

ALMA MATER STUDIORUM · UNIVERSITÀ DI BOLOGNA

SCUOLA DI SCIENZE
Corso di Laurea in Informatica

**Visualizzazione multiplatforma
di modelli geologici
tramite Python**

Relatore:
Prof.
Federico Montori

Presentata da:
Gabriele Buttinoni

Correlatori:
Dott. Maria Rosaria Tondi
Ing. Giampaolo Zerbinato

Sessione: I
Anno Accademico 2023/2024

*No one knows what the future holds,
that's why its potential is **infinite**.
-Rintarō Okabe in Steins;Gate*

El Psy Kongroo

Abstract

La visualizzazione dei dati geologici è fondamentale per comprendere le strutture del sottosuolo e prendere decisioni informate nei settori della geofisica, dell'estrazione mineraria e della gestione delle risorse naturali. Con l'aumento del volume e della complessità dei dati raccolti, si è reso necessario lo sviluppo di strumenti di visualizzazione sempre più avanzati. Questa tesi presenta lo sviluppo di un programma *open-source* basato su Python, progettato per la visualizzazione tridimensionale e multi-piattaforma dei modelli geologici. Utilizzando librerie come **Pandas** e **Plotly**, il programma permette di caricare, elaborare e visualizzare dati geologici tramite modelli bidimensionali e tridimensionali, offrendo interazioni dinamiche come la rotazione, lo zoom e la visualizzazione delle coordinate geografiche dei punti di interesse. Particolare attenzione è stata data all'accessibilità del progetto, con l'implementazione della scala di colori *Viridis* per garantire una corretta percezione anche per le persone affette da daltonismo. Il programma sviluppato rappresenta una soluzione gratuita, flessibile e potente per la visualizzazione dei dati geologici, rendendolo uno strumento utile per ricercatori, studenti e professionisti del settore.

Keywords: *Python, Plotly, Big Data, Data Visualization, Volume Rendering, Geospatial Data, Visualization Tools*

Indice

1	Introduzione	1
1.1	Contesto e rilevanza	1
1.2	Revisione della letteratura	1
1.3	Motivazioni	2
2	Stato dell'Arte della Visualizzazione dei Dati	3
2.1	Introduzione	3
2.2	Tecniche di Visualizzazione dei Dati	3
2.2.1	Ranking dei Dati	4
2.2.2	Evoluzione dei Dati	5
2.2.3	Distribuzione dei Dati	7
2.2.4	Mappe e Flusso dei Dati	8
2.2.5	Correlazione dei Dati	10
2.3	Strumenti di Visualizzazione	12
2.3.1	Librerie di Visualizzazione	12
2.3.2	Strumenti per la Visualizzazione Interattiva	13
2.3.3	Piattaforme Web-Based	13
2.4	Sfide Attuali nella Visualizzazione dei Dati	13
2.4.1	Gestione di grandi volumi di dati e scalabilità	14
2.4.2	Percezione umana	14
2.4.3	Automazione	15
2.4.4	<i>User-oriented</i>	15
2.4.5	Privacy e sicurezza	16
2.4.6	Nuove applicazioni e domini emergenti	16
2.5	Visualizzazione dei dati nelle geoscienze	16
2.5.1	Tecnologie di Visualizzazione	16
2.5.2	Metodi di Visualizzazione	17
2.5.3	Applicazioni nella Ricerca e nell'Industria	17

3	Panoramica delle soluzioni esistenti	19
3.1	Matlab	19
3.2	ArcGIS	20
3.3	Geosoft Oasis montaj	22
3.4	Alternative open-source	23
3.4.1	QGIS	23
3.4.2	Paraview	24
3.4.3	GNU Octave	26
3.4.4	Python con librerie di visualizzazione	27
4	Implementazione della visualizzazione	29
4.1	Introduzione	29
4.2	Funzionamento del programma	29
4.2.1	Caricamento e pre-elaborazione dei dati	30
4.2.2	Creazione del modello	30
4.3	Accessibilità del progetto sviluppato	35
4.3.1	Problemi delle scale dei colori per il daltonismo	35
4.3.2	Utilizzo della scala di colori <i>Viridis</i>	36
4.4	Vantaggi rispetto alle alternative esistenti	37
5	Conclusioni con possibili sviluppi futuri	39
5.1	Possibili sviluppi futuri	40
5.1.1	Supporto per ulteriori tipi di dati	40
5.1.2	Implementazione di ulteriori modelli	40
5.1.3	Ottimizzazione delle prestazioni	40
5.1.4	Interfaccia utente migliorata	41
5.1.5	Visualizzazione di modelli su webserver	41

Capitolo 1

Introduzione

La visualizzazione dei dati geologici è fondamentale per comprendere le strutture del sottosuolo e prendere decisioni informate nei settori della geofisica, dell'estrazione mineraria e della gestione delle risorse naturali. Negli ultimi decenni, il volume e la complessità dei dati raccolti hanno richiesto lo sviluppo di strumenti di visualizzazione sempre più avanzati. Questo progetto si inserisce in questo contesto, con l'obiettivo di fornire uno strumento open-source, accessibile e potente per la visualizzazione interattiva dei dati geologici.

1.1 Contesto e rilevanza

La geofisica è una disciplina che si basa in modo preponderante sull'analisi dei dati per comprendere i processi e le strutture del sottosuolo. La visualizzazione dei dati geologici non solo aiuta a interpretare meglio le informazioni raccolte, ma facilita anche la comunicazione dei risultati a una vasta gamma di *stakeholder*, inclusi scienziati, ingegneri e responsabili delle decisioni. L'importanza della visualizzazione è cresciuta con l'aumento della complessità dei modelli geologici e della quantità di dati disponibili grazie a tecniche di acquisizione avanzate, come la sismica tridimensionale e le indagini geofisiche ad alta risoluzione.

1.2 Revisione della letteratura

Studi recenti hanno dimostrato come le tecniche di visualizzazione avanzate possano migliorare significativamente l'interpretazione dei dati sismici, gravimetrici e magnetometrici. Ad esempio, la visualizzazione volumetrica e la realtà aumentata sono state utilizzate per esplorare modelli complessi del sottosuolo, offrendo nuove prospettive e intuizioni [16]. Altri studi hanno evidenziato l'importanza di strumenti interattivi che

permettono agli utenti di manipolare i dati in tempo reale per esplorare diverse ipotesi e scenari [18].

La visualizzazione volumetrica è particolarmente utile per rappresentare dati tridimensionali, permettendo agli utenti di esplorare e analizzare le strutture interne dei modelli geologici. Ad esempio, i dati sismici possono essere visualizzati come volumi che mostrano variazioni di velocità o densità all'interno della Terra, offrendo una comprensione più dettagliata delle strutture sotterranee.

La realtà aumentata (AR) è un'altra tecnologia emergente che sta rivoluzionando la visualizzazione geologica. L'AR consente agli utenti di sovrapporre modelli geologici virtuali sull'ambiente reale, migliorando l'analisi sul campo e facilitando la comunicazione dei risultati a un pubblico più ampio.

1.3 Motivazioni

Le motivazioni sottostanti la creazione di questo programma sono molteplici. In primo luogo, esiste una necessità crescente di strumenti di visualizzazione accessibili e gratuiti che possano essere utilizzati da studenti e ricercatori con risorse economiche limitate. In secondo luogo, il programma mira a offrire una maggiore flessibilità rispetto alle soluzioni proprietarie, consentendo agli utenti di personalizzare le visualizzazioni e integrare facilmente altri strumenti di analisi. Infine, l'obiettivo a lungo termine è quello di creare un software ricco e completo in grado di fornire tutti gli strumenti di base necessari a uno studioso per visualizzare e interpretare le misurazioni effettuate.

Il programma elaborato può essere utilizzato su diverse piattaforme, inclusi sistemi operativi come Windows, macOS e Linux, nonché su macchine ad alte prestazioni (HPC - High Performance Computing), il che lo rende uno strumento versatile e adatto a una vasta gamma di applicazioni geofisiche. L'utilizzo di Python come linguaggio di programmazione principale offre ulteriori vantaggi in termini di flessibilità e integrazione con altre librerie di analisi dei dati, come *NumPy*, *SciPy* e *Pandas*.

In conclusione, questo progetto rappresenta un passo avanti significativo nella visualizzazione dei dati geologici, fornendo uno strumento potente e accessibile che può essere utilizzato da una vasta gamma di utenti per visualizzare al meglio le formazioni del sottosuolo e abbassare al massimo la soglia di difficoltà necessaria ad ottenere modelli completi.

Capitolo 2

Stato dell'Arte della Visualizzazione dei Dati

2.1 Introduzione

La visualizzazione dei dati è un campo cruciale nell'analisi dei dati, che permette di rappresentare graficamente le informazioni per facilitare la comprensione e l'interpretazione. Con l'aumento esponenziale dei dati generati in vari settori, la necessità di tecniche di visualizzazione efficaci è diventata più evidente. Questo capitolo fornisce una panoramica sulle tecniche, gli strumenti e le sfide attuali della visualizzazione dei dati, oltre a discutere le direzioni future e le opportunità di ricerca.

2.2 Tecniche di Visualizzazione dei Dati

Le tecniche di visualizzazione dei dati sono essenziali per rappresentare in modo efficace le informazioni, riuscendo sia a veicolare più informazioni rispetto alle *summary statistics*, sia (spesso) a renderle accessibili anche a un pubblico non esperto. Un importante aspetto da tenere in considerazione è come vengono percepite le visualizzazioni dall'occhio umano. Infatti, come evidenziano Franconeri et al. [9] esiste un ordinamento di precisione secondo il quale il visualizzatore è in grado di indicare verbalmente il rapporto tra due valori (figura 2.1): i metodi più efficaci sono posizione, lunghezza e area, mentre angoli e intensità vengono distinti con maggiore difficoltà.

Le tecniche di visualizzazione possono essere classificate in diverse categorie, tra cui distribuzione dei dati, correlazione, ranking, evoluzione, mappe e flusso dei dati.

Per scrivere i successivi paragrafi dedicati alle diverse tipologie di grafici esistenti ci affideremo al libro “*Better Data Visualizations: a Guide for Scholars, Researchers and Wonks*” di Jonathan Schwabish [34].

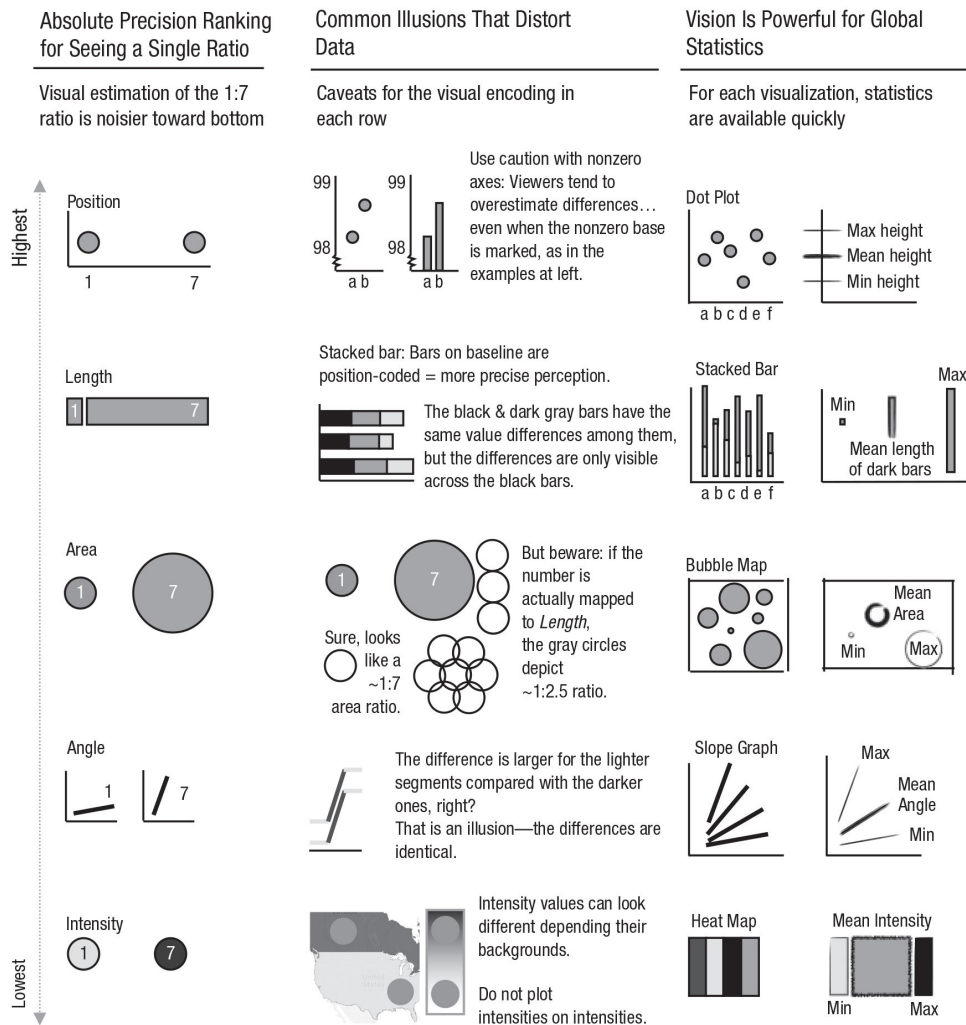


Figura 2.1: Classifica di precisione [9]

2.2.1 Ranking dei Dati

La visualizzazione del ranking dei dati viene spesso realizzata tramite grafici a barre e a colonne, che mostrano la classifica degli elementi in base a vari criteri. Questi grafici sono utili per confrontare categorie o entità e per evidenziare le differenze di performance. Ad esempio, un grafico a barre può essere utilizzato per mostrare le vendite mensili di diversi prodotti, facilitando l'identificazione dei *best-seller*. Questa tecnica aiuta a prendere decisioni strategiche basate su confronti visivi chiari e immediati.

L'orientamento del grafico può essere indifferentemente orizzontale o verticale, mentre la scelta di posizionare le barre su un cerchio (*radial layout*) è da evitare perché porta a una percezione distorta. Molti esperti di visualizzazione dati concordano sulla necessità

di iniziare gli assi a 0, in modo da non enfatizzare in modo eccessivo le differenze tra i dati e distorcerne la percezione (figura 2.2). Inoltre, l'utilizzo delle linee della griglia, se realizzate con colori poco accesi in modo da non distrarre l'attenzione del visualizzatore dai dati, sono un utile strumento di aiuto, soprattutto per interpretare i dati posti più lontano dalle etichette degli assi.

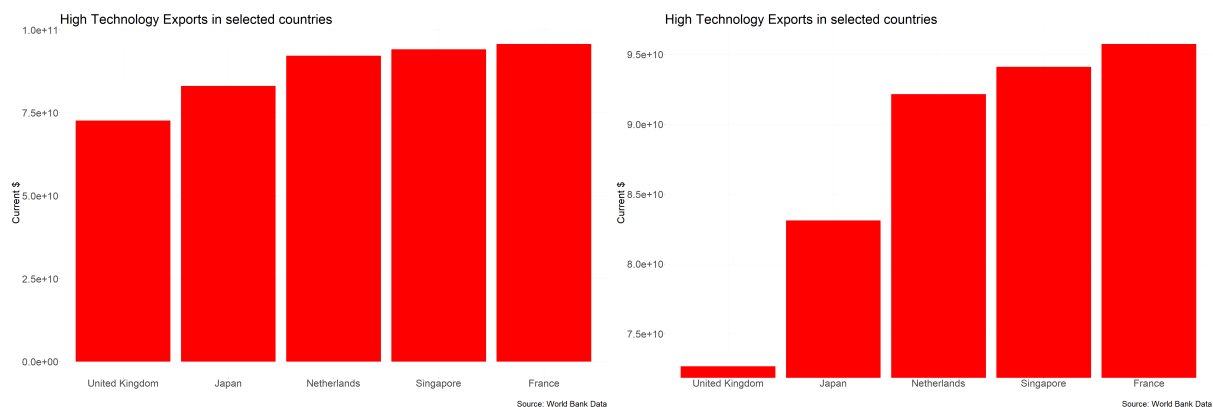


Figura 2.2: Grafici a barre con asse che parte da 0 e non

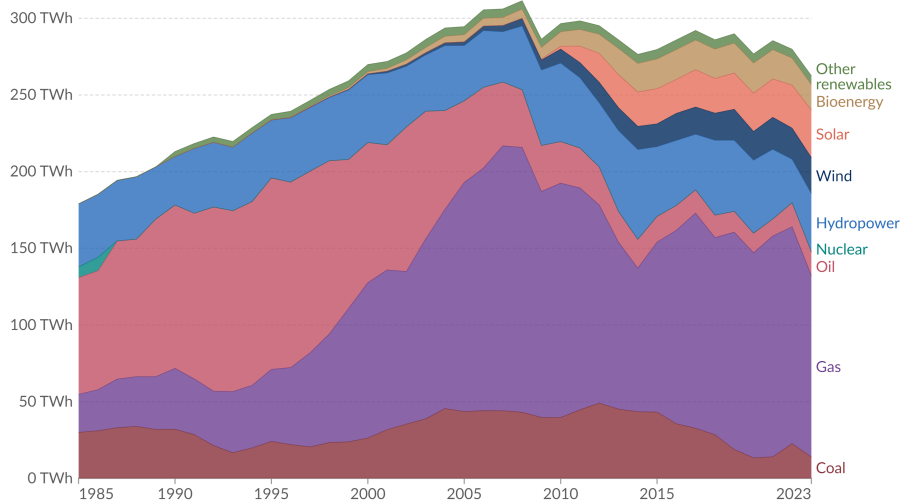
2.2.2 Evoluzione dei Dati

Le tecniche di visualizzazione per l'evoluzione dei dati, come le *line charts* e i grafici ad area, sono utilizzate per mostrare i cambiamenti nel tempo. Questi strumenti sono fondamentali per analizzare le tendenze temporali e per prevedere sviluppi futuri. Ad esempio, un grafico a linee può rappresentare l'andamento delle temperature giornaliere su un periodo di anni, evidenziando pattern stagionali o anomalie climatiche. Visualizzare l'evoluzione dei dati nel tempo è essenziale per monitorare i progressi e per adattare le strategie in base alle tendenze osservate.

Nei grafici ad area, che possono essere pensati come dei grafici a barre con barre infinitamente sottili, è preferibile iniziare gli assi a 0. Questo tipo di grafico è particolarmente utilizzato nel formato *stacked*, nel quale vengono visualizzate varie serie temporali che vengono sommate. Un esempio applicativo può essere la visualizzazione della produzione elettrica nel tempo disaggregata per tipo di fonte (figura 2.3)

Electricity production by source, Italy

Measured in terawatt-hours¹.



Data source: Ember (2024); Energy Institute - Statistical Review of World Energy (2023) OurWorldInData.org/energy | CC BY
Note: "Other renewables" include waste, geothermal, wave, and tidal.

Figura 2.3: Esempio di grafico ad area [32]

Nei grafici a linee non è necessario iniziare gli assi a 0, ma è invece fondamentale la scelta della scala sull'asse verticale, che porta a percezioni molto diverse dei cambiamenti e dovrebbe essere personalizzata in base al contesto (figura 2.4).

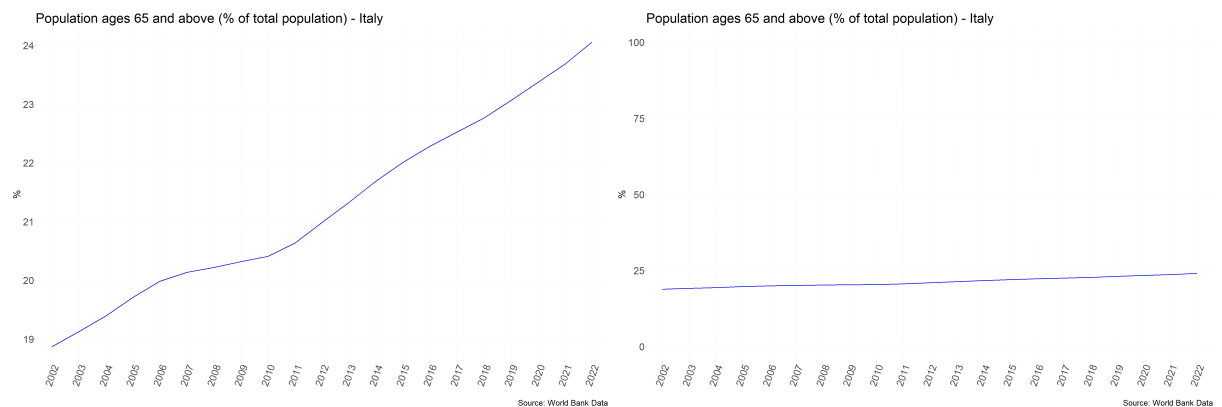


Figura 2.4: Grafici a linee con asse verticale differente

In caso si voglia comparare l'andamento del tempo di più variabili con unità di misura differenti è preferibile non adottare i grafici *dual-axis*, perchè di difficile lettura e perchè il punto di incontro tra le linee, anche se privo di significato, diventa il punto focale del grafico.

Anche in questo caso esiste la possibilità di adottare una struttura circolare, più difficile da interpretare ma di grande impatto visivo, se non si è interessati ai valori specifici di ogni istante temporale. Un esempio significativo è la *Climate spiral* della NASA [24], che applica questo tipo di grafico alle anomalie globali delle temperature rispetto al periodo 1951-1980.

2.2.3 Distribuzione dei Dati

Le tecniche di visualizzazione per la distribuzione dei dati includono istogrammi e box-plot, che aiutano a esplorare la relazione tra variabili numeriche. Questi strumenti permettono di visualizzare la variabilità e la distribuzione dei dati (simmetrica oppure *skewed*), identificando rapidamente tendenze e outliers.

Gli istogrammi sono tipi particolari di grafici a barre nei quali la variabile in analisi viene suddivisa in intervalli e l'altezza (o lunghezza) delle barre mostra la frequenza assoluta dei dati. In questo tipo di grafico è fondamentale la scelta del numero di intervalli da utilizzare (figura 2.5): un numero troppo basso può nascondere i pattern della distribuzione, mentre uno troppo elevato può rendere difficoltosa l'interpretazione della distribuzione complessiva.

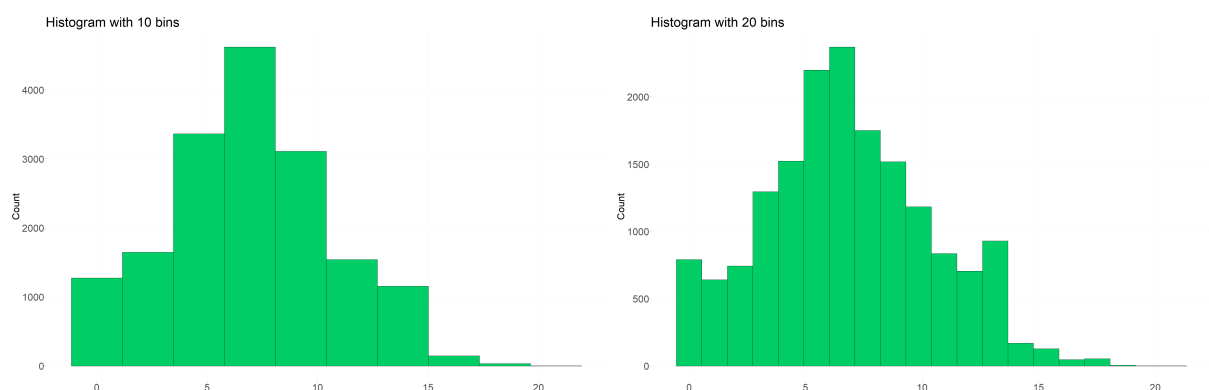


Figura 2.5: Istogrammi con un diverso numero di intervalli

I box-plot (figura 2.6) sono grafici molto sintetici che mostrano solo alcuni specifici punti della distribuzione e sono generalmente più adatti a un pubblico esperto. Questo tipo di grafico, può avere un orientamento sia verticale sia orizzontale; per semplicità descriviamo la sua versione verticale. Consiste in un rettangolo i cui lati inferiore e superiore rappresentano rispettivamente il primo e il terzo quartile della distribuzione; una linea orizzontale all'interno indica la mediana. Due linee (*whiskers*) connettono i lati orizzontali del rettangolo ai valori più estremi, che generalmente vengono posti a 1.5 volte il *range* interquartile (differenza tra primo e terzo quartile). I valori più estremi, detti *outliers*, vengono rappresentati con dei punti. La questione principale da tenere a

mente se si sceglie di utilizzare questo tipo di visualizzazione è che non viene visualizzata l'intera distribuzione: bisogna quindi essere certi di non stare omettendo informazioni rilevanti [34].

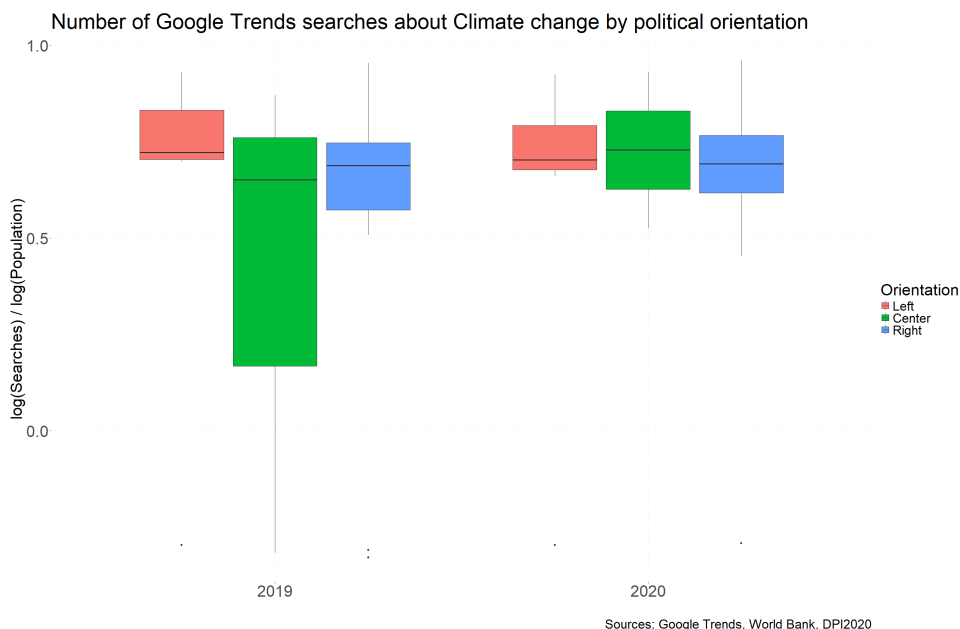


Figura 2.6: Esempio di box-plot

2.2.4 Mappe e Flusso dei Dati

Le mappe geografiche e i diagrammi di flusso sono strumenti potenti per visualizzare i dati spaziali e il flusso delle informazioni.

Le mappe geografiche più diffuse sono le cosiddette *choropleth maps* (figura 2.7), che usano colori, sfumature o pattern sulle unità geografiche per mostrare quantità o intensità. Il problema principale legato a questo tipo di rappresentazione è la difficoltà nel trovare stati di piccole dimensioni, soprattutto se i dati sono a livello mondiale. Inoltre, la dimensione geografica dei vari paesi non ha alcun legame con l'importanza dei dati rappresentati, il che può indurre una distorsione a livello di interpretazione. Nonostante possa venire istintiva l'idea di creare una mappa quando si hanno dei dati geografici, in assenza di evidenti pattern geografici, questo tipo di visualizzazione potrebbe risultare di scarsa utilità.

MDG4.A: Child mortality rate, 2020



Millennium Development Goal (MDG) 4.A was to reduce global child mortality (children under 5 years old) rates by two-thirds from 1990 to 2015. The global target was therefore a reduction to 3 deaths per 100 live births by 2015. This target was missed.

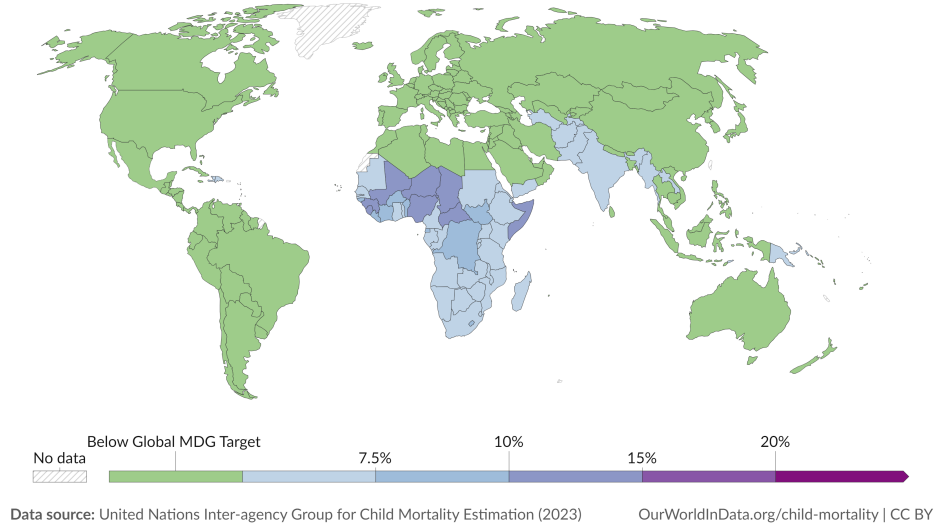


Figura 2.7: Esempio di mappa geografica con un chiaro trend spaziale [6]

I diagrammi di flusso (figura 2.8), invece, rappresentano i movimenti e le interazioni tra diversi elementi, come il flusso di materiali in una catena di approvvigionamento. Questi strumenti aiutano a comprendere la dimensione spaziale e dinamica dei dati, mettendo in luce i vari *steps* che compongono un processo e facilitando l'identificazione di pattern e anomalie. Nei diagrammi di flusso, l'uso delle forme permette di differenziare le varie fasi in base a cosa rappresentano: ad esempio, possono essere utilizzati di default dei rettangoli, dei rombi per i punti focali e degli ovali per indicare l'inizio e la fine del processo.

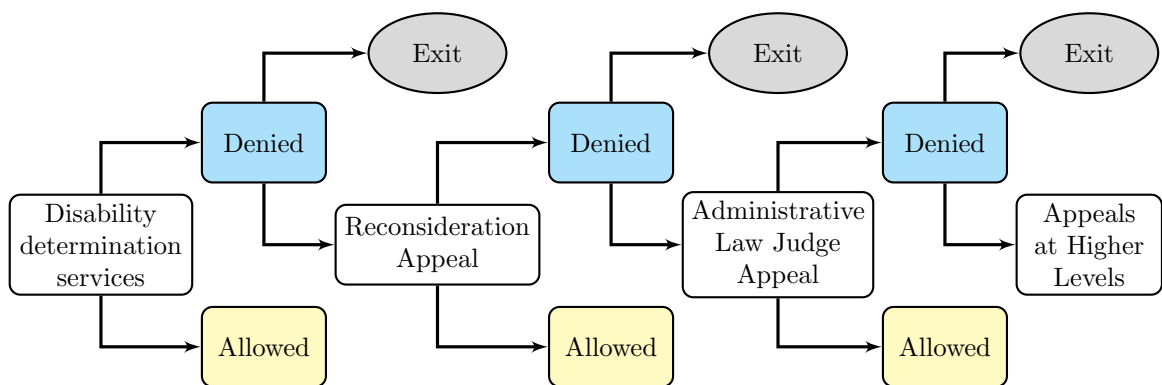


Figura 2.8: Esempio di diagramma di flusso [34]

2.2.5 Correlazione dei Dati

Le tecniche per la correlazione dei dati, come i grafici a dispersione (*scatterplot*) e le mappe di calore (*heatmap*), consentono di visualizzare le relazioni tra variabili diverse. Questi strumenti aiutano a identificare se esiste una relazione positiva, negativa o neutra tra le variabili.

I grafici a dispersione (figura 2.9) sono il tipo di visualizzazione più diffusa e sono facilmente interpretabili anche da un pubblico non esperto: un'indagine del Pew Research Center del 2015 [13] ha evidenziato che il 63% dei rispondenti era in grado di interpretare questo tipo di grafico in modo corretto. In questa visualizzazione una variabile è rappresentata sull'asse orizzontale e una su quello verticale. Se le due variabili si muovono nella stessa direzione, ovvero se all'aumentare dell'una aumenta anche l'altra, allora sono legate da una correlazione positiva; se quando una aumenta l'altra diminuisce la correlazione è negativa. Se non esiste alcun pattern evidente le due variabili non sono correlate. Per rendere il grafico più facilmente interpretabile può essere inserita una linea di regressione che mostra la direzione generale della relazione.

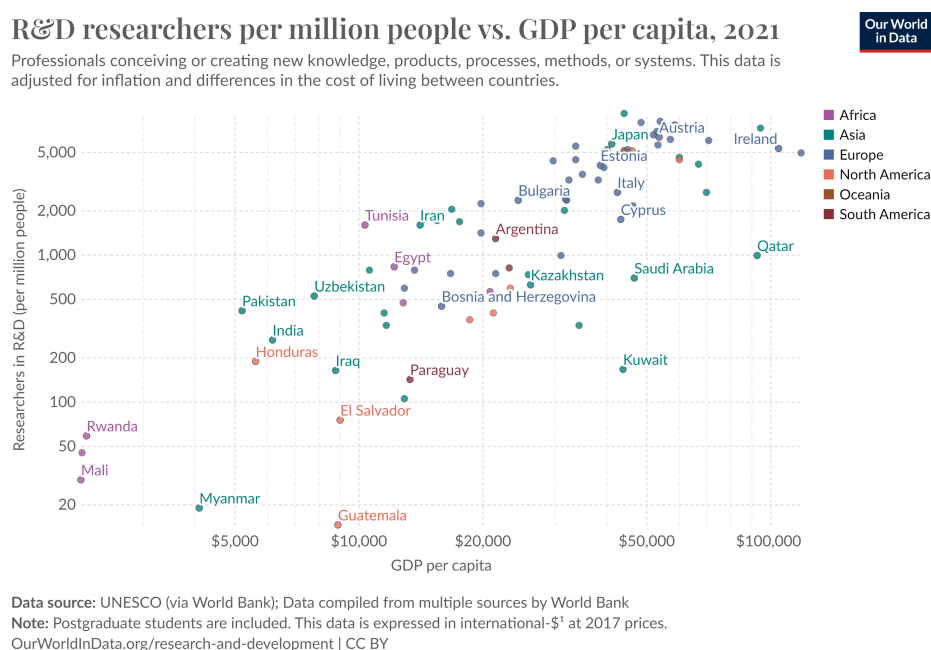


Figura 2.9: Esempio di grafico a dispersione con correlazione positiva [31]

Se si vuole analizzare la correlazione tra più di due variabili (purché siano di tipo quantitativo) è consigliabile affidarsi a una matrice di correlazione. Le celle di questa matrice rappresentano l'intensità della correlazione tra le coppie di variabili, spesso misurata tramite il coefficiente di correlazione di Pearson. La visualizzazione grafica di questa

matrice può avvenire tramite una mappa di calore (figura 2.10), dove alle colorazioni più intense corrispondono relazioni più forti tra le variabili: generalmente le correlazioni positive vengono rappresentate con gradazioni di rosso mentre quelle negative con gradazioni di blu.

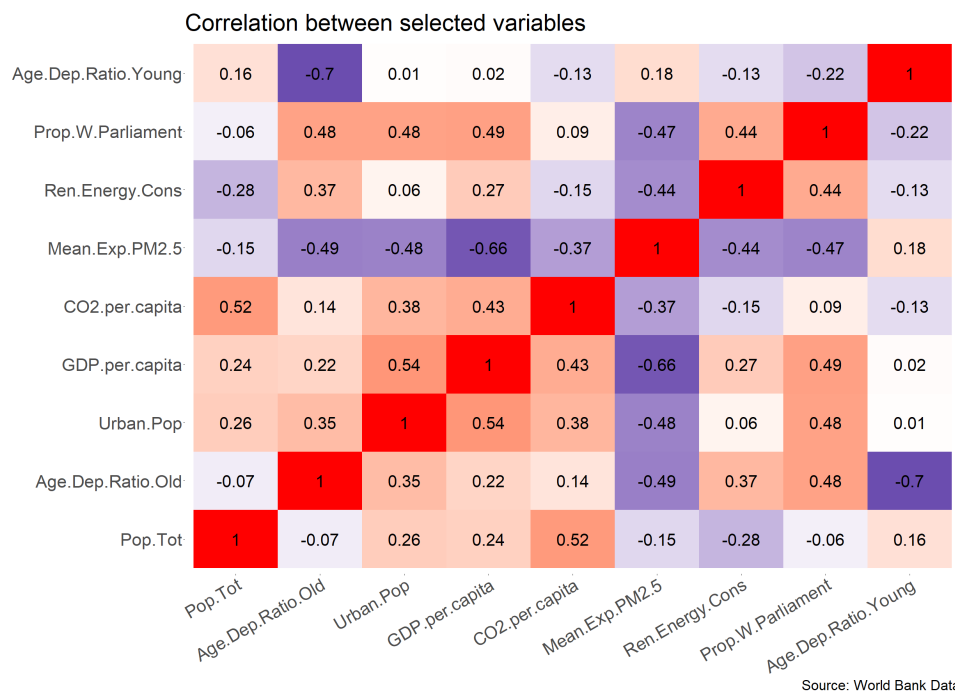


Figura 2.10: Esempio di *heatmap*

2.3 Strumenti di Visualizzazione

Gli strumenti di visualizzazione sono fondamentali per trasformare i dati grezzi in rappresentazioni grafiche informative. Esistono numerose librerie e piattaforme che supportano la visualizzazione dei dati, che si differenziano principalmente in base alla loro flessibilità e semplicità di utilizzo (figura 2.11). In seguito analizziamo alcune delle soluzioni più diffuse.

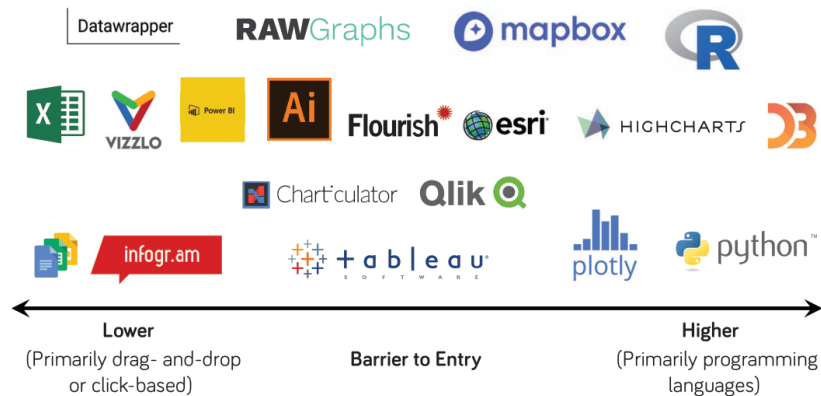


Figura 2.11: Classificazione degli strumenti di visualizzazione per semplicità di utilizzo

2.3.1 Librerie di Visualizzazione

I linguaggi di programmazione open-source più utilizzati per creare visualizzazioni di dati sono R, Javascript e Python.

R è un linguaggio di programmazione specificamente progettato per analisi statistiche. La libreria `ggplot2` [11, 4] permette di creare diversi tipi di grafici a partire da tre componenti fondamentali e necessariamente presenti: un dataset, un sistema di coordinate e una geometria, che definisce il modo in cui verranno visualizzati i valori (e quindi il tipo di grafico desiderato). La componente geometrica può essere personalizzata, definendo ad esempio dimensioni e colori attraverso il parametro “*aesthetics*”.

Javascript è un linguaggio di programmazione nato per creare pagine web animate. La libreria `d3.js` [25] è estremamente flessibile essendo di basso livello: infatti, a differenza delle librerie di *charting*, le visualizzazioni devono essere create a partire dalle componenti più elementari. Per questo motivo approcciarsi a questo tipo di visualizzazione è particolarmente difficoltoso.

Python è un linguaggio di programmazione che può essere applicato a molteplici domini. La libreria `matplotlib` [22] permette di visualizzare i dati in una figura costituita da uno o più assi e un'area nella quale i punti possono essere specificati per mezzo di coordinate. La libreria `seaborn` [35] permette di effettuare operazioni statistiche su *dataframes* a *array*, basandosi poi su `matplotlib` per disegnare i grafici.

2.3.2 Strumenti per la Visualizzazione Interattiva

Strumenti come Tableau, Power BI, e Plotly offrono interfacce intuitive per creare visualizzazioni interattive.

Tableau [37] permette di creare e condividere visualizzazioni interattive in modo semplice e intuitivo. Infatti, è basato sul *drag-and-drop* e quindi non necessita dell'apprendimento di un linguaggio specifico per poterlo utilizzare, il che lo rende adatto a ogni tipo di utente. È una piattaforma a pagamento, ma ne esiste anche una versione gratuita con poche limitazioni: Tableau Public.

Power BI [29] permette di creare visualizzazioni e *dashboard* in modo simile a Tableau. È di proprietà di Microsoft e perciò si integra con gli altri programmi del pacchetto Office, in particolare Excel. Anche in questo caso esistono sia una versione gratuita sia diverse opzioni a pagamento.

Plotly [28] è una libreria gratis e *open-source* di Python, che consente di creare grafici interattivi che possono essere incorporati in applicazioni web.

2.3.3 Piattaforme Web-Based

Le piattaforme web-based, come Google Charts e Datawrapper, permettono di condividere e interagire con le visualizzazioni online.

Google Charts [15] utilizza la tecnologia HTML5 e SVG e permette di visualizzare i dati nelle pagine web tramite un gran numero di grafici interattivi pronti all'uso. Può essere quindi considerato una valida alternativa di più semplice impiego all'utilizzo della libreria *d3.js*.

Datawrapper [5] è una piattaforma online gratuita (esistono anche opzioni a pagamento) che permette di creare grafici, tabelle e mappe interattivi e personalizzabili. È pensato anche per utenti senza esperienza di programmazione, il che lo rende ideale per giornalisti e professionisti che devono presentare dati in modo efficace e rapido. Permette inoltre di collegare le visualizzazioni a fonti di dati dinamiche, senza il bisogno di modificare manualmente il grafico.

2.4 Sfide Attuali nella Visualizzazione dei Dati

Nonostante i progressi significativi, ci sono diverse sfide che devono essere affrontate nel campo della visualizzazione dei dati. “*The Big Data Visual Exploration and Analytics workshop*” (BigVis) è un *workshop* con cadenza annuale, che dal 2018 si occupa di affrontare le più recenti sfide e problematiche legate all'esplorazione e all'analisi dei dati. In particolare, l'edizione del 2020 ha previsto una serie di contributi sulla visualizzazione dei dati da parte di esperti del settore [1]. Nei paragrafi successivi cercheremo di riassumere quelle che sono state evidenziate come le principali sfide e i possibili sviluppi nella

visualizzazione dei dati, mettendo in luce i punti di vista dei diversi autori partecipanti al *workshop*¹.

2.4.1 Gestione di grandi volumi di dati e scalabilità

I moderni sistemi di visualizzazione devono essere in grado di gestire dataset con miliardi di oggetti, ma allo stesso tempo fornire una risposta visiva in pochi millisecondi.

Al giorno d'oggi, la maggior parte dei dataset non solo è multidimensionale e affetta da “rumore”, ma è anche multiutente e viene aggiornata di frequente: come afferma Michael Stonebraker pensare ai dati come qualcosa di statico è molto pericoloso. Ciò si collega alla questione affrontata da Stratos Idreos, che vede la visualizzazione dei dati come una problematica *data-intensive* (opposta a *computer-intensive*). In quest'ottica, la scelta del corretto metodo di archiviazione per il tipo di analisi che si vuole effettuare è fondamentale. Studiare il design di strutture di dati che possano sia rispondere alle nuove richieste di accesso sia fornire risposte interattive in poco tempo è perciò un'enorme opportunità per la *community* di *data management*. Un metodo di archiviazione più efficace potrebbe superare i limiti di piattaforme come Tableau e Power BI, che nel caso di dataset di grandi dimensioni riducono fortemente la libertà di azione dell'utilizzatore (Tim Kraska).

La sfida più rilevante è la scalabilità, che in questo ambito consiste nella capacità di assicurare l'interattività anche nel momento in cui si utilizzano grandi volumi di dati. Una possibile soluzione, proposta da Jean-Daniel Fekete, è la computazione progressiva, che consiste nel suddividere le operazioni in una serie di brevi *chunks* che producono stime sempre più precise anziché eseguire tutto in un unico *step* che potrebbe richiedere un tempo di esecuzione eccessivamente lungo.

2.4.2 Percezione umana

Come abbiamo citato nel capitolo relativo alle tecniche di visualizzazione dei dati, è di fondamentale importanza di comprendere la percezione visiva, l'attenzione e la comprensione umane. In particolare, anziché affidarsi a degli esperimenti, come viene spesso fatto nel campo della visualizzazione, la conoscenza di questi aspetti può fare la differenza tra una visualizzazione di impatto e una di scarsa rilevanza. Questo aspetto diventa sempre più rilevante all'aumentare della complessità del dataset. Antti Oulasvirta propone perciò di usare modelli presi in prestito dalla psicologia per adattare le visualizzazioni agli individui per le quali sono pensate.

¹Alla presentazione del “*BigVis2000 Challenges Report*” [1] hanno partecipato: Gennady Andrienko, Natalia Andrienko, Steven Drucker, Jean-Daniel Fekete, Danyel Fisher, Stavros Idreos, Tim Kraska, Guoliang Li, Kwan-Liu Ma, Jock D. Mackinlay, Antti Oulasvirta, Tobias Schreck, Heidrun Schumann, Michael Stonebraker, David Auber, Nikos Bikakis, Panos K. Chrysanthis, George Papastefanatos, Mohamed A. Sharaf

Modelli efficaci da un punto di vista cognitivo possono essere utilizzati, come suggeriscono Gennady e Natalia Andrienko, durante tutto il processo analitico per supportare l'utente ogni volta che deve prendere una decisione.

2.4.3 Automazione

La maggior parte degli autori concorda che le visualizzazioni dei dati dovrebbero diventare maggiormente automatizzate, suggerendo, ad esempio, quali tipi di visualizzazioni utilizzare. In generale, il programma utilizzato dovrebbe fornire all'utente sufficiente supporto nel processo di analisi, supportando visualizzazioni efficienti e personalizzabili.

All'aumento delle dimensioni e dell'eterogeneità dei *dataset*, gli utenti possono incontrare delle difficoltà nell'esplorare i dati in modo efficace e impiegare più tempo a raggiungere i loro obiettivi. Tobias Schreck fa notare come la necessità e il potenziale di una interazione assistita siano concetti evidenti, la cui implementazione, però, è tutt'altro che intuitiva. Uno dei possibili ostacoli al processo di automazione è la grande quantità di metodi di analisi utilizzabili nei vari campi di ricerca, che rendono difficile un possibile processo di unificazione (Heidrun Schumann).

2.4.4 *User-oriented*

Gennady e Natalia Andrienko affermano che al giorno d'oggi le tecnologie sono già così semplici da utilizzare che chiunque può analizzare i dati. Questo può portare a un errato utilizzo delle visualizzazioni a causa della scarsa conoscenza dell'intero processo di analisi.

Allo stesso tempo, però, la maggior parte degli esperti concorda sulla necessità di rendere le applicazioni più *user-oriented*. Ciò permetterebbe gli utenti poco esperti di conoscere le proprietà e la complessità dei dati in modo più naturale e intuitivo. Per raggiungere questo scopo, Danyel Fisher manifesta la necessità di capire quali siano gli effettivi bisogni degli utenti, quali domande potrebbero porre e quale sia il livello di precisione desiderato.

Interfacce utente capaci di supportare visualizzazioni efficaci permetterebbero agli utenti di concentrarsi sulle problematiche che necessitano attenzione umana e capacità decisionale, lasciando al computer le elaborate analisi statistiche. In questo modo si integrerebbero in modo ottimale le abilità complementari di uomo e macchina (Jock D. Mackinlay).

Tim Kraska fa notare come, però, venga spesso sviluppato per primo il sistema e solo in un secondo momento ci si concentri sull'interfaccia, mentre pensando il processo all'opposto si potrebbe costruire un sistema che supporti effettivamente l'utente.

2.4.5 Privacy e sicurezza

La protezione della privacy e la sicurezza dei dati sono questioni cruciali, specialmente quando si lavora con dati sensibili, anche se, come evidenzia Guoliang Li, la maggior parte dei sistemi di visualizzazione non ne tiene conto. La visualizzazione dei dati deve garantire che le informazioni personali e sensibili siano adeguatamente protette. Ciò include l'implementazione di tecniche di anonimizzazione e la gestione sicura dei dati. Inoltre, è importante considerare le implicazioni etiche della visualizzazione dei dati, assicurandosi che le visualizzazioni non possano essere utilizzate in modo improprio o portare a interpretazioni errate che potrebbero danneggiare gli individui o le organizzazioni.

2.4.6 Nuove applicazioni e domini emergenti

L'innovazione nelle tecniche di visualizzazione e lo sviluppo di nuovi strumenti sono fondamentali per migliorare l'efficacia delle rappresentazioni grafiche. Le tecnologie emergenti come la realtà aumentata e virtuale offrono nuove modalità per interagire con i dati, permettendo agli utenti di esplorare le informazioni in modo immersivo.

Inoltre, come suggerisce Steven Drucker, l'applicazione della *data visualization* al *machine learning* è una grande sfida, che permetterebbe di migliorare l'interpretazione dei risultati.

In generale, la visualizzazione dei dati viene sempre più applicata a qualsiasi campo, da quelli di maggiore rilevanza sociale, come il cambiamento climatico, la sanità o l'educazione, al contesto industriale. Un'applicazione innovativa viene presentata da Guoliang Li ed è nel *debugging*: le visualizzazioni possono aiutare a trovare *bug* nel *workflow* (*bug-debugging*) oppure errori nei risultati finali (*data debugging*).

2.5 Visualizzazione dei dati nelle geoscienze

La visualizzazione dei dati ha assunto un ruolo cruciale nelle geoscienze, una disciplina che analizza fenomeni complessi e multidimensionali. Con l'avvento delle tecnologie moderne, la capacità di interpretare grandi volumi di dati geospaziali e temporali è diventata essenziale per comprendere processi naturali e antropogenici. Questo capitolo esplorerà lo stato dell'arte della visualizzazione dei dati nelle geoscienze, discutendo le tecnologie, i metodi e le applicazioni principali che stanno rivoluzionando il campo.

2.5.1 Tecnologie di Visualizzazione

Negli ultimi anni, le tecnologie di visualizzazione si sono evolute rapidamente, offrendo strumenti sempre più sofisticati per l'analisi dei dati geoscientifici. I Sistemi di Informazione Geografica (GIS) sono tra questi strumenti fondamentali. I GIS consentono la manipolazione e l'interpretazione di dati georeferenziati e sono utilizzati in svariati

ambiti, dalla geologia alla gestione delle risorse idriche, facilitando la visualizzazione di mappe tematiche e l'analisi spaziale avanzata [21, 3]. Un'altra tecnologia chiave sono i Modelli Digitali del Terreno (DTM), che rappresentano la superficie terrestre in forma tridimensionale. Utilizzati per analisi topografiche e geomorfologiche, i DTM offrono una visione dettagliata della morfologia del terreno e sono essenziali per studi sul rischio idrogeologico, erosione e monitoraggio ambientale [8, 19].

Le tecnologie 3D e la Realtà Virtuale (VR) stanno aprendo nuove possibilità per l'immersione nei dati geoscientifici, permettendo di esplorare ambienti complessi in modo interattivo. Queste tecnologie trovano applicazione nella formazione, nella pianificazione urbana e nella gestione delle risorse naturali, offrendo un'esperienza più immersiva e dettagliata nella comprensione dei fenomeni geologici [17, 20].

2.5.2 Metodi di Visualizzazione

I metodi di visualizzazione dei dati geoscientifici variano a seconda della natura dei dati e degli obiettivi dell'analisi. Le mappe tematiche rappresentano variabili specifiche, come la distribuzione delle rocce, i tipi di suolo o la concentrazione di inquinanti. Questo metodo facilita l'interpretazione e permette di comunicare informazioni complesse in modo intuitivo [40]. I grafici e i diagrammi, come istogrammi, scatter plot e diagrammi a torta, sono strumenti utilizzati per rappresentare dati quantitativi e qualitativi, offrendo una visualizzazione chiara delle relazioni tra variabili e supportando l'identificazione di pattern e trend [41].

Un altro metodo significativo è rappresentato dalle animazioni, che permettono di visualizzare i cambiamenti temporali nei dati, come l'evoluzione dei fenomeni climatici o l'attività sismica. Le animazioni offrono una comprensione dinamica dei processi e la possibilità di realizzare simulazioni previsionali, essenziali per l'analisi e la gestione di fenomeni naturali complessi [27, 2].

2.5.3 Applicazioni nella Ricerca e nell'Industria

La visualizzazione dei dati nelle geoscienze trova applicazione in numerosi settori, contribuendo a una migliore comprensione e gestione delle risorse naturali. Nella geologia e nell'esplorazione mineraria, la visualizzazione avanzata aiuta nella mappatura geologica e nella localizzazione di giacimenti minerari, aumentando l'efficienza esplorativa e riducendo i costi e i rischi operativi [36]. Nel campo della climatologia e degli studi ambientali, strumenti di visualizzazione sono utilizzati per monitorare cambiamenti climatici, analizzare dati meteorologici e studiare impatti ambientali. Questo consente una migliore previsione dei fenomeni estremi e supporta le politiche ambientali [14].

Anche la gestione delle risorse idriche beneficia enormemente della visualizzazione dei dati. Visualizzare i dati idrologici è essenziale per la gestione delle risorse idriche e la

prevenzione di alluvioni e siccità, consentendo una pianificazione sostenibile delle risorse e la mitigazione dei disastri naturali.

Capitolo 3

Panoramica delle soluzioni esistenti

Nel campo della geofisica, esistono diversi software proprietari e open-source utilizzati per l'elaborazione e la visualizzazione dei dati. Tra i più noti software proprietari troviamo Matlab, ArcGIS e Geosoft Oasis montaj, mentre le più comuni soluzioni open-source sono QGIS, Paraview, Octave e Python. Ciascuno di questi strumenti presenta punti di forza e limitazioni.

3.1 Matlab

Matlab [38] (figura 3.1) è uno dei software più utilizzati per l'elaborazione dei dati scientifici. Offre una vasta gamma di funzioni integrate per la manipolazione dei dati, la visualizzazione e l'analisi numerica. La documentazione ufficiale di Matlab sottolinea che Matlab è progettato per l'analisi e il calcolo numerico, offrendo un ambiente interattivo e un linguaggio di programmazione di alto livello che permette di eseguire operazioni complesse su matrici e array. Inoltre, Matlab include *toolboxes* specializzati per diverse applicazioni, come l'elaborazione di segnali, l'analisi statistica e l'elaborazione delle immagini.

Un esempio di utilizzo di Matlab è nell'analisi dei dati sismici, dove i *toolboxes* permettono di applicare filtri avanzati e tecniche di inversione per migliorare l'interpretazione dei dati.

- **Costo:** Matlab è un software proprietario con costi di licenza elevati, il che lo rende poco accessibile a studenti e ricercatori con budget limitati, nonostante l'esistenza di piani a prezzi agevolati pensati per la ricerca accademica. Inoltre, per accedere alle *toolboxes*, è richiesto l'acquisto di licenze annuali specifiche.
- **Compatibilità:** Nonostante sia compatibile con la maggior parte dei sistemi operativi alcune funzioni potrebbero richiedere componenti hardware avanzate.

- **Flessibilità:** Può risultare meno flessibile rispetto alle soluzioni open-source in termini di personalizzazione e integrazione con altri strumenti.
- **Metodo di utilizzo:** Matlab richiede una curva di apprendimento significativa, specialmente per utenti nuovi alla programmazione. Gli utenti devono scrivere script complessi per visualizzare dati, il che può rappresentare un ostacolo per chi non ha esperienza di programmazione.

Rispetto a Matlab il software presentato in questo documento risulta essere molto più semplice da usare, in quanto non richiede di scrivere codice, e soprattutto non richiede alcun investimento economico. Inoltre, le funzionalità interattive di Matlab sono molto difficili da implementare, al contrario di quelle di Plotly. Infine, il codice sorgente è totalmente a disposizione dell'utente.

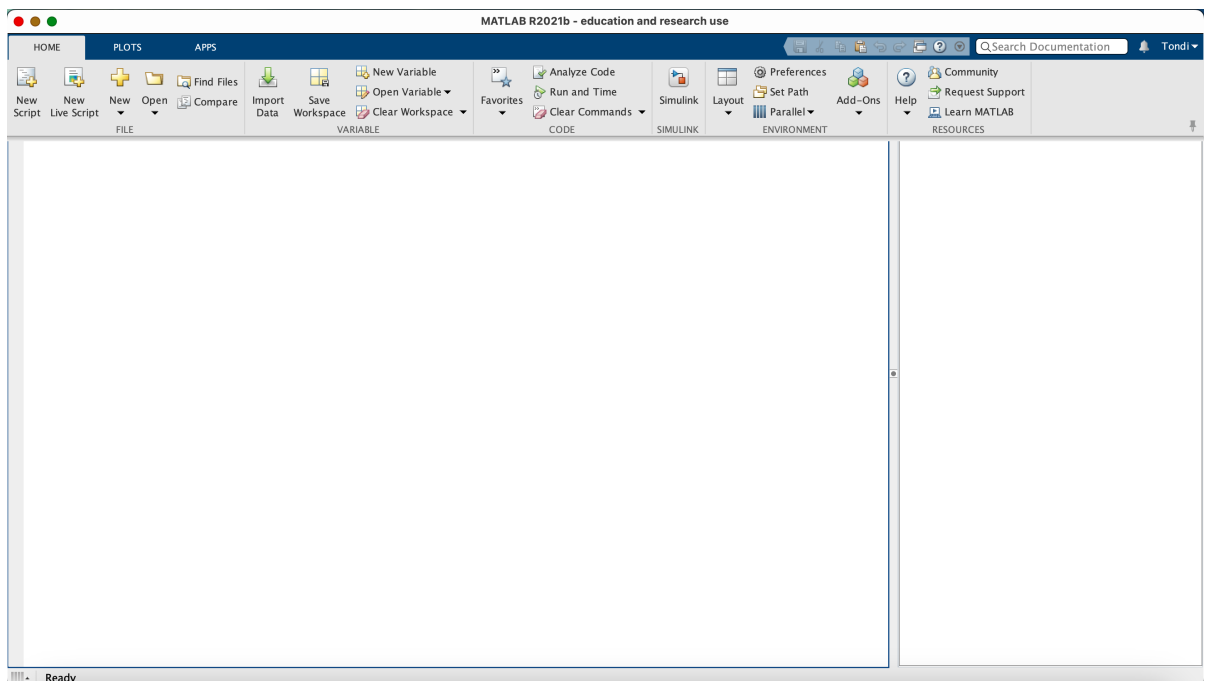


Figura 3.1: Interfaccia MATLAB

3.2 ArcGIS

ArcGIS [7] (figura 3.2) è un software sviluppato da Esri per la creazione e gestione di mappe e dati geospaziali. È ampiamente utilizzato in geologia, geografia e altre discipline correlate. Secondo la documentazione di Esri, ArcGIS fornisce una piattaforma completa per la gestione dei dati geospaziali, dando la possibilità agli utenti di raccogliere,

archiviare, analizzare e visualizzare informazioni spaziali. Inoltre, ArcGIS offre strumenti di *geoprocessing* che permettono di automatizzare i flussi di lavoro e migliorare l'efficienza operativa.

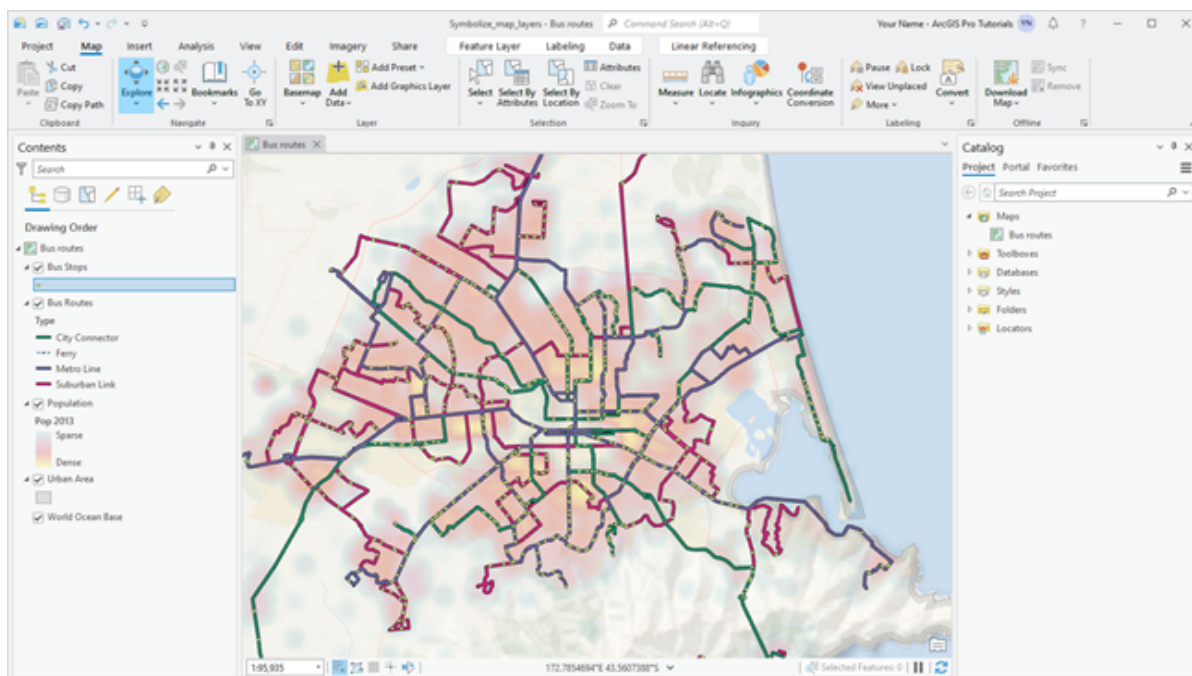


Figura 3.2: Interfaccia ArcGIS

ArcGIS è particolarmente utile per la creazione di mappe dettagliate e per l'analisi spaziale avanzata. Ad esempio, è ampiamente utilizzato per la gestione delle risorse naturali, la pianificazione urbana e la mitigazione dei rischi ambientali.

- **Costo:** Come Matlab, ArcGIS è un software costoso, con licenze che possono rappresentare un investimento significativo. Nonostante la presenza di piani più accessibili, l'utilizzo di molte funzioni richiede il possesso di licenze più costose. ArcGIS fornisce anche funzionalità avanzate online accessibili tramite un sistema di crediti acquistabili separatamente.
- **Compatibilità:** Nonostante ArcGIS funzioni sia su Mac che su Windows, richiede comunque un'attrezzatura hardware di fascia alta.
- **Flessibilità:** Nonostante l'interfaccia user-friendly, ArcGIS può risultare rigido in termini di personalizzazione delle funzionalità avanzate senza utilizzare estensioni a pagamento.

- **Metodo di utilizzo:** ArcGIS utilizza principalmente un'interfaccia utente grafica (GUI) che facilita l'interazione con i dati spaziali. Tuttavia, la complessità del software richiede una formazione specifica per sfruttare appieno le sue capacità.

Il programma proposto, al contrario di ArcGIS è gratuito e rispetto ad ArcGIS è molto più semplice da utilizzare. Inoltre, grazie a Python, il software è disponibile su molti sistemi operativi diversi, senza richiedere prestazioni elevate.

3.3 Geosoft Oasis montaj

Geosoft Oasis montaj [10] (figura 3.3) è uno strumento specializzato per l'interpretazione e la visualizzazione dei dati geofisici. La documentazione di Geosoft indica che Oasis montaj è progettato per gestire grandi volumi di dati geofisici, offrendo strumenti avanzati per l'elaborazione e l'analisi dei dati di esplorazione. Inoltre, il software include moduli specializzati per diverse tecniche geofisiche, come la magnetometria, la gravimetria e la sismica.

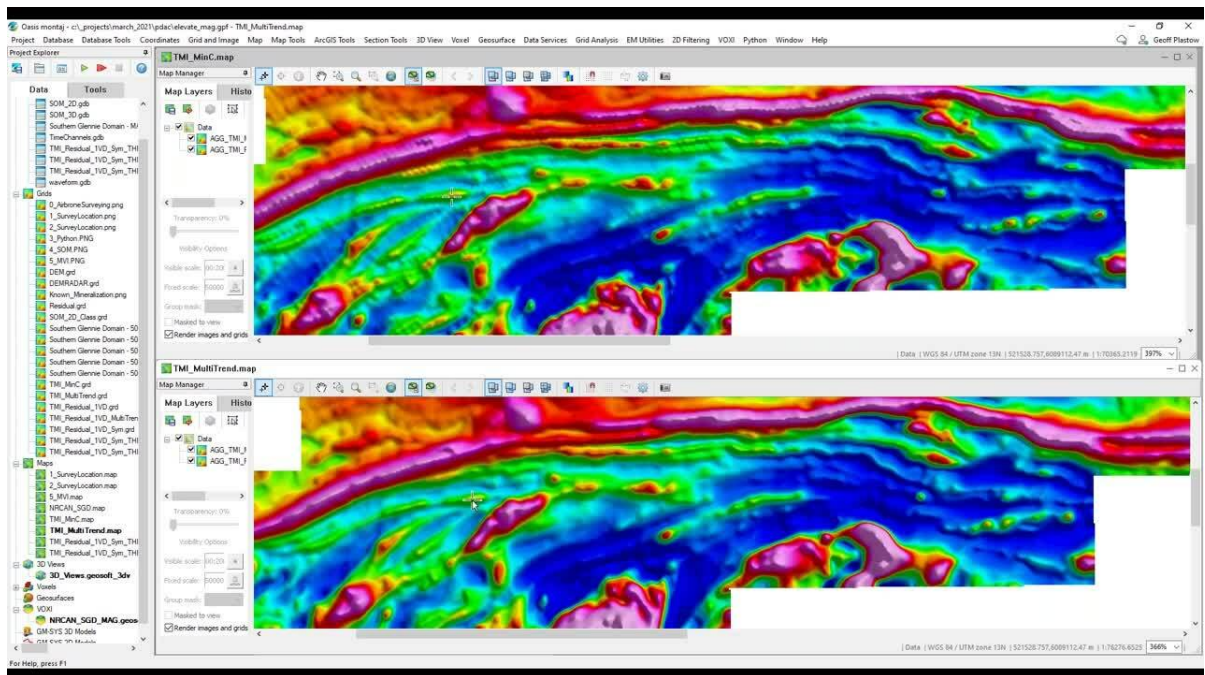


Figura 3.3: Interfaccia Oasis Montaj

Un esempio di applicazione di Geosoft Oasis montaj è nell'esplorazione mineraria, dove il software viene utilizzato per interpretare i dati geofisici e identificare potenziali aree di interesse.

- **Costo:** Come gli altri strumenti proprietari, il costo delle licenze può essere proibitivo per alcune tipologie di utenti.
- **Compatibilità:** Oasis Montaj è compatibile solamente con Windows 11 e pertanto richiede l'utilizzo di macchine virtuali per essere impiegato su macchine con sistemi operativi diversi, riducendo così di molto l'appetibilità a una grande fetta di utenti.
- **Metodo di utilizzo:** Geosoft Oasis montaj combina un'interfaccia utente grafica con strumenti di analisi specializzati. Nonostante la sua potenza, l'uso efficace del software richiede una conoscenza approfondita dei processi geofisici e una formazione specifica.
- **Supporto e comunità:** Ha una solida base di utenti e offre supporto tecnico e risorse di formazione.

Oasis Montaj pecca in compatibilità, infatti il software proposto è disponibile su praticamente tutti i sistemi operativi in circolazione. Inoltre, Oasis Montaj richiede molta più esperienza nel settore per poter ottenere dei risultati. Infine, il programma proposto è gratuito e open-source.

3.4 Alternative open-source

3.4.1 QGIS

QGIS [30] (figura 3.4) è un software mantenuto da una comunità di sviluppatori ed è il sistema di informazioni geografiche (GIS) più diffuso al mondo. È open-source e offre funzionalità avanzate per la gestione e l'analisi dei dati geospaziali. Secondo la documentazione ufficiale, QGIS supporta numerosi formati di dati raster e vettoriali, nonché funzionalità di geoprocessing tramite la libreria integrata di GDAL/OGR.

QGIS è particolarmente utile per la creazione di mappe tematiche e per l'analisi spaziale. Ad esempio, è utilizzato in progetti di monitoraggio ambientale e gestione delle risorse naturali.

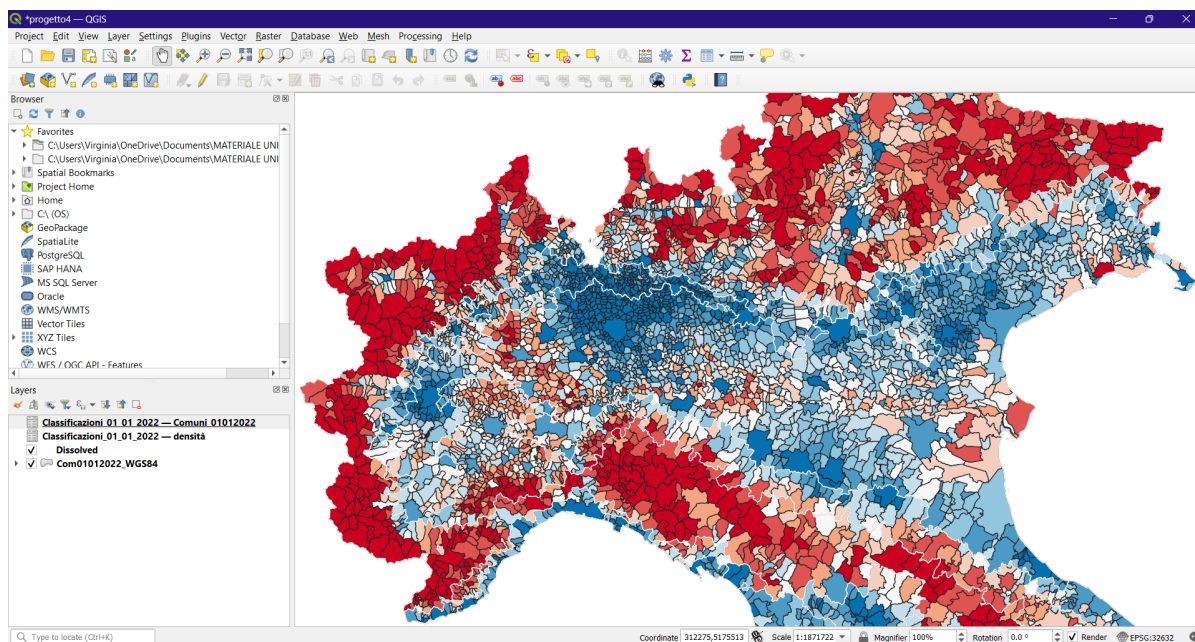


Figura 3.4: Interfaccia QGIS

- **Compatibilità:** QGIS è scritto in C++ e Python, il che lo rende compatibile con tutti i sistemi operativi più diffusi.
- **Estensibilità:** QGIS supporta *plugin* che possono essere sviluppati dagli utenti per estendere le funzionalità del software. *Plugin* di particolare rilevanza o diffusione vengono talvolta inseriti nelle successive versioni del software.
- **Interfaccia utente:** Pur essendo potente, l'interfaccia utente di QGIS può risultare poco intuitiva e più dispersiva rispetto a quella dei software proprietari.
- **Supporto:** Sebbene esista una vasta comunità di utenti, il supporto tecnico professionale potrebbe essere limitato rispetto ai software proprietari.

QGIS ha *feature* molto simili al software proposto, ma la differenza principale si ha nella semplicità di utilizzo, infatti QGIS è molto più difficile da utilizzare se non si ha esperienza nell'utilizzo di software simili.

3.4.2 Paraview

Paraview [26] (figura 3.5) è un software open-source per la visualizzazione di dati scientifici che permette di gestire grandi volumi di dati e creare visualizzazioni avanzate. La documentazione di Paraview afferma che il software è progettato per visualizzare

dati di grandi dimensioni su piattaforme di calcolo ad alte prestazioni. Inoltre, offre un'interfaccia utente intuitiva per creare visualizzazioni complesse e personalizzabili.

Paraview è ampiamente utilizzato nella ricerca scientifica per visualizzare simulazioni complesse e dati volumetrici.

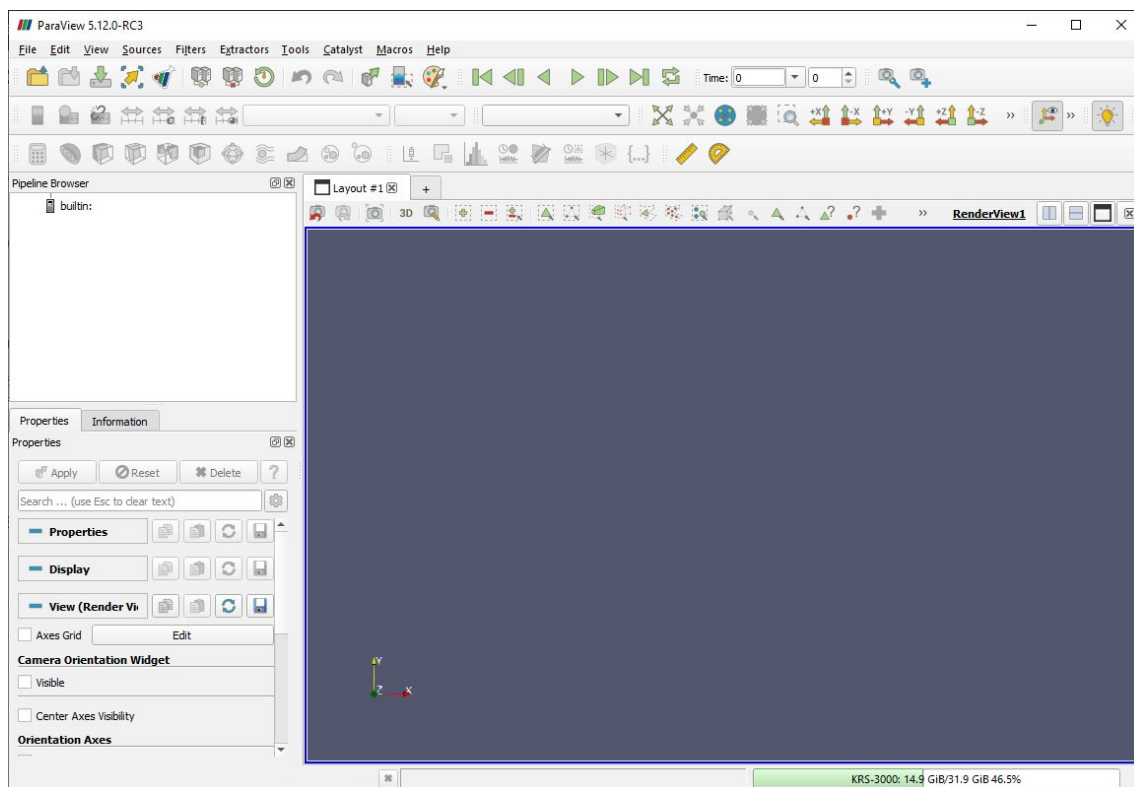


Figura 3.5: Interfaccia Paraview

- **Compatibilità:** Paraview è compatibile con la maggior parte dei sistemi operativi, nonché con *cluster* e computer ad alte prestazioni.
- **Estensibilità:** Gli utenti possono sviluppare *plugin* personalizzati per estendere le funzionalità del software.
- **Interfaccia utente:** La GUI e le funzionalità avanzate possono richiedere un significativo investimento di tempo per essere padroneggiate.
- **Potenza di elaborazione:** Supporta la visualizzazione di dataset di grandi dimensioni attraverso tecniche di rendering distribuito.
- **Interattività:** Offre strumenti interattivi per esplorare e analizzare i dati in 3D.

La soluzione proposta, a differenza di Paraview, è molto più semplice in quanto a GUI e modalità di utilizzo, in modo molto simile a QGIS.

3.4.3 GNU Octave

GNU Octave [12] (figura 3.6) è un software di calcolo numerico che fornisce un ambiente di programmazione simile a quello di Matlab. È ampiamente utilizzato per la risoluzione di problemi numerici, l'elaborazione dei dati e la visualizzazione. Secondo la documentazione di Octave, GNU Octave è progettato per fornire un linguaggio di programmazione ad alto livello principalmente destinato al calcolo numerico. Inoltre, Octave può essere utilizzato sia come linguaggio di programmazione interattivo sia per eseguire script.

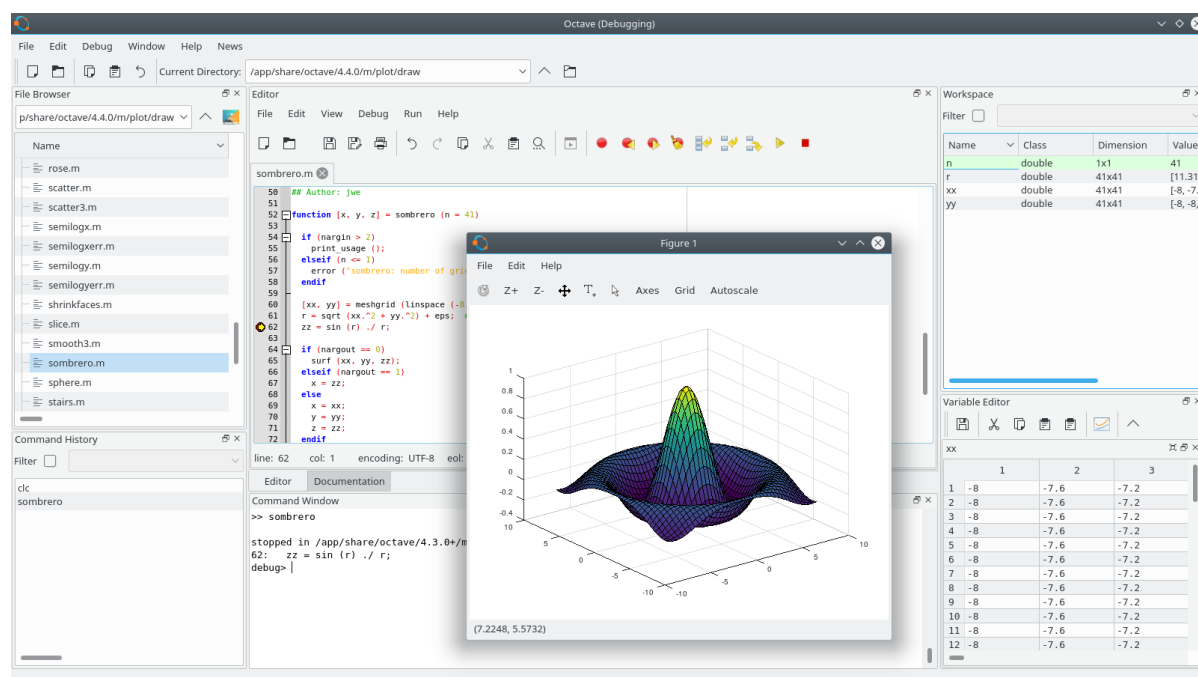


Figura 3.6: Interfaccia GNU Octave

Octave è utilizzato in numerosi campi scientifici e ingegneristici per la sua capacità di eseguire calcoli numerici complessi, elaborare i segnali e creare grafici.

- **Compatibilità:** GNU Octave è compatibile con la maggior parte dei sistemi operativi. Inoltre, è altamente compatibile con Matlab, permettendo l'esecuzione di molti script Matlab senza modifiche significative.

- **Interfaccia utente:** L'interfaccia utente di Octave è minima. Essendo meno sviluppata rispetto a quella di Matlab, le stesse operazioni effettuate su questo software risultano essere più difficili da eseguire.
- **Metodo di utilizzo:** Octave utilizza un linguaggio di programmazione molto simile a Matlab, il che facilita la transizione per gli utenti Matlab.

Rispetto a GNU Octave, la soluzione presentata non richiede conoscenze di programmazione e risulta essere molto più semplice da utilizzare.

3.4.4 Python con librerie di visualizzazione

Python, insieme a librerie come `Matplotlib` [22], `Plotly` [28] e `Mayavi` [23], offre un potente ambiente open-source per la visualizzazione dei dati. La sua documentazione indica che `Matplotlib` è una libreria completa per creare visualizzazioni statiche, animate e interattive in Python. `Plotly`, invece, permette di creare grafici interattivi e dashboard che possono essere visualizzati in un browser web. `Mayavi` è una libreria per la visualizzazione scientifica in 3D che integra la potenza di visualizzazione di VTK (*The Visualization Toolkit*) con la semplicità di utilizzo di Python.

Python con le sue librerie è ampiamente utilizzato per la creazione di visualizzazioni personalizzate e interattive in vari campi scientifici e industriali.

- **Compatibilità:** Le librerie Python sono compatibili con tutti i principali sistemi operativi, rendendo le soluzioni sviluppate accessibili su diverse piattaforme.
- **Flessibilità:** Python è altamente flessibile e permette di personalizzare le visualizzazioni secondo le esigenze specifiche dell'utente.
- **Supporto:** Python e le sue librerie sono ampiamente supportati dalla comunità open-source.
- **Metodo di utilizzo:** Nonostante la flessibilità, l'apprendimento di Python e delle sue librerie può richiedere tempo, soprattutto per utenti nuovi alla programmazione.
- **Integrazione:** Facilita l'integrazione con altri strumenti e flussi di lavoro di analisi dei dati.

Il software proposto non richiede conoscenze di programmazione, in quanto è scritto in Python ma non richiede che l'utente scriva del codice da sé.

Capitolo 4

Implementazione della visualizzazione

4.1 Introduzione

In questo capitolo viene illustrata l'implementazione del programma di visualizzazione dei dati geologici sviluppato nell'ambito di questo progetto. Il programma utilizza diverse librerie Python avanzate per il caricamento, la pre-elaborazione, la modellazione e la visualizzazione interattiva dei dati. Le sezioni successive descrivono il flusso di lavoro del programma, le tecniche utilizzate per creare modelli bidimensionali e tridimensionali dei dati, e le soluzioni adottate per garantire l'accessibilità del progetto, in particolare per le persone affette da daltonismo. Infine, viene fatta una comparazione con le alternative esistenti, mettendo in luce i vantaggi offerti dal nostro programma, tra cui la gratuità, la compatibilità, la flessibilità, la visualizzazione avanzata e l'inclusività.

I dati utilizzati sono frutto di misurazioni ed elaborazioni riguardanti l'area degli Appennini Meridionali [39.8294 42.1706],[13.9549 17.0451]. I dati misurati vengono elaborati attraverso un processo di inversione sequenziale integrata che combina informazioni sismologiche e gravimetriche [39]; il software implementato utilizza i dati derivanti da questo processo di elaborazione.

4.2 Funzionamento del programma

Il programma è stato sviluppato utilizzando diverse librerie Python, tra cui NumPy, Pandas, Matplotlib e Plotly.

La costruzione del *software* è avvenuta in un'ottica *user-oriented*, concentrandosi sulla possibilità di fornire all'utente un'interfaccia interattiva di facile utilizzo.

Il flusso di lavoro principale del programma può essere descritto attraverso le fasi elencate in seguito.

4.2.1 Caricamento e pre-elaborazione dei dati

Per prima cosa vengono importati i dati in formato CSV tramite la libreria `Pandas`. Il dataset deve contenere le coordinate geografiche o cartesiane dei punti (X, Y e Z) e un valore calcolato (densità o velocità sismica). La pulizia dei dati consiste nell'esclusione dei valori di bordo, da non considerare poiché conseguenza dell'elaborazione dei dati gravimetrici.

4.2.2 Creazione del modello

I modelli vengono creati utilizzando la libreria `Plotly` e in particolare il pacchetto `graph_objects (go)`. L'utente ha la possibilità di personalizzare i valori della legenda e di optare per una visualizzazione accessibile.

Le visualizzazioni realizzate prevedono l'utilizzo dell'intensità cromatica per rappresentare i valori delle aree geologiche. Questo tipo di rappresentazione, pur trovandosi al livello più basso nella classifica di precisione della percezione visiva, è particolarmente utile per fornire una visione generale dell'area in analisi. In particolare, l'applicazione a dati geologici assicura la presenza di trend spaziali, per i quali l'utilizzo delle gradazioni di colore risulta particolarmente efficace.

Modello tridimensionale

I dati vengono visualizzati in formato quadridimensionale (figura 4.1), con l'intensità del valore misurato espressa tramite una scala di colori continua. L'intervallo di valori fornito in input agisce da filtro sui dati da visualizzare, permettendo di visualizzare, se si desidera, alcune formazioni geologiche specifiche. È possibile interagire con il modello creato tramite rotazione, zoom e spostamento; inoltre, posizionando il cursore su un punto specifico ne vengono visualizzate le coordinate geografiche e il valore ad esse associato.

In questo modello può essere inserita la topografia, se fornita in input tramite un sistema di coordinate in un file CSV addizionale. Per rendere i valori della topografia più distinguibili, questi vengono scalati in modo da rendere la scala sopra la superficie terrestre confrontabile con quella sotterranea, che arriva fino a 50 km di profondità: in particolare, i valori di superficie vengono rappresentati in scala 4:1.

Un'altra opzione di visualizzazione del modello tridimensionale è a sezioni (figura 4.2), che si ottiene iterando sui valori unici delle coordinate. Per questo tipo di grafico si è scelto di adottare una scala di colori discreta, che rende la visualizzazione più facilmente interpretabile. L'interazione consiste nello scorrimento di una barra orizzontale che mostra le sezioni corrispondenti a diverse coordinate.

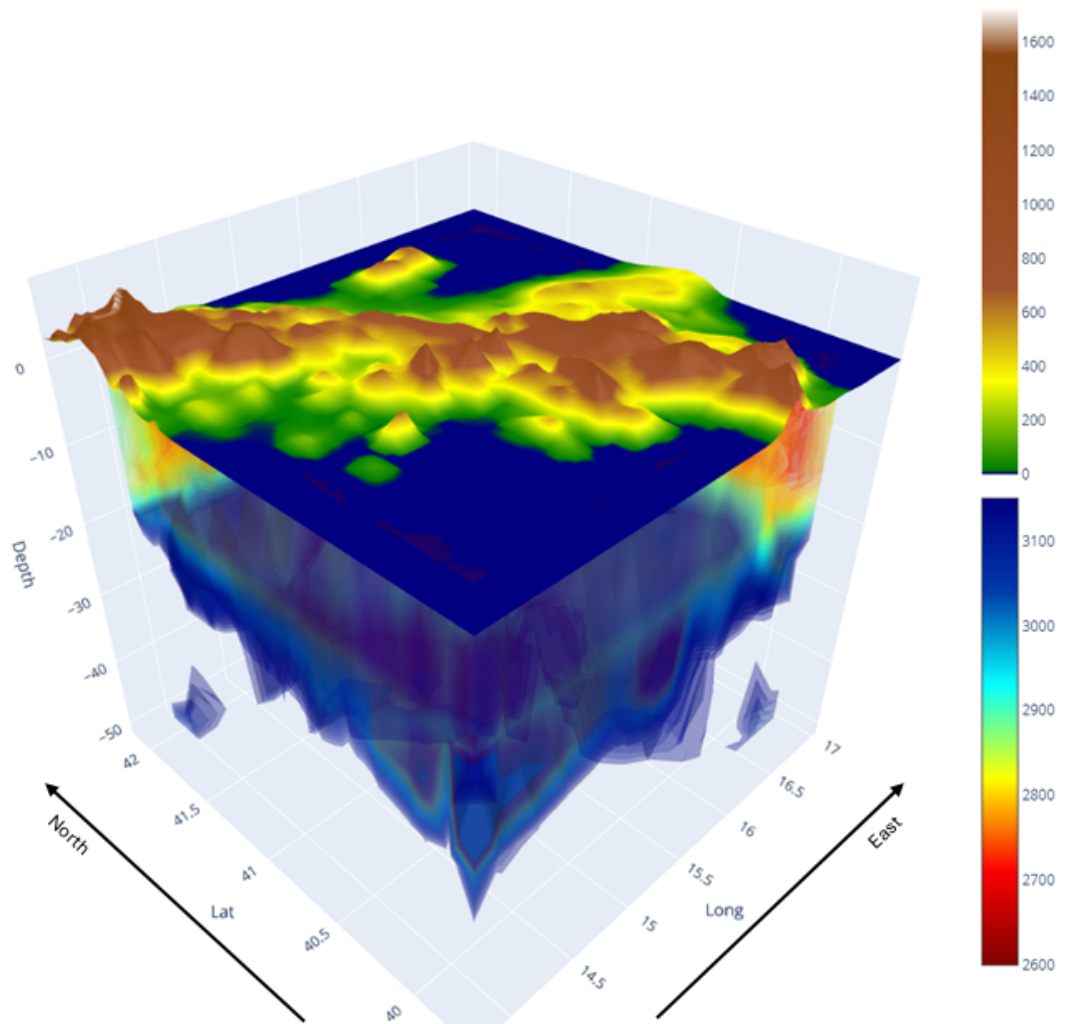


Figura 4.1: Modello tridimensionale con topografia

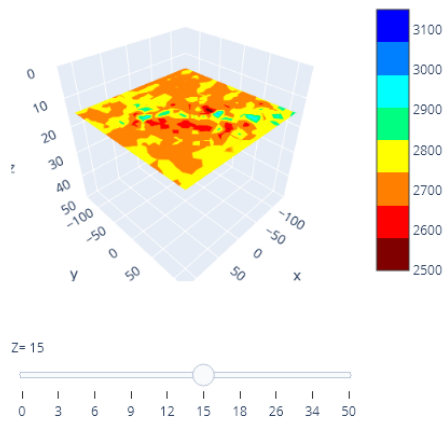
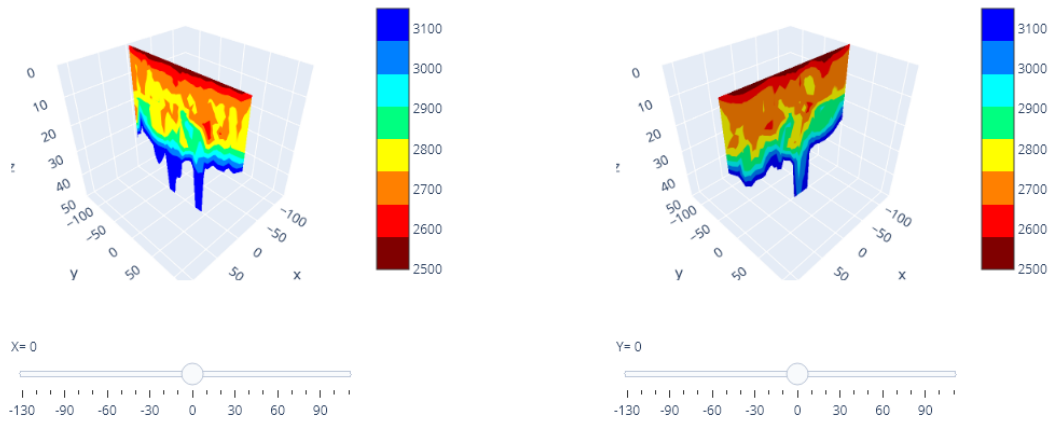
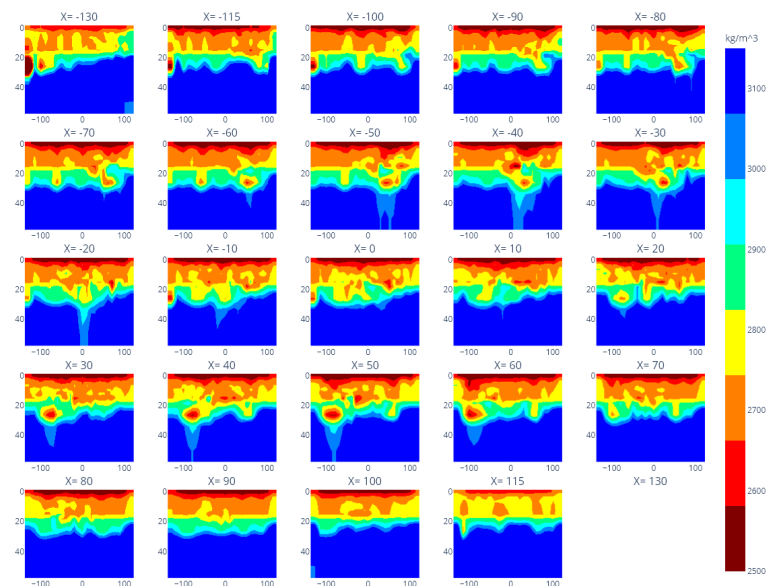
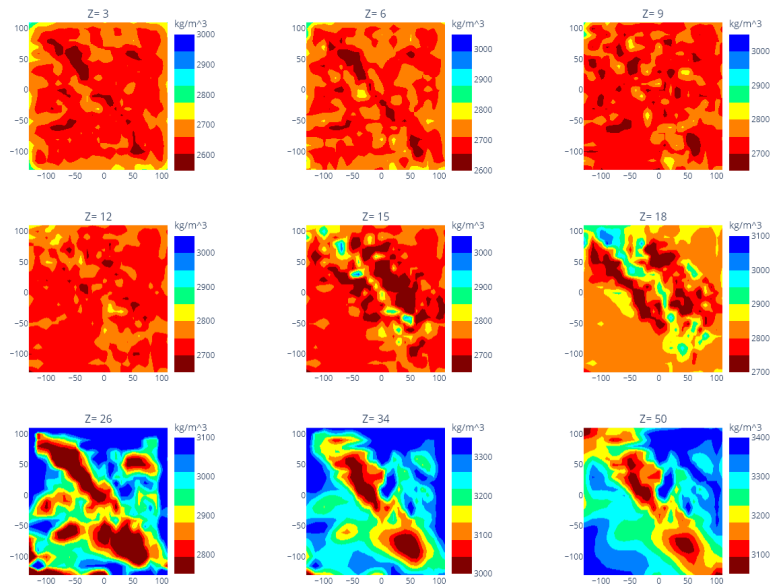


Figura 4.2: Sezioni

Modello bidimensionale



(a) Sezioni lungo l'asse x



(b) Sezioni lungo l'asse z

Figura 4.3: Griglie di sezioni

Le sezioni di terreno possono essere visualizzate anche in una griglia di *heatmap* (figura 4.3). I grafici che la compongono possono avere una legenda comune o ad hoc per ciascuno di essi; la *colorscale* è discreta.

Un'altra opzione di visualizzazione si può ottenere grazie all'utilizzo di uno *slider* (figura 4.4), che permette di visualizzare una *heatmap* bidimensionale in corrispondenza delle diverse coordinate. La scala di colori è di tipo discreto e ogni *heatmap* ha la sua scala personalizzata.

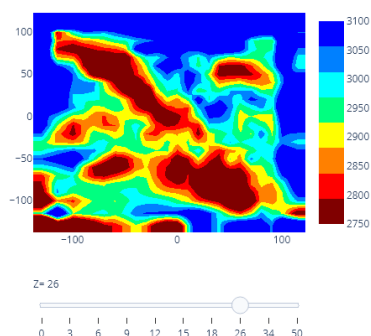


Figura 4.4: Modello bidimensionale con barra di scorrimento

Per finire, è anche possibile definire un profilo lungo il quale effettuare il taglio del modello tridimensionale: deve essere fornito in input tramite un file CSV. Il programma restituisce una vista bidimensionale del profilo (figura 4.5), che, come nei casi precedenti, è interattiva e personalizzabile; la *colorscale* è discreta.

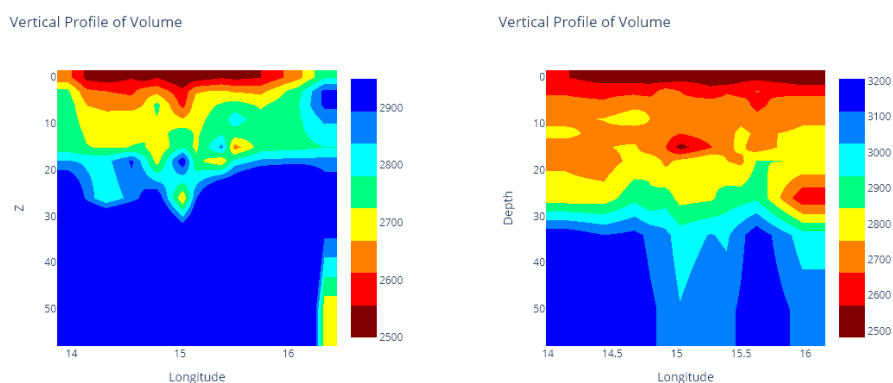


Figura 4.5: Profili di taglio del modello tridimensionale

4.3 Accessibilità del progetto sviluppato

4.3.1 Problemi delle scale dei colori per il daltonismo

Un aspetto importante nella visualizzazione dei dati è l'accessibilità per le persone affette da daltonismo. Il daltonismo è una condizione che altera la percezione dei colori e può essere di diversi tipi:

- **Deuteranopia:** Mancanza di percezione del verde.
- **Protanopia:** Mancanza di percezione del rosso.
- **Tritanopia:** Mancanza di percezione del blu.

Le scale di colori utilizzate nelle visualizzazioni possono risultare indistinguibili per le persone affette da questi tipi di daltonismo. Per questo motivo, è importante scegliere scale di colori che siano facilmente distinguibili da tutti gli utenti.

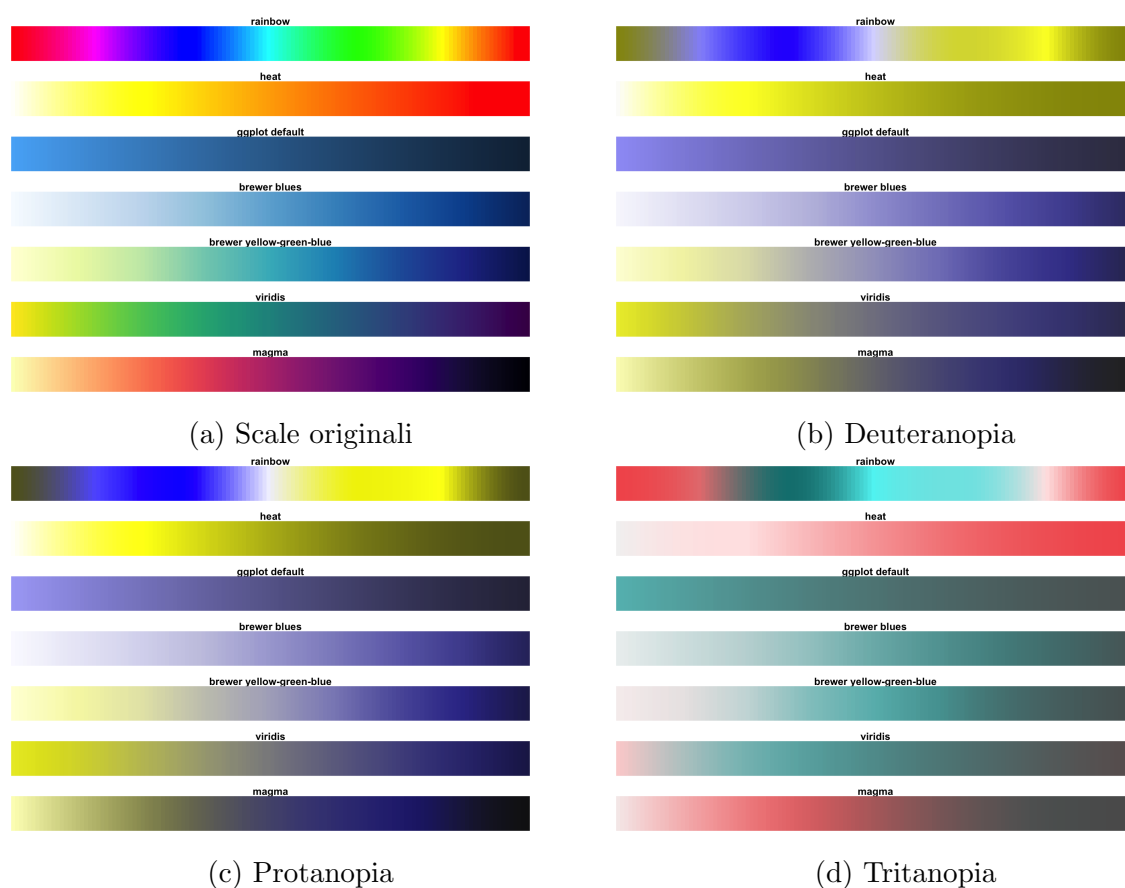


Figura 4.6: Paragone scale più diffuse per ogni tipo di daltonismo [33]

4.3.2 Utilizzo della scala di colori *Viridis*

Per garantire l'accessibilità del programma sviluppato, è stata scelta la scala di colori *Viridis*. *Viridis* è una scala di colori progettata per essere facilmente distinguibile anche per le persone affette da daltonismo. Infatti, dall'immagine (figura 4.6) si può notare come le scale *Viridis* e *Magma* (appartenenti al pacchetto di scale di *Viridis*) vengano percepite comunque con contrasti chiari e luminosità graduale, in modo molto simile per quanto riguarda deuteranopia e protanopia, mentre risulta essere meno efficace per la tritanopia. Trattandosi quest'ultima di una condizione estremamente rara la scala *Viridis* rimane la migliore scelta di scala nell'ambito dell'accessibilità [33]. Questa scala è continua e percepita uniformemente, il che la rende ideale per rappresentare variazioni nei dati geologici (figura 4.7).

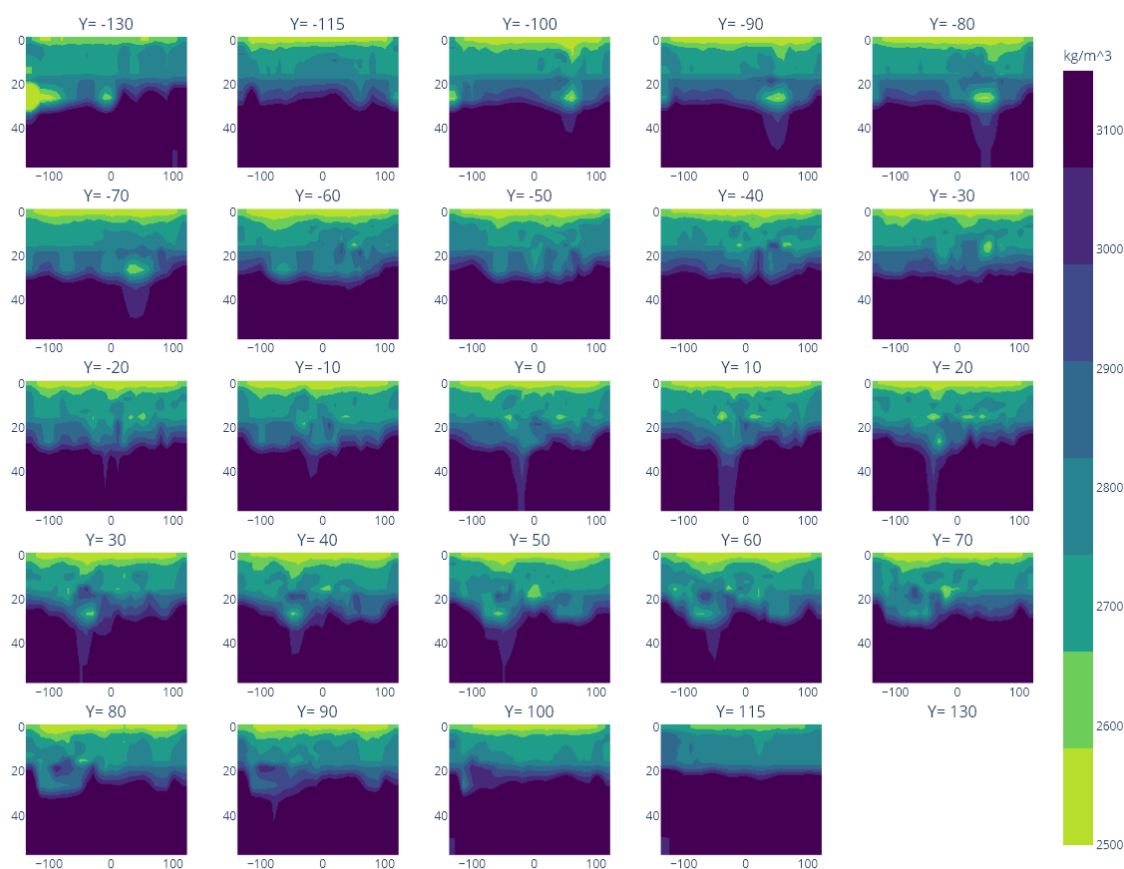


Figura 4.7: Sezioni lungo l'asse y con scala di colori *Viridis*

4.4 Vantaggi rispetto alle alternative esistenti

Il programma sviluppato presenta diversi vantaggi rispetto alle soluzioni descritte nel capitolo precedente:

- **Gratuità e accessibilità:** Essendo *open-source*, il programma è completamente gratuito e può essere utilizzato, modificato e distribuito liberamente. Questo lo rende particolarmente attraente per studenti, ricercatori e professionisti con budget limitati.
- **Compatibilità:** Il programma è compatibile con tutti i principali sistemi operativi (Windows, macOS, Linux) e può essere eseguito su macchine ad alte prestazioni (HPC), rendendolo versatile e adatto a diversi ambienti di lavoro.
- **Flessibilità e personalizzazione:** Grazie all'utilizzo di Python, il software può essere facilmente personalizzato e ampliato con ulteriori funzionalità secondo le esigenze specifiche dell'utente. La comunità Python offre una vasta gamma di librerie che possono essere integrate per migliorare ulteriormente le capacità del programma.
- **Visualizzazione avanzata:** Utilizzando la libreria `Plotly`, il programma permette la creazione di visualizzazioni interattive e dettagliate, migliorando l'analisi e l'interpretazione dei dati. Le visualizzazioni possono essere esplorate dinamicamente dall'utente, offrendo un'esperienza *user-friendly*.
- **Accessibilità del colore:** L'uso della scala di colori *Viridis* garantisce che le visualizzazioni siano accessibili anche alle persone affette da daltonismo, migliorando l'inclusività del programma.
- **Semplicità d'uso:** Il programma non richiede nessuna conoscenza particolare all'utente in quanto utilizza solo file di dati o input testuali semplici, il che lo rende adatto all'utilizzo da parte di utenti poco esperti nella programmazione e nell'utilizzo dei computer in generale.

Capitolo 5

Conclusioni con possibili sviluppi futuri

Il progetto presentato in questa tesi rappresenta una potenziale alternativa potente, flessibile ed accessibile nel campo della visualizzazione dei dati geologici.

L'obiettivo del progetto era quello di sviluppare un software che fosse in grado di usare efficacemente dati di natura geologica per rappresentare accuratamente dei modelli; questi ultimi dovevano essere di vario tipo e altamente personalizzabili dall'utente. Uno degli aspetti chiave era quello di rendere il software *user-friendly* e disponibile gratuitamente, in modo che potesse essere utilizzato da molti ricercatori in ambito geologico.

L'obiettivo è stato in gran parte raggiunto, con anche alcune *feature* aggiuntive che puntano a rendere il progetto espandibile nel futuro prossimo. Il programma sviluppato è infatti in grado di fornire molti tipi di modelli utilizzando file di dati in input in formato CSV e di esportare tali modelli sotto forma di immagini e file HTML. I modelli forniti dal programma sono di molteplici tipologie:

- Modelli tridimensionali che permettono di rappresentare i dati fornendo una visione totale del sottosuolo e della topografia della superficie.
- Modelli bidimensionali dinamici, che grazie alle animazioni permettono di vedere il modello tridimensionale una sezione alla volta, fornendo così una vista più dettagliata e isolata delle zone d'interesse.
- Modelli bidimensionali statici, che in modo simile ai precedenti permettono di vedere tutte le sezioni contemporaneamente sotto forma di *heatmap* in un unico grafico.
- Sezioni di taglio discrete, che grazie a un ulteriore set di dati contenente le coordinate dei punti da seguire per tagliare il volume, presenta una visione bidimensionale della sezione del volume lungo il taglio desiderato dall'utente.

Il progetto ha rispettato tutti i principali obiettivi prefissati: infatti, grazie all'utilizzo esclusivo di Python il programma è eseguibile sulla maggior parte dei sistemi operativi esistenti, comprendendo anche le macchine ad alte prestazioni (HPC). L'utilizzo della libreria Plotly rende i modelli prodotti totalmente interattivi, rendendo la visualizzazione chiara e accattivante. Un ulteriore obiettivo raggiunto è la facilità di utilizzo, infatti il software prodotto non richiede nessuna conoscenza pregressa nella programmazione per poter ottenere i risultati desiderati: ciò lo rende accessibile a chiunque. Infine, un'ulteriore caratteristica implementata è quella della completa disponibilità del codice, il che permette a un utente più esperto di personalizzare al massimo le visualizzazioni e i metodi di calcolo: consente quindi agli sviluppatori di aggiungere con facilità nuove funzioni per eventuali sviluppi futuri necessari.

Riassumendo, il programma sviluppato si impone come *competitor open-source* e *user-friendly* a software specifici per la visualizzazione di dati geologici, andando così ad offrire una valida opzione per il pubblico.

5.1 Possibili sviluppi futuri

Nonostante i risultati ottenuti, ci sono molte aree in cui il progetto può essere ulteriormente sviluppato e migliorato. Di seguito sono elencate alcune possibili direzioni future:

5.1.1 Supporto per ulteriori tipi di dati

La possibilità di fornire al software altri tipi di dati per ottenere dei modelli di diversa natura potrebbe essere una caratteristica che renderebbe questo programma ancora più versatile e completo, accogliendo una più vasta gamma di utenti.

5.1.2 Implementazione di ulteriori modelli

L'aggiunta di altri tipi di modelli di visualizzazione potrebbe permettere di aumentare le opzioni di utilizzo e di espandere il target di utenti anche ad altri ambiti, oltre alla geologia.

5.1.3 Ottimizzazione delle prestazioni

Ottimizzare gli algoritmi di elaborazione dei dati potrebbe ridurre i tempi di calcolo e permetterebbe la visualizzazione di una quantità maggiore di dati in un minor lasso di tempo. L'utilizzo di librerie specializzate potrebbe permettere di visualizzare dati con una densità molto più alta senza complicazioni.

5.1.4 Interfaccia utente migliorata

Al momento il programma fornisce un'interfaccia utente minima, che si riduce a un'interazione testuale. Un possibile progresso potrebbe essere quello di implementare una vera e propria GUI per rendere il programma più facile e veloce da utilizzare e per presentare al meglio tutte le funzionalità che propone all'utente

5.1.5 Visualizzazione di modelli su webserver

Uno degli sviluppi più interessanti che possono riguardare questo progetto è quello di sviluppare una pagina web in cui pubblicare giornalmente modelli elaborati dal software e permettere di utilizzare il programma tramite una semplice pagina web. Questo progresso è stato già in parte attuato, infatti è stata creata una pagina HTML dove possono essere visualizzati tutti i modelli ottenuti dal software in un unico posto, la pagina creata è ancora statica e viene visualizzata solamente in locale, ma in futuro potrà essere usata per permettere agli utenti di accedere alle funzioni del software da pagina web. Ciò è stato reso possibile grazie alla flessibilità di Plotly, che permette di esportare ogni modello ottenuto come HTML, rendendolo esplorabile nel browser.

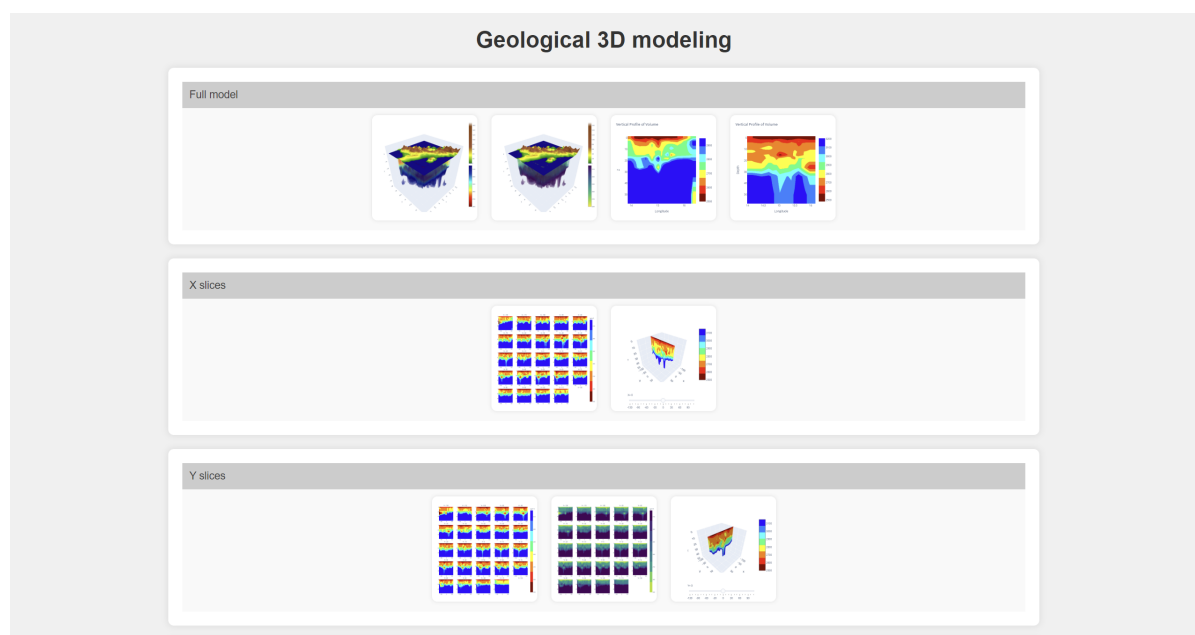


Figura 5.1: Pagina HTML

Gli sviluppi futuri proposti possono ulteriormente migliorare il programma, rendendolo uno strumento indispensabile per ricercatori, studenti e professionisti nel campo della geofisica.

Bibliografia

- [1] BigVis 2020. *Big Data Visualization and Analytics: Future Research Challenges and Emerging Applications*. URL: <https://bigvis.imsi.athenarc.gr/bigvis2020/program.html>.
- [2] G. Andrienko, N. Andrienko e S. Wrobel. “Visual analytics tools for analysis of movement data”. In: 9.2 (2007). ISSN: 1931-0145. DOI: 10.1145/1345448.1345455.
- [3] P.A. Burrough, R.A. McDonnell e C.D. Lloyd. *Principles of Geographical Information Systems*. OUP Oxford, 2015. ISBN: 9780198742845. URL: <https://books.google.it/books?id=kvoJCAAAQBAJ>.
- [4] *Data visualization with ggplot2: cheatsheet*. URL: <https://github.com/rstudio/cheatsheets/blob/main/data-visualization.pdf>.
- [5] *Data Wrapper*. URL: <https://www.datawrapper.de/>.
- [6] S. Dattani et al. “Child and Infant Mortality”. In: *Our World in Data* (2023). URL: <https://ourworldindata.org/child-mortality>.
- [7] Esri. *ArcGIS Documentation*. 2023. URL: <https://www.esri.com/en-us/arcgis/products/arcgis-pro/resources>.
- [8] I. S. Evans. “Geomorphometry and landform mapping: What is a landform?” In: *Geomorphology* 137.1 (2012), pp. 94–106. ISSN: 0169-555X. DOI: <https://doi.org/10.1016/j.geomorph.2010.09.029>.
- [9] S.L. Franconeri et al. “The Science of Visual Data Communication: What Works”. In: *Psychological Science in the Public Interest* 22.3 (2021), pp. 110–161. DOI: <https://doi.org/10.1177/152910062111051956>.
- [10] Geosoft Inc. *Geosoft Oasis montaj Documentation*. 2023. URL: <https://www.geosoft.com/products/oasis-montaj>.
- [11] *ggplot2*. URL: <https://ggplot2.tidyverse.org/>.
- [12] GNU Octave Development Team. *GNU Octave Documentation*. 2023. URL: <https://www.gnu.org/software/octave/doc/interpreter/>.

- [13] Pew Research Center; Sara Kehulani Goo. *The art and science of the scatterplot*. URL: <https://www.pewresearch.org/short-reads/2015/09/16/the-art-and-science-of-the-scatterplot/>.
- [14] M. F Goodchild. “Citizens as sensors: the world of volunteered geography”. In: *GeoJournal* 69.4 (2007), pp. 211–221. DOI: <https://doi.org/10.1007/s10708-007-9111-y>.
- [15] *Google Charts*. URL: <https://developers.google.com/chart>.
- [16] Philippe Cerfontaine Hans-Balder Havenith e Anne-Sophie Mreyen. “How virtual reality can help visualise and assess geohazards”. In: *International Journal of Digital Earth* 12.2 (2019), pp. 173–189. DOI: [10.1080/17538947.2017.1365960](https://doi.org/10.1080/17538947.2017.1365960).
- [17] T. Höllerer e S. Feiner. *Mobile augmented reality*. Vol. 21. CRC Press Florida, 2004, pp. 221–260. URL: https://web.cs.wpi.edu/~gogo/courses/imgd5100_2012f/papers/Hollerer_AR_2004.pdf.
- [18] Diane J. Janvrin, Robyn L. Raschke e William N. Dilla. “Making sense of complex data using interactive data visualization”. In: *Journal of Accounting Education* 32.4 (2014), pp. 31–48. ISSN: 0748-5751. DOI: <https://doi.org/10.1016/j.jaccedu.2014.09.003>.
- [19] Z. Li, C. Zhu e C. Gold. *Digital Terrain Modeling: Principles and Methodology (1st ed.)*. CRC Press, 2005. DOI: [10.1201/9780203357132](https://doi.org/10.1201/9780203357132).
- [20] C. Lin e R. B. Loftin. “Application of virtual reality in the interpretation of geoscience data”. In: (1998), pp. 187–194. DOI: [10.1145/293701.293736](https://doi.org/10.1145/293701.293736).
- [21] P. Longley. *Geographical Information Systems and Science*. Wiley, 2005. ISBN: 9780470870006. URL: <https://www.wiley.com/en-us/Geographic+Information+Science+and+Systems%2C+4th+Edition-p-9781119128458>.
- [22] Matplotlib Development Team. *Matplotlib Documentation*. 2023. URL: <https://matplotlib.org/stable/contents.html>.
- [23] Mayavi Development Team. *Mayavi Documentation*. 2023. URL: <https://docs.enthought.com/mayavi/mayavi/>.
- [24] NASA. *Video: Climate Spiral (1880-2022)*. URL: <https://science.nasa.gov/resource/video-climate-spiral-1880-2022/>.
- [25] Observable. *d3.js*. URL: <https://d3js.org/>.
- [26] ParaView Development Team. *ParaView Documentation*. 2023. URL: <https://www.paraview.org/documentation/>.
- [27] D. J. Peuquet. *Representations of space and time*. Guilford Press, 2002. ISBN: 9781138871502.

- [28] Plotly Technologies Inc. *Plotly Documentation*. 2023. URL: <https://plotly.com/python/>.
- [29] *Power BI*. URL: <https://app.powerbi.com/>.
- [30] QGIS Development Team. *QGIS Documentation*. 2023. URL: <https://docs.qgis.org/>.
- [31] H. Ritchie, E. Mathieu e M. Roser. “Research and Development”. In: *Our World in Data* (2023). URL: <https://ourworldindata.org/research-and-development>.
- [32] H. Ritchie, P. Rosado e M. Roser. “Energy”. In: *Our World in Data* (2023). URL: <https://ourworldindata.org/energy>.
- [33] B. Rudis, N. Ross e S. Garnier. *Introduction to the viridis color maps*. URL: <https://cran.r-project.org/web/packages/viridis/vignettes/intro-to-viridis.html>.
- [34] J. Schwabish. *Better Data Visualizations : A Guide for Scholars, Researchers, and Wonks*. Columbia University Press, 2021. ISBN: 9780231550154. URL: <https://search.ebscohost.com/login.aspx?direct=true&db=e000xww&AN=2453484&site=ehost-live&scope=site>.
- [35] *seaborn*. URL: <https://seaborn.pydata.org/>.
- [36] J. F Shroder. *Treatise on geomorphology*. Academic Press, 2013. ISBN: 9780080885223.
- [37] *Tableau*. URL: <https://www.tableau.com/>.
- [38] The MathWorks, Inc. *MATLAB Documentation*. 2023. URL: <https://www.mathworks.com/help/matlab/>.
- [39] R. Tondi et al. “Parallel, ‘large’, dense matrix problems: Application to 3D sequential integrated inversion of seismological and gravity data”. In: *Computers & Geosciences* 48 (2012), pp. 143–156. ISSN: 0098-3004. DOI: <https://doi.org/10.1016/j.cageo.2012.05.026>.
- [40] E. R. Tufte. *The visual display of quantitative information*. Graphics Press, 2001. ISBN: 9781930824133. URL: <https://faculty.salisbury.edu/~jtanderson/teaching/cosc311/fa21/files/tufte.pdf>.
- [41] L. Wilkinson. *The grammar of graphics*. Springer New York, NY, 2005. ISBN: 9780387245447. DOI: 10.1007/0-387-28695-0.

Ringraziamenti

Per prima cosa vorrei ringraziare il mio relatore Prof. Federico Montori per avermi aiutato nella stesura della tesi e avermi seguito in questo percorso.

Ringrazio anche i miei due correlatori Dott. Maria Rosaria Tondi e Ing. Giampaolo Zerbinato per avermi guidato nello sviluppo di questo progetto e per avermi dedicato il loro tempo anche a lavoro ultimato.

Ringrazio anche tutta la mia famiglia per avermi sempre supportato e spronato durante questo percorso, nonostante gli inciampi.

Un ringraziamento particolare va al mio amico di sempre Alessio, grazie per tutti questi anni passati a divertirci e a incitarci l'un l'altro e al mio amico Niccolò, grazie per avermi aiutato a staccare la spina in momenti molto pesanti e per avermi aiutato durante tutti questi anni.

Infine, il ringraziamento più importante va alla mia fidanzata Virginia, senza di te non sarei mai riuscito a finire questa avventura universitaria. Grazie di cuore per esserci sempre stata da 5.300588 anni a questa parte e per avermi sempre dato fiducia.