Università degli Studi di Milano-Bicocca

Corso di Laurea in Data Science



La pericolosità nei paesi globali

Rogerio De Oliveira Novara 814582

Michele Bonera 890256

Alex Marino 817440



Indice

Introduzione

- 1. Progettazione dell'infografica
 - 1.1. Gli obiettivi e le domande di ricerca
 - 1.2. La raccolta dei dati
 - 1.3. Target di riferimento
- 2. Realizzazione della data visualization
 - 2.1. Accessibilità
 - 2.2 Modifiche alle infografiche
- 3. Valutazione della qualità della data visualization
 - 3.1. Valutazione euristica
 - 3.2. Test utente
 - 3.3. Questionario psicometrico
- 4. Conclusioni e implementazioni future
- 5. Fonti

Introduzione

L'analisi svolta in questo report studia la pericolosità nei paesi del mondo rispetto a due punti di vista: geopolitico e interno. Nel primo caso abbiamo utilizzato indicatori aggregati a livello nazionale. In particolare, ci siamo basati sul Global Peace Index (GPI). Il GPI è prodotto dall'Institute for Economics and Peace ed è la misura di riferimento mondiale per determinare l'attitudine di un determinato paese a essere considerato pacifico. Il GPI racchiude al suo interno il 99,7% della popolazione mondiale, utilizzando 23 indicatori qualitativi e quantitativi provenienti da fonti altamente rispettate, e misura lo stato di pace in tre domini: il livello di sicurezza e protezione della società, l'entità del conflitto interno e internazionale in corso e il grado di militarizzazione. Tale indice assume valori maggiori in paesi meno pacifici. [1]

Successivamente, abbiamo analizzato la pericolosità interna delle nazioni considerando le morti verificatesi in modo violento. Nello specifico, per questo scopo vengono utilizzati indicatori quali il numero di morti violente per 100.000 abitanti e il numero di omicidi volontari per 100.000 abitanti.

Per il presente report sono state realizzate due infografiche con lo scopo di raccontare graficamente la pericolosità nei paesi del mondo. La costruzione delle due visualizzazioni è stata svolta seguendo principalmente due passaggi: sono stati raccolti dati e informazioni utili sul tema e in seguito sono state progettate visualizzazioni che riuscissero a comunicare il messaggio prestabilito in modo chiaro.

Le visualizzazioni realizzate sono state valutate attraverso la valutazione euristica, con la finalità di individuare le problematiche presenti nelle infografiche. È stato inoltre realizzato un test utente, con lo scopo di valutare l'efficacia e l'efficienza delle data visualization. Infine, per ottenere un quadro generale di valutazione delle infografiche è stato realizzato un questionario psicometrico.

1. Progettazione dell'infografica

1.1 Gli obiettivi e le domande di ricerca

Partendo dai dati a nostra disposizione l'obiettivo della nostra data visualization è quello di indagare e fornire agli utenti un quadro della condizione sociale, economica e politica sussistente nelle diverse nazioni Per lo svolgimento di tale studio sono stati selezionati alcuni indicatori del Worldwide Governance Indicators (WGI) che permettono di cogliere i tratti generali della sicurezza, tra i quali:

- Political Stability and Absence of Violence/Terrorism: riflette la stabilità politica e l'assenza di violenza/terrorismo misurando la percezione della probabilità di instabilità politica e/o violenza a sfondo politico, compreso il terrorismo.
- Voice and Accountability: riflette la percezione della misura in cui i cittadini di un Paese hanno libertà di espressione, libertà di associazione e vi sono mezzi di comunicazione liberi.
- Regulatory Quality: Riflette le percezioni sulla capacità del governo di formulare e attuare politiche e regolamenti solidi che consentono e promuovono lo sviluppo del settore privato.
- Rule of Law: riflette la percezione della misura in cui gli individui hanno fiducia e rispettano le regole della società. In particolare: la qualità dell'esecuzione dei contratti, i diritti di proprietà, la polizia e i tribunali, nonché la probabilità di crimini e violenze.
- Control of Corruption: riflette le percezioni sulla misura in cui il potere pubblico viene esercitato per il guadagno privato, comprese le forme di corruzione sia piccole che grandi.

 Così come anche la "cattura" dello stato da parte delle élite e degli interessi privati.
- Tasso delle morti violente: indice relativo alle morti avvenute in circostanze violente per 100'000 abitanti.

Tali indici hanno un range che va da -2.5 a +2.5, dove a valori al di sotto dello zero corrisponde una situazione negativa dei parametri stimati.

Le **domande di ricerca** sono le seguenti:

- 1. Quali sono i Paesi con un elevato tasso di morti violente e qual è la sua evoluzione nel tempo?
- 2. Il livello di pace di uno Stato può essere correlato al grado stimato di libertà di stampa e di espressione all'interno dei suoi confini? Tale grado influenza le altre variabili?

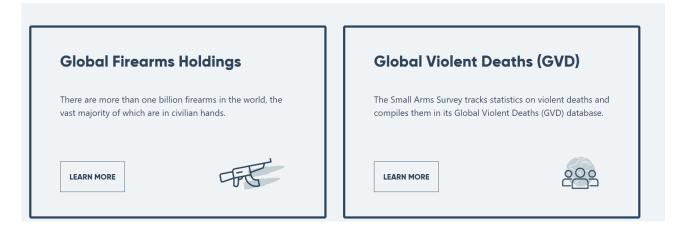
Mentre le **ipotesi** da verificare risultano:

- All'aumentare della libertà di espressione aumenta anche il livello di pace del Paese. Questa ipotesi si basa sul fatto che ci si aspetta che dove si hanno politiche meno repressive, maggiore rispetto per l'individuo e stabilità politica, si sviluppi anche una cultura per la pace più solida
- 2. Esistenza di una correlazione fra il livello di pace di una Nazione e il tasso di morti violente al suo interno. Ci si aspetta infatti che l'aumento delle morti violente sia direttamente proporzionale al peggioramento del Global Pace Index.

1.2 La raccolta dei dati

La fase successiva al concepimento dell'idea è stata quella di ricerca delle fonti dati utili alla realizzazione dell'infografica. A tale fine per le informazioni sul tema della sicurezza interna è stato utilizzato il dataset fornito Small Arm Survey, ente che raccoglie ogni anno dati relativi ad armi e alle violenze che avvengono all'interno degli Stati. Sono stati quindi scelti i due database più funzionali per la nostra ricerca: il *Global Firearms Holdings* e il *Global Violent Deaths* [5].

Considerando che alcune variabili contenute nel database *Global Firearms Holdings* risultavano parzialmente incomplete si è ritenuto opportuno mantenere solo quella relativa alla stima di armi da fuoco detenute in ogni paese. Per quanto concerne il database relativo alle *Global Violent Deaths* sono stati sfruttati tutti i dati relativi alla serie storica dei valori assunti dal tasso di morti violente per ogni Paese.



[rog] La fase di raccolta dei dati per lo studio della pace di un paese si è basata sui dati del Global Pace Index, prodotti dall'Institute for Economics and Peace (IEP), dal sito "Vision of Humanity"[1]. Gli indicatori del Worldwide Governance Indicators sono invece stati scaricati dalla Wolrdbank [2].

Inoltre, all'interno delle visualizzazioni sono stati mostrati anche i conflitti attualmente in atto, la cui fonte utilizzata per ottenere tali informazioni è stata wikipedia [4]. Infine, a supporto grafico, si sono utilizzate anche le informazioni riguardo le tipologia di governo politico adottato dal paese e divise in quattro macro categorie:

- Autocrazia chiusa
- Autocrazia elettorale
- Democrazia elettorale
- Democrazia liberale

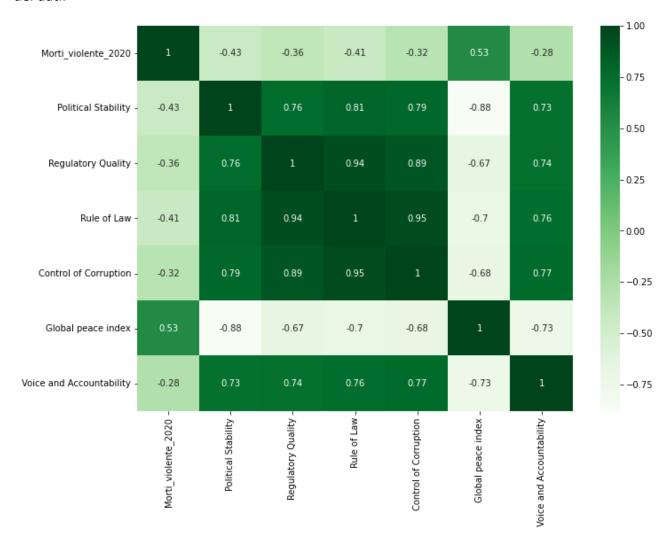
Tali dati si basano dalla classificazione Regimes of the World (RoW), che misura i sistemi politici con i dati del progetto Varieties of Democracy (V-Dem) [3].

1.3 Target di riferimento

Vista la natura della nostra ricerca il target di riferimento sono coloro che risultano interessati ad avere informazioni dettagliate su alcune variabili che influenzano le condizioni di vita dei vari Paesi del globo. In particolare potrebbe essere uno strumento utile per enti pubblici al fine di sensibilizzare la popolazione sulla situazione del proprio Stato, ma anche per società private che intendono espandersi internazionalmente.

2. Realizzazione della Data Visualization

Per rispondere alle domande poste come obiettivo si è iniziato con una analisi della correlazione dei dati.

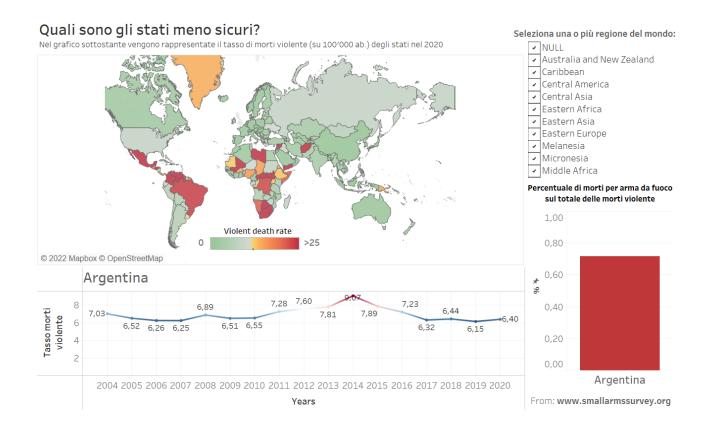


Da questo primo correlogramma si può notare subito come il Global peace index sia molto correlato con gli indicatori del WGI e in particolare con il "Voice and Accountability" e "Political Stability". Tuttavia, non è molto correlato con il tasso di morti violente.

Successivamente alla fase di raccolta dati e la prima analisi preliminare è stato necessario decidere quali grafici adottare per raffigurare ed esporre al meglio le informazioni a nostra disposizione. Per quanto riguarda la prima infografica (sicurezza interna) è stato scelto di adottare una mappa coropletica, in quanto è più immediata la visualizzazione delle differenze esistenti tra i diversi stati in termini di morti violente avvenute nel 2020.

I vari Paesi sono stati evidenziati da una banda di colori divergente che passa dal verde al rosso in modo da rappresentare in modo netto il valore assunto dal tasso di morti violente per ciascun Paese.

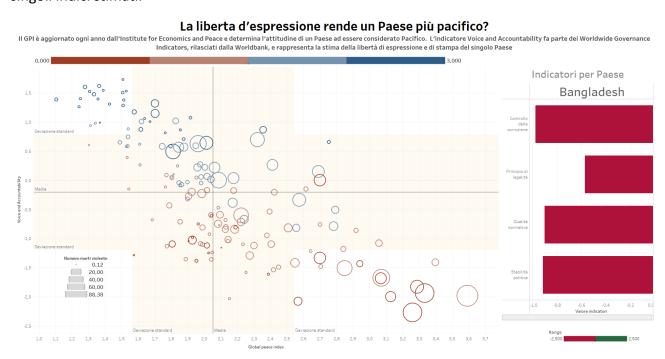
Alla mappa è stata affiancata una serie storica rappresentante l'evoluzione del tasso di morti violente dal 2004 al 2020: tale grafico mostra l'andamento per un singolo Paese una volta che questo viene selezionato sulla mappa. Anche in questo caso la linea della serie storica è stata colorata utilizzando una banda di colori divergente dal blu al rosso, così da evidenziare il punto di minimo relativo (blu) e il punto di massimo relativo (rosso) nell'intervallo temporale d'indagine. Infine è stato inserito un barplot, specifico per ogni Paese, che denota la percentuale di morti per via di armi da fuoco sul totale delle morti violente. Tale grafico ha la funzione di mostrare quanto le armi da fuoco siano presenti all'interno di uno Stato e quanto queste vengano utilizzate per compiere violenze.



Per il tema della correlazione fra la libertà d'espressione e il livello dell'indice di pace di un Paese, si è ritenuto di utilizzare un diagramma a dispersione così da poter mostrare immediatamente all'osservatore diversi aspetti di interesse dei dati. Tramite questa rappresentazione può infatti subito notare il grado di correlazione fra il Global Peace Index (sull'asse delle ascisse) e l'indicatore

"Voice and Accountability" (sull'asse delle ordinate) e come essi si distribuiscono. Questa scelta permette all'utente di individuare subito come i paesi con un sistema politico democratico si trovino sulla parte sinistra e alta del grafico, che corrisponde quindi ai paesi con un alto grado di libertà d'espressione e un elevato livello di pace. Infine, all'interno dello scatterplot si è inserita un'ulteriore dimensione, ovvero la grandezza dei punti, che serve per collegare il tema del livello di pace definito dall'indice e il numero di morti violente interne al Paese. Anche se, come si è potuto vedere dalla matrice di correlazione il GPI e il tasso di morti violente non è molto correlato, grazie a questa dimensione dello scatterplot si nota come sulle code dei dati viene rispettata l'ipotesi fatta a priori. Ovvero, sulla coda sinistra, fuori dalle bande della devianza, si hanno i paesi con una maggiore livello di pace e di libertà d'espressione e si può notare come abbiano tutti un basso tasso di morti violente. Di contro, sulla coda destra dei dati, si hanno i paesi con il più alto tasso di morti violente.

In questa Dashboard è stato poi affiancato un grafico a barre orizzontale che permette all'utente di conoscere i valori degli ulteriori indicatori del WGI. Questo grafico permette di capire rapidamente, grazie anche al supporto dei colori, se il Paese selezionato ha un valore positivo o negativo per i singoli indici stimati.



2.1 Accessibilità

Data la necessità di rendere disponibile a tutti l'interpretazione corretta delle infografiche è stato necessario intervenire sulla scala di colori utilizzata. Infatti, nonostante Tableau permetta di visualizzare le informazioni di ogni Paese semplicemente selezionando con il cursore il relativo pallino o la sua collocazione sulla mappa, è stato deciso di scegliere delle palette che definissero i colori in modo chiaro anche per le persone affetti da daltonismo. Tale scelta è stata anche motivata dalla centralità assunta dai colori nella mappa coropletica, che ricoprono un aspetto fondamentale per l'impatto visivo assorbito dall'utente.

Per la prima infografica la palette utilizzata è stata la seguente:

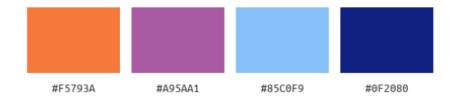
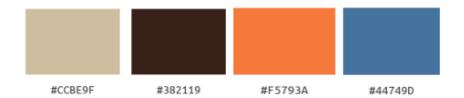


Fig 2.1 - Palette colori finali

Discorso equivalente per quanto riguarda il grafico a dispersione. Dato che nella seconda infografica le variabili che definiscono i colori non sono continue ma categoriche sono stati scelti alcuni colori da diverse palette in modo tale da massimizzare il contrasto visivo nello scatter plot e facendo sì che la visualizzazione potesse essere inclusiva anche per soggetti affetti da daltonismo.



3. Valutazione della qualità della data visualization

Per valutare la qualità e l'efficacia delle infografiche realizzate sono state adottate le seguenti metodologie:

- valutazione euristica, con la finalità di individuare le problematiche presenti nelle visualizzazioni e eventuali migliorie;
- test utente, che ha lo scopo di valutare l'efficacia (in termini di tempo necessario per eseguire un task) e l'efficienza (misurando il numero di task eseguiti con successo) delle infografiche;
- questionario psicometrico, per ottenere un quadro generale di valutazione della data visualization;

3.1 Valutazione euristica

Per la valutazione euristica sono stati intervistati 4 utenti, ai quali è stato chiesto di interagire con entrambe le visualizzazione per un breve lasso di tempo e di esprimere ad alta voce i propri commenti e pensieri su ciò che vedevano. Interpretando i comportamenti e le affermazioni degli utenti abbiamo identificato una lista di problemi di usabilità.

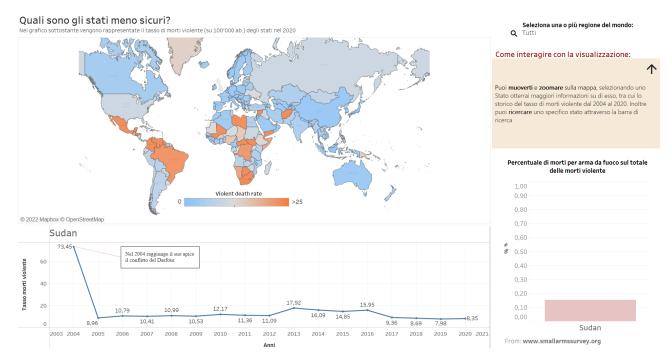
Utente 1	L'utente non mostra problematiche nell'interazione con la mappa, tuttavia nel momento in cui si focalizza sulla serie storica non comprende la presenza della scala di colori, ipotizzando l'esistenza di una variabile che non è ben specificata sul grafico
Utente 2	Dopo una rapida interazione con il secondo grafico l'utente non comprende che maggiore è il valore del GPI, minore è la stabilità socio-politica dello stato, di conseguenza fraintende totalmente lo scatterplot.
Utente 3	Pone molta attenzione alla variazione degli indicatori del grafico a barre nella seconda visualizzazione evidenziando però che in mancanza di una specifica descrizione essi non risultano del tutto chiari.
Utente 4	Interagendo con la prima infografica impiega diverso tempo per ricercare uno specifico stato e sottolinea che la barra di selezione risulta pesante e troppo generica (meglio direttamente sugli stati che sulle regioni). Inoltre nella seconda infografica non comprende il significato dei colori utilizzati nello scatterplot.

Modifiche alle infografiche:

In seguito ai risultati ottenuti dagli utenti con la valutazione euristica si è proceduto a modificare i grafici precedentemente creati.

Nella prima visualizzazione si è ritenuto opportuno introdurre una barra di ricerca a digitazione per gli Stati, così da rendere la ricerca sulla mappa più immediata.

Inoltre sono stati modificati i colori applicati alla mappa così da renderla accessibile anche ai soggetti daltonici. Si è anche introdotto un riquadro apposito a scomparsa per spiegare come interagire con l'infografica, in modo da semplificare il processo di familiarità dell'utente con essa. Infine l'ultima modifica apportata ha riguardato l'introduzione all'interno della serie storica di una serie di informazioni inerenti alle principali cause che hanno spinto alla crescita o alla decrescita del tasso di morti violente nei primi 15 stati per valore di quest'ultimo.

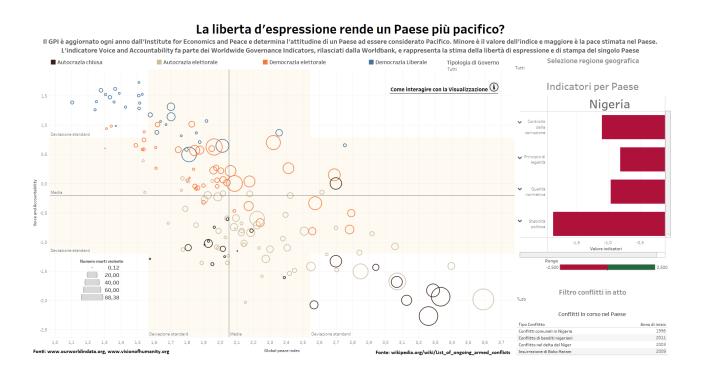


Per il diagramma a dispersione di sono quindi modificati i valori della variabile relativa alla tipologia della politica del paese, passando quindi da una variabile discreta a una categoriale. In questo modo la legenda e anche i colori sono diventati molto più intuitivi e di facile lettura.

Per quanto riguarda la problematica relativa all'interpretazione del GPI che risultava di difficile comprensione si è aggiunto nel sottotitolo una breve descrizione del suo funzionamento, aggiungendo quindi che "minore è il valore dell'indice e maggiore è la pace stimata nel Paese".

Successivamente, si è corretto anche l'aspetto relativo alla descrizione degli indicatori del WGI e cosa essi rappresentino. Questo problema è stato risolto aggiungendo un tasto alla sinistra di ogni singolo indicatore che apre una descrizione relativa al suo funzionamento.

Inoltre, la visualizzazione è stata migliorata aggiungendo le fonti in basso al grafico in modo da essere il più possibile trasparenti. È stato aggiunto anche un nuovo menù a tendina che permette di filtrare i paesi in base a fatto che siano o no impegnati in conflitti e sotto ad esso uno spazio dedicato a rappresentare quali siano i conflitti a cui partecipa e quando sono cominciati.



3.2 Test utente

Nella fase di valutazione successiva all'individuazione e correzione delle problematiche siamo passati alla fase della somministrazione di specifici task a 12 utenti. Per ogni utente abbiamo tenuto nota del tempo impiegato per portare a termine il compito a loro assegnato così da valutare *l'efficienza* delle visualizzazioni. Inoltre, abbiamo preso in considerazione il numero di task correttamente svolti in modo da valutare invece *l'efficacia* delle visualizzazioni.

Nella tabella di seguito vengono descritti i compiti richiesti agli utenti:

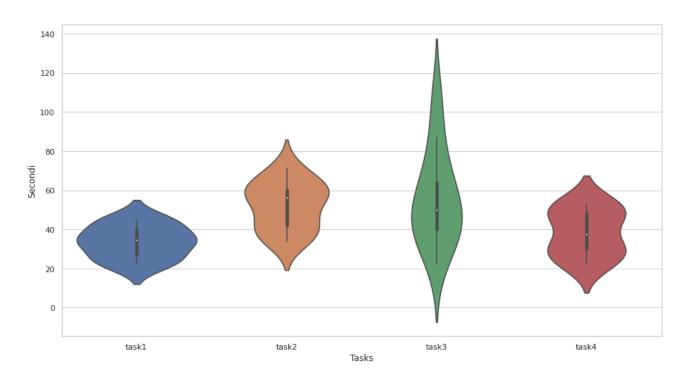
Task 1	Ricercare l'aiuto su come interagire con i grafico in modo da essere autonomo in caso di necessità	, , ,
Task 2	Valutare se l'utente è in grado di capire in quale delle due visualizzazioni si trovano le informazioni richieste. Inoltre permette di vedere se l'utente è in grado di utilizzare la	Salvador presenta la percentuale maggiore di morti per armi da

	ricerca per paesi.	
Task 3	Permette di capire se l'utente è in grado di utilizzare correttamente i filtri inseriti all'interno della visualizzazione.	Sapendo che la Nigeria è una Democrazia elettorale e ha dei conflitti in corso, quanti conflitti aveva in corso nel 2020?
Task 4	Questo task permette di capire se l'utente è in grado di identificare le eventuali cause che hanno influito ad aumentare il tasso di morti violente in un determinato anno nella storia di un Paese.	Quale evento storico ha determinato un aumento delle morti violente in Libia e in che anno?

I task sono stati proposti ad un pubblico misto di utenti senza esperienza o con una modesta esperienza in tableau. I task proposti sono in ordine crescente di difficoltà nell'utilizzo delle infografiche.

I risultati per determinare l'efficienza e l'efficacia ottenuti dai 12 utenti è riportata nella seguente tabella:

Utenti	Task 1	Task 2	Task 3	Task 4
Utente 1	32"	58''	73"	43"
Utente 2	24"	38"	42''	31"
Utente 3	38"	66"	87''	52"
Utente 4	45"	34"	53"	31"
Utente 5	44"	46"	37"	27"
Utente 6	22"	37"	60''	46"
Utente 7	28"	61"	108"	53"
Utente 8	23"	58"	36"	47''
Utente 9	35"	60''	41"	23"
Utente 10	44"	71"	61"	53"
Utente 11	34"	54"	22"	32"
Utente 12	35"	43"	47"	22"



I risultati sono rappresentati con l'utilizzo dei violin plot. Questi grafici hanno il vantaggio di mostrare in modo intuitivo la distribuzione dei dati identificando così rapidamente i casi unimodali, bimodali, ecc. Inoltre, al loro interno sono raffigurati dei boxplot che danno informazioni riguardo alla mediana, valore minimo e valore massimo.

Dal primo violin plot è possibile notare come la maggior parte degli utenti abbia impiegato circa 30 secondi a rispondere a questa domanda. Inoltre, questo primo grafico è anche il più compresso con valori minimi e massimi poco distanti, confermando l'idea che la prima domanda fosse la più semplice e che quindi è facile identificare il supporto al grafico dato dalle istruzioni sull'interazione. Il secondo task è invece risultato essere molto il più complesso da risolvere. Questo perché, anche se non è il task con valori di tempo più alti per il suo completamento, vediamo come sia quello con la mediana più alta. Questo però può essere giustificato dal fatto che è l'unico task che richiede la scrittura dei paesi e gli utenti possono confondersi e perdere del tempo anche nel capire che i paesi vanno scritti in inglese e non in italiano.

Dal terzo grafico è facile notare come il task 3 sia stato maggiormente impegnativo per alcuni utenti, portandoli al completamento in molto più tempo rispetto agli altri task proposti. Tuttavia, la mediana è in linea con gli altri task.

Infine, il grafico evidenzia come il task 4 sia bimodale. Questo significa che la maggior parte degli utenti si è concentrata sui 25 e i 45 secondi.

3.3 Questionario

Al termine della somministrazione del test utente, è stato deciso di utilizzare come questionario psicometrico per la valutazione della qualità della data visualization quello che adotta la scala Cabitza-Locoro.

Tale strumento si compone di due sezioni. La prima contiene una serie di aggettivi (utile, chiara, informativa, bella, intuitiva) relativi alla data visualization che l'utente deve valutare fornendo un valore che si trova su una scala da 1 (pochissimo) a 6 (moltissimo).

Nella seconda sezione viene richiesto all'utente di valutare la qualità complessiva percepita della data visualization, sempre attraverso una scala di valori che parte da 1 (bassissimo) e arriva a 6 (altissimo).

A tale questionario hanno partecipato 25 soggetti, la maggior parte dei quali sono studenti universitari (triennale e magistrale) e i restanti sono adulti.

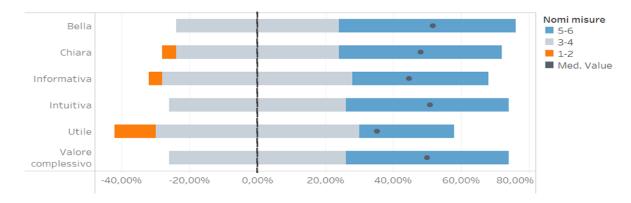
I risultati per la prima infografica sono i seguenti:

	Utile	Chiara	Informativa	Bella	Intuitiva	Valore complessivo
1	1	0	0	0	0	0
2	2	1	1	0	0	0
3	6	4	5	4	3	4
4	9	8	9	8	10	9
5	5	7	6	7	7	7
6	2	5	4	6	5	5

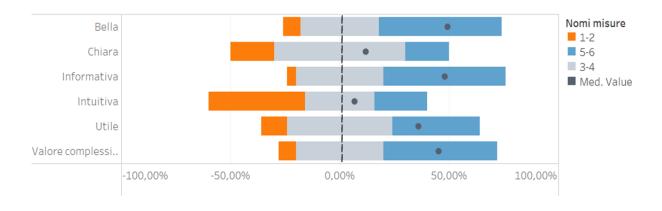
Per la seconda infografica i risultati sono invece i seguenti:

	Utile	Chiara	Informativa	Bella	Intuitiva	Valore complessivo
1	0	0	0	0	2	0
2	3	5	1	2	8	2
3	3	12	6	4	5	3
4	9	3	4	5	4	7
5	9	4	9	8	4	10
6	1	1	5	6	2	3

Di seguito l'analisi dei risultati del questionario, rappresentati tramite dei divergent stacked bar charts.

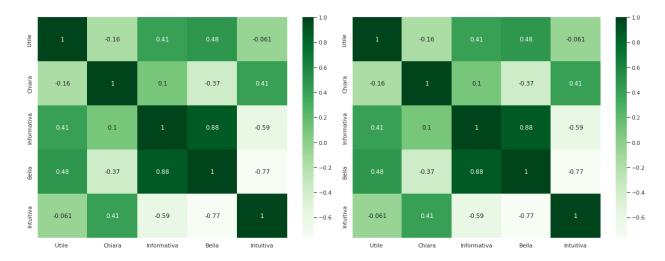


Risultati questionario psicometrico - Divergent Stacked Bar Chart, prima visualizzazione



Risultati questionario psicometrico - Divergent Stacked Bar Chart, seconda visualizzazione

Il primo grafico mostra risultati più che buoni per tutte le dimensioni, in particolare l'infografica risulta bella, intuitiva, chiara e informativa. Il grafico relativo alla seconda visualizzazione invece mostra valori mediamente inferiori. Notiamo che l'infografica, come potevamo aspettarci, risulta meno intuitiva e chiara. Tuttavia, rimangono alti i valori degli altri aspetti.



Grazie ai correlogrammi è possibile evidenziare la correlazione ottenuta dai diversi punteggi. Dal primo correlogramma si vede come "Bella" sia fortemente correlata con "Intuitiva" e "Chiara". Il secondo correlogramma invece dimostra come ci sia forte correlazione fra "Informativa" e "Bella".

4. Conclusioni

In seguito alle analisi svolte la visualizzazione ha mostrato come si distribuiscono i Paesi in base al loro grado di GPI e di libertà di espressione. Inoltre, è stato reso più chiaro quali paesi abbiano un tasso elevato di morti violente grazie alla mappa coropletica e permettendo così di rispondere alla prima domanda di questo report. La seconda domanda, invece, riguardava la correlazione fra la libertà d'espressione interna di un Paese e il livello di pace stimato dall'indice GPI. Dal correlogramma prima e dallo scatterplot poi, è possibile vedere come al diminuire della libertà di stampa diminuisca anche il livello di pace di un paese. Inoltre agli estremi dello scatterplot, è possibile vedere come sulla coda sinistra si hanno i paesi con una maggiore livello di pace e di libertà d'espressione e si può notare come abbiano tutti un basso tasso di morti violente. Di contro, sulla coda destra dei dati, si hanno i paesi con il più alto tasso di morti violente.

5. Fonti

- [1] Vision Of Humanity: https://www.visionofhumanity.org/maps/#/
- [2] Wolrdbank: http://info.worldbank.org/governance/wgi/
- [3] Our World in Data https://ourworldindata.org/
- [4] wikipedia: https://en.wikipedia.org/wiki/List_of_ongoing_armed_conflicts
- [5] Small Arms Survey: https://www.smallarmssurvey.org