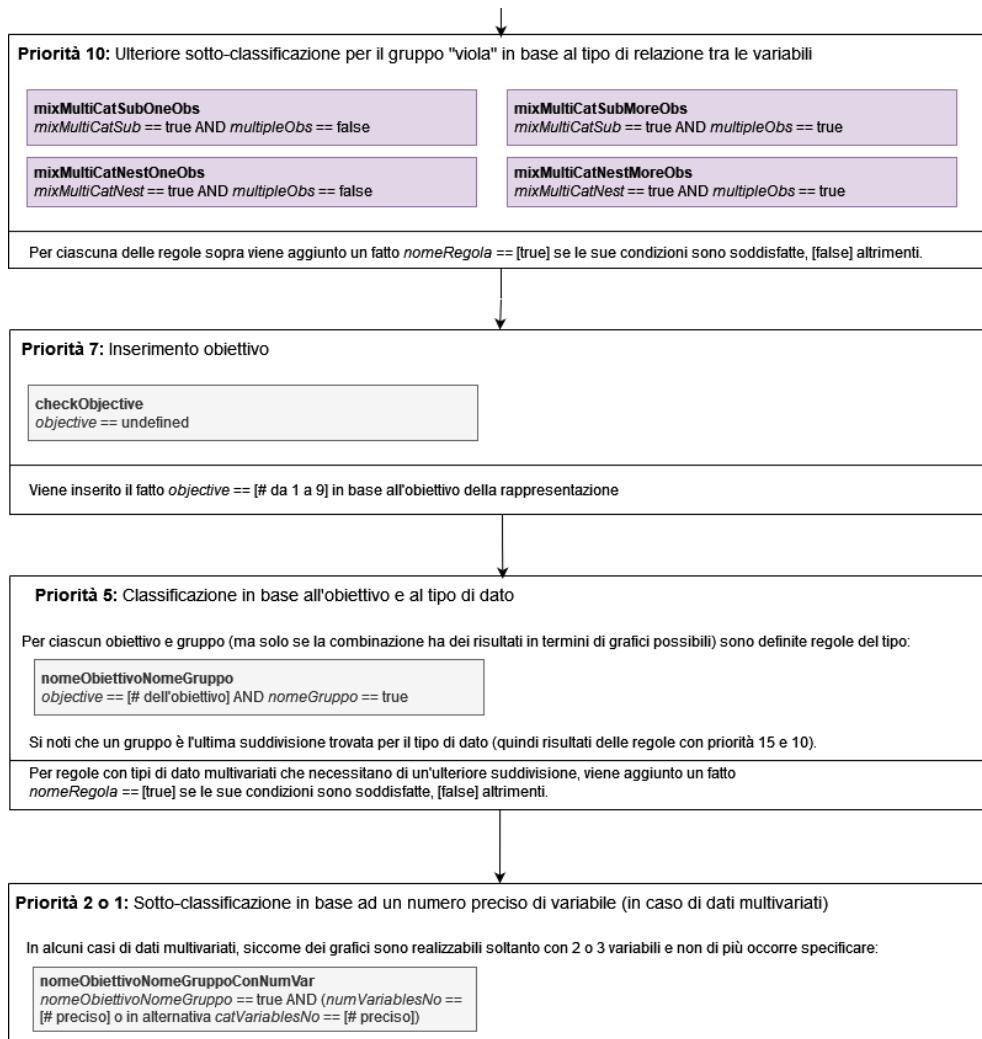


Figura 2.3: Sistema gerarchico di priorità per le regole di Chart-chooser - priorità 20 e 15



**Figura 2.4:** Sistema gerarchico di priorità per le regole di Chart-chooser - priorità dalla 10 all'1

Dalla figura 2.4 si osserva un'eccezione rispetto al funzionamento descritto in precedenza per quanto riguarda il livello di *priorità* 7. In questo caso, il livello non effettua alcuna suddivisione, bensì si occupa solamente di recuperare il numero di obiettivo, necessario per poter proseguire con la classificazione. Tale configurazione permette di valutare le *regole* relative al tipo di dato mentre il numero di obiettivo viene ricavato dall'elaborazione del *LLM* dell'input utente, anziché attendere passivamente il completamento di quest'ultima.

### Proposta di grafici

Le *regole* di *priorità* minore o uguale a 5, se soddisfatte, innescano degli *eventi* che hanno come parametri il nome dei grafici adatti al caso d'uso specificato dalla *regola* stessa. Il *motore*, dunque, al termine della valutazione delle *regole*, controlla gli *eventi*



emessi e mostra all'utente i parametri di questi, ovvero il nome dei grafici trovati (se presenti).

I grafici individuabili sono riportati nell'[Appendice A](#) del documento.

### Limitazioni dello strumento

Come già accennato precedentemente *Chart-chooser* è uno strumento prototipale, presenta infatti diverse limitazioni.

Innanzitutto, l'applicativo è disponibile soltanto da riga di comando e non dispone di un'interfaccia grafica. Inoltre, esso richiede che i dati siano disponibili in locale e che siano formattati secondo uno schema specifico (una specifica struttura [JSON](#)). L'utente deve, altresì, essere a conoscenza del *dataset* per poter rispondere correttamente alle domande relative a esso.

Ciononostante, *Chart-chooser* ha il potenziale per essere migliorato e offrire un'esperienza utente più semplice e intuitiva. Una possibile direzione per uno sviluppo futuro potrebbe essere l'implementazione di un *parsing* automatico dei dati e la capacità di rilevare e interpretare informazioni chiave direttamente dai *dataset*, riducendo così la dipendenza dall'intervento dell'utente.

## Capitolo 3

# Infografiche

*In questo capitolo verranno esaminate le caratteristiche funzionali e di design delle infografiche e presentate le loro applicazioni. Verrà inoltre fornita una classificazione delle infografiche basata sulla loro struttura e output visivo.*

### 3.1 Definizione e applicazioni

#### 3.1.1 Cos'è un'infografica e in cosa si distingue dalla data visualization

Analogamente alla [Data Visualization](#), le infografiche si occupano di rappresentare visivamente i dati al fine di comunicarli in maniera più semplice e accessibile. Esse trasformano i dati in idee chiare, permettendo di inquadrare facilmente un fenomeno, una situazione o un processo, coadiuvando così la presa di decisioni ponderate sul tema presentato. Queste idee sono generate a partire dall'interpretazione, formalizzazione e contestualizzazione dei dati, che spesso sono disomogenei per formato, tipo di contenuto e origine. Tale interpretazione ha il compito di svelare il significato nascosto nel *dataset* e selezionare le connessioni o fenomeni più interessanti per il destinatario della comunicazione. In altre parole, l'infografica non si limita a mostrare i dati, ma li organizza e li presenta in modo da renderne più evidente significato e importanza.

Per raggiungere questo obiettivo, le infografiche combinano elementi testuali e grafici per presentare le informazioni; usando le parole di Edward R. Tufte “un'infografica mostra visivamente grandezze misurate mediante l'uso combinato di punti, linee, un sistema di coordinate, numeri, simboli, parole, ombreggiature e colore”<sup>1</sup>. Pertanto, le infografiche rigettano la tradizione di “purezza grafica” (i.e. rappresentazione dei dati basata solo su grafici, diagrammi e mappe) imposta alla [Data Visualization](#) e abbracciano, invece, un insieme molto più ampio di tecniche di comunicazione visiva, come illustrazioni, immagini, etc. Inoltre, il focus dell'infografica non è solamente l'oggetto da rappresentare, come invece accade con la [Data Visualization](#). Infatti, le infografiche prendono in considerazione anche il destinatario della comunicazione, scegliendo e disponendo gli elementi per costruire una narrazione. Possiamo infatti pensare all'infografica come a una forma di racconto, di lettura che non è di per sé né imparziale né completa, ma si limita a fare luce ed esporre i punti salienti individuati nel

---

<sup>1</sup>Tufte, *The Visual Display of Quantitative Information, Second Edition*.

*dataset*. In realtà, tale caratteristica delle infografiche è proprio ciò che maggiormente le distingue dalla [Data Visualization](#). Infatti, mentre le prime sono progettate per presentare i dati raccontandoli attraverso storie specifiche, guidando il lettore con una narrazione, la seconda è più orientata verso l'esplorazione autonoma delle informazioni da parte del fruitore, che deve analizzare e scoprire i dati per conto proprio. Questa differenza nel livello di analisi richiesto si ripercuote anche sul tipo di pubblico di ciascuna tipologia di rappresentazione. Nello specifico, le infografiche, grazie alla loro capacità di semplificare e contestualizzare i dati, hanno un pubblico più ampio, in quanto sono accessibili anche a coloro che non possiedono capacità analitiche avanzate.

Nonostante queste differenze, infografiche e [Data Visualization](#) sono strettamente connesse e complementari. Le infografiche, pur essendo maggiormente orientate alla narrazione, possono includere elementi che stimolano l'esplorazione autonoma dei dati. Allo stesso modo, le *visualizzazioni dei dati* possono contenere aspetti narrativi che facilitano l'interpretazione.

### 3.1.2 Applicazioni e vantaggi

Le infografiche sono strumenti versatili che vengono utilizzati in una vasta gamma di settori e contesti, in particolare al fine di analizzare o presentare dati. Offrono infatti vantaggi significativi che le rendono particolarmente adatte a questi scopi.

Per quanto riguarda l'**analisi dei dati**, le infografiche sono utili per la cosiddetta [Exploratory Data Analysis](#), dove il fruitore non ha una domanda precisa a cui sta cercando risposta, bensì vuole semplicemente scoprire cosa emerge di interessante dal *dataset*.

Per quanto riguarda invece la **presentazione dei dati**, va da sé che le infografiche siano perfette per tale scopo, in quanto grazie alla narrazione riescono a comunicare in maniera efficace e abbracciare in tal modo un pubblico molto ampio.

In generale, quale sia l'applicazione, l'uso di infografiche consente anche di:

- **Facilitare la scoperta di informazioni** nascoste e **stimolare la curiosità** per una ricerca futura;
- **Semplificare la comprensione**, facilitando così il raggiungimento della *saggezza*;
- **Raccontare una storia** che permetta una visione d'insieme della situazione, fenomeno o processo presentato.

## 3.2 Classificare le infografiche

### 3.2.1 Metodi di classificazione

Una prima classificazione delle infografiche può essere fatta **in base all'output visivo**. Nello specifico, si hanno:

- **Infografiche statiche**, in cui le informazioni sono rappresentate e visibili tutte in un'unica volta, avendo così un impatto più veloce e immediato sul pubblico.
  - Esempio: infografiche presenti nei giornali.
- **Infografiche animate**, in cui le informazioni sono presentate sequenzialmente in maniera consistente oppure contengono altri tipi di animazione.

- Esempio: infografiche realizzate tramite presentazioni multimediali.
- **Infografiche interattive**, in cui le informazioni sono presentate in base a scelte operate dall'utente.
  - Esempio: infografiche online in cui è l'utente a selezionare il dettaglio a partire da un insieme complesso di informazioni visualizzabili.

Un'altra classificazione può essere fatta **in base all'obiettivo**. Nello specifico, si hanno:

- **Infografiche basate su processi ordinati**, le quali vengono usate per mostrare passo dopo passo un processo o una sequenza ordinata di informazioni, ad esempio per insegnare qualcosa di nuovo all'utente (e.g. ricette).
- **Infografiche a mo' di lista**, in questo caso le informazioni sono rappresentate sotto forma di elenco testuale, eventualmente accompagnato da icone. Per costruzione, queste infografiche permettono di mostrare molte informazioni e, allo stesso tempo, di farle comprendere velocemente.
- **Infografiche basate sul tempo**, in cui viene mostrata una sequenza cronologica di eventi che delinea come si evolve una storia o un soggetto nel tempo. Quest'ultimo è solitamente rappresentato tramite una linea, chiamata appunto "linea del tempo".
- **Infografiche basate sul confronto**, esse comparano due diversi argomenti in contrasto ponendoli uno a fianco all'altro. Ciò risulta molto utile per mostrare quali elementi i due hanno in comune oppure per quanto divergono uno rispetto all'altro.
- **Infografiche basate su una gerarchia**, le quali vengono utilizzate per mostrare informazioni organizzate per livelli e il collegamento tra questi.
- **Infografiche basate su informazioni correlate poste allo stesso livello**, in cui si vuole comunicare un nuovo concetto o dare una panoramica su un argomento; infatti, tali infografiche sono anche dette "Infografiche informative". Solitamente, lo spazio viene diviso in sezioni con un proprio titolo descrittivo.

### 3.2.2 Infographic-helper

Come accennato in precedenza, l'infografica è una forma di racconto e, pertanto, la scelta e scrittura della storia da comunicare sono centrali nella sua costruzione. Nel corso dello stage si è dunque voluto sviluppare uno strumento prototipale, *Infographic-helper*, che potesse aiutare il "narratore" in tale compito. Nello specifico, questo strumento consente di validare la narrativa sviluppata, oltre a individuarne le parti e l'obiettivo principale. Per quanto riguarda il linguaggio di programmazione utilizzato, l'intero strumento è implementato usando [JS](#).

Lo strumento è disponibile a [questo link](#).

### Validazione della narrativa

Uno degli obiettivi di *Infographic-helper* è garantire che la narrativa dell'infografica sia persuasiva e venga recepita efficacemente dal pubblico. Questi obiettivi coincidono con quelli della retorica, la quale fornisce i principi per costruire messaggi convincenti e chiari, che espongono il punto di vista del comunicatore in maniera comprensibile. Questi principi non solo sono applicabili alla pura comunicazione verbale, ma si rivelano utili anche per le infografiche. Infatti, nonostante queste siano rappresentazioni visive, esse sono comunque una forma di narrazione e comunicazione e pertanto oggetto di retorica.

Aristotele, pioniere di quest'arte, definisce suddetti principi come *ethos*, *pathos* e *logos*. Nello specifico:

- ***Ethos***, comprende tutto ciò che stabilisce o aumenta la credibilità e autorità del comunicatore e delle fonti;
- ***Pathos***, comprende tutto ciò che suscita emozioni nel pubblico, che crea una connessione emotiva con gli spettatori;
- ***Logos***, comprende tutto ciò che fa parte di un'argomentazione razionale, logica e coerente, come dati e statistiche.

*Infographic-helper* si ispira dunque a questi principi aristotelici per migliorare la costruzione delle storie delle infografiche. Nello specifico, controlla che questi elementi siano presenti nel testo e, eventualmente, ne suggerisce di altri.

In aggiunta a questi elementi retorici, *Infographic-helper* tiene conto anche del target dell'infografica, ovvero il pubblico a cui è destinata, per poter contestualizzare meglio la storia. Inoltre, lo strumento controlla anche se siano presenti eventuali elementi estranei o irrilevanti per il tema principale della storia.

Queste funzionalità sono implementate attraverso interrogazione a un [LLM](#), nello specifico un modello a 7 miliardi di parametri, utilizzato attraverso [llama.cpp](#) e reso disponibile dall'azienda ospitante.

Il *prompt* utilizzato è il seguente:

```

1 Questa è una conversazione tra User e Llama, un chatbot amichevole.
   Llama è disponibile, onesto, bravo a scrivere e non manca mai di
   rispondere immediatamente e con precisione a qualsiasi
   richiesta.
2
3 User:
4 Devo presentare il seguente testo a ${target}.
5 ${ethos}.
6
7 Testo:
8 ${storia}
9
10 Domande:
11 1. Supponi che un elemento sia di "pathos" se può indurre il
   pubblico a sentire (o non sentire) una connessione emotiva con
   il contenuto.
12 Riporta tutti gli elementi (frasi, espressioni e parole) di "
   pathos" del Testo, SOLO se hanno senso nel contesto del
   Testo. Suggestisci altre frasi incisive di "pathos" da poter
   inserire nel Testo.
13 2. Supponi che un elemento sia di "ethos" se può indurre il
   pubblico a ritenere che l'autore sia (o meno) affidabile e
   credibile.
```

```

14      Riporta tutti gli elementi (frasi, espressioni e parole) di "
        ethos" del Testo, SOLO se hanno senso nel contesto del Testo
        . Suggestisci altre frasi incisive di "ethos" da poter
        inserire nel Testo basandoti sul mio ruolo SE rilevante al
        tema del Testo.
15 3.  Supponi che un elemento sia di "logos" se può indurre il
        pubblico a credere che l'argomentazione sia (o meno) logica e
        supportata da prove adeguate.
16      Riporta tutti gli elementi (frasi, espressioni e parole) di "
        logos" del Testo, SOLO se hanno senso nel contesto del Testo
        . Suggestisci altre frasi incisive di "logos" da poter
        inserire nel Testo.
17 4.  Supponi che un elemento sia "estraneo" se può ritenersi estraneo
        rispetto al tema principale del Testo oppure può ritenersi un'
        assurdità.
18      Riporta tutti gli elementi (frasi, espressioni e parole) "
        estraneo" presenti nel Testo.
19
20 [...continua...]

```

dove `$target` rappresenta il pubblico di destinazione dell'infografica e `$ethos` rappresenta il ruolo di chi espone la narrativa, inseriti per una maggiore contestualizzazione della storia poi riportata (inserita al posto di `$storia`).

### Ricerca del layout da utilizzare

L'altro fine di *Infographic-helper* è quello di classificare la storia in base alla “classificazione per obiettivo” (definita alla sezione 3.2.1) e trovare le sue principali parti costitutive. Tali informazioni consentono di individuare il *layout* più adatto per raccontare visivamente la storia e definire l'argomento specifico da trattare in ciascuna sezione di tale *layout*.

Le possibili disposizioni sono descritte alla sezione 3.3.1.

Queste funzionalità sono anch'esse implementate interrogando il LLM descritto precedentemente.

In questo caso, il *prompt* utilizzato è il seguente:

```

1  [...segue...]
2  5.  Quali sono le parti in cui si divide il testo? Per ogni parte,
        dimmi solo un titolo breve.
3  6.  Tola l'introduzione, identifica la relazione generale tra le
        parti come una delle seguenti:
4      - parti di una lista,
5      - parti di una linea del tempo,
6      - parti poste a confronto,
7      - parti che compongono gerarchia,
8      - parti correlate poste allo stesso livello,
9      - parti di un processo ordinato.
10
11 Fornisci solamente le risposte nel seguente formato JSON:
12 { "pathos": { "presente": boolean, "elementi": [string], "
        suggerimenti": [string] }, "ethos": { "presente": boolean, "
        elementi": [string], "suggerimenti": [string] }, "logos": { "
        presente": boolean, "elementi": [string], "suggerimenti": [
        string] }, "elementi_estraneo": {"presente": boolean, "elementi
        ": [string]}, "parti": [string], "relazione": "string" }
13
14
15 Llama:

```

Si noti inoltre che è stata definita una [grammatica](#) corrispondente allo schema menzionato nel *prompt* (fornito per ulteriore chiarezza), la quale consente di ottenere l'output del [LLM](#) nel formato fisso [JSON](#) indicato. Questo approccio facilita il *parsing* dei dati, garantendo risposte precise con un formato standard, e ne agevola l'eventuale integrazione futura con un'interfaccia grafica.

### Limitazioni dello strumento

Come già accennato precedentemente *Infographic-helper* è uno strumento prototipale, presenta infatti diverse limitazioni.

Innanzitutto, l'applicativo è disponibile soltanto da riga di comando e non dispone di un'interfaccia grafica. Inoltre, esso non ha la capacità di individuare molteplici obiettivi. Potrebbero essere infatti presenti diverse macro-sezioni nella narrativa con obiettivi distinti. Lo strumento, inoltre, non è capace di valutare la veridicità del contenuto della storia, ma valida soltanto il modo in cui questo è comunicato. Pertanto, è compito dell'utente assicurarsi che quanto scritto corrisponda a verità.

Ciononostante, *Infographic-helper* ha il potenziale per essere migliorato e offrire un'esperienza utente più semplice e intuitiva.

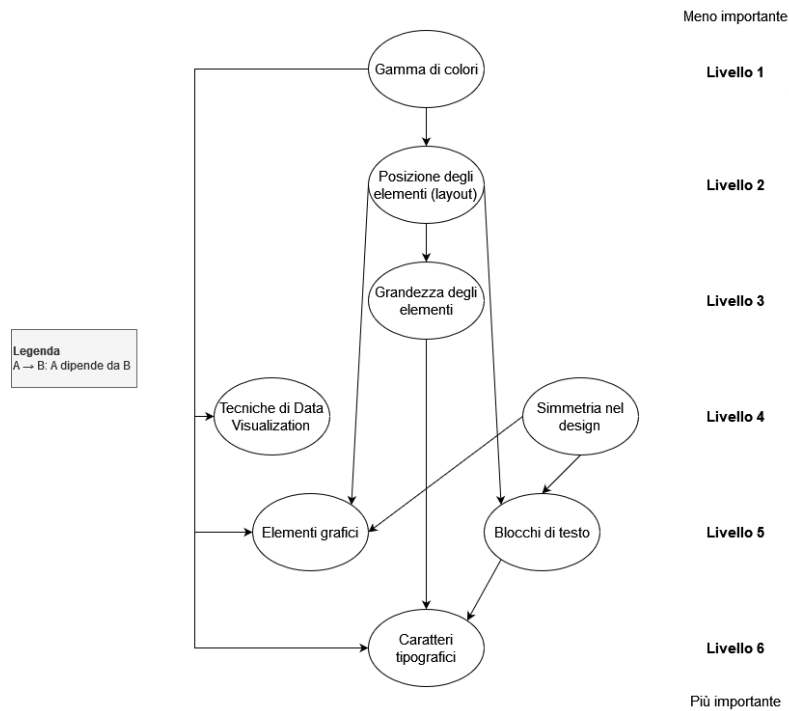
## 3.3 Design delle infografiche

### 3.3.1 I fattori che influenzano il design

Studi<sup>2</sup> hanno permesso di ordinare gerarchicamente i fattori che influenzano la struttura del design delle infografiche. Tale gerarchia è suddivisa in sei livelli come segue:

---

<sup>2</sup>Oleksandr Tymchenko, Svitlana Vasiuta e Orest Khamula. «Optimization of the Mathematical Model of Factors of Composite Design of Infographic». In: *2018 IEEE 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT)*. 2018.



**Figura 3.1:** Gerarchia dei fattori che influenzano il design delle infografiche

Vediamo dunque come applicare al meglio tali fattori nella costruzione del design di un'infografica.

### I colori

Il colore è un elemento cruciale nel design delle infografiche e, pertanto, è da utilizzare con cura. Esso infatti, se scelto in maniera appropriata, può migliorare la leggibilità e facilitare la comprensione e memorizzazione dell'informazione.

Generalmente, si utilizza il colore per catturare l'attenzione, evidenziare dati specifici e rafforzare il messaggio narrativo. Quest'ultimo punto, in particolare, è possibile in quanto ogni colore porta con sé un significato simbolico, influenzato da cultura e geografia, che permette di evocare sensazioni e valori specifici nel pubblico. Di seguito sono riportati i valori più spesso associati ai diversi colori, con riferimento alla cultura dell'Occidente:

- Colori caldi:
  - **Rosso:**
    - \* Valori: energia, temperamento forte, passione.
    - \* In linea con tali valori, viene utilizzato principalmente per attirare l'attenzione.
  - **Arancione:**
    - \* Valori: ottimismo, vitalità, energia.
    - \* In linea con tali valori, viene generalmente utilizzato per temi giovanili e/o dinamici.



- **Giallo:**
  - \* Valori: leggerezza, felicità, vitalità.
  - \* In linea con tali valori, viene anch'esso utilizzato principalmente per attirare l'attenzione, specie per temi innovativi.
- Colori freddi:
  - **Verde:**
    - \* Valori: salute, natura, tranquillità.
    - \* In linea con tali valori, viene generalmente utilizzato per temi riguardanti la sanità, salute ed ecologia.
  - **Blu:**
    - \* Valori: calma, stabilità.
    - \* In linea con tali valori, viene utilizzato ad esempio da banche o compagnie assicurative.
  - **Viola:**
    - \* Valori: eleganza, nobiltà, spiritualità.
    - \* In linea con tali valori, viene generalmente utilizzato per temi riguardanti il settore della bellezza o del lusso.
- Colori neutri:
  - **Marrone:**
    - \* Valori: terra, calore.
    - \* In linea con tali valori, viene generalmente utilizzato per temi riguardanti la natura.
  - **Bianco:**
    - \* Valori: purezza, pulizia, semplicità.
    - \* In linea con tali valori, viene generalmente utilizzato per temi riguardanti la sanità.
  - **Nero:**
    - \* Valori: ribellione, formalità, potenza.
    - \* In linea con tali valori, viene generalmente utilizzato per temi di protesta o riguardanti il settore del lusso.
  - **Grigio:**
    - \* Valori: modernità, raffinatezza.
    - \* In linea con tali valori, viene generalmente utilizzati per temi riguardanti tecnologia o moda.

In ogni caso, è fondamentale mantenere un tono di colore coerente nell'infografica, preferibilmente usando tonalità sobrie.

Si aggiunge infine che, come per la [Data Visualization](#), anche in questo caso, è importante scegliere una gamma di colori che sia accessibile anche a persone daltoniche o *color-deficient*.

## I layout

Gli elementi di un'infografica possono essere disposti in un massimo di sei livelli:

- **titolo principale;**
- **sottotitolo;**
- **contenuto principale e secondario;**
- **elementi di supporto**, come ad esempio divisori, contenitori etc.;
- **elementi di dettaglio**, come ad esempio le note a margine;
- **comandi** di interazione e *highlights* dei dati.













Per quanto riguarda nello specifico la struttura del contenuto principale e secondario, le infografiche sono solite utilizzare una “composizione a *panel*”, ovvero si hanno diverse componenti grafiche, dette *gruppi visivi*, leggibili in sequenza. Questo approccio differisce notevolmente dalle pratiche comunemente usate nella [Data Visualization](#), dove tutte le sfaccettature del tema d'interesse sono presentate in un'unica composizione grafica.

La direzione e ordine in cui questa sequenza di *gruppi visivi* viene presentata all'interno dell'infografica dipende dalla storia che si vuole raccontare. Infatti, per trasmettere efficacemente le informazioni e la narrativa, è necessario avere una struttura semantica sottostante all'infografica che collega i *gruppi visivi* in una determinata maniera dipendente dal messaggio. Questa struttura è detta [Visual Information Flow \(VIF\)](#), letteralmente “flusso delle informazioni visive”. Tale flusso è spesso rimarcato ed evidenziato tramite l'uso di “suggerimenti narrativi”, che possono essere espliciti, come frecce o descrizione testuali, oppure impliciti, se provengono da vari principi di design come ad esempio i principi di *Gestalt* (approfonditi alla sezione [3.3.3](#)).

Alcuni studi<sup>3</sup> hanno permesso di individuare i seguenti *pattern* principali di [VIF](#), distinti in base alla loro *backbone shape*, ovvero la forma della linea lungo cui la maggior parte dei *gruppi visivi* è allineata nella narrazione.

---

<sup>3</sup>Min Lu et al. «Exploring Visual Information Flows in Infographics». In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 2020.

Backbone Shape	Pattern	Backbone Shape	Pattern
	Clock		Down-ladder
	Dome		Up-ladder
	Bowl		Landscape
	Left-wing		Portrait
	Right-wing		Pulse
	Star		Spiral

**Tabella 3.1:** Backbone shape dei pattern di Visual Information Flow (VIF)

Le informazioni, dunque, saranno disposte all'interno dell'infografica a seconda dell'obiettivo della storia da raccontare. Per la definizione degli obiettivi si rimanda alla sezione 3.2.1.

Nello specifico, vengono definite le seguenti disposizioni generali, a cui si possono facilmente far corrispondere i diversi *pattern* di VIF sopraelencati.

- **Disposizione a lista:**
  - Applicazione: infografiche che hanno come obiettivo rappresentare un elenco (e.g. orari o programmi), ma anche infografiche basate su una gerarchia.
  - Direzione: la storia viene presentata a partire dall'alto verso il basso. In alcuni casi, potrebbe anche essere orientata da sinistra verso destra.
  - *Pattern* VIF:
    - \* **Portrait;**
    - \* **Landscape;**
    - \* **Down-ladder.**

- Disposizione **a passi**:
  - Applicazione: infografiche che rappresentano processi ordinati.
  - Direzione: la storia viene presentata dal basso verso l'alto e/o seguendo un percorso a zigzag che parte da sinistra e, generalmente, termina al lato opposto.
  - *Pattern VIF*:
    - \* **Pulse**;
    - \* **Spiral**;
    - \* **Up-ladder**.
- *Timeline*:
  - Applicazione: infografiche basate sul tempo.
  - Direzione: la storia viene presentata da sinistra verso destra oppure dall'alto verso il basso, solitamente avendo come riferimento una singola linea rappresentante il tempo.
  - *Pattern VIF*:
    - \* **Portrait**;
    - \* **Landscape**.
- Disposizione **a mandala**:
  - Applicazione: infografiche basate su informazioni correlate poste allo stesso livello.
  - Direzione: la storia si sviluppa dal centro verso l'esterno, seguendo un *pattern* radiale che si estende nelle quattro direzioni diagonali o, più in generale, verso tutti i lati in modo sistematico.
  - *Pattern VIF*:
    - \* **Clock**;
    - \* **Star**;
    - \* Variazioni dei suddetti modelli, come: **dome**, **bowl**, **left-wing** e **right-wing**.
- Disposizione **a matrice e divisa in due**:
  - Applicazione: infografiche basate sul confronto.
  - Direzione: in caso di confronto tra due elementi lo spazio viene diviso verticalmente in due, mentre in caso di quattro si ha una suddivisione data dai quattro angoli, a formare una sorta di piano cartesiano.
  - *Pattern VIF*:
    - \* **Clock**.

Si noti che i *pattern VIF* potrebbero essere applicati in maniera composita per diverse macro-aree dell'infografica.

### La grandezza degli elementi

La scelta della dimensione degli elementi deve essere fatta con cura per poter garantire l'accessibilità dei dati, una simmetria nel design e poter trasmettere correttamente l'ordine di importanza dei dati presentati.

Innanzitutto, è essenziale che tutti gli elementi abbiano dimensioni sufficienti a essere visti e/o letti agevolmente. Ciò è particolarmente importante per infografiche web destinate alla visualizzazione su schermi con dimensioni diverse, come computer e dispositivi mobili.

In secondo luogo, gli elementi e dati più rilevanti dovrebbero avere dimensioni maggiori per poter attirare l'attenzione del lettore, mentre i dettagli secondari possono essere più piccoli. La scala degli elementi è importante dunque anche per dare una sorta di "gerarchia visiva", che permette di strutturare meglio l'infografica, facilitando il confronto tra elementi con stesso livello ed evidenziando sezioni e relazioni tra le parti.

### Le tecniche di visualizzazione dei dati inserite

Un elemento determinante nella presentazione dei dati nelle infografiche è rappresentato dalle tecniche di [Data Visualization](#) utilizzate e, nello specifico, dalla scelta dei grafici e diagrammi. Questo argomento è già stato affrontato nella sezione [2.3](#), che riguarda la classificazione dei grafici e la loro scelta tramite lo strumento *Chart-chooser*.

### La simmetria nel design

La simmetria è un altro principio fondamentale nel design delle infografiche, poiché contribuisce a creare un senso di equilibrio e ordine visivo che può aiutare l'occhio del lettore a orientarsi all'interno dell'infografica.

Tale equilibrio può essere raggiunto in due diversi modi:

- ***Symmetrical balance***: si distribuiscono gli elementi in modo simmetrico rispetto a un asse centrale, orizzontale o rispetto a un punto centrale. L'obiettivo è creare una composizione visivamente stabile, elegante e prevedibile. Tale approccio porta a creare infografiche con *layout* molto strutturati e ordinati.
  - Tale metodo è visibile nella maggior parte dei *pattern* [VIF](#), come ad esempio *portrait* (asse orizzontale), *landscape* (asse verticale) o *layout* circolari (attorno a punto centrale).
- ***Asymmetrical balance***: si accetta un certo livello di asimmetria e dinamicità nel *layout* dell'infografica, pur mantenendo una sensazione generale di equilibrio visivo. In questo approccio, la scelta di proporzioni e distribuzione degli elementi è più flessibile, tuttavia si pone comunque una particolare attenzione a evitare la presenza di parti nell'infografica che appaiano troppo pesanti o trascurate.
  - Esistono modelli [VIF](#), come *up-ladder* e *down-ladder*, dove gli elementi sono decentrati e posizionati sul lato opposto del titolo per mantenere l'equilibrio.

### Gli elementi grafici

Oltre a grafici, diagrammi ed elementi testuali, l'infografica può contenere anche altre tecniche di comunicazione visiva. Tra queste, ad esempio, vengono spesso utilizzati illustrazioni, icone e simboli.

Tali tecniche consentono di facilitare la memorizzazione, grazie alla loro potenza evocativa e alla loro capacità di richiamare oggetti facilmente riconoscibili dall'uomo.

Alcune linee guida da seguire per ottimizzarne l'uso, e costruire così una comunicazione efficace, sono:

- Seguire i **principi di Gestalt** (meglio approfonditi alla sezione 3.3.3);
- **Selezionare immagini evocative** che possano provocare un'esperienza sensoriale all'utente;
- Fare particolare **attenzione alla risoluzione dell'immagine** in modo che questa sia sempre visibile e nitida.

### I blocchi di testo

È bene ricordare che l'infografica non è lo strumento perfetto per ogni situazione. L'infografica, infatti, è utile solo quando si dispone di sufficienti informazioni per raccontare una storia, senza però eccedere. Un testo troppo lungo può infatti creare "rumore" e compromettere l'efficacia della comunicazione.

Generalmente i blocchi di testo presenti in un'infografica sono:

- Il **titolo**;
- Il **sottotitolo**, che specifica i contenuti annunciati con il titolo e pertanto non può esistere senza quest'ultimo;
- L'**introduzione**, inserita solitamente vicino al titolo e posta in corsivo, specifica ulteriormente il tema dell'infografica e funge da "guida" per la sua lettura;
- Il **testo principale**, che serve a spiegare meglio le informazioni visualizzate con altri mezzi (e.g. grafici);
- Le **note a piè di pagina**, che servono a dare informazioni a carattere formale (e.g. fonti) all'utente e dunque utilizzano il più piccolo carattere leggibile possibile per evitare di distrarlo.

Tra questi, la parte forse più importante degli elementi testuali di un'infografica sono i titoli (sia quello globale sia quelli, seppur in minor misura, dei singoli blocchi). Le intestazioni devono, infatti, essere incisive e invogliare il fruitore dell'infografica a continuare a visionare la presentazione. A tal fine, i titoli sono generalmente formulati in uno dei seguenti modi:

- **Sotto forma di domanda**, si pone un interrogativo curioso o interessante che stimoli il lettore a cercarne la risposta nei dati presentati.
- **Includendo la statistica principale**, evidenziando il dato presentato più significativo.
- **Come confronto**, si mettono in relazione fin da subito due o più elementi di cui si mostreranno differenze e similitudini.
- **Come introduzione a un elenco numerato**, il titolo funziona come introduzione a una lista (e.g. "Top X").

### I caratteri tipografici

L'elemento più importante nel design di un'infografica è la scelta del carattere tipografico, il quale influisce pesantemente sulla leggibilità e comprensibilità dell'infografica. Nello specifico, la gamma di caratteri utilizzati non deve essere troppo ampia (max. 3 o 4); bisogna invece concentrarsi sulla scelta del tipo di carattere, il quale può avere un impatto significativo sul design complessivo dell'infografica.

A tal fine, si possono applicare le seguenti regole:

- Per quanto riguarda l'uso di **caratteri graziati** (*serif*), essi sono da utilizzare quando si vuole conferire serietà e autorevolezza al testo. Tuttavia, possono risultare meno leggibili su schermo.
  - Esempio: Times New Roman.
- Per quanto riguarda l'uso di **caratteri senza grazie** (*sans serif*), essi sono da utilizzare per testi brevi e/o su schermi, in quanto più leggibili.
  - Esempio: Arial (adatto per testi lunghi, essendo più stretto), Verdana (ideale per frasi ad effetto), Helvetica.
- Per quanto riguarda l'uso di **caratteri decorativi**, essi sono spesso difficili da utilizzare e generalmente, dunque, sono da evitare. Possono essere considerati, seppur con prudenza, per casi specifici per riflettere il tema discusso nella presentazione.
  - Esempio: caratteri calligrafici o fumettistici.
- Per quanto riguarda l'uso del **grassetto** (*bold*), è da utilizzare per evidenziare componenti testuali brevi e focalizzare l'attenzione su di essi.
- Per quanto riguarda l'uso del **corsivo** (*italic*), è da utilizzare per citazioni o termini particolari.
- Per quanto riguarda l'uso del **maiuscolo**, è da utilizzare per dare maggiore importanza o evidenziare testi brevi, come i titoli, in quanto corrisponde ad un "innalzamento della voce".

Sempre con riferimento ai caratteri, è importante seguire anche le seguenti regole:

- Per quanto riguarda la **sottolineatura**, nel web questa è da evitare per evidenziare elementi (preferendo invece il grassetto o il maiuscolo), in quanto solitamente rappresenta un collegamento ipertestuale.
- Per quanto riguarda l'**interlinea**, conviene che la distanza tra una riga e l'altra sia abbastanza ampia da consentire una lettura agevole.
- Per quanto riguarda l'**allineamento**, si utilizza solitamente il "centrato" per i titoli, specie se non accompagnati da ulteriore testo. Per il resto, si preferisce invece l'allineamento "a sinistra".
- Per quanto riguarda la **lunghezza**, testi troppo lunghi non sono letti volentieri dal pubblico e, inoltre, necessiterebbero di caratteri più piccoli, meno leggibili.

### 3.3.2 Interattività

Per quanto riguarda la navigazione ed esplorazione delle infografiche si applica generalmente il cosiddetto **principio delle informazioni visive** (postulato dall'informatico statunitense Ben Shneiderman), che afferma: “prima panoramica, zoom e filtri, poi dettagli su richiesta”. Si devono dunque presentare i punti salienti e più rilevanti, mentre ulteriori informazioni sono fornite agli utenti solo attraverso l'esplorazione dell'infografica.

Le tecniche di esplorazione e navigazione possibili sono le seguenti:

- **Scorrimento e *panning***, il primo utilizzato generalmente su siti web con orientamento verticale, mentre il secondo viene usato in genere su mappe per spostarsi;
- **Zoom**;
- **Apertura e chiusura**, ad esempio per aprire o chiudere finestre *pop-up* di dettaglio;
- **Classificazione e riordino**, ad esempio riordinando i dati contenuti in una tabella;
- **Ricerca e filtro**.

L'utente può impiegare tali tecniche interagendo con l'interfaccia nei seguenti modi:

- **Tramite input diretti**, l'utente dà un input diretto, ad esempio clicca un pulsante, e l'infografica si adatta di conseguenza;
- **Tramite richieste**, l'utente “dialoga” tramite apposite finestre con la presentazione e ne cambia, in tal modo, i parametri;
- **Tramite manipolazione**, l'utente cambia la struttura o l'aspetto della visualizzazione, ad esempio tramite il filtraggio dei dati rappresentati;
- **Tramite esplorazione**, l'utente visualizza l'infografica secondo un suo percorso, ad esempio muovendosi a piacimento all'interno di modelli 3D.

Applicare queste tecniche aiuta a rendere la presentazione dei dati più coinvolgente e chiara, contribuendo così a raggiungere meglio gli obiettivi delle infografiche. Tuttavia, per un loro giusto impiego, è necessario seguire questi principi:

- **Visibilità**, ossia rendere il più visibile possibile la funzionalità di un oggetto in modo che sia più facile per gli utenti riconoscerlo e capirne la funzionalità. Tale norma può essere implementata tramite la forma dell'oggetto, che spesso suggerisce visivamente la funzionalità che esso permette (e.g. pulsante sul web che somigli a uno reale), oppure semplicemente tramite la mera organizzazione degli elementi visivi. Infatti, se un'informazione è fondamentale per la comprensione dell'intera infografica, essa dovrebbe essere sempre visibile e non nascosta da uno strato di interattività.
- **Feedback**, ciò significa che a ogni azione dell'utente si dovrebbe avere una reazione che indichi l'esito dell'operazione.
- **Vincoli**, ovvero devono essere posti dei vincoli in modo da orientare la navigazione dell'utente.
- **Uniformità**, ciò implica che entità analoghe dovrebbero somigliarsi.



### 3.3.3 Principi di Gestalt

La *Gestalt* (letteralmente “forma” o “schema” in tedesco) è una corrente psicologica che analizza e studia il funzionamento del cervello umano per cercare di prevedere come questo percepisce e organizza le informazioni visive. Nello specifico, i principi definiti da questa psicologia si basano sulla capacità del cervello di riconoscere schemi e classificare elementi visivi basandosi su differenze e somiglianze, anche prima che venga prestata attenzione consapevole ai dettagli specifici (si parla, infatti, di “percezione pre-attentiva”). Per tali motivi, incorporare questi concetti nella progettazione delle infografiche potrebbe migliorare l’efficacia comunicativa e, al contempo, rendere l’interfaccia grafica più intuitiva per l’utente.

Vediamo dunque come applicare le leggi di *Gestalt* al caso specifico del design di infografiche:

- **Legge di pregnanza**, siccome il cervello umano tende a elaborare i *pattern* regolari e ordinati più velocemente, si organizzino i dati logicamente e in maniera semplice. Si tenga inoltre conto che, sempre per questo principio, se si hanno linee o forme regolari con degli spazi, questi vengono mentalmente chiusi.
- **Legge di continuità** (di direzione), siccome il cervello umano tende a raggruppare oggetti vicini o allineati, si organizzino i dati di uno stesso gruppo concettuale vicini tra loro e con uno stesso allineamento. Tale configurazione permette di facilitare il processo di raggruppamento, comparazione e conseguente comprensione da parte dell’utente.
- **Legge di similarità**, siccome elementi con caratteristiche simili (e.g. colore, forma, dimensione, orientamento) vengono percepiti come un unico gruppo, si utilizzino uguali caratteristiche per stabilire relazioni tra i vari elementi.
- **Legge del punto focale**, simmetricamente alla legge precedente, si utilizzino caratteristiche diverse per aumentare il divario tra i vari gruppi ed evidenziare le parti più importanti, creando punti focali che attirino l’attenzione dell’utente.
- **Legge degli isomorfi**, siccome gli eventi e gli elementi sono interpretati dagli utenti in base alla loro esperienza passata, si tenga conto dei condizionamenti dati da cultura e dalle abitudini e convenzioni adottate.
- **Legge dell’immagine e dello sfondo**, siccome gli elementi vengono interpretati in maniera diversa in base alle relazioni con lo sfondo, si assicuri di avere un grande contrasto tra le informazioni e lo sfondo se le prime sono molto importanti, mentre si possono integrare dati nello sfondo se questi non sono particolarmente rilevanti.
- **Legge del fato o del destino comune**, siccome un gruppo di elementi simili è percepito come una struttura strettamente correlata, si utilizzino orientamenti, direzioni o movimenti per stabilire o interrompere le relazioni tra i componenti.

## 3.4 Come costruire un'infografica

### 3.4.1 L'Infomodel

Per garantire un prodotto finale di comunicazione efficace, la progettazione di un'infografica deve seguire un preciso schema, detto *Infomodel*<sup>4</sup>. Nello specifico, questo modello si compone di una serie ordinata di passaggi:

1. **Identificazione degli interlocutori;**
2. **Identificazione della storia e dei suoi effetti;**
3. **Identificazione del *dataset*;**
4. **Analisi dei dati;**
5. **Rappresentazione della storia.**

Si approfondiscono tali passaggi nei paragrafi successivi.

#### Identificazione degli interlocutori

L'obiettivo dell'infografica è rappresentare l'argomento secondo le preferenze degli interlocutori. Nello specifico, si vogliono presentare le informazioni rispecchiando le inclinazioni cognitive e decisionali della maggior parte del pubblico sul tema.

**Inclinazioni cognitive.** Quando parliamo di “inclinazioni cognitive” intendiamo, generalmente, le modalità di comprensione e di analisi dei fenomeni. Tali inclinazioni possono essere riassunte con il [Myers-Briggs Type Indicator \(MBTI\)](#), strumento che divide le preferenze degli utenti in quattro coppie di opposti. Tali coppie sono:

- **Estroversione o Introversione**, fornisce una preferenza riguardante il modo di trarre energia dell'individuo. Se ci si concentra più sul mondo esterno che quello interiore si tende al primo valore, in caso contrario al secondo.
- **Sensitività o Intuizione**, fornisce una preferenza riguardante la modalità di acquisizione delle informazioni. Se ci si concentra più sui fatti concreti che sul quadro generale e sulle possibili implicazioni si tende al primo valore, in caso contrario al secondo.
- **Ragionamento o Sentimento**, fornisce una preferenza sulle modalità di decisione e raggiungimento di conclusioni. Se si adotta un approccio più oggettivo che empatico si tende al primo valore, in caso contrario al secondo.
- **Giudicare o Percepire**, fornisce una preferenza sui modi di rapportarsi con il mondo esterno. Se non si è flessibili e non si è aperti a nuove informazioni si tende al primo valore, in caso contrario al secondo.

Ogni individuo tende a preferire uno dei due valori di ciascuna coppia. La combinazione di queste preferenze corrisponde a un tipo di personalità, il quale influenza l'approccio dell'individuo alla comprensione e all'analisi delle informazioni. Infatti, ciascun diverso tipo si comporta diversamente e ha differenti interessi, opinioni e motivazioni. Pertanto, la consapevolezza delle differenze tra i tipi può aiutare a strutturare l'infografica in modo più efficace.

---

<sup>4</sup>Paolo Bottazzini e Michele Gotuzzo. *Design della mente. Infografica e data visualization*. Press Grafica, Srl, 2014.

**Inclinazioni decisionali.** Quando parliamo invece di “inclinazioni decisionali” intendiamo le modalità di apprendimento con cui un individuo riesce ad arrivare più facilmente alla *saggezza*. Tali inclinazioni possono essere riassunte con il modello *4MAT* sviluppato da Bernice McCarthy, il quale individua quattro diversi stili di apprendimento basati sulle domande che gli individui si pongono riguardo al contenuto del materiale di studio. A ciascun stile è dunque facilmente associabile una modalità di narrazione che si concentra sul rispondere a tali domande. Questi stili sono:

#### 1. Apprendimento “innovativo”

- Principale interrogativo: “perché?”
- Tipo di individui: comprende tutte le persone che tendono a interrogarsi e riflettere sulla rilevanza e importanza concreta di un argomento prima di affrontarlo. Essi vogliono sapere il significato e i motivi per cui un fatto è accaduto.
- Tipo di narrazione: la spiegazione, in quanto racconta il significato e le ragioni.

#### 2. Apprendimento “analitico”

- Principale interrogativo: “cosa?”
- Tipo di individui: comprende tutte le persone che tendono a raccogliere fatti per approfondire la loro conoscenza dell’argomento. Essi vogliono dunque saperne il più possibile di un determinato concetto e analizzarlo in dettaglio.
- Tipo di narrazione: una descrizione più completa possibile della questione che fornisca tutti gli attori della storia.

#### 3. Apprendimento basato sul “buonsenso”

- Principale interrogativo: “come?”
- Tipo di individui: comprende tutte le persone che preferiscono sperimentare e mettere in pratica prima di formarsi una piena opinione in merito all’argomento. Essi vogliono dunque capire come possono utilizzare le informazioni per poter migliorare le proprie competenze.
- Tipo di narrazione: un racconto che svisceri il modo in cui si svolge il processo descritto, specificando requisiti, dettagli e attori in gioco.

#### 4. Apprendimento “dinamico”

- Principale interrogativo: “cosa succede se?”
- Tipo di individui: comprende tutte le persone il cui apprendimento è dato da una scoperta autonoma dettata dalla propria intuizione. Essi sono dunque interessati a capire le possibili conseguenze future e applicazioni pratiche delle diverse situazioni per adattarsi ai cambiamenti nell’ambiente o alle nuove possibilità che queste potrebbero comportare.
- Tipo di narrazione: una storia che esplora e incorpora tutte le opzioni e possibilità derivanti dalle scelte presentate.

Per un apprendimento accessibile a tutti, il *4MAT* prevede la combinazione di questi stili nell’ordine presentato sopra. Nello specifico, si dovrebbe iniziare con una spiegazione del perché l’argomento è rilevante e perché il pubblico dovrebbe essere interessato ad

ascoltarlo, seguito da una descrizione dettagliata e oggettiva dei dati e delle informazioni in merito, proseguendo con un'esplorazione dei risvolti pratici di quanto specificato e concludendo con una panoramica delle possibili conseguenze e cambiamenti futuri.

Tuttavia, per ottimizzare ulteriormente la comunicazione, è importante adattare la narrazione al tipo di pubblico specifico, approfondendo maggiormente la parte dettata dallo stile predominante tra gli interlocutori. Ciò consente di raggiungere una maggiore efficacia del messaggio e facilitarne una comprensione più profonda da parte del pubblico.

### Identificazione della storia e dei suoi effetti

**Formulare una storia.** Essenziale per una buona infografica è l'individuazione di una narrativa coinvolgente. A tal fine, si utilizza la struttura classica del racconto, ovvero un arco narrativo sviluppato in tre atti come segue:

- **Inizio:** presentazione dell'equilibrio iniziale e della sua rottura. Viene fornita una panoramica dello stato dell'arte dell'argomento e si mostrano gli agenti coinvolti. Successivamente si espongono il motivo o la complicazione che spinge verso il cambiamento, solitamente sotto forma di domanda.
- **Sviluppo:** esplorazione delle "peripezie" degli agenti coinvolti, ovvero vengono espone le azioni compiute da questi per cercare di rispondere alla domanda iniziale. Questo atto comporta dunque la crescita dell'azione, il suo climax e termina con la decrescita dell'azione stessa. È importante che questo processo non sia troppo affrettato: se la risposta alla domanda iniziale è fornita troppo presto, la storia rischia di diventare noiosa o banale.
- **Conclusione:** ripristino dell'equilibrio attraverso le azioni svolte dagli agenti, si torna dunque a una nuova stabilità che costituisce l'epilogo della storia.

Da tale descrizione, possiamo dunque identificare i principali elementi di una storia in agenti e azioni. Gli agenti sono i soggetti o le forze che influiscono o vengono influenzati dai dati, mentre le azioni sono gli eventi o le decisioni che modellano la storia stessa. Si noti, inoltre, che quest'ultime devono seguire i seguenti criteri:

- trasformare gli agenti o la situazione, ogni azione deve dunque generare un **cambiamento**;
- trasmettere un significato, ogni azione è dunque collegata a un **tema**;
- basarsi su dettagli rilevanti, ogni azione è dunque compiuta con **coerenza**;
- seguire delle regole ed essere credibili, ogni azione deve dunque rispettare il principio di **plausibilità**.

**Gli effetti della storia.** È cruciale considerare non solo l'esperienza utente dell'infografica al momento dell'interazione, ma anche come tale esperienza sarà ricordata dall'utente in futuro. Diventa dunque necessario empatizzare con i valori e le aspirazioni del pubblico per poter realizzare un design che provochi reazioni istantanee, siano esse viscerali (come risposta a colori, materiali etc.) o fisiche (e.g. cliccare un pulsante), e che al contempo innesti memorie e associazioni permanenti relative all'infografica.

Dal punto di vista pratico, tale processo può essere realizzato creando degli archetipi di utenti che aiutino a comprendere come persone con diversi desideri e abilità potrebbero approcciarsi all'infografica. A tal fine, si noti che è particolarmente importante ricostruire quale siano storia, emozioni, risorse e obiettivi di queste “persone” in modo tale da poter capire il perché di alcune loro azioni. Meno interessante, invece, è individuare quale siano le singole azioni effettuate.

### Identificazione del dataset

Si scelgono dei *dataset* in base ai punti precedenti e facendo attenzione all'autorevolezza delle fonti e alla rappresentatività del campione. Successivamente, è necessario pulire i dati per renderli coerenti con le convenzioni adottate globalmente e per sostituire eventuali incoerenze di notazione con le corrette.

A seconda dei dati disponibili scelti, si possono rivelare e mettere in luce diversi aspetti dei dati stessi. Tali aspetti identificano gli obiettivi di visualizzazione, come definiti al paragrafo [Obiettivi della visualizzazione](#).

### Analisi dei dati e rappresentazione della storia

Per comprendere appieno le implicazioni tecniche delle scelte fatte nei punti precedenti e poter rappresentare efficacemente la storia e i dati scelti, è essenziale analizzare il *dataset* con attenzione, considerando la qualità, rappresentatività e compatibilità reciproca dei dati.

Innanzitutto, è necessario esaminare approfonditamente il dataset ed eseguire delle visualizzazioni preliminari, annotandosi le prime impressioni. Ciò può aiutare a identificare potenziali *bias* e a ottenere una comprensione iniziale del *dataset*. Successivamente, è necessario cercare e identificare dei *pattern* nei dati, modificando la visualizzazione di conseguenza. Si vogliono dunque mettere in luce tali regolarità, utilizzando ad esempio tecniche di zoom, aggregazioni, filtri o riduzione del rumore, facendo attenzione a mantenere l'integrità dei dati. Infine, è cruciale assicurarsi che l'analisi effettuata sia orientata verso il pubblico giusto, in modo tale da garantire che le informazioni siano rilevanti e comprensibili per gli utenti.

Una volta fatto ciò, è possibile combinare le varie visualizzazioni e aggiungere ulteriori elementi di supporto. In tale processo, vengono applicate alcune tecniche narrative al fine di comunicare efficacemente i dati attraverso una rappresentazione coinvolgente della storia.

Tali tecniche comprendono, ad esempio, metafore narrative e strutture retoriche, le quali, grazie al loro impatto emotivo e cognitivo, facilitano la creazione di una narrazione suggestiva e persuasiva.

## 3.4.2 Principi e linee guida

### Linee guida generali

Alcuni studi<sup>5</sup> identificano dei principi che le infografiche, come strumenti di visualizzazione delle informazioni, dovrebbero seguire:

- **Mostrare i dati in modo coerente e favorirne il confronto;**

---

<sup>5</sup>Tufte, *The Visual Display of Quantitative Information, Second Edition*.

- **Focalizzarsi sul contenuto** piuttosto che sulla metodologia grafica con cui questo è rappresentato;
- **Evitare di distorcere i dati** per confermare le proprie tesi;
- **Sviluppare la narrazione su diversi livelli di dettaglio**, per permettere un approfondimento di questa solo se il lettore lo desidera;
- **Seguire uno scopo**, che deve essere riflesso sia a livello informativo che decorativo;
- **Integrare descrizioni statistiche e verbali** per avere un quadro completo del contesto nel quale i dati presentati si inseriscono.

Grazie al soddisfacimento di tali requisiti è possibile creare uno strumento di comunicazione che permetta di esporre efficientemente ed efficacemente idee complesse in maniera chiara e semplice.

### Decidere la complessità dell'infografica

Quando si sviluppa un'infografica è fondamentale adattarne la complessità al lettore medio. A tal fine, alcuni studi<sup>6</sup> identificano una serie di coppie di caratteristiche (descritte nella forma [*più complesso, ma approfondito* - *più semplice, ma superficiale*]) da bilanciare per trovare un equilibrio di complessità:

- **Astrazione - Raffigurazione**: si ha un'infografica figurativa quando si utilizzano immagini o fotografie che mostrano il soggetto in modo dettagliato e mimetico. Se, invece, si utilizzano pittogrammi o simboli, che richiedono dunque un'ulteriore elaborazione per essere interpretati, si avrà un'infografica più vicina al lato dell'astrazione e dunque più complessa.
- **Funzionalità - Decorazione**: si ha un'infografica più decorativa se si hanno molti elementi visivi non direttamente utili alla comprensione dei dati, come decorazioni artistiche e ornamenti, i quali riducono la complessità informativa generale dell'infografica.
- **Densità - Leggerezza**: si ha una maggiore densità o leggerezza in base allo spazio utilizzato per rappresentare i dati, naturalmente più spazio è utilizzato maggiore sarà la complessità informativa in quanto, semplicemente, sono mostrate più informazioni.
- **Multidimensionalità - Unidimensionalità**: queste caratteristiche riguardano i livelli di approfondimento di un'infografica. Avendo un solo livello (unidimensionalità), l'infografica sarà più superficiale e dunque più semplice e veloce da comprendere.
- **Originalità - Familiarità**: si hanno infografiche tendenti alla familiarità se usano forme grafiche comuni (e.g. grafico a torta), le quali essendo già note al lettore possono essere comprese più facilmente.
- **Novità - Ridondanza**: si hanno infografiche ridondanti se i concetti vengono spiegati e ripetuti più volte attraverso diverse metodologie, consentendo ai lettori di chiarire eventuali incomprensioni.

---

<sup>6</sup>Alberto Cairo. *L'arte funzionale Infografica e visualizzazione delle informazioni*. Pearson Italia, SpA, 2013.

Nello specifico, tali caratteristiche sono da valutare nel seguente modo:

1. Innanzitutto, si pensa a *densità - leggerezza* e *multidimensionalità - unidimensionalità*, cercando di non sottostimare la capacità di comprensione dei lettori. Nello specifico, si organizza l'infografica in più livelli sistemati in ordine logico, includendo solo i sottostrati necessari per raggiungere l'obiettivo narrativo preposto. Il livello più esterno deve fungere da punto di ingresso dell'infografica, fornendo una sintesi chiara e concisa delle informazioni presentate, comprese introduzione e punti salienti.
2. Una volta determinate tali caratteristiche, si pensa a *funzionalità - decorazione* e *astrazione - raffigurazione*, definendo dapprima struttura e organizzazione dei dati e successivamente la loro rappresentazione. Solo in un secondo momento si scelgono eventuali elementi decorativi aggiuntivi.
3. Infine, si pensa a *originalità - familiarità* e *novità - ridondanza*, cercando di garantire la comprensione dei concetti esposti ma anche di evitare la noia a cui la ripetizione di questi potrebbe portare. In linea generale, meno comune è la forma grafica scelta più ridondanze dovranno essere inserite.

## Capitolo 4

# Applicazione pratica di infografica web con D3.js

*In questo capitolo verrà presentato il processo di sviluppo - comprendente progettazione, codifica e validazione - di un'infografica web sviluppata durante il corso dello stage. Tale infografica ha lo scopo di presentare il corso di laurea triennale di Informatica dell'Università degli Studi di Padova.*

### 4.1 Progettazione dell'infografica

La scelta di realizzare un'infografica sul corso di laurea triennale in Informatica dell'Università degli Studi di Padova è stata motivata dall'esperienza personale dello stagista come studente del corso, prossimo alla sua conclusione. Questo *background* ha consentito di trattare il tema con l'*ethos* e l'autorevolezza necessari per creare un'infografica informativa e credibile. Una solida conoscenza e competenza sull'argomento della presentazione sono, infatti, essenziali per garantire la qualità e l'accuratezza delle informazioni esposte.

Inoltre, la disponibilità di dati pubblici e di fonti autorevoli, come i report dell'Università stessa e le statistiche di Almalaurea (il più importante consorzio interuniversitario italiano), ha garantito anche l'affidabilità del contenuto, contribuendo alla scelta di questo argomento.

Per quanto riguarda il processo di progettazione dell'infografica, questo ha seguito gli *step* delineati dall'*Infomodel*, come descritto nel capitolo precedente. Si riportano di seguito le varie scelte progettuali effettuate per ciascuna fase del processo.

#### 4.1.1 Identificazione degli interlocutori

##### Target

Il target di riferimento dell'infografica sono principalmente gli individui che desiderano proseguire la loro formazione con un percorso universitario e sono interessati in particolare all'ambito informatico. Inoltre, l'infografica potrebbe anche attrarre studenti che già frequentano il corso, i quali potrebbero essere interessati a visualizzare e analizzare le statistiche relative al loro programma di studi.



### Scelte individuate a partire dalle inclinazioni cognitive

Per quanto riguarda le loro inclinazioni cognitive, possiamo ragionevolmente supporre che la maggior parte degli interlocutori tendano maggiormente al *ragionamento* piuttosto che al *sentimento*. Questo poiché il campo dell'informatica richiede una forte capacità di analisi logica e *problem solving*, elementi che sono tipicamente associati al ragionamento analitico.

Alla stessa maniera, è probabile che i futuri studenti di informatica siano più inclini al *percepire* che al *giudicare*, essendo flessibilità e apertura all'innovazione fondamentali per questo settore in continua evoluzione.

Per quanto riguarda invece *sensitività o intuizione*, è probabile che la maggior parte degli interlocutori tenda verso il secondo, in quanto avere una visione globale e una capacità di immaginare le possibili conseguenze risultano essenziali per lo sviluppo di programmi informatici.

L'infografica è quindi progettata tenendo conto di queste caratteristiche. Nello specifico, essa fornirà una panoramica generale del corso di laurea, illustrando anche le potenziali conseguenze e benefici (*intuizione*). Si cercherà, inoltre, di fornire molti dati oggettivi (*ragionamento*), pur inserendo elementi di *pathos*, necessari per connettersi emotivamente con gli utenti. Infine, verranno inclusi elementi innovativi nella presentazione per stimolare l'interesse e l'*engagement* degli interlocutori (*percepire*).

### Scelte individuate a partire dalle inclinazioni decisionali

Per quanto riguarda invece le inclinazioni decisionali degli interlocutori, si considerano tutti gli stili di apprendimento, facendo però una particolare attenzione all'*apprendimento analitico*. Si suppone infatti che la maggior parte degli individui interessati all'informatica sia incline a tale approccio, che predilige la scomposizione delle informazioni in parti più gestibili e l'uso di ragionamenti logici per risolvere problemi complessi. Questo tipo di procedimento è infatti alla base del funzionamento di molti algoritmi informatici.

Si struttura la storia rispondendo alle seguenti domande:

- **Perché?**

- Si inserisce una spiegazione iniziale che illustra l'importanza e rilevanza del corso e i motivi per cui sceglierlo.

- **Cosa?**

- Si dedica una grossa fetta dell'infografica a descrivere il corso, fornendo tutti i dati possibili per averne un quadro generale.

- **Come?**

- Si inseriscono le opinioni degli studenti laureati all'interno dell'infografica, in modo tale da avere anche una prospettiva più pratica fornita da chi ha vissuto in prima persona l'esperienza del corso.

- **Cosa succede se?**

- Si esplorano le opportunità e opzioni sia accademiche sia professionali che il corso apre al suo completamento.

### 4.1.2 Identificazione della storia e dei suoi effetti

#### Formulazione della storia

La storia viene formulata a partire dalle risposte alle domande definite nella sezione precedente, a formare un arco narrativo di tre parti come segue:

- **Introduzione al corso:**
  - La prima parte dell'infografica fornisce una panoramica generale del corso. Viene illustrata brevemente l'importanza di esso, indicando le sue principali aree di studio e i motivi per cui potrebbe essere una scelta valida.
  - Questa parte risponde alla domanda “perché?”.
- **Descrizione del corso:**
  - La seconda parte esplora i vari aspetti del corso in modo più approfondito. Vengono forniti dettagli sulla durata del corso, il voto medio degli studenti, i requisiti necessari, i luoghi di studio e le materie ed esperienze pratiche offerte dal corso. Inoltre, si delineano anche i profili tipici degli iscritti. Si include anche la prospettiva degli studenti laureati, che forniscono la loro opinione sui vari argomenti trattati.
  - Questa parte risponde alle domande “cosa?” e “come?”.
- **Valutazione finale e opportunità post-laurea:**
  - L'ultima parte dell'infografica analizza se il corso rappresenti effettivamente una scelta valida, esaminando le opportunità professionali e accademiche che si presentano dopo il completamento e fornendo l'opinione finale dei laureati sul corso.
  - Questa parte risponde alla domanda “cosa succede se?”.

Nello specifico, si ha la seguente storia:

- **Primo atto dell'arco narrativo: introduzione al corso**
  - ***Il corso in breve***  
*Il corso di laurea fa parte del dipartimento di matematica ‘Tullio-Levi-Cita’ dell’Università degli Studi di Padova e dura 3 anni.*  
*Il corso di laurea consente di ottenere una competenza e metodologia di lavoro essenziale per adattarsi a diverse situazioni sia presenti che future. Ciò è molto importante, specie nel campo dell’informatica che, come sappiamo, è continuamente stravolto da nuove scoperte.*  
*Ciò è possibile grazie ad un corso di laurea che offre una solida base teorica e pratica, combinando l’apprendimento di nozioni e fondamenti classici di matematica e informatica con corsi di taglio progettuale e stage obbligatorio. I contenuti del corso di laurea hanno la certificazione di qualità concessa dal GRIN, ente nazionale che si occupa di vagliare, organizzare e promuovere attività scientifiche e didattiche informatiche. Inoltre, l’università di Padova ottiene buoni ranking (4a) nell’ambiente nazionale anche secondo il QS.*
- **Secondo atto dell'arco narrativo: descrizione del corso**

– **Cosa ti serve per iniziare?**

- \* *Diploma di maturità quinquennale (o equipollenti), non è necessario aver studiato materie specifiche precedentemente, si partirà dalle basi.*
- \* *Superare il test d'ingresso TOLC-I in quanto il corso ha accesso a numero programmato.*
- \* *Avere la giusta attitudine: un'attenzione all'innovazione, un interesse per la tecnologia, una curiosità scientifica, un pensiero astratto e logico, un'inclinazione al problem solving.*

– **Chi ti accompagnerà?**

*La maggior parte degli studenti sono dei ragazzi giovani, di circa 19 anni, provenienti da istituti tecnici o licei scientifici. Tuttavia, se sei una ragazza, se non sei più così giovane, o se la tua formazione pre-universitaria non segue il percorso tradizionale, non preoccuparti! Il corso accoglie studenti con storie diverse, offrendo opportunità di successo a tutti coloro che sono motivati e appassionati.*

– **Dove si va?**

*L'università dispone di una vasta area, con quasi 800.000 ettari di proprietà che include sia edifici moderni che con secoli di storia. Tuttavia, tu non dovrai correre da una parte all'altra della città per arrivare a lezione. Il corso, infatti, si svolge a Padova, all'interno di poche aule situate nella zona del Piovego.*

– **Cosa imparerai?**

*Dei 180 CFU totali, 48 sono dedicati alla parte matematica, 102 all'informatica e 12 allo stage (i rimanenti per tesi, abilità linguistica di inglese ed esami a scelta).*

*Per alcuni ambiti quali programmazione a oggetti, basi di dati, tecnologie web e ingegneria del software vengono svolti progetti per sperimentare e toccare con mano quanto appreso. Nell'ultimo caso, vi è anche la collaborazione con aziende esterne, che propongono i progetti, rendendo l'esperienza ancora più vicina alla realtà del campo professionale.*

*Anche lo stage inoltre copre un ruolo fondamentale nel corso essendo poi su quello che si baserà la tesi di laurea.*

*Per quanto riguarda il carico di studio, esso viene ritenuto adeguato da quasi il 90% degli studenti.*

– **Ma è difficile?**

*La maggior parte degli studenti si laurea entro i tre anni prestabiliti; tuttavia, la durata degli studi è mediamente 4.3 anni. Per quanto riguarda il voto medio questo è 96.8.*

*Quindi potremmo considerare il corso generalmente abbastanza difficile.*

• **Terzo atto dell'arco narrativo: valutazione finale e opportunità post-laurea**

– **Finalmente laureati! E poi?**

*Si può proseguire con la magistrale, della durata di due anni, per arricchire la propria formazione ed acquisire competenze specifiche.*

*Alternativamente o parallelamente, si può iniziare a lavorare. I principali sbocchi lavorativi riguardano lo sviluppo di app software e la gestione di reti informatiche. Non ci si deve preoccupare di non trovare lavoro! Infatti, il*

*tasso di disoccupazione è basso: si aggira intorno al 3.9%.*

*La maggior parte dei laureati sceglie quest'ultima strada, perseguendo professioni tecniche o intellettuali. Nel caso, invece, di proseguimento degli studi con la laurea magistrale, solitamente il corso scelto rappresenta il proseguimento naturale di quanto studiato, confermando la scelta effettuata con la triennale.*

– **Quindi, ne vale la pena?**

*Circa il 90% degli studenti si ritiene soddisfatto abbastanza del corso, nello specifico il 46% si dichiara decisamente entusiasta.*

*Complessivamente l'80% si iscriverebbe nuovamente al corso.*

La storia è stata anche data in input a *Infographic-helper* il quale ha fornito elementi di *pathos*, *ethos* e *logos* da integrare. Ciò ha portato al contenuto presentato sopra. Lo strumento ha inoltre permesso la suddivisione in sezioni, sopra anticipate da un titolo in grassetto per una maggiore visibilità. Per quanto riguarda l'obiettivo, *Infographic-helper*, ha classificato la storia come “informazioni correlate”; tuttavia, si può individuare un “processo ordinato” che è l'obiettivo di visualizzazione che viene effettivamente perseguito.

### 4.1.3 Identificazione del dataset

I dati sono stati presi dai report dell'università, disponibili a [questo link](#) e dai dati forniti da Almalaurea, disponibili a [questo link](#). I dati considerati sono quelli relativi all'anno più recente disponibile (2022 e/o 2023), per assicurare che le informazioni siano aggiornate e riflettano la situazione attuale.

Altri dati, riguardanti ad esempio le materie, sono stati ricavati manualmente dalla pagina del corso di laurea (disponibile a [questo link](#)).

Per quanto riguarda i luoghi di studio, sono stati anch'essi ricavati manualmente a partire dall'esperienza personale dello stagista-studente e dei colleghi, cercando di inserire tutte le strutture utili presenti nelle vicinanze.

In base ai diversi *dataset* disponibili e grazie allo strumento *Chart-chooser*, si identificano i diversi aspetti di questi che si vogliono mettere in luce con la visualizzazione:

- *Composizione* per quanto riguarda i dati sulle materie di studio, le opportunità post-laurea e alcune delle opinioni degli studenti;
- *Flussi* per quanto riguarda i profili degli studenti;
- *Grandezze* per quanto riguarda i dati sulla difficoltà del corso;
- *Divergenza* per quanto riguarda la maggior parte dei dati sull'opinione degli studenti;
- *Spazi* per quanto riguarda i luoghi.

### 4.1.4 Analisi dei dati e rappresentazione della storia

#### Analisi dei dati

Per quanto riguarda l'analisi dei dati, essendo questi provenienti da fonti autorevoli, possono essere considerati di qualità. Inoltre, il numero limitato di enti distributori di tali fonti e l'elaborazione dei dati che loro stessi forniscono ha permesso di non

avere particolari problemi di notazione. È stato, infatti, necessario solo un *parsing* dei dati: per quanto riguarda i dati dell'Università di Padova, sono state escluse le informazioni non pertinenti al corso di laurea; mentre, per i dati di Almalaurea, è stato necessario strutturare i dati in sezioni, domande e relative risposte. Questo ha permesso di ottenere dati utili e pertinenti alla rappresentazione.

Ulteriori analisi effettuate utili per il design dell'infografica sono riportate successivamente per i vari fattori di interesse.

### Scelte progettuali che influenzano la rappresentazione

Si riportano di seguito le scelte progettuali effettuate per ciascuno dei fattori che influenzano il design dell'infografica e, di conseguenza, la rappresentazione della storia.

**I Colori.** I colori principali scelti per l'infografica sono:

- **Rosso scuro e sue sfumature**, in quanto colore del logo dell'Università degli Studi di Padova. Ciò permette di mantenere una certa coerenza visiva con il tema presentato e, allo stesso tempo, attrarre l'attenzione del pubblico. Inoltre, le sfumature più tendenti all'arancione aggiungono un tocco giovanile e dinamico, riflettendo il carattere innovativo del settore dell'informatica.
- **Gradazioni di grigio**, utilizzato in maniera simmetrica ai rossi sui grafici per mostrare divergenze oppure per rappresentare altri elementi che si scostano dall'obiettivo principale della singola visualizzazione. Le gradazioni di grigio simboleggiano tecnologia, quindi perfettamente in tema, e contemporaneamente offrono anche un contrasto neutro che bilancia l'uso del colore principale.
- **Tonalità di bianco e nero**, utilizzate per gli sfondi e per il colore del testo. Si precisa che la scelta del colore dei vari blocchi di testo è stato scelto in maniera tale da rispettare lo standard AA definito dal [Web Content Accessibility Guidelines \(WCAG\)](#).

**Il layout.** Si è utilizzato lo strumento *Infographic-helper* per individuare le varie sezioni della storia, come descritto alla sezione [Formulazione della storia](#), e l'obiettivo principale. Tale obiettivo della presentazione corrisponde a "infografiche basate su un processo ordinato". Seguendo i principi enunciati nel capitolo precedente, si sceglie una *disposizione a passi*. Nello specifico, si adotta il *pattern VIF spiral*, che consente di implementare uno scorrimento verticale, poiché il *layout* è sviluppato anch'esso verticalmente. Tale tipo di scorrimento risulta essere più agevole e familiare agli utenti di quello orizzontale, essendo la convenzione nella maggior parte dei siti web. È importante notare che l'infografica richiede necessariamente uno scorrimento per essere visualizzata nella sua interezza poiché contiene numerose informazioni.

**La grandezza degli elementi.** La grandezza degli elementi è influenzata principalmente dalle inclinazioni cognitive e decisionali del pubblico target. Infatti, per rispondere alla predominanza di ragionamento e apprendimento analitico, si attribuisce maggiore importanza ai dati oggettivi, rappresentandoli attraverso grafici di dimensioni maggiori. Al contrario, le opinioni degli studenti, che si basano più sulla pratica e sull'esperienza soggettiva, sono rappresentate con dimensioni minori e possono essere visualizzate in modo opzionale.

**Le tecniche di visualizzazione dei dati inserite.** Si è utilizzato lo strumento *Chart-chooser* per individuare le tecniche di [Data Visualization](#) più adatte. Le visualizzazioni risultanti scelte sono le seguenti:

- Per la visualizzazione dei dati riguardanti il profilo degli studenti si è scelto di utilizzare un diagramma Sankey.
- Per la visualizzazione dei dati geografici si è utilizzata una semplice mappa con evidenziati i luoghi d'interesse.
- Per la visualizzazione dei dati riguardanti le materie di studio si è scelto di utilizzare un *treemap*.
- Per la visualizzazione dei dati riguardanti la regolarità (intesa come durata) negli studi si è scelto di utilizzare un semplice grafico a barre.
- Per la visualizzazione dei dati riguardanti le opportunità post-laurea si è utilizzato un diagramma di Venn, corredato da dei *waffle chart*.
- Per la visualizzazione dei dati riguardanti l'opinione degli studenti si è scelto di utilizzare degli *diverging stacked bar charts* per opinioni divergenti, altrimenti dei *waffle chart* per opinioni non classificabili solamente in positive e negative.

Si precisa che, essendo lo strumento prototipale e non compatibile con il formato dei dati, il numero di variabili è stato inserito manualmente nel codice per ottenere dei risultati corretti.

**La simmetria del design.** Per la struttura generale dell'infografica si è impiegata la *symmetrical balance*. Nello specifico, il *layout* utilizzato ha consentito di allineare gli elementi sia a sinistra che a destra in modo equilibrato. Questa scelta di design è stata fatta per rappresentare al meglio la storia dell'infografica (tramite il *layout* scelto) e ottenere una distribuzione visiva armoniosa e ordinata, che crei un senso di equilibrio e coerenza per una maggiore facilità nell'interpretazione delle informazioni.

**Gli elementi grafici.** Si è scelto di inserire come elementi grafici delle icone piuttosto che immagini realistiche in quanto più adatte a rappresentare concetti astratti e categorie, che costituiscono la maggior parte delle informazioni da visualizzare. Per la parte rimanente non astratta e non categorica si è comunque scelto di usare icone per coerenza con il resto dell'infografica.

Si è inoltre prestata particolare attenzione alla risoluzione di tali elementi grafici, al fine di garantire una corretta visione su tutti i tipi di dispositivo. Infatti, si usa nella stragrande maggioranza dei casi il formato [Scalable Vector Graphics \(SVG\)](#), che permette di scalare l'elemento senza perderne la qualità.

**I blocchi di testo.** I blocchi di testo "standard" sono inseriti nell'infografica come segue:

- **I titoli:** è presente un titolo generale dell'infografica e uno per ogni sezione. Il titolo generale introduce semplicemente il tema; per quanto riguarda invece i titoli delle sezioni, questi sono formulati sotto forma di domanda, in modo tale da generare interesse nell'utente.

- **I sottotitoli:** sono presenti solo per le sezioni e non per l'infografica nel suo insieme. In particolare, essi specificano il contenuto della sezione rispondendo brevemente alla domanda del titolo con la statistica principale.
- **L'introduzione:** è presente all'inizio dell'infografica e fornisce una panoramica generale del corso di laurea presentato.
- **Il testo principale:** esso viene diviso nelle varie sezioni ed è visualizzato in multipli contenitori separati per avere una visione più strutturata dei vari dati.
- **Le note a piè di pagina:** servono a indicare le fonti da cui vengono presi i dati.

Oltre a questi blocchi, è presente del testo anche in:

- **Note sul testo e sui grafici:** sono presenti dei *pop-up* e *tooltip* visibili selezionando o posizionandosi col cursore in alcune parti del testo e dei grafici. Essi forniscono del testo informativo sull'elemento in questione.
- **Chat:** è presente un *chatbot* dove si può inviare e ricevere blocchi di testo per avere informazioni sull'infografica e orientarsi all'interno di essa. Tali testi sono inseriti in contenitori stile "messaggio".

In generale, il testo presente nell'infografica è molto conciso. La priorità è, infatti, la visualizzazione diretta dei dati, con opzioni di approfondimento testuali disponibili su richiesta. Ciò viene fatto per evitare un eccessivo rumore visivo e mantenere l'efficacia comunicativa.

**I caratteri tipografici.** Si utilizzano caratteri *sans-serif*, essendo un'infografica web e composta da poco testo. La scelta degli specifici caratteri è stata dettata da preferenze stilistiche dello stagista, tuttavia vengono comunque inserite delle opzioni secondarie compatibili con tutti i browser in modo tale da garantire l'accessibilità dell'infografica.

Per mettere in luce alcuni dati, sono stati utilizzati caratteri più spessi, talvolta accompagnati da diversi colori e dimensioni per evidenziarne ulteriormente l'importanza.

Per quanto riguarda invece l'allineamento del testo, questo è generalmente centrato per testi inseriti in contenitori specifici, altrimenti l'allineamento segue la struttura del *layout* al fine di mantenere un'armonia e un equilibrio visivo complessivi.

### Interattività

Per quanto riguarda gli elementi interattivi presenti all'interno dell'infografica, si hanno:

- **Panning e zoom,** inseriti sui grafici per aumentarne la visibilità su vari dispositivi;
- **Apertura e chiusura di *pop-up* e *tooltip*,** utilizzati per riportare dettagli e fornire un ulteriore livello di approfondimento all'informazione;
- **Ricerca,** implementata attraverso un *chatbot* che riporta alla sezione dell'infografica in cui si discute dell'argomento richiesto;
- **Filtro,** utilizzato per mostrare le materie a seconda dell'affinità (a temi informatici, matematici o altro) oppure a seconda dell'anno accademico in cui vengono offerte;
- **Piccole animazioni grafiche,** utilizzate in alcuni grafici per migliorarne la comprensione, attirando l'attenzione sugli elementi chiave.

### Complessità dell'infografica

Per quanto riguarda *densità* - *leggerezza*, l'infografica tende verso una maggiore complessità, sono presenti infatti numerose informazioni. Dal punto della *multidimensionalità* - *unidimensionalità*, l'infografica è organizzata su un massimo di due livelli: uno generale che mostra visivamente i dati e uno secondario che ne specifica i valori. Pertanto, si ha una struttura abbastanza semplice.

Per quanto concerne *astrazione* - *raffigurazione*, l'infografica contiene solamente icone e nessuna immagine realistica. Inoltre, la maggior parte degli elementi inseriti sono utili per comprendere l'infografica, rendendola più *funzionale* che *decorativa*.

Per quanto riguarda invece *originalità* - *familiarità*, l'infografica combina e bilancia elementi non familiari, come ad esempio il diagramma Sankey, con elementi ad uso comune, come ad esempio il diagramma di Venn. Infine, in termini di *novità* - *ridondanza*, viene inserito un unico elemento di ridondanza, ovvero il "secondo livello" che fornisce informazioni testuali sulla parte visiva della presentazione.

Tale configurazione di caratteristiche crea un'infografica abbastanza complicata. Tuttavia, si ritiene che il target di riferimento possa essere in grado di comprenderla agevolmente, essendo potenziali futuri studenti universitari e dunque inclini a gestire informazioni complesse.

Inoltre, per una maggiore facilità di comprensione e navigazione, viene incluso uno strumento - il *chatbot* - che consente di ricevere risposta ai propri dubbi e/o trovare l'informazione ricercata all'interno dell'infografica.

## 4.2 Implementazione e uso dell'infografica

Innanzitutto, è necessario premettere che sono disponibili due versioni dell'infografica, differenziate solamente dall'algoritmo che caratterizza il *chatbot*.

- La prima versione, da qui in poi chiamata **versione base**, è stata sviluppata dallo stagista e implementa l'algoritmo solamente attraverso codice [JS](#).
- La seconda versione, da qui in poi chiamata **versione integrata**, implementa una ricerca ibrida attraverso i linguaggi [Python](#) e [Structured Query Language \(SQL\)](#). Tale versione è così chiamata in quanto integra in sé il codice prodotto dal collega Fabio Meneghini come parte della sua tesi.

Per una descrizione dettagliata delle differenze tra le due versioni si rimanda alla sezione apposita [Chatbot](#).

Per una visione completa dell'infografica e per esplorarne le funzionalità interattive, è possibile accedere al prodotto finale attraverso [questo link](#) per la versione base o [quest'altro link](#) per la versione integrata.

### 4.2.1 Strumenti e tecnologie utilizzati

#### Tutte le versioni

**HTML + CSS + JS.** Il progetto di stage utilizza [HTML](#), [CSS](#) e [JS](#) per costruire e stilizzare l'infografica. Nello specifico, si usa [HTML](#) per definire la struttura della pagina web, [CSS](#) per controllarne l'aspetto visivo e [JS](#) per aggiungerci interattività. Inoltre, attraverso [JS](#) vengono integrati moduli e librerie per la [Data Visualization](#) e, nel caso della versione base, si utilizza tale linguaggio anche per gestire le richieste del *chatbot*.



**D3.js.** Il progetto di stage applica concretamente quanto ricercato utilizzando la libreria **JS D3.js**, che riveste dunque un ruolo fondamentale nello sviluppo. Tutte le rappresentazioni di **Data Visualization** sono infatti implementate attraverso tale libreria. **D3.js** consente di creare grafici e visualizzazioni scalabili, in formato **SVG**, e interattive, oltre che altamente personalizzabili. Ciò permette di rendere maggiormente coinvolgente e dinamica l'esperienza utente. Inoltre, questa libreria fornisce anche dei metodi per il *parsing* dei dati, permettendo dunque di manipolarli con maggiore facilità e precisione.

Per la creazione del diagramma Sankey, inserito nell'infografica per rappresentare il profilo degli studenti, è stato necessario utilizzare l'estensione di **D3.js d3-sankey**. Tale *plugin* permette di creare diagrammi Sankey interattivi e proporzionali in maniera efficiente e precisa, migliorando la comprensione dei dati.

Si è inoltre utilizzato il modulo **d3-legend**, che estende **D3.js** per la creazione di legende personalizzate e interattive. L'utilizzo di tale modulo ha consentito di creare le legende più velocemente e con maggiore precisione, migliorando la chiarezza e l'interpretazione dei dati visualizzati.

La libreria è disponibile al link <https://d3js.org/>, mentre la documentazione di d3-sankey è disponibile a <https://github.com/d3/d3-sankey> e del modulo d3-legend al link <https://d3-legend.susielu.com/>.

**Venn.js.** Per la creazione del diagramma di Venn che rappresenta le strade intraprese (lavoro o proseguimento della formazione) dai laureati alla fine del corso, si è utilizzata questa libreria. Venn.js è una libreria **JS**, basata su **D3.js**, che è progettata proprio per la creazione di diagrammi di Venn. In particolare, consente di visualizzare le relazioni e le intersezioni tra insiemi e renderle interattive, permettendo agli utenti di esplorare le sovrapposizioni e le esclusioni tra i gruppi di dati in modo dinamico. Inoltre, la libreria garantisce che le dimensioni delle aree nel diagramma siano proporzionali ai valori dei dati, migliorando l'interpretazione delle informazioni.

La libreria è disponibile al link <https://github.com/benfred/venn.js>.

**Leaflet.** Per la creazione e visualizzazione della mappa delle strutture dell'Università, è stata utilizzata la libreria Leaflet in combinazione con **D3.js**. Leaflet è una libreria **JS open-source** progettata proprio per gestire e visualizzare dati geospaziali su mappe interattive. Questa combinazione delle due tecnologie consente di visualizzare mappe dettagliate e dinamiche, sfruttando la capacità di Leaflet di gestire elementi "mappe" e la potenza di **D3.js** per creare visualizzazioni personalizzate.

Si è inoltre impiegato il modulo **OverlappingMarkerSpiderfier** di Leaflet per gestire la sovrapposizione dei *marker* (i.e. le puntine) sulla mappa, consentendo una visualizzazione più chiara e completa dei punti di interesse.

La libreria è disponibile a <https://leafletjs.com/>, mentre la documentazione del modulo OverlappingMarkerSpiderfier è disponibile al seguente link <https://github.com/jawj/OverlappingMarkerSpiderfier-Leaflet>.

Sono inoltre necessarie le seguenti tecnologie, diverse in base alla versione scelta.

#### Versione base

Si utilizzano:

- Node.js

- Framework **JS** multiplatforma e *open-source* tra i più utilizzati. In particolare, si tratta di un ambiente di esecuzione che permette di eseguire codice **JS** come un qualsiasi linguaggio di programmazione, al di fuori dei browser. Per l'infografica, in particolare, è stato utilizzato il modulo di Node.js **http-server**, che consente di servire file statici tramite un server HTTP. Questo strumento è fondamentale per visualizzare l'infografica nel browser, permettendo così l'esecuzione e la visualizzazione delle rappresentazioni create con **D3.js**.
- Node.js è disponibile al link <https://nodejs.org/>, mentre per maggiori dettagli su http-server si visiti <https://www.npmjs.com/package/http-server>.

### Versione integrata

Si utilizzano:

- **Python:**
  - Linguaggio di programmazione ad alto livello, molto versatile e ampiamente utilizzato. Viene adoperato per elaborare i dati e interfacciarsi con il *database* PostgreSQL per rispondere alle richieste dell'utente fatte via *chatbot*.
- **PostgreSQL** (anche detto semplicemente Postgres):
  - Sistema di gestione di *database* relazionali *open-source* conforme a **SQL**. Nell'infografica viene utilizzato per la memorizzazione e la gestione di testi contenenti le informazioni sull'infografica e le sezioni da cui sono ricavate. Postgres viene utilizzato insieme all'estensione **pgvector** che consente la gestione di vettori e operazioni di similarità. Questa estensione viene usata nel progetto per memorizzare i vettori di **embedding** riferiti ai suddetti testi.
  - PostgreSQL è disponibile al seguente link <https://www.postgresql.org/>, mentre pgvector è disponibile a <https://github.com/pgvector/pgvector>.
- **pgAdmin:**
  - Strumento *open-source* di gestione e amministrazione per Postgres. Esso fornisce un'interfaccia grafica per la gestione dei *database*. È utilizzato nel progetto per facilitare la gestione e la configurazione del *database* PostgreSQL.
  - La documentazione di pgAdmin è disponibile a <https://www.pgadmin.org/>.
- **psycopg2:**
  - Libreria **Python** per l'interazione con *database* PostgreSQL. Essa rende possibile e facilita l'esecuzione di *query* e la gestione di dati in un *database* PostgreSQL tramite **Python**.
  - La documentazione di psycopg2 è disponibile al link <https://www.psycopg.org/>.
- **Txtai:**
  - Modulo **Python** per la gestione di **embedding** e ricerca semantica. Viene utilizzato per generare e gestire gli **embedding** dei testi memorizzati in PostgreSQL. Questo strumento aiuta, dunque, a trovare risposte più pertinenti e precise grazie alla rappresentazione dei testi in uno spazio vettoriale.

- La documentazione di Txtai è disponibile al link <https://neuml.github.io/txtai/>.

- **Flask:**

- Micro-framework per [Python](#) utilizzato per costruire applicazioni web. In questo progetto, Flask è impiegato per creare il server web che gestisce le richieste provenienti dal *chatbot* dell'infografica, restituendo la risposta al messaggio dell'utente dopo aver richiamato gli algoritmi di ricerca. Siccome il server Flask e l'infografica risiedono su domini diversi, è necessario utilizzare anche l'estensione **Flask-CORS** che consente di gestire le richieste [cross-origin](#).
- La documentazione di Flask è disponibile a <https://flask.palletsprojects.com/>, mentre Flask-CORS è disponibile al link <https://github.com/corydolphin/flask-cors>.

Si precisa, infine, che nella versione integrata del progetto, si utilizza **Python HTTP server** (incluso nella libreria standard di [Python](#)) per gestire le richieste HTTP e visualizzare i contenuti web dell'infografica. Questo approccio ha scopo analogo a `http-server` di Node.js nella versione base, ma sfrutta le capacità di [Python](#), integrandosi senza soluzione di continuità con le altre tecnologie utilizzate in questa versione. Consente, dunque, di mantenere un'infrastruttura coerente ed evitare l'uso di tecnologie aggiuntive.

### 4.2.2 Implementazione

Si presentano di seguito i dettagli implementativi dell'infografica. Saranno descritti i vari componenti del prodotto e come essi sono stati realizzati per garantire un'esperienza utente coinvolgente, oltre che informativa.

#### Elaborazione e gestione dei dati

Per leggere i dati e utilizzarli in [JS](#) si sono utilizzati le seguenti funzioni di [D3.js](#):

- `d3.json()`, usato per caricare i dati [GeoJSON](#) relativi alle strutture d'interesse dell'Università.
- `d3.dsv()`, impiegato per leggere i dati in formato [Comma-Separated Values \(CSV\)](#) che utilizzano “;” come delimitatore. Questo metodo è utilizzato sia per i dati delle materie del corso sia per i dati provenienti dai report dell'Università degli Studi di Padova.
- `d3.text()`, utilizzato per leggere i dati forniti da Almalaurea. Sebbene questi dati siano in formato [CSV](#), la loro struttura complessa, organizzata in sezioni, domande e risposte, richiede una lettura come testo e una successiva strutturazione manuale. I dati vengono, infatti, elaborati e organizzati attraverso una funzione creata appositamente.

Una volta che i dati sono stati letti ed eventualmente elaborati, essi vengono passati alle varie funzioni che creano gli elementi di [Data Visualization](#). Ciò è possibile utilizzando le funzioni `.then()` per gestire le promesse restituite dalle operazioni di lettura dei dati, assicurando che le elaborazioni e le visualizzazioni siano fatte solamente dopo il completo caricamento.

### Struttura dell'infografica nel suo insieme

**Struttura generale del documento HTML.** Il documento HTML dell'infografica è organizzato nelle seguenti parti principali:

- **Intestazione (<head>):** contiene i metadati del documento, nello specifico il titolo della pagina, le descrizioni e le parole chiave per i motori di ricerca. Contiene, inoltre, i fogli di stile [CSS](#) e i collegamenti a *font* esterni.
- **Corpo (<body>):** contiene il contenuto visivo della pagina ed è strutturato come segue:
  - **Chatbot (<div>):** contiene la parte relativa al *chatbot*. Maggiori informazioni sono disponibili alla sezione apposita [Chatbot](#).
  - **Header (<div>):** contiene la sezione introduttiva. Questa sezione è posta al di fuori del <main> in quanto non è solo parte del flusso informativo dell'infografica ma ne è anche l'introduzione. Ciò comporta, dal punto di vista stilistico, una grafica più impattante e non conforme al resto del flusso, in quanto deve catturare immediatamente l'attenzione dell'utente. Tale caratteristica stilistica si riflette dunque anche nella struttura stessa del documento.
  - **Contenuto principale (<main>):** contiene la parte principale, il cuore, del documento. Maggiori informazioni sono disponibili al paragrafo apposito [Struttura del <main>](#).
  - **Footer (<div>):** contiene le fonti delle informazioni presentate nell'infografica. Questa sezione è posizionata al di fuori del <main> in quanto non è direttamente parte del flusso informativo dell'infografica, ma semplicemente ne fornisce informazioni di corredo.

Alla fine del corpo del documento sono inclusi anche i collegamenti agli script [JS](#), sia di librerie esterne sia di script implementati dallo stagista.

**Struttura del <main>.** Il contenuto principale dell'infografica è suddiviso in diverse sezioni, ciascuna rappresentata da un tag <div> con attributo `class="section"`. Tramite tale attributo, infatti, è possibile, tra le altre cose, applicare il *layout* scelto per l'infografica, allineando il contenuto delle sezioni a destra e a sinistra in maniera alternata.

```

1 #content > .section:nth-child(odd) { /*odd sections*/
2     background-image: url(../images/bg-connection-overlay.svg),
3       linear-gradient(to right, var(--light-neutral), white);
4     background-position: left;
5     background-repeat: no-repeat;
6     background-size: contain;
7     text-align: left;
8     align-items: flex-start;
9 }
10 #content > .section:nth-child(even) { /*even sections*/
11     background-image: url(../images/bg-connection-overlay-rotated.
12       svg), linear-gradient(to right, white, var(--light-neutral))
13       ;
14     background-position: right;
15     background-repeat: no-repeat;
16     background-size: contain;

```

```

15     text-align: right;
16     align-items: flex-end;
17 }

```

Dal codice si nota che le varie sezioni, oltre a un diverso allineamento, hanno anche un diverso sfondo. Nello specifico, vengono utilizzati la stessa immagine e lo stesso gradiente per tutte le sezioni, ma cambiato di direzione, in modo da allinearlo con il contenuto delle sezioni stesse. Questo approccio aiuta a mantenere una coerenza visiva e, al contempo, rafforzare l'impatto del *layout*.



Figura 4.1: Esempio di cambio sfondo tra sezioni

Per quanto riguarda la composizione interna delle sezioni dell'infografica, queste sono contenitori *flex* a direzione verticale e comprendono un titolo (<h2>), un sottotitolo (<h3>) e una parte di contenuto principale (<div>). Se sono presenti informazioni di corredo, come le opinioni degli studenti sui vari temi, queste sono incluse in ulteriori <div> di appendice.

All'interno dei <div> di contenuto principale, si trovano solitamente i grafici più rilevanti e/o i contenitori del testo informativo. Le appendici, invece, contengono i grafici di corredo più piccoli.

In generale, tutti i grafici sono racchiusi in <div> con attributo `class="dataviz-wrapper"`, ad eccezione della mappa dei luoghi, che è gestita a sé da un contenitore dedicato. Tale struttura consente di creare simultaneamente gli [SVG](#) per i grafici di [Data Visualization](#) che hanno la stessa dimensione. Nello specifico, si avranno [SVG](#) per i grafici principali di grandi dimensioni, mentre per le appendici si avrà altezza minore. Nel caso siano presenti più grafici affiancati, le dimensioni degli [SVG](#) sono ridotte anche per larghezza. Si riporta di seguito il codice utilizzato per tale obiettivo.

```

1  d3.selectAll(".section-content > .dataviz-wrapper")
2    .append("svg")
3    .attr("preserveAspectRatio", "xMinYMin meet")
4    .attr("viewBox", '0 0 ${widthBig} ${heightBig}');
5
6  d3.selectAll(".section-appendix > .dataviz-wrapper")
7    .append("svg")
8    .attr("preserveAspectRatio", "xMinYMin meet")
9    .attr("viewBox", '0 0 ${widthBig+margin} ${heightSmall+margin}');
10
11 d3.selectAll(".section-content-gridlike > .dataviz-wrapper")
12   .append("svg")
13   .attr("preserveAspectRatio", "xMinYMin meet")
14   .attr("viewBox", '0 0 ${widthSmall+margin} ${heightSmall+margin}');

```

La creazione dei grafici in sé viene poi effettuata selezionando tali [SVG](#) a partire dall'*id* della sezione di riferimento.

Questo approccio consente dunque una facile aggiunta e creazione dinamica degli [SVG](#), che sia coerente con il resto dell'infografica e che non comporti ripetizioni inutili di codice.

**Elementi interattivi non specifici.** Sono presenti alcuni elementi interattivi ricorrenti in più parti dell'infografica, quali:

- **Tooltip su testo**, usati per spiegare e/o approfondire alcune parti del testo. Tali *tooltip* sono visibili quando l'utente passa il cursore su parti specifiche indicate da una "i" cerchiata.

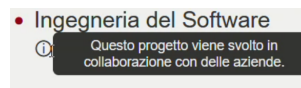


Figura 4.2: Esempio di tooltip su testo

Tale funzionalità è implementata tramite una combinazione di [JS](#) e [CSS](#).

- **Tooltip su grafici**, usati per dettagli aggiuntivi sui dati quando l'utente passa il cursore sui grafici. Questi *tooltip* sono utili per comprendere meglio le informazioni visualizzate, visionando il valore esatto dei dati.

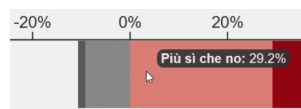


Figura 4.3: Esempio di tooltip su un grafico

Tale funzionalità è implementata tramite una combinazione di [JS](#) e [CSS](#), utilizzando anche [D3.js](#) per la selezione dell'elemento del grafico in questione.

- **Contenitori collassabili**, usati per contenere i grafici rappresentanti le opinioni degli studenti sui vari temi. Questi contenitori possono essere aperti o chiusi tramite apposito pulsante (+ o -), consentendo di visualizzare o nascondere rapidamente queste informazioni extra. Questa funzionalità è particolarmente utile in un'infografica *densa* e lunga come la presente, in quanto permette di rendere la navigazione più fluida e migliorare così esperienza utente.



Figura 4.4: Esempio di contenitore "collassato"

Tale funzionalità è implementata tramite una combinazione di [JS](#) e [CSS](#).

- **Panning e zoom**, utilizzati per esplorare i grafici principali di ogni sezione. Essi permettono agli utenti di ingrandire e spostare la visualizzazione per esaminare dettagli specifici oppure per adattare la visualizzazione su schermi di dimensioni più ridotte.

Tali funzionalità sono implementate attraverso [D3.js](#) e, in particolare, tramite `d3.zoom()`.

- **Scorrimento orizzontale**, utilizzato per esplorare i grafici di corredo, che hanno un'ampia larghezza e un'altezza non elevata. È necessaria tale funzionalità per poter vedere questi grafici in dispositivi con schermi piccoli, specie se sviluppati verticalmente come gli smartphone.

Tale funzionalità è implementata tramite una regola **CSS** e, in particolare, ponendo **overflow: auto;**.

Un altro elemento interattivo, non associabile a nessuna sezione specifica, è il pulsante che consente di aprire il *chatbot*.



**Figura 4.5:** Pulsante per l'apertura del chatbot.

Questo pulsante si trova in posizione “galleggiante” nell'angolo in basso a destra dell'infografica. Tuttavia, è possibile anche spostare tale pulsante, trascinandolo a piacimento all'interno della schermata. Ciò consente di accedere facilmente alla chat in qualsiasi momento e, al contempo, evitare che questo pulsante copra parti importanti del contenuto.

### Struttura dell'infografica per sezione

Si riportano di seguito le varie sezioni dell'infografica.

Si segnala che lo sfondo utilizzato nelle sezioni del contenuto principale non è visibile nelle figure qui presentate a causa della strumentazione utilizzata per la loro generazione. Si prega, dunque, di fare riferimento alla figura 4.1 per visualizzarne i dettagli.

**Introduzione.** La sezione introduttiva dell'infografica è illustrata nella figura seguente 4.6.



**Figura 4.6:** Sezione introduttiva dell'infografica

L'elemento principale di tale sezione è il titolo, inserito in un tag `<h1>` e caratterizzato da dimensioni molto elevate che scalano in base alle dimensioni del dispositivo,

mantenendo però il forte impatto visivo. La sezione in sé, grazie a specifiche regole [CSS](#), ricopre l'intero schermo, andando così a rafforzare ancora di più l'impatto visivo.

Si è inoltre deciso di utilizzare uno sfondo rosso per questa sezione che, non solo amplifica l'attenzione sul titolo, ma anche la differenzia visivamente dal resto del flusso informativo, rispondendo così alle esigenze progettuali.

In aggiunta, sono stati inseriti i loghi dell'Università e del dipartimento del corso per conferire autorità e contestualizzare il corso stesso. Tuttavia, questi loghi sono stati posizionati in secondo piano o marginalmente, poiché l'obiettivo principale è valorizzare il corso in sé per le sue caratteristiche e opportunità, piuttosto che per l'affiliazione con l'Università stessa.

**Requisiti.** La sezione riguardante i requisiti di accesso al corso è illustrata nella figura seguente [4.7](#).

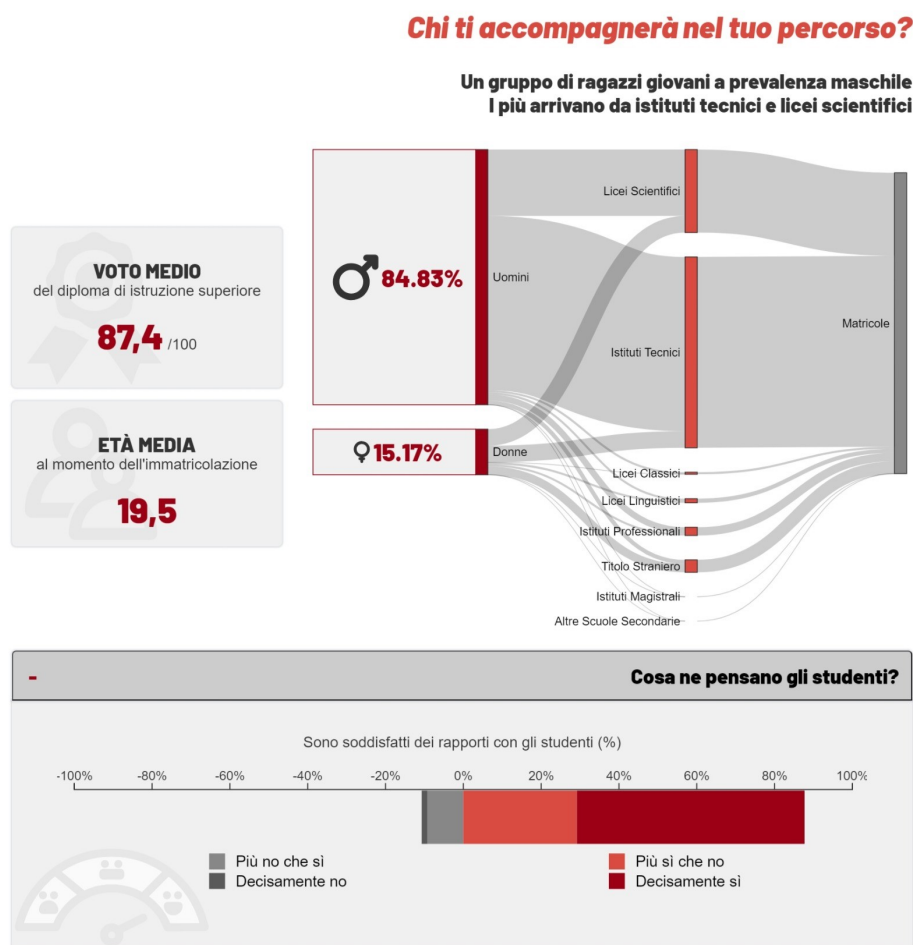


**Figura 4.7:** Sezione dell'infografica sui requisiti del corso

Si ha semplicemente un `<div>` con *layout* a griglia che contiene e organizza i singoli contenitori dei requisiti. Tale *layout* è progettato per adattarsi automaticamente alle dimensioni dello schermo, disponendo i requisiti verticalmente su schermi più stretti e orizzontalmente su schermi più ampi. Ciò garantisce una leggibilità ottimale e una disposizione chiara dei contenuti in tutte le situazioni.

**Profilo studenti.** La sezione riguardante il profilo degli studenti è illustrata nella figura seguente [4.8](#).

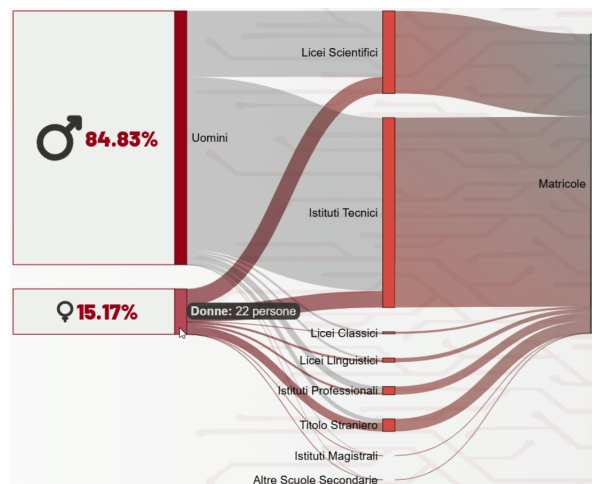




**Figura 4.8:** Sezione dell'infografica sul profilo degli studenti

Questa sezione fornisce una panoramica del profilo degli studenti, in particolare fornisce informazioni riguardanti il voto e l'età media tramite contenitori e dettagli sul sesso e la scuola di provenienza tramite un diagramma di Sankey.

Tale grafico, oltre agli elementi di interazione generale, ne presenta uno ulteriore: posizionando il cursore sui nodi del diagramma si ha un'animazione che colora gradualmente i flussi derivanti da tale nodo. Ciò consente di rendere più leggibili ed evidenziare ulteriormente i dati che l'utente sta visionando. La figura 4.9 mostra l'aspetto del diagramma dopo tale animazione.



**Figura 4.9:** Diagramma di Sankey sul profilo degli studenti animato

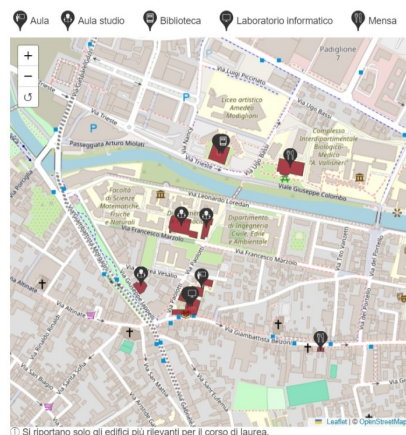
In appendice, su un contenitore collassabile, viene infine riportato un grafico di corredo riguardante l'opinione degli studenti sui rapporti con gli altri studenti.

Si noti che, anche in questo caso, per garantire la leggibilità della sezione anche per schermi di diverse dimensioni, la parte principale della sezione viene adattata in base al dispositivo utilizzato. Nello specifico, per schermi più stretti, i contenitori dei dati e il diagramma di Sankey vengono posti uno sopra l'altro, piuttosto che affiancati.

**Luoghi di interesse.** La sezione riguardante le strutture universitarie d'interesse è illustrata nella figura seguente [4.10](#).

## Dove si va?

### A Padova, città dei "gran dotori"



Il totale delle superfici universitarie supera i

**791.000 mq**  
(115 campi da calcio)

Si ha un rapporto di **100 mq per dipendente** e **10 mq per studente**

L'Università degli Studi di Padova nei suoi

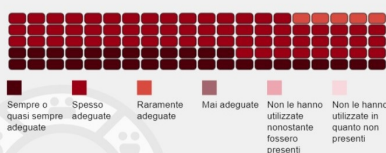
**otto secoli**

di storia ha plasmato il tessuto urbano della città con un mix di edifici storici e moderni.

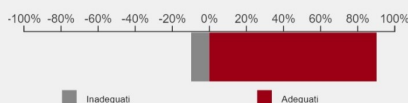
La maggior parte ha **funzione didattica**, ma numerosi sono anche gli uffici, laboratori, archivi, musei, sale convegni e spazi all'aperto.

### Cosa ne pensano gli studenti sulle aule?

Valutazione delle aule (per 100 fruitori)

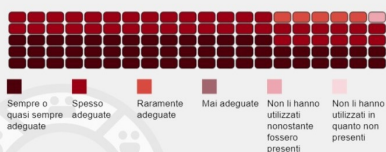


Valutazione degli spazi dedicati allo studio individuale (per 100 fruitori)



### E su laboratori e biblioteche?

Valutazione delle attrezzature per le altre attività didattiche (laboratori, attività pratiche, ...) (per 100 fruitori)



Valutazione dei servizi di biblioteca (prestito/consultazione, orari di apertura, ...) (per 100 fruitori)

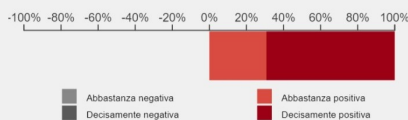


Figura 4.10: Sezione dell'infografica sui luoghi d'interesse

Questa sezione fornisce una panoramica delle strutture dell'Università che potrebbero essere interessanti per un futuro studente del corso. In particolare, fornisce informazioni riguardanti le proprietà dell'Università nel loro insieme tramite contenitori e dettagli sulle strutture di interesse tramite una mappa.

In questa mappa, le strutture vengono evidenziate in rosso per metterle in risalto e visualizzarne più velocemente la dimensione. Inoltre, vengono inserite delle puntine

che indicano la tipologia di struttura; la cui legenda è riportata al di sopra della mappa. Cliccando su tali puntine è possibile vedere i dettagli riguardanti la struttura corrispondente, inclusi nome e tipologia, nonché eventuali altri dettagli utili. Inoltre, la struttura viene evidenziata in un altro colore che ne enfatizza ulteriormente la selezione. Ciò permette di aumentare la leggibilità della mappa e, al contempo, di fornire ulteriori informazioni. Se una stessa struttura ha tipologia mista, ovvero ha più usi, vengono mostrate più puntine sovrapposte. Tuttavia, è possibile vedere individualmente ciascuna puntina cliccando sull'insieme di queste. La figura 4.11 mostra l'interazione.

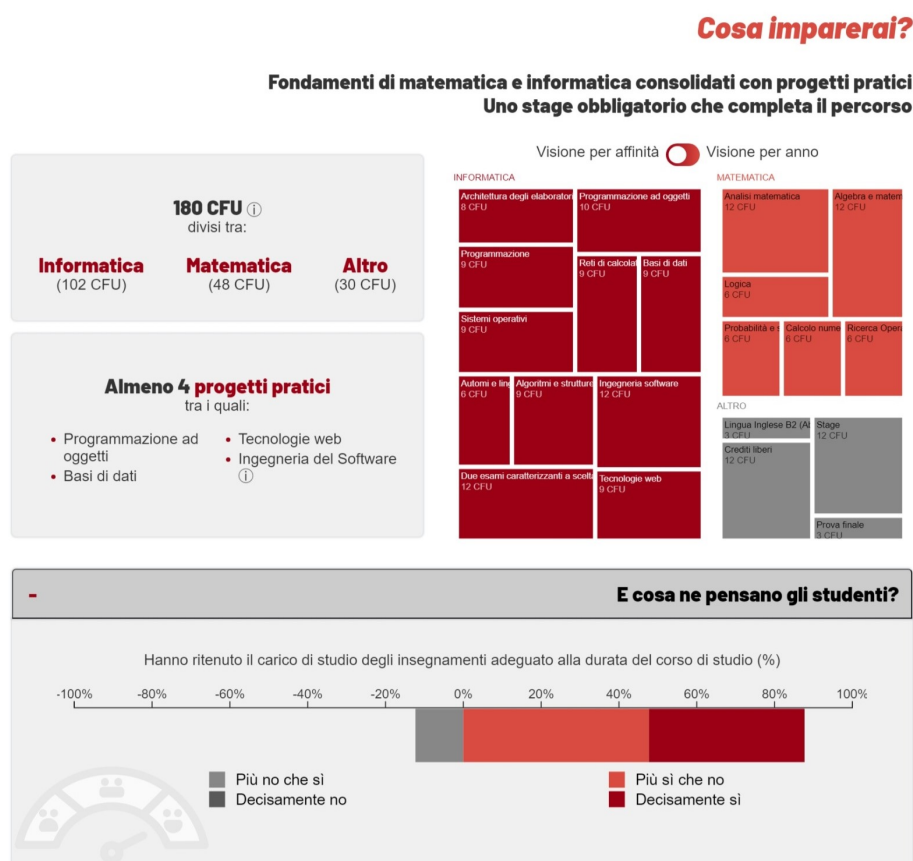


**Figura 4.11:** Mappa dei luoghi dell'Università - visualizzazione individuale delle puntine e pop-up

In appendice, su un contenitore collassabile, vengono riportati diversi grafici di corredo riguardanti l'opinione degli studenti sui vari tipi di strutture.

Si noti che, anche in questo caso, per garantire la leggibilità della sezione anche per schermi di diverse dimensioni, la parte principale della sezione viene adattata in base al dispositivo utilizzato. Nello specifico, per schermi più stretti, i contenitori dei dati e la mappa vengono posti uno sopra l'altro, piuttosto che affiancati.

**Materie di studio.** La sezione riguardante le materie di studio è illustrata nella figura seguente 4.12.



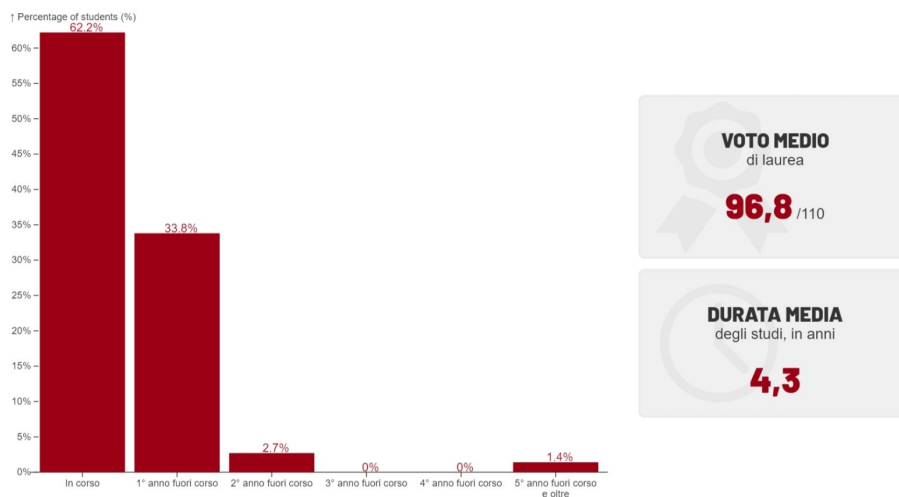
**Figura 4.12:** Sezione dell'infografica sulle materie di studio

Questa sezione fornisce una panoramica delle materie del corso. In particolare, fornisce delle informazioni riguardanti i progetti pratici e una quantificazione generale di impegno tramite contenitori, mentre utilizza un *Treemap* per visualizzare le materie specifiche e il loro peso sul corso.

Tale grafico consente una visualizzazione sia per affinità con temi matematici, informatici o altro, sia per anno di offerta. Tale configurazione è possibile grazie ad uno *switch* posto al di sopra del grafico.

Si noti che, anche in questo caso, per garantire la leggibilità della sezione anche per schermi di diverse dimensioni, la parte principale della sezione viene adattata in base al dispositivo utilizzato. Nello specifico, per schermi più stretti, i contenitori dei dati e il *treemap* vengono posti uno sopra l'altro, piuttosto che affiancati.

**Regolarità negli studi ed esiti.** La sezione riguardante il percorso universitario degli studenti (per durata e voto) è illustrata nella figura seguente 4.13.

**Ma è difficile?****Non sarà una passeggiata, ma sarà soddisfacente****Figura 4.13:** Sezione dell'infografica sul percorso universitario per durata e voto

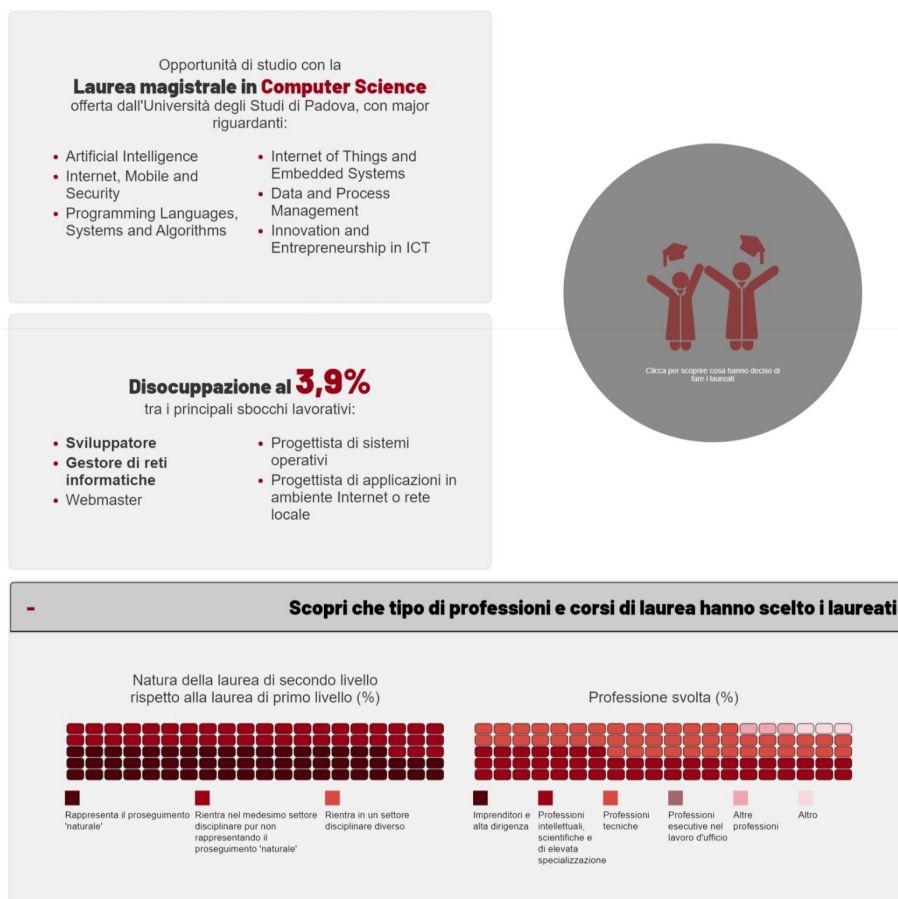
Questa sezione presenta in maniera molto semplice la riuscita e regolarità del percorso di studio degli studenti. In particolare, si utilizza un grafico a barre per rappresentare la durata del percorso, mostrando le quantità di studenti in corso e fuori corso; mentre, si utilizzano dei contenitori per mostrare le statistiche medie relative a voto e durata complessiva.

Si noti che, anche in questo caso, per garantire la leggibilità della sezione anche per schermi di diverse dimensioni, la parte principale della sezione viene adattata in base al dispositivo utilizzato. Nello specifico, per schermi più stretti, i contenitori dei dati e il grafico vengono posti uno sopra l'altro, piuttosto che affiancati.

**Opportunità post-laurea.** La sezione riguardante i percorsi intrapresi post-laurea (lavorativi o formativi) è illustrata nella figura seguente [4.14](#).

## Finalmente laureati! E poi?

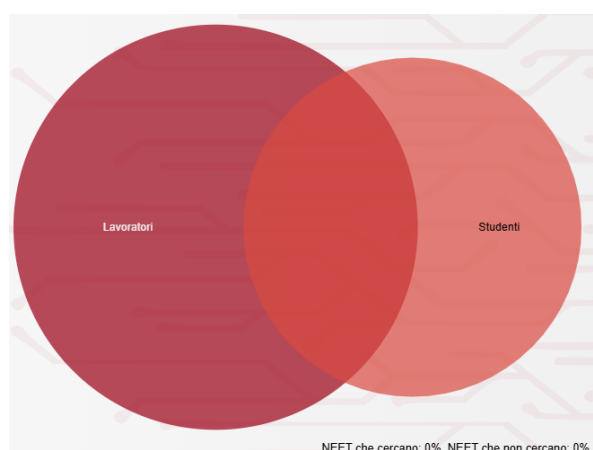
Continuare a studiare o iniziare a lavorare, le opportunità non mancano



**Figura 4.14:** Sezione dell'infografica riguardante i percorsi intrapresi post-laurea

Questa sezione presenta una panoramica delle possibilità per gli studenti dopo la laurea. In particolare, si utilizza un diagramma di Venn per visualizzare la proporzione tra studenti e lavoratori. Si noti che tale statistica non è mostrata fin da subito; infatti, il diagramma mostra inizialmente solo un cerchio con un'icona di due laureati festeggianti. Questa scelta è volta ad attirare l'attenzione dell'utente e fare appello alle sue emozioni, creando una visualizzazione più coinvolgente e accattivante. Le proporzioni sono visibili semplicemente cliccando sul cerchio stesso, ottenendo un diagramma di Venn proporzionale come illustrato nella figura 4.15.

Oltre a tale grafico, sono presenti dei contenitori che descrivono i possibili percorsi, lavorativi o formativi, che i laureati possono intraprendere.



**Figura 4.15:** Diagramma di Venn sulla proporzione di studenti e lavoratori dopo l'animazione

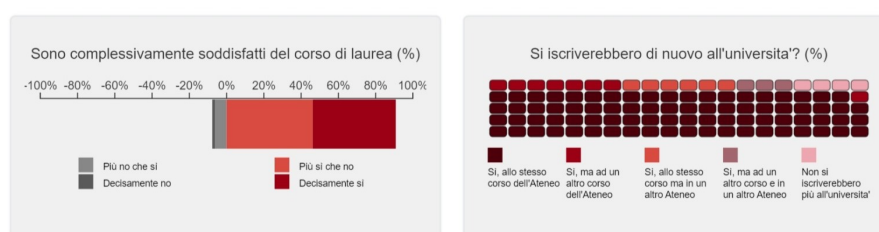
In appendice, su un contenitore collassabile, vengono infine riportati diversi grafici di corredo riguardanti cosa effettivamente hanno scelto di fare gli studenti in caso di proseguimento con gli studi e/o lavoro.

Si noti che, anche in questo caso, per garantire la leggibilità della sezione anche per schermi di diverse dimensioni, la parte principale della sezione viene adattata in base al dispositivo utilizzato. Nello specifico, per schermi più stretti, i contenitori dei dati e il diagramma di Venn vengono posti uno sopra l'altro, piuttosto che affiancati.

**Opinioni finali.** La sezione riguardante l'opinione sul corso dei laureati è illustrata nella figura seguente 4.16.

### Quindi, ne vale la pena?

#### Gli studenti dicono di sì



**Figura 4.16:** Sezione dell'infografica sull'opinione finale dei laureati

Questa sezione presenta in maniera molto semplice le opinioni finali degli studenti sul corso. In particolare, si utilizzano due grafici rispettivamente per visualizzare la soddisfazione generale dei laureati e per indicare se, alla luce della loro esperienza, quest'ultimi sceglierebbero di iscriversi nuovamente al corso.

Si noti che, anche in questo caso, per garantire la leggibilità della sezione anche per schermi di diverse dimensioni, la sezione viene adattata in base al dispositivo utilizzato.



Nello specifico, per schermi più stretti, i contenitori dei grafici disposti verticalmente, piuttosto che orizzontalmente

**Fonti.** Le note a piè di pagina, contenti le fonti, sono illustrate nella figura seguente 4.17.



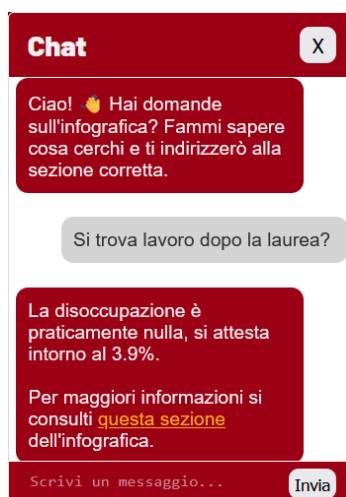
**Figura 4.17:** Sezione dell'infografica sulle fonti

Tale sezione utilizza uno sfondo rosso per segnalare la distinzione dal normale flusso informativo dell'infografica, similmente a quanto accade per la sezione introduttiva. I link alle fonti sono riportati in giallo per differenziarli dal testo normale e, al contempo, garantirne la leggibilità rispetto al colore di sfondo.

### Chatbot

L'infografica dispone di un *chatbot* progettato per rispondere ai dubbi degli utenti e per facilitare la navigazione all'interno dell'infografica, essendo essa molto lunga e dunque potenzialmente disorientante.

L'accesso al *chatbot* avviene tramite un pulsante (visibile nella figura 4.5), il quale apre una finestra nell'angolo in basso a destra dell'infografica. All'interno di tale finestra gli utenti possono porre domande e ricevere una breve risposta dal *chatbot* per il loro interrogativo. Quest'ultimo, inoltre, rimanda anche alla sezione dell'infografica che tratta l'argomento della domanda per avere delle informazioni più dettagliate.



**Figura 4.18:** Esempio di finestra della chat

Dal punto di vista implementativo, gli utenti inseriscono le loro domande in un campo di testo (<textarea>) inserito all'interno di un tag [HTML](#) <form>. Quando l'utente invia il messaggio, sia tramite pulsante apposito sia tramite "Invio" da tastiera,

viene inviato anche il modulo e, di conseguenza, richiamata una funzione [JS](#) che gestisce le funzionalità della chat. Nello specifico, tale funzione di occupa di:

- Cancellare il contenuto del tag `<textarea>` e riscriverlo, invece, all'interno dell'area della chat in un nuovo `<div>` rappresentante il messaggio inviato dall'utente.
- Aggiungere un ulteriore `<div>` all'area di chat che simula una risposta in arrivo dal *bot*. Contiene infatti tre puntini animati a indicare che il messaggio è in fase di elaborazione. L'animazione di tali puntini è realizzata attraverso transizioni [CSS](#).
- Richiamare la funzione che genera la risposta al messaggio inviato e inserire tale risposta al posto dei tre puntini.

Si noti che la creazione dei messaggi all'interno della chat viene effettuata tramite [D3.js](#), che consente un inserimento e aggiornamento dinamico degli elementi.

Per quanto riguarda invece la funzione di generazione della risposta, questa varia in base alla versione dell'infografica utilizzata. Nei paragrafi successivi verranno riportati i dettagli specifici del funzionamento di ciascuna versione.

**Generazione della risposta nella versione base.** Viene richiamata la funzione `answer(q)` passando il messaggio dell'utente come parametro. Tale funzione viene eseguita all'interno di `setTimeout()` per simulare un'elaborazione del messaggio, consentendo anche la visualizzazione dei tre puntini di caricamento anziché mostrare immediatamente la risposta. Ciò consente di fornire un'interazione più naturale e dunque migliorare l'esperienza utente. La funzione `answer(q)` viene definita come segue:

```

1 function answer(q) {
2   // Remove stop-words of user question
3   let filteredUserQ = removeStopWords(q);
4
5   // Remove stop-words from faq questions
6   let filteredQs = faq.map(el =>
7     removeStopWords(el.question)
8   );
9
10  // Stemming of user question
11  var stem = PorterStemmerIt.newStemmer('italian').stem;
12  filteredUserQ = filteredUserQ.map(el => stem(el));
13
14  // Stemming of faq questions
15  filteredQs = filteredQs.map(el => el =
16    el.map(item => stem(item))
17  );
18
19  // Indexing: for each word in filteredUserQ returns indexes of
20    filteredQs that has that word
21  let invertedIndex = Array.from({ length: filteredUserQ.length },
22    () => []);
23  filteredQs.forEach((q, q_index) => {
24    filteredUserQ.forEach((word, i_word) => {
25      if(q.includes(word)) {
26        invertedIndex[i_word].push(q_index);
27      }
28    })
29  });
30 }
```

```

28
29     const qsIndexes = getQsIndexCorrespondence(invertedIndex);
        // sorted array of elements [q_index, terms_count_in_q_index
        // _question]
30
31     if (qsIndexes.length > 0) {
32         const highestCorr = faq[parseInt(qsIndexes[0][0])];
33         return highestCorr.answer + '<br /><br /> Per maggiori
            informazioni si consulti <a href="${highestCorr.section
            }">questa sezione</a> dell'infografica.';
34     }
35     else {
36         return "Spiacente, non sembra esserci questa informazione
            nell'infografica. &#128542;";
37     }
38 }

```

dove `faq` è una variabile contenente un array di oggetti nel formato `[question: "string", answer: "string", section: "string"]`, che rappresenta una serie di domande frequenti che l'utente potrebbe fare, assieme alle relative risposte e all'*id* della sezione corrispondente nel documento [HTML](#). Tale array è ricavato da un file [JSON](#) tramite la funzione `d3.json()` fornita da [D3.js](#). È importante notare che il *chatbot* diventa utilizzabile solo dopo che questo file è stato completamente caricato. Per quanto riguarda invece le altre funzioni richiamate, si hanno:

- `removeStopWords(text)`, implementata dallo stagista, che serve a rimuovere le [stop-words](#) dal testo. Tali parole sono rimosse per migliorare l'efficacia della ricerca. A tal fine, il testo viene anche "normalizzato", ovvero viene tolta la punteggiatura e la stringa viene trasformata tutta in minuscolo. Alla fine del processo, `removeStopWords(text)` restituisce un array contenente le parole rimaste dopo le suddette operazioni.
- `stem(word)`, fornita da una libreria esterna, che serve a effettuare lo [stemming](#) della parola.
- `getQsIndexCorrespondence(indexPerWord)`, implementata dallo stagista, dove `indexPerWord` è un array in cui ogni elemento rappresenta una parola della frase da confrontare, nel caso specifico una parola della domanda dell'utente. Dal punto di vista pratico, ciascuno di questi elementi è costituito da un array che contiene gli indici dei documenti - nel caso specifico gli indici delle domande in `faq` - in cui la suddetta parola è presente. La funzione conta quante volte uno stesso indice di documento appare e ordina i risultati in base al numero di occorrenze. Si ottiene dunque, nel caso specifico, un array ordinato del tipo `[[indice domanda FAQ, numero di occorrenze di parole del messaggio dell'utente nella domanda]]`.

La funzione `answer(q)` si occupa dunque di:

- Rimuovere le [stop-words](#) sia dalla domanda dell'utente che dalle domande frequenti.
- Applicare il processo di [stemming](#), sempre sia dalla domanda dell'utente che dalle domande frequenti.
- Cercare nelle domande frequenti così ottenute quelle che contengono le parole dell'utente risultanti dalle operazioni sopra.

- Contare le occorrenze di queste parole nelle domande e determinare quale domanda ha il maggior numero di corrispondenze.
- Ritornare una risposta (sotto forma di [HTML](#)), composta da `answer` di `faq` relativa alla domanda individuata e un paragrafo aggiuntivo con il collegamento alla sezione dell'infografica relativa (preso da `section` di `faq`). Nel caso non ci siano corrispondenze in nessuna domanda del `faq`, viene restituito un messaggio di errore.

**Generazione della risposta nella versione integrata.** Innanzitutto, si precisa che, prima di utilizzare questa versione, è necessario aver configurato correttamente un *database* con il file contenente le informazioni e le sezioni. Questo passaggio è fondamentale per assicurare che il sistema possa accedere e recuperare i dati necessari a generare risposte pertinenti.

Per quanto riguarda la generazione della risposta in sé, si effettua una richiesta tramite `fetch` all'*endpoint* fornito dal server Flask, passando il messaggio dell'utente come corpo della richiesta. La risposta al messaggio viene quindi elaborata dal server tramite la funzione `Python response_generator` (implementata dal collega Fabio Meneghini) e restituita, poi, in risposta alla richiesta nello stesso formato fornito da `answer(q)` nella versione base. Si noti che, essendo `fetch()` asincrona, non è necessaria alcuna simulazione del tempo di elaborazione. Infatti, mentre il server elabora la richiesta, viene correttamente mostrato il messaggio con i tre puntini di caricamento e, solo una volta ricevuta la risposta, essa viene inserita come messaggio nella chat.

Per quanto riguarda il funzionamento specifico della funzione `response_generator`, essa utilizza un modello di *sentence similarity* (per determinare la vicinanza semantica di parole o frasi) e il metodo [Ranking BM25](#) per cercare informazioni nel *database*. Si noti che quest'ultimo metodo viene implementato tramite *query SQL*, la quale utilizza una funzione nativa di Postgres che realizza anche la rimozione delle *stop-words* e lo *stemming*.

Successivamente, `response_generator` combina i risultati ottenuti dalle suddette tecniche utilizzando un algoritmo. A tal proposito, sono disponibili due diverse opzioni: [Reciprocal Rank Fusion \(RRF\)](#) o [Distribution-Based Score Fusion \(DBSF\)](#). Sebbene sia possibile usare entrambi gli algoritmi, per questo caso specifico si consiglia di utilizzare [RRF](#), poiché più veloce e, considerando la brevità dei testi forniti, l'uso di [DBSF](#) non presenterebbe differenze significative in termini di qualità dei risultati. Ciò permette di realizzare una ricerca ibrida, che combina la ricerca semantica, realizzata grazie all'elaborazione e analisi degli [embedding](#), con un metodo statistico, il [Ranking BM25](#), che considera la frequenza e rilevanza delle parole usate dall'utente.

## 4.3 Validazione

Non è stato possibile implementare alcun metodo di validazione dell'infografica. Tuttavia, per eventuali future implementazioni, sono stati individuati i seguenti metodi:

- **Validazione A/B:** questo metodo prevede la creazione di due versioni dell'infografica, a partire dagli stessi dati, ma con lievi variazioni nel design o nei contenuti. Tali versioni sono visualizzate da due diversi gruppi di utenti, i quali forniscono un *feedback* su ciascuna versione. Questo approccio consente, dunque,

di determinare quale sia il design più efficace tramite un riscontro immediato, ottimizzando così l'usabilità dell'infografica. Attualmente, questo metodo potrebbe essere applicato per testare e confrontare la performance e usabilità del *chatbot*, essendone disponibili due versioni.

- **Integrazione di un *mouse tracker*:** questo metodo prevede l'utilizzo di un *mouse tracker*, ovvero uno strumento che consente di memorizzare i movimenti del cursore degli utenti. Ciò consentirebbe di ottenere dei dati su come gli utenti interagiscono con l'infografica, i quali potrebbero essere analizzati per determinare le parti più coinvolgenti o problematiche dell'infografica stessa. Una volta rilevati tali elementi, diventa dunque possibile migliorare la presentazione a partire proprio dal comportamento reale dei fruitori di questa.
- **Memorizzazione delle domande degli utenti:** questo metodo prevede l'integrazione al *chatbot* di un sistema che registri le domande poste dagli utenti. Grazie a tale strumento sarebbe, infatti, possibile raccogliere e analizzare le informazioni su quali siano le parti dell'infografica meno chiare o che necessitano di ulteriori dettagli. Tale approccio permetterebbe dunque di apportare delle modifiche puntuali all'infografica, migliorandone la chiarezza e completezza generale, rispondendo direttamente alle esigenze degli utenti.

## Capitolo 5

# Conclusioni

*In questo capitolo verranno delineati i risultati raggiunti durante lo stage e verranno fornite alcune considerazioni finali.*

### 5.1 Rendiconto dei risultati

#### 5.1.1 Vincoli e obiettivi generali

I vincoli temporali e tecnologici del progetto di stage sono stati rispettati. Infatti, sono state effettuate 306 ore di lavoro in totale e si sono utilizzate, tra le altre tecnologie, [HTML](#), [CSS](#), [JS](#) e [D3.js](#) per lo sviluppo pratico di un'infografica.

Per quanto riguarda gli obiettivi generali, essi sono stati raggiunti pienamente.

Infatti, lo stagista ha acquisito una solida comprensione del campo della [Data Visualization](#) e delle infografiche, riflessa nei capitoli centrali del presente documento.

Inoltre, è stata acquisita una competenza avanzata nell'uso di [D3.js](#), attraverso la lettura di risorse di studio e attraverso l'applicazione pratica, concretizzata tramite lo sviluppo di elementi di [Data Visualization](#) sia presenti nell'infografica sia indipendenti da essa, realizzati a scopo auto-formativo.

Infine, sono state rafforzate anche le competenze di sviluppo web dello stagista, tramite la creazione in [HTML](#), [CSS](#) e [JS](#) dell'esempio di infografica descritto nel precedente capitolo.

#### 5.1.2 Attività svolte e obiettivi specifici

Lo sviluppo del progetto di stage ha portato ad approfondire e analizzare alcuni aspetti inizialmente non previsti o sottovalutati. Ciò ha comportato il prolungamento di alcune attività e, dunque, la riduzione del tempo disponibile per le successive.

Nello specifico, per quanto riguarda l'attività di "Identificazione dei criteri per la scelta del tipo di grafico più adatto, considerando la struttura dei dati e gli obiettivi della visualizzazione", oltre all'identificazione in sé si è sviluppato lo strumento *Chart-chooser* per automatizzare tale processo. L'attività "Ricerca approfondita di informazioni e guide riguardanti il concetto di infografica", invece, ha rivelato un aspetto nella creazione di infografiche che non era stato considerato inizialmente, ovvero l'importanza della

storia. Si è dunque sentita l'esigenza di sviluppare uno strumento, *Infographic-helper*, che aiutasse gli utenti nella costruzione di tale storia.

Lo sviluppo di tali strumenti ha comportato uno slittamento delle attività successive, impedendo la completa realizzazione dell'attività di "Implementazione dei metodi di validazione individuati".

Per quanto riguarda gli obiettivi specifici, si riporta di seguito il loro stato di soddisfacimento alla fine del progetto:

Identificativo	Tipo	Descrizione dell'obiettivo	Stato
O01	Obbligatorio	Comprendere i principi fondamentali della <a href="#">Data Visualization</a> .	Soddisfatto
O02	Obbligatorio	Selezionare il grafico più adatto per tipo di struttura dati e obiettivo della visualizzazione.	Soddisfatto
O03	Obbligatorio	Realizzare grafici informativi e interattivi utilizzando <a href="#">D3.js</a> .	Soddisfatto
O04	Obbligatorio	Realizzare un'infografica web funzionante in <a href="#">HTML</a> , <a href="#">CSS</a> e <a href="#">JS</a> che visualizzi i risultati degli algoritmi di intelligenza artificiale e <a href="#">Machine Learning</a> .	Soddisfatto
O05	Obbligatorio	Creare un <i>template</i> dell'infografica.	Soddisfatto <sup>1</sup>
O06	Obbligatorio	Individuare dei metodi di validazione dell'infografica.	Soddisfatto
D01	Desiderabile	Sviluppare infografiche alternative.	Non soddisfatto
D02	Desiderabile	Implementare un <i>chatbot</i> con <a href="#">LLM</a> che interagisca con l'utente sull'infografica.	Soddisfatto parzialmente, non si utilizza <a href="#">LLM</a>

**Tabella 5.1:** Stato di soddisfacimento degli obiettivi a fine stage

## 5.2 Considerazioni finali

Pur non avendo potuto completare tutte le attività previste, lo stagista si ritiene soddisfatto dello stage svolto sia a livello lavorativo che umano.

<sup>1</sup>Vengono identificati dei "template" generali per i diversi obiettivi delle infografiche, risultanti più utili di un singolo template relativo all'esempio pratico di infografica sviluppato.

Tutti i colleghi, infatti, si sono dimostrati disponibili, rendendo possibile lavorare in un ambiente sereno e amichevole.

Dal punto di vista tecnico, lo stage ha concesso l'opportunità di arricchire il proprio bagaglio di competenze, sia consolidando quanto imparato durante il corso di laurea sia apprendendo nuove conoscenze. Le basi fornite dal corso, tuttavia, hanno facilitato l'acquisizione di queste nuove competenze, rendendo il processo di apprendimento più agevole.

Inoltre, lo stage ha contribuito a colmare anche alcune lacune pratiche, come la gestione degli orari lavorativi e l'organizzazione della ricerca e auto-formazione.



## Appendice A

## Appendice A

### A.1 Grafici risultanti da Chart-chooser

Di seguito si riporta una tabella contenente tutti i grafici possibili individuati, insieme ad un breve consiglio di utilizzo e alle *regole* del [sistema di regole](#) di *Chart-chooser* che possono portare all'uso di tale grafico. Per quanto riguarda la denominazione delle regole si faccia riferimento alle figure [2.3](#) e [2.4](#).

Si precisa che i grafici identificati sono frutto dell'analisi di diverse risorse, principalmente *from Data to Viz*. URL: <https://www.data-to-viz.com/> e *Visual Vocabulary*. URL: <https://ft-interactive.github.io/visual-vocabulary/>.

Nome	Regole	Consiglio d'uso
Diverging Bar	<ul style="list-style-type: none"><li>• deviationNumUniv</li></ul>	Quando la precisione dei valori è importante, quando i valori sono discreti, quando il numero dei valori non è troppo elevato.
Diverging Area Chart	<ul style="list-style-type: none"><li>• deviationNumUniv</li></ul>	Quando si vuole evidenziare un trend, con valori continui o serie temporali.
Diverging Stacked Bar	<ul style="list-style-type: none"><li>• deviationNumMultiOrd</li><li>• deviationMixMultiNumOneObs</li></ul>	Quando si vuole evidenziare la divergenza per componenti di un gruppo, quando il numero delle variabili non è elevato.
Butterfly Chart	<ul style="list-style-type: none"><li>• deviationNumMultiOrd</li></ul>	Quando si hanno due variabili numeriche rappresentanti gruppi contrastanti.

Nome	Regole	Consiglio d'uso
Connected Scatterplot	<ul style="list-style-type: none"> <li>• correlationNumMultiOrd</li> <li>• correlationMixMultiNumOrd</li> <li>• timeNumMultiOrd</li> <li>• timeMixMultiNumOrd</li> </ul>	Quando l'ordine dei punti è importante, quando si vogliono mostrare come si evolvono le relazioni tra un punto e l'altro.
Scatterplot	<ul style="list-style-type: none"> <li>• correlationNumTwoVarUnord</li> </ul>	Quando si vogliono identificare anomalie, si vuole visualizzare la distribuzione dei punti.
Correlogram	<ul style="list-style-type: none"> <li>• correlationNumLotsVarUnord</li> <li>• correlationMixMultiNumUnord</li> </ul>	Quando si vuole esaminare correlazioni a coppie tra molte variabili.
Bubble Chart	<ul style="list-style-type: none"> <li>• correlationNumThreeVarUnord</li> <li>• rankingNumThreeVarUnord</li> <li>• magnitudeNumThreeVarUnord</li> </ul>	Quando si vuole mostrare la correlazione tra tre variabili simultaneamente.
Line+Column Chart	<ul style="list-style-type: none"> <li>• correlationNumLotsVarUnord</li> <li>• timeNumMultiOrd</li> </ul>	Quando si vuole mostrare evoluzioni della relazioni tra dati con due diverse unità di misura.
Ordered Bar	<ul style="list-style-type: none"> <li>• rankingCatUniv</li> <li>• rankingMixUnivOneObs</li> </ul>	Quando si hanno tante categorie.
Ordered Column	<ul style="list-style-type: none"> <li>• rankingCatUniv</li> <li>• rankingMixUnivOneObs</li> </ul>	Quando si hanno poche categorie.
Lollipop Chart	<ul style="list-style-type: none"> <li>• rankingCatUniv</li> <li>• rankingMixUnivOneObs</li> <li>• timeNumMultiOrd</li> <li>• magnitudeCatUniv</li> <li>• magnitudeMixUnivOneObs</li> </ul>	Da usare come alternativa all' <i>Ordered Bar</i> o all' <i>Ordered Column</i> .
Slope Chart	<ul style="list-style-type: none"> <li>• rankingCatMultiInd</li> <li>• rankingMixMultiNumOneObs</li> <li>• timeCatMultiInd</li> <li>• timeMixMultiNumOneObs</li> </ul>	Quando si vogliono visualizzare dei cambiamenti discreti nel tempo, quando si hanno poche categorie.
Dot Strip Plot	<ul style="list-style-type: none"> <li>• rankingMixUnivMoreObs</li> <li>• distributionNumLotsVarUnord</li> </ul>	Quando si vogliono visualizzare ranking e distribuzione di dataset ampi.

Nome	Regole	Consiglio d'uso
Histogram	<ul style="list-style-type: none"> <li>• distributionNumUniv</li> <li>• distributionNumTwoVarUnord</li> <li>• distributionMixUnivMoreObs</li> <li>• distributionMixUnivOneObs</li> </ul>	Quando si vuole visualizzare frequenze su intervalli, quando si hanno pochi punti, quando si vuole una rappresentazione semplice adatta a tutti.
Density Plot	<ul style="list-style-type: none"> <li>• distributionMixUnivMoreObs</li> <li>• distributionNumUniv</li> </ul>	Quando si ha una distribuzione non categorizzata in bin, quando si vuole mostrare la "forma" dei dati.
Cumulative curve	<ul style="list-style-type: none"> <li>• distributionNumUniv</li> <li>• distributionMixUnivMoreObs</li> </ul>	Quando si vuole mostrare le proporzioni dei dati e le loro tendenze.
Boxplot	<ul style="list-style-type: none"> <li>• distributionNumLotsVarUnord</li> <li>• distributionMixUnivMoreObs</li> <li>• distributionMixMultiNumUnord</li> <li>• distributionMixMultiCatNestMoreObs</li> <li>• distributionMixMultiCatSubMoreObs</li> </ul>	Quando si hanno pochi punti, quando si vogliono mostrare statistiche come mediana, quartili e casi anomali.
Violin Plot	<ul style="list-style-type: none"> <li>• distributionNumLotsVarUnord</li> <li>• distributionMixUnivMoreObs</li> <li>• distributionMixMultiNumUnord</li> <li>• distributionMixMultiCatNestMoreObs</li> <li>• distributionMixMultiCatSubMoreObs</li> </ul>	Quando si hanno tanti punti, quando si vuole combinare <i>Boxplot</i> con <i>Density Plot</i> .
2D Density Plot	<ul style="list-style-type: none"> <li>• distributionNumTwoVarUnord</li> <li>• distributionMixMultiNumUnord</li> </ul>	Quando si hanno tanti punti, quando si ha un dataset ampio.
Ridgeline	<ul style="list-style-type: none"> <li>• distributionNumLotsVarUnord</li> <li>• distributionMixUnivMoreObs</li> </ul>	Quando si hanno tanti punti, quando si vogliono visualizzare più distribuzioni.
Calendar Heatmap	<ul style="list-style-type: none"> <li>• timeNumUniv</li> </ul>	Quando si hanno dati giornalieri e si vogliono visualizzare pattern che si ripetono.
Seismogram	<ul style="list-style-type: none"> <li>• timeNumUniv</li> </ul>	Quando si vogliono mostrare oscillazioni o cambiamenti repentini o alti nel tempo.

Nome	Regole	Consiglio d'uso
Line Chart	<ul style="list-style-type: none"> <li>• timeNumMultiOrd</li> <li>• timeMixMultiNumOrd</li> </ul>	Quando si vogliono mostrare dei trend continui nel tempo.
Area Chart	<ul style="list-style-type: none"> <li>• timeNumMultiOrd</li> </ul>	Quando si vogliono mostrare cambiamenti cumulativi nel tempo.
Fan Chart	<ul style="list-style-type: none"> <li>• timeNumMultiOrd</li> <li>• timeMixMultiNumOrd</li> </ul>	Quando si hanno dati predittivi e si vuole rappresentare incertezza o un range di possibili risultati.
Priestley Timeline	<ul style="list-style-type: none"> <li>• timeCatUniv</li> </ul>	Quando si vuole mostrare una sequenza di eventi nel tempo.
Circle Timeline	<ul style="list-style-type: none"> <li>• timeMixUnivMoreObs</li> </ul>	Quando si vuole mostrare la grandezza di valori discreti che cambiano nel tempo.
Stacked Area Chart	<ul style="list-style-type: none"> <li>• timeMixMultiNumOrd</li> </ul>	Quando si vogliono mostrare cambiamenti cumulativi nel tempo enfatizzandole la grandezza, quando si hanno poche categorie.
Stream Graph	<ul style="list-style-type: none"> <li>• timeMixMultiNumOrd</li> </ul>	Alternativa a <i>Stacked Area Chart</i> che enfatizza il flusso di cambiamenti cumulativi nel tempo, quando si hanno poche categorie.
Pie Chart	<ul style="list-style-type: none"> <li>• compositionCatUniv</li> <li>• compositionMixUnivOneObs</li> </ul>	Quando si ha un dataset piccolo, quando si vogliono evidenziare le proporzioni.
Donut Chart	<ul style="list-style-type: none"> <li>• compositionCatUniv</li> <li>• compositionMixUnivOneObs</li> </ul>	Alternativa a <i>Pie Chart</i> , quando si ha un dataset piccolo, quando si vogliono evidenziare le proporzioni.
Waterfall	<ul style="list-style-type: none"> <li>• compositionCatUniv</li> <li>• compositionMixUnivOneObs</li> <li>• flowCatUniv</li> </ul>	Quando si vogliono mostrare come i valori influiscono nel totale.

Nome	Regole	Consiglio d'uso
Waffle Chart	<ul style="list-style-type: none"> <li>• compositionCatUniv</li> <li>• compositionMixUnivOneObs</li> </ul>	Quando si vuole mostrare la composizione di una categoria in griglia.
Venn Diagram	<ul style="list-style-type: none"> <li>• compositionCatMultiInd</li> </ul>	Quando si vogliono mostrare le parti che si sovrappongono tra diverse categorie.
Treemap	<ul style="list-style-type: none"> <li>• compositionCatMultiNest</li> <li>• compositionMixMultiCatNestOneObs</li> </ul>	Quando si vogliono mostrare dati gerarchici, quando si vogliono mostrare tanti valori in uno spazio compatto.
Sunburst	<ul style="list-style-type: none"> <li>• compositionCatMultiNest</li> <li>• compositionMixMultiCatNestOneObs</li> </ul>	Quando si vogliono mostrare dati gerarchici.
Icicle	<ul style="list-style-type: none"> <li>• compositionCatMultiNest</li> <li>• compositionMixMultiCatNestOneObs</li> </ul>	Alternativa al Sunburst, quando si vogliono mostrare dati gerarchici con settori rettangolari.
Stacked Column Chart	<ul style="list-style-type: none"> <li>• compositionCatMultiSub</li> <li>• compositionMixMultiCatSubOneObs</li> </ul>	Quando si vuole mostrare la composizione di diverse categorie o su periodi di tempo.
Stacked Bar Chart	<ul style="list-style-type: none"> <li>• compositionCatMultiSub</li> <li>• compositionMixMultiCatSubOneObs</li> <li>• magnitudeMixMultiCatSubOneObs</li> </ul>	Quando si vuole mostrare la composizione di diverse categorie.
Proportional Stacked Bar	<ul style="list-style-type: none"> <li>• compositionCatMultiSub</li> <li>• compositionMixUnivOneObs</li> <li>• compositionCatMultiSub</li> <li>• magnitudeMixUnivOneObs</li> </ul>	Quando si vuole mostrare la contribuzione relativa delle categorie sul totale.
Dendogram	<ul style="list-style-type: none"> <li>• compositionCatMultiNest</li> <li>• compositionMixMultiCatNestOneObs</li> </ul>	Quando si vuole rappresentare una composizione gerarchica.
Circular Packing	<ul style="list-style-type: none"> <li>• compositionCatMultiNest</li> <li>• compositionMixMultiCatNestOneObs</li> </ul>	Quando si vuole rappresentare una composizione gerarchica enfatizzando la grandezza della categoria.

Nome	Regole	Consiglio d'uso
Pictogram	<ul style="list-style-type: none"> <li>• magnitudeCatUniv</li> <li>• magnitudeMixUnivOneObs</li> </ul>	Quando si vuole rappresentare quantità con simboli.
Grouped Barplot	<ul style="list-style-type: none"> <li>• compositionCatMultiSub</li> <li>• magnitudeMixMultiNumOneObs</li> <li>• magnitudeMixMultiCatSubOneObs</li> </ul>	Quando si vuole comparare la grandezza di gruppi tra loro collegati, quando si hanno pochi gruppi.
Radar Chart	<ul style="list-style-type: none"> <li>• magnitudeMixMultiNumOneObs</li> </ul>	Quando si vuole comparare multiple variabili per diverse categorie.
Dot Density	<ul style="list-style-type: none"> <li>• spatialNumTwoVarUnord</li> </ul>	Quando si vogliono visualizzare punti su una mappa, quando si vuole mostrare la densità in diverse regioni.
Proportional Symbol	<ul style="list-style-type: none"> <li>• spatialNumThreeVarUnord</li> </ul>	Quando si vuole mostrare la magnitudine del dato su una mappa.
Flow Map	<ul style="list-style-type: none"> <li>• spatialNumFourFiveVarUnord</li> <li>• spatialCatTwoVarInd</li> <li>• spatialMixTwoCatInd</li> </ul>	Quando si vogliono visualizzare un flusso tra diversi punti sulla mappa.
Basic Choropleth	<ul style="list-style-type: none"> <li>• spatialCatUniv</li> <li>• spatialMixUnivOneObs</li> </ul>	Quando si vuole mostrare la distribuzione di una variabile tra diverse regioni geografiche.
Equalized Cartogram	<ul style="list-style-type: none"> <li>• spatialCatUniv</li> <li>• spatialMixUnivOneObs</li> </ul>	Quando si vuole mostrare la distribuzione di una variabile tra diverse regioni geografiche in maniera uniforme.
Contour Map	<ul style="list-style-type: none"> <li>• spatialCatUniv</li> <li>• spatialMixUnivOneObs</li> </ul>	Quando si vuole mostrare la divergenza di variabili continue su un'area geografica.
Scaled Cartogram	<ul style="list-style-type: none"> <li>• spatialCatUniv</li> <li>• spatialMixUnivOneObs</li> </ul>	Quando si vuole rappresentare la magnitudine di una variabile su una mappa.

Nome	Regole	Consiglio d'uso
Heatmap	<ul style="list-style-type: none"> <li>• flowCatMultiInd</li> <li>• flowCatMultiSub</li> <li>• flowMixMultiCatSubOneObs</li> <li>• flowMixMultiCatInd</li> </ul>	Quando si vuole visualizzare l'intensità di flussi in forma tabulare.
Sankey Diagram	<ul style="list-style-type: none"> <li>• flowCatMultiInd</li> <li>• flowCatMultiSub</li> <li>• flowMixMultiNumOneOb</li> <li>• flowMixMultiCatSubOneObs</li> <li>• flowMixMultiCatInd</li> </ul>	Quando si vuole visualizzare la struttura del flusso, quando si vuole mostrare la distribuzione proporzionale del flusso.
Network Diagram	<ul style="list-style-type: none"> <li>• flowCatMultiInd</li> <li>• flowMixMultiCatInd</li> </ul>	Quando si vogliono mostrare connessioni tra più entità enfatizzando comunità e intensità dei flussi.
Chord Diagram	<ul style="list-style-type: none"> <li>• flowCatMultiInd</li> <li>• flowMixMultiCatInd</li> </ul>	Quando si vogliono visualizzare connessioni tra più entità enfatizzando l'intensità tra le connessioni.
Arc Diagram	<ul style="list-style-type: none"> <li>• flowCatMultiInd</li> <li>• flowMixMultiCatInd</li> </ul>	Quando si vogliono visualizzare connessioni tra più entità enfatizzando l'intensità tra le connessioni.

**Tabella A.1:** Grafici risultanti da Chart-chooser

# Acronimi e abbreviazioni

**CSS** Cascading Style Sheets. 2–4, 48, 52, 54–56, 66, 70, 71, 81

**CSV** Comma-Separated Values. 51, 81

**DBSF** Distribution-Based Score Fusion. 68, 82

**ERP** Enterprise Resource Planning. 1, 82

**FSL** Few-Shot Learning. 12, 82

**HTML** Hypertext Markup Language. 2–4, 48, 65, 67, 68, 70, 71, 81, 82

**ICL** In-Context Learning. 12, 82

**ICT** Information and Communication Technology. 1, 83

**JS** JavaScript. 2–4, 12, 20, 48–52, 54, 66, 70, 71, 81, 83

**LLM** Large Language Model. 3, 4, 12, 13, 16, 21–23, 71, 83

**MBTI** Myers-Briggs Type Indicator. 34, 83

**RRF** Reciprocal Rank Fusion. 68, 84

**SQL** Structured Query Language. 48, 50, 68, 84

**SVG** Scalable Vector Graphics. 46, 49, 53, 54, 81, 84

**VIF** Visual Information Flow. 26–29, 45, 84

**WCAG** Web Content Accessibility Guidelines. 45, 84



# Glossario

**big data** Grande quantità di dati informatici, estesa sia in termini di volume sia di varietà sia per velocità con cui vengono forniti. Richiedono tecnologie e metodi analitici specifici per poterne estrarre informazioni utili. [5](#)

**business intelligence** Tecnologie e processi aziendali volti a raccogliere ed analizzare i dati al fine di prendere decisioni più informate e sviluppare strategie aziendali. [1](#)

**chartjunk** Termine coniato dallo statistico Edward R. Tufte per indicare gli elementi presenti in una visualizzazione che hanno solamente funzione decorativa o, più in generale, tutti gli elementi che possono essere omessi senza perdere informazioni. [8](#), [9](#)

**cross-origin** Termine che si riferisce alla politica di sicurezza dei browser web che gestisce le richieste tra domini differenti. Nello specifico quando un'applicazione web fa una richiesta a un dominio diverso da quello da cui è stata originariamente caricata, si ha una richiesta *cross-origin*. [51](#)

**CSS** Linguaggio usato per definire la formattazione di documenti scritti in un linguaggio di *markup*, come ad esempio [HTML](#). Nello specifico, vengono definite delle regole che dicono al browser come gli elementi della pagina devono essere visualizzati. [80](#)

**CSV** Formato di file basato su file di testo, solitamente utilizzato per l'importazione ed esportazione di dati tabulari. Ogni riga della tabella viene rappresentata da una riga di dati nel file *csv* e ogni campo in quella riga è separato da una virgola. [80](#)

**D3.js** Libreria [JS](#) *open source* usata per la creazione, a partire da dati organizzati, di visualizzazioni interattive che siano visibili attraverso un comune browser. Il suo nome, infatti, sta per *Data-Driven Documents* (documenti basati sui dati). In particolare, consente di creare grafici interattivi e dinamici sul web, utilizzando [HTML](#), [SVG](#) e [CSS](#). [ii](#), [2-4](#), [49-51](#), [54](#), [66](#), [67](#), [70](#), [71](#)

**Data Science** Disciplina scientifica che combina competenze matematiche, statistiche, informatiche e di visualizzazione al fine di interpretare ed estrarre conoscenza dai dati. [7](#), [8](#), [82](#)

**Data Visualization** Processo di rappresentazione visiva dei dati attraverso grafici, tabelle e altre forme di visualizzazione grafica. Viene utilizzato per facilitare l'interpretazione e l'analisi delle informazioni, permettendo di identificare tendenze,

*pattern* e *insight* in modo chiaro e intuitivo, migliorandone la comprensione. [ii](#), [2–8](#), [18](#), [19](#), [25](#), [26](#), [29](#), [46](#), [48](#), [49](#), [51](#), [53](#), [70](#), [71](#), [82](#)

**DBSF** Algoritmo utilizzato per combinare i risultati provenienti da più sistemi di ricerca. Ad esempio, può essere utilizzato per combinare la ricerca semantica con metodi statistici, realizzando così una ricerca ibrida. Esso utilizza i punteggi ottenuti dai diversi sistemi e li normalizza, indipendentemente dalla media della loro distribuzione, ottenendo così una classifica priva di *bias* di normalizzazione. [80](#)

**e-business** Termine generale che comprende tutte le attività commerciali che possono svolgersi online, ossia attraverso Internet e altre reti telematiche. Comprende anche l'*e-commerce*. [1](#)

**embedding** Rappresentazione di oggetti (come parole o frasi) in uno spazio vettoriale continuo. Queste rappresentazioni vettoriali catturano le relazioni e le somiglianze semantiche tra gli oggetti. Tali strutture sono fondamentali per compiti come la classificazione del testo, la ricerca per similarità e i sistemi di raccomandazione. [50](#), [68](#)

**ERP** *Software* utilizzati per la gestione quotidiana di imprese. Essi integrano tutti i processi e funzioni aziendali rilevanti, come ad esempio vendite, acquisti, gestione magazzino, finanza o contabilità. [80](#)

**Exploratory Data Analysis** Tecnica del campo della [Data Science](#), usata per analizzare e indagare i set di dati e riassumerne le caratteristiche principali. Spesso impiega metodi di [Data Visualization](#). [19](#)

**FSL** Strategia di apprendimento automatico che utilizza un numero molto limitato di esempi di addestramento. Si sfruttano le informazioni apprese precedentemente da compiti o grandi set di dati per aiutare nella classificazione di nuove classi per le quali vengono forniti solo alcuni esempi ciascuna. È un sottoinsieme del *n-shot learning*. [80](#)

**GeoJSON** Formato basato su [JSON](#), progettato per rappresentare dati geospaziali. È ampiamente utilizzato per archiviare e scambiare informazioni geografiche e geometrie spaziali. [51](#)

**grammatica** Nel presente documento, termine utilizzato nel senso di “grammatica formale” e indica un insieme di regole e sintassi formali utilizzate per definire e vincolare il tipo di output che un modello di linguaggio può generare. [23](#)

**HTML** Linguaggio di *markup* ampiamente usato per i documenti web. Stabilisce tramite dei tag cosa un browser web deve mostrare a un visitatore al suo arrivo in un sito. [80](#)

**ICL** Tecnica di [Prompt Engineering](#) che consiste nell’inserire direttamente nel *prompt* una dimostrazione del compito o esempio specifico di risposta che il modello di Intelligenza Artificiale deve compiere/restituire. [80](#)

**ICT** Tecnologie riguardanti i sistemi integrati di telecomunicazione, i computer, le tecnologie audio-video e relativi *software*. In generale, dunque, tutte le tecnologie che permettono agli utenti di creare, immagazzinare e scambiare informazioni. 80

**information overload** Anche detta “sovraccarico informativo”, indica una serie di problemi psicologici, produttivi e decisionali che si verificano a causa di un’esposizione quotidiana ad un numero troppo elevato di informazioni e stimolazioni. Si concretizza negli individui con una difficoltà nell’attenzione e nella comprensione di argomenti, nonché nella presa di decisioni. 5

**JS** Linguaggio di programmazione ad alto livello interpretato e orientato agli eventi. Viene comunemente utilizzato per aggiungere interattività e dinamicità alle pagine web, ma può essere utilizzato anche per lo sviluppo di applicazioni web, mobile e desktop. 80

**JSON** Formato di scrittura (basato su **JS**) che si caratterizza per la sua grande leggibilità, dovuta ad una sintassi semplice e facilmente interpretabile dai computer. Viene utilizzato in diversi contesti, tra cui lo sviluppo web e lo scambio di dati tra applicazioni. 13, 17, 23, 67, 82, 83

**json-rules-engine** Libreria **JS** che permette di definire e gestire regole basate su **JSON** per la costruzione di un *sistema di regole*. Consente, dunque, di implementare logiche di business complesse attraverso un formato di regole dichiarative, definite in maniera semplice e leggibile. 13

**llama.cpp** Strumento *open-source* progettato per eseguire modelli di Intelligenza Artificiale direttamente in C/C++. Consente anche di ricevere in risposta dal modello un output strutturato. 12, 21

**LLM** Tipologia di modello di Intelligenza Artificiale progettato per comprendere e generare testo con una capacità simile a quella umana. Utilizza reti neurali profonde per apprendere da grandi quantità di testi regole linguistiche, coerenza del contesto e stili di scrittura. 80

**Machine Learning** Branca dell’Intelligenza Artificiale che permette ai sistemi di apprendere da grandi set di dati e migliorare automaticamente dalle esperienze, senza ulteriore programmazione. Utilizzano algoritmi statistici e modelli matematici per identificare schemi da queste grandi quantità di dati e fare previsioni basate su nuovi input. 4, 71

**MBTI** Indicatore psicologico che identifica, tramite questionari, quattro caratteristiche psicologiche che, a seconda del modo in cui queste si presentano, determinano il modo in cui una persona si rapporta con il mondo e la vita in generale. La valutazione di tali caratteristiche porta all’identificazione di un tipo di personalità (tra i 16 disponibili, determinati da tutte le possibili combinazioni di queste caratteristiche). 80

**Prompt Engineering** Processo di progettazione e ottimizzazione delle istruzioni fornite a un modello di Intelligenza Artificiale, volto ad ottenere risposte più accurate e pertinenti da questi modelli. La qualità delle risposte, infatti, può variare di molto a seconda della formulazione del *prompt*.. 12, 82

**Python** Linguaggio di programmazione ad alto livello, molto versatile e ampiamente utilizzato. Supporta diversi paradigmi di programmazione, come quello *object-oriented*, quello imperativo e quello funzionale. Inoltre, offre una tipizzazione dinamica forte e presenta una sintassi molto chiara e leggibile. 48, 50, 51, 68

**Ranking BM25** Algoritmo di *ranking* per il reperimento dell'informazione. Esso identifica la pertinenza di un documento rispetto a una certa *query* e ordina i documenti sulla base dei loro punteggi di rilevanza. Tali punteggi sono determinati tenendo conto della frequenza con cui il termine di ricerca compare nel documento, della lunghezza del documento e della lunghezza media di tutti i documenti disponibili. 68

**RRF** Algoritmo utilizzato per combinare i risultati provenienti da più sistemi di ricerca. Ad esempio, può essere utilizzato per combinare la ricerca semantica con metodi statistici, realizzando così una ricerca ibrida. Esso utilizza i *rank* dei documenti ottenuti dai diversi sistemi e li aggrega in una classifica unificata, migliorando così la qualità dei risultati di ricerca complessivi. 80

**server farm** Insieme di server collocati in un unico ambiente fisico. Vengono utilizzate per poter centralizzare la gestione, la manutenzione e la sicurezza dei server stessi. 1

**sistema di regole** Sistema esperto che consente di compiere in automatico certe azioni prestabilite, evitando così l'intervento manuale di figure esperte. Partendo da una serie di dati in input, mette in atto autonomamente procedure di inferenza, di logica, giungendo così alla risoluzione di problemi complessi. 12–14, 73, 83

**SQL** Linguaggio standardizzato usato per *database* basati sul modello relazionale. Consente la creazione e la gestione delle strutture di *database*, come tabelle e indici. Nello specifico, consente di eseguire anche operazioni di interrogazione, aggiornamento, inserimento e cancellazione dei dati nelle tabelle. 80

**stemming** Processo che riduce le parole alla loro radice o forma base. In operazioni di confronto tra testi, questo è utile in quanto permette di accorpare varianti di una parola alla stessa radice, migliorando la precisione in tali confronti. 67, 68

**stop-words** Parole comuni, come ad esempio articoli o congiunzioni, che non si riferiscono ad un argomento specifico e dunque non contribuiscono al significato del testo. In operazioni di confronto tra testi, tali parole sono spesso rimosse per migliorare l'efficacia di queste operazioni. 67, 68

**SVG** Formato di file utilizzato per descrivere grafica vettoriale bidimensionale. In particolare, consente di creare immagini scalabili senza perdita di qualità, adatte per essere visualizzate su dispositivi di diversi formati e dimensioni. 80

**VIF** Modo in cui gli elementi grafici vengono organizzati e collegati tra loro per guidare l'utente attraverso una storia visiva. Tale struttura consente di comprendere e interpretare le informazioni con maggior chiarezza e coerenza. 80

**WCAG** Parte di una serie di linee guida per l'accessibilità dei siti Web, pubblicate dal *World Wide Web Consortium* (W3C). 80

# Bibliografia

## Riferimenti bibliografici

- Bottazzini, Paolo e Michele Gotuzzo. *Design della mente. Infografica e data visualization*. Press Grafica, Srl, 2014 (cit. a p. 34).
- Brand, Willemien. *Visual Thinking: Empowering People and Organisations through Visual Collaboration*. BIS Publishers, 2017.
- Cairo, Alberto. *L'arte funzionale Infografica e visualizzazione delle informazioni*. Pearson Italia, SpA, 2013 (cit. a p. 38).
- Hinderman, Bill. *Building Responsive Data Visualization for the Web*. John Wiley & Sons, Ltd, 2015.
- Lu, Min et al. «Exploring Visual Information Flows in Infographics». In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 2020 (cit. a p. 26).
- Lupton, Ellen. *Design is storytelling*. Cooper Hewitt, 2024.
- Tufte, Edward R. *The Visual Display of Quantitative Information, Second Edition*. Graphics Press, LLC, 2006 (cit. alle pp. 7, 18, 37).
- Tymchenko, Oleksandr, Svitlana Vasiuta e Orest Khamula. «Optimization of the Mathematical Model of Factors of Composite Design of Infographic». In: *2018 IEEE 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT)*. 2018 (cit. a p. 23).

## Siti web consultati

- Almalaurea - Informatica (L-31), Università degli Studi di PADOVA. URL: [https://apex.cca.unipd.it/pls/apex/f?p=144:32:::::P32\\_CODICIONE,P32\\_COD\\_CDS,P32\\_CODICE\\_SEDE,P32\\_TIPO\\_CORSO:0280106203100001,SC1167,PD,L2023](https://apex.cca.unipd.it/pls/apex/f?p=144:32:::::P32_CODICIONE,P32_COD_CDS,P32_CODICE_SEDE,P32_TIPO_CORSO:0280106203100001,SC1167,PD,L2023).
- d3 SVG Legend (v4). URL: <https://d3-legend.susielu.com/>.
- d3-sankey. URL: <https://github.com/d3/d3-sankey>.
- D3.js. URL: <https://d3js.org/>.
- Flask. URL: <https://flask.palletsprojects.com/>.
- flask-cors. URL: <https://github.com/corydolphin/flask-cors>.

*from Data to Viz*. URL: <https://www.data-to-viz.com/> (cit. a p. 73).

*Infogram - What is data visualization?* URL: [https://infogram.com/blog/what-is-data-visualization/?\\_gl=1\\*1rypyyw\\*\\_up\\*MQ..\\*\\_ga\\*MTg5MzY2NzU2MS4xNzE5MzMxMTgw\\*\\_ga\\_LD50PRQER7\\*MTcxOTMzMTE3OS4xLjAuMTcxOTMzMTE3OS4wLjAuMA...](https://infogram.com/blog/what-is-data-visualization/?_gl=1*1rypyyw*_up*MQ..*_ga*MTg5MzY2NzU2MS4xNzE5MzMxMTgw*_ga_LD50PRQER7*MTcxOTMzMTE3OS4xLjAuMTcxOTMzMTE3OS4wLjAuMA...)

*json-rules-engine*. URL: <https://github.com/CacheControl/json-rules-engine>.

*Leaflet*. URL: <https://leafletjs.com/>.

*Medium - Infographics: Purpose, elements and types*. URL: <https://medium.com/@tetracubetech/infographics-purpose-elements-and-types-ae8bfc7cd89b>.

*Node.js*. URL: <https://nodejs.org/>.

*npm - http-server*. URL: <https://www.npmjs.com/package/http-server>.

*Overlapping Marker Spiderfier for Leaflet*. URL: <https://github.com/jawj/OverlappingMarkerSpiderfier-Leaflet>.

*pgAdmin*. URL: <https://www.pgadmin.org/>.

*pgvector*. URL: <https://github.com/pgvector/pgvector>.

*PostgreSQL*. URL: <https://www.postgresql.org/>.

*psycopg*. URL: <https://www.psycopg.org/>.

*Quillbot - Evaluating Appeals to Ethos, Logos, and Pathos*. URL: <https://quillbot.com/courses/english-literacy-and-composition-b/chapter/text-evaluating-appeals-to-ethos-logos-and-pathos/>.

*txtai*. URL: <https://neuml.github.io/txtai/>.

*Università degli Studi di Padova - Dati statistici*. URL: <https://www.unipd.it/dati-statistici>.

*venn.js*. URL: <https://github.com/benfred/venn.js>.

*Visual Vocabulary*. URL: <https://ft-interactive.github.io/visual-vocabulary/> (cit. a p. 73).