

**POLITECNICO DI MILANO**

**Scuola di Ingegneria Industriale e dell'Informazione  
Corso di Laurea Magistrale in Computer Science and Engineering  
Dipartimento di Elettronica, Informatica e Bioingegneria**



**Una comparazione tra tecniche tradizionali e  
innovative di Data Visualization**

**Relatore: Prof. Emanuele DELLA VALLE  
Correlatore: Ing. Andrea POZZETTI**

**Tesi di Laurea Magistrale di:  
Alessandro Donini - matr. 836802**

**Anno Accademico 2015 - 2016**



*Alla mia famiglia,  
a Martina, i miei amici  
e in memoria di Iole ...*



# Ringraziamenti

In primo luogo desidero ringraziare il Professor Emanuele Della Valle, il quale mi ha dato la grande opportunità di approfondire un argomento che suscitava un forte interesse in me. E' stato un lavoro lungo e impegnativo, ma ha permesso di stimolare ulteriormente questa mia passione. Sono grato anche ad Andrea Pozzetti e Andrea Vaccarella, per la loro guida e il loro supporto in numerose situazioni lungo tutto il periodo legato alla ricerca. Ringrazio tutti i componenti dei *Bargis' Lovers*, un gruppo di ingegneri ormai, ma prima ancora di amici sinceri, dai quali ho potuto imparare tantissimo, e crescere al punto da arrivare fino a qui. In particolare vorrei ringraziare Leonardo, Lorenzo e Rosa, con i quali ho passato senza ombra di dubbio i periodi più belli, più divertenti e più stressanti di tutta la carriera universitaria. Ringrazio Simone, per esserci stato in molti momenti difficili e per avermi insegnato che arrabbiarsi spesso e volentieri è bene. Ringrazio ancora Gian, sempre presente e disponibile in ogni situazione, soprattutto nelle lunghe notti in patio dove abbiamo dato vita a numerosi progetti. Ringrazio gli amici di sempre, per essersi sbronzati anche per me nelle notti di luna piena mentre io ero in patio con Gian. I ragazzi di Cremona, o Piacenza, insomma da quelle parti lì, che nonostante il poco tempo passato assieme sono riusciti a lasciare un bellissimo ricordo. Ringrazio Martina, che mi è stata accanto in ogni momento felice, triste, nervoso, felice, nervoso ... nervoso ... e poi ancora nervoso, senza però mai abbandonarmi. Ultima ma non per questo meno importante, ringrazio la mia famiglia, per avermi sempre sostenuto e aiutato in situazioni che da solo non avrei mai potuto superare.

Grazie a tutti, grazie per avermi sempre spronato a dare il massimo, perché se ci si mette tutto l'impegno, alla fine ne vale la pena.

Ale



# Abstract

*Data Visualization* is becoming one of the most important phenomenon that characterizes the Data Analysis environment. If until a few years ago it was enough to use bar and pie charts to analyze and visualize data, with the rapid growth in data production that we are experiencing today, new techniques need to be explored. The benefits introduced by the techniques of *Data Visualization*, which exploits the ability of the human perceptual system, are becoming even more important and have captured the attention of academics, as a result, the number of techniques available in the market have significantly increased.

In the light of these considerations, it is interesting address the following questions: “*are taxonomies and the automatic graphics systems detections existing today still the right tools to identify which is the best technique to use?*” and “*are we confident that the traditional techniques used until now, are still the most appropriate?*”

The first part of this thesis focuses on the description of the exploratory phase, in order to understand the context and how to use the offered advantages. The second part discusses the study of the aggregate type of data that has been analysed, and how, based on this, we have addressed the investigation to identify all suitable *Data Visualization* techniques, both *traditional* that *innovative*. It follows the definition of the user-oriented methodology adopted to compare the visualization found, which is divided into three surveys. A first study to compare traditional chart used to show the type of data of our case study. A second one in which the previous best techniques are compared with those innovative. Finally a comparison of the two methods that were found to be the most effective during the previous investigations. This has led to the development of a Web Application, where it is possible to view and interact with the visualizations, and identify the features of interest.

The objective of this thesis is to call into question the traditional techniques, demonstrating that innovative techniques, until now rarely used, are able to show correct data through a more effective representation, thus allowing to improve the process related to analysis of data. Such comparison was performed through the definition of a user-oriented methodology, which can fit into a broader context and make a contribution in the field of *Data Visualization* techniques.



# Sommario

L'importanza assunta dalla *Data Visualization* negli ultimi anni è uno dei fenomeni che più caratterizza l'ambiente della analisi dei dati. Se fino a pochi anni fa si poteva considerare sufficiente l'utilizzo di diagrammi a barre e di grafici a torta per effettuare analisi e mostrare informazioni, con la rapida crescita nella produzione dei dati a cui stiamo assistendo questo non può più essere vero. I benefici introdotti dalle tecniche di *Data Visualization*, le quali sfruttano le capacità del sistema percettivo umano, sono sempre più apprezzati, e con l'aumento di interesse da parte della ricerca, sono aumentate notevolmente le tecniche oggi disponibili.

Alla luce di questo, risulta interessante porsi le seguenti domande: *“le tassonomie e i sistemi automatici di individuazioni dei grafici oggi esistenti, sono ancora strumenti adatti per identificare quale sia la tecnica migliore da utilizzare?”* e *“siamo sicuri che le tecniche tradizionali, utilizzate fino ad oggi, siano ancora le più appropriate?”*

La prima parte di questa tesi si concentra sulla descrizione della fase esplorativa, volta a conoscere il contesto e a comprendere come sfruttare i vantaggi offerti. Nella seconda parte viene descritto lo studio della tipologia di dato aggregato che abbiamo analizzato, e di come, sulla base di questo, abbiamo affrontato la fase di ricerca per l'identificazione di tutte le tecniche di *Data Visualization* adatte, sia *tradizionali* che *innovative*. Segue la definizione della metodologia user-oriented adottata per la comparazione di queste, la quale è suddivisa in tre indagini. Una prima indagine per confrontare le visualizzazioni tradizionali utilizzate per mostrare la tipologia di dato presa in esame. Una seconda nella quale vengono confrontate le precedenti migliori tecniche con quelle innovative. Infine è stato effettuato il confronto e lo studio dei due metodi che sono risultati essere i più efficaci durante le indagini precedenti. Questo è stato possibile grazie allo sviluppo di una applicazione Web tramite la quale è possibile visionare e interagire con le visualizzazioni e identificarne tutte le caratteristiche.

L'obiettivo è quello di mettere in discussione le tecniche tradizionali, dimostrando che grafici innovativi, fino ad oggi poco utilizzati, sono capaci di mostrare i dati in maniera corretta attraverso una rappresentazione più efficace, permettendo così di migliorare i processi relativi alla analisi dei dati. Tale comparazione è stata effettuata attraverso la definizione di una metodologia user-oriented, la quale può inserirsi in un contesto più ampio e contribuire allo studio delle tecniche di *Data Visualization*.

# Indice

<b>Ringraziamenti</b>	<b>I</b>
<b>Abstract</b>	<b>III</b>
<b>Sommario</b>	<b>V</b>
<b>1 Introduzione</b>	<b>1</b>
1.1 Lavori correlati e motivazioni . . . . .	3
1.2 Descrizione del problema . . . . .	4
1.3 Soluzione proposta . . . . .	5
1.4 Struttura della Tesi . . . . .	6
<b>2 Stato dell'Arte</b>	<b>7</b>
2.1 Il valore dei dati . . . . .	8
2.2 Data Visualization . . . . .	10
2.2.1 Eccellenza nei grafici . . . . .	11
2.2.2 Integrità dei grafici . . . . .	11
2.2.3 Utilizzo degli spazi e dei colori . . . . .	12
2.3 La percezione umana . . . . .	14
2.4 Interagire coi dati . . . . .	16
2.5 Lavori correlati . . . . .	18
2.5.1 Tassonomie . . . . .	19
2.5.2 Altri lavori correlati . . . . .	20
<b>3 Definizione del Problema e Metodologia Proposta</b>	<b>23</b>
3.1 Il problema . . . . .	24
3.2 Metodologia proposta . . . . .	25
3.2.1 Definizione del caso studio . . . . .	25
3.2.2 Individuazione dei grafici . . . . .	25
3.2.3 Comparazione dei grafici . . . . .	26

<b>4</b>	<b>Soluzione del Problema</b>	<b>29</b>
4.1	Definizione del caso studio . . . . .	30
4.1.1	MasterChef Italia . . . . .	30
4.1.2	MasterChef Italia e i Social . . . . .	30
4.1.3	Struttura del dato . . . . .	31
4.2	Identificazione dei grafici . . . . .	32
4.2.1	Stacked Bar Graph (o Stacked Bar Chart) . . . . .	34
4.2.2	Histogram . . . . .	35
4.2.3	Donut Chart . . . . .	36
4.2.4	Pie Chart . . . . .	37
4.2.5	Radar Chart . . . . .	38
4.2.6	Radial Column Chart . . . . .	39
4.2.7	Parallel Sets . . . . .	40
4.2.8	Circle Packing . . . . .	41
4.2.9	Heat Map . . . . .	42
4.3	Comparazione dei grafici . . . . .	43
4.3.1	Comparazione tra visualizzazioni tradizionali . . . . .	43
4.3.2	Comparazione con visualizzazioni innovative . . . . .	47
4.4	Comparazione tra le visualizzazioni migliori . . . . .	50
4.4.1	Introduzione alle tecnologie . . . . .	50
4.4.2	Descrizione dell'applicazione . . . . .	52
4.4.3	Requisiti di interfaccia utente . . . . .	54
4.4.4	Requisiti funzionali . . . . .	60
4.4.5	Struttura delle domande . . . . .	66
<b>5</b>	<b>Analisi dei Risultati e Valutazioni</b>	<b>69</b>
5.1	Risultati della comparazione tra visualizzazioni tradizionali . . . . .	70
5.2	Risultati della comparazione con visualizzazioni innovative . . . . .	75
5.3	Risultati della comparazione tra le visualizzazioni migliori . . . . .	83
5.3.1	Risultati prima visualizzazione proposta . . . . .	84
5.3.2	Risultati seconda visualizzazione proposta . . . . .	87
<b>6</b>	<b>Conclusioni e Direzioni Future</b>	<b>91</b>
6.1	Conclusioni . . . . .	92
6.2	Limiti . . . . .	93
6.3	Direzioni future . . . . .	94
	<b>Bibliografia</b>	<b>95</b>

# Elenco delle figure

1.1	Mancanza di conoscenza . . . . .	1
2.1	Bar Chart di <i>William Playfair</i> . Importazioni ed esportazioni della Scozia nel 1781 per 17 paesi . . . . .	10
2.2	Varietà delle tecniche di Data Visualization . . . . .	18
4.1	Logo MasterChef Italia . . . . .	30
4.2	Stacked Bar Graph (o Stacked Bar Chart) . . . . .	34
4.3	Histogram . . . . .	35
4.4	Donut Chart . . . . .	36
4.5	Pie Chart . . . . .	37
4.6	Radar Chart . . . . .	38
4.7	Radial Column Chart . . . . .	39
4.8	Parallel Sets . . . . .	40
4.9	Circle Packing . . . . .	41
4.10	Heat Map . . . . .	42
4.11	Grafici confrontati durante la comparazione tra visualizzazioni tradizionali . . . . .	44
4.12	Grafici confrontati durante la comparazione con visualizzazioni innovative . . . . .	48
4.13	MockUp Home - Start . . . . .	55
4.14	MockUp Home - Descrizione della Ricerca . . . . .	55
4.15	MockUp Home - Link ai risultati . . . . .	56
4.16	MockUp Home - Chi siamo . . . . .	56
4.17	MockUp Introduction Area - Descrizione dell'attività richiesta all'utente . . . . .	57
4.18	MockUp Introduction Area - Descrizione del caso studio . . . . .	57
4.19	MockUp Visualization Area - Autovalutazione . . . . .	58
4.20	MockUp Visualization Area - Stacked Bar Graph . . . . .	58
4.21	MockUp Visualization Area - Circle Packing . . . . .	59
4.22	MockUp Thanks Area - Ringraziamenti finali . . . . .	59

4.23	Sequence Diagram - Introduzione all'attività richiesta . . . . .	61
4.24	Sequence Diagram - Questionario . . . . .	63
4.25	Sequence Diagram - Visualizzazione dei grafici . . . . .	65
4.26	Distorsioni grafiche del Circle Packing . . . . .	67
5.1	Grafici confrontati durante la comparazione tra visualizzazioni tradizionali . . . . .	70
5.2	Distribuzione Demografica . . . . .	71
5.3	Risultati relativi alle reazioni degli intervistati - Comparazione tra visualizzazioni tradizionali . . . . .	72
5.4	Risultati ottenuti dal confronto tra grafici simili - Comparazione tra visualizzazioni tradizionali . . . . .	73
5.5	Risultati delle domande di controllo - Comparazione tra visualizzazioni tradizionali . . . . .	74
5.6	Grafici confrontati durante la comparazione con visualizzazioni innovative . . . . .	75
5.7	Livello di conoscenza degli intervistati - Comparazione con visualizzazioni innovative . . . . .	76
5.8	Grafici conosciuti dagli intervistati suddivisi per livello di conoscenza - Comparazione con visualizzazioni innovative . . . . .	77
5.9	Cosa gli intervistati comprendono dal grafico mostrato rispetto allo Stacked Bar Chart - Comparazione con visualizzazioni innovative . . . . .	78
5.10	Risultati relativi ai grafici intuitivi e non intuitivi - Comparazione con visualizzazioni innovative . . . . .	79
5.11	Risultati relativi all'appeal dei grafici - Comparazione con visualizzazioni innovative . . . . .	79
5.12	Risultati delle domande di controllo - Comparazione con visualizzazioni innovative . . . . .	81
5.13	Distribuzione degli intervistati in base a livello di conoscenza nella Data Visualization . . . . .	83
5.14	Chi è lo Chef più popolare su Instagram? - Prima visualizzazione proposta . . . . .	84
5.15	Chi è lo Chef più popolare alle 18:00? - Prima visualizzazione proposta . . . . .	85
5.16	Credi che il grafico che ti ho mostrato sia intuitivo? - Prima visualizzazione proposta . . . . .	86
5.17	Sei in grado di identificare con facilità i valori singoli e totali? - Prima visualizzazione proposta . . . . .	86

5.18	Chi è lo Chef più popolare su Instagram? - Seconda visualizzazione proposta . . . . .	87
5.19	Chi è lo Chef più popolare alle 18:00? - Seconda visualizzazione proposta . . . . .	88
5.20	Credi che il grafico che ti ho mostrato sia più intuitivo rispetto al precedente? - Seconda visualizzazione proposta . . . . .	89
5.21	Sei in grado di identificare i valori singoli e totali con una maggiore facilità rispetto al grafico che ti ho mostrato in precedenza? - Seconda visualizzazione proposta . . . . .	89





# Capitolo 1

## Introduzione

*“Despite their potential, the benefits of data visualization are undermined today by a general lack of understanding.”*

Stephen FEW

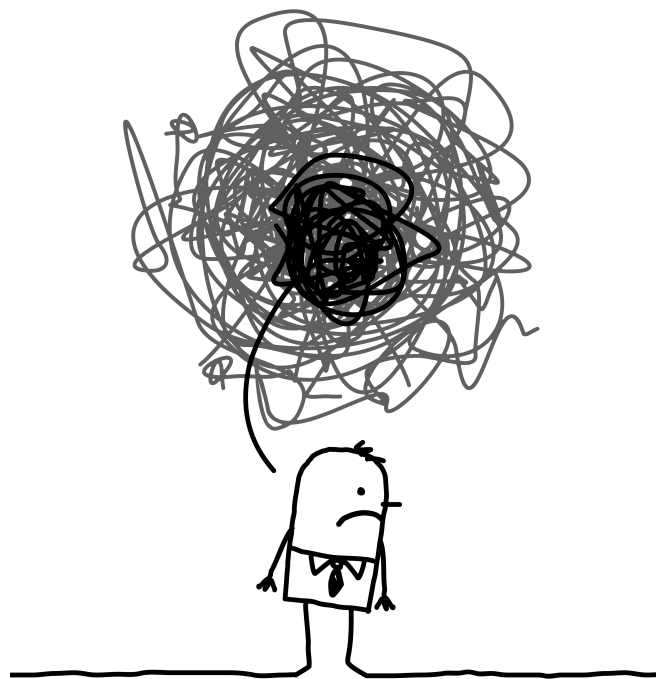


Figura 1.1: Mancanza di conoscenza

Oggi giorno stiamo assistendo ad una crescita esponenziale nella produzione dei dati, ed ora più che mai è di primaria importanza, non solo creare infrastrutture in grado di gestire tali cumuli, ma anche progettare strumenti in grado di trasformare queste grandi quantità in informazione. L'incontro tra gli ambiti di ricerca della *Computer Science*, *Information Design* e *Human Perception* ha dato il via a quella che oggi è conosciuta come *Data Visualization*.

La *Data Visualization* è un campo di ricerca multidisciplinare, che si focalizza sullo studio e sulla creazione di rappresentazioni visive dei dati [42], con l'obiettivo di estrapolare informazioni e permettere l'analisi di queste in modo chiaro ed efficiente.

Negli ultimi anni questa disciplina è diventata sempre più apprezzata per i benefici che sta apportando nell'ambito della analisi dei dati. Essa infatti fornisce ottimi strumenti sia per trasformare i dati in informazioni, che per comprendere e condividere in maniera più accessibile queste con altre persone. Nonostante il suo potenziale però, tali vantaggi sono ancora oggi sottovalutati a causa di una generale disinformazione nei confronti della materia. Inoltre, molte tendenze in questa area, stanno producendo l'effetto opposto rispetto a quello desiderato, creando confusione piuttosto che comprensione. Per sfruttare i suoi benefici quindi, dobbiamo capire e abbracciare ciò che realmente funziona, gettando in mare le cose superflue che minano la comprensione dei dati attraverso l'uso delle visualizzazioni [10].

Un'ulteriore elemento penalizzante è dato dal forte aumento delle tecniche di visualizzazione oggi esistenti, causato da un crescente interesse da parte della *ricerca* e della *Business Intelligence*. Questi fattori vanno in direzione opposta rispetto all'obiettivo della *Data Visualization*: fornire strumenti volti ad un'interpretazione chiara ed efficiente del significato che racchiudono i dati. Tale operazione però sta diventando sempre più complessa data la grande varietà di tecniche oggi a disposizione (utilizzate in differenti scenari), la quale causa significative difficoltà nella progettazione delle rappresentazioni dei dati. Risulta quindi necessario sapere quale di queste sia la più adatta a seconda del contesto in cui ci troviamo [21].

## 1.1 Lavori correlati e motivazioni

Con la crescente attenzione nei riguardi di questa materia, è diventato sempre più importante il ruolo delle tassonomie, in quanto, studio delle teorie e delle regole di classificazione. Queste tassonomie, chiamate “*Visualization Taxonomies*”, hanno principalmente due scopi:

1. **Guidare l’utente.** Le persone al di fuori delle “*visualization community*” hanno difficoltà nel trovare in letteratura le giuste informazioni, utili per una corretta progettazione delle visualizzazioni, se queste non sono categorizzate in modo significativo.
2. **Guidare la ricerca.** I ricercatori necessitano di comprendere in quale contesto si inserisce il loro lavoro di ricerca e di individuare persone che svolgono ricerche simili alle loro. Inoltre la ricerca si focalizza maggiormente e cresce più rapidamente quando aumenta l’interesse e la compressione per una area specifica. [36].

La maggior parte delle “*Visualization Taxonomies*” oggi esistenti si basano sulla tipologia di dato coinvolta. I dati possono avere molte caratteristiche incluse [36, 40]:

- Numero di variabili indipendenti
- Numero di variabili dipendenti
- Tipologia di variabile (scalare, discreta, nominale)

Altre tassonomie invece, come quella proposta da *Melanie Tory* e *Torsten Möller* si basano sul modello dei dati [36], mentre altre ancora sono basate sulla tipologia di grafico o sul tipo di informazione che si intende estrapolare.

In aggiunta alle tassonomie, esistono anche strumenti di selezione automatica delle visualizzazioni, i quali permettono di identificare, sulla base della struttura del dato fornito in ingresso, la rappresentazione grafica più adatta. Questi sistemi vengono oggi utilizzati in numerosi applicativi, come ad esempio in *Tableau* [33], uno strumento di visual analytics il quale, fornite le dimensioni da analizzare, suggerisce le visualizzazioni più adeguate tra quelle contenute nel proprio catalogo.

Sia le tassonomie che i sistemi di selezione automatica, i quali verranno descritti in dettaglio nel Capitolo 2, forniscono una buona base di partenza per identificare le visualizzazioni più adatte al proprio scopo, ma considerano solo tecniche tradizionali, e citando la definizione di *Data Visualization* data da *Stephen Few* “...*communicate what we have discovered to others.*” [10], la

manca a tutti questi schemi credo sia il fatto di non considerare se l'utente finale, “...to others”, sia realmente in grado di “leggere”, in maniera rapida ed efficiente, la visualizzazione proposta.

## 1.2 Descrizione del problema

Oggi, la *Data Visualization* sta rapidamente prendendo il posto come pilastro portante nell'ambito della *Business Intelligence*. Con la crescente consapevolezza dei grandi vantaggi che si possono ottenere attraverso questo strumento, le aziende sono sempre più alla ricerca di persone che lavorino con i dati, come ad esempio gli analisti, ovvero figure in grado di sfruttare al meglio i benefici offerti. Inoltre, sta diventando un'area di ricerca importante anche nell'ambito accademico. Molte università hanno ora facoltà focalizzate sullo studio della visualizzazione dei dati, ed alcune di queste offrono eccellenti programmi a servizio degli studenti, i quali consentono di produrre utili studi di ricerca e prototipi interessanti [10].

L'interesse verso questa disciplina quindi è cresciuto notevolmente, ed è facile mostrare come la scelta della visualizzazione più appropriata possa fornire la chiave per una migliore comprensione dei dati, beneficiando conseguentemente di un'analisi più accurata e rendendo più efficienti i processi decisionali [29].

Nella Sezione 1.1 è stata descritta quella che consideriamo essere una lacuna riguardo gli studi fino ad ora effettuati in tale disciplina: comprendere se l'utente finale, cioè colui che dovrà effettuare le analisi e prendere delle decisioni sulla base di queste, sia effettivamente in grado di analizzare la rappresentazione dei dati individuata attraverso gli strumenti oggi disponibili, in maniera rapida ed efficiente.

Possiamo quindi iniziare a introdurre quelle che sono state le nostre domande preliminari alla base di questa ricerca: “*le tassonomie e i sistemi automatici di individuazioni dei grafici oggi esistenti, sono ancora strumenti adatti per identificare quale sia la tecnica migliore da utilizzare?*” e “*siamo sicuri che le tecniche tradizionali, utilizzate fino ad oggi, siano ancora le più appropriate?*”

## 1.3 Soluzione proposta

Sulla base delle nostre domande, ci siamo imposti l'obbiettivo di trovare una valida metodologia user-oriented per identificare le migliori tecniche di visualizzazione dati tra l'ampio catalogo di quelle oggi offerte, focalizzandoci sulla percezione e sulla lettura del dato. La metodologia che abbiamo adottato è suddivisa in 5 fasi:

1. **Caso studio.** Il primo passo è stato definire un dominio di interesse, individuando la tipologia di dato da studiare e il tipo di informazione da trasmettere.
2. **Esplorazione.** Nella fase esplorativa è stato necessario identificare tutte le possibili tecniche adatte a mostrare la tipologia di dato scelta, selezionando sia quelle tradizionali, generalmente utilizzate e in accordo con lo stato dell'arte, che tra quelle innovative, non generalmente utilizzate ma idonee a mostrare il dato grazie alla loro struttura.
3. **Indagine con dati fittizi.** Per effettuare questa indagine abbiamo definito un questionario ad hoc. L'obbiettivo è quello di collezionare i feedback degli utenti riguardo alle tecniche di visualizzazione selezionate per il nostro caso studio. Queste sono state create in modo da mostrare il dato nella maniera più chiara possibile così da esaltarne i pregi.
4. **Indagine con dati reali.** Una volta selezionate le tecniche che hanno riscosso un maggiore successo in termini di comprensione dei dati e di efficacia di lettura, abbiamo creato un ultimo questionario. Al contrario del precedente, le visualizzazioni sono state progettate in modo da renderle interattive, attingendo a dati provenienti da fonti reali e mettendo così in evidenza non solo i pregi ma anche i difetti.
5. **Valutazioni finali.** Una volta collezionati tutti i feedback derivanti dai precedenti step, è necessario analizzarli, così da verificare se le nostre domande e la nostra ricerca ha avuto effettivamente un riscontro positivo.

## 1.4 Struttura della Tesi

La presente tesi è strutturata nel seguente modo:

- **Capitolo 2: Stato dell'Arte**

Nel secondo capitolo si introduce al lettore il tema della *Data Visualization*, descrivendo lo stato dell'arte con un breve accenno di storia, per poi focalizzarci maggiormente sulla progettazione delle visualizzazioni e sui lavori correlati alla nostra ricerca. Il capitolo contiene inoltre informazioni circa gli studi relativi alla percezione umana dei grafici e una breve introduzione sull'importanza dei dati.

- **Capitolo 3: Definizione del Problema e Metodologia Proposta**

Nel terzo capitolo introduciamo in modo più circostanziato la lacuna che ha dato il via alla ricerca affrontata in questa tesi, dando una definizione più ampia del problema e una descrizione dettagliata dei requisiti e del procedimento proposto per la risoluzione di questo.

- **Capitolo 4: Soluzione del Problema**

Nel quarto capitolo viene trattato invece l'esperimento che abbiamo effettuato per la validazione della nostra proposta, dove è descritto nel dettaglio il caso studio e come questo è stato gestito durante tutte le fasi della ricerca, focalizzandoci su quelli che sono stati gli aspetti implementativi principali.

- **Capitolo 5: Analisi dei Risultati e Valutazioni**

Nel quinto capitolo sono invece descritte le osservazioni e le valutazioni dei risultati raccolti durante tutta le fasi dell'indagine, con la definizione degli strumenti utilizzati.

- **Capitolo 6: Conclusioni e Direzioni future**

Infine, nel sesto e ultimo capitolo, vengono descritte quelle che sono le conclusioni che abbiamo tratto a valle di questa ricerca, alcune osservazioni in merito all'efficacia del metodo proposto e i possibili sviluppi futuri del lavoro svolto.

## Capitolo 2

# Stato dell'Arte

*“Graphical excellence is that which gives to the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space.”*

Edward R. TUFTE

In questo capitolo viene fornita una panoramica generale della situazione corrente nel mondo della *Data Visualization*. Per comprendere a pieno il contesto da cui trae origine questa disciplina, è indispensabile, così come anticipato nel capitolo precedente, intrecciare informazioni che derivano da 3 ambiti di ricerca: *Computer Science*, *Information Design* e *Human Perception*.

Inizieremo introducendo il concetto che sta alla base della *Data Visualization*, ovvero il concetto di dato. Successivamente vedremo come i primi statisti, grazie all'utilizzo di tecniche visive, furono in grado estrapolare informazioni dai dati in maniera rapida ed efficiente. Infine verrà descritto come l'utilizzo delle nuove tecnologie abbia introdotto l'interazione coi dati e modificato il concetto di analisi.

## 2.1 Il valore dei dati

I dati sono diventati un fattore così importante nella nostra società tanto quanto lo è il lavoro, il capitale, e la terra [39]. Ma che cosa sono i dati? E perché sono così importanti?

Un dato (dal latino *datum* che significa letteralmente “fatto”) è una descrizione elementare, spesso codificata, di una cosa, di una transazione, di un avvenimento o di altro [43]. Il dato rappresenta quindi qualcosa che si può registrare, analizzare e riorganizzare. Trasformare un’attività elementare in dato, quale la temperatura o la data di acquisto di un libro, significa convertirla in una forma quantitativa, in modo che possa essere tabulata, permettendone l’analisi [23]. La successiva elaborazione di questi dati consente poi di ottenere informazioni, e oggi, grazie alle nuove tecnologie di analisi, tali informazioni diventano sempre più importanti e rappresentative della realtà. L’informazione quindi descrive il contesto in cui i dati sono raccolti, codificandoli in una forma tale che può essere compresa con il solo intelletto.

Queste informazioni possono essere ricavate dai luoghi più improbabili, come dimostrato da *Shigeomi Koshimizu*, professore dell’*Advanced Institute of Industrial Technology* di Tokyo. Il Prof. *Koshimizu* ha identificato nel modo in cui si siede un individuo, una grande fonte di informazioni. La distribuzione del peso, la postura e i contorni del corpo possono essere quantificati e tabulati. *Koshimizu*, assieme al suo team di ingegneri, ha analizzato i dati derivanti dalla misurazione della pressione esercitata in 360 punti del sedile di un’auto. Questo è stato possibile grazie all’utilizzo di appositi sensori e ha indicizzato ciascuna misurazione su una scala da 1 a 256. Il risultato così ottenuto ha permesso di identificare un codice digitale univoco per ogni singolo individuo. Grazie all’applicazione di tali risultati è stato costruito un sistema, il quale, in un esperimento, è riuscito a distinguere i singoli soggetti all’interno di un gruppo con una precisione del 98% [26]. Grazie all’utilizzo di questa tecnologia è possibile oggi ottenere, attraverso l’analisi dei dati della posizione di seduta, un servizio prezioso. Le potenzialità scaturite da questa analisi sono svariate. Per esempio si è in grado di percepire un lieve scivolamento in avanti del guidatore dovuto alla stanchezza, il quale può far scattare il sistema di sicurezza di un’automobile attivando un allarme o i freni, oppure riconoscere il legittimo proprietario di un’auto, bloccando la vettura nel caso in cui alla guida ci sia un’altra persona. Con questa ricerca il professore *Koshimizu* ha dimostrato come, una cosa che non era mai stata considerata una fonte di dati, possa contenere al suo interno informazioni molto preziose [23].



Parole, localizzazione e interazioni costituiscono tre tipologie di fenomeni fondamentali che possono essere trasformati in informazioni. *Google*, attraverso un'analisi delle parole contenute nei libri, effettuata tramite software ottici, è riuscita a catalogare gran parte dei libri esistenti, permettendo così delle ricerche basate sulle parole contenute. I sistemi GPS che sfruttano la rete mobile permettono di individuare la nostra posizione in maniera quasi istantanea, abilitando nuovi servizi. Ad esempio, aziende come *UPS* possono monitorare la posizione dei propri furgoni ottimizzando i percorsi per effettuare consegne più rapide, mentre altre aziende hanno potuto sfruttare le informazioni derivanti dalla nostra posizione per avviare campagne pubblicitarie più mirate. Processando enormi quantità di dati sono stati sperimentati alcuni strumenti in grado di identificare coloro che hanno contratto l'influenza prima ancora che sapessero di essere ammalate, questo grazie alla sola analisi dei dati provenienti dalle chiamate e dalle posizioni fornite dai telefoni cellulari [14].

Nuove tipologie di dati si possono oggi ottenere attraverso il web, che trova nei social network la sua fonte principale. Questi dati riguardano i nostri sentimenti, le nostre idee e le nostre amicizie. Più o meno consciamente forniamo a siti come *Facebook*, *Twitter* o *LinkedIn*, per citarne alcuni tra i più famosi, sempre più informazioni personali. Le potenziali applicazioni di queste piattaforme sono infinite, e vanno oltre la condivisione delle foto e i "mi piace". Si pensi per esempio alle analisi che le aziende effettuano sulla base dei commenti dei propri utenti col fine di comprendere se una campagna pubblicitaria è stata efficace, o ancora alle analisi dei tweet riguardanti un film per capire se questo sarà di successo oppure un fallimento.

Quando saremo in grado di analizzare tutta questa mole di dati, potremo far luce sulle dinamiche sia sociali che del singolo individuo nella loro interezza [23]. Il loro valore sta diventando sempre più importante, sia dal punto di vista sociale che da quello economico. La quotazione in borsa di molte aziende si basa sul volume di dati posseduto e le società sono disposte a pagare ingenti quantità di denaro per averne anche solo una parte. *Amazon* ha fatto del suo database uno dei punti di forza dell'azienda. Al suo interno contiene informazioni sui gusti e sugli interessi personali di molti di noi, e grazie a questi dati le aziende possono ad esempio gestire le proprie campagne di marketing.

Quello che emerge quindi, è che i dati a disposizione sono completamente inutili senza una profonda analisi e comprensione, ed è necessario avere gli strumenti adatti per elaborarli, analizzarli e trasformarli in informazioni col fine di generare conoscenza. Questo permetterà di comprendere a fondo dinamiche altrimenti non comprensibili e scoprirne di nuove.

## 2.2 Data Visualization

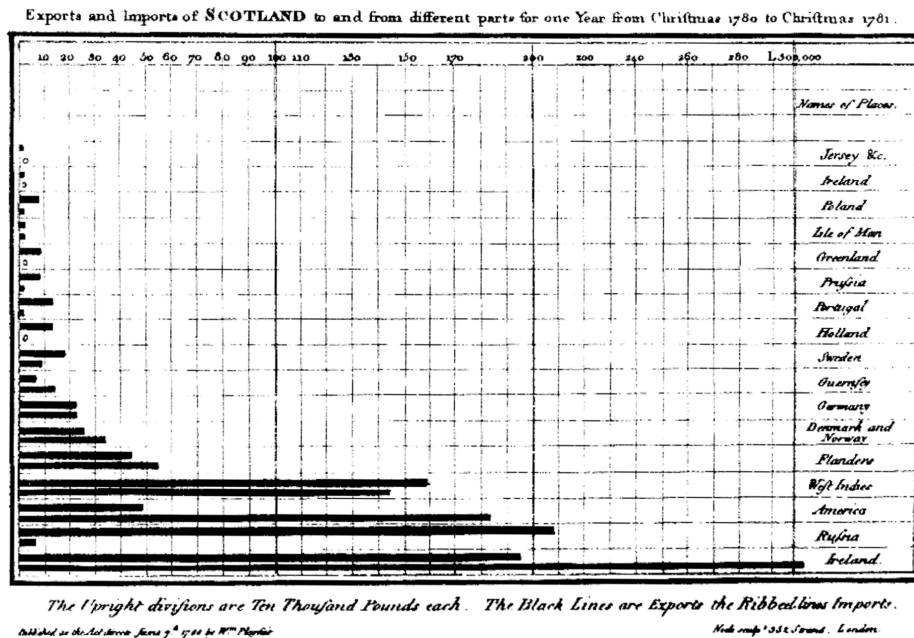


Figura 2.1: Bar Chart di William Playfair. Importazioni ed esportazioni della Scozia nel 1781 per 17 paesi

Circa 200 anni fa *William Playfair* (1786), uno statistico Scozzese, iniziò ad utilizzare le prime forme di rappresentazioni grafiche per analizzare i dati [9]. *Playfair*, cercando un metodo semplice per sostituire una tabella contenente numeri con una rappresentazione grafica, diede il via ad un ambito che ancora oggi necessita di numerose ricerche per beneficiare a pieno di tutte le sue potenzialità. Le moderne visualizzazioni offrono molto più che un metodo per sostituire una semplice tabella. Al loro meglio, i grafici sono strumenti che abilitano il ragionamento e l'esplorazione delle informazioni quantitative [37].

Nel libro *The Visual Display of Quantitative Information* (1983) di *Edward R. Tufte* vengono descritti gli errori più comuni e le procedure corrette da utilizzare durante la progettazione delle visualizzazioni. Anche se risale a più di 30 anni fa, questo libro è considerato ancora oggi una delle basi essenziali per lo studio della *Data Visualization*. Le teorie esposte fanno riferimento all'utilizzo dell'inchiostro. Oggi, a differenza di allora, l'inchiostro è stato sostituito con il pixel, ma le regole e le indicazioni fornite da *Tufte* risultano essere ancora valide.

### 2.2.1 Eccellenza nei grafici

L'eccellenza nella progettazione dei grafici consiste nel comunicare in maniera chiara, precisa ed efficiente idee complesse, fornendo all'osservatore il maggior numero di informazioni nel minor tempo possibile e con il minor utilizzo di "inchiostro" [37]. I grafici devono:

- Mostrare i dati.
- Indurre l'osservatore a riflettere sul significato dei dati, piuttosto che sul design del grafico, sulla tecnologia usata per la produzione della visualizzazione o su altro.
- Evitare di distorcere quello che i dati hanno da dire.
- Presentare molti numeri in uno spazio ridotto.
- Rendere grandi dataset coerenti.
- Favorire agli occhi dell'osservatore il confronto di differenti dati.
- Rivelare i dati con differenti livelli di dettaglio.
- Avere uno scopo ragionevolmente chiaro: descrizione, esplorazione, tabulazione o decorazione
- Essere strettamente integrato con la descrizione statistica e verbale di un insieme di dati.
- Rivelare i dati.

### 2.2.2 Integrità dei grafici

La prima cosa che viene in mente a molte persone quando pensano a grafici statistici è "falsità". Senza dubbio molti grafici mostrano distorsioni significative dei dati, rendendo difficile all'osservatore il compito di estrapolare le corrette informazioni. I grafici non sono molto differenti dalle parole in questo senso, e utilizzati in modi specifici possono essere sfruttati per ingannare le persone.

Un grafico non produce distorsioni se la rappresentazione dei dati risulta consistente con la rappresentazione numerica. La rappresentazione numerica è la misurazione fisica della superficie del grafico rappresentante la quantità numerica rilevata dai dati. Quando si presentano distorsioni grafiche, l'unico modo per fornire all'osservatore informazioni corrette e ridurre le ambiguità è l'utilizzo delle etichette. Per evitare la presenza di queste distorsioni e

mantenere integra una visualizzazione, può risultare utile seguire i 5 seguenti principi:

- La rappresentazione dei numeri, in quanto misurazione fisica della superficie del grafico, deve essere direttamente proporzionale alla quantità numerica rappresentata.
- Un'etichettatura chiara, dettagliata e approfondita dovrebbe essere utilizzata solo per risolvere le ambiguità e le distorsioni grafiche.
- Mostra le variazioni dei numeri, non la variazione del design.
- Il numero di variabili raffigurate non deve superare il numero di dimensioni dei dati.
- Le visualizzazioni non devono citare i dati fuori dal contesto.

Nonostante questi principi risalgono a più di 30 anni fa, le condizioni sotto la quale molte visualizzazioni ancora oggi sono prodotte, presentano una grande mancanza di competenze negli illustratori che si occupano della progettazione di queste. Queste condizioni portano alla creazione di grafici che non rappresentano la realtà, che utilizzano un design inadatto e che mancano delle informazioni rilevanti. Per ridurre la produzione di visualizzazioni mediocri sono richieste competenze non solo artistiche ma anche statistiche. Oggi, molti lavori che riguardano la progettazione delle visualizzazioni, come ad esempio nelle pubblicazioni giornalistiche, sono sotto la direzione di una figura artistica. Permettere a illustratori con sole competenze artistiche di controllare il design e il contenuto di grafici statistici è come permettere ad un tipografo di controllare lo stile, il contenuto e la modifica della prosa. E' quindi necessario che un maggior numero di esperti con conoscenze trasversali partecipi alla progettazione dei grafici [37].

### 2.2.3 Utilizzo degli spazi e dei colori

La gran parte dei pixel utilizzati per una visualizzazione deve rappresentare le informazioni contenute nei dati, e variare quando questi variano. Chiamiamo *Data-Pixel* (in analogia con il "*Data-Ink*" definito da *Tufte*) i pixel non cancellabili, che rappresentano il cuore di un grafico, i pixel non ridondanti, corrispondenti alla variazione dei numeri che rappresentano. Nessuno di questi pixel può essere cancellato se non a patto di perdere informazioni. La presenza di ogni pixel in un grafico richiede una ragione, e la ragione dovrebbe essere quella di rappresentare informazioni.

Oltre ai *Data-Pixel* esistono anche i *non-Data-Pixel*. I *non-Data-Pixel* corrispondono ai pixel non strettamente legati al dato quantitativo, come ad esempio le griglie. A differenza dei precedenti, è bene limitare questa tipologia e cancellarli ove possibile entro limiti ragionevoli. Pixel che falliscono nell'obiettivo di mostrare dati o informazioni utili alla loro lettura, spesso finiscono per ingombrare spazio in maniera non giustificata. Altre tipologie di pixel che ricadono sotto questa categoria sono i *Data-Pixel ridondanti* come ad esempio le etichette in cima ad un Bar Chart o le ombreggiature [37].

Applicare questi semplici principi produce sostanziali cambiamenti nel risultato finale della progettazione di un grafico, dando luogo a visualizzazioni più chiare ed efficaci.

Strettamente legati ai principi sui pixel abbiamo i principi relativi ai colori.

L'utilizzo dei colori riguarda ogni aspetto della *Data Visualization*. C'è stata più ricerca sull'utilizzo del colore che in qualsiasi altra questione percettiva. Grazie al loro utilizzo è possibile mostrare variazioni nei dati, evidenziare informazioni rilevanti e distinguere differenti dimensioni. La teoria che sta dietro allo studio sull'utilizzo del colore richiederebbe più di un capitolo [40]. Riportiamo quindi alcune linee guida rilevanti:

- Utilizzare colori più saturi quando si vuole codificare piccoli simboli, linee sottili o altri settori di piccole dimensioni. Utilizzare colori meno saturi per la codifica di aree di grandi dimensioni.
- Garantire un adeguato contrasto di luminanza al fine di definire e risaltare le caratteristiche importanti.
- Se grandi superfici sono state definite utilizzando colori luminosi quasi equi, considerare l'utilizzo di linee di confine sottili con grandi differenze di luminanza (dai colori delle aree) per aiutare a definire le forme.
- Se si utilizza la saturazione dei colori per codificare quantità numeriche, utilizzare una maggiore saturazione per rappresentare le quantità più grandi. Evitare l'uso di una sequenza di saturazione per codificare più di tre valori.

La Data Visualization trova quindi le sue radici nei primi statisti, i quali erano alla ricerca di metodi per analizzare i dati, che fossero più efficaci delle rappresentazioni tabellari. L'evoluzione e l'utilizzo di questi metodi, grazie anche all'aiuto degli illustratori, è diventato sempre più massiccio, e con l'avvento dei computer e dei display ad alta definizione, le possibilità offerte da questo ambito sono sempre maggiori.

## 2.3 La percezione umana

Alla base di una visualizzazione abbiamo simboli, proprietà grafiche ed elementi che richiedono il controllo dei processi umani come ad esempio un testo [22]. I processi visivi umani lavorano sostanzialmente su due livelli: *automatici* e *controllati* [30]. I processi automatici, i quali lavorano sulle proprietà visive come la posizione e i colori, sono fortemente parallelizzabili ma hanno potenza limitata; i processi controllati invece, lavorano per esempio con i testi, permettendo potenti operazioni, ma hanno capacità limitata.

Le proprietà grafiche invece, fanno riferimento ad un set limitato, set identificato da Bertin per primo, migliorato da Mackinlay poi [7, 22]. Una visualizzazione elementare è composta quindi dalla seguenti proprietà: un insieme di forme (le quali possono essere punti, linee, aree, superfici e volumi), posizione nello spazio e nel tempo, un set di proprietà visive collegate alle codifiche della retina (come il colore e la dimensione), proprietà di connessione e proprietà di inclusione.

In definitiva una visualizzazione è formata da [7]:

- Forme (punti, linee, area, superficie, volume)
- Elaborazione automatica delle proprietà grafiche:
  - Posizione: (X,Y,Z,T)
  - Codifiche della retina
  - Connessioni
  - Inclusioni
- Elaborazione controllata delle proprietà grafiche (colori, dimensioni, forme, orientamento, texture)

Sulla base delle proprietà appena descritte, le domande cruciali a cui rispondere sono: come si individua il modo migliore per trasformare i dati in qualcosa che la gente possa comprendere? e come si conduce al processo decisionale ottimale?

I diagrammi sono costituiti da simboli, e i simboli sono basati su interazioni sociali. Il significato di un simbolo è generalmente creato per convenzione, stabilita nel corso di una comunicazione da persona a persona. Lo studio dei simboli e di come questi si trasmettono cade sotto il nome di *semiotica*. Il significato è stato creato dalla società. Possiamo interpretare un'altra cultura solo nel contesto della nostra e usando gli strumenti della lingua che conosciamo. Le lingue sono mezzi convenzionali di comunicazione in cui i significati dei simboli sono stati stabiliti in maniera indipendente

l'una dall'altra. Il punto è che nessuna rappresentazione è “migliore” di un'altra. Tutte le rappresentazioni hanno un valore. Questo punto di vista è molto criticato e sostituito dall'idea che è possibile avere una nuova semiotica basata non su affermazioni filosofiche che sostengono l'utilizzo arbitrario dei simboli, ma su un'evidenza scientifica [40].

*“le persone possono interpretare i simboli senza la necessità di un training ”*

La parola *sensoriale* è utilizzata per riferirsi ai simboli e agli aspetti delle visualizzazioni che traggono la loro forza espressiva dalla loro capacità di utilizzare la potenza di elaborazione percettiva del cervello, senza la necessità di avere conoscenze a priori. La parola *arbitraria* viene utilizzata invece per definire gli aspetti delle rappresentazioni che devono essere appresi al fine di estrapolare informazioni da essa. La maggior parte delle visualizzazioni sono ibride e posseggono proprietà sia sensoriali che arbitrarie [40]. Tali proprietà sono riassumibili nei seguenti punti:

- Proprietà sensoriali delle rappresentazioni.
  - Permettono di essere comprese senza conoscenze a priori.
  - Resistenti a denotazioni alternative.
  - Immediatezza sensoriale.
  - Validità cross-culturale.
  
- Proprietà arbitrarie delle rappresentazioni.
  - Difficili da leggere.
  - Facili da dimenticare.
  - Incorporate nella cultura e architettura.

Se adesso andiamo a considerare l'interazione tra l'uomo e le visualizzazioni, e consideriamo questa come un sistema, è possibile identificare due componenti critiche: il sistema visivo umano, cioè un pattern finder flessibile accoppiato ad un meccanismo decisionale, e la vasta quantità di informazioni a cui si può accedere attraverso un computer. La visualizzazione risulta essere l'interfaccia tra le due parti. Migliorare questa interfaccia può migliorare in maniera sostanziale le prestazioni dell'intero sistema. Fino a poco tempo fa, il termine visualizzazione significava costruire di un'immagine visiva nella mente [38]. Oggi invece è qualcosa di più simile ad una rappresentazione grafica di dati o concetti. Così, dall'essere una costruzione interna della mente, una visualizzazione è diventata un manufatto esterno che supporta il processo decisionale [40].

## 2.4 Interagire coi dati

L'esplorazione di una grande mole di informazioni è ancora oggi una sfida molto complicata, anche se problematiche come le architetture parallele, le connessioni di rete a banda larga, i dischi di grandi dimensioni ad alta velocità e i moderni sistemi di gestione dei database hanno fatto notevoli passi in avanti. Questi progressi hanno lasciato molti utenti con la sensazione che di non poter analizzare questi grandi flussi di informazioni. La chiave dietro a questo problema risiede nel comprendere la grande capacità di elaborazione delle informazioni che l'uomo possiede. Presentando in modo visivo le informazioni e consentendo all'utente di avere un controllo dinamico, è possibile esplorare ampi spazi di informazione e facilitare la comprensione di questi [28, 31]. In pochi decimi di secondo, l'uomo può ricordare immagini e identificare anomalie [3].

Un concetto fondamentale quindi è quello di supportare la navigazione, la quale si distingue dalla composizione di query e dal recupero delle informazioni a causa della sua enfasi sulla possibilità di filtrare rapidamente i dati per ridurre set di risultati, affinare in maniera progressiva i parametri di ricerca, riformulare in maniera continua gli obiettivi e permettere la scansione in maniera visiva dei dati per identificare i risultati.

Psicologi della percezione, statisti e designer come *Bertin* [4], *Cleveland* [8] e *Tufte* [37] hanno contribuito e offerto delle preziose indicazioni su come presentare in maniera grafica le informazioni, ma con le potenzialità introdotte oggi dai computer, viene data la possibilità ai progettisti di *user interface* di analizzare e scoprire nuove dinamiche relative alle visualizzazioni. Il successo relativo alla manipolazione delle interfacce utente è un indicatore di quanto sia potente l'utilizzo del computer per mostrare le informazioni in maniera grafica. Spesso si dice che *un'immagine vale più di mille parole*, e per molte operazioni (ma non tutte) è chiaro che una rappresentazione visiva risulta notevolmente più facile da analizzare rispetto ad un testo o un report. Con l'aumento della potenza di calcolo dei computer e della risoluzione degli schermi, la *Data Visualization* e le interfacce utente hanno la possibilità di ritagliarsi un ruolo importante. Immaginiamo di avere su uno schermo la mappa degli Stati Uniti. Attraverso l'utilizzo del computer è possibile selezionare una qualsiasi città direttamente dalla mappa e ottenere per esempio le informazioni turistiche che la riguardano. Certamente una persona che conosce il nome ma non la sua posizione preferirebbe avere un elenco in ordine alfabetico tra cui scegliere. Tale rappresentazione risulta quindi utile per fornire un contesto e favorire l'orientamento, permettendo la selezione di regioni e fornire feedback dinamici sui cambiamenti, come ad



esempio il meteo. Le visualizzazioni scientifiche hanno la potenza di rendere gli atomi, il cosmo, e i fenomeni tridimensionali visibili e comprensibili. Le visualizzazioni di informazioni astratte invece permettono di rilevare pattern, cluster e outliers nell'ambito della statistica. Le potenzialità di questo strumento sono applicabili in svariati ambiti, e soprattutto hanno tutte le caratteristiche per essere fortemente accessibili. L'uomo possiede abilità percettive che ancora oggi sono sottoutilizzate dalla visualizzazioni proposte. Un utente è in grado di scansionare, riconoscere e ricordare immagini molto rapidamente, può identificare cambiamenti nelle dimensioni, nei colori, nelle forme e nei movimenti. Può individuare un singolo pixel, anche all'interno di display megapixel, ed è in grado effettuare delle azioni come ad esempio trascinare un oggetto sopra un altro. Fino a non molto tempo fa, le interfacce utente sono state orientate verso rappresentazioni testuali, quindi le rappresentazioni visive hanno ancora molte opportunità da offrire. Ci sono molte linee guida da seguire quando si progetta un interfaccia utente, come ad esempio le *5S* proposte da *Garret* [12], ma i principi base legati alla progettazione di una visualizzazione possono essere riassunti in quello che *Shneiderman* ha definito Visual Information Seeking Mantra [32]:

*“Overview first, zoom and filter, then details-on-demand”*

Per creare una visualizzazione interattiva, Shneiderman ha definito 7 linee guida principali:

1. Overview: dare una panoramica della visualizzazione.
2. Zoom: focalizzarsi sugli oggetti di interesse.
3. Filter: filtrare gli oggetti non di interesse.
4. Details-on-demand: selezionare un oggetto o un gruppo e fornire i dettagli quando necessario.
5. Relate: mostrare le relazioni tra gli oggetti.
6. History: mantenere una cronologia delle azioni per permettere di tornare indietro e ripetere le azioni passate.
7. Extract: consentire l'estrazione di sotto-collezioni di dati sotto specifici parametri di ricerca.

Un altro trend che sta prendendo piede, non riguarda solo la possibilità di interagire con le visualizzazioni, ma anche l'utilizzo di animazioni che riguardano il movimento di oggetti all'interno di queste. Tali animazioni

hanno l'obiettivo di mostrare le relazioni tra i dati e vedere come queste variano nel tempo. Per anni sono stati utilizzati grafici a linea. Queste rappresentazioni funzionano perfettamente quando ci si focalizza su dati strettamente orientati al tempo, come ad esempio, la misura del traffico web valutata in intervalli temporali uguali. Ma risulta ancora efficace quando vogliamo analizzare differenti relazioni tra più valori e vedere come queste variano ad esempio in un anno? Si prenda la relazione tra le spese del marketing e il fatturato. Il miglior modo per analizzare questa relazione in un determinato istante di tempo è utilizzare lo Scatter Plot, con le spese di marketing misurate lungo l'asse X e il fatturato lungo l'asse Y. Se vogliamo ora vedere come la correlazione di queste due variabili sia cambiata nel tempo, non è possibile farlo tramite l'utilizzo di una linea. Per sopperire a questo problema possiamo animare tale visualizzazione, permettendo che i dati rappresentati si muovano all'interno del grafico, rivelando così la loro variazione e il loro legame nel tempo. Alcuni dei migliori esempi relativi all'uso delle animazioni sono stati sviluppati da [www.GapMinder.org](http://www.GapMinder.org), i quali fanno uso di questa tecnica per mostrare dati importanti del mondo, come il rapporto tra il reddito dei paesi e la mortalità infantile, e di come il mondo sia cambiato in questo senso negli ultimi 30 anni [10].

## 2.5 Lavori correlati

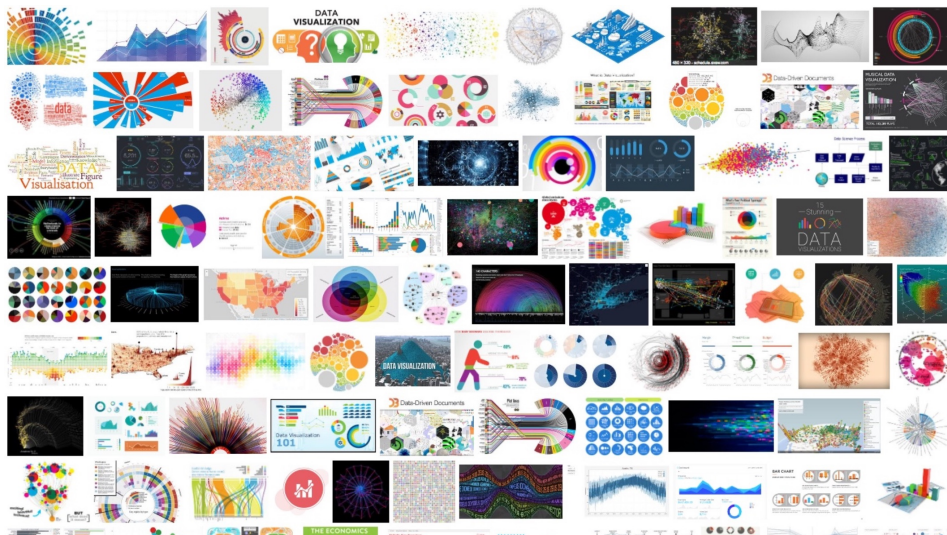


Figura 2.2: Varietà delle tecniche di Data Visualization

Uno dei maggiori problemi relativi l'esplorazione e l'analisi dei dati si presenta quando siamo di fronte ad una grande mole di informazioni, le quali possono risultare difficili da visualizzare in modo efficiente, causando così confusione durante l'analisi e nascondendo ciò che è realmente importante dietro ad un muro di disordine. Questo è un problema che sta riscuotendo molta attenzione da parte della comunità accademica. Molte tecniche di rappresentazione dati sono ancora oggi in fase di studio e sofisticati algoritmi sono stati sviluppati per domare la grande mole di dati, riducendone la quantità in modo da evitare la perdita di significato, o riducendo l'ingombro visivo della visualizzazione stessa attraverso nuovi approcci per il posizionamento degli oggetti, per un utilizzo migliore del colore o tramite altri attributi visivi come la trasparenza [10]. Tutti questi fattori sono tenuti in considerazione durante la fase di selezione di una tecnica di visualizzazione dati. Non stupisce quindi che tale selezione non sia banale, considerando anche l'ampio catalogo oggi disponibile tra cui scegliere. Di conseguenza, avere uno strumento che aiuti a identificare le metodologie più adatte a mostrare le informazioni è sempre più importante. Tale strumento è possibile trovarlo nelle tassonomie e negli studi relativi alla selezione automatica dei grafici, grazie al quale è possibile individuare metodologie per associare una determinata tipologia di dato alla tecnica di visualizzazione più appropriata. In letteratura sono state presentate diverse tassonomie e metodologie per identificare le visualizzazioni ottimali a seconda del contesto, ma le più apprezzate e le più rilevanti sono risultate essere quelle di:

- [41] *Stephen Wehrend e Clayton Lewis*
- [36] *Melanie Tory e Torsten Möller*
- [32] *Ben Shneiderman*
- [22] *Jock Mackinlay*
- [29] *Philip K. Robertson*

### 2.5.1 Tassonomie

L'obiettivo di *Stephen Wehrend e Clayton Lewis* era quello di sviluppare un "catalogo" per gli utenti, tramite il quale fosse possibile cercare e condividere le tecniche di visualizzazione sulla base di specifiche tipologie di dato e di compiti. Tale catalogo doveva risolvere in maniera rapida il problema legato all'identificazione delle tecniche di visualizzazione. La loro proposta dunque si è basata su una classificazione delle tecniche di visualizzazione

indipendente dal dominio. Per identificare una visualizzazione, l'utente deve suddividere il proprio problema in sottoproblemi. Questi sottoproblemi devono poi essere descritti in termini di oggetti da rappresentare e di operazioni che la tecnica di visualizzazione deve supportare. Sulla base di queste definizioni vengono quindi identificate le visualizzazioni più adatte all'interno del catalogo, per poi combinarle assieme e ottenerne una in grado di rispondere al problema originario [41].

La tassonomia proposta *Melanie Tory e Torsten Möller* invece si basa sulla classificazione degli algoritmi di visualizzazione piuttosto che sui dati. Gli algoritmi sono suddivisi in categorie sulla base delle ipotesi che questi fanno riguardo ai dati che vengono visualizzati; queste serie di assunzioni vengono chiamate *design models*. Poiché questa tassonomia si basa sul design models, secondo le autrici risulta essere più flessibile rispetto alle tassonomie esistenti e inoltre considera il modello concettuale dell'utente, sottolineando l'aspetto umano della visualizzazione. I design models sono classificati in base al fatto che i dati siano di discreti o continui [36].

### 2.5.2 Altri lavori correlati

Una dei metodo più apprezzati per identificare una visualizzazione a seconda del contesto in cui ci troviamo è stata presentata da *Ben Shneiderman*. *Shneiderman* definisce delle direttive da dover seguire mentre si progetta, o si sceglie, il metodo di rappresentazione dati. Lui stesso definisce queste direttive come Visual Information Seeking Mantra: "Overview first, zoom and filter, then details-on-demand". L'obiettivo di questo lavoro è quello di dare un punto di partenza per comprendere il ricco e variegato insieme delle tecniche di *Data Visualization* che sono state proposte negli ultimi anni.

Nel suo documento, Shneiderman, descrive 7 attività ad alto livello (overview, zoom, filter, details-on-demand, relate, history, and extract), dipendenti dal dominio, applicate a 7 tipologie di dati differenti:

- Dati in 1 dimensione
- Dati in 2 dimensioni
- Dati in 3 dimensioni
- Dati multidimensionali
- Dati temporali
- Dati strutturati a rete (Network Data)

- Dati strutturati ad albero (Tree Data)

Applicando le attività proposte alla tipologia di dato che vogliamo mostrare, è possibile quindi identificare molte criticità e definire requisiti funzionali ad alto livello che la visualizzazione scelta dovrà soddisfare [32].

A differenza dei lavori correlati descritti fino ad ora, gli studi di *Jock Mackinlay* e *Philip K. Robertson* si differenziano rispetto ai precedenti. L'obiettivo di *Mackinlay* era quello di gettare le basi per un sistema in grado di scegliere le tecniche "ottimali" di visualizzazione. Due problemi si sono sollevati da questo obiettivo: la codifica dei criteri di progettazione grafica in una forma utilizzabile dallo strumento di presentazione, e la generazione di una grande varietà di grafici in modo che lo strumento di presentazione sia in grado di ospitare una vasta gamma di informazioni. L'approccio descritto nel suo articolo si basa sulla considerazione che le tecniche di rappresentazione grafica sono paragonabili a frasi del linguaggio grafico. I problemi di progettazione grafica sono codificati seguendo i criteri di espressività ed efficacia per i linguaggi grafici. I criteri di espressività determinano se un linguaggio grafico può esprimere le informazioni desiderate. I criteri di efficacia invece determinano se un linguaggio grafico sfrutta sia le capacità in output dello strumento che il sistema visivo umano. Una grande varietà di visualizzazioni può essere generata in modo sistematico utilizzando un'algebra in grado di comporre un piccolo insieme di linguaggi grafici primitivi. Tecniche di intelligenza artificiale sono state utilizzate per implementare un piccolo prototipo di presentazione chiamato APT, che si basa sulla composition algebra e sui criteri di progettazione grafica [22].

*Robertson* ha approfondito un lavoro molto simile a quello di *Mackinlay*, aggiungendo però la possibilità di comprendere il contesto di utilizzo e permettere all'utente di scegliere (e modificare) più tecniche di visualizzazione ottimali. Nel suo articolo infatti, formalizza il processo per la creazione di un framework generico, in grado di incrociare la tipologia di informazione a cui siamo interessati con le abilità e le proprietà delle differenti tecniche di visualizzazione, col fine di selezionare la tecnica più adatta [29].



## Capitolo 3

# Definizione del Problema e Metodologia Proposta

*“The speed at which you can combine all your data, the speed at which you process and visualize it, and most importantly the speed at which you reach actionable insights. The faster you can juggle information, the more data and data combinations you can afford to cover, the more unexpected the insights you will find.”*

I. EREZ

Come anticipato nel Capitolo 1, alla base della realizzazione di questa tesi si trova la volontà di rispondere alle seguenti domande: *“le tassonomie e i sistemi automatici di individuazioni dei grafici oggi esistenti, sono ancora strumenti adatti per identificare quale sia la tecnica migliore da utilizzare?”* e *“siamo sicuri che le tecniche tradizionali, utilizzate fino ad oggi, siano ancora le più appropriate?”*.

Per rispondere a queste domande abbiamo deciso di proporre un metodologia user-oriented efficace per identificare quelle che sono le visualizzazioni più indicate a mostrare una determinata tipologia di dato, verificando se, i grafici tradizionali, ovvero quelli storicamente usati, sono ancora oggi i più adatti, oppure se le visualizzazioni innovative, introdotte recentemente, possono essere delle ottime candidate per sostituire le prime.

### 3.1 Il problema

Lo studio di nuove tecniche di visualizzazione ha effetto in diversi ambiti, in quanto queste sono collegate alla nostra innata abilità di identificare pattern in maniera rapida, e di convertire questi in informazioni utili per effettuare analisi [34]. Questo implica che l'utilizzo di punti, linee, numeri, simboli e colori in un determinato modo piuttosto che in un altro ha un forte impatto sulla percezione delle informazioni ed è di primaria importanza farne un uso corretto.

La scienza cognitiva però ha problemi fondamentali che sono ancora oggi da risolvere. Spesso il miglioramento dei sistemi cognitivi comporta un'ottimizzazione nella ricerca dei dati e rende più facile identificare modelli e pattern importanti. Come già descritto nel Capitolo 2, un individuo che lavora con strumenti di *Data Visualization* genera un sistema cognitivo in cui i componenti critici sono, da un lato, il sistema visivo umano, cioè un pattern finder flessibile accoppiato ad un meccanismo decisionale, e, dall'altro lato, la potenza computazionale e la vasta quantità di informazioni di un computer accoppiato al World Wide Web. La visualizzazione è l'interfaccia tra le due parti. Migliorare questa interfaccia può migliorare in maniera sostanziale le prestazioni dell'intero sistema. Fino a poco tempo fa, il termine visualizzazione significava costruire di un'immagine visiva nella mente [38]. Oggi invece è qualcosa di più simile ad una rappresentazione grafica di dati o concetti. Così, dall'essere una costruzione interna della mente, una visualizzazione è diventata un manufatto esterno che supporta il processo decisionale [40].

La quantità e l'eterogeneità dei dati a disposizione è in forte crescita. In relazione a questa grande disponibilità di dati e informazioni, diventa sempre più importante la fase di sintetizzazione e soprattutto di comunicazione dei dati, in modo da poterli utilizzare all'interno dei processi decisionali [34].

Dal punto di vista tecnologico questo si rispecchia nella progettazione di infrastrutture e applicazioni capaci di supportare questa mole di dati e di trasformare questi in visualizzazioni in grado di fornire in maniera rapida ed efficiente informazioni utili e corrette. Strumenti come *Tableau* [33], *QlikView* [20] e *Power BI* [24], per citarne alcuni, offrono tali possibilità, e librerie *JavaScript* come *D3.js* [6] aprono la le porte a visualizzazioni limitate soltanto dalla nostra creatività. Il problema però è che molto spesso, soprattutto nel settore del *Business*, c'è una forte carenza di conoscenza riguardo al mondo della *Data Visualization* e non è raro trovarsi di fronte ad un utilizzo scorretto di determinate visualizzazioni, le quali producono inevitabilmente un effetto opposto rispetto a quello desiderato: confusione



invece che comprensione.

Riuscire quindi a identificare quali tra le nuove visualizzazioni recentemente introdotte siano maggiormente indicate in un determinato contesto, e quali di queste siano più adatte a sostituire le visualizzazioni tradizionali, può fornire un grande aiuto nel mondo sia della *Business Intelligence* che della ricerca, migliorando l'intero sistema cognitivo e raggiungendo lo scopo per cui la Data Visualization è nata:

*"The objective is to make it easier to access more information and to assimilate and interact with it faster to gain understanding and then make decisions [10]."*

## 3.2 Metodologia proposta

### 3.2.1 Definizione del caso studio

La metodologia proposta ha inizio con la definizione di un caso studio, quindi con l'identificazione dell'informazione che vogliamo ottenere a partire dai dati a disposizione. Come introdotto nel Capitolo 1 è necessario studiare il dato per capirne le sue caratteristiche ed individuare così la quantità e la tipologia di informazioni che è in grado di fornire. L'obiettivo finale sarà quello di estrapolare da queste, la maggior quantità di conoscenza. Questa fase inciderà su tutte le scelte che verranno effettuate nelle successive fasi.

### 3.2.2 Individuazione dei grafici

Definito il caso studio è necessario individuare i grafici più adatti a mostrare la tipologia di dato preso in esame, affinché questi siano in grado di fornire e trasmettere l'informazione voluta.

Oggi per identificare le soluzioni grafiche adatte a mostrare una determinata tipologia di dato, ci si può basare sulle tassonomie esistenti, come quelle presentate nel Capitolo 2. Questo lavoro però può risultare molto lungo e dispendioso. I tool di *Data Visualization* oggi in commercio, la maggior parte dei quali contiene sistemi di identificazione automatica dei grafici, possono fornire un buon punto di partenza alternativo, ma offrono un numero limitato di visualizzazioni, e spesso propongono solo quelle tradizionali. Questo, però, va nella direzione opposta rispetto all'obiettivo della nostra ricerca. Uno strumento di partenza che abbiamo trovato molto utile e intuitivo è possibile trovarlo al link: <http://www.datavizcatalogue.com>.

Originariamente nato come strumento di riferimento per *Severino Ribeca*, creatore di questo lavoro e Graphic Designer all'Academy Class, tale

catalogo fornisce un numero elevato di tecniche di visualizzazione, ed è in costante aggiornamento, diventando un ottimo punto di partenza per identificare quali visualizzazioni utilizzare in svariate situazioni.

La suddivisione proposta in tale catalogo è rispetto alla funzione dei grafici. Per funzione si intende la tipologia di informazione che è in grado di trasmettere la visualizzazione. Per esempio alcune delle funzioni che possiamo trovare sono *Confronti*, nella quale ricadono i metodi di visualizzazione che aiutano a mostrare le differenze o le similitudini tra i valori, o le *Relazioni*, dove è possibile trovare quelle tecniche che aiutano a identificare le relazioni tra i dati.

Una volta estratte le rappresentazioni più indicate al nostro caso studio, è necessario uno sforzo progettuale, identificando quelle visualizzazioni che generalmente non vengono ritenute adatte rispetto al caso studio in esame, ma che grazie a modifiche mirate possono essere in grado di comunicare l'informazione voluta.

### 3.2.3 Comparazione dei grafici

Una volta selezionati i grafici è necessario confrontarli. L'approccio proposto consiste nella stesura di un'indagine la quale è suddivisa in 3 fasi differenti.

Lo strumento base utilizzato per la maggior parte delle indagini è il questionario. Questo strumento fornisce l'opportunità di ottenere una classificazione omogenea dei dati raccolti, in quanto il modello è uniforme per tutti gli intervistati e consente a tutti gli intervistatori di porre sempre le stesse domande, nella medesima sequenza.

Importante durante la stesura dei questionari sono i collaudi (o fase pilota). Questi collaudi sono da svolgere eventualmente a più riprese, intervallati da opportune modifiche della stesura originaria. In questo modo è più facile che il questionario possa rimanere inalterato nel corso di eventuali ripetizioni dell'indagine, in quanto buona parte dei problemi è stata eliminata in seguito ai collaudi della fase pilota. Bisogna però porre molta attenzione quando si modifica il questionario da un'indagine all'altra, poiché si rischia di perdere parte della confrontabilità.

Le domande formulate possono essere suddivise nelle seguenti categorie: domande dirette, domande filtro e domande di controllo. Queste tre categorie saranno poi suddivise in domande aperte e a risposta multipla.

Andiamo ora ad analizzare le tre fasi in cui è suddivisa l'indagine:

### **Comparazione tra visualizzazioni tradizionali**

Nella prima fase l'obiettivo è confrontare tra loro tutte le visualizzazioni tradizionali, generalmente utilizzate per visualizzare il dato del caso studio selezionato. Questo confronto deve avvenire valutando sia l'appeal che queste visualizzazioni hanno sull'utente, sia l'efficacia con la quale le informazioni vengono trasmesse, ovvero come l'utente recepisce le informazioni, con che rapidità, e soprattutto se le recepisce in maniera corretta. Per effettuare questo confronto è bene mostrare le visualizzazioni in modo da farne risaltare i pregi, confrontandole così al massimo della loro efficacia.

Altra considerazione importante da fare riguarda la struttura delle domande. E' importante che queste non diano troppi dettagli riguardo all'informazione che stiamo mostrando. Questo serve a valutare come in assenza di un contesto, l'utente percepisce l'informazioni trasmesse, e quindi quanto una visualizzazione sia auto-descrittiva.

### **Comparazione con visualizzazioni innovative**

Nella seconda fase l'obiettivo è confrontare le visualizzazioni che sono risultate essere le più efficaci nel precedente step con quelle innovative, che generalmente non vengono utilizzate. L'obiettivo di questa fase è lo stesso della precedente, quello che cambia sono le tecniche di *Data Visualization* messe a confronto. Valgono quindi ancora le considerazioni fatte sulla modalità con cui vengono mostrati i grafici e sulle domande da effettuare agli intervistati.

### **Comparazione tra le visualizzazioni migliori**

Nella terza ed ultima fase invece, l'obiettivo è quello di mettere a confronto le due tecniche di *Data Visualization* che sono risultate essere le migliori dalla precedente indagine, in termini di efficacia e correttezza di lettura. In questa fase, al contrario delle precedenti, è importante attingere a dati reali e permette all'utente di interagire con le visualizzazioni. In questo modo è possibile non solo confrontare i pregi delle due tecniche, come nelle precedenti indagini, ma anche i limiti e i difetti, valutando così la risposta dell'utente in più situazioni e quindi l'efficacia di una visualizzazione nel suo complesso.

### **Correlazione con il background degli utenti**

Sfruttando questo approccio, abbiamo pensato di condurre uno studio secondario. Tale studio ha l'obiettivo di valutare se, il livello di conoscenza

dell'intervistato in tale ambito ed il background che egli possiede, possa influire sull'efficacia che un determinato grafico ha su di lui.

Se tale teoria fosse confermata, oltre ad identificare una metodologia per sostituire e verificare l'efficacia delle tecniche di visualizzazione tradizionali, potremmo gettare le basi per una sorta di *Recommender System*, ovvero un sistema capace di consigliare il grafico più adatto in base all'utente con cui si interfaccia.

## Capitolo 4

# Soluzione del Problema

*“Graphical elegance is often found in simplicity of design and complexity of data.”*

Edward R. TUFTE

In questo capitolo verrà descritto come abbiamo applicato la soluzione proposta per la risoluzione del problema affrontato nella nostra ricerca, descrivendo nel dettaglio il caso studio preso in esame e come questo è stato trattato durante le differenti fasi. Ci focalizzeremo sugli aspetti implementativi principali e sulle problematiche incontrate durante l'indagine.

L'esperienza di ricerca condotta, ha seguito fedelmente i passi descritti nel capitolo precedente: è stato quindi definito un caso studio, il quale ha guidato le fasi successive. In queste fasi siamo andati prima a individuare le visualizzazioni sia tradizionali che innovative tramite il quale mostrare i dati del nostro caso studio, e poi a comparare queste attraverso differenti indagini, con l'obiettivo finale di individuare la tecnica più efficace e più chiara.

## 4.1 Definizione del caso studio

Per verificare la fattibilità della soluzione proposta, abbiamo preso come caso studio i dati corrispondenti al numero di posts pubblicati sui tre principali Social Network: *Twitter*, *Instagram*, *Facebook*, nella quale sono stati menzionati i giudici di *MasterChef Italia*.

Prima di proseguire, introduciamo una rapida panoramica del contesto in cui ricadono le informazioni che andremo a mostrare.

### 4.1.1 MasterChef Italia



*Figura 4.1: Logo MasterChef Italia*

MasterChef Italia è la versione italiana del talent show culinario d'origine britannica MasterChef. La trasmissione vede susseguirsi una serie di prove in cui i concorrenti devono dimostrare le proprie abilità in cucina. In seguito al casting da parte della produzione, vi è la fase delle selezioni. Esse consistono in una serie di prove iniziali svolte da numerosi aspiranti cuochi.

Il programma si imposta sull'alternanza costante di due tipologie di episodi, differenziate dalle prove svolte in esse. Al termine di ciascun episodio si vede solitamente l'eliminazione di un concorrente, fino ad arrivare alla prova finale dove viene decretato il vincitore.

Chi eliminare ad ogni puntata è compito dei giudici che risultano essere il ristoratore Joe Bastianich e i cuochi Bruno Barbieri, Carlo Cracco e Antonino Cannavacciuolo [45].

### 4.1.2 MasterChef Italia e i Social

In questi anni MasterChef Italia ha riscosso molto successo. Si pensi che l'edizione del 2015/2016 ha avuto un numero di telespettatori pari a 1.191.000 con uno share del 4,16% [45]. Tutto questo successo si è anche riscontrato sui Social Networks. Sono molte oramai le pagine dedicate a questo show

Nome Chef | Ora e data ( in ms ) | # posts su Twitter | # posts su Instagram | # posts su Facebook

Tabella 4.1: Struttura dei dati

dove gli spettatori possono commentare e condividere i pensieri, oltre che sulle proprie pagine personali. Tutte queste attività producono una grande quantità di dati. Con gli strumenti adatti a monitorare e prelevare questi è possibile produrre analisi per ottenere molte informazioni e trasformarle in conoscenza, come ad esempio l'utilizzo di tecniche di *Sentiment Analysis*, ovvero tecniche che fanno uso dell'elaborazione del linguaggio naturale per identificare ed estrarre informazioni soggettive, per capire quale personaggio sta riscontrando un maggiore successo tra gli spettatori.

### 4.1.3 Struttura del dato

L'attività di estrapolazione dei dati dai Social Networks non è stata oggetto della nostra ricerca, in quanto non è stata ritenuta rilevante ai fini dello studio proposto da questa tesi, la quale si è focalizzata maggiormente sulla progettazione delle visualizzazioni e sul confronto di queste.

I dati sono stati estrapolati dalla *Fluxedo Srl*, azienda che grazie al progetto *SocialOmers* [35] si occupa dell'analisi dei dati provenienti dai social. Ogni record del dataset contiene le informazioni relative al nome dello Chef e il numero di posts in cui questo è stato menzionato suddiviso per Social Network (*Twitter*, *Instagram*, *Facebook*). Questi sono stati aggregati ogni mezzora per ogni singolo Chef e . Più nello specifico, i posts presi in esame sono relativi al giorno 3 Marzo 2016 e in una manipolazione successiva sono stati ulteriormente aggregati per ora, mantenendo la suddivisione per Social Network e per Chef. Il dato finale è stato quindi tabulato secondo la struttura mostrata nella Tabella 4.1.

Con questi dati a disposizione ci siamo imposti l'obiettivo di mostrare le informazioni riguardanti la popolarità dei giudici di *MasterChef Italia*. Con popolarità si intende il numero di posts dove è stato menzionato il singolo Chef. Più posts riguardo ad un Chef sono presenti, più questo è ritenuto popolare. Tali informazioni verranno mostrate non solo in un determinato istante, ma anche lungo l'arco temporale di un'intera giornata, in modo da analizzare come la popolarità viene distribuita nel tempo.

## 4.2 Identificazione dei grafici

Per identificare i grafici adatti a mostrare il dato del nostro caso studio, è stata effettuata una ampia ricerca preliminare delle metodologie esistenti oggi in letteratura.

Questo studio si è basato inizialmente sulla tassonomia online consultabile al link: <http://www.datavizcatalogue.com/>. Partendo da questo lavoro siamo andati a migliorare e catalogare a livello funzionale le visualizzazioni oggi presenti, tralasciando quelle che nascono dalla fusione di più tecniche.

Il lavoro, durato poco meno di 2 mesi, ha dato frutto ad una “piccola wikipedia delle tecniche di visualizzazione dati” consultabile al link: [http://fluxedomanager.cloudapp.net:8080/projects/visualization-thesis/wiki/Datavisualization\\_Catalogue](http://fluxedomanager.cloudapp.net:8080/projects/visualization-thesis/wiki/Datavisualization_Catalogue).

Di seguito viene riportato l’elenco delle 55 visualizzazioni censite:

*Arc Diagram, Area Graph, Bar Chart, Box and Whisker Plot, Bubble Chart, Bubble Map, Bullet Chart, Candlestick Chart, Chord Diagram, Choropleth Map, Circle Packing, Connection Map, Density Plot, Donut Chart, Dot Map, Dot Matrix Chart, Error Bars, Flow Chart, Flow Map, Gantt Chart, Heat Map, Histogram, Illustration Diagram, Kagi Chart, Line Graph, Marimekko Chart, Multi-set Bar Chart, Network Diagram, Nightingale Rose Chart, Non-ribbon Chord Diagram, Open-high-low-close Chart, Parallels Coordintaes Plot, Parallel Sets, Pictogram Chart, Pie Chart, Population Pyramid, Proportional Area Chart, Radar Chart, Radial Bar Chart, Radial Column Chart, Sankey Diagram, Scatterplot, Span Chart, Spiral Plot, Stacked Area Graph, Stacked Bar Graph, Stream Graph, Sunburst Diagram, Timeline, Tree Diagram, Treemap, Venn Diagram, Violin Plot, Vonroi Diagram, Word Cloud.*



Sulla base del censimento effettuato e delle informazioni raccolte, abbiamo selezionato le visualizzazioni più adatte a mostrare il dato e l'informazione relativa al nostro caso studio:

- Stacked Bar Graph (o Stacked Bar Chart)
- Histogram
- Donut Chart
- Pie Chart
- Radar Chart
- Radial Column Chart
- Parallel Sets
- Circle Packing
- Heat Map

Andiamo ad analizzarle nel dettaglio (le immagini e le informazioni seguenti provengono da <http://www.datavizcatalogue.com> [27]).

### 4.2.1 Stacked Bar Graph (o Stacked Bar Chart)

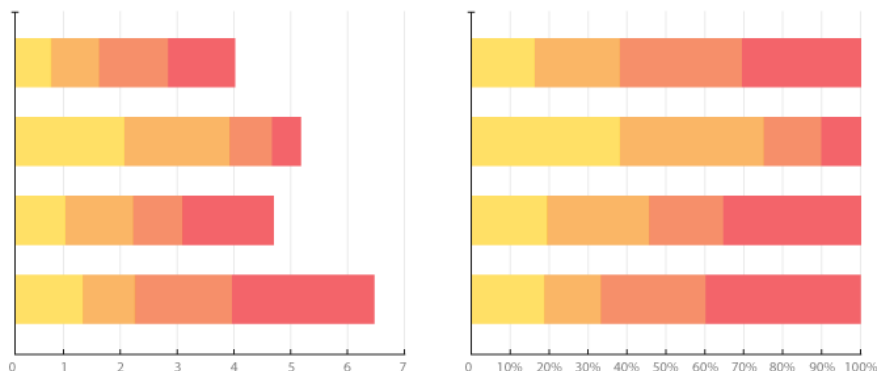


Figura 4.2: Stacked Bar Graph (o Stacked Bar Chart)

Lo **Stacked Bar Graphs** è un grafico utilizzato per mostrare i dati attraverso l'utilizzo di barre posizionate una affianco all'altra. Viene tipicamente impiegato per rappresentare come una categoria più grande è suddivisa in categorie più piccole e quale sia il rapporto che ciascuna sotto-categoria ha rispetto al totale. Esistono due tipi di Stacked Bar Graphs:

**Simple Stacked Bar Graphs.** In questo grafico le barre appartenenti alle sotto-categorie sono collocate una dopo l'altra, e la loro lunghezza corrisponde al valore della quantità che rappresentano. La somma di tutte le sotto-categorie corrisponde al valore totale della macro-categoria. Questa versione è ideale per confrontare le quantità sia tra le macro-categorie che tra le sotto-categorie.

**100% Stack Bar Graphs.** In questo grafico la barre corrispondenti alle macro-categorie hanno tutte la stessa lunghezza e rappresentano il 100% del loro valore. La lunghezza delle sotto-categorie invece è data dal valore percentuale che queste hanno rispetto al totale di riferimento. Questo rende più facile vedere le differenze relative tra le quantità dei diversi gruppi.

Uno dei maggiori difetti di questo metodo di visualizzazione dati è il fatto che diventa di difficile lettura in due situazioni: con l'aumentare del numero dei segmenti e nel caso in cui ci siano all'interno di una macro-categoria valori significativamente più piccoli rispetto ad altri. Inoltre il confronto tra ogni segmento non è immediato, in quanto questi non sono allineati rispetto ad una base comune.

**Funzioni:** Confronti - Relazioni - Patterns - Proporzioni - Parte del totale

## 4.2.2 Histogram

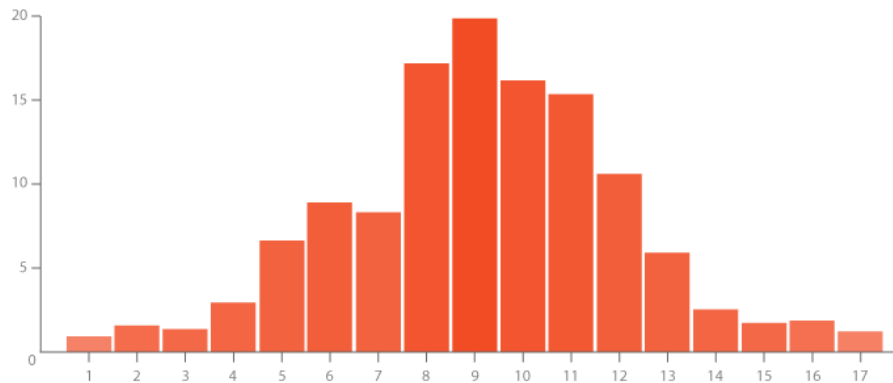


Figura 4.3: Histogram

Un **Histogram** ( conosciuto comunemente in italiano sotto il nome di istogramma ) permette di visualizzare la distribuzione dei dati su un intervallo continuo o su un determinato periodo di tempo. Ogni barra nell'istogramma rappresenta la frequenza tabellare in ogni intervallo. La superficie totale dell'istogramma equivale al numero di dati.

Gli istogrammi contribuiscono a dare una stima di dove sono concentrati i valori, quali sono gli estremi e se ci sono eventuali lacune o valori insoliti. Sono anche utili per fornire una visione approssimativa della distribuzione di probabilità.

**Funzioni:** Confronti - Dati nel tempo - Distribuzione - Patterns - Probabilità - Range

### 4.2.3 Donut Chart

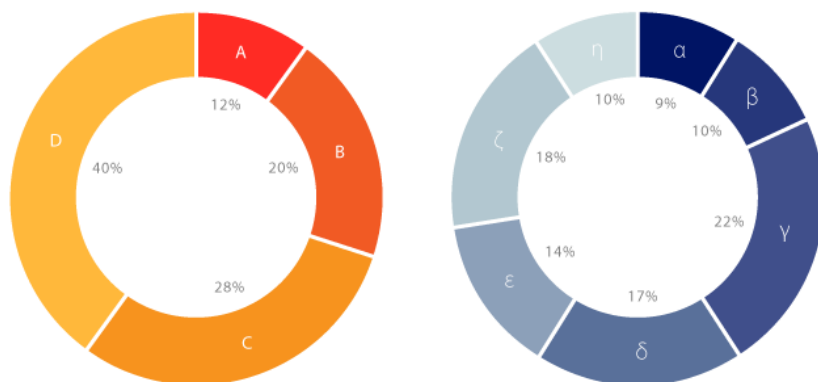


Figura 4.4: Donut Chart

Il **Donut Chart** è essenzialmente un grafico a torta a meno di una zona al centro.

Tuttavia, questi grafici hanno un leggero vantaggio rispetto ai loro cugini, i quali sono spesso criticati sia per la focalizzazione sulle dimensioni relative delle parti rispetto al totale, sia perché quando vengono effettuati confronti tra più grafici a torta, spesso si incontrano difficoltà di lettura causati dalla mancanza di indicazioni relative alla quantità rappresentata dal totale.

Un Donut Chart affronta in parte questo problema de-enfatizzando l'uso della superficie, permettendo all'utente di concentrarsi maggiormente sui cambiamenti dei valori. Infatti questa visualizzazione si focalizza sulla lettura della lunghezza degli archi, piuttosto che sul confronto delle proporzioni come accade nel grafico a torta.

Inoltre i Donut Charts possono occupare uno spazio ridotto rispetto ai grafici a torta, sfruttando il loro centro per mostrare informazioni utili come KPI.

**Funzioni:** Confronti - Parte del totale - Proporzioni

#### 4.2.4 Pie Chart



Figura 4.5: Pie Chart

Ampiamente utilizzato nelle presentazioni e negli uffici, i **Pie Charts** (grafici a torta) contribuiscono a mostrare proporzioni e percentuali tra le categorie, dividendo un cerchio in segmenti proporzionali. La lunghezza di ogni arco rappresenta la proporzione di ciascuna categoria rispetto al totale, rappresentato dal cerchio nella sua interezza e pari al 100%. I grafici a torta sono ideali per dare al lettore una rapida idea della distribuzione dei dati. Tuttavia i principali aspetti negativi dei grafici a torta sono:

- Impossibilità di mostrare un numero elevato di valori, poiché se questo dovesse aumentare notevolmente, la dimensione di ciascun segmento risulterebbe sempre più piccola. Questo li rende inadatti per analizzare un numero elevato di categorie.
- Occupano molto spazio, di conseguenza le loro alternative, come un 100 % Stack Bar Graphs, sono più consigliate in quanto sfruttano in maniera più efficiente lo spazio.
- Non sono adatti per effettuare confronti accurati tra gruppi di Pie Charts. Questo perché risulta difficile distinguere la dimensione degli oggetti rispetto all'area.

Nonostante ciò, confrontare una data categoria (una fetta) all'interno del totale di un singolo grafico a torta, spesso può risultare molto efficace.

**Funzioni:** Confronti - Parte del totale - Proporzioni

## 4.2.5 Radar Chart

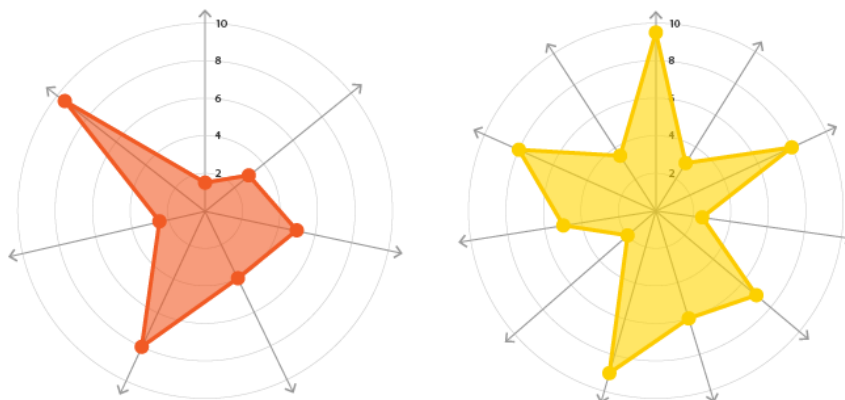


Figura 4.6: Radar Chart

Conosciuto anche come: Spider Chart, Web Chart, Polar Chart, Star Plots, il **Radar Chart** è un grafico che permette di confrontare contemporaneamente più variabili quantitative. La sua struttura permette di vedere quali variabili hanno valori simili e se esistono outliers. IL Radar Chart è inoltre utile per verificare quale variabile ha una valutazione alta/bassa all'interno di un dataset, rendendolo ideale per mostrare le performance.

Ad ogni variabile è associato un asse che ha inizio dal centro. Gli assi sono posizionati in maniera radiale, equidistanti l'uno dall'altro e condividono la stessa scala. Le linee circolari che connettono gli assi tra loro sono spesso usate come guide. Il valore di ogni variabile è mostrato lungo il proprio asse, e tutti i valori del dataset sono connessi tra loro formando un poligono.

I principali difetti di questo grafico sono i seguenti:

- La formazione di più poligoni sullo stesso Radar Chart rende difficile la lettura dei dati.
- L'utilizzo di un numero elevato di variabili porta alla creazione di altrettanti assi rendendo ancora una volta il grafico di difficile lettura.

E' buona norma quindi, mantenere questo grafico il più semplice possibile, e limitare il numero di variabili.

**Funzioni:** Confronti - Relazioni - Patterns

## 4.2.6 Radial Column Chart

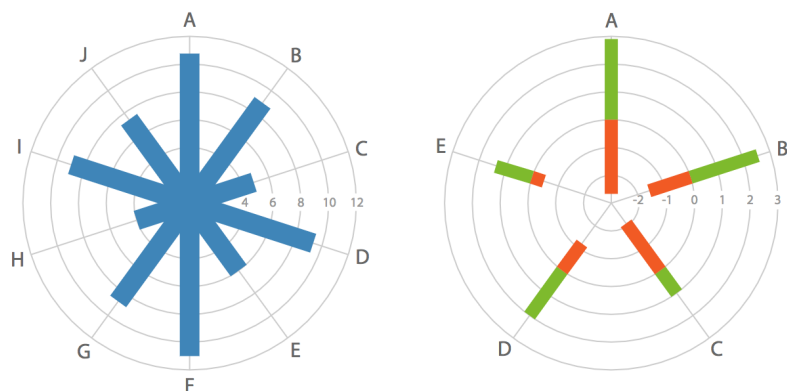


Figura 4.7: Radial Column Chart

Conosciuto anche come Circular Column Graph o Star Graph, il **Radial Column Graph** utilizza una griglia concentrica per mostrare barre. Ogni cerchio del grafico rappresenta un valore della scala adottata, mentre le rette equidistanti che partono dal centro vengono utilizzate per identificare le varie categorie. Le barre lungo le rette infine rappresentano la quantità associata alla categoria.

Solitamente il valore più basso della scala parte dal centro e aumenta ad ogni cerchio. Con il Circular Column Graph però è possibile mostrare anche valori negativi. Posizionando lo zero in uno qualsiasi dei cerchi si può attribuire ai cerchi interni valori negativi.

Le barre normalmente partono dal centro e si estendono verso l'esterno, ma è possibile anche mostrare dei range di valori attribuendo punti di partenza differenti ad ogni categoria. Infine è possibile “impilare” le barre come viene fatto nello Stacked Bar Graph.

**Funzioni:** Confronti

## 4.2.7 Parallel Sets

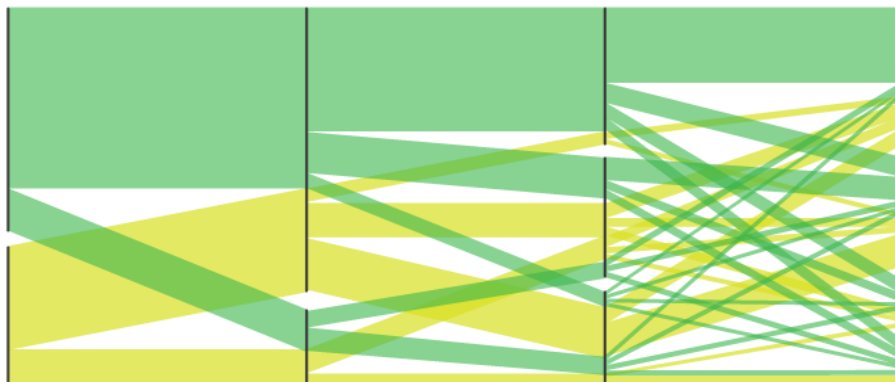


Figura 4.8: Parallel Sets

Il **Parallel Set** è un grafico simile al Sankey Diagrams, un grafico comunemente usato per mostrare i flussi e la loro quantità. A differenza di questo però non fa utilizzo di frecce e i percorsi vengono divisi ad ogni linea verticale. Queste linee corrispondono a dimensioni/dataset, i cui valori/categorie sono rappresentati dai vari segmenti. L'ampiezza di ogni segmento e il flusso che parte da questo, è determinato dalla porzione proporzionale rispetto al totale individuato dall'intera linea. Ogni flusso può essere colorato per mostrare e confrontare la distribuzione tra le differenti categorie.

**Funzioni:** Confronti - Distribuzione - Flussi - Processi - Proporzioni



## 4.2.8 Circle Packing

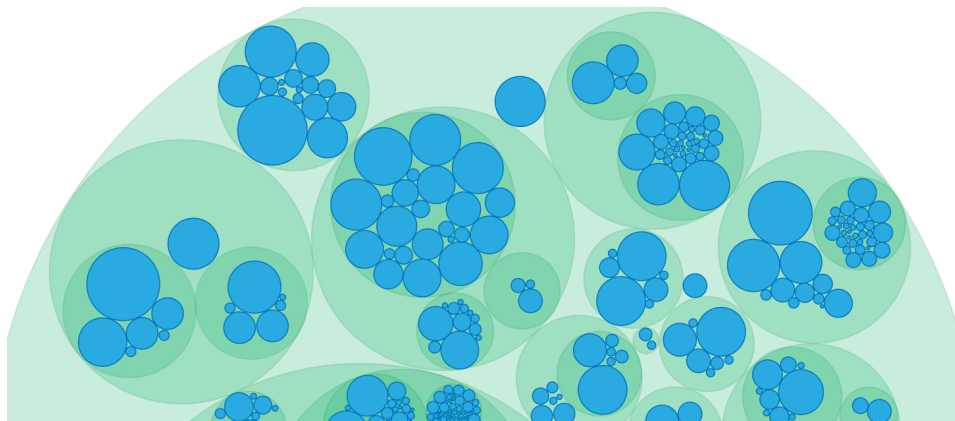


Figura 4.9: Circle Packing

Anche conosciuto sotto il nome di Circular Treemap, il **Circle Packing** è una variazione della Treemap, una tecnica di visualizzazione utilizzata per rappresentare le strutture gerarchiche; a differenza di quest'ultima però, il Circle Packing utilizza cerchi al posto dei rettangoli.

Il contenuto all'interno di ogni cerchio rappresenta un livello inferiore nella gerarchia. Continuando il paragone con la Treemap, ogni ramo dell'albero è rappresentato con un cerchio, e i suoi rami secondari sono rappresentati come cerchi all'interno di esso. L'area di ogni cerchio può essere utilizzata per rappresentare un valore arbitrario aggiuntivo, come per esempio una quantità. Il colore può essere usato anche per assegnare categorie o per rappresentare un'altra variabile tramite diverse tonalità.

Nonostante il gradevole aspetto, il Circle Packing non sfrutta in maniera efficiente lo spazio, al contrario della Treemap, in quanto esiste una grande quantità di spazio vuoto tra i vari cerchi. Tuttavia, questa visualizzazione rivela la struttura gerarchica in maniera migliore rispetto alla Treemap.

**Funzioni:** Gerarchie - Proporzioni

## 4.2.9 Heat Map

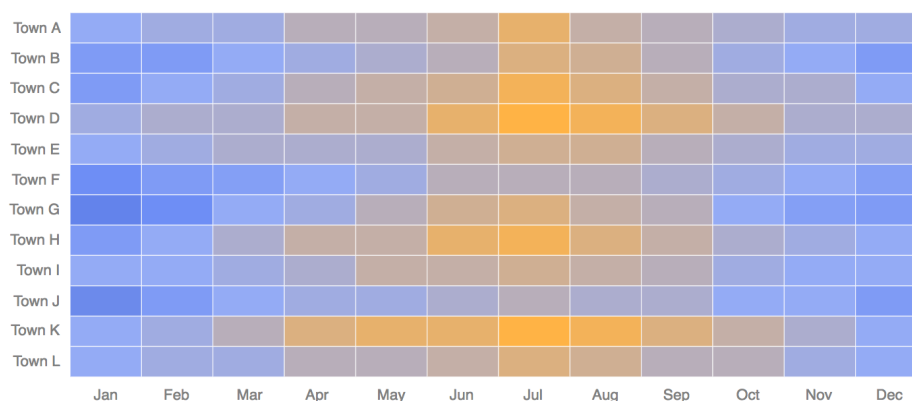


Figura 4.10: Heat Map

La **Heat Map** mostra i dati attraverso variazioni di colore. Tale visualizzazione è utile per esaminare in maniera incrociata dati multivariati, attraverso il posizionamento delle variabili sulle righe e sulle colonne, e colorando le celle all'interno della tabella. E' un ottimo strumento per verificare se alcune variabili sono simili ad altre e per identificare possibili correlazioni tra queste.

Tipicamente, tutte le righe sono assegnate ad una sola categoria, mentre sulle colonne ne viene assegnata un'altra. Le singole righe e colonne sono poi suddivise in sottocategorie. Le celle all'interno della tabella sono codificate attraverso l'utilizzo di colori che possono essere sia in base alla categoria, che in base ad dati numerici suddivisi da una scala di colore. I dati contenuti nelle celle sono relativi alla relazione delle variabili connesse alla riga e alla colonna considerata. Per facilitarne la lettura è necessario l'utilizzo di una legenda che mostri la codifica dei colori adottati.

A causa del forte legame con l'utilizzo dei colori, la Heat Map è un grafico adatto a mostrare una vista generale dei valori numerici dei dati, in quanto risulta difficile identificare in maniera accurata la differenza delle sfumature di colore e quindi l'estrazione dettagliata delle informazioni (a meno che non venga introdotto all'interno della cella il valore numerico esatto).

**Funzioni:** Confronti - Dati nel tempo - Patterns - Relazioni

## 4.3 Comparazione dei grafici

Una volta identificate le visualizzazioni più adatte a mostrare i dati a nostra disposizione, è stato necessario confrontarle seguendo le fasi introdotte nel Capitolo 3. I questionari sono stati creati attraverso l'utilizzo di Google Forms, un servizio gratuito di Google grazie al quale è possibile effettuare sondaggi accessibili da qualsiasi dispositivo. Per ottenere il maggior numero di risultati abbiamo deciso di effettuare le indagini sfruttando i Social Networks. Il questionario è stato quindi condiviso con i gruppi di Ingegneria Informatica di Milano, di Como e con la mia rete sociale personale, in modo da ottenere un target di persone ben distribuito. Durante la seconda indagine i dati raccolti non hanno sfruttato solo i canali descritti precedentemente, ma hanno interessato anche una serie di interviste effettuate durante l'evento fieristico SMAU svoltosi a Milano il 25, 26 e 27 Ottobre 2016.

Andiamo ora ad analizzare come abbiamo affrontato queste fasi della nostra ricerca.

### 4.3.1 Comparazione tra visualizzazioni tradizionali

Nella comparazione tra visualizzazioni tradizionali l'obiettivo è confrontare i grafici comunemente usati per mostrare i dati presi in esame. Nel nostro caso, nella quale vogliamo analizzare una quantità aggregata rispetto a tre dimensioni, quali i 3 Social Network, per ognuno dei giudici di MasterChef, le visualizzazioni tradizionali utilizzate per mostrare queste informazioni sono le seguenti: *Stacked Bar Graph (o Stacked Bar Chart)*, *Histogram*, *Donut Chart*, *Pie Chart*, *Radar Chart*, *Radial Column Chart*. Per confrontare queste tecniche di visualizzazione abbiamo seguito i seguenti step:

1. Progettazione Grafici
2. Raccolta Feedback

### Progettazioni Grafici

In questa fase abbiamo progettato e adattato le visualizzazioni tradizionali in base al nostro caso studio. Per lo sviluppo di tali grafici è stato utilizzato Adobe Photoshop CC 2017, in modo da avere una maggiore flessibilità e rapidità durante la progettazione e lo sviluppo. Abbiamo seguito le linee guida introdotte nel Capitolo 2 ottenendo così i grafici mostrati in fig. 4.11

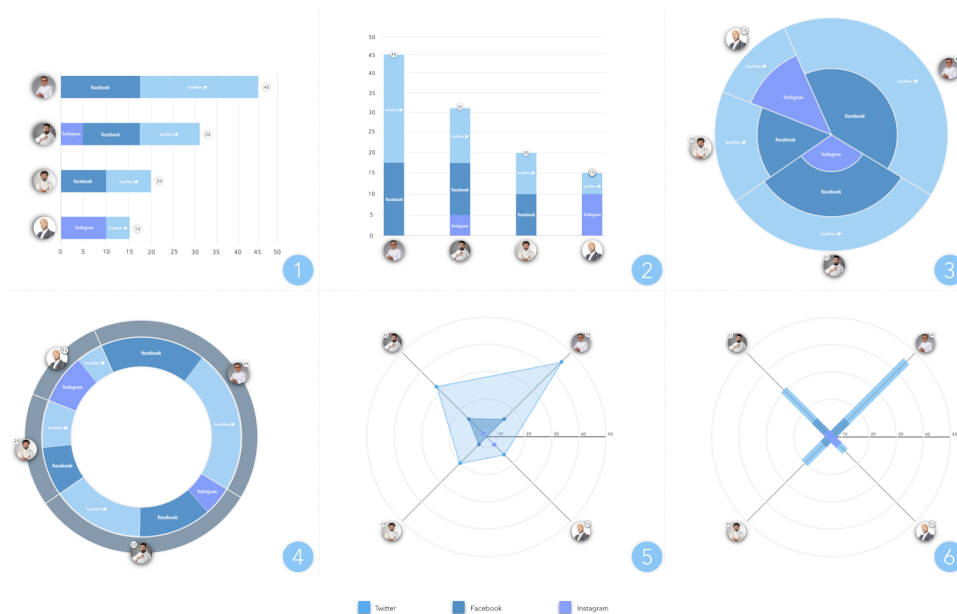


Figura 4.11: Grafici confrontati durante la comparazione tra visualizzazioni tradizionali

## Raccolta Feedback

A seguito della progettazione dei grafici siamo andati a confrontare questi con l'obiettivo di raccogliere feedback degli utenti. Come introdotto precedentemente abbiamo creato e distribuito questionari agli utenti. Nella sezione successiva è descritta la struttura del questionario, entrando nel dettaglio di ogni singola domanda, spiegando il perché di tale scelta e l'obiettivo che sta dietro ad ognuna di queste. Il questionario proposto in questa fase è stato suddiviso in 6 step più una fase preliminare introduttiva:

## Informazioni degli intervistati

In questo step introduciamo l'utente nel contesto. Viene spiegato l'obiettivo della ricerca relativa alla raccolta dati tramite l'utilizzo del questionario proposto e l'ambito di ricerca in cui questo è coinvolto. Successivamente raccogliamo alcune informazioni demografiche, quali il sesso la professione, la tipologia di educazione e il paese. E' stato deciso di raccogliere tali informazioni per verificare se queste fossero in qualche modo correlate con le risposte fornite nei passi successivi.

### **Conoscenza degli intervistati**

In questo step viene chiesto all'intervistato di elencare il nome di tutte le tecniche di *Data Visualization* di cui è a conoscenza. La definizione di *Data Visualization* non è stata fornita di proposito e non è stato inserito un elenco tra cui scegliere. Questo per verificare effettivamente quanto l'intervistato abbia confidenza con l'argomento. Come nello step precedente questa informazione verrà poi utilizzata per verificare se esiste un certo livello di correlazione tra le risposte nei seguenti step, e il livello di conoscenza e confidenza che l'intervistato ha con tale argomento.

### **Aspettative degli intervistati**

Questo step inizia con la seguente domanda: *If a monitor is showing the popularity of a person, which graph do you expect to see ?* Con questa domanda l'obiettivo è valutare cosa l'intervistato si aspetta di vedere nel momento in cui gli dovesse essere mostrato un grafico che rappresenta la "popolarità". Anche in questo caso non abbiamo dato nessuna definizione, in modo da non introdurre informazioni aggiuntive riguardanti l'analisi dei grafici che verranno mostrati nei seguenti step, ma valutare in assenza di informazioni cosa si aspetta di vedere l'utente.

Successivamente viene mostrato un grafico tra i sei mostrati in Figura 4.11. La scelta di quale grafico mostrata viene fatta in maniera casuale. Tale decisione è stata dettata dalla necessità di ridurre al più possibile effetti bias dovuti ad una possibile acquisizione di confidenza col grafico mostrato in questa fase rispetto a quelli mostrati negli step successivi, la quale potrebbe deviare la scelta delle risposte alla domande seguenti. Una volta mostrato il grafico, viene chiesto all'intervistato quali informazioni è in grado di estrapolare. Anche in questo caso non vengono fornite informazioni aggiuntive, così da verificare quanto sia auto-esplicativa la visualizzazione proposta.

### **Reazioni degli intervistati**

In questo step invece l'obiettivo è mettere a confronto tutti i grafici illustrati in Figura 4.11. Prima di mostrare le visualizzazioni viene fatta un introduzione all'intervistato nella quale gli viene spiegato cosa sta per vedere, il contesto, il significato del dato visualizzato e che tipo di informazione vogliamo trasmettere. Successivamente vengono mostrati i 6 grafici contemporaneamente e gli viene chiesto di sceglierne uno per ognuno dei seguenti criteri: quello più intuitivo, quello meno intuitivo e infine quello con un maggiore

appeal. Abbiamo deciso di confrontare i grafici tutti assieme per verificare come l'intervistato reagisce di fronte alla possibilità di scegliere una visualizzazione tra un set che mostra le stesse quantità e le stesse informazioni ma in modi differenti.

### **Confronti tra grafici simili**

A questo punto abbiamo confrontato le visualizzazioni a coppie. Sempre in riferimento alla Figura 4.11 le coppie formate sono state: il grafico 1 con il 2, il 3 con il 4 e infine il 5 con il 6. Per ognuna di queste coppie abbiamo chiesto all'intervistato quale visualizzazione tra quelle coinvolte fosse maggiormente leggibile e intuitiva. Questo è stato fatto per essere in grado di stilare una classifica e confrontare tra loro visualizzazioni simili nella forma e nella struttura e verificare infine se una tecnica, simile ad un'altra, sia più leggibile rispetto a questa.

### **Domande di controllo**

In questo step vengono invece effettuate domande di controllo a risposta multipla, finalizzate alla verifica dell'attendibilità delle risposte. Vengono in primo luogo riproposti i 6 grafici contemporaneamente. Successivamente si chiede all'intervistato di scegliere quale visualizzazione secondo lui sia quella più adatta a mostrare la popolarità complessiva sui Social Network. Dopodiché, per verificare l'attendibilità della risposta, viene chiesto all'utente di indicare quale sia lo Chef più popolare sui Social Networks. Per aiutare l'intervistato nella scelta, viene fornita una immagine informativa raffigurante l'associazione del nome dello chef con il volto corrispondente. Poiché tali visualizzazioni non mostrano soltanto la popolarità considerando l'insieme di tutti Social Networks coinvolti, ma mostrano anche il dettaglio per singolo Social, viene chiesto all'utente di scegliere da quale grafico è in grado di estrapolare in maniera più rapida e chiara tale informazione. Come per la domanda precedente, successivamente si chiede all'intervistato di indicare in quale Social lo chef Joe Bastianich sia più popolare, in modo da validare la risposta della domanda precedente.

### **Feedback intervistati**

Nell'ultimo step infine, lasciamo all'utente la libertà di fornire qualsiasi tipo di feedback riguardante i grafici, se l'utilizzo di altri colori può aiutare nella leggibilità ed eventualmente quali. Si conclude infine il questionario chie-

dendo la mail nel caso l'utente volesse partecipare ai successivi questionari ed essere a conoscenza dei risultati ottenuti.

### 4.3.2 Comparazione con visualizzazioni innovative

In questa seconda fase dell'indagine l'obiettivo principale è quello di confrontare le visualizzazioni che sono risultate essere le più efficaci e le più apprezzate nella fase precedente con quelle innovative, le quali rappresentano quella tipologia di visualizzazioni non generalmente utilizzate per mostrare le informazioni del nostro caso studio. Sulla base dei feedback e dei suggerimenti ottenuti nell'intervista precedente, siamo andati ad apportare al questionari delle miglioni e delle correzioni rispetto alla stesura originale, in modo da rendere le domande più appropriate e più rilevanti per il nostro studio. Il tutto senza però modificare la struttura di base, così da non perdere la possibilità di confrontarlo con quello del passo precedente. I grafici messi a confronto in questa fase sono quindi i seguenti: *Stacked Bar Graph (o Stacked Bar Chart)*, *Parallel Sets*, *Pie Chart*, *Donut Chart*, *Circle Packing*, *Heatmap* In analogia con il precedente questionario, per confrontare queste tecniche di visualizzazione abbiamo effettuato i seguenti step:

1. Progettazione Grafici
2. Raccolta Feedback

#### **Progettazioni Grafici**

A differenza della precedente intervista, nella quale abbiamo progettato le visualizzazioni basandoci unicamente sulle linee guida presenti in letteratura, in questa fase abbiamo tenuto in considerazione anche i feedback ottenuti dagli intervistati. Sono state apportate delle modifiche nei colori utilizzati e applicato alcuni accorgimenti grafici per migliorarne la lettura, il tutto senza stravolgere la struttura di base utilizzata precedentemente. Per lo sviluppo di tali grafici è stato utilizzato ancora una volta Adobe Photoshop CC 2017. I grafici così ottenuti sono raffigurati nella Figura 4.12.

#### **Raccolta Feedback**

Una volta progettati i grafici quello che rimane da fare, così come nella Fase 1, è confrontare questi e ottenere i feedback degli utenti. Andiamo ad analizzare ora la struttura del questionario entrando nel dettaglio di ogni singola domanda, spiegando il perché di tale scelta e l'obiettivo che sta dietro ad ognuna di esse. Il questionario proposto in questa fase è sempre suddiviso in 6 step più una fase preliminare introduttiva:

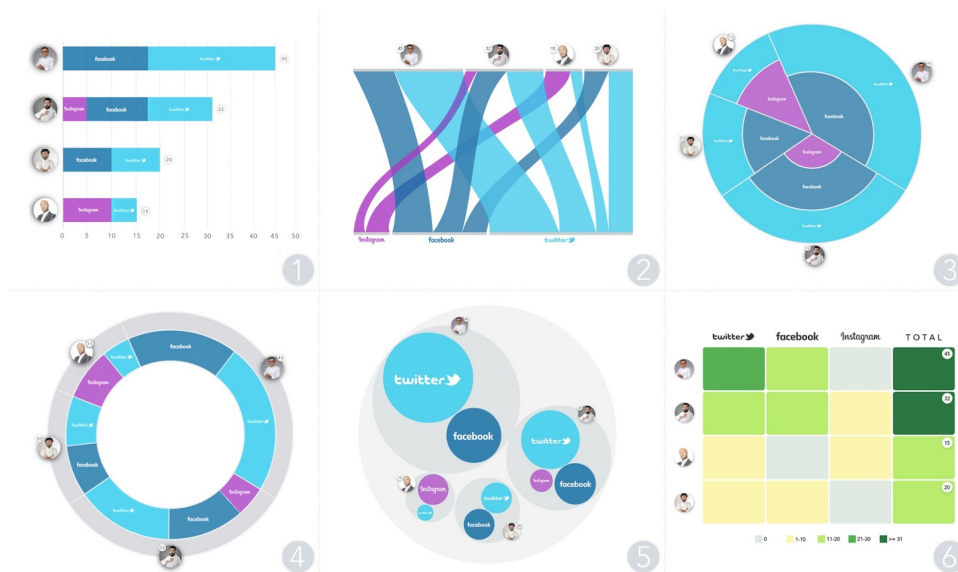


Figura 4.12: Grafici confrontati durante la comparazione con visualizzazioni innovative

### Informazioni degli intervistati

In questa fase introduciamo l'utente nel contesto. Viene quindi spiegato l'obiettivo che sta dietro alla raccolta dati tramite l'utilizzo di questo questionario proposto e l'ambito di ricerca in cui questo è coinvolto. Successivamente raccogliamo alcune informazioni demografiche, quali l'età il sesso e il paese di origine. A differenza del precedente questionario però, chiediamo all'intervistato quale sia il proprio livello di conoscenza nell'ambito della *Data Visualization* e nella *Data Analysis*. Questa scelta è stata dettata dalla volontà di verificare se esistono correlazioni tra il background e la dimestichezza che un utente ha in questo ambito e l'efficacia che un determinato grafico ha su di lui.

### Conoscenza degli intervistati

Successivamente viene chiesto all'intervistato di elencare il nome di tutte le tecniche di *Data Visualization* che conosce. Di proposito non è stata data la definizione di *Data Visualization* e non è stato inserito un elenco tra cui scegliere. Questo per verificare effettivamente quanto l'intervistato abbia confidenza con l'argomento. Come per la fase introduttiva questa informazione verrà poi utilizzata per verificare se esiste un certo livello di



correlazione tra le risposte dei seguenti step, e il livello di conoscenza e confidenza che l'intervistato ha con tale argomento. Inoltre viene utilizzata come domanda di controllo rispetto alle autovalutazioni fatte al precedente step.

### **Aspettative degli intervistati**

Il secondo step inizia con la seguente domanda: *If a monitor is showing the popularity of a person, which graph do you expect to see?* Con questa domanda l'obiettivo è valutare cosa l'intervistato si aspetta di vedere nel momento in cui gli dovesse essere mostrato un grafico che rappresenta la "popolarità". Anche in questo caso non abbiamo dato nessuna definizione, questo per non introdurre informazioni aggiuntive riguardanti l'analisi dei grafici che verranno mostrati nei seguenti step, ma valutare in assenza di informazioni cosa si aspetta di vedere l'utente.

A seguito di tale domanda viene mostrato all'intervistato il grafico *Stacked Bar Graph (o Stacked Bar Chart)*. Una volta mostrato il grafico viene chiesto quali informazioni è in grado di estrapolare. Anche in questo caso non vengono fornite informazioni aggiuntive, questo per verificare quanto sia auto-esplicativa la visualizzazione proposta.

Successivamente viene proposto un grafico tra quelli mostrati in Figura 4.12, ad eccezione dello *Stacked Bar Graph (o Stacked Bar Chart)* in quanto mostrato subito prima, e si chiede all'utente se dal grafico riesce ad estrapolare le stesse informazioni rispetto al precedente, oppure se identifica informazioni differenti.

La scelta di quale grafico proporre viene fatta in maniera casuale. Tale scelta è stata dettata dalla necessità di ridurre al più possibile effetti bias dovuti ad una possibile maggiore leggibilità di tale grafico negli step successivi, la quale, potrebbe deviare l'intervistato nella scelta delle risposte alla domande, visto il maggiore livello di confidenza preso con la visualizzazione. Oltre a questo l'obiettivo è riuscire a confrontare tutti i grafici mostrati in Figura 4.12 con lo *Stacked Bar Graph (o Stacked Bar Chart)* in quanto questo è risultato essere il grafico più efficace durante l'intervista precedente.

### **Rimanti fasi**

Per tutte le successive fasi dell'indagine, l'ordine e le tipologie di domande fatte sono rimaste invariate rispetto al questionario per confrontare le tecniche tradizionali, a meno dei grafici utilizzati.

## 4.4 Comparazione tra le visualizzazioni migliori

In questa sezione andremo a descrivere come è stato affrontato l'ultimo step della nostra ricerca, focalizzandoci sulle tecnologie e sulle scelte implementative.

L'obiettivo di questa terza ed ultima fase è quello di confrontare i grafici che sono risultati essere i più efficaci durante le precedenti indagini, ovvero lo *Stacked Bar Graph* (o *Stacked Bar Chart*) e il *Circle Packing*. A differenza di queste però, vogliamo analizzare sia i pregi che i difetti delle visualizzazioni, in modo da valutare come l'utente reagisce non solo quando l'informazione è rappresentata in modo chiaro, ma anche quando essa risulta di difficile lettura. Per fare questo abbiamo deciso di implementare le visualizzazioni attingendo da dati reali e di rendere queste interattive, così da mettere l'utente in condizioni di navigare e interagire con esse.

Per confrontarle e valutarle come appena descritto, abbiamo deciso di sviluppare una applicazione web, così da renderle portabili e accessibili ad un maggior numero di utenti. Questo lavoro ha portato alla realizzazione di *DataVisualization Research*, un applicazione web accessibile tramite il link <https://datavizresearch.herokuapp.com>, nella quale è possibile navigare e interagire con le visualizzazioni risultate essere le migliori per mostrare il dato del nostro caso studio, accompagnate da una serie di domande col fine di raccogliere i dati necessari per effettuare le analisi conclusive. Come funzionalità secondaria, è stato reso possibile consultare i risultati ottenuti dalle due interviste precedenti. . Andiamo ora a vedere quali sono stati i passi che hanno portato allo sviluppo di questa applicazione.

### 4.4.1 Introduzione alle tecnologie

Il primo problema che abbiamo dovuto affrontare durante il lavoro di sviluppo dell'applicazione, è stato quello di decidere quali tecnologie utilizzare. Dopo un fase preliminare di ricerca, abbiamo identificato in *JavaScript* la tecnologia più adatta per la raggiungimento del nostro scopo, in quanto questa fornisce un gran numero di librerie adatte per la creazione di visualizzazioni interattive. Inoltre è possibile accedere ad una vasta documentazione e numerose community dove trovare molte informazioni a riguardo. Confrontando e analizzando le varie librerie a disposizione, abbiamo scelto di sfruttare due di queste: *D3.js* [5, 6] e *React.js* [16]. Di seguito viene fornita un breve descrizione di tali strumenti e del perché abbiamo scelto di utilizzarli.

## JavaScript

JavaScript (in breve *JS*) è un linguaggio di programmazione ad alto livello, interpretato, dinamico e debolmente tipizzato [11]. E' un linguaggio di scripting orientato agli oggetti e agli eventi, comunemente utilizzato nella programmazione Web lato client per la creazione, in siti web e applicazioni web, di effetti dinamici interattivi tramite funzioni di script invocate da eventi innescati a loro volta in vari modi dall'utente sulla pagina web in uso (mouse, tastiera, caricamento della pagina ecc...) [44].

Per la nostra applicazione abbiamo usato ECMAScript 2015 (ES6), in breve ES2015, il nuovo standard JS che è stato rilasciato nel giugno del 2015. ES2015 è stato un aggiornamento molto significativo per questo linguaggio, il primo aggiornamento importante dal 2009. L'aggiornamento di queste nuove caratteristiche nei principali motori di JavaScript è ancora in corso, questo significa che molti browser non supportano pienamente il nuovo standard. Per utilizzare quindi la nuova versione del codice bisogna effettuare una transpilazione (nota anche come compilazione source-to-source). Il transpilatore utilizzato è stato Babel [1].

Insieme a JS abbiamo usato JSX, un'estensione della sintassi JavaScript che permette di sfruttare al meglio i benefici offerti da React [15].

La scelta di questa tecnologia è stata quindi dettata dal fatto che fornisce ottimi strumenti per creare animazioni e interazioni sul web, fornendo un gran numero di librerie ad hoc per la gestione di visualizzazioni interattive.

## D3.js

D3.js, indicato anche semplicemente come D3, è una libreria JavaScript per la creazione di visualizzazioni dei dati. La sigla D3 fa riferimento al nome completo dello strumento, *Data-Driven Documents*. Quello che serve per sfruttare questa libreria sono i dati, i quali devono essere forniti dallo sviluppatore, e la creazione di documenti basati sul Web, il che significa tutto ciò che può essere renderizzato da un browser web, come documenti *HTML* e *SVG*. D3 fa da ponte tra queste due cose, collegando i dati ai documenti. Naturalmente, il nome funziona anche come allusione intelligente per la rete di tecnologie di base dello stesso strumento: *W3*, o *World Wide Web*, o, oggi, semplicemente il *Web*.

L'autore principale di D3 è *Mike Bostock*. Il progetto è interamente open source e disponibile su GitHub. D3 è rilasciato sotto licenza BSD, quindi si può utilizzare, modificare e adattare sia per uso non commerciale che commerciale, senza alcun costo [25].

Utilizzare questa libreria per creare le nostre visualizzazioni è stato quasi un obbligo. D3 oggi probabilmente fornisce il maggior numero di funzioni e strumenti per la manipolazione dei dati e per creare visualizzazioni. In aggiunta a questo esiste una community fortemente sviluppata dove è possibile trovare una gran mole di documentazione, dalla quale prendere molti spunti e informazioni [18].

## React.js

Anche se JavaScript nasce e viene utilizzato come linguaggio di scripting, integrato quindi, all'interno di un altro programma [44], oggi è possibile scrivere un'intera applicazione web con questo linguaggio. L'utilizzo di alcune librerie però può aiutare e semplificare tutto il processo di sviluppo.

Oggi esistono centinaia di librerie/framework JavaScript, e tra le più famose e utilizzate troviamo sicuramente Angular.js e React.js [2, 13, 16].

React.js è una libreria open-source, parte del grande catalogo di librerie offerte da Facebook [17], utilizzata per la creazione di interfacce utente [16], e basata sulla sintassi JSX [15], estensione della sintassi JavaScript.

Le motivazioni che ci hanno portato a scegliere questa libreria sono le seguenti:

- è stata creata e utilizzata da Facebook anche per il proprio sito web, questo probabilmente significa avere supporto e aggiornamenti per un lungo periodo.
- è utilizzata da grandi brand come Instagram, Netflix, Uber etc... [19].
- è open-source ed ha grande community di supporto.
- è una libreria molto leggera e flessibile.

### 4.4.2 Descrizione dell'applicazione

Nella presente sezione introduciamo quelli che sono gli aspetti generali dell'applicazione e i suoi requisiti.

#### Obiettivo

L'obiettivo di questa applicazione è quello di mettere a confronto, attraverso l'utilizzo di domande specifiche, le visualizzazioni selezionate e descritte all'inizio di questa sezione: lo *Stacked Bar Graph* (o *Stacked Bar Chart*) e il *Circle Packing*.

## **Interfaccia Utente/Sistema**

In quanto applicazione web, è necessario che sia presente una connessione internet per accedere ad essa. Gli utenti possono interfacciarsi a questa attraverso i browser più utilizzati quali: *Chrome, Safari e Firefox*, sia tramite PC che dispositivi mobili.

## **Servizi di Hosting**

Front End:

- URL : <https://datavizresearch.herokuapp.com>
- PaaS : Heroku
- Source : <https://www.heroku.com/>

Back End:

- URL : <http://datavizresearchbackendpy.tech/>
- IaaS : Digital Ocean
- VM OS : Ubuntu
- Versione : 14.04.5
- Web Server : Apache 2
- Source : <https://www.digitalocean.com/>

## **Funzionalità Principali**

Le funzioni principali dell'applicazione sono le seguenti:

- Consultazione dei risultati dei precedenti questionari.
- Visualizzazione e interazione con le due tecniche di Data Visualization selezionate.
- Raccolta dei feedback derivanti dalle domande presenti nell'applicazione.

## **Caratteristiche degli Utenti**

Per accedere all'applicazione gli utenti devono essere in grado di connettersi a internet attraverso l'utilizzo di un browser. Non è necessaria nessuna registrazione per sfruttare le funzionalità offerte dall'applicazione. Tali funzionalità sono accessibili con facilità, in quanto il sistema è in grado di guidare l'utente durante tutte le fasi di interazione.

## Vincoli

Lo sviluppo e la manutenzione dell'applicazione *DataVisualization Research* deve rispettare i seguenti vincoli:

- Limitazioni Hardware: L'applicazione non necessita di alcun vincolo a livello hardware.
- Interfacce con Servizi Esterni: L'applicazione non sfrutta nessun servizio esterno.
- Requisiti Linguaggi ad Alto Livello: Lo sviluppo e la gestione dell'applicazione richiede conoscenze riguardo a: JSX, React.js, D3.js, HTML, CSS, SASS e Python.
- Requisiti di Affidabilità: In situazioni di gestione normale del flusso di dati non è richiesto nessun vincolo per assicurare il corretto funzionamento, se non quello di avere una connessione internet attiva nel momento dell'invio dei dati dal Front End al Back End.
- Altre Considerazioni: L'applicazione è facile da usare e deve rimanere tale. L'utente non spende molto tempo per interagire con essa, e in aggiunta l'applicazione ha un aspetto gradevole e piacevole. Il design è user-friendly, col fine di rendere l'esperienza utente il migliore possibile.

### 4.4.3 Requisiti di interfaccia utente

L'applicazione si suddivide in 4 aree principali:

- Home: area in cui viene presentato il progetto di ricerca, l'obiettivo, il lavoro svolto fino ad ora e infine chi siamo.
- Introduction Area: area dove viene fornita all'utente una spiegazione dettagliata del caso studio e ulteriori dettagli riguardo l'obiettivo della nostra ricerca.
- Visualization Area: la Visualization Area è l'area più importante. In questo punto dell'applicazione infatti vengono esposti all'utente i quesiti necessari per il confronto delle visualizzazioni e viene data la possibilità di navigare e interagire con esse.
- Thanks Area: area conclusiva dell'applicazione, dove viene ringraziato l'utente per il tempo dedicato a questa ricerca e viene invitato, se interessato, a tornare alla home per visualizzare i risultati ottenuti nelle precedenti indagini.

## User Interface - MockUp

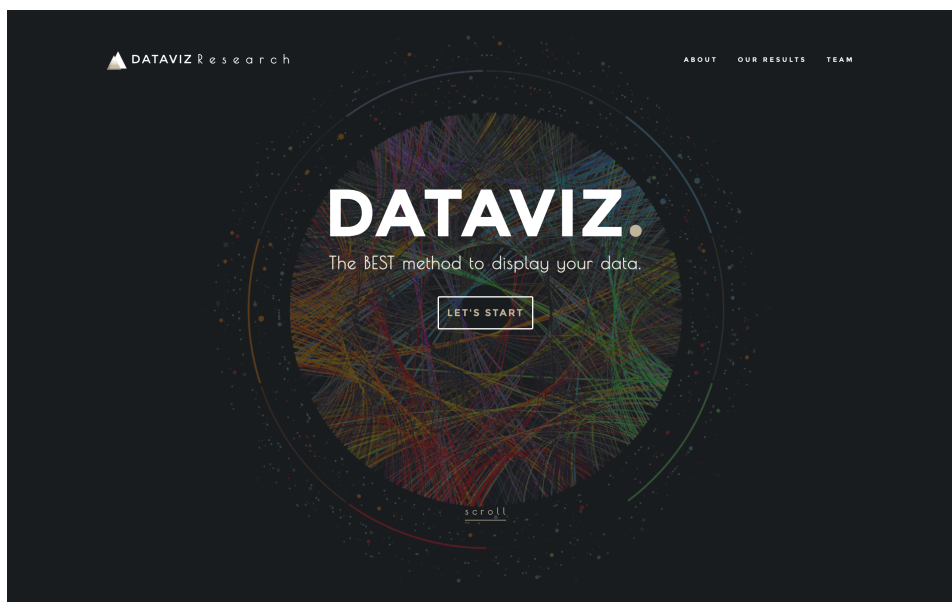


Figura 4.13: MockUp Home - Start

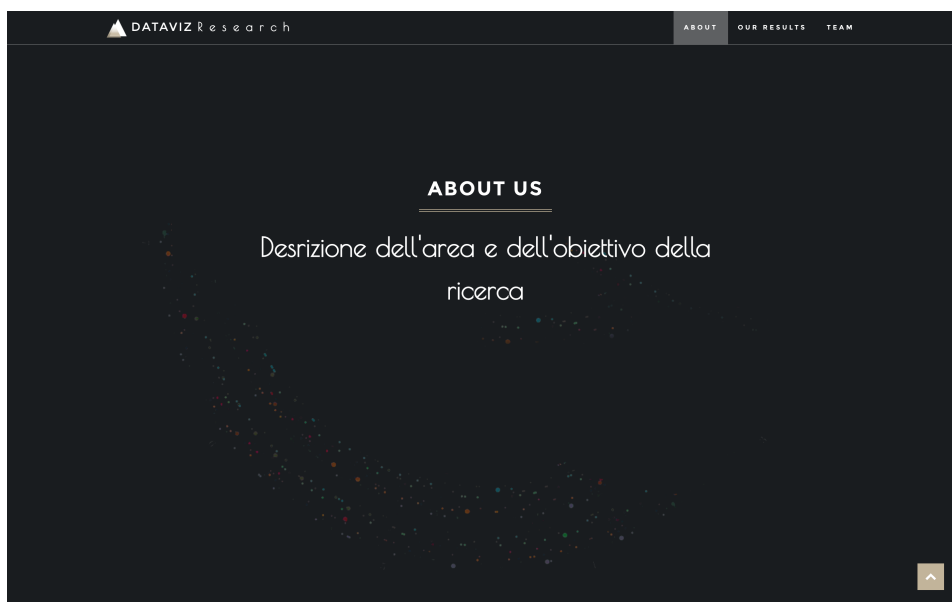


Figura 4.14: MockUp Home - Descrizione della Ricerca

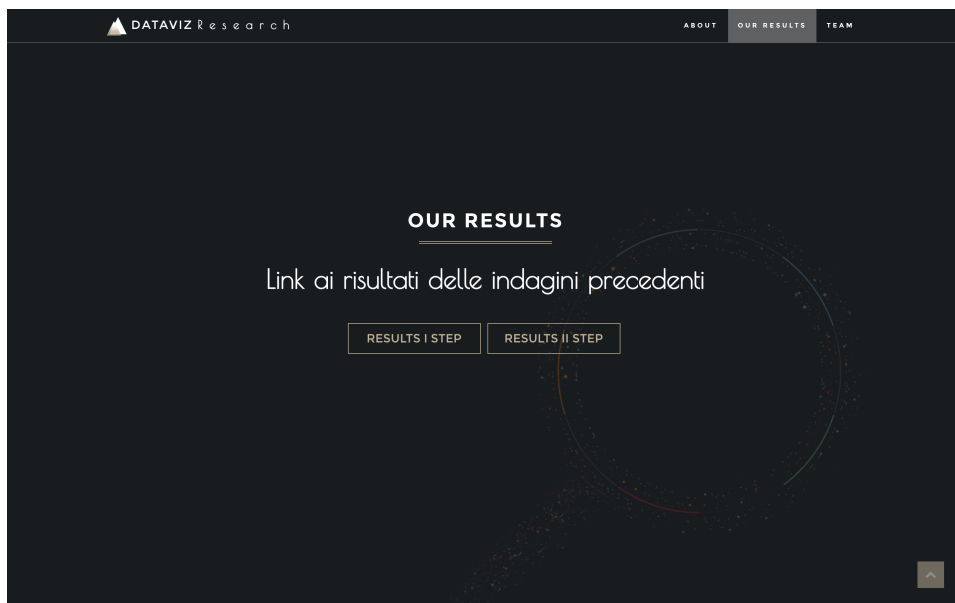


Figura 4.15: MockUp Home - Link ai risultati

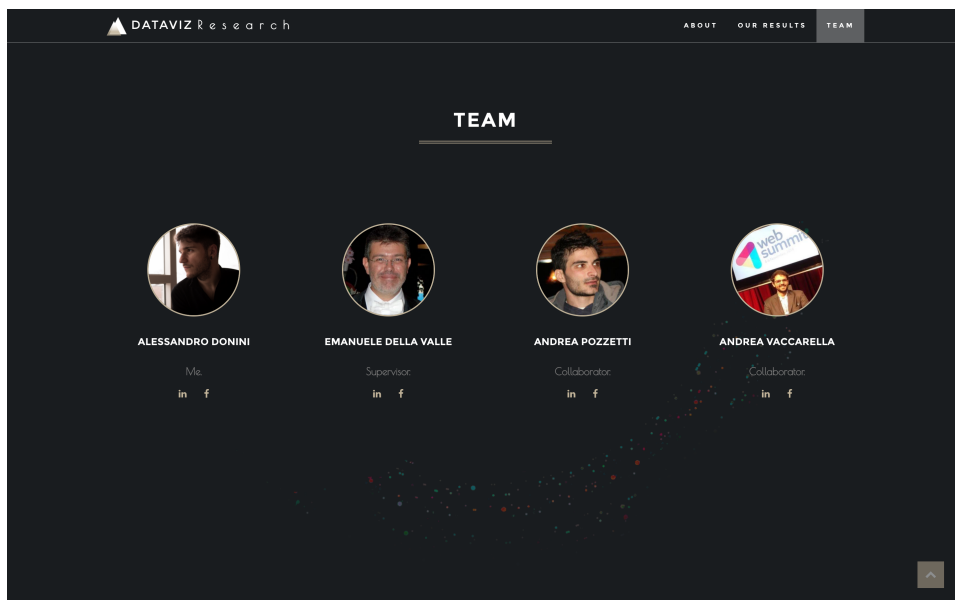


Figura 4.16: MockUp Home - Chi siamo



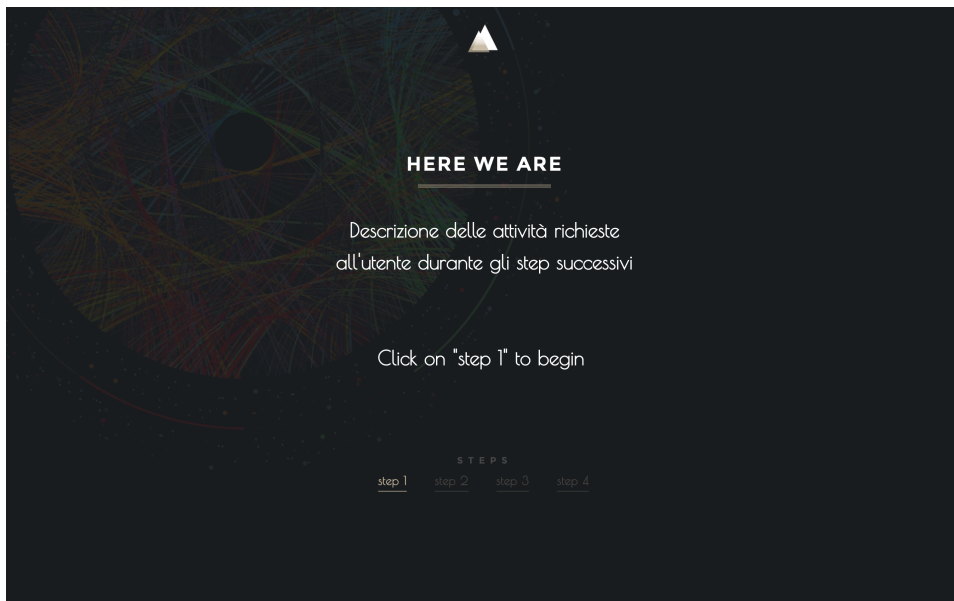


Figura 4.17: MockUp Introduction Area - Descrizione dell'attività richiesta all'utente

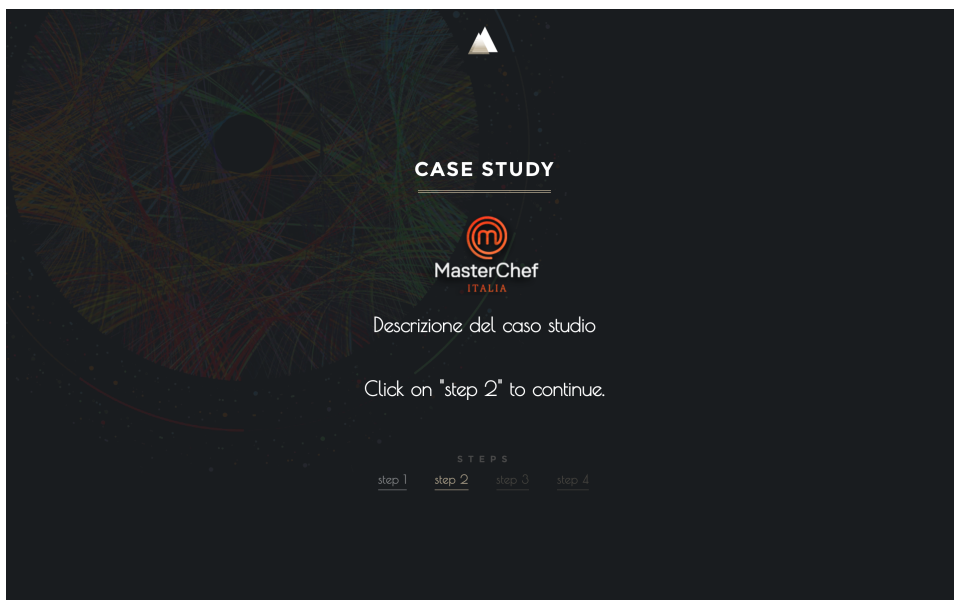


Figura 4.18: MockUp Introduction Area - Descrizione del caso studio

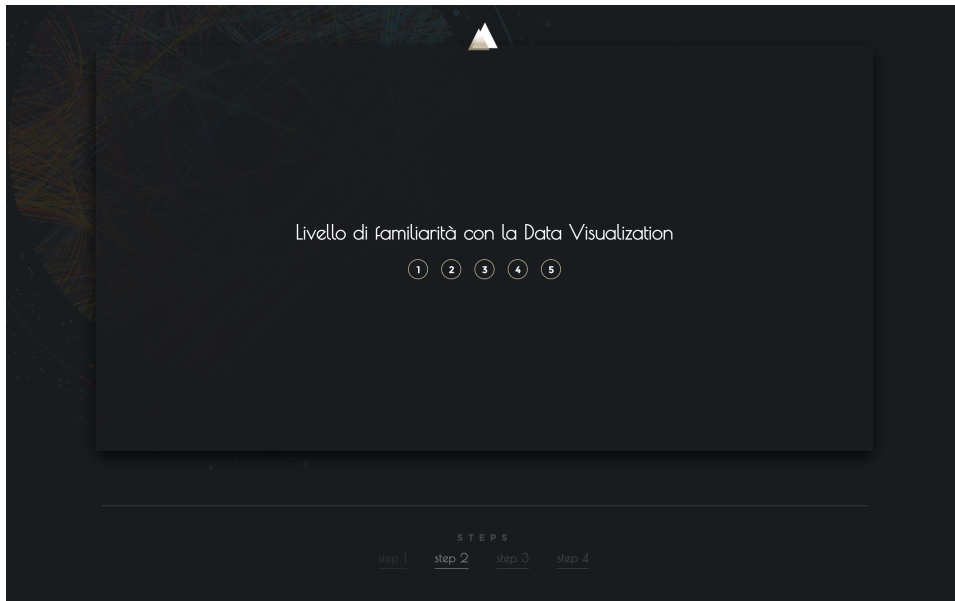


Figura 4.19: MockUp Visualization Area - Autovalutazione



Figura 4.20: MockUp Visualization Area - Stacked Bar Graph

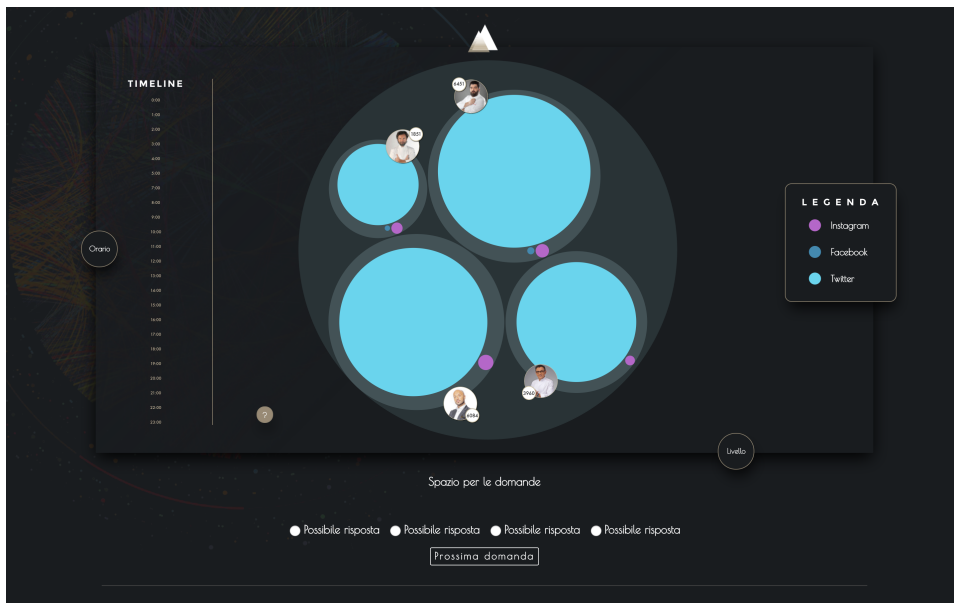


Figura 4.21: MockUp Visualization Area - Circle Packing

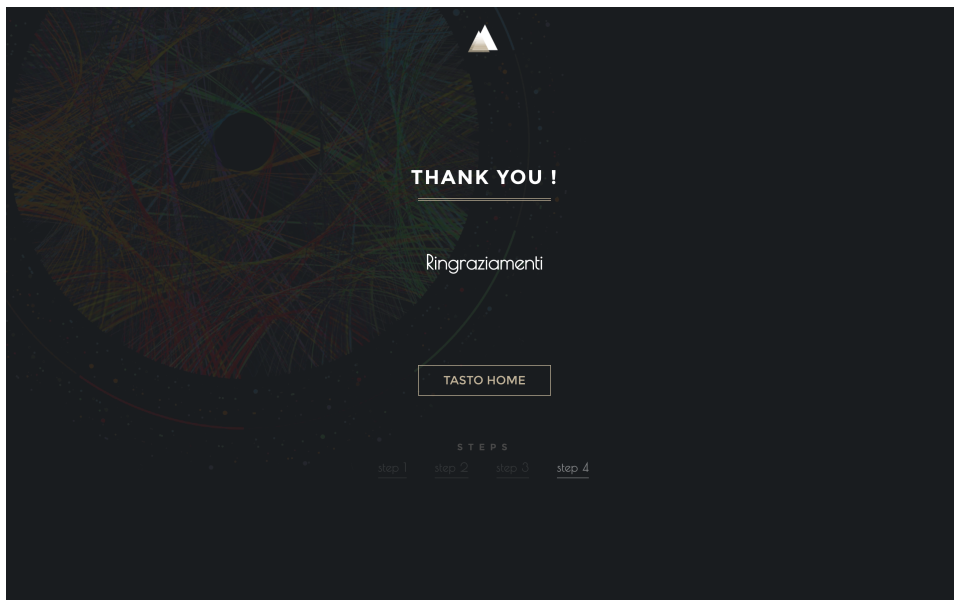


Figura 4.22: MockUp Thanks Area - Ringraziamenti finali

#### 4.4.4 Requisiti funzionali

L'applicazione fornisce a tutti gli utenti le medesime funzionalità. Per analizzare come l'utente interagisce con essa, di seguito sono rappresentati e descritti i principali casi d'uso attraverso l'utilizzo di Sequence Diagram. I casi d'uso analizzati coprono le principali azioni che l'utente è in grado di effettuare all'interno dell'applicazione quali:

1. Introduzione all'attività richiesta.
2. Visualizzazione dei grafici.
3. Questionario.

##### Introduzione all'attività richiesta

- Nome: Introduzione all'attività richiesta
- Attori: Utente
- Condizioni di ingresso:  
L'utente si è collegato correttamente all'applicazione attraverso il link <https://datavizresearch.herokuapp.com>.
- Flusso degli eventi:  
Il sistema mostra la Home Page. L'utente preme sul tasto start. Il sistema salva il nuovo stato dell'applicazione e mostra all'utente la schermata introduttiva dove vengono forniti dettagli riguardo all'obiettivo della ricerca e alle attività richieste. In seguito l'utente preme sul tasto relativo allo step successivo. Il sistema aggiorna lo stato dello step, e mostra le informazioni relative al nuovo stato dove viene fornita all'utente una descrizione del caso studio.
- Condizioni di uscita:  
L'utente preme sul tasto corrispondente allo Step 2 o su quello corrispondente alla home.
- Eccezioni: Non ci sono eccezioni

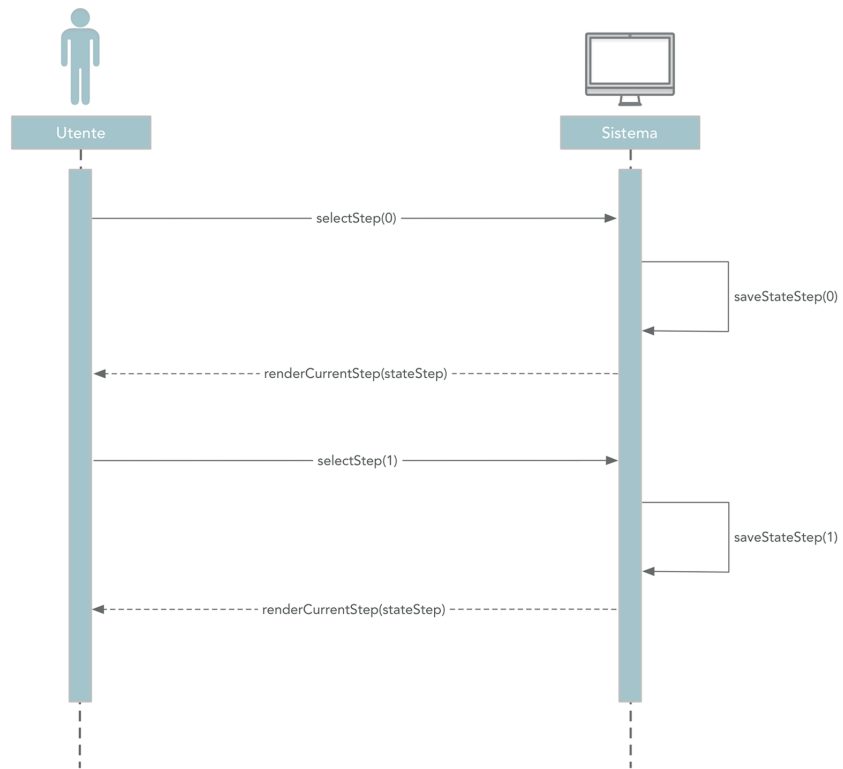


Figura 4.23: Sequence Diagram - Introduzione all'attività richiesta

## Questionario

- Nome: Questionario
- Attori: Utente
- Condizioni di ingresso:  
L'utente ha selezionato il proprio livello di conoscenza della materia o ha concluso la prima parte del questionario e ha premuto il tasto relativo allo step 3.
- Flusso degli eventi:  
Il sistema mostra all'utente una domanda alla volta. L'utente seleziona la risposta e preme sul tasto per confermare questa. Ogni volta che la risposta è stata selezionata e il tasto di conferma viene premuto, il sistema salva la risposta, aggiorna lo stato relativo alle domande e mostra quindi quella successiva.
- Condizioni di uscita:  
L'utente risponde a tutte le domande dello Step in cui si trova. Il tasto per passare allo step successivo si attiva e l'utente preme questo per procedere.
- Eccezioni:  
Se viene premuto il tasto di invio senza aver selezionato nessuna risposta, il sistema restituisce un messaggio di errore in cui comunica all'utente di rispondere alla presente domanda prima di procedere. Se l'utente torna alla home le risposte fino a quel momento date vengono perse e viene chiesto all'utente di rispondere nuovamente.

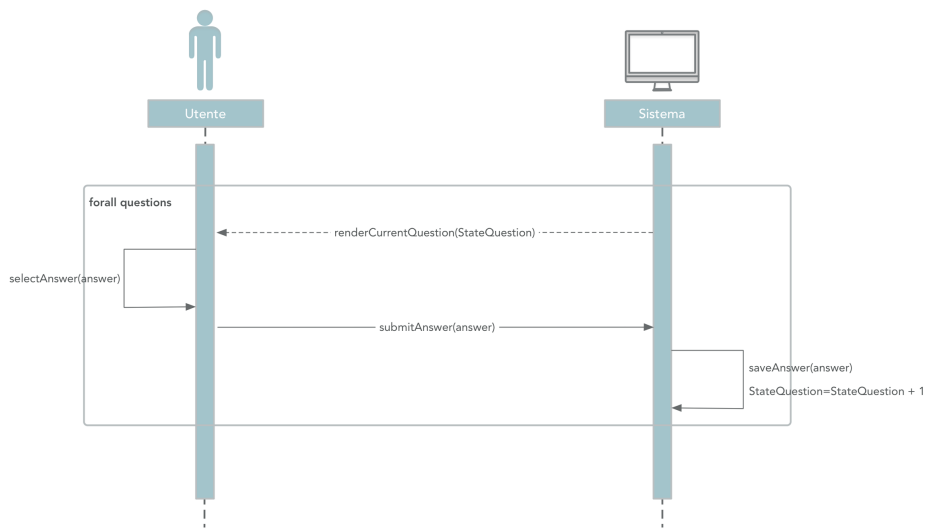


Figura 4.24: Sequence Diagram - Questionario

## Visualizzazione dei grafici

- Nome: Visualizzazione dei grafici
- Attori: Utente
- Condizioni di ingresso:  
Lo stato dello step è uguale a 1 e l'utente ha premuto sul pulsante dello step successivo.
- Flusso degli eventi:  
L'utente seleziona il livello più appropriato al rispetto alla conoscenza che egli ha della materia. Il sistema salva il livello selezionato nel proprio stato. Se il livello è 1 o 2 o 3 il sistema restituisce la pagina contenente lo Stacked Bar Chart, altrimenti il Circle Packing. Vengono poi effettuate le domande secondo il flusso descritto nel Sequence Diagram relativo al questionario. Finito questo l'utente seleziona lo step successivo. Il sistema aggiorna lo stato relativo agli step e mostra il grafico successivo secondo una logica inversa rispetto alla precedente. L'utente effettua nuovamente il questionario. Un volta concluso il sistema salva e invia tutte le risposte al server le quali vengono salvate in un database.
- Condizioni di uscita:  
L'utente ha risposto a tutte le domande e preme sul tasto corrispondente allo Step 4.
- Eccezioni:  
Se l'utente preme sul tasto relativo alla home prima dell'invio da parte del sistema al server le risposte date vengono perse e viene richiesto all'utente di ripetere la procedura appena descritta.



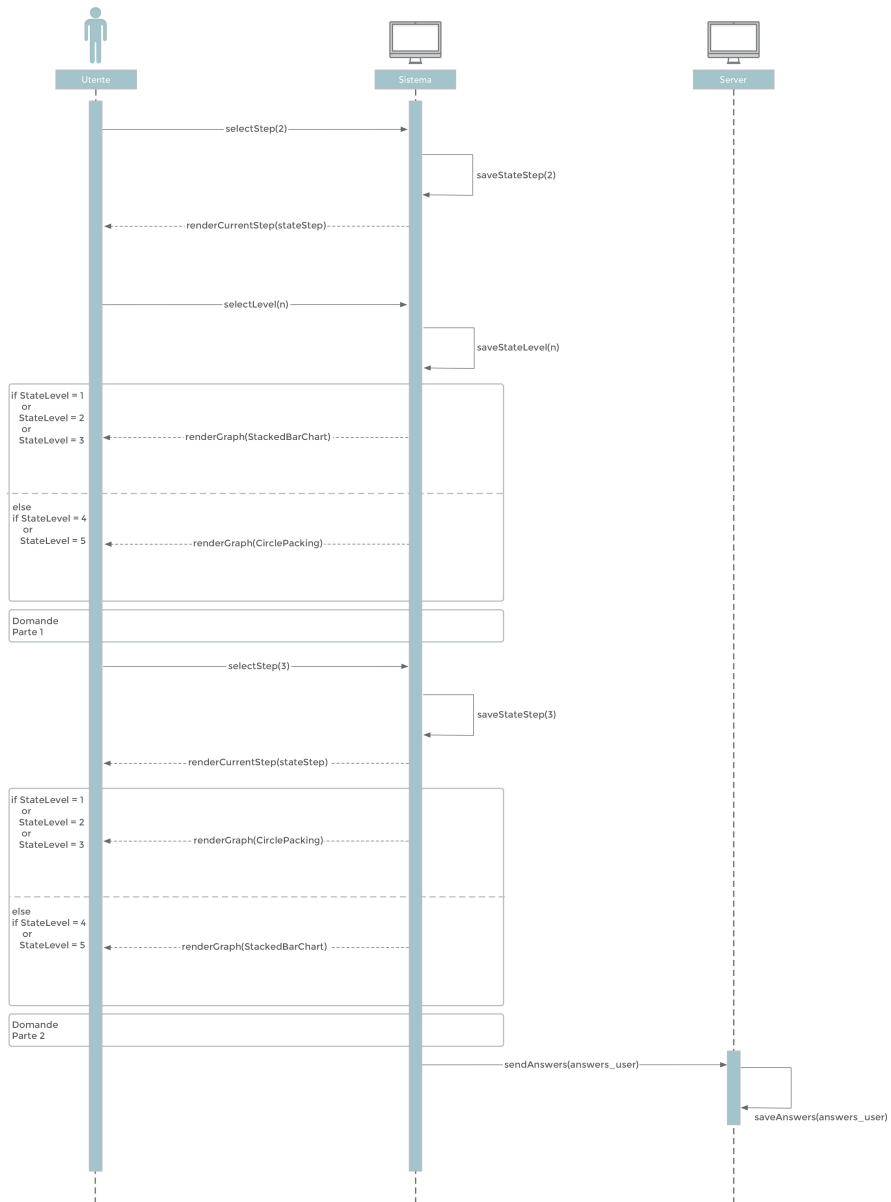


Figura 4.25: Sequence Diagram - Visualizzazione dei grafici

#### 4.4.5 Struttura delle domande

Di seguito sono riportate le domande effettuate all'interno del questionario:

Questionario Parte 1

1. Can you tell who is more popular on Instagram looking at the graph I have showed you?
2. Can you tell who is more popular at 18:00?
3. Do you think this graph is intuitive? (You don't need a lot of explanation to understand it)
4. Can you easily see and evaluate individual and total values? (The number of posts on a Social regarding a Chef and the sum of all posts on all Social of this Chef)

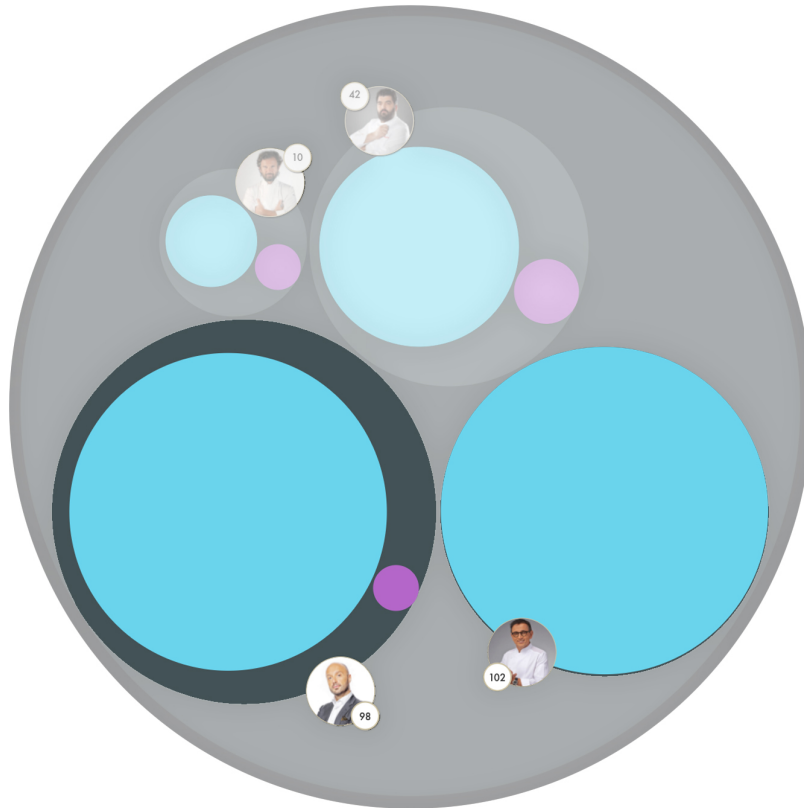
Questionario Parte 2

1. Can you tell who is more popular on Instagram looking at the graph I have showed you?
2. Can you tell who is more popular at 18:00?
3. Do you think is more intuitive this graph than the previous one? (You need less explanation to understand it)
4. Is this graph more readable than the previous one? (You can evaluate and see the individual and total values in a better way)

La prima domande in entrambi i questionari mira a valutare una delle problematiche che si incontrano quando si lavora con più dimensioni: analizzare le piccole quantità. Nel nostro caso specifico, chiediamo all'utente di capire chi è più popolare su Instagram poiché tale dimensione risulta essere di due ordini di grandezza inferiore rispetto al totale di ogni Chef. Nella seconda domanda invece andiamo a valutare quello che è un limite del Circle Packing. Tale visualizzazione converte un quantità nell'area del cerchio che la rappresenta e racchiude tutte le dimensioni in un unico cerchio corrispondente allo chef a cui fanno riferimento. Quando una di queste viene a mancare, inevitabilmente il cerchio esterno risulta di dimensioni inferiori, creando una distorsione grafica che può portare ad errori di lettura, così come mostrato nella Figura 4.26. In tale figura si può notare come lo chef che ha un numero totale di 102 posts è rappresentato con un cerchio di area inferiore rispetto allo chef cha ha un totale di 98 posts. In queste situazioni risulta quindi necessario l'utilizzo delle label. Nella terza e nella quarta

domanda invece andiamo a raccogliere e confrontare quelle che sono state le sensazioni degli utenti nei confronti delle visualizzazioni proposte.

I risultati ottenuti da questo sondaggio, così come quelli ottenuti dai precedenti, sono analizzati in dettaglio nel capitolo seguente.



*Figura 4.26: Distorsioni grafiche del Circle Packing*



## Capitolo 5

# Analisi dei Risultati e Valutazioni

*“Consider adopting novel design solutions only when the estimated payoff is substantially greater than the cost of learning to use them.”*

Colin WARE

In questo capitolo sono descritte le valutazioni e le analisi dei risultati ottenuti a seguito della nostra ricerca. Tali valutazioni sono state affrontate al termine di ogni indagine. Le analisi relative alle prime due sono state utilizzate per identificare le tecniche di visualizzazione più efficaci. Una volta individuate, queste sono state confrontate nella terza indagine, la quale ha permesso di raccogliere i dati per trarre le conclusioni e rispondere alle domande che ci siamo posti a monte della ricerca trattata in questa tesi.

## 5.1 Risultati della comparazione tra visualizzazioni tradizionali

Nella prima indagine le visualizzazioni messe a confronto sono quelle riportate nella Figura 5.1. I numeri assegnati ad ogni grafico saranno utilizzati come riferimento nei grafici relativi alle analisi dei risultati.

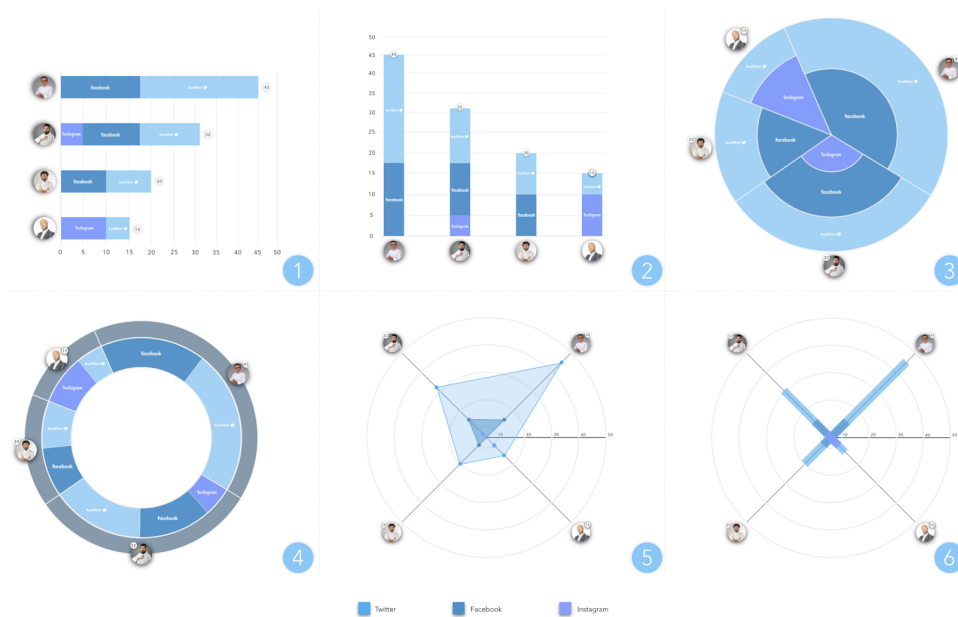


Figura 5.1: Grafici confrontati durante la comparazione tra visualizzazioni tradizionali

Il numero di intervistati è stato pari a 32, suddivisi demograficamente secondo la distribuzione mostrata in Figura 5.2. L'obiettivo dietro la raccolta di queste informazioni è quello di identificare una correlazione con le risposte riguardanti le domande successive. A seguito delle analisi dei dati però non è stata identificata nessuna correlazione.

Le risposte ottenute nel secondo step dell'indagine il quale mirava a valutare il livello di conoscenza degli intervistati e le aspettative di questi (vedi sezione 4.3.1) si sono dimostrate destrutturate e non analizzabili, quindi non utilizzabili per effettuare studi rilevanti.

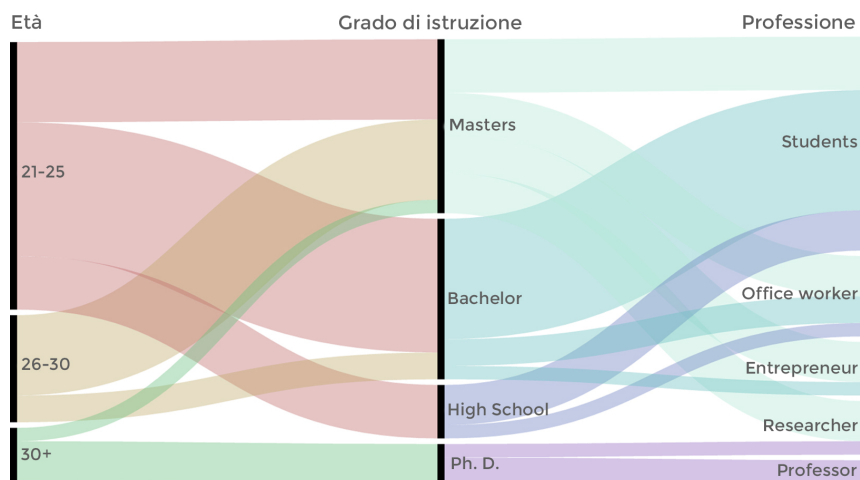
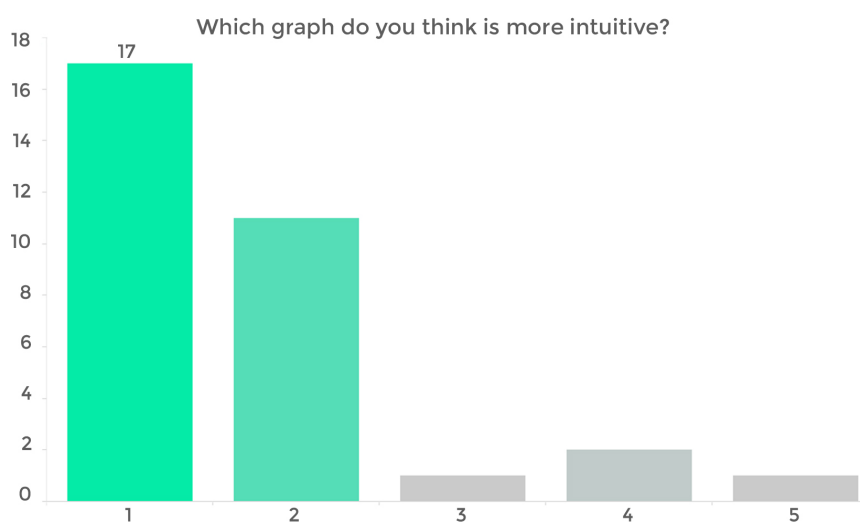


Figura 5.2: Distribuzione Demografica

La fase relativa alle reazioni degli intervistati invece ha permesso di effettuare una selezione preliminare delle visualizzazioni. Analizzando i grafici mostrati in Figura 5.3, dove sull'asse X sono riportati i numeri corrispondenti ai grafici, si nota come lo Stacked Bar Chart è risultato essere il grafico più intuitivo e con un maggiore appeal. Il Radar Chart e il Radial Column Chart invece, oltre ad essere risultati i meno intuitivi in fase di lettura, sono stati valutati anche come grafici con il minore appeal.



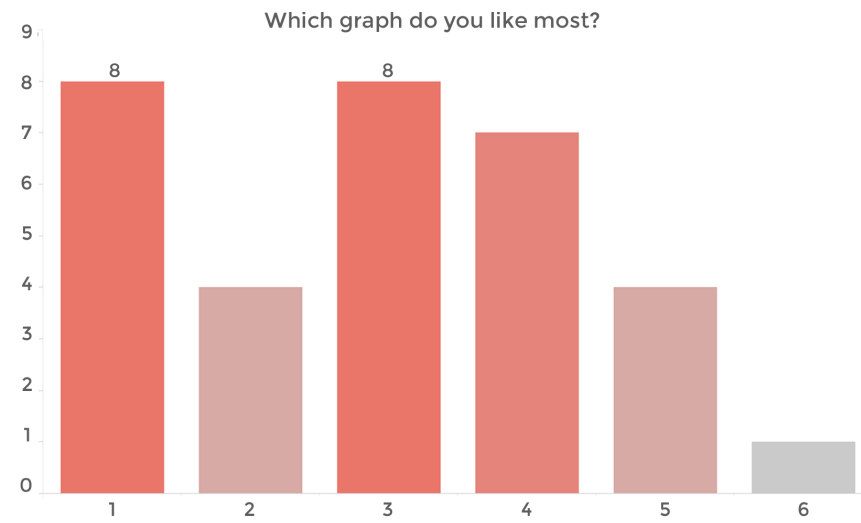
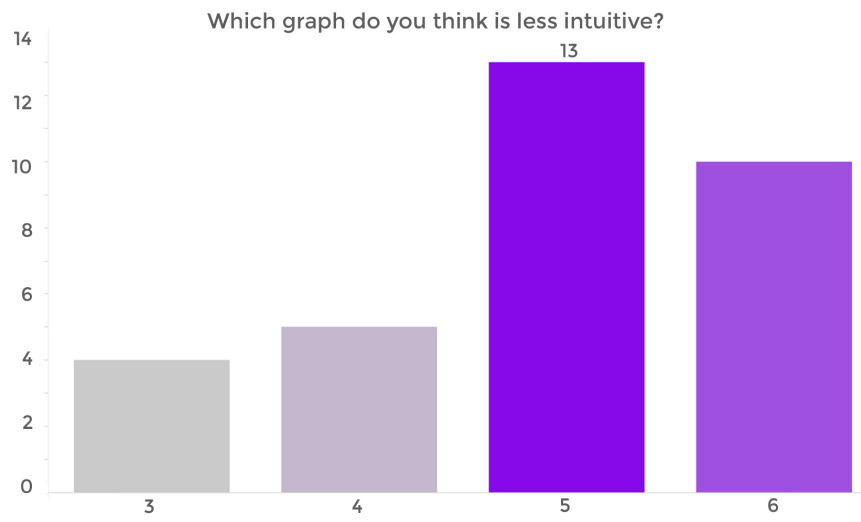


Figura 5.3: Risultati relativi alle reazioni degli intervistati - Comparazione tra visualizzazioni tradizionali



Per quanto concerne il confronto tra coppie di grafici i risultati ottenuti sono mostrati in Figura 5.4. I numeri indicati nella legenda rappresentano il grafico considerato, con riferimento alla Figura 5.1. Da questi confronti si nota come lo Stacked Bar Chart sia stato molto più apprezzato rispetto all'Istogramma. Questo risultato era previsto poiché l'Istogramma risulta più indicato per mostrare andamenti o distribuzioni dei dati, mentre lo Stacked Bar Chart permettere di analizzare e confrontare in maniera più rapida i dati quantitativi. Come evidenziato sempre in Figura 5.4 non sono state riscontrate significative differenze tra le altre coppie di grafici. Tale risultato è in linea con le nostre aspettative, in quanto, le visualizzazioni coinvolte nel confronto, condividono fortemente la struttura e il metodologia con la quale vengono visualizzati i dati. Nel successivo step, nella quale sono state

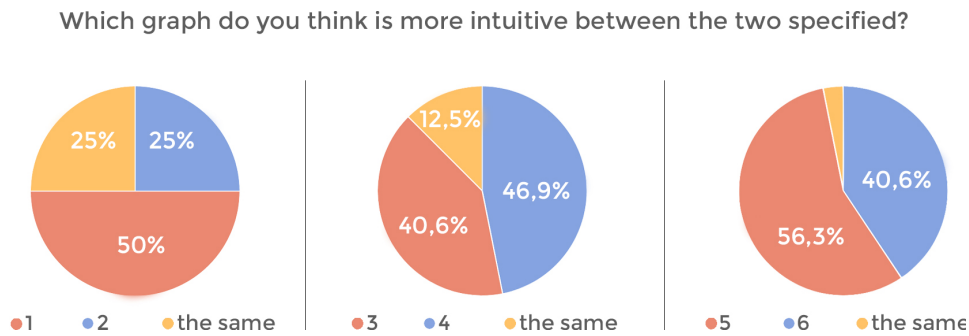


Figura 5.4: Risultati ottenuti dal confronto tra grafici simili - Comparazione tra visualizzazioni tradizionali

effettuate domande di controllo, si sono evidenziati risultati inaspettati. In questa fase l'obiettivo era quello di far scegliere all'utente la visualizzazione più adatta per estrapolare una specifica informazione, e sulla base di tale scelta verificare se fosse in grado di leggerla correttamente. Le scelte degli utenti sono mostrate in figura 5.5. I numeri sull'asse Y sono relativi ai grafici coinvolti nell'analisi. Come si può notare, il grafico 1, corrispondente allo Stacked Bar Chart, e il 4, corrispondente al Donut Chart, sono risultati essere i grafici scelti maggiormente. Queste scelte sono state poi verificate con le domande di controllo. Le verifiche effettuate per la prima domanda mostrata in figura 5.5 hanno avuto un riscontro positivo con una percentuale del 91%, mentre quella relative alla seconda domanda del 97%. Gli utenti che hanno selezionato il grafico 1 o 4 hanno risposto correttamente alle domande di verifica con una percentuale del 100%. Da questo step si è evidenziato come il Donut Chart, il quale non era emerso in maniera significativa du-

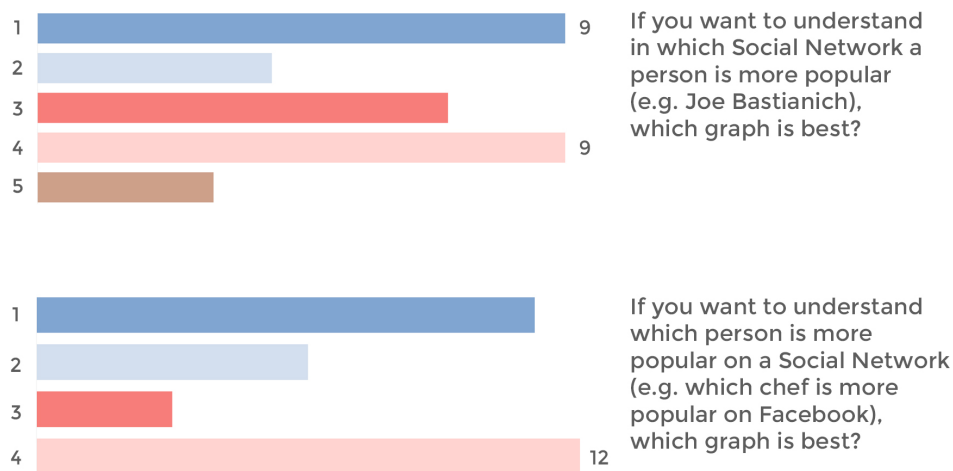


Figura 5.5: Risultati delle domande di controllo - Comparazione tra visualizzazioni tradizionali

rante le precedenti analisi, sia risultato essere il grafico maggiormente scelto nel momento in cui viene chiesto all'utente di estrapolare informazioni.

Nell'ultima parte dell'indagine abbiamo raccolto i feedback degli utenti. Il tema principalmente criticato dagli intervistati è stato l'utilizzo dei colori, tema sottovalutato ma molto sentito. Quello che è emerso è che per facilitare la lettura dei grafici, gli utenti si aspettano di vedere associato ad un ogni Social il colore rappresentativo di questo. Tale feedback è stato accolto positivamente e applicato nell'indagine successiva.

## 5.2 Risultati della comparazione con visualizzazioni innovative

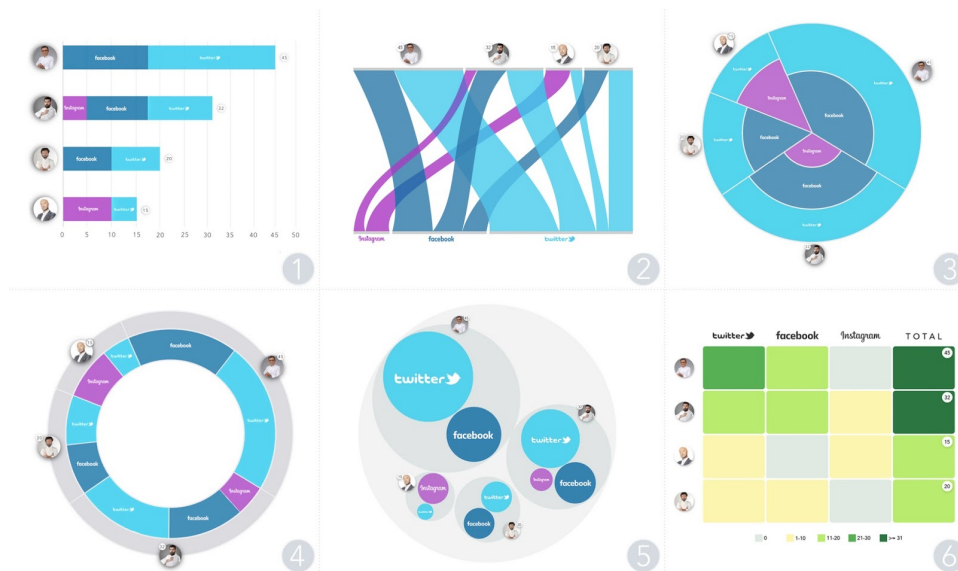


Figura 5.6: Grafici confrontati durante la comparazione con visualizzazioni innovative

Sulla base dei feedback e delle valutazioni effettuate nel precedente step, abbiamo ridisegnato e confrontato i grafici mostrati in figura 5.6. Come per la precedente analisi i numeri assegnati ad ogni grafico saranno utilizzati come riferimento nei grafici relativi alle analisi dei risultati.

Il numero di intervistati è stato di 97, di cui 33 relativi alle interviste effettuate durante la SMAU, mentre i restanti 64 sono stati raggiunti attraverso i Social.

La prima analisi effettuata riguarda la distribuzione del livello di conoscenza degli intervistati nell'ambito sia della Data Visualization che della Data Analysis. Come mostrato in figura è possibile vedere che la maggior parte degli utenti si è posizionata tra i livelli 2 e 3. Queste informazioni, così come nella prima indagine, sono state raccolte per verificare l'esistenza di una qualche correlazione con la capacità di lettura delle informazioni. Anche questa volta però, non state evidenziate correlazioni significative.

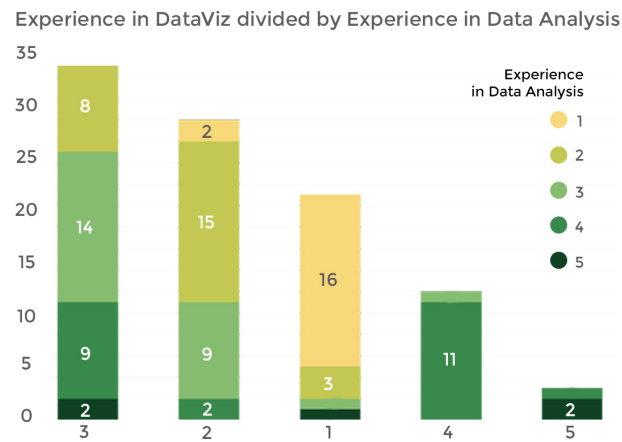
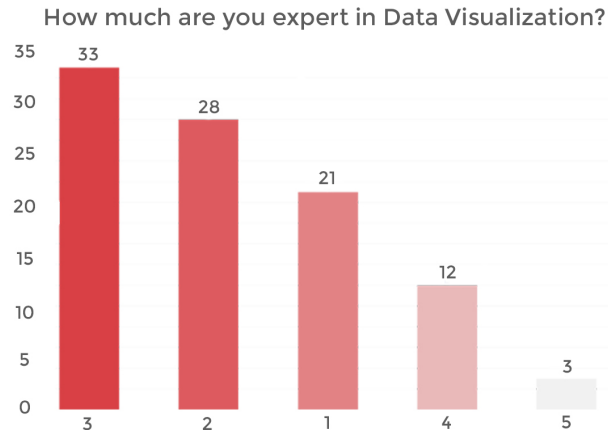
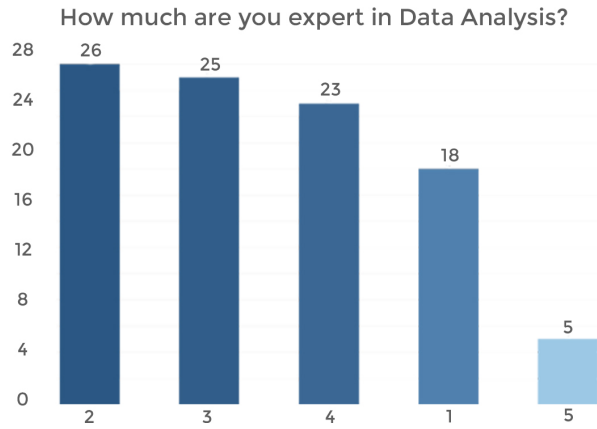


Figura 5.7: Livello di conoscenza degli intervistati - Comparazione con visualizzazioni innovative

Nel grafico rappresentato in figura 5.8 invece sono mostrati i nomi dei grafici conosciuti dagli intervistati suddivisi per livello di conoscenza nell'ambito della Data Visualization.

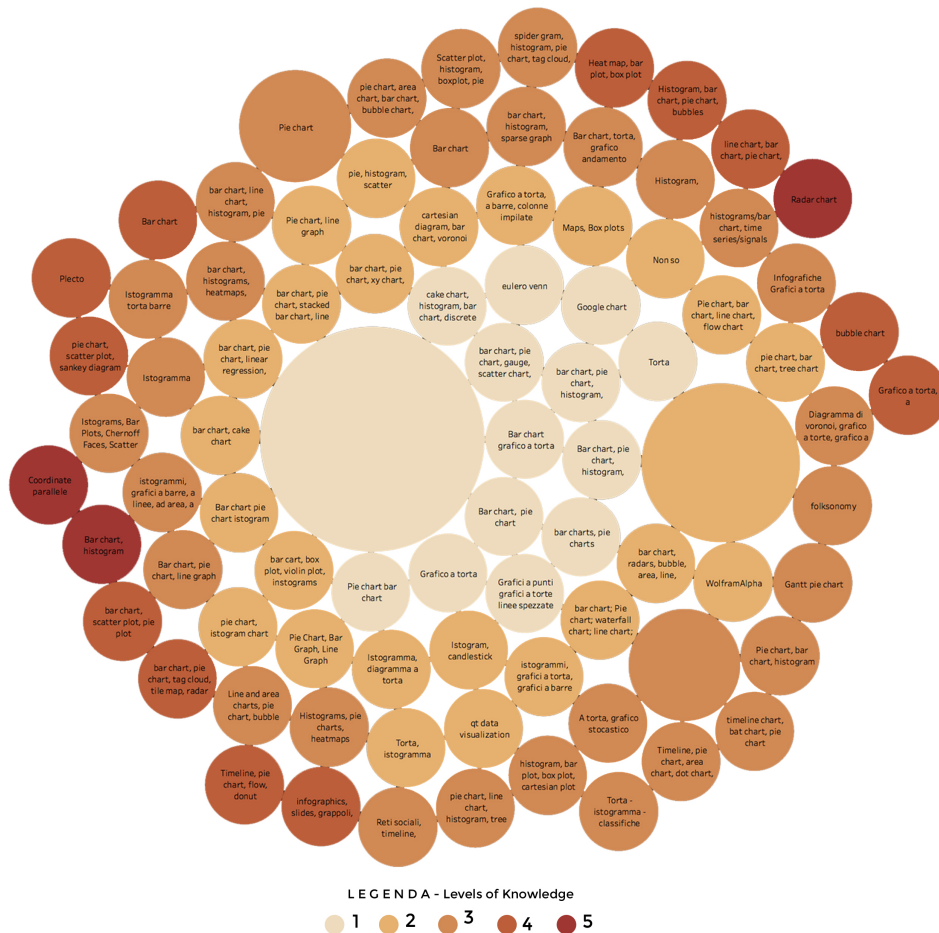


Figura 5.8: Grafici conosciuti dagli intervistati suddivisi per livello di conoscenza - Comparazione con visualizzazioni innovative

Successivamente abbiamo analizzato cosa l'utente si aspetta di vedere se dovesse essergli mostrato un grafico che rappresenta la popolarità. Il 54% degli intervistati ha risposto con Bar Chart, il 14% il Pie Chart e il 18% non ha risposto, mentre il rimanente 14% ha risposto con altro. Questo risulta essere un indicatore molto importante, in quanto conferma che, indipendentemente dal livello di conoscenza, la maggior parte degli utenti si aspetta di vedere un Bar Chart, il quale è il grafico storicamente più utilizzato per mostrare e confrontare informazioni quantitative.

Difronte allo Stacked Bar Chart senza informazioni aggiuntive, la maggior parte degli utenti ha capito che mostrava la popolarità dei giudici. Solo una piccola percentuale ha anche capito che si trattava del numero dei posts in cui questi erano citati.

Rispetto allo Stacked Bar Chart, in Figura 5.9 viene evidenziato se di fronte ad altri grafici gli utenti riescono ad estrapolare le stesse informazioni.

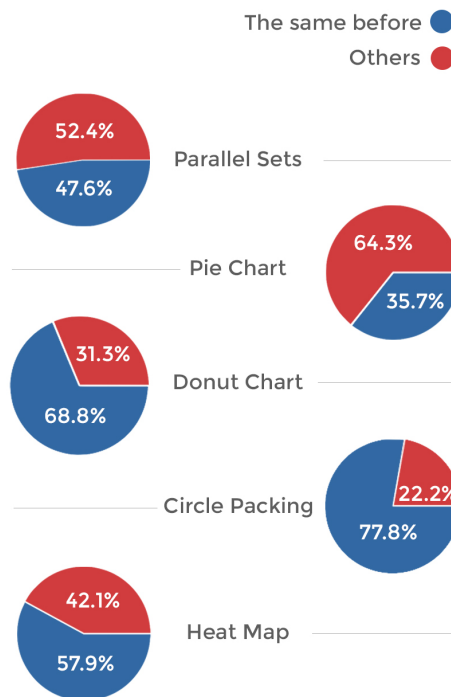


Figura 5.9: Cosa gli intervistati comprendono dal grafico mostrato rispetto allo Stacked Bar Chart - Comparazione con visualizzazioni innovative

Quello che emerge è che il grafico dai cui gli utenti riescono a intuire maggiormente la stessa tipologia di informazioni è il Circle Packing.

Lo step successivo è stato analizzare quelli che secondo gli utenti sono i grafici più intuitivi e meno intuitivi, in modo da valutare quali di questi diano una visione chiara dell'informazione che vogliamo mostrare. Da questa analisi, i cui risultati sono mostrati in Figura 5.10, il grafico che è risultato essere il più intuitivo è stato ancora una volta lo Stacked Bar Chart. Un risultato inaspettato però è stato quello riguardante il Circle Packing, il quale ha raccolto il 15% delle preferenze, posizionandosi sopra il Donut Chart e il Pie Chart. I grafici risultati essere meno intuitivi invece sono

stati il Parallel Sets e la Heat Map, raccogliendo rispettivamente il 55% e il 37% dei consensi.

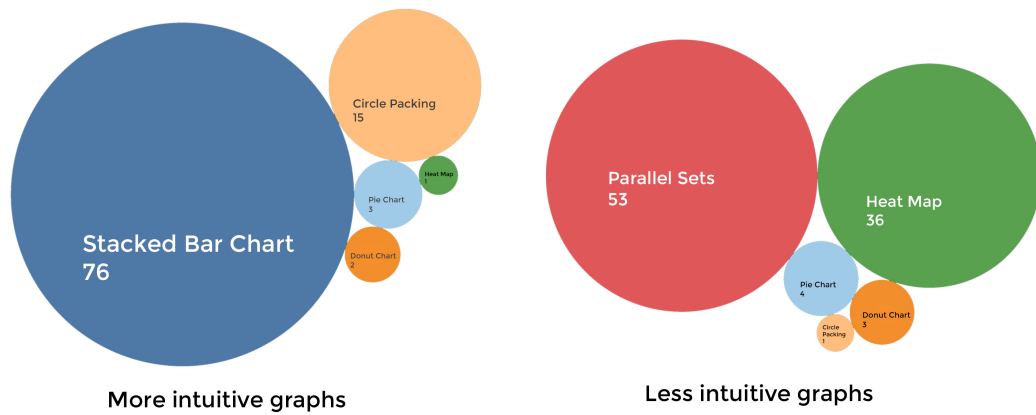


Figura 5.10: Risultati relativi ai grafici intuitivi e non intuitivi - Comparazione con visualizzazioni innovative

Per quanto riguarda invece l'appeal delle visualizzazioni, con nostra sorpresa il Circle Packing è risultato essere il migliore. Oltre ad essere gradevole per l'utente, il Circle Packing ha anche riscontrato nel 77,8% degli utenti la stessa efficacia di lettura dello Stacked Bar Chart.

#### The most ❤️ graph

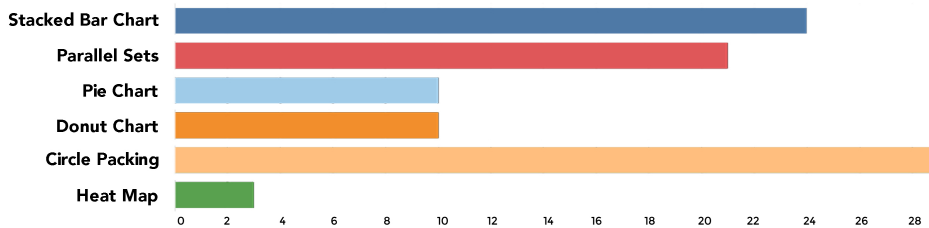


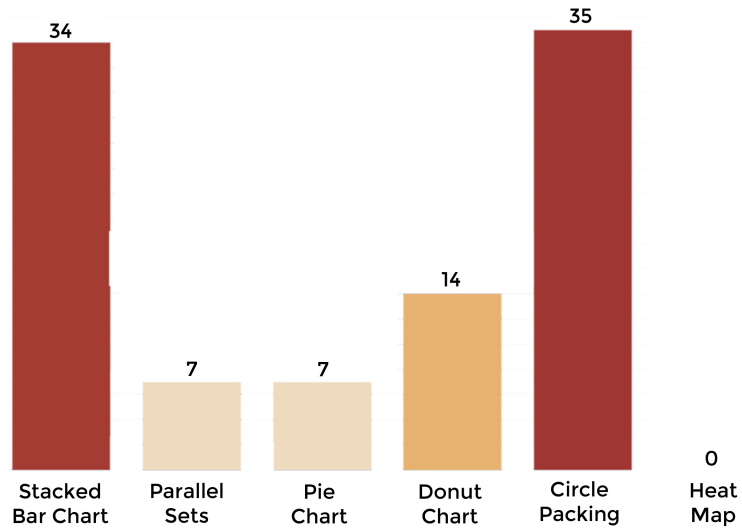
Figura 5.11: Risultati relativi all'appeal dei grafici - Comparazione con visualizzazioni innovative

Come nella precedente indagine, successivamente abbiamo fatto scegliere all'utente la visualizzazione più adatta per estrapolare una specifica informazione, e sulla base di tale scelta abbiamo verificato se fosse in grado di leggerla correttamente. Le scelte degli utenti sono mostrate in Figura 5.12. Come si può notare, lo Stacked Bar Chart e il Circle packing, sono risultati

essere ancora una volta i grafici maggiormente scelti. Queste scelte sono state poi verificate con le domande di controllo. Nelle verifiche effettuate per la prima domanda mostrata in figura 5.12 lo stacked Bar Chart ha avuto un riscontro positivo del 92% mentre il Circle Packing del 100%. Per quanto riguarda la seconda domanda invece lo Stacked Bar Chart ha avuto un riscontro positivo del 98% mentre il Circle Packing del 93%.



We want to know in wich Social Network a person is more popular. Wich graph is best?



We want to know wich person is more popular on a Social Network. Wich graph is best?

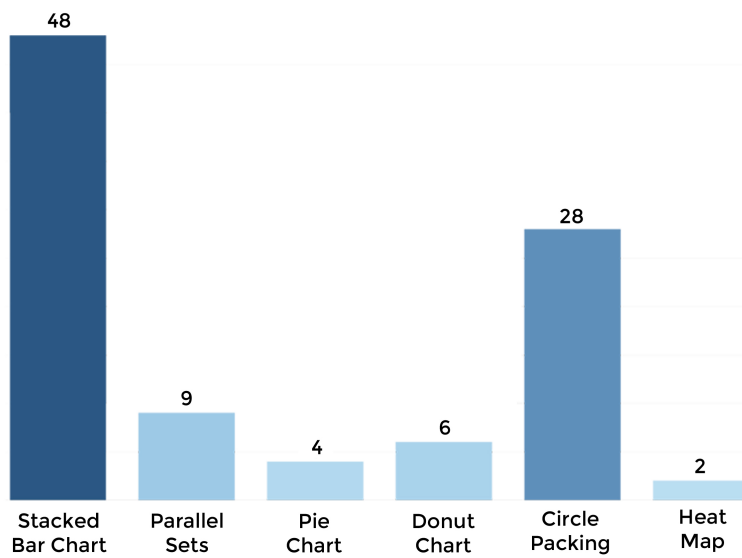


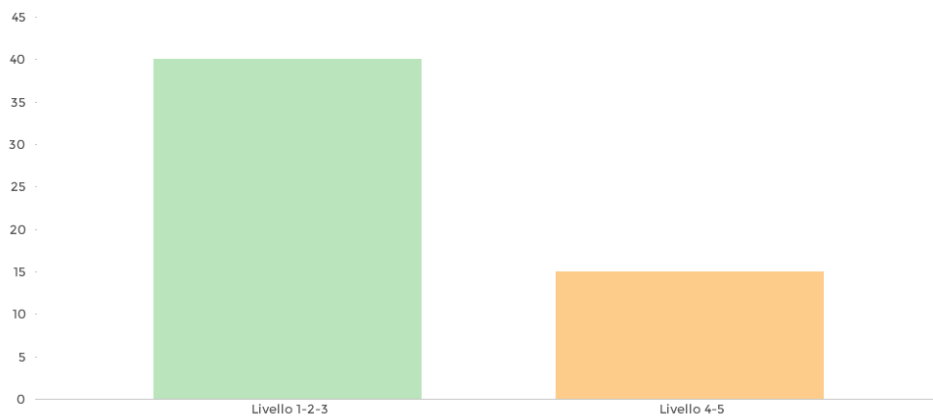
Figura 5.12: Risultati delle domande di controllo - Comparazione con visualizzazioni innovative

A seguito delle prime due indagini possiamo quindi trarre le prime conclusioni. L'analisi dei risultati ha confermato che lo Stacked Bar Chart è il grafico più efficace per mostrare e confrontare i dati del nostro caso studio. Questo risultato non è sorprendente in quanto tale tecnica tradizionale è la più utilizzata fino ad oggi e quindi anche la più conosciuta. Risulta quindi comprensibile che di fronte alla scelta di differenti grafici, gran parte degli utenti abbia estrapolato le informazioni dalla visualizzazione più familiare. Ad ogni modo, il Circle Packing si è rivelata essere un'ottima alternativa. Questo perché è risultata essere più gradevole e quindi in grado di attirare maggiormente l'attenzione, mantenendo un ottimo livello di efficacia di lettura e mostrando tutte le informazioni del nostro caso studio in maniera chiara e rapida.

Nei risultati della terza ed ultima indagine, andremo ad analizzare i feedback ottenuti dal confronto tra lo Stacked Bar Chart e il Circle Packing, valutandone i pregi e difetti, con l'obiettivo di verificare se il nostro candidato sia in grado di sostituire completamente la tecnica tradizionale, lo Stacked Bar Chart.

### 5.3 Risultati della comparazione tra le visualizzazioni migliori

In questa fase conclusiva della nostra ricerca abbiamo raccolto 55 feedback. Sotto l'ipotesi di una correlazione tra livello di conoscenza e grafico più adatto, la prima distinzione che siamo andati ad analizzare riguarda proprio il livello di conoscenza degli utenti contattati. In Figura 5.13 viene mostrato come sono distribuiti gli intervistati coinvolti in questa indagine. Si nota uno sbilanciamento verso i livelli uguale a 1, 2 o 3.



*Figura 5.13: Distribuzione degli intervistati in base a livello di conoscenza nella Data Visualization*

Le analisi seguenti sono state suddivise in secondo i 2 step presenti nell'applicazione web tramite la quale abbiamo raccolto i feedback. Distinguiamo quindi tra i risultati ottenuti dalla prima visualizzazione proposta e la seconda visualizzazioni proposta

### 5.3.1 Risultati prima visualizzazione proposta

In questa fase se l'utente ha selezionato un livello pari a 1, 2 o 3 si è trovato di fronte allo Stacked Bar Chart, se invece ha selezionato 4 o 5 gli è stato mostrato il Circle Packing. Le domande in entrambe i casi sono le medesime. La prima domanda effettuata è stata “Chi è lo Chef più popolare su Instagram?”, la cui risposta corretta era “Joe Bastianich”. I risultati ottenuti sono mostrati in Figura 5.14. Analizzando le risposte date dagli intervistati, si nota come il Circle Packing abbia ottenuto una percentuale maggiore di risposte esatte. Questo risultato è conforme alle nostre aspettative, in quanto per rispondere a questa domanda, era necessario analizzare dei valori che avevo un ordine di grandezza molto inferiore rispetto al totale. Tale risultato conferma quindi che in queste situazioni lo Stacked Bar Chart risulta di difficile lettura, preferendo a questo il Circle Packing, il quale permette di mettere in evidenza anche dati che hanno valori con ordini di grandezza inferiori rispetto al totale.

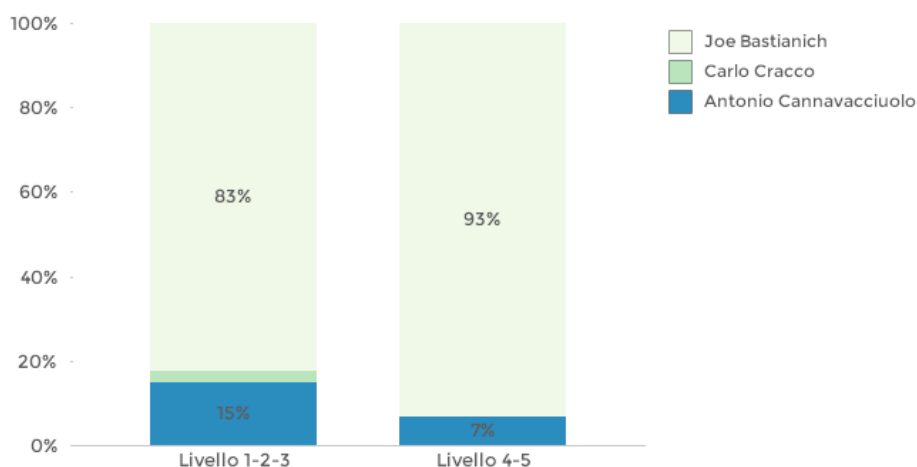


Figura 5.14: Chi è lo Chef più popolare su Instagram? - Prima visualizzazione proposta

La seconda domanda effettuata è stata “Chi è lo Chef più popolare alle 18:00?” Con questa domanda l’obiettivo era analizzare il difetto relativo al Circle Packing descritto alla fine della sezione 4.4.5 del Capitolo 4. Anche in questa situazione il risultato ottenuto equivale alle nostre aspettative. La risposta corretta a questa domanda era “Bruno Barbieri” e come ci aspettavamo tale informazione è stata letta con facilità attraverso l’utilizzo dello Stacked Bar Chart, mentre nel Circle Packing solo il 70% degli intervistati è stato in grado di estrapolare l’informazione corretta, contro il 94% dello

Stacked Bar Chart. Tale risultato è illustrato in Figura 5.15 La terza e la

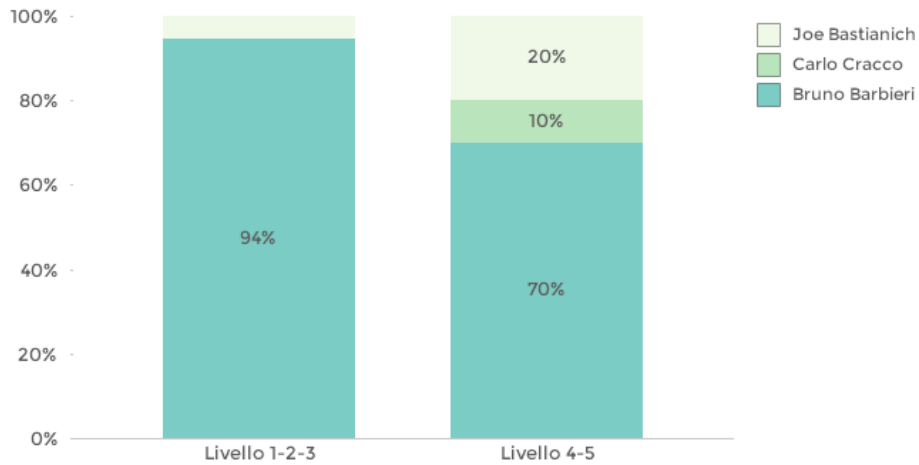


Figura 5.15: Chi è lo Chef più popolare alle 18:00? - Prima visualizzazione proposta

quarta domanda vanno ad analizzare le sensazioni degli utenti. In Figura 5.16 e 5.17 sono rappresentati risultati. Quello che emerge è che la maggiore parte degli intervistati che si sono trovati di fronte allo Stacked Bar Chart, ovvero gli intervistati aventi livello 1, 2 o 3, lo hanno trovato intuitivo. Una percentuale degli stessi utenti però non ha trovato questo grafico di facile lettura sui singoli valori, confermando quello che era emerso dai dati estratti nella prima domanda. Per gli utenti che si sono trovati di fronte al Circle Packing, ovvero gli intervistati aventi livello 4 o 5, invece non è emersa nessuna preferenza significativa, in quanto circa il 50% degli intervistati ha risposto sì, e il rimanente no, ad entrambe le domande.

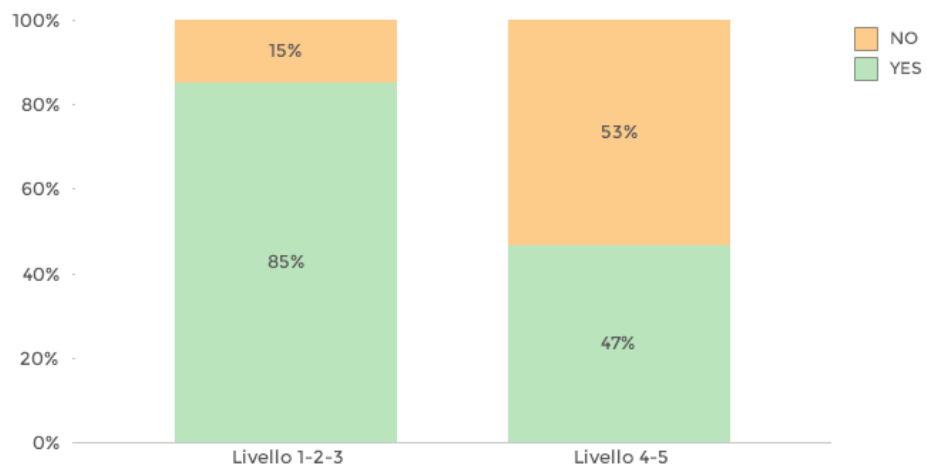


Figura 5.16: Credi che il grafico che ti ho mostrato sia intuitivo? - Prima visualizzazione proposta

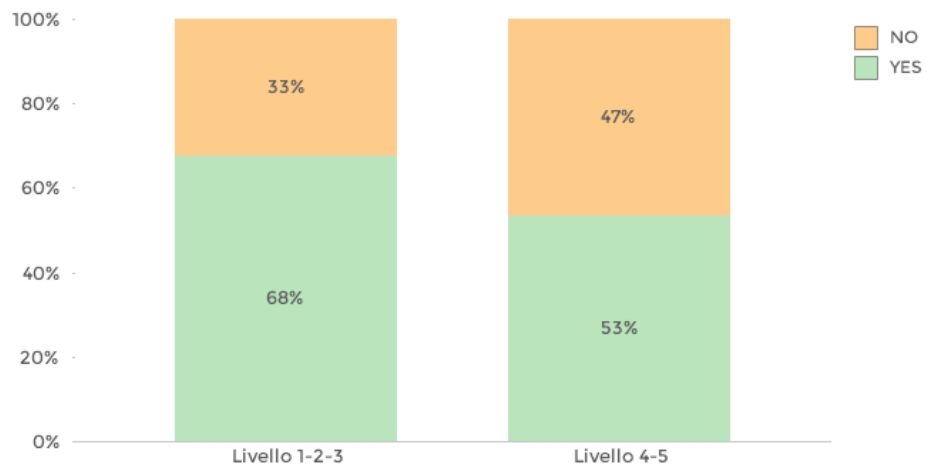


Figura 5.17: Sei in grado di identificare con facilità i valori singoli e totali? - Prima visualizzazione proposta

### 5.3.2 Risultati seconda visualizzazione proposta

La differenza rispetto alla parte precedente è nel grafico mostrato agli utenti. Se l'utente aveva selezionato un livello pari a 1, 2 o 3 arrivati a questo punto si sarà trovato di fronte allo Circle Packing, se invece aveva selezionato 4 o 5 gli è stato mostrato lo Stacked Bar Chart. Le domande in entrambe i casi sono le medesime.

La prima domanda effettuata è “Chi è lo Chef più popolare su Instagram?”, la cui risposta corretta era “Joe Bastianich”. I risultati sono mostrati in Figura 5.18. Analizzando le risposte date dagli intervistati e confrontandole con quelle date nella prima domanda della precedente parte, si nota che gli utenti di livello 4-5, i quali ora stanno visualizzando lo Stacked Bar Chart, hanno dato una percentuale inferiore di risposte corrette rispetto alla medesima domanda nella fase precedente, nella quale erano di fronte al Circle Packing. Questo risultato è una conferma ulteriore del fatto che lo Stacked Bar Chart risulta di difficile lettura in presenza di valori di ordini di grandezza inferiori rispetto al totale.

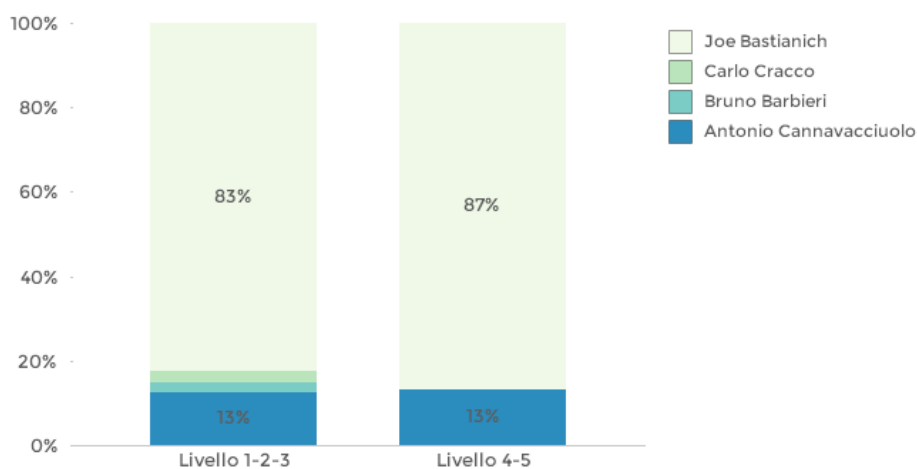
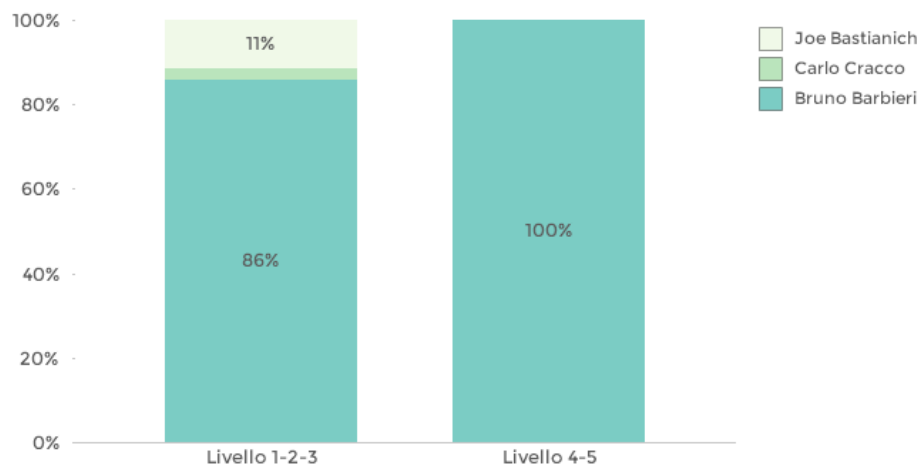


Figura 5.18: Chi è lo Chef più popolare su Instagram? - Seconda visualizzazione proposta

Nella seconda domanda, “Chi è lo Chef più popolare alle 18:00?”, i cui risultati sono mostrati in Figura 5.19 l'86% degli utenti di livello 1, 2 o 3, i quali erano di fronte al Circle Packing, hanno risposto correttamente con una percentuale maggiore rispetto agli utenti di livello 4 o 5 durante la fase precedente.



*Figura 5.19: Chi è lo Chef più popolare alle 18:00? - Seconda visualizzazione proposta*

La terza e la quarta domanda di questa parte ha l'obiettivo di confrontare le sensazioni degli utenti riguardo al grafico appena analizzato rispetto al quello che hanno visualizzato nella precedente fase. In figura 5.20 e 5.21 sono rappresentati risultati. Quello che emerge è che la maggiore parte degli intervistati ha ritenuto lo Stacked Bar Chart migliore su entrambe le metriche di valutazione.



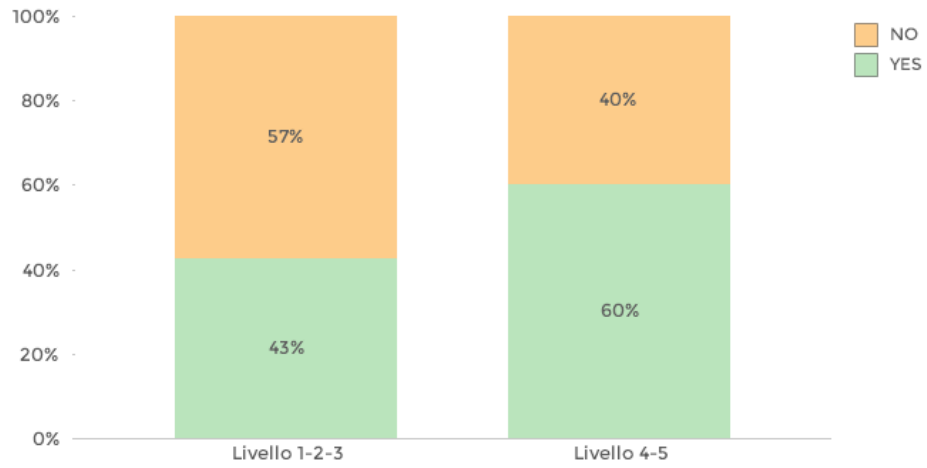


Figura 5.20: Credi che il grafico che ti ho mostrato sia più intuitivo rispetto al precedente? - Seconda visualizzazione proposta

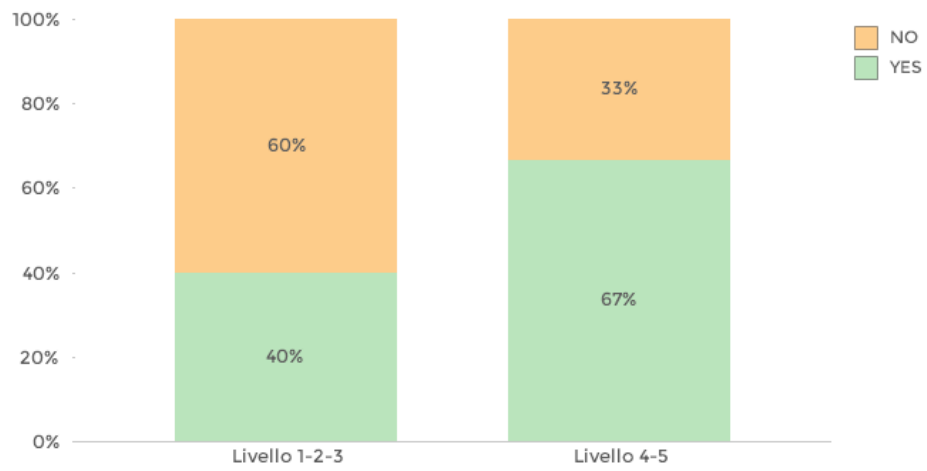


Figura 5.21: Sei in grado di identificare i valori singoli e totali con una maggiore facilità rispetto al grafico che ti ho mostrato in precedenza? - Seconda visualizzazione proposta



## Capitolo 6

# Conclusioni e Direzioni Future

*“Above all else show the data.”*

Edward R. TUFTE

Alla base del lavoro di ricerca svolto e descritto in questa tesi, c'è la volontà di dare un contributo nell'ambito della Data Visualization, focalizzandoci sulle problematiche che affliggono l'individuazione delle tecniche adatte a mostrare i dati e sensibilizzando l'importanza che questa disciplina ricopre nello studio e nella analisi dei dati. L'obiettivo principale è quello di rispondere alle domande definite nel Capitolo 1: *“le tassonomie e i sistemi automatici di individuazioni dei grafici oggi esistenti, sono ancora strumenti adatti per identificare quale sia la tecnica migliore da utilizzare?”* e *“siamo sicuri che le tecniche tradizionali, utilizzate fino ad oggi, siano ancora le più appropriate?”* mettendo così in discussione le tecniche tradizionali, e dimostrando che tecniche innovative, fino ad oggi poco utilizzate, sono capaci di mostrare i dati in maniera corretta attraverso rappresentazioni più efficaci. In questo capitolo finale sono quindi descritte quelle che sono le conclusioni che abbiamo tratto a valle di questa ricerca, le osservazioni e i limiti relativi al metodo proposto e i possibili sviluppi futuri legati all'indagine effettuata.

## 6.1 Conclusioni

Il lavoro svolto verte essenzialmente su due fasi: una prima in cui ci siamo concentrati sulla esplorazione del dominio applicativo, ricercando le informazioni relative all'utilizzo delle visualizzazioni in contesti differenti, e di come queste siano catalogate e identificate dai metodi automatici. A questa segue una seconda, più pratica, costituita da un'esperienza implementativa volta alla realizzazione di un'indagine, con l'obiettivo di creare una metodologia valida per confrontare e identificare le tecniche di *Data Visualization* più adatte a seconda del contesto e del dominio di interesse relativo alla analisi.

La fase esplorativa ha permesso di scoprire conoscenze determinanti per la realizzazione delle indagini. In primo luogo è stato possibile identificare le visualizzazioni, sia tradizionali che innovative, adatte a mostrare le informazioni del caso studio preso in esame. E' stato necessario poi progettare queste visualizzazioni, e la definizione e l'identificazione di linee guida sia dal punto di vista grafico che della percezione dell'uomo, sono state essenziali per avere la certezza di mostrare i dati in maniera corretta. Questa fase esplorativa ha permesso anche di identificare e confermare le lacune alla base della nostra domanda di ricerca, riscontrando un mancato studio strettamente legato alla percezione dei dati e al confronto delle visualizzazioni tradizionali con quelle recentemente introdotte.

L'insieme di conoscenze così acquisite ci ha portato, nella fase implementativa, a realizzare le visualizzazioni descritte nei capitoli precedenti, con le quali rispondere alle domande poste a monte della ricerca.

A seguito dell'indagine si è provato che alcune delle nuove tecniche di visualizzazione possono essere considerate ottime varianti di quelle generalmente utilizzate, permettendo in alcune situazioni, di offrire una efficacia di lettura maggiore. I risultati ottenuti infatti, hanno dimostrato come il Circle Packing, il quale viene generalmente utilizzato per mostrare strutture gerarchiche, sia risultato essere efficace nel mostrare dati quantitativi multidimensionali tanto quanto lo è Stacked Bar Chart. Oltre ad avere un maggiore appeal rispetto a quest'ultimo, il Circle Packing è risultato essere più efficace nella lettura di dati parziali aventi un ordine di grandezza inferiore rispetto al totale mostrato.

In conclusione, alla luce dei risultati registrati in fase di valutazione delle visualizzazioni, possiamo affermare che le domande alla base di questa tesi hanno avuto un riscontro positivo. Possiamo ritenerci soddisfatti del lavoro svolto e della metodologia adottata durante tutto il lavoro, la quale può inserirsi in un contesto più ampio e contribuire allo studio di ulteriori tecniche di visualizzazione dati. Vale la pena sottolineare che la ricerca nel

mondo della *Data Visualization* rappresenta oggi un'area in grande espansione in virtù dell'importanza sempre maggiore che questi strumenti stanno assumendo nel settore Business e accademico. Quanto condotto rappresenta solo un nuovo punto di partenza che può essere approfondito in diverse direzioni e portare a scoprire nuove forme di conoscenza legate al mondo delle visualizzazioni, esattamente come, durante questo nostro studio, sono state identificate tecniche fino ad ora non utilizzate per mostrare determinate informazioni.

## 6.2 Limiti

Come ogni lavoro di ricerca, lo studio che abbiamo effettuato è soggetto a limiti e imperfezioni. In primo luogo il confronto di grafici innovativi con grafici tradizionali è inevitabilmente soggetto a effetti bias dovuti ad un maggiore utilizzo di determinate visualizzazioni, le quali risultano essere più facili da interpretare anche a causa di una maggiore confidenza e familiarità che l'intervistato ha con essa. Tale osservazione è stata riscontrata anche nella nostra indagine, in quanto, nonostante il Circle Packing abbia permesso di identificare determinate informazioni con una maggiore efficacia, questo è stato valutato dagli intervistati meno intuitivo rispetto allo Stacked Bar Chart. Non sappiamo quanto la familiarità degli utenti con un determinato grafico possa aver influito sui dati raccolti, ma è bene tener presente questo fattore.

Un ulteriore limite relativo a questa ricerca riguarda la scalabilità del metodo user-oriented proposto. Infatti una ricerca su larga scala effettuata secondo la procedura utilizzata, richiederebbe uno sforzo significativo. In primo luogo sarebbe necessario identificare un gran numero di casi studio, così da poter confrontare le visualizzazioni in un numero di contesti tale da censire ogni possibilità di analisi; questo comporta anche avere la possibilità di accedere a dati che permettano tale varietà. Secondo, per ognuna di queste analisi, è necessario raggiungere un quantità di intervistati tale da avere un campione rappresentativo significativo. Questa ultima problematica può risultare quasi più complicata della prima poiché difficilmente una persona è disposta ad effettuare indagini se non interessata all'argomento.

### 6.3 Direzioni future

Il principale sviluppo futuro relativo all'analisi fin qui condotta consiste nell'allargare ulteriormente lo spettro dei dati e delle informazioni considerate, anche se questo può risultare dispendioso per le motivazioni indicate nel precedente paragrafo. Ulteriori studi però possono permettere di identificare metodologie e tecniche fino ad oggi non considerate, così come accaduto durante la nostra indagine.

Per quanto riguarda l'ambito applicativo, invece, lo sviluppo futuro indubbiamente più vicino a questa ricerca risiede nell'aiuto che questa può fornire durante la definizione delle tassonomie e dei sistemi automatici di selezione delle visualizzazioni. Tali sistemi troppo spesso non considerano le nuove tecniche oggi esistenti, e identificare una metodologia che permetta di validarle e di verificare se queste possano essere migliori di quelle tradizionali, può contribuire alla costruzione di strumenti più efficienti ed efficaci.

Un'altro elemento che sarebbe interessante approfondire riguarda la correlazione tra l'utente e il grafico. Se fosse possibile identificare fattori che possano condurre all'individuazione di una correlazione tra le caratteristiche dell'utente e la visualizzazione dalla quale riesce ad estrapolare informazioni in maniera efficace, sarebbe possibile utilizzare e ampliare ulteriormente la metodologia proposta. Dai dati da noi raccolti non è stato possibile identificare risultati significativi relativi a questo aspetto, ma solo delle piccole tendenze. Crediamo però che, nonostante la percezione di forme e colori sia sia relativa al sistema percettivo dell'uomo, possano esistere determinati fattori, relativi alla singola persona, che vanno ad influire su tale sistema. Se tale teoria fosse confermata, si potrebbero aprire le porte ad ulteriori studi di ricerca nell'ambito della *Data Visualization* e portare così alla creazione di *sistemi di raccomandazione*, i quali potrebbero essere integrati nei sistemi oggi esistenti, con l'obiettivo di migliorare le aree di analisi tramite l'utilizzo delle visualizzazioni.

# Bibliografia

- [1] Babel: a javascript compiler. <https://babeljs.io>.
- [2] The definitive source of the best javascript libraries, frameworks, and plugins. <https://www.javascripting.com>.
- [3] Christopher Ahlberg and Ben Shneiderman. Visual information seeking: Tight coupling of dynamic query filters with starfield displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '94*, pages 313–317, New York, NY, USA, 1994. ACM.
- [4] Jacques Bertin. *Semiology of Graphics*. University of Wisconsin Press, 1983.
- [5] Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. D3 data-driven documents. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2301–2309, December 2011.
- [6] Mike Bostock. D3 data-driven documents. <https://d3js.org>.
- [7] S. K. Card and J. Mackinlay. The structure of the information visualization design space. In *Proceedings of the 1997 IEEE Symposium on Information Visualization (InfoVis '97)*, INFOVIS '97, pages 92–, Washington, DC, USA, 1997. IEEE Computer Society.
- [8] William S. Cleveland. *Visualizing Data*. Hobart Press, 1993.
- [9] William S. Cleveland and Robert McGill. Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, pages 531–554, 1984.
- [10] Stephen Few. *Data visualization past, present, and future*. 2007.

- [11] David Flanagan. *JavaScript: The Definitive Guide*. O'Reilly & Associates, Inc., Sebastopol, CA, USA, 3rd edition, 1998.
- [12] Jesse James Garrett. *The Elements of User Experience: User-Centered Design for the Web*. New Riders Publishing, Thousand Oaks, CA, USA, 2002.
- [13] Google. Angular.js. <https://angularjs.org>.
- [14] R.L. Hotz. The really smart phone. *Wall Street Journal*, 23, 2011.
- [15] Facebook Inc. Facebook open source project. <https://facebook.github.io/react/docs/jsx-in-depth.html>.
- [16] Facebook Inc. React: a javascript library for building user interface. <https://facebook.github.io/react>.
- [17] Facebook Inc. React: a javascript library for building user interface. <https://code.facebook.com/projects>.
- [18] Github Inc. D3 - tutorials. <https://github.com/d3/d3/wiki/Tutorials>.
- [19] Github Inc. Sites using react. <https://github.com/facebook/react/wiki/sites-using-react>.
- [20] QlikTech International. See the whole story. <http://www.qlik.com>.
- [21] Muzammil Khan and Sarwar Shah Khan. Data and information visualization methods, and interactive mechanisms: A survey. *International Journal of Computer Applications*, 34(1):1–14, 2011.
- [22] Jock Mackinlay. Automating the design of graphical presentations of relational information. *ACM Trans. Graph.*, 5(2):110–141, April 1986.
- [23] V. Mayer-Schönberger and K. Cukier. *BIG DATA: Una rivoluzione che trasformerá il nostro modo di vivere e già minaccia la nostra libertà*. 2013.
- [24] Microsoft. Bring your data to life. <https://powerbi.microsoft.com>.
- [25] Scott Murray. *Interactive Data Visualization for the Web*. O'Reilly Media, Inc., 2013.
- [26] "Nikkei". Car seat of near future ids driver's backside. 2011.



- [27] Severino Ribecca. The data visualization catalogue. <http://www.datavizcatalogue.com>.
- [28] George G. Robertson, Stuart K. Card, and Jack D. Mackinlay. Information visualization using 3d interactive animation. *Commun. ACM*, 36(4):57–71, April 1993.
- [29] P. K. Robertson. A methodology for choosing data representations. *IEEE Computer Graphics and Applications*, 11(3):56–67, May 1991.
- [30] Richard M. Shiffrin and Walter Schneider. Controlled and automatic human information processing: Ii. perceptual learning, automatic attending and a general theory. pages 127–190, 1977.
- [31] Ben Shneiderman. *Designing the User Interface (2Nd Ed.): Strategies for Effective Human-computer Interaction*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1992.
- [32] Ben Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *Visual Languages, 1996. Proceedings., IEEE Symposium on*, pages 336–343. IEEE, 1996.
- [33] TABLEAU SOFTWARE. Make your data make an impact. <https://www.tableau.com>.
- [34] BITBANG SRL. Digital intelligence, 2015. <http://www.digital-intelligence.it>.
- [35] Fluxedo Srl. Data analytics per tutti. <http://www.socialometers.com>.
- [36] Melanie Tory and Torsten Moller. Rethinking visualization: A high-level taxonomy. In *Information Visualization, 2004. INFOVIS 2004. IEEE Symposium on*, pages 151–158. IEEE, 2004.
- [37] Edward R. Tufte. *The Visual Display of Quantitative Information - Second edition*. Graphics Press, Cheshire, CT, USA, 2001.
- [38] Little W., Fowler H. W., and J. Coulson. *Shorter Oxford English dictionary Vol. 2*. 3 edition, 1972.
- [39] Wolfgang Wahlste, Jose Maria Cavanillas, and Edward Curry. *New Horizons for a Data-Driven Economy*. 2010.
- [40] Colin Ware. *Information Visualization: Perception for Design*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 3 edition, 2012.

- [41] Stephen Wehrend and Clayton Lewis. A problem-oriented classification of visualization techniques. In *Proceedings of the 1st Conference on Visualization '90*, VIS '90, pages 139–143, Los Alamitos, CA, USA, 1990. IEEE Computer Society Press.
- [42] Wikipedia. Data visualization — wikipedia, the free encyclopedia, 2016. [https://en.wikipedia.org/w/index.php?title=Data\\_visualization&oldid=756493348](https://en.wikipedia.org/w/index.php?title=Data_visualization&oldid=756493348).
- [43] Wikipedia. Dato — wikipedia, l'enciclopedia libera, 2016. <http://it.wikipedia.org/w/index.php?title=Dato&oldid=79609859>.
- [44] Wikipedia. Javascript — wikipedia, l'enciclopedia libera, 2017. <http://it.wikipedia.org/w/index.php?title=JavaScript&oldid=86197458>.
- [45] Wikipedia. Masterchef italia — wikipedia, l'enciclopedia libera, 2017. [http://it.wikipedia.org/w/index.php?title=MasterChef\\_Italia&oldid=86265170](http://it.wikipedia.org/w/index.php?title=MasterChef_Italia&oldid=86265170).