RIGA TECHNICAL UNIVERSITY

Faculty of Computer Science, Information Technology and Energy

Institute of Applied Computer Systems

Samay Kochhar

Student of the academic bachelor study program "Computer Systems",

Student ID No 211ADB013

COMPARISON OF DATA VISUALISATION TOOLS: MEETING REQUIREMENTS FOR DATA STORYTELLING.

BACHELOR THESIS

Signature not validated

Digitally signed by SAMAY KOCHHAR
Date: 2024.05.28 15:50:23 EEST

Scientific adviser: assist. prof. Dr.sc.ing.

VITA ŠAKELE

RIGA 2024

RIGA TECHNICAL UNIVERSITY
FACULTY OF COMPUTER SCIENCE, INFORMATION TECHNOLOGY AND ENERGY

Institute of Applied Computer Systems

Work Performance and Assessment Sheet of the Bachelor Thesis

The author of the graduation thesis:

Student: Samay Kochhar_____ _____ (signature, date)

The graduation thesis has been approved for the defence:

Scientific adviser:

Scientific adviser: assist. prof. Dr.sc.ing. VITA ŠAKELE

(signature, date)

ABSTRACT

Keywords: Data Storytelling, Data Visualisation tools, Comparison of data Visualisation tools, Interactive visualisation, Animated visualisation

Data storytelling has been catching people's attention in recent times. It is used everywhere in our daily life to the most important areas of decision-making. It is used to present and communicate complex and difficult-to-understand data so that the audience can understand effectively. Yet, it is hard to find an appropriate tool that supports the requirements for data storytelling. This bachelor thesis aims to compare and analyze various data visualization tools based on meeting specific requirements for data storytelling. The first task in this research is to define the requirements for data storytelling keeping various factors in consideration such as audience type, size of audience, goal, and communication methods. This thesis can serve as a checklist for selecting the tools for data storytelling. Afterwards, an extensive literature review needs to be conducted on existing research on data storytelling support in different data visualization tools. Experiments are then conducted to compare and analyze various data visualization tools based on features and techniques for data storytelling. Through these experiments, recommendations are provided for specific goals and scenarios. This research is a contribution to the field of data visualization and community as the recommendation for selecting the appropriate tool for different scenarios and features. The findings of this study potentially benefit both academicians and practitioners in the field of data visualization and data storytelling. It leads to effective data storytelling, organizations and individuals can effectively communicate their data insights, leading to better decision-making and improved understanding of complex data.

This thesis comprises 4 chapters spanning over a total of 59 pages. Additionally, it contains 2 tables, 25 figures, and references to 78 information sources.

Atslēgvārdi: Datu stāstīšana, Datu vizualizācijas rīki, Datu vizualizācijas rīku salīdzināšana, Interaktīva vizualizācija, Animēta vizualizācija

Datu stāstīšana pēdējā laikā ir piesaistījusi cilvēku uzmanību. To izmanto visur mūsu ikdienas dzīvē svarīgākajās lēmumu pieņemšanas jomās. To izmanto, lai parādītu un viegli interpretētu sarežģītus un grūti saprotamus datus, lai ikviens to varētu efektīvi saprast. Tomēr ir grūti atrast piemērotu rīku, kas atbalstītu datu stāstīšanas prasības. Šī bakalaura darba mērķis ir salīdzināt un analizēt dažādus datu vizualizācijas rīkus, pamatojoties uz specifiskām datu stāstīšanas prasībām. Pirmais šī pētījuma uzdevums ir definēt prasības datu stāstīšanai, ņemot vērā dažādus faktorus, piemēram, auditorijas veidu, auditorijas lielumu, mērķi un komunikācijas metodes. Tas kalpos kā kontrolsaraksts rīku atlasei. Pēc tam ir jāveic plašs literatūras apskats par esošajiem pētījumiem par datu stāstīšanas atbalstu dažādos datu vizualizācijas rīkos. Pēc tam tiek veikti eksperimenti, lai salīdzinātu un analizētu dažādus datu vizualizācijas rīkus, pamatojoties uz datu stāstīšanas funkcijām un metodēm. Izmantojot šos eksperimentus, tiks sniegti ieteikumi konkrētiem mērķiem un scenārijiem. Šis pētījums būs lielisks ieguldījums datu vizualizācijas un kopienas jomā kā ieteikums, izvēloties piemērotu rīku dažādiem scenārijiem un funkcijām. Šī pētījuma rezultāti potenciāli dos labumu gan akadēmiķiem, gan praktiķiem datu vizualizācijas un datu stāstīšanas jomā. Tas nodrošinās efektīvu datu stāstīšanu, organizācijas un privātpersonas varēs efektīvi paziņot savus datu ieskatus, tādējādi uzlabojot lēmumu pieņemšanu un uzlabojot izpratni par sarežģītiem datiem.

Šis darbs sastāv no 4 nodaļām, kas kopumā aptver 59 lappuses. Turklāt tajā ir 2 tabulas un 25 attēli ar 78 atsaucēm.

TABLE OF CONTENTS

INTRODUCTION

Data storytelling is an important aspect of communicating the data but there is not much research done on the requirements for data storytelling and comparison of data storytelling tools. Which data visualisation tool should be used for data storytelling and how well each tool supports data storytelling features and techniques?

In today's world the amount of data is increasing at a high pace and so are the data-informed decisions. The more the data, the harder it becomes to effectively communicate and explain it for this thus arises the need for data storytelling. In day-to-day lives, from small to big businesses and organisations data-informed decisions are made. The role of data storytelling is to somehow convey this complex information generated from raw data into an interesting and engaging narrative so that it can easily be understood by the audience.

There are various data visualization tools, and it is hard to choose one according to the user requirements. There is a need to analyze and compare them based on their support for data storytelling. This makes the topic of this research highly important and useful. These tools are known for their various features to explore, manage, manipulate, and generate insights but the features to convey these insights are still missing.

Therefore, providing recommendations for choosing the right tool for data storytelling forms the justification for the research topic. There are also many outside factors in data storytelling like audience type, size, goal, and communication method that are overlooked as to how well each tool supports these methods. Important techniques and features have also been overlooked. Such analysis is currently missing in existing research. Data contributes to a variety of fields. The results of this study could bring a change in their way of communication and lead to more informed decision-making.

The **goal** of the thesis is to study and compare how various data visualisation tools support the requirements of data storytelling.

The **tasks** of the bachelor thesis:

- To define the requirements for data storytelling by analysing research papers and conducting a literature review on common storytelling methods and techniques.
- To conduct a literature review to identify the existing research on data storytelling support in data visualization tools.

- To conduct experiments to compare existing data visualisation tools to measure how well they meet the defined requirements.
- To provide recommendations for selecting the appropriate data visualisation tool for specific data storytelling goals and features.
- To validate the provided recommendation through practical demonstration and scenario application.

The bachelor's thesis contains 4 chapters and the structure of the thesis is as follows:

- The first chapter discusses the concept of data storytelling, its use, importance, and impact on decision-making.
- The second chapter discusses about the types of data storytelling including the role of interactivity, various methods, concepts, and factors affecting data storytelling.
- Chapter three of this thesis includes information about data visualisation's role in data storytelling and creating requirements for data visualisation tools.
- The fourth chapter contains the experimental part of the thesis including metrics for comparing the tools, and dataset, analysis of selected tools and features, providing recommendations and validating using practical demonstration.

In the end, the author of the thesis provides the conclusion of the comparison of data visualisation tools on meeting the requirements for data storytelling and its features. The author also provides the limitations of this thesis and directions for further research in this area.

1. INTRODUCTION TO DATA STORYTELLING

Data storytelling is a relatively new term, but it's been around for ages. This picture explains that people are getting to know about its importance and is growing at a fast pace. In the research paper (Ojo & Heravi, 2017) it is mentioned that there are limited publications on data storytelling. In early July 2016, only 202 documents were returned when "Data storytelling" was searched in Google Scholar. In April 2017, it returned only 316. The graph below is made using Google Ngram Viewer. This shows the frequency of search of the term 'Data Storytelling'. It shows data from 2010 to 2019. The data earlier than 2010 is very low with not much growth and data after 2019 is not available. The graph shows a high upward trend in the last 4 years showing interest in growth and relevance. The trend suggests it is a potential area of future research in the direction of tools, techniques, frameworks etc.

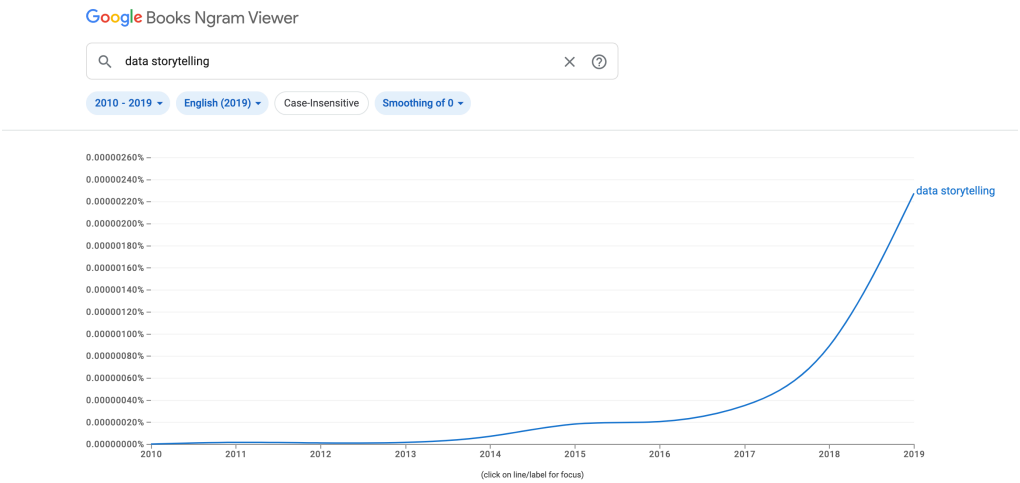

Fig. 1.1. Data Storytelling term search on Google (made using Google Ngram Viewer)

Segel & Heer mention in their paper that as our knowledge about narrative visualisations improves, it will also open new possibilities for visualisation/storytelling tool research. (Segel & Heer, 2010). It is one of the first papers to explore the direction of data storytelling, has been extensively cited, and remains integral and significant in this topic.

According to the Power BI website "Data storytelling is the process of crafting a captivating narrative from complex data and analytics that help tell user story and influence and inform a particular audience." (Team, n.d.). It's a combination of data

analysis with the art of storytelling involving communicating insights, patterns, and trends from the data in an interesting and captivating narrative that is both understandable and actionable for the stakeholders. This approach bridges the gap between quantitative data experts and decision-makers by presenting findings in an engaging and accessible manner. It injects a human touch into the complex and technical world of raw data, making it more relatable and understandable. Data Storytelling is comprised of three main elements: Data, Visuals and Narrative.

According to Gemignani, data storytelling is using traditional storytelling-narrative flow, context and structure to drive the audience through visualisation using narrative and data and provide relevant insights (Gemignani, 2021). The audience is the first priority, as the story is created to convince the audience using data. The storytelling is more than data, Data visualisation combined with narrative and graphics together guide the audience to clear insights and help them to take necessary action (Gemignani, 2021).

1.1 Evolution of Storytelling

The art of Data Storytelling dates back to tribal communities, who utilized stories as a means of passing down knowledge to the next generation. This tradition has evolved into a widespread practice, as it can be seen in stories depicted at cave paintings and even utilized in modern advertising campaigns. Presentations, too, have embraced storytelling as an effective means of delivery. And now, with the rise of social media, storytelling has taken centre stage, with the focus on crafting captivating and engaging stories in content creation (Kernbach, 2018). Storytelling is an ancient art rooted in our common human culture, as well as in our physiology and psychology. According to Bran Ferren, co-chairman and chief creative officer of Applied Minds, Inc., it is even considered the world's second-oldest profession (Gershon & Page, 2001b). Data storytelling is akin to traditional human storytelling, with the added benefits of providing a deeper understanding and supporting evidence through the use of visual aids such as charts and graphs (Team, n.d.).

Stories convey content and messages in a subtle, subconscious manner, making it easy for the audience to absorb without exerting much effort, as compared to facts and statistics (Kernbach, 2018). Stories are assumed to appeal to the experiential system of personality because of their emotional engagement and representation of events in a

manner like how they are experienced in real life (Epstein, 1994). Stories communicate who did what, when and why and act as thought machines through which (Brinker, 1988) individuals test out ideas and feelings about something and try to learn more about it (mcLellan, 2006). There are 4 types of basic stories: descriptive, explicative, argumentative, and narrative (Brinker, 1988) which have been adapted by Dan Roam into Report, for describing facts, Explanation, for explicative, Pitch, for argumentative, and Drama, for narrative (Roam, 2014).

A good story always connects a cause to an effect (Herman & Schank, 1996). Each effect has a cause as something is related to another. This is also true vice versa. A good story neither states the obvious nor uses common sense as a universal yardstick (Dahlstrom, 2014). In the case of common sense, then the speaker may misunderstand the audience's understanding level due to which the audience can lose their attention. Even this sentence has a cause-and-effect relationship. This relationship is used even in daily life English sentences. Furthermore, presenting information already known to the audience may reduce engagement. A great example of this is scientific stories (Matei & Hunter, 2021).

A story to be called as a good story it must draw audience interest. This can be created by adding the surprise element. Good stories should violate audience expectations because which makes them curious (Matei & Hunter, 2021). To find this reason for the violation of the expectations, questions are triggered in the brain of the audience till they get this answer. The question of why is created leads to the creation of the surprise element in the audience's mind. To find that answer to the question the audience will listen to the story to the end. The answer to that question is used as a lesson (Matei & Hunter, 2021).

The best stories excel because they induce a change in the audience through learning (Matei & Hunter, 2021). An example of this is when an elder tells bedtime stories to a child. The child gets curious and asks questions in return. Usually, the answer to the question carries the learning for the child. Similar is the case with author-driven narratives. (Matei & Hunter, 2021).

1.2. Use of data storytelling in different sectors

Storytelling is used in knowledge management, corporate culture, branding and advertising it engages content and connects it with the audience in a meaningful and

memorable way (Kernbach, 2018). Storytelling can be considered an informal practice that facilitates the transfer of tacit knowledge (Kernbach, 2018). In the article What storytelling can do for information visualisation Gershon & Page mention that 'we all began our lives getting most of our information visually'(Gershon & Page, 2001). This makes it the most familiar way for humans to communicate information (Gershon & Page, 2001). Researchers can use stories not only to support discussion and decision-making but also for process analysis (Kosara & Mackinlay, 2013).

Data storytelling is used in different sectors but it comes with its own unique challenges. Sometimes it is tempting to pick only data sets that support the story but looking at the whole picture is more beneficial. Having loads of data sets available can be overwhelming but it helps the narrator to structure the story and how he would like to tell the story (Team, n.d.). Complexity, too many analytics software options and no idea what to choose, alignments and a shared vision (Contextualize Your Data with the Storytelling Studio | Toucan, n.d.). User has no idea about what features are required according to their goal. It is the final step and most of the time people put all of their time and energy into analysis and finding insights that they forget to focus on communication.

1.3. Importance and Impact of Data Storytelling

The reason for writing this thesis on Data Storytelling is due to its importance and benefits in the different sections. In addition to offering an incredibly popular way of conserving data and passing it on, stories also provide the connective tissue between facts that make them memorable (Austin, 2011). Stories are an effective means of communication, packaging information and knowledge in a way that is easy for the other person to understand (Kosara & Mackinlay, 2013). While raw data gives tells 'what', data storytelling allows us to understand "why", "when", "where", and "what to do about it (Contextualize Your Data with the Storytelling Studio | Toucan, n.d.). Data storytelling plays a crucial role in making data more accessible, understandable, and actionable. It bridges the gap between data analysis and communication, enabling effective storytelling with data (Sun et al., 2022; Amini et al., 2017). Shape the narrative of the story in such a way that it is possible to conclude the purpose in a single sentence.

Data storytelling has a huge impact on the audience. It leads them to actions, helps to communicate insights easily, bridges the gap between the user and the

stakeholders by highlighting patterns and trends, and engaging the audience and building emotions. Data storytelling has the following benefits:

- Highlight Patterns and Trends - Visualizations with storytelling can show patterns, trends, and relationships in the data that might not be seen in numbers or text (Lee, et al., 2015). They provide context, help the audience to understand the results and make the data easier to understand so the audience can process information.
- Engaging the audience - Interesting narratives hold better retention of the audience for a longer period than just visuals. Hence making it more impactful (Lee, et al., 2015). Complicated information is simplified according to audience understanding so they can engage with the content and make critical decisions quicker and more confidently (Team, n.d.). Visual data stories engage the audience by capturing their attention and maintaining their interest throughout the narrative. This engagement leads to better retention of information, as people are more likely to remember stories compared to raw data or statistics (Sun et al., 2022; Amini et al., 2017).
- Information exploration and interaction - Interactive visualizations allow the audience to explore the data on their own terms. They can manipulate the visual elements, filter the data, and interact with different aspects of the visualization. This interactivity provides the audience to play with the data and discover new insights, enhancing their understanding and involvement in the storytelling process (Lee, et al., 2015).
- Build emotions and influence the audience - Stories can produce emotions among the audience and results in retention of information for a longer period and generate a positive impact on them. They have the power to generate emotions, and curiosity, and create a memorable experience (Lee, et al., 2015). It provides a human touch to the data (Team, n.d.). Data storytelling can change perceptions, decisions, and opinions and help people make well-thought-out actions (Sun et al., 2022; Amini et al., 2017).

- Effective communication and Understanding - Combining data visualisation with narrative techniques results in easy communication of complex information making it easier for the audience to receive the insights and messages. It allows the audience to see patterns, trends, and relationships in the data, leading to deeper insights and better decision-making (Sun et al., 2022; Amini et al., 2017). People from different disciplines come together to collaborate and exchange their knowledge to build more engaging, impactful, and soulful stories (Sun et al., 2022; Amini et al., 2017) and work towards a common goal. Storytelling in the context of information visualisation plays a crucial role in extracting relevant insights through the visualization and results in better retention of it. Storytelling through data visualisation can make data more intelligible, provide insight, and convey knowledge easily and simply (Cruz & Machado, 2011b).
- Lead to more questions - A well-presented data story is likely to lead to new produce more questions in the mind of the audience. Effective presentations must create memories (Kosara & Mackinlay, 2013). Stories naturally lead to questions, which lead to discussions, which lead to deeper analysis (Kosara & Mackinlay, 2013). It builds credibility in the industry and make the user become the leader of the story (Team, n.d.).
- Improved decision-making and enhanced understanding - Visualizations and narrative structure of the story directly support each other. Visualizations are visual representations that are easier to understand and interpret. Stories add narrative to the visuals that help the audience get the information more quickly and effectively (Lee, et al., 2015). Explaining complex information in the story makes it more relatable and understandable for all types of audiences. It adds value to data and insights. Helps in interpreting complex information and highlighting essential key points for the audience (Team, n.d.). Understanding the story behind the data helps stakeholders, employees, customers, investors, businesses, and organizations in making well-

informed decisions through well-constructed data visualizations and stories highlighting trends, anomalies, and patterns (Few, 2009).

2. TYPES OF DATA STORYTELLING

Narrative visualization techniques combine data visualization and narrative techniques including sequencing, narration, and pacing to deliver an engaging experience and better comprehension of the data. The art of visualizing information involves transforming data, information, and knowledge into a visually digestible format that taps into the power of the human visual system. The ultimate purpose is to empower the user/viewer to not just look at the information, but to truly grasp, comprehend, and derive meaning from it (Gershon & Page, 2001). There are two major aspects of data storytelling: Narrative data Storytelling and Visual data storytelling. These both go hand in hand. Studies have shown that data visualization with narrative could reveal information effectively (Gershon & Page, 2001) and enhance readers engagement, memory, comprehension, and communication (Sun et al., 2022; M. A. Borkin, 2015; Jones, 2012; Segel & Heer, 2010). With the help of technology, information is available with a click this doesn't make it easier instead it makes it harder. As there is more and more information the more difficult it is to filter the most important information. Kelliher and Slaney (2012) classify data stories into four categories (Kelliher & Slaney, 2012):

1. 'Informative': Presenting data to support or challenge claims.

2.'Explanatory': Uncovering deeper meanings and connections hidden within the data.

3. 'Persuasive': Emphasizing the significance of the findings to shape opinions or actions.

4. 'Entertaining': Presenting data engagingly and enjoyably.

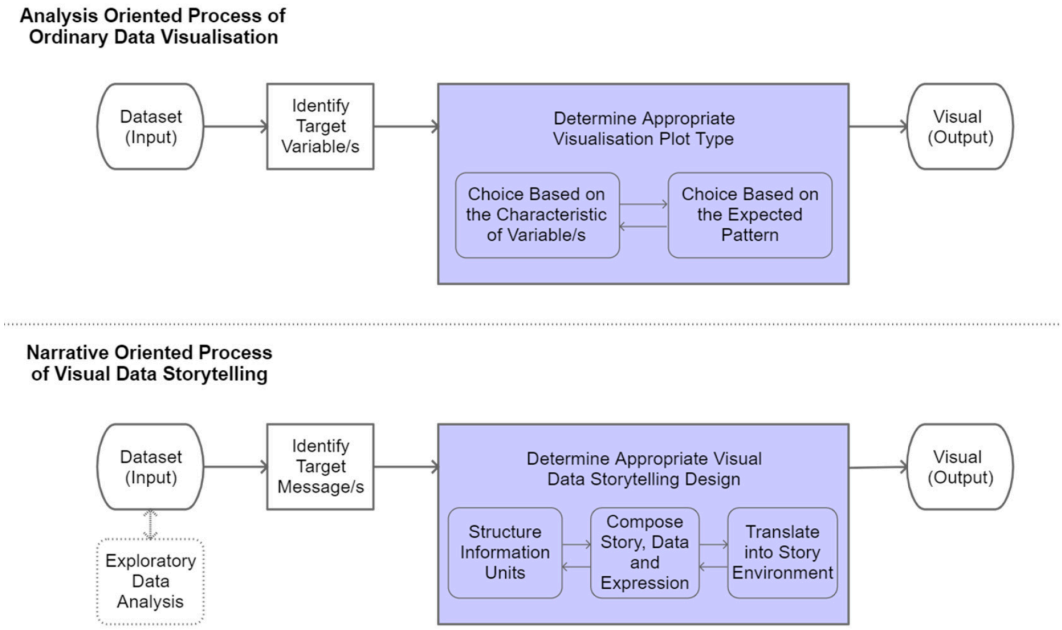

Fig. 2.1. Comparison of Visual Data Storytelling with Normal Data Visualisation (adopted from(Zhang et al., 2022))

2.1.Types of Data storytelling and Analysis

There are two types of fundamental approaches to understanding and working with data: exploratory and explanatory data analysis (Knaflic, 2015). In exploratory analysis, all the information is provided, and the audience needs to find the important things by themselves in explanatory it is needed to jot down the important information from all the data.

The objective of explanatory analysis is to explain or clarify the patterns and relationships in the data. It is used when the user already has a good understanding of the data and what he wants to communicate or trying to answer. The focus is on confirming hypotheses and answering specific questions.

The objective of exploratory analysis is to understand patterns, relationships, anomalies etc. explore data and find useful insights. This process is used at the beginning when the narrator is just trying to get a sense of data. The user often tries to create a variety of visualisations to understand different aspects of data. The focus is on generating hypotheses and more about asking questions than answering them.

Visual data stories are widely used in data-driven storytelling to communicate complex insights and support data understanding (Sun et al., 2022). A visual data story is a series of connected data facts shown in the form of a narrative visualisation, which

is usually used to help with information communication (Lee, et al., 2015). A visual data story includes a set of story pieces, and specific facts backed up by data presented in a way to deliver an intended message (Lee, et al., 2015; Sun et al., 2022). The author of this research article mentions that the term "storytelling" is used everywhere but there is no clear idea of what a visual data story consists of (Lee, et al., 2015). They focus on narrowing the scope of the data story by distinguishing between visual data story and data visualization. This includes the process of finding insights from data (explore data), converting these insights into a narrative (make a story) and communicating this narrative to an audience (tell a story). Context determines the way a visual story is presented (Lee, et al., 2015).

In the article Generative Storytelling for Information Visualisation the author mentions an important conceptual framework: generative storytelling. "This approach aims to build various stories conveying the same fabula from a given dataset" (Cruz & Machado, 2011b). Data fabulas are used as a set of time-ordered events caused or experienced by actors. Data fabulas are used to create stories by transforming the discrete fabula into a narrative with a continuous timeline. Finding a fabula from the dataset helps to maintain the chronological order of events. Actors are the agents that perform the actions (Cruz & Machado, 2011b). Users can also consider rows as incidents or events and columns as the actors that produce some actions. This concept allows to alter the dramatic meaning of the story without changing semantic meaning and it takes a flexible approach. It also provides room for expressiveness with 'the aim to have a balance between data portrayal and expressiveness'. This is one of the important features of this concept. The framework uses two models:

1. The event model is used to create the timeline of the story. Allowing the user to manipulate the time of the events and action model works on implementing behaviours of 'actors' also representing the fabula's actions (Cruz & Machado, 2011b).
2. The data story should have a logical narrative structure. The flow and sequence of the information should be designed in such a way that it is easy for the audience to follow and understand (Sun et al., 2022).

'Rule of three' says that a piece of argument should be supported by three pieces of evidence as this increases the charismatic appeal of the storyteller. The rule of three, the use of analogies and metaphor, the appeal to higher ambition and addition, when delivering the story, using a low voice, both for male and female storytellers, using

gestures to connect with the audience and making conscious breaks to gain momentum (Derrico, et al., 2013). Stories are best remembered if they have one emotional highlight and a meaningful end (Eppler & Kernbach, 2016).

2.2. Role of Interactivity in Data Storytelling

In the research paper Exploring Data in Virtual Reality (Millais et al., 2018) authors compared VR visualisations with 2D visualisations and concluded that VR visualisations increased dimensionality and sense of view helps in data exploration resulting in more satisfaction and fewer incorrect insights. It suggests that VR and interaction help to communicate insights better (Millais et al., 2018).

In the research paper (Ojo & Heravi, 2017), the authors analysed 44 papers on Data Analysis (63%) and Data Visualisation (65%) are considered essential technologies for effectively presenting data through storytelling techniques. The percentage refers to the number of storytelling examples of a particular category from the set of the 44 papers analysed. Map Visualisation (26%) is also used as a supporting technology. Some examples of the tools included are ArcGIS, Mapbox, mapper, Open Street Maps AI, and Google Maps. For Data visualisation, Tableau Public and D3.js are the most common data visualisation frameworks employed. Studies also show that the common representation style of stories is using Annotated graph/map (77%), then web app (14%), and Image (11%) while graphics, slideshow and animation are at 9%, 7% respectively. The studies show that there is a certain level of Interactivity in 59% of the stories of which 27% of them allow the user to search, filter and select while the other 7% have only limited interactivity (Ojo & Heravi, 2017). These stories share a key characteristic: they uncover hidden or contradictory information not apparent in surface-level data or news reports. Unlike technical reports, these stories take a journalistic approach, providing insights and context beyond the immediate headlines (Ojo & Heravi, 2017).

In the article ‘What Storytelling can do for Information Visualization’ the authors suggest Starting presentations with stories rather than slides (Gershon & Page, 2001). Using visual media, such as animations and comics, to present complex information. Integrating storytelling elements, such as zooming options and continuity, into data presentations (Gershon & Page, 2001).

Improving how data is shown is an ongoing discussion. The use of storytelling to boost users' participation, interaction, and understanding has been suggested. Some techniques combined with narrative styles are messages and interactivity. Researchers recognized seven genres of narrative visualization that can be suitable depending on the data explored: Magazine-style articles, Charts with notes, Poster separated into sections, Flow charts, Comic strips, Slide shows and Video/Animation (Segel & Heer, 2010). A bar graph can use notes to share a new story, offering facts viewers would not see alone and animations can improve graphics. Annotations can even be combined with highlights, animated transitions, and interactivity. A study aimed at identifying which tools work best for building stories would be fascinating, considering the user experience, the data used, and the story sought. It is very important to employ annotations and highlights in any type of graph, as they provide insights for the reader that could not otherwise be observed (Borges et al., 2022). Tools like Tableau have some resources for crafting narrative visualizations, but they are still quite restricted in that area (Borges et al., 2022).

Simple Interactivity plays an important role in telling stories with data. Interactivity allows creators to personalize the experience for different viewers based on what's in a data source. This helps share tailored views of the data with each person, making the information more useful and engaging. Interactivity also provides a chance to involve viewers with data in new ways, building on established video methods to tell an interesting and customized story. By changing how data is presented to be not just informative but also fun, interactivity makes storytelling an effective way to motivate many people to engage with data (Hook, 2018).

Exploration and explanation are increasingly supported through interaction in data stories. Techniques like details on demand and highlight hover help readers interact with these stories. They enhance the experience and facilitate understanding. Interaction can strengthen how readers connect with stories. It may boost involvement and reinforce memory of the information (Shi et al., 2022).

Dimara and Perin grouped interactions into three types: data actions, viewing data actions, and non-data actions. Interaction type ranges from basic to complex. It is only possible through actions with a data interface (Dimara & Perin, 2017). When Segel and Heer first suggested the idea of narrative visualization, interaction was seen as a key part of the design (Segel & Heer, 2010). This included seven common interactions like hovering, filtering, and navigation. Stolper and others (Stolper et al., 2016)

identified and described storytelling techniques used in new data stories. These techniques were put into four groups: communicating the story, explaining the data, connecting separate story parts, improving structure and navigation, and allowing controlled exploration.

Interaction can also provide direct benefits to story readers, including deepening the comprehension and recall of information, eliciting deeper reflection, and augmenting levels of engagement. Specifically, (Hohman, et al., 2020) found interactive articles have five unique benefits. They can link people and facts, make systems more fun, prompt self-reflection, personalize the reading experience, and reduce mental strain. Audiences desire captivating visuals, innovative experiences, engaging narratives, and immersive interactions. (Dykes, 2019)

In the analysis of 45 author-driven data stories and 20 data storytelling techniques. They identified many data storytelling techniques that are used in the examples. New ways of sharing information and navigating through stories have emerged, such as using scrolling to move the story forward and start animations. These stories included slideshows (interactive, recorded live and saved online, or just saved online), animated data videos, and interactive timelines. Visuals are usually added as static pictures because support for visualization is basic (Stolper et al., 2016).

They have divided the storytelling techniques into 4 categories which covered 20 techniques. According to the literature review and the author's work. It could be shortlisted down to 14 techniques commonly used in data storytelling.

Segel & Heer Codes	Our Techniques
Genre	
Magazine Style	(Majority of stories)
Annotated Graph/Map	Text annotation on visualization
Partitioned Poster	
Flow Chart	Flowchart arrows
Comic Strip	
Slide Show	Next/previous buttons, some scrolling
Film/Video/Animation	Narration
Visual Narrative	
Visual Structuring	
Establishing Shot/ Splash Screen	
Consistent Visual Platform	
Progress Bar/Timebar	
Checklist Progress Tracker	Breadcrumbs, Section header, buttons, Menu selection, Timeline, Geographic map
Highlighting	Highlighting elements
Close-Ups	
Feature Distinction	
Character Direction	
Motion	
Audio	
Zooming	
Transition Guidance	Linking through color, Linking through animation
Familiar Objects (with Cuts)	
Viewing Angle	
Viewer (Camera) Motion	
Continuity Editing	
Object Continuity	
Animated Transitions	
Narrative Structure	
Ordering	(Focus on linear stories)
Random Access	
User-Directed Path	
Linear	
Interactivity	
Hover Highlighting / Details	Tooltips
Filtering / Selection / Search	
Navigation Buttons	Next/previous buttons
Very Limited Interactivity	Dynamic query, Embedded exploratory visualization
Explicit Instruction	
Tacit Tutorial	
Stimulating Default Views	
Messaging	
Captions / Headlines	Labeling
Annotations	Text annotation on visualization
Accompanying Article	Separate exploratory visualization
Multi-Messaging	
Comment Repetition	
Introductory Text	Textual narrative
Summary / Synthesis	Textual narrative

Fig. 2.2. Analysis of Data Storytelling examples (adopted from (Segel & Heer, 2010; Stolper et al., 2016))

In the above figure, the author of this thesis depicts the interaction features and techniques for data storytelling mentioned by (Segel & Heer, 2010; Stolper et al., 2016).

Techniques and features discussed above by (Stolper et al., 2016) are explained in the following paragraphs containing the use and effect of each. The techniques that help in communicating insights are (Stolper et al., 2016):

Textual Annotation: Annotations can be labelled above or below charts, titled slides or sections. There are three properties of annotation: content, location and time. They provide additional information and explanations for specific data points. For example, If there is a sudden increase in the sales of a product, the reason 'why' can be explained using annotations.

Highlighting elements: Segel and Heer have mentioned 7 types of highlights. (Close-ups, Feature distinction, character direction, motion, audio, and zooming) (Segel & Heer, 2010). It allows the user to draw the audience's attention to important parts of the visualisation. It provides the opportunity to communicate effectively and can be helpful during explanatory analysis.

Labels: Labels help the audience to easily read and interpret the chart. It can be in any form as axes, titles, captions or summaries. It shows what is what and explains the visualisation making it more understandable.

Animation: Animation allows the audience to experience data visualisation and makes it more engaging. It helps to illustrate dynamic data and data changing with time.

Dynamic query: filtering/selection/ search. Dynamic query refers to the process of filtering, selection and searching in data. It allows the audience to look for something they are searching for.

Scrollytelling: As the reader scrolls up or down, changes occur in the story set by the user. This helps the user to drive narrative with the visualisation leading to a narrative-driven story.

Changing viewing angle: Looking at the visualisation from different viewing angles provides new insights and perspectives that can not be gained from a single angle. The viewer can rotate or move a chart in any direction.

Embedding visualisations: Embedding visualisation is a necessary feature when it comes to sharing visualisations. It allows the user to interact with the data without needing the required tool to communicate the insights.

Interactive tooltips: Tooltips are Additional details details shown when the reader hovers over an element in the visualisation. It also can be shown when Hovering and selecting. It is similar to annotation but it is used in the exploratory analysis. It only shows when the user interacts with the visualisation.

Consistent colour scheme: A Consistent colour scheme allows the user to relate or divide different data points. It can also be used to draw the audience's attention to specific points.

Next/previous button/ Flowchart arrows: Navigation buttons allow the audience to move forward or backward or if they want to look at specific visualisation or slide.

Breadcrumbs: Breadcrumbs are also navigation buttons to go forward or backwards. They also give an additional hierarchical structure to the story.

Navigation through geographic map: This navigation method as the name suggests refers to the navigation through the geographic map.

Timeline: Timeline is a useful feature when it comes to explaining time series data. The audience can go to any specific time to see the visualisation.

The research paper by (Stolper et al., 2016) mentions that most of the tools do not support most of the techniques mentioned above by the visualisation software. But in the meantime, of 7 years. Things have changed a lot. Tools have now started to cope with the requirements and storytelling techniques.

2.3. Methods and Concept for Data Storytelling

A story contains components that form the story (structures, elements, and concepts) and those that influence that 'telling' part of storytelling (people, tools, and channels) (Lee, et al., 2015). There are various methods and concepts mentioned in the resources on Data storytelling. But the most commonly used and known are the following resources:

1. **The Data Storytelling Arc (Knaflic, 2015)**:

This method is a blueprint for tailoring data narratives and provides a logical flow to data stories. It mainly consists of three sections:

- **Setup:** This is the starting phase where the audience is provided with the data context and the dataset's origin, nature, and general overview.
- **Conflict:** This stage highlights the central insights or discrepancies in the data. Any surprising findings, trends or anomalies that require further exploration or action are also communicated.
- **Resolution:** This is the conclusion phase where the narrator draws the actionable insights or potential steps for action.

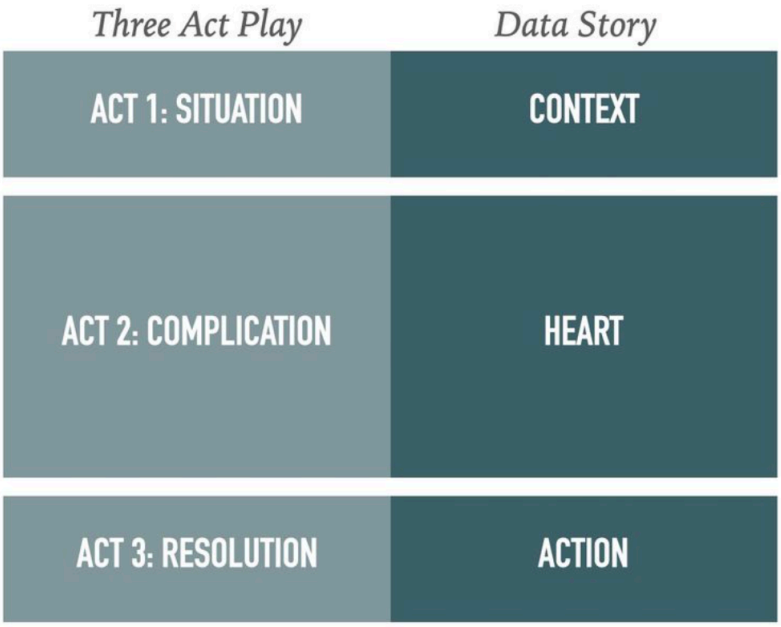

Fig. 2.3. Three-Act Structure of Data Storytelling (adopted from (Gemignani, 2021))

In the setup phase histograms, pie charts or other broad visual overviews are used to provide a sense of data distribution. In the conflict phase, more detailed visualizations are used to pinpoint any particular anomalies or trends such as scatter plots, heatmaps, line charts etc. In the resolution phase, more summarized charts can be used (Knaflic, 2015).

In the exposition phase area charts or geographical maps can be used. Multi-layered visualization can provide drill-down insights into rising action. Attention-grabbing visuals like treemaps or radial charts can be used to present insight into the climax phase. Simple yet impactful visualisations can be used to drive takeaway points in Falling action and denouement (Segel & Heer, 2010).

2. Data-Ink Ratio Concept

Data ink Ratio is a concept of minimalism in data visualisation introduced by Edward Tufte. It points to the idea that any non-data ink (or pixels) affects the story's efficiency in communicating and conveying the message. By maximizing this ratio, one can maximize the audience's attention and eliminate distractions. Various visualisation tools offer amazing designs and decorative options. These options are tempting to use, but the author advises restraint from doing it (Tufte, 2001). Tools that support minimalistic designs with high-contrast visuals with limited colours, emphasizing data

points allowing of removal of gridlines, excessive labels, or unnecessary decorations support this principle (Tufte, 2001).

3. Storytelling Canvas

'Storytelling canvas' (Kernbach, 2018) consists of general conditions of the topic that the story is about, the identification of the audience and their needs, using the empathy map canvas, (Gray, 2017) and the goal should be achieved by telling the story and by anticipating what the audience should think, feel, and know after they have been exposed to the story. The story elements consist of an overarching three-level structure of beginning, middle and end in line with the three-act model of storytelling (Mckee, 1997). Three elements can be used to create an impactful start which is 'start with why' (S.Sinek, 2009) 'common ground' and 'one big idea' (Duarte, 2013). Start with why helps to focus on why it is important. Common ground is to find something common to relate. One big idea refers to saying everything in one complete sentence. The seven elements abbreviated as the "SUCCESS" formula (Brinker, 1988) can be used to form the middle part of the story where S for Simplicity, U for Unexpectedness, C for concreteness, C for Credibility, E stands for emotions, S for storyline and the last S for "Star moment". 'Reward' and 'Call to action' can be the two things to finish it off with a high. Reward is the benefit to the audience, or anything related to the audience and a Call to action makes the audience do something. (Kernbach, 2018).

4. Visual Data Storytelling Process (VDSP)

During the Explore and Analyze phase, relevant data excerpts are collected through the analysis. These Data excerpts are assembled sequentially, creating logical links, developing a flow, defining a central message, and crafting the conclusion. Users get the plot of the story which describes the meanings and relationships between the story elements in terms of timing cause and effect patterns etc. Then in the last stage comes the data storytelling part where the storyteller creates the presentation, shares the story, and receives feedback. It involves developing visual representations, interactions, annotations, or narration (Lee, et al., 2015).

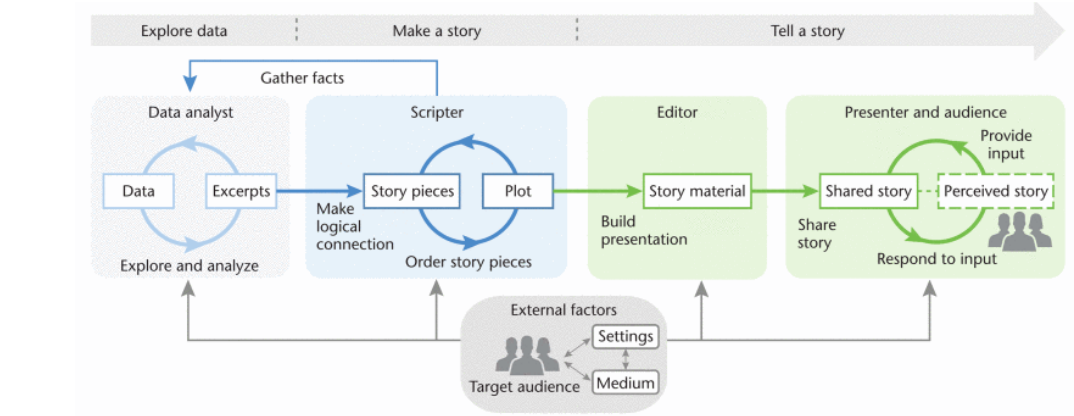

Fig. 2.4. Visual Data Storytelling Process (VDSP) components (adopted from (Lee, et al., 2015))

Fig. 2.4. shows, the process is divided into three parts: Explore data, construct the story and narrate the story. It covers the full process of data storytelling. There are four major roles: Data analyst, scripter, editor, and presenter (Lee, et al., 2015).

5. Microsoft's Data Storytelling Process

There are three essential elements to the data storytelling process: narrative, visuals, and data. These three elements combine to create a well-rounded anecdote of the theory and surely create an emotional response.

Build a narrative: When telling the story leverage data as pillars to support user insights. Develop insight by distilling complex information into informative insights to help the audience understand users' points of view. Narrative and context will drive the linear nature of data storytelling (Team, n.d.).

Use visuals to enlighten: By linking visual elements, such as charts and graphs, to the narrative, the audience is captivated and exposed to hidden insights that serve as key data to support the theory. Rather than relying on a single data point to bolster the theory, showcasing a range of data, from the smallest details to the big picture, allows the audience to truly grasp and appreciate the perspective (Team, n.d.).

Show data to support: While humans may not possess a natural affinity for analytics, the addition of augmented analytics can be a major driver for a good outcome. Narrative illuminates crucial insights through tangible data and thoughtful critique. Integrate business analytic tools to extract deeper meaning and provide context, enhancing the overall impact of the data story (Team, n.d.).

6. User-centred Interactive Data Storytelling Framework

A framework for data storytelling should encompass a plan for structuring information, methods for mapping visual data units, and suitable approaches grounded in user-focused design. To cultivate this framework, the study aims to fuse the technique of data visualisation with ideas sourced from spheres like communication and information theory, coupled with artistic methods and aesthetics studies. The key elements of this framework are (Zhang and Lugmayr, 2019):

- Data snippets act as essential elements to transmit the details hidden in data sources.
- Key designing aspects required for creating an interactive data narrative.
- Integrating theories from different fields within the frame of data storytelling.
- Functional strategies for integrating data into a narrative.

It also focuses on creating simple yet complex visualisation, prioritising the important elements. But this framework is currently in progress, and it will be evaluated based on user study in future (Zhang and Lugmayr, 2019).

2.4. Factors Important for Data Storytelling

Many things affect the process of data storytelling. The choice of genre, as well as the presentation medium, affects content, as well as what the audience gets from the process, following the Canadian literary and media scholar (and professor at the University of Toronto) Marshal McLuhan's well-known pre-PC-era insight: "The medium is the message (Gershon & Page, 2001). The particular genre of presentation can have a positive as well as negative effect on the understanding process for the audience (Gershon & Page, 2001). The different factors that affect the data storytelling process are:

- Data Accuracy and Integrity - If the data is not accurate then the insights generated can be misleading making the stakeholders take poor decisions which can negatively affect the user and the company. Consistent data ensures that everyone is on the same page and reduces confusion. All the insights and story of data should be clearly communicated in the narrative.
- Audience Understanding - Consider the background, knowledge, education level, age, profession, and interests of the audience. Know the

domain-specific knowledge and technical understanding of the intended recipients of the story. Each person has a different level of data literacy, cognitive understanding and cultural sensitivities. Tailor the data story to their level of understanding and use language and visuals that resonate with them. This will make it both accessible and meaningful (Sun et al., 2022). The story should resonate with the audience and make sense of the information. The audience should be segmented into categories such as experts, the general public, executives, etc. Understand what is important for each of the audience (e.g., simplicity, detail, speed). Such as for the domain level expert storytellers can delve deep into the topic whereas for the general audience, more explanation and simplification would be required. Usually, scientific visualizations are intended for highly educated scientists but that is not the case in Data Storytelling. There are people with all types and levels of personal skills, education and tastes. (Gershon & Page, 2001).

- Communication method - Different communication methods require different and unique forms of narrative methods to convey the story's insights. Commonly used communication method are live presentations, reports, dashboards, videos and other interactive methods. The narrative of the story can be either lead by the user or the audience depending on the exploratory or explanatory.
- Size of audience -The data story depends on the size of the audience for example. individual, small group, large auditorium, vast online audiences etc. The mode of presentation and interaction is dependent on the size of the audience. A large amount of audience needs a broader overview and low interactivity while a smaller group might need deep discussions.
- Appropriate visualization - Choose appropriate visual representations for data. Select charts, graphs, maps, or other visual elements that effectively communicate the information and support the narrative (Sun et al., 2022).
- Interaction with visualization - Audience retains more information through interactivity but it requires high-capability tools and proper

design to prevent misunderstanding. High interactivity can also distract the audience from the goal.

- Relevance -Add relevance to the story by connecting it to real-world situations. Help the audience understand why the data is important and how it relates to their lives or the broader context (Sun et al., 2022).
- Time aspect of visualization - Temporal data storytelling requires tools or methods that can efficiently depict change over time, such as line graphs, animations, or time sliders. Flourish uses animated visualisations to achieve audiences attention for a longer duration of time. This leads to higher interaction and retain ability of the communicated information. More about this is mentioned in under experimentation section in this research paper.
- Analysing complex data - More complex data need advanced visualization techniques, aggregation or multiple visuals. Hence, it is limited to using only tools that allow data manipulation for complex, multifaceted, or large amounts of data. To extract reliable and significant insights from data, statistical techniques, equations, filters, and computations are employed to analyze the data. A well-designed data storytelling tool should enable the import, manipulation and delve into data from different origins and arrangements. It should also support a wide range of basic and advanced analysis features including aggregation, correlation, regression, clustering, and others (Presentation, 2023).

3. DATA VISUALIZATION TOOLS ROLE IN DATA STORYTELLING

A visualization can be thought of as a fusion of two critical elements - design and presentation - working together to create a compelling narrative (Strothotte & Strothotte, 2003). The visualization includes annotations (labels, pointers, text, and such) or narration to highlight and emphasize the intended messages and to avoid ambiguity (Lee, et al., 2015). A study exploring the impact of embellishments on memory found that infographics, with their dynamic and visually appealing designs, were significantly more memorable compared to simplistic charts (Bateman, et al., 2010). Finding the right balance between text and visualization is crucial, as excessive text can distract from the underlying data while too little text can leave the audience struggling to make sense of the connections (Kosara & Mackinlay, 2013). Regardless of how useful or amazing a tool may be, it will never have the same understanding of the data and narrative as the narrator (Knaflic, 2015).

3.1. Existing research on data visualization tools in data storytelling

There are a lot of auto-data story design tools that depend on the users to make critical decisions. Most of them generate poor-quality stories and lack feelings whereas man-made data stories are soulful and insightful but the process is relatively tough and sometimes requires experience and expertise which makes it inefficient (Sun et al., 2022).

For this type of communication, individuals have frequently relied on tools such as PowerPoint and Excel, as well as conventional dashboard and business intelligence systems. Though these tools offer solutions, they fall short of achieving the perfect balance between the informational and exploratory aspects of a powerful data story. However, a new set of tools is on the rise, designed specifically for data storytelling. These innovative tools boast user-friendly visualizations, seamless integration of text and visuals, smooth flow of narrative, interconnected stories, user-friendly authoring, and effortless sharing (Gemignani, 2023). On the Juiceanalytics website, the author Zach Gemigani mentions about four unique categories he identified to compare and divide different data visualization tools (Gemignani, 2023).

- **Tools for Guided Analytics:**

Combination of exploratory data visualisation with explanatory text and graphical elements. The interactivity features of these tools provide them an edge against traditional visualisation tools.

Juicebox: Juicebox is an all-in-one platform. It creates world-class data visualisation, interactive applications, reports and performance dashboards. It can combine various images, narrative text, and interactive visualisations to tell memorable stories with the data. It is for non-technical business users for eg. Business users who work more with Excel as compared to SQL. It's easy to use and learn. It is cost-effective and affordable for any size organization. It provides a balance between communication and analysis. (*Juicebox Vs. Competition — Juice Analytics*, n.d.). It is one of the tools where they provide the comparison with the tableau and Power BI in terms of purpose, communication method, audience and users. It has one of the best materials regarding data storytelling. Be it lessons or examples they have got a lot of informative details. It has received the award of top 50 Data visualisation vlog. Natville technology company of the year.

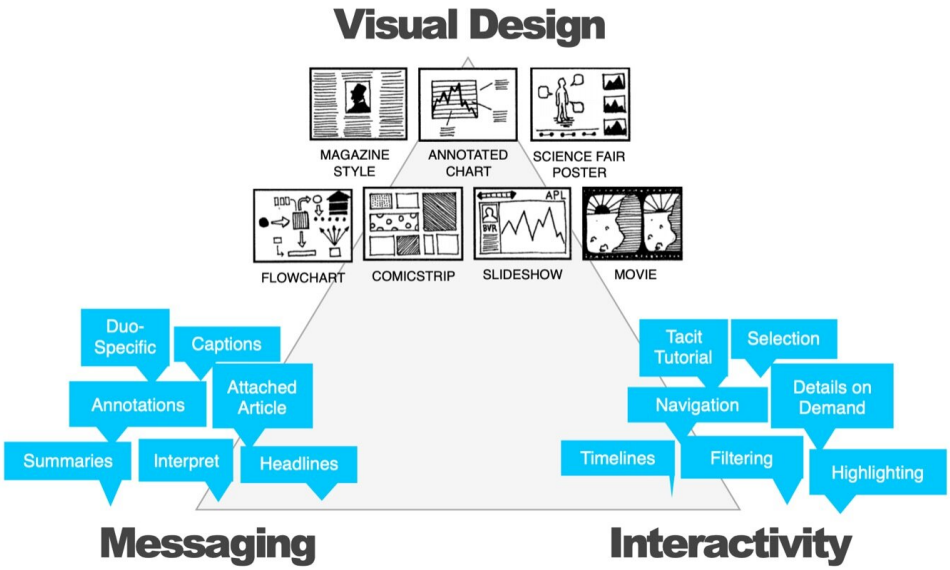

Fig. 3.1. Messaging and Interactive features in Visual design (adopted from (Gemigani, 2023))

Features: No special features but contains interactivity options, Texts and annotations options available, Easy to navigate UI/UX.

Strengths: Lightweight, easy editing, automatically connected visualisations, explaining, non-technical users, any size of the organisation (Gemignani, 2023) (*Juicebox Vs. Competition — Juice Analytics*, n.d.).

Toucan Toco

Dynamic features such as sharing, annotation, and drill-in story views allow to convey a holistic understanding of a subject. Using built-in story creation tools, the context can be set, a comprehensive glossary of terms can be created and compelling calls to action can be included to enhance the value of content (Gemignani, 2023). Turn insights into action by providing a step-by-step experience for users. The interconnection of charts, filtering and the ability to enable drill-downs enable a highly user-friendly flow of insights for the audience. It is possible to select the most suitable chart for a specific task with the help of the Guided Design feature. Built-in designs promote consistency to ensure users are not perplexed by any unexpected layout choices. Platform offers a versatile colour palette, from backgrounds to chart colours, text, and controls. To increase accessibility and readability it offers effortless translation to any language and manages it with customizable user rolls. It offers the possibility to create own data or fake data to simulate a production dashboard for conveying the message effectively. There is no code data preparation possibility. They have a mobile version available and there is no need for a desktop to work. (Contextualize Your Data With the Storytelling Studio | Toucan, n.d.).

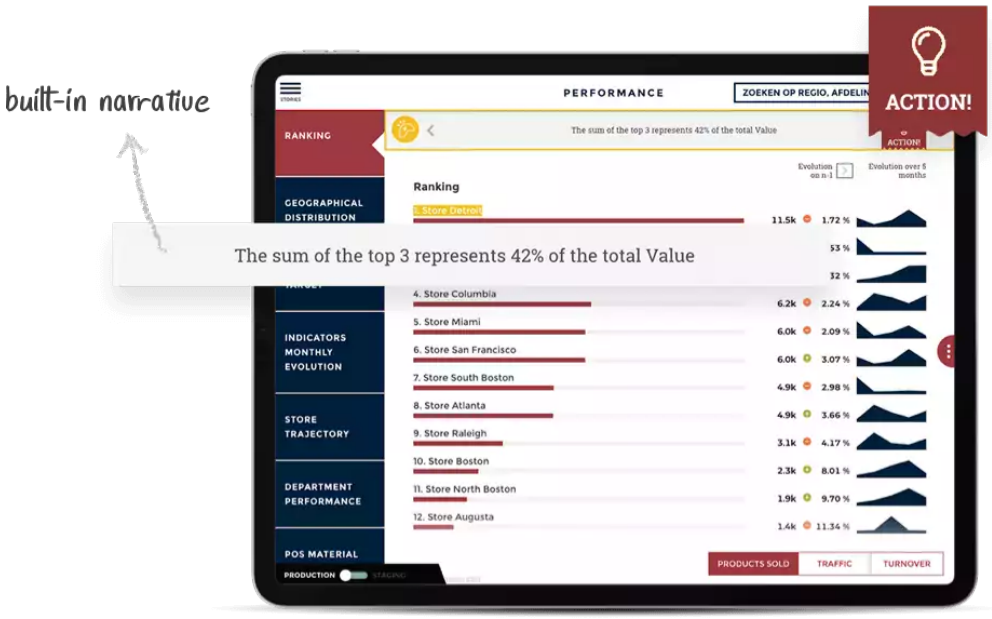

Fig. 3.2. Built-in narrative example from Toucan Toco (adopted from (Contextualize Your Data With the Storytelling Studio | Toucan, n.d.))

Feature: It has its own Storytelling studio which has a built-in Storytelling framework, narrative and guided design. It has a built-in collaborative environment

with features of Live chat, Sharing and Annotation, and a glossary. It has also integration with Microsoft Teams and Slack.

Strengths: Dashboard-style layout; user management features; sharing via presentation for sharing (Gemignani, 2023). They provide a speciality in embedded analytics. It has Toucan visual query which allows filtering, connecting charts, adding filters and enabling drill-down operations.

Nugit

Enhance data sharing experience with teammates and audience through automated tools for making dynamic data stories on the internet and through email (Gemignani, 2023). Say goodbye to manual processes and hello to engaging and seamless communication. Attractive design with the help of amazing text features makes this tool worth considering. They have a drag-and-drop story builder with an intuitive interface that helps build stories from scratch. Explore a vast selection of exclusive infographics showcasing English language summarises and highlights. ChatGpt 4 powered Natural Language Processing generate automatic insights (The Data Storytelling Platform. Put a Human Perspective on the Data That Drives Your Business., 2018).

Feature: It has Chatgpt-powered NLP (Natural Language processing). They claim to make data stories with zero clicks (automated stories). It can automatically suggest headlines, and summaries, highlight key strengths, and the possibility to deep dive into data with prompts even with auto-updating features. They are good for marketing stories as they support many social media data sources directly. Support of ChatGPT. It can generate insights in many languages.

Strengths: Live API integrations, report/email automation, automated natural language generation, infographic-style graphics, GPT 4 support (Gemignani, 2023).

Flow immersive

This dynamic platform specializes in creating captivating 3D visual displays that are seamlessly incorporated with engaging videos and Prezi-style presentations. Flow has provided users with the capability to craft, deliver, and distribute immersive and interactive data stories through an online medium, a pre-recorded video, or during a live meeting (Gemignani, 2023). It can be used for futuristic or immersive experiences but it generates complex visualisation and can’t be used standalone or need a bit of work.

Feature: It uses extended reality (XR), Flow Augmented reality (AR), 3D visualisations and Artificial Intelligence. It is good for large audiences. It contains futuristic features for interactivity and collaboration. They have a partnership with META. It is suitable for multidimensional data. It can filter and users can ask questions using Voice (microphone). Users can immerse themselves in the visualisation. Because of its exceptional features, people can connect anywhere using headsets. There is a 3D approach which makes it more fun to interact.

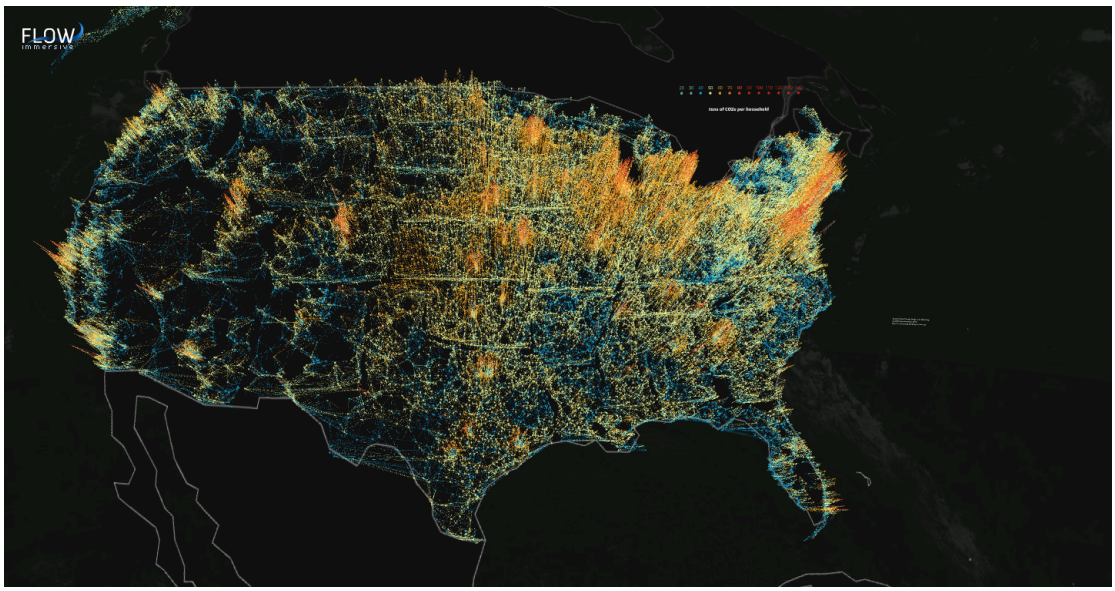

Fig. 3.3. Example of geospatial visualisation in Flow Immersive (adopted from (Flow - Immersive Data Visualizations, n.d.))

Strengths: Multi-dimensional visuals for showing many points positioned in 3-dimensional space, Augmented reality and AI and 3-D visualisation support (Gemignani, 2023).

- **Stand-alone visualisations:**

Flexible and stunning designs to help create personalized visualizations. It is used in web pages or online to support facts.

Flourish

Their racing bar chart is widely recognized, however, they offer a variety of other visually appealing choices. Present data with ease, whether on the internet or in person, using dynamic slideshows. Add a touch of excitement to the website by crafting captivating "scrolly" narratives without needing to code. Elevate content beyond typical visualizations and incorporate a range of interactive elements, from quizzes to carousels. They have a running bar chart which can be used to measure a metric with

time. Magic charts are used to provide multi-direction view to column charts making it more interesting. Animated 3D maps can be used for geospatial or locations (Features, n.d.).

Feature: They have provided various templates for various scenarios which is very useful and the user just needs to put data. There is a possibility of embedding. One of the creative ways to share the message. Famous for time series data using racing bar charts. Make it fun to interact using animated visualisations.

Strengths: Animated visualization, easy embedding, fine-grain configuration of visualizations, No coding or Software installation, free plan for education (Gemignani, 2023).

RAWGraphs

It's a simple open-source project that can easily create downloadable images for embedding in webpages. RAWGraphs states that 'it is the missing link between spreadsheets and data visualisation' (Gemignani, 2023). But it doesn't contain any special features for data storytelling. It is just another data visualisation tool that is hassle-free and makes fast advanced visualisation.

Strengths: Open source, lightweight editing, advanced visualizations, No software required, Free to use for everyone (Gemignani, 2023).

Qlik sense

Storytelling feature: They have a feature of snapshot which is just a screenshot at a particular incident in time and it supports editing with texts and annotations. But also there is a limitation of this feature as it is just a screenshot it does not change with data update. This feature is similar to Tableau story points but there is a possibility to navigate whereas it's just a screenshot.

Design over data: They are excluded from the categories list as the tools are not related to data storytelling they are from the design perspectives only. They don't have options to clear the requirements. None of the tools mention anything about data storytelling.

- **Data storytelling as a feature:**

One of the largest platforms in terms of business intelligence and visual analytics. Used for various purposes and data storytelling is used as a feature.

Tableau Story Points

One of Tableau's standout features is Story Points, which enables the presenter to create a dynamic story within Tableau by using a series of visualizations that

highlight different points or perspectives. These "story points" serve as individual dashboards, allowing for a more engaging and fluid presentation (Lee, et al., 2015). Story points provide an opportunity to create a narrative with data without any requirement of programming skills (Build Narratives in Your Dashboards With Story Points in Tableau 8.2, n.d.). One of the industry leaders in visual analytics, showing the early realisation of the need for data storytelling. The feature is yet to achieve its potential. Their attention is towards PowerPoint export options instead of data storytelling. A story is a sequence of visualisations, skillfully crafted to communicate valuable information. Every single sheet in the story is known as a story point, and it provides unique opportunities for users to actively engage with the story, interact with it, do exploration and ask new questions. (*Stories*, n.d.) It is easy to present as it consists of a presentation mode. Tableau's unique "Story points" feature allows for the creation of a sequence of slide-based dashboards, adding a new dimension to data storytelling. By utilizing their advanced capabilities, this tool greatly enhances the efficiency of data analysis through speedy data manipulation and the ability to easily publish dashboards. However, due to its technical and complex nature, Tableau is best suited for a specialized and deeply analytical audience. With dedicated time and effort, it has the potential to accomplish virtually any data-related task. Toucan mentions on their website that ‘In the past, many companies have relied on cookie-cutter Business Intelligence tools such as Tableau Software, Qlik, and Business Objects. However, these tools have a limited scope and are primarily geared towards data experts.’’ (Contextualize Your Data With the Storytelling Studio | Toucan, n.d.).

Feature: It has been mentioned in many resources. One of the earliest to mention data storytelling. Tableau data stories an another feature that automatically creates a narrative from the dashboard.

Strengths: Market leader, Huge customer base, first ones to use a storytelling feature.

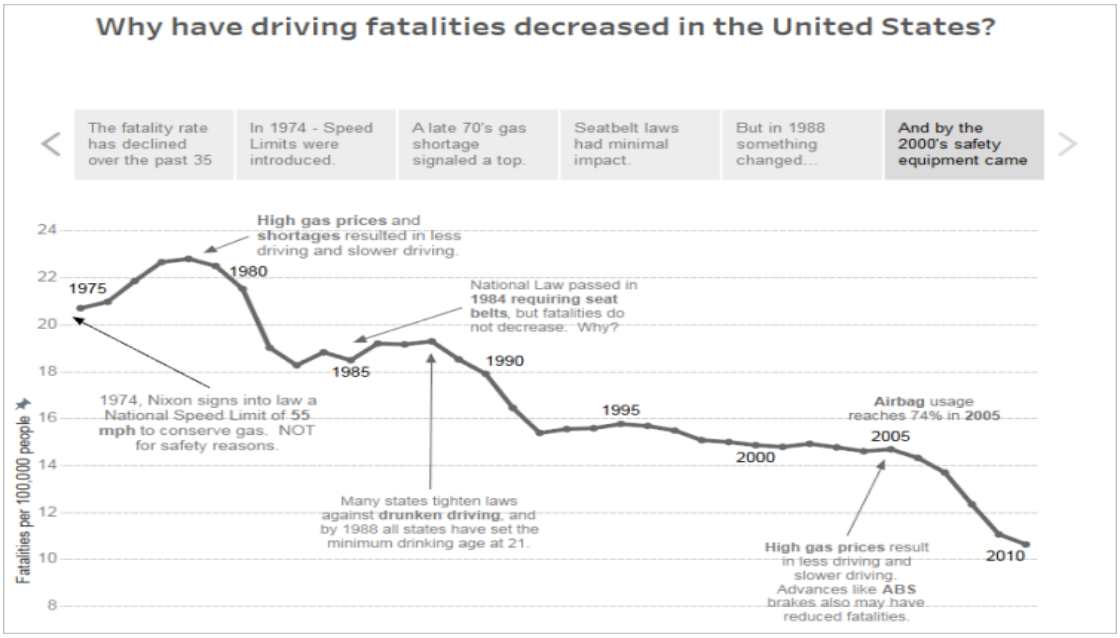

Fig. 3.4. Example of story points feature in tableau (adopted from(Stories, n.d.))

ArcGIS Story Maps:

Feature: It is compatible with GIS (Geographical Information System) which makes it easy to add places to the story, It provides context with places. Possibility to create custom maps also adding story points, pop-ups and arrows to enhance the story

Speciality: Effective for geographical data as it is a product of ESRI(Environmental Systems Research Institute) which specializes in GIS software.

Mapbox: It is similar to ArcGIS and can be a good tool for geographical data. It also has boundaries with multiple data layers of administrative, legislative, locality, postal, and statistical It also comes with the possibility to add live traffic data. It is a mapping tool but is also used for data analysis purposes. The story is broken down into chapters. It mentions data storytelling on its website. It can be used for geospatial analysis

Storymap js: Storymap JS uses a timeline and can be used for a time-series data analysis. It is used to create narratives that use maps to tell stories. It is an effective tool when it comes to explanatory data analysis.

Online tools e.g., Many Eyes offer a set of visualizations based on predefined templates. But they just have limited options for mouse over or clicks. They are suitable for simple stories. For more complex interactions and animations, people prefer libraries such as D3. But they require programming experience (Lee, et al., 2015).

SketchStory enhances the art of storytelling for presenters by captivating their audience's focus and fostering a sense of excitement, even though its capabilities for note-taking and participation are somewhat restricted (Lee, et al., 2013).

Gapminder (http://gapminder.org) utilizes animated presentations featuring bubble charts and scatterplots where point size corresponds to a specific value, such as population, to depict changes over time. While visually appealing and attention-grabbing, these animated transitions can hinder people's capacity to track trends (Robertson, et al., 2008).

Erato: It is a human-machine cooperative data story editing system' (Sun et al., 2022). Users are needed to provide a given number of keyframes which are used as the brief description of the topic and the structure of the data story. This system lets users create insightful data stories. It uses an interpolation algorithm to fill the gap between two keyframes (key data facts) (Sun et al., 2022).

- **Stories with words:**

The main focus point is data insights are converted to context, stories and sentences.

Sisense : (https://www.sisense.com/blog/sisense-narratives-telling-story-data/)

Features: It has a feature named Sisense Narratives. It uses NLG (Natural language generation) to find insights and convert them into easy-to-understand text (S. Team, 2022),

Speciality: Using AI to generate narrative insights just with the click of a button. There is an option to select sentiments and set the verbosity.

Phrazor: (https://phrazor.ai/docs/report-configuration/analyses)

Feature: AI-generated insights as well as narratives. It can be used as a plugin in Tableau and PowerBI with a Python Library as well. The analysis can be derived into 6 different types – Descriptor, Change, Compare, Target, Budget, and Trend (*Drag & Drop Analyses | Phrazor*, n.d.).

Specialty: Drag and Drop features with advanced functionality. Plus can be used as a plugin. It has a feature called Ask Phrazor where user can ask anything from data (*Drag & Drop Analyses | Phrazor*, n.d.).

Experimentation on goal, audience, size of audience, and communication method is done by user study. It is the output and can be measured by the people viewing the visualisation tools and judging it. This should require people and experts to observe it. Also, there are no features in the tools for this kind of scenario. It is more

related to narrative storytelling and is qualitative. It can use be used for validation, Rather it is needed to compare tools for storytelling so they can be divided into three categories: a comparison of normal visualisation tools with emphasis on storytelling suppose interactivity, story point features etc., tools for specific purposes like flourish for time changing or for multidimensional data. The third is the narrative part of the data storytelling. It comes under narrative data storytelling. This can together cover the whole storytelling process and then at the end, there can be recommendations for tools. The aim is to cover a broader spectrum of topics and include storytelling-specific tools.

This chapter includes the analysis of all the tools that mention data storytelling on their website or are mentioned in the resources. This chapter contains the Specialty and features of each tool for data storytelling. The tools are divided into three parts: Business Intelligence tools, tools that contain special features for data storytelling, and tools that help in creating the narrative for the story. Out of all these, tools are selected based on experimentation and justification criteria. Then the comparison is conducted.

3.2. Requirements for Data Storytelling Tools

The formulation of the requirements for data storytelling tools is based on a wide-ranging analysis of different techniques and concepts in the field of data storytelling. This analysis is derived from both theoretical frameworks and practical applications discussed in available literature and research. Moreover, it considers the characteristics and functionalities that data visualization tools present according to their official websites. These insights were consolidated to come up with a detailed checklist that can be used to systematically evaluate and shortlist suitable data storytelling tools. This part shall outline the major requirements that were identified through this thorough process.

- **Interactivity** - Allow users to interact with the story such as filtering or drilling down to specifics, giving more space for exploratory analysis, especially with digital platforms. Interaction with the data and self-exploring can enhance understanding and engagement.
- **Annotations** - Annotations are emphasized in many data storytelling methods. Effectively communicating insights just from conveying information is not enough. The use of annotations using texts, images or graphics provides additional information.

- **Appropriate visualization** - The selection of data visualization must align with the nature of the data and the story's purpose. Complex visualizations are not always a good choice, sometimes, simple are much more effective in conveying the message. Choosing the right visualisation: Each chart is used for a different purpose and varies with the data and message. It emphasizes different elements or relationships. Preferably use visualisations that are used more frequently and simpler to understand (Choosing the Right Chart | Explore Workspace, n.d.).

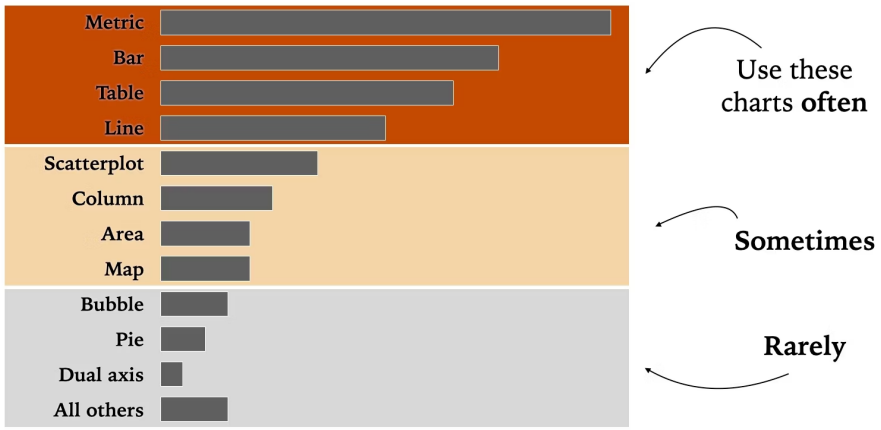

Fig. 3.5. Descending order of visualisations according to their usage (adopted from (Choosing the Right Chart | Explore Workspace, n.d.))

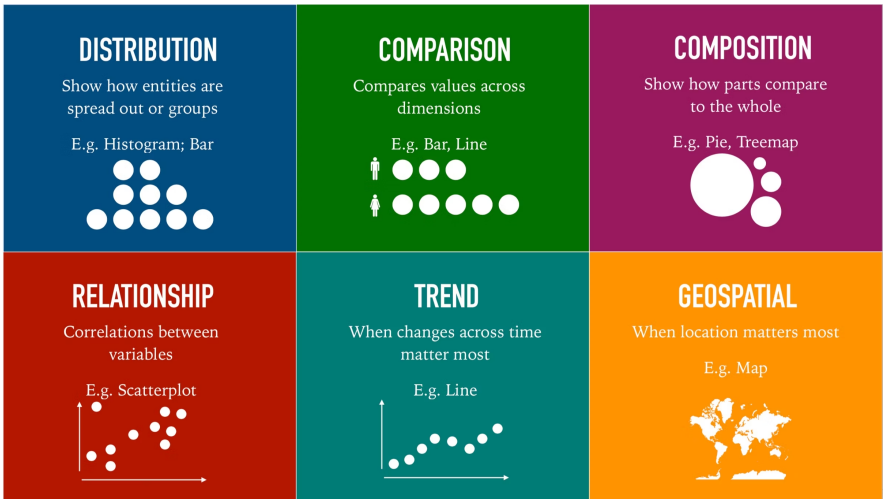

Fig. 3.6. Different types of visualisation categories (adopted from (Choosing the Right Chart | Explore Workspace, n.d.))

To find the right visualisation the user needs to first define the question and then decide a visualisation that can answer the question most effectively. Take the help of

keywords (eg. Trend, change, rank, compare, where, location (geospatial), most common) to identify the best chart (Choosing the Right Chart | Explore Workspace, n.d.).

- **Narrating Data Story flow** – A data story has a starting, middle and end. It is structured in such a way that the starting is with providing context, presenting the main insights, and concluding call to action leading the audience through a journey. Data storytelling tool should offer a variety of features, such as text, audio, video, and interactive elements, to enhance data stories. It should help in the formation of the narrative (Presentation, 2023).
- **Tailored for the audience** - Adjust the depth of the story depending on the audience's familiarity with the topic, their data literacy, understanding level and the complexity of it. Making data more relatable or reconnecting it to reality with examples helps to tell a more compelling data story. Providing a reference to something the audience is familiar with makes it relatable. Specificity is the soul of narrative. It brings information to life (*Relatable and Specific | Explore Workspace*, n.d.).
- **Accessibility** - Data stories accessibility to diverse audiences is paramount, ensuring considerations for colour blindness, screen readers and other accessibility tools. The ability to create visualisations on the browser itself and without downloading tool is a great benefit.
- **Feedback Mechanism and Collaboration** - Collaborating on data involves the active exchange, review, and enhancement of data narratives with peers, clients, or vested parties. A data storytelling tool that can efficiently publish, embed, export or dispatch data tales through diverse channels and platforms must streamline collaboration. Furthermore, it should grant the ability to oversee the accessibility, authorizations, and input provided by collaborators (Presentation, 2023). Providing a comfortable mechanism with the narrative for the audience to ask questions, give feedback, or opportunity to delve deeper into the topic can enhance the effectiveness of the story. Embedding and Collaboration features are mentioned in a lot of tools websites.

Embedding allows users to use the visualisation in multiple platforms. Collaboration supports engagement and enhances communication adding important feedback. Real-time collaboration saves time and increases user satisfaction.

- **Data Storytelling Guidance** – To understand data storytelling as a feature, a good tool should not only provide tutorials, examples, and feedback but also provide documentation of the tool and data storytelling (Presentation, 2023).
- **Clear Objective** - Each data story must have a clear aim or question it needs to address. A well defined objective ensures that the narrative remains focused and makes the visualisation productive. It could be either identifying a trend influencing decision making or explaining a phenomenon.

4. EXPERIMENTAL TOOL EVALUATION

This chapter contains the experimental part of this thesis. The tools analysed from the previous chapter are shortlisted according to the requirements checklist and the justification provided in the following subchapter. An example dataset is selected to perform experimentation, any dataset can be used as the goal is to compare the features and create the visualization. The selected dataset is about the AQI values of the countries which tells about the air pollution level in each country. The features selected above are taken and time to create visualisation, learning time and errors are also noted but it is mentioned in the appendix as it could be subjective. This chapter also includes the analysis of each tool while creating the visualisation and further challenges and limitations or strengths. Then at the end of the analysis recommendation for selecting each tool according to the features is provided. Results are then validated according to the real-life scenario.

4.1. Tools selection criteria

Many tools have strong support for textual narratives and notes but lack support for interaction. Tableau has a feature called Story Points, which allows a user to create scene-based stories made up of visuals or dashboards. Advances in web-based visual technology (e.g., the fast adoption of D3) allowed novel dynamic data-driven storytelling on the web. New York Times uses an internal tool named Mr Chartmaker to create and embed charts, which includes support for annotations and highlights, access to different screen sizes, and the ability to include interactive elements (Stolper et al.; 2016). With advancements in research and technology, nowadays people can create data stories with little or no programming skills even including animations and interactivity. Timeline JS allows users to create interactive timelines. Infogram and infoactive are used for creating interactive infographics (Stolper et al.; 2016).

There are limited research articles available regarding the comparison of data storytelling tools. Also, there are limited resources regarding the comparison of data visualisation tools. However, many websites have mentioned tools specifically for data storytelling. Also for renowned visualisation tools, the materials are easily available. The author understands there aren’t any experiments performed so the aim is to try to select the tools on the criteria that are necessary for comparison on data storytelling.

Checklist for selection of tools based on requirements for data storytelling:

From the above chapter of the thesis following requirements for the tools are deduced to select and compare different storytelling tools.

1. Interactivity: A large number of research papers have mentioned the importance of interactivity.
2. Texts, Images, graphics and annotations: A lot of research has emphasized the use of semantics to convey information. Alone data is not enough. Context provides additional detail to data.
3. Narration creation: This is the part that has a major role in differentiating data storytelling from data visualisation.
4. Sharing and collaboration: Sharing and collaboration are essential when it comes to the aim of communicating the information to the stakeholders.
5. Accessibility: Accessibility refers to the ability to reach everyone without any challenges of disability or language. It refers to the ability to change language and make data available for the screen readers.

Tools selection

1. There should be available resources to justify the selected tools.
2. Verify and choose tools that mention data storytelling on their website.
3. They should have special features related to data storytelling.

The author of this thesis has tried to cover a broad spectrum of tools according to the purposes and requirements. In the thesis, the focus is on interactivity and annotation, text and highlighting as is mentioned in many research papers, for example.

After going through the features of all the tools. There are 3 common features of data storytelling. It is possible to divide all the tools into three categories.

1. Business Intelligence tools
2. Interactivity
3. NLG (Natural Language Generation)

In this thesis, the emphasis is more on the interactive tools. Interactive tools are divided into three categories based on data, goals and features:

a) AR/VR 3D Visualisations

- Flow Immersive
- Virtualitics

b) Geospatial visualisations
    - ArcGis
    - Mapbox
    - Flowimmersive
    - Kepler.gl
c) Animated visualisations
    - Flourish
    - Living Charts
    - Tableau
    - D3.js

4.2. Measurement criteria

Visualizing information for various purposes presents a unique challenge in terms of evaluation, as it offers varying perspectives and is driven by different objectives (Pousman, 2007).

In his book Usability Engineering, Jakob Nielsen mentioned that task completion time is an important metric for evaluating software tools including data visualisation tools measuring their usability, efficiency, and satisfaction of the user. It talks about learnability (the time to learn the software and do the task), efficiency (high level of productivity), Errors (low error rate) and overall satisfaction as important. The author mentions that the time taken to measure certain test tasks is the best way to measure efficiency. One of the important factors to measure the usability of the tools is the number of errors encountered. When describing errors within a system, they are commonly defined as any actions that do not result in the intended goal (Nielsen, 1994).

Segel & Heer have mentioned in their article Narrative Visualisation that Messaging involves using text to provide observations and explanations about images. Often, this text is presented in familiar formats, including headlines, captions, labels, and annotations (Segel & Heer, 2010). By incorporating interactivity, the viewer can manipulate the visualization in multiple ways. From navigation buttons to hover highlighting, hover details-on-demand to filtering, searching to drill-down, and zooming to time sliders, there are numerous options for interactivity in narrative visualization (Segel & Heer, 2010).

To stand out, visual analytics tools need to bolster the seamless and flexible usage of visualizations at speeds that are in sync with the promptness of human consciousness. In the article, Interactive Visualisation of Large Datasets authors talk about the problem of slow response time with Interactive visualisations with large datasets. It is mentioned that When query response time exceeds a few seconds, interaction becomes cumbersome, and displaying all data points simultaneously can overload the visualization, leading to confusion. Also, interactive visualisations are used for exploratory analysis a lot of times which includes showing all the data. This leads to overloading the visualisation (Godfrey et al., 2016). However it is not very useful in this experimentation.

In the article Measuring Effective Data Visualisation author Ying Zhu has provided an approach to measure the effectiveness of data visualisation based on three principles: accuracy, utility and efficiency. Utility can be measured by creating benchmark tasks using a benchmarked data set and a utility score is provided based on the number of tasks completed. Efficiency is measured using task completion time, eye movements and learning curve. These are the quantitative measures. He also suggested qualitative measures which include user response using interviews, observation and expert-novice difference ("Advances in Visual Computing," 2007). The author of this thesis has not taken a separate number of tasks rather the full process of creating the data visualisation is taken into account. The time taken to convert raw data into usable form is not considered as it is not dependent on the tool and does not affect it in any way. The time taken and error can together give an idea about the Ease of use and efficiency of the tool. It is important to have some quantitative benchmarks to compare. Yet, these metrics are subjective and can be affected by the author's knowledge and abilities. However, the author is more focused on the relative time and big differences. Also, the more important part of the thesis is the table which analyses the support of data storytelling features in data visualisation tools. The features and techniques are selected from the literature review of the two major contributing papers in the field of data storytelling. The list of the features is updated with the new features and techniques included in the web pages of the data visualisation tools.

Metrics chosen for experimentation:

1. Quantitative:
    - Time taken to create visualisation

- Tool learning time
- Errors

2. Qualitative

Data storytelling features in visualisation tools

These metrics will be used by the author in the upcoming chapter of this thesis.

4.3. Dataset and Laptop Specification

Link for dataset: https://www.kaggle.com/code/azminetoushikwasi/daily-aqi-geospatial-analysis

This dataset from Kaggle – 'Daily AQI Geospatial analysis' is used as a sample dataset for the experimentation of the tools. It is a reputed dataset with 7,700+ views. It daily records the Air Quality Index (AQI) values of different locations. The AQI value of a location provides information about the Air quality of that location. This data can be used both for Animation and geospatial datasets. As it contains information about the particular values every day for all the scheduled locations. It can be used to show many types of visualisations and for different goals. Also, this is an important topic as it is related to the health effects. Usually, it ranges from 0 to 500+. 1 being the least polluted while 500 or more is highly polluted. The lowest value it contains is 1 and the highest is 963. The limitation of this dataset is that it does not contain information about the latitude and longitudinal data. The additional information is added from external sources which are in point form.

There are multiple types of locational information possible in geographical data. Point (latitude and longitude coordinates), polygon (coordinates of the area), line data (linear), S2 layer, Icon, Cluster, Hexbin, grid, H3, heatmap, arc and trip layer.

The author has chosen point data as it is commonly used and easily available. It is supported by many tools. Also, to add polygon and other location information it was not easy to find accurate latitudinal and longitudinal data.

Author of this have used the Macbook Air M1 Chip (2020) model for performing these experiments. For on-browser tasks, Google Chrome (Version 123.0.6312.58 (Official Build) (arm64)) is used.

RAM: 8 GB

Storage: 256 GB

4.4. Analysis of Tools

In this subchapter, the author will compare the tools along with the methods discussed in the above chapters. Under the visualizations, there is a link provided that takes to the external webpage where it is possible to interact with the visualisation.

4.4.1 Flourish

The first tool used is Flourish, with version v15.8.0. It came into the limelight during the COVID-19 period and is now one of the most used visualisations of flourish. The plan is to create a Racing Bar chart. It is used to show temporal data as an animation. It provides the use of templates that allow the user to easily create visualisations in just some clicks. It has a feature of Auto-set column bindings. When the user uses a template and uploads data. It automatically interprets data and selects the appropriate column for the visualisation. It is responsive for different screen sizes. It has a help option for each visualisation. There is a possibility to change the default colour by colour overriding feature. It also has an accessibility option with a screen reader mode which makes it accessible to readers with disability. It requires to put data data directly to the editor, there is no possibility of manipulating, clean and transforming data. The data is prepared using an Excel pivot table and then copy-pasted to the Flourish. It is available in the browser and the UI is easy to navigate. It allows to filter data using labels. This allows the user to filter using categories as done in the visualisation. It offers many other storytelling features like scrollytelling and embedding directly, but it is in the pro version.

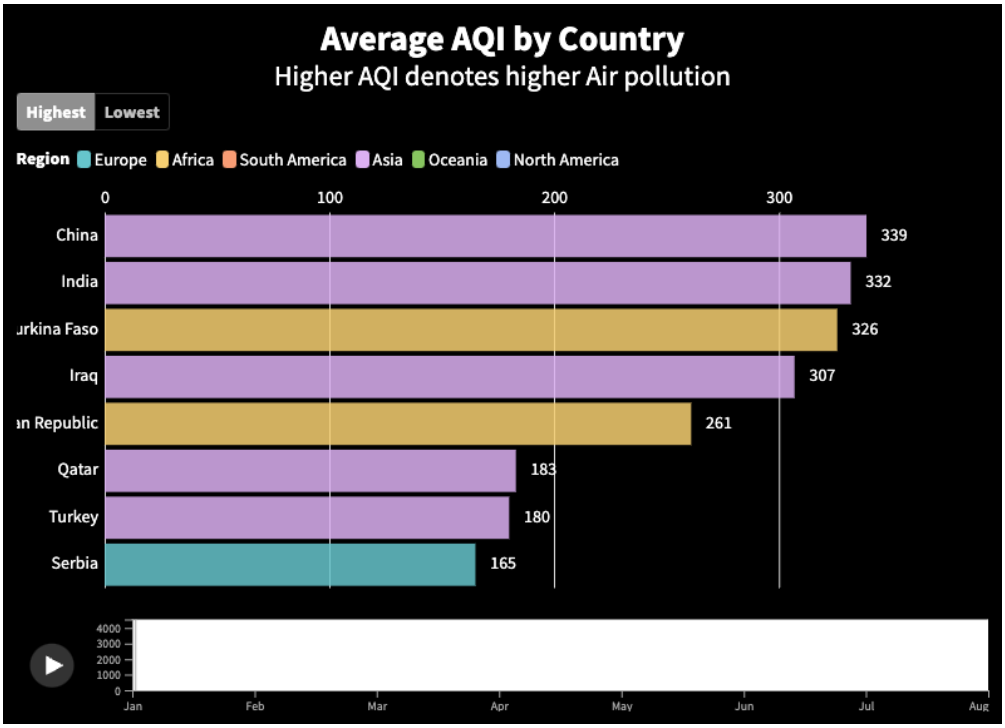

Fig. 4.1. Racing bar chart created on Flourish

Link for visualization: https://public.flourish.studio/visualisation/17274044/

Limitation:

Flourish templates have data size limitations. If the dataset is larger than the given limit. The data can fail to load and crash. The amount of data it can handle is quite low. However, if data is filtered or a certain range of data is needed, it can be used.

Template	Data limit
Line, Bar, Pie	Around 5,000 rows (10,000 if using a filter)
Bar chart race	Around 1,000 rows
Survey	Around 10,000 rows (20,000 if using a filter)
Scatter (in WebGL mode)	Up to 200,000 rows
Network graph	Up to 10,000 rows
3D map	Up to 2,000,000 rows
GeoJSON files	Up to 5MB
Data explorer	Up to 1,000 rows

Fig. 4.2 Data size limitation in Flourish on Templates (adopted from (*How Much Data Flourish Templates Can Handle*, n.d.))

4.4.2. Living Charts

It has a similar interface to Flourish. But the UI and features are not as good as Flourish. It has a watermark and ads. Not as many features but it still looks cluttered. No feature for a timeline. It offers only a desktop view. It takes less time to learn as it is similar to flourish. If someone has experience with flourish, then it is easy to use this and vice versa. It offers to export for free, as well as a video with an option to select the fps. That is unique to flourish but nothing else. In terms of features, functions, and UI it is better than it. It only contains 4 types of Animated Charts. The author struggled with working with categories which is why it took more time to create visualisation than learning the tool itself. The visualizations don't look similar because both tools contain different features.

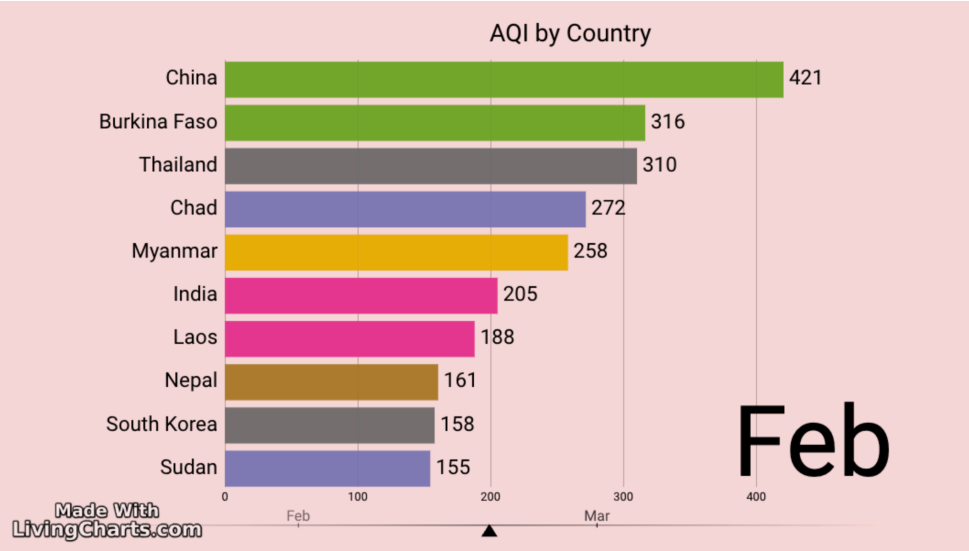

Fig 4.3: Visualisation created on LivingCharts

Link for visualization: https://livingcharts.com/edit/vb9SUJuaaSBAIvvuyJzd

Limitation:

Living charts create cluttered visualisations that can distract the audience from the goal. Flourish is a better tool in all aspects except it provides a video export option. But that can also be done using screen recording.

4.4.3 Tableau

Tableau is a tool that is mentioned in most of the resources as it was the first to add data storytelling features to its functionality. Tableau has in Total of 3 features: Data Story points, Animation, and AutoCaption generation. With the possibility to add animation to many charts. There is no specific visualisation like a racing bar chart. A racing bar chart can be created by adding animation to a horizontal bar chart.

It has a feature to help users select the right chart according to the measure and dimensions. It is helpful for beginners to start the process of creating data visualisation without getting affected by the number of features. The author had already used Tableau previously but did not learn it through the documentation. This can create variables in learning time and could be different for someone who might use it for the first time.

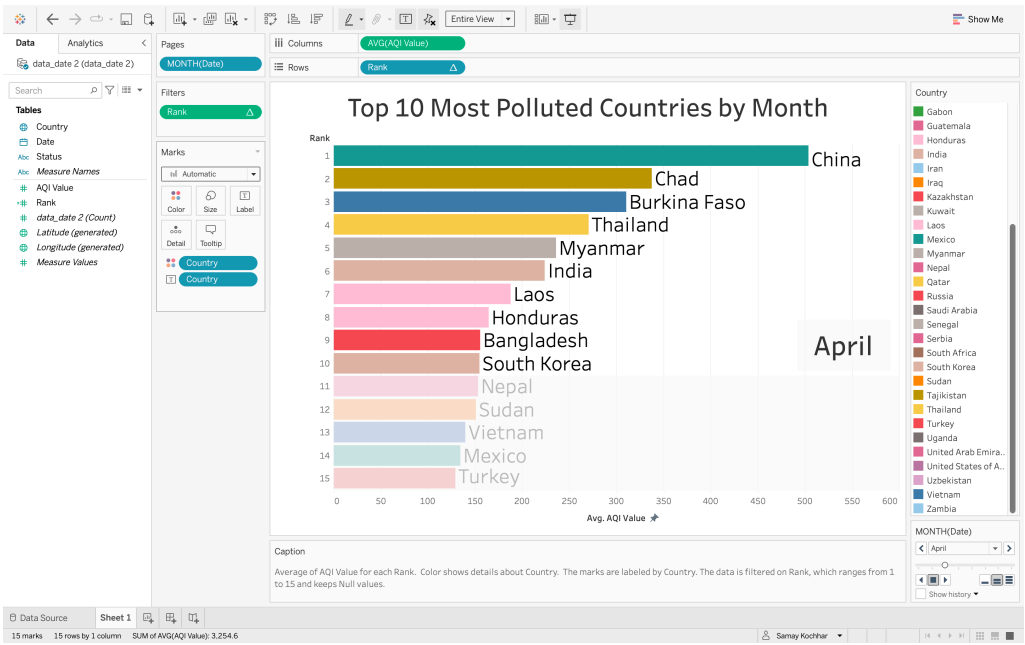

Fig 4.4. Animated bar chart created on Tableau Public

Link for visulization

https://public.tableau.com/app/profile/samay.kochhar1226/viz/Top10MostPollutedCountriesbyMonth/Sheet1?publish=yes

Limitation:

The author encountered some errors while uploading data. One ended up even closing the software. It was due to the Null Values. It led to the loss of full progress. Also, there is no AutoUpdate feature for automatic saving.

4.4.4. D3.js

The author has included D3.js as it is mentioned in many resources. Flourish, Tableau, and many other tools use D3.js as their language. Users can run d3.js on the Observable platform in the form of notebooks like Jupyter Notebook. On Observable there are multiple templates already provided and ready to use but it is advised to understand the code first unlike flourish and living charts where it is possible to make

changes to the table directly. It allows users to add, clean and transform data that is not there in both above-mentioned tools. As D3.js is used to create a lot of tools, it is possible to make custom animations, add annotations, highlights etc. But it requires a great amount of knowledge. It consists of many charts focussing on animation and interactivity. The author has used a template by Mike Bostock (https://observablehq.com/@d3/bar-chart-race?intent=fork) transformed data to a similar format and made necessary changes according to visualisation.

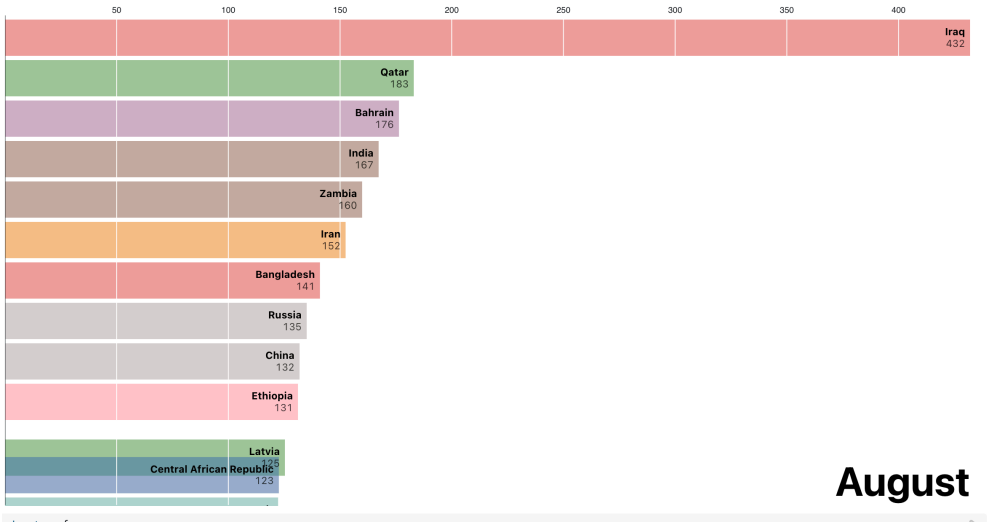

Fig 4.5. Animated bar chart created on Tableau Public

Link for visualization: https://observablehq.com/@samaythesis/bar-chart-race

Limitation:

D3.js is a tool that is recommended when needing customisation else it is preferred to use flourish. As it is a whole new language, users need to spend a high amount of time learning it.

4.4.5. Arcgis

ArcGIS is one of the leaders in the analysis of geographical Data. It offers a wide variety of geographical charts with different colour filters. It provides the user with several Styles when a data layer is added. Users can manipulate the data to fit the dataset according to the visualisation.

The visualisation is created on ArcGIS online as the desktop or student version requires a paid subscription. ArcGIS online can create good visualisations for geographical data. It provides various features and charts but each of them includes

different functions. If the user chooses Color and amounts, there are 2 styles each emphasizing Color and size respectively. If the focus is on the colour side, then it is not possible to classify the data or add a filter but if the user chooses size, then there are not many colour options. The user must select manually. Each time there is a change in the chart style, the previous chart customization is lost. Also, the Open Map viewer option is used, and complete progress is lost (remember to manually save the progress at each step). At login, the user gets 400 credit points to use. It requires more than that to add big data without longitudinal and latitudinal information. The missing information about latitude and longitude values is added using external websites. It allows to add add popups, annotations, highlights, and filters. However, the author could not add titles or filter data using the classification. It offers storytelling functionality, but it is in the paid version.

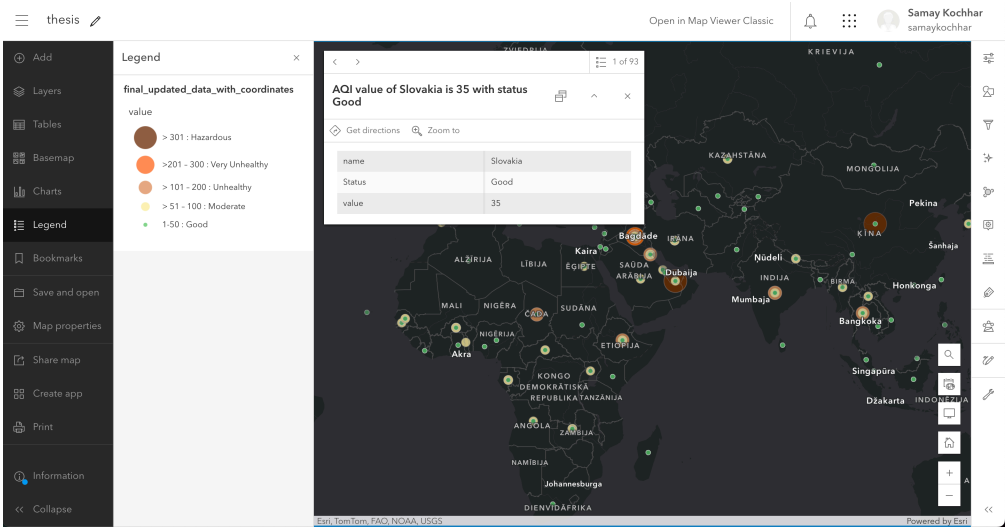

Fig 4.6. Map Viewer Classic interface in ArcGIS

4.4.6. Mapbox

Mapbox offers various functionalities for data storytelling. It uses Javascript (JS) as its language, and it has a separate Mapbox JS library which adds interactivity to a chart. To load data, it needs similar data as ArcGIS with latitude and longitude information. It is also possible to add time series data with location info. It gives the possibility to show animation visualisation of geographical data change over time. But for the selected dataset choropleth map is the best option for explaining and comparison. It has many types of map styles and layers. Users can also add multiple layers, 3d lightning and terrain. It has 8 types of projections of the earth but not for

specific locations. It has an advantage over the other tools when it comes to world data. It has smooth transitions while changing the camera perspectives. The process of creating visualisations in this tool is divided into several steps. First is saving the dataset in the Tileset format. This can be done using a Tileset viewer. It took 48 seconds to load and create a tileset of the dataset. Once the tileset is created, it is added to Mapbox Studio for preparing the visualisation.

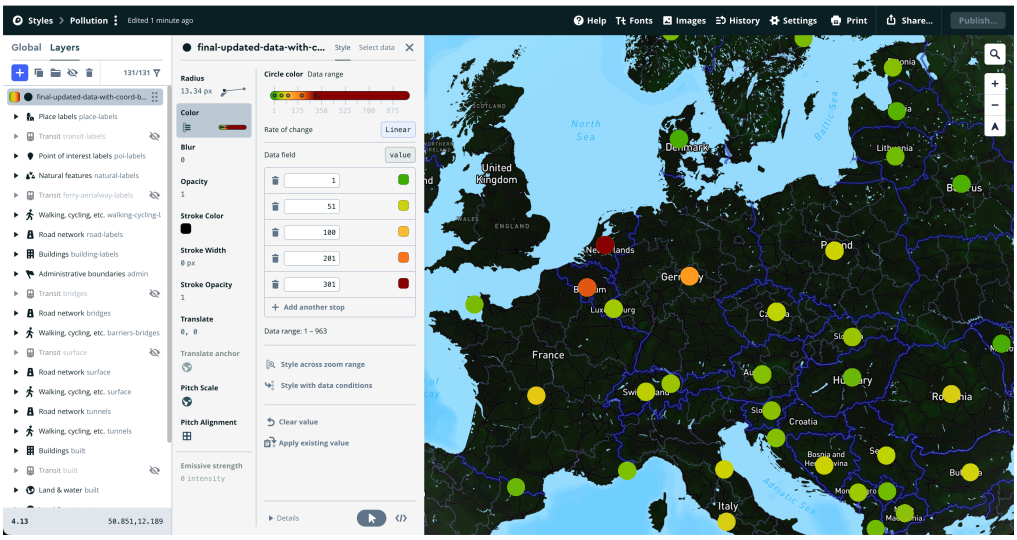

Fig 4.7. Visualisation created on Mapbox

4.4.7. Kepler.gl

The experimentation is performed on version 3.0.0. After uploading the dataset, it automatically detects the dataset type. It is possible to filter the data easily and tooltips provide additional information. It offers limited features, but they are very easy to find and use. It has a Geocoder function that allows the search of any coordinates and location. There are 7 types of map layers, 7 types of map styles and 12 types of data layers available to style the visualisation. There is also a possibility of adding a custom base map. The author had to make the graph 3 times as while saving the progress gets lost. It is because the link redirects to the official website when the save button is pressed. If visualisation is downloaded as an image, it doesn't show the data points. There are multiple language options to display the data.

Fig 4.9. Data Visualisation created on Kepler.gl

Link for visualisation: file:///Users/samaykochhar/Downloads/kepler.gl.html

4.4.8. FlowImmersive-

Flow Immersive is one of the market leaders in data visualisation tools when it comes to handling multidimensional data. It has the support of virtual reality. The author has used the paid version. He requested it while receiving the demo. Fortunately, It was provided with a paid subscription till July. It has many features related to storytelling. But it takes time to learn this tool and get used to the features. Users can create stories like scrollytelling and there is a possibility to communicate insights using presentation flow. It is possible to give access to the people and they can interact with the flow and immerse themselves in the visualisation. Users can create tooltips, pop-ups, and annotations easily. It automatically saves the project after each change, and this is very useful as the user doesn't need to worry about it as was the case for the author in many tools. It also automatically detects datatype. There are multiple templates to choose from, also to create custom templates as well. It has 2D and 3D options and allows one to change camera angle and look from multiple perspectives. There are 9 location maps and 3 styling options.

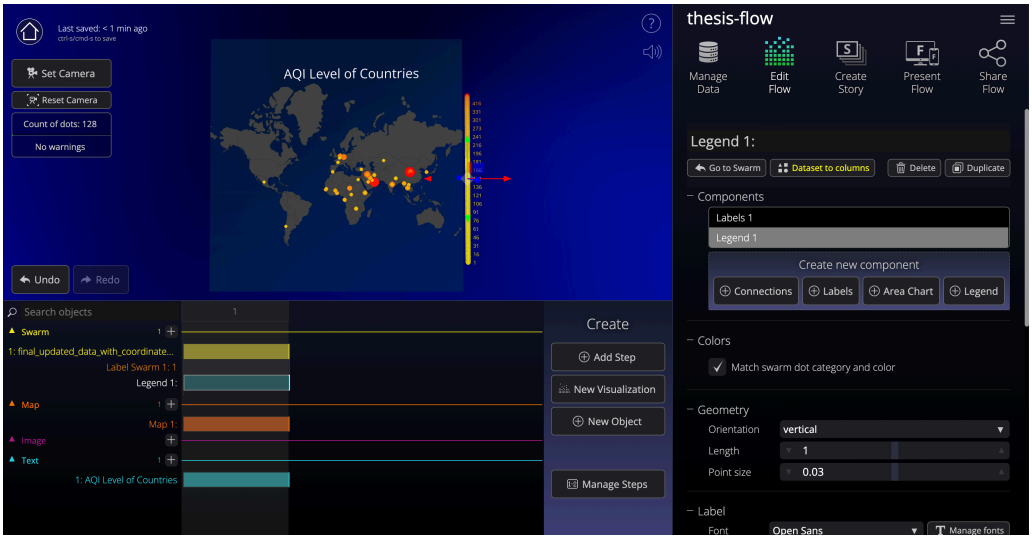

Fig 4.10. Flow created in Flow Immersive

Link for visualization: https://app.flow.gl/flow/lw0s9u7y

4.5. Feature Analysis

The feature analysis done in this bachelor thesis is the most essential part of this as it objectively analyses the support of various data storytelling features and techniques in the above-mentioned data visualisation tools. The features and techniques selected for the comparison are selected from the recognized resources in the field of data storytelling. This comparison is qualitative since there is no way to measure the effectiveness of these features and how well a tool supports them. The table below shows the ability of the tools, highlighting key functionalities such as textual annotation, dynamic queries, and geographical navigation. This helps the user to select the tools that align with the narrative, goal and understanding of the audience. This analysis is important because it enables the user to formulate the narrative and allows the audience to make well-informed decisions. The green colour in the table signifies the support of a particular feature, the red colour signifies that the tool doesn't support that feature, whereas yellow shows that either the tool requires a paid subscription or in the D3.js case high knowledge and amount of time.

Table 4.1

The analysis of support of storytelling features in Visualisation tools

Tools	Flourish	Living Charts	Tableau	D3.js	ArcGIS	Mapbox	Kepler.gl	Flowimersive
Textual annotation								
Highlighting elements								
Labelling								
Animation								
Dynamic query								
Strollytelling								
360* viewing angle								
Embedding visualisations								
Tooltips								
Color scheme								
Next/previous button and arrows								
Breadcrumbs								
Geographical navigation								
Timeline								

This table shows the comparison of data storytelling tools based on support for storytelling features.

4.6. Recommendations

This subchapter contains the recommendations that are formed after the comparative analysis of all the tools. The 8 tools compared in table 4.1 are based on 14 selected features. To select the tools for the scenario, it is necessary to understand the question and the aim of the visualisation. Then, the necessary features are selected which are to be included in the visualisation. The support for the storytelling features is checked from the above table together with the time to create the visualisation and learning time to determine the tool selected. The author of this thesis provides recommendations that should be considered in mind before selecting the tool. These recommendations are the following:

Purpose of visualisation - To create an effective data storytelling visualisation, it is necessary to understand the goal of the visualisation. Whether it is explanatory or exploratory. Presentation flow, Scrollytelling and Timeline can be good features for explanatory analysis whereas animation, 360 viewing angle and navigational features help in exploratory data story. One is used for providing proof and evidence and the other is used to explore and find patterns and insights. For explanatory data labelling, legends, and highlighting elements are not that important. If the goal is to explain Tableau is better as it contains highlighting, and the dataset can also be cleaned and transformed while Flourish is a better option for exploratory analysis. D3.js can also be used to create custom visualisations and data stories.

Understanding the Data and Types - Data will define which visualisation can be made and which tool should be selected for creating it. Multidimensional data type is best worked in flow immersive. ArcGIS is better suitable for geographical data and navigation compared to Mapbox and Kepler.gl. As many tools are made using d3.js, it is possible to add custom features by learning and writing code. It is highly suitable for creating interactive visual narratives whereas Living Charts and Kepler.gl have limited feature support. Evaluating the nature and complexity of the data helps to understand the advantages and limitations. Flourish struggles with handling complex data.

Necessary features - It is necessary to understand which feature and process is important for the user. Flourish, Tableau, D3.js and Flow Immersive are more feature-rich and cover a broad spectrum of data storytelling requirements like adding legends, filters and annotations as compared to other tools. For highly interactive web-based

stories D3.js is ideal. If the user needs additional features or has the skill and doesn't want to pay for any subscription to the other tools it is recommended to use D3.js.. However, it takes more time to learn and create visualisation as features make it more complex. Flourish, ArcGIS, and Flow Immersive have scrollytelling and embedding features.This makes it easier to communicate insights.

Time to create visualisation - Some tools offer quick and easy setup with pre-built templates. The time to create the visualisation also depends on how much time it takes to learn the tool. D3.js and flow-immersive take a lot of time to learn and create the visualisation as compared to others. A lot of business users might not prefer putting that much time and effort into learning new tools as they tend to spend more time making decisions. They could use Tableau and flourish as they are versatile and offer a broad spectrum of data storytelling features. They are easy to use and have a steep learning curve. Whereas living charts, kepler.gl and Mapbox are easier to learn and create visualisations. They also provide templates to make the process smoother.

Templates vs. Customization - Many tools mentioned above can create interaction and animation but are easier to use in Flourish. It has templates that are smooth to operate and work well. Tableau allows all the processes of data analysis as well as generating data visualisation to data presentation. This covers the whole process and it is helpful in business problems. Living charts use templates with the least features. D3.js consists of both templates and provides room for customization. A lot of tools use Javascript at the backend hence, if the user possesses the skill then it is possible to create custom visualisation. In geographic data, ArcGIS and Mapbox can be used. They also provide templates and a little bit of customization. Flow immersive includes a good amount of features and is complicated.

Compatibility and Cost-Flow Immersive is a futuristic tool that can require AR/VR to take its full potential. But to take full advantage of this tool, there is a requirement for the necessary equipment to support it. This can add additional cost to the budget. Many companies have set tools for their all tasks in that perspective Tableau and D3.js can be used. Embedding visualisation is

4.7. Validation

Scenario: For the verification scenario the aim is to take a real real-life Business scenario. Forage provides Virtual internship opportunities where various companies

host their programs giving real-life job simulations. The goal is to do the task given by TATA. It covers the full process of data analysis from preparing questions, understanding the dataset, finding insights, to giving presentations. The dataset is also provided. The first task is to understand the dataset and prepare questions to ask the CEO and CMO of the company. The dataset comprises 5,41,910 rows and 8 columns including Invoice no, Stockcode, Description, quantity, InvoiceDate, Unit price, CustomerID and Country. They have provided the requirements and questions. The task is to fulfil the requirements and answer their questions using visuals. From the insights generated, they need to create an expansion strategy. 'The executives want to analyse the trends and the breakdown by different categories so that they have clarity on how the revenue is being generated and what are the main factors affecting the online store' (TATA) (Forage, n.d.).

The first task is to clean the data by removing negative values from price and Quantity. For this task, both Tableau and Flourish can be used since the goal is to explore and we have time series data. Flourish can be preferred for the features but since it has a large amount of data. Due to flourish limitation Tableau is used for this task. Templates can be used to create visualisations easily. In the end, scrollytelling can also be used to explain interactively. To solve this question, need to create a revenue column.

Data type – Time series data

The time given to solve the tasks is not explicitly divided but the time to solve this course is 4-5 hours so let's assume it is the time.

Goal: Exploratory analysis and find insights.

Features required: Dynamic query, Embedding visualization and navigational features.

Under the visualisation link is provided to explore it.

The four questions that must be solved are:

Question 1

The CEO of the retail store is interested in viewing the time series of the revenue data for the year 2011 only. They would like to view granular data by looking into revenue for each month. The CEO is interested in viewing the seasonal trends and wants to dig deeper into why these trends occur. This analysis can be helpful for the CEO to forecast for the next year (*Forage*, n.d.).

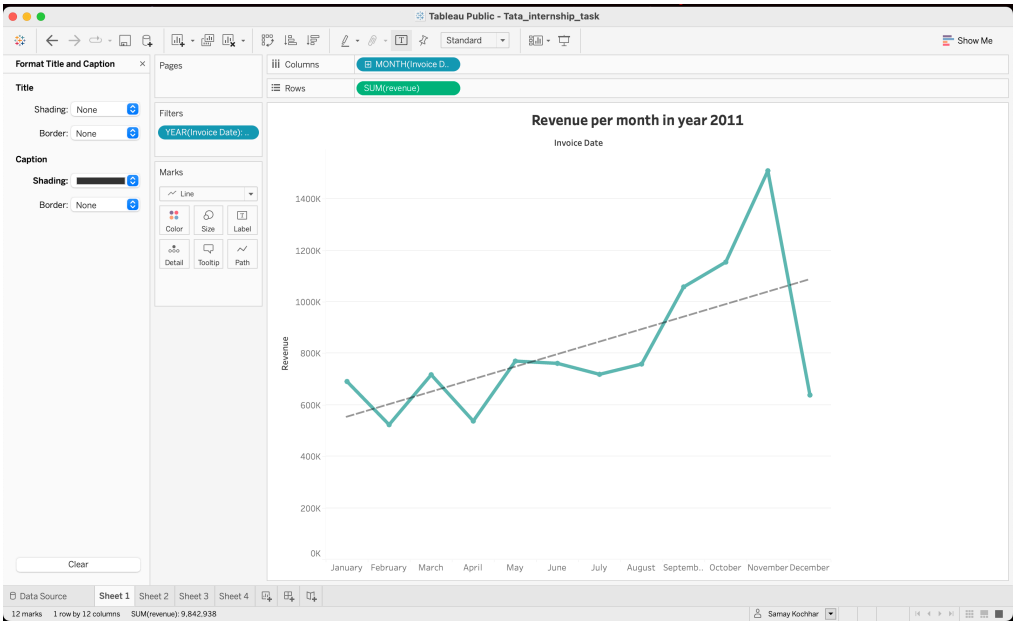

Fig 4.11. Line chart showing revenue per month in 2011

Link for visualisation:

https://public.tableau.com/views/Tata_internship_task/Sheet1?:language=en-GB&publish=yes&:sid=&:display_count=n&:origin=viz_share_link

Question 2

The CMO is interested in viewing the top 10 countries that are generating the highest revenue. Additionally, the CMO is also interested in viewing the quantity sold along with the revenue generated. The CMO does not want to have the United Kingdom in this visual (*Forage*, n.d.).

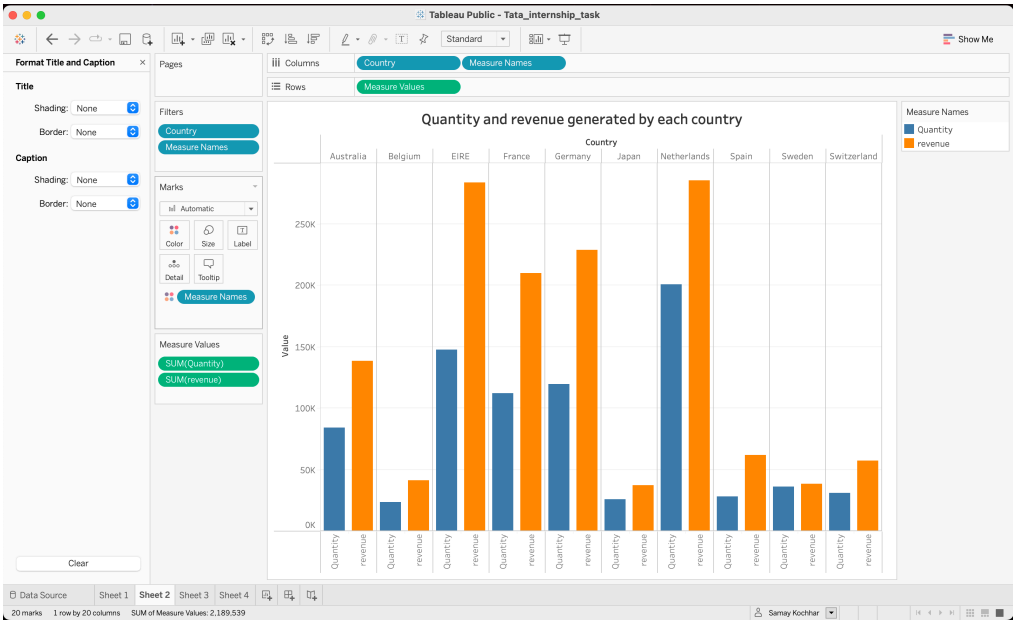

Fig 4.12. Top 10 countries generating the highest revenue

Link for visulisation: https://public.tableau.com/app/profile/samay.kochhar1226/viz/Tata_internship_task/Sheet2?publish=yes

Question 3

The CMO of the online retail store wants to view the information on the top 10 customers by revenue. They are interested in a visual that shows the greatest revenue-generating customer at the start and gradually declines to the lower revenue-generating customers. The CMO wants to target the higher revenue-generating customers and ensure that they remain satisfied with their products. (*Forage*, n.d.)

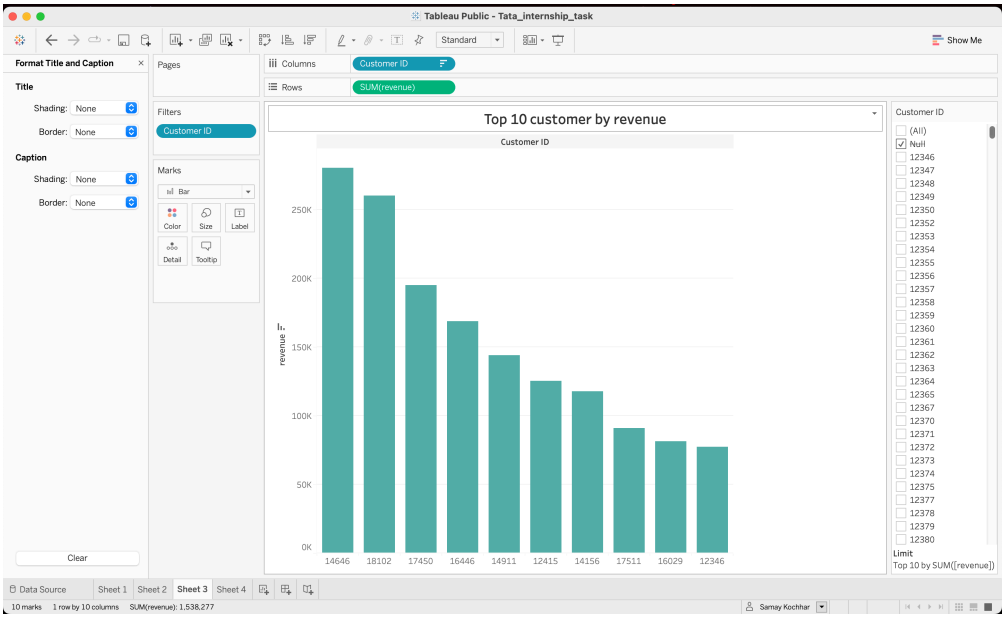

Fig 4.13. Top 10 customers by revenue

Link for visualisation: https://public.tableau.com/app/profile/samay.kochhar1226/viz/Tata_internship_task/Sheet3?publish=yes

Question 4

The CEO is looking to gain insights into the demand for their products. They want to look at all countries and see which regions have the greatest demand for their products. Once the CEO gets an idea of the regions that have high demand, they plan to initiate an expansion strategy this allows the company to target these areas and generate more business from these regions. The company wants to view the entire data on a single view without the need to scroll or hover over the data points to identify the demand. There is no need to show data for the United Kingdom as the CEO is more interested in viewing the countries that have expansion opportunities (Forage, n.d.).

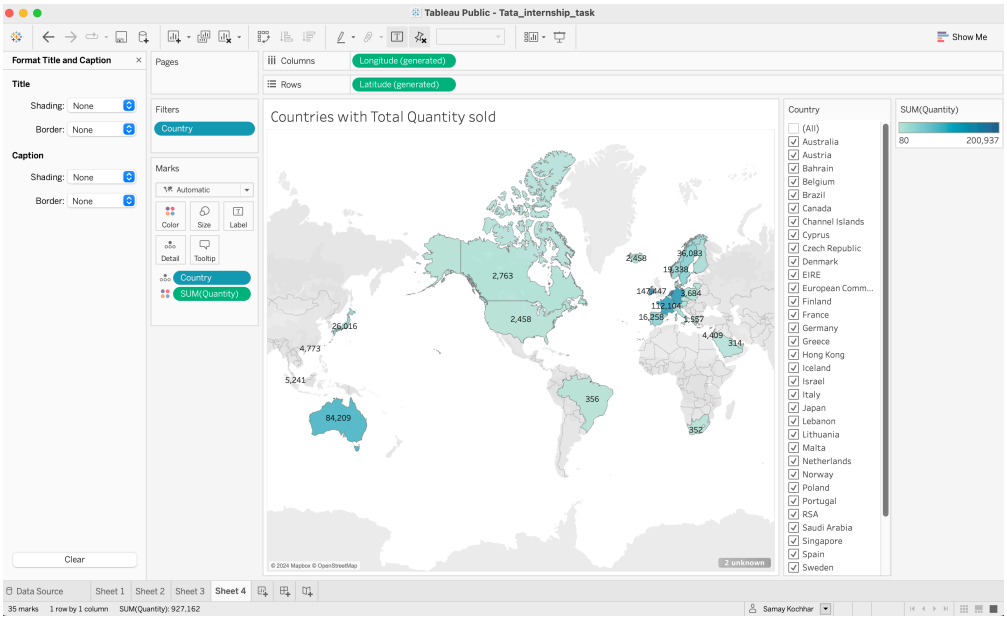

Fig 4.14. Geographic map of countries showing quantity sold

Link for visualisation: https://public.tableau.com/app/profile/samay.kochhar1226/viz/Tata_internship_task/Sheet4?publish=yes

CONCLUSION

The comparison and analysis of data visualisation tools on meeting requirements of data storytelling provides important information about the support of data storytelling features and techniques present in the current data visualisation tools. As mentioned at the start of this bachelor thesis, this field is still developing, there are a limited number of resources that include quantitative methods for comparison of these tools. This thesis is written by conducting extensive research on this topic over the past year, covering most of the available literature in data storytelling. The article 'Narrative visualisation: telling stories with data' by Segel & Heer and the book 'Storytelling with Data' by Nussbaumer Knaffic are two highly cited resources on this concept. Stopler et al have analysed 45 data storytelling examples combined with 20 techniques suggested by Segel & Heer, together with the research done by going through the webpages of many data visualization tools containing information about data storytelling forms the metrics for comparing features and techniques. Given the limited existing information on tools, the research included reviewing reputable websites and tool documentation.

After going through the features available in the tools, they are divided into three categories: Business Intelligence tools, Interactivity and NLG. The use of interactivity in data storytelling is a find as a crucial element mentioned in multiple papers. It is also emphasized in this thesis. The tools are then categorized and shortlisted using the requirements checklist and justification criteria. A requirement checklist is prepared by analyzing various factors and methods of data storytelling. A total of 8 tools are included in the final comparison. Despite multiple efforts to contact some tools, such as Virtualitics, ArcGIS and Flourish, did not receive any response. However, not all the efforts were in vain, as Flowimmersive provided a paid subscription also helped with the doubts during the demo. This could be a significant issue as it affects both the user's experience and the company's reputation.

Feature analysis is the highlight of the thesis as it provides an objective way to compare the tools. Tools like Flourish and Tableau are good solutions for the corporate market. D3.js stands out when it comes to creating custom visualisation. It has unparalleled ability, yet it is hard to use. ArcGIS is recognized for handling geographical data. Flowimmersive is proficient in managing big data and multidimensional analysis. The thesis then provides recommendations based on the

result of the feature analysis. Validation is performed using real-life scenarios, after analysing the scenario features are selected and the tool is chosen accordingly. However, there is no evaluation method for verifying this which is a limitation.

Quantitative metrics such as learning time and visualisation creation time are essential in the analysis. However, these metrics are subjective and could include many variables dependent on the author's abilities. This could potentially distract from the feature perspectives, that's why it is included in the appendix as time is a useful factor in business decision-making. This thesis includes both qualitative and quantitative assessments. The insights provided can guide future research and practical applications in this developing field.

Visualization systems do not usually incorporate help for making a story phase and they particularly fall short in supporting people in the steps to collect and organize excerpts from data exploration that are potentially interesting for the final visual story (Lee, et al., 2015). Also, they don't provide any support for building a story plot. There are research opportunities to further explore ways to provide guidelines for making a good story, suggesting different plots, and letting people experiment with several plots and select the most compelling one (Lee, et al., 2015). Finding a way, how the audience interact with the visualisation can form a useful methodology to measure interactivity. Tools have been selected based on the requirements, some high-technical tools, visualisation libraries and old-school reporting tools may not be included in the comparison. Some tools didn't meet the requirements but the author has included them as they don't market data storytelling as a feature, but they do have support for it.

LIST OF REFERENCES

A. Satyanarayan and J. Heer, "Authoring Narrative Visualizations with Ellipsis", Computer Graphics Forum (Proc. EuroVis), vol. 33, no. 3, pp. 361-370, 2014.

Best practices for telling great stories. (n.d.). Tableau. https://help.tableau.com/current/pro/desktop/en-us/story_best_practices.htm

B. Lee, R.H. Kazi and G. Smith, "SketchStory: Telling More Engaging Stories with Data through Freeform Sketching", IEEE Trans. Visualization and Computer Graphics, vol. 19, no. 12, pp. 2416-2425, 2013.

B. Lee, N. H. Riche, P. Isenberg and S. Carpendale, "More Than Telling a Story: Transforming Data into Visually Shared Stories," in IEEE Computer Graphics and Applications, vol. 35, no. 5, pp. 84-90, Sept.-Oct. 2015, doi: 10.1109/MCG.2015.99.

Borges, M., Correa, C.M., & Silveira, M.S. (2022). Fundamental elements and characteristics for telling stories using data. J. Interact. Syst., 13, 77-86

Build narratives in your dashboards with Story Points in Tableau 8.2. (n.d.). Tableau. https://www.tableau.com/blog/82-preview-tell-story-your-data-story-points-30761

Brinker, K. (1988). *Linguistische Textanalyse,* Dt Blindenstudienanst.

Cairo, A. (2016). The truthful art: data, charts, and maps for communication. https://openlibrary.org/books/OL26268908M/The_Truthful_Art

Choosing the Right Chart | explore Workspace. (n.d.). Juicebox. https://explore.myjuicebox.io/a/choosing_the_right_chart/?_ga=2.213549844.140790763.1699533971-1923208024.1697102191

Contextualize Your Data with the Storytelling Studio | Toucan. (n.d.). https://www.toucantoco.com/en/product/storytelling-studio

Cruz, P., & Machado, P. (2011b). Generative storytelling for information visualization. IEEE Computer Graphics and Applications, 31(2), 80–85. https://doi.org/10.1109/mcg.2011.23

C. Tonkinwise, A taste for practices: Unrepressing style in design thinking, Design Studies, vol. 32, no. 6, pp. 533-545, 2011.

Dahlstrom, M. F. 2014. Using narratives and storytelling to communicate science with nonexpert audiences. Proceedings of the National Academy of Sciences 111 (Supplement_4):13614–20.

Drag & Drop analyses | Phrazor. (n.d.). https://phrazor.ai/docs/report-configuration/analyses

D. Roam, Show and tell: how everybody can make extraordinary presentations, UK:Penguin, 2014.

Dykes, B. (2019). Effective data storytelling: How to Drive Change with Data, Narrative and Visuals. John Wiley & Sons

Eckerson, W. W. (2010). Performance Dashboards: Measuring, Monitoring, and Managing Your Business. John Wiley & Sons.

E. Dimara and C. Perin. What is interaction for data visualization? IEEE transactions on visualization and computer graphics, 26(1):119–129, 2019

Empathy Map Canvas, Gray.D, 2017 http://gamestorming.com/empathy-map-canvas-006/.

Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. American Psychologist, 49(8), 709–724. https://doi.org/10.1037/0003-066X.49.8.709

F.Amini,N.H.Riche,B.Lee,A.Monroy-Hernandez,andP.Irani.Author- ing data-driven videos with dataclips. IEEE Transactions on Visualization and Computer Graphics, 23(1):501–510, 2017

Features. (n.d.). Flourish. https://flourish.studio/features/

F D'Errico, R Signorello, D Demolin and I Poggi, "The Perception of Charisma from Voice: A Cross-Cultural Study", Affective Computing and Intelligent Interaction (ACII) 2013 Humaine Association Conference, pp. 552-557, 2013.

Few, S. (2009). Now You See it: Simple Visualization Techniques for Quantitative Analysis

F.Hohman , M.Conlen, J.Heer, and D.H.P.Chau. Communicating with interactive articles. Distill, 5(9):e28, 2020 (Hohman, et al., 2020)

Flow - Immersive data visualizations. (n.d.). Flow - Immersive Data Visualizations. https://a.flow.gl/flow/nlm9e69b0/display

Forage. (n.d.). Forage. https://www.theforage.com/virtual-experience/MyXvBcppsW2FkNYCX/tata-group/data-visualisation-p5xo/creating-effective-visuals

G. Dove and S. Jones. Narrative visualization: Sharing insights into complex data. Interfaces and Human Computer Interaction (IHCI), pp. 21–23, 2012.

Gemignani, Z. (2021, November 5). *Getting Started with Data Storytelling — Juice Analytics*. Launch Customer-Facing Data Products – Juice Analytics. https://www.juiceanalytics.com/writing/getting-started-with-data-storytelling

Gemignani, Z. (2023, January 15). *15 Best Data Storytelling Tools (updated for 2023) — Juice Analytics*. Launch Customer-Facing Data Products – Juice Analytics. https://www.juiceanalytics.com/writing/best-data-storytelling-solutions

Gershon, N., & Page, W. (2001). What storytelling can do for information visualization. Communications of the ACM, 44(8), 31–37. https://doi.org/10.1145/381641.381653

Gray, J., Chambers, L. and Bounegru, L., 2012. The data journalism handbook: How journalists can use data to improve the news. " O'Reilly Media, Inc."

G. Robertson, "Effectiveness of Animation in Trend Visualization", Trans. Visualization and Computer Graphics, vol. 14, no. 6, pp. 1325-1332, 2008.

How much data Flourish templates can handle. (n.d.). Flourish. https://helpcenter.flourish.studio/hc/en-us/articles/8827938638991-How-much-data-Flourish-templates-can-handle

Herman, D., & Schank, R. C. (1996). Tell me a story: A new look at real and artificial memory. SubStance, 25(1), 140. https://doi.org/10.2307/3685240

H. McLellan, "Corporate storytelling perspective", The Journal for Quality and Production, vol. 29, no. 1, pp. 17-20, 2006.

Jakob Nielsen. 1994. Usability Engineering. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, adoption barriers and myths of open data and open government. Information Systems Management, 29(4), 258-268.

Jonathan Hook. 2018. Facts, Interactivity and Videotape: Exploring the Design Space of Data in Interactive Video Storytelling. In Proceedings of the 2018 ACM International Conference on Interactive Experiences for TV and Online Video

(TVX '18). Association for Computing Machinery, New York, NY, USA, 43–55. https://doi.org/10.1145/3210825.3210826

Juicebox vs. Competition — Juice Analytics. (n.d.). Launch Customer-Facing Data Products – Juice Analytics. https://www.juiceanalytics.com/juicebox-vs-competition

Kelliher, A., & Slaney, M. (2012). Tell me a story. IEEE MultiMedia, 19(1), 4. https://doi.org/10.1109/mmul.2012.13

Krug, S. (2014). Don't make me think, revisited: A Common Sense Approach to Web Usability. Pearson Education.

Mackinlay, J., Hanrahan, P., & Stolte, C. (2007). Show Me: Automatic Presentation for Visual Analysis. IEEE Trans. Visualization and Computer Graphics, 13(6), 1137–1144.

M. A. Borkin, Z. Bylinskii, N. W. Kim, C. M. Bainbridge, C. S. Yeh, D. Borkin, H. Pfister, and A. Oliva. Beyond memorability: Visualization recognition and recall. IEEE Transactions on Visualization and Computer Graphics, 22(1):519–528, 2015.

M. Austin, Useful Fictions: Evolution Anxiety and the Origins of Literature, Univ. Nebraska Press, 2011.

McKee, R. (1997). *Story: Substance, structure, style, and the principles of screenwriting*. HarperCollins.

Meixner, B., Matusik, K., Grill, C. and Kosch, H., 2014. Towards an easy to use authoring tool for interactive non-linear video. Multimedia Tools and Applications, 70, pp.1251-1276.

Metrics: Your Story Characters | explore Workspace. (n.d.). Juicebox. https://explore.myjuicebox.io/a/metrics_your_story_characters/?_ga=2.44220708.140790763.1699533971-1923208024.1697102191

M. J. Eppler and S. Kernbach, Dynagrams: Enhancing design thinking through dynamic diagrams, Design Studies, vol. 47, pp. 91-117, 2016.

Miller, Z. (2014). Data Visualization and the News: A Critical Analysis. Journal of Communication, 64(2), 320-349

Murdoch, T.B. and Detsky, A.S., 2013. The inevitable application of big data to health care. Jama, 309(13), pp.1351-1352.

N. Duarte, Resonate: Present visual stories that transform audiences, John Wiley & Sons, 2013.

Nussbaumer Knaflic, C. (2015). Storytelling with data (C. N. Knaflic, Ed.). John Wiley & Sons.

Ojo, A., & Heravi, B. R. (2017). Patterns in award winning data storytelling. Digital Journalism, 6(6), 693–718. https://doi.org/10.1080/21670811.2017.1403291

Patrick Millais, Simon L. Jones, and Ryan Kelly. 2018. Exploring Data in Virtual Reality: Comparisons with 2D Data Visualizations. In Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems (CHI EA '18). Association for Computing Machinery, New York, NY, USA, Paper LBW007, 1–6. https://doi.org/10.1145/3170427.3188537

P. Godfrey, J. Gryz and P. Lasek, "Interactive Visualization of Large Data Sets," in IEEE Transactions on Knowledge and Data Engineering, vol. 28, no. 8, pp. 2142-2157, 1 Aug. 2016, doi: 10.1109/TKDE.2016.2557324

Pousman, Z.; Stasko, J.; Mateas, M. Casual Information Visualization: Depictions of Data in Everyday Life. IEEE Trans. Vis.Comput. Graph. 2007, 13, 1145–1152. [CrossRef] [PubMed]

Presentation, D. (2023, October 23). What are the key features of a good data storytelling tool and why are they important? www.linkedin.com. https://www.linkedin.com/advice/0/what-key-features-good-data-storytelling-tool

Relatable and Specific | explore Workspace. (n.d.). Juicebox. https://explore.myjuicebox.io/a/relatable_and_specific/?_ga=2.18989368.1407907 63.1699533971-1923208024.1697102191

R. Kosara and J. Mackinlay, "Storytelling: The Next Step for Visualization," in Computer, vol. 46, no. 5, pp. 44-50, May 2013, doi: 10.1109/MC.2013.36.

S. Bateman, "Useful Junk? The Effects of Visual Embellishment on Comprehension and Memorability of Charts", Proc. ACM Conf. Human Factors in Computing Systems (CHI 10), pp. 2573-2582, 2010.

Segel, E., & Heer, J. (2010). Narrative Visualization: Telling Stories with Data. IEEE Transactions on Visualization and Computer Graphics, 16(6), 1139–1148. https://doi.org/10.1109/tvcg.2010.179

S. Kernbach, "Storytelling Canvas: A Visual Framework for Developing and Delivering Resonating Stories," 2018 22nd International Conference Information Visualisation (IV), Fisciano, Italy, 2018, pp. 390-395, doi: 10.1109/iV.2018.00073.

Sorin Adam Matei & Lucas Hunter (2021) Data storytelling is not storytelling with data: A framework for storytelling in science communication and data journalism, The Information Society, 37:5, 312-322, DOI: 10.1080/01972243.2021.1951415

S. Sinek, Start with why: How great leaders inspire everyone to take action, Penguin, 2009.

Stolper, Charles D, et al. Emerging and Recurring Data-Driven Storytelling Techniques: Analysis of a Curated Collection of Recent Stories. 2016. MSR-TR-2016-14

Shi, Y., Gao, T., Jiao, X., & Cao, N. (2022). Breaking the fourth wall of data stories through interaction. IEEE Transactions on Visualization and Computer Graphics, 1–11. https://doi.org/10.1109/tvcg.2022.3209409

Stories. (n.d.). Tableau. https://help.tableau.com/current/pro/desktop/en-us/stories.htm#:~:text=Each%20individual%20sheet%20in%20a,new%20questions%20of%20the%20data

Sun, M., Cai, L., Cui, W., Wu, Y., Shi, Y., & Cao, N. (2022). ERATO: Cooperative data story editing via fact interpolation. IEEE Transactions on Visualization and Computer Graphics, 1–11. https://doi.org/10.1109/tvcg.2022.3209428

Team, P. B. (n.d.). *What is data storytelling?* https://powerbi.microsoft.com/en-us/data-storytelling/

Team, S. (2022, March 18). SiSense Narratives: telling a story with your data. Sisense. https://www.sisense.com/blog/sisense-narratives-telling-story-data/

Technics Publications. (2021, September 2). How to choose exactly the right data story for your audience. Data Management. https://www.techtarget.com/searchdatamanagement/feature/How-to-choose-exactly-the-right-data-story-for-your-audience#:~:text=The%20importance%20of%20stakeholders,failure%20of%20a%20given%20initiative

The Data Storytelling Platform. Put a human perspective on the data that drives your business. (2018, January 29). Nugit. https://www.nugit.co/

T. Strothotte and C. Strothotte, Seeing between the Pixels, Springer-Verlag, 2003.

Tufte, E. R. (2001). The visual display of quantitative Information PAPERBACK: Second Edition PAPERBACK

Yangjinbo Zhang and Artur Lugmayr. 2020. Designing a User-Centered Interactive Data-Storytelling Framework. In Proceedings of the 31st Australian Conference on Human-Computer-Interaction (OzCHI '19). Association for Computing Machinery, New York, NY, USA, 428–432. https://doi.org/10.1145/3369457.3369507

Zhang, Y., Reynolds, M. A., Lugmayr, A., Damjanov, K., & Hassan, G. M. (2022). A visual data Storytelling framework. Informatics (Basel), 9(4), 73. https://doi.org/10.3390/informatics9040073

Zhu, Y. (2007). Measuring Effective Data Visualization. In: Bebis, G., *et al.* Advances in Visual Computing. ISVC 2007. Lecture Notes in Computer Science, vol 4842. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-76856-2

APPENDIXES

Appendix 1

Quantitative experimentation of data storytelling tools

This appendix contains the table which includes information about the learning time, visualisation creation time, and the number of errors that occurred during creating the visualisation. The goal is to get an idea about each tool. This experimentation is subjective as it is measured by only 1 user which is the author of this thesis. Therefore, it is included in the appendix. Performing this test on more users can provide a clearer picture and accurate data. Time is an important factor when making decisions. Hence, it is an important criterion to be measured.

Table 4.2

Comparison of data storytelling tools

Tools	Learning time	Creation time	No, of Error
Flourish	1 hr 23 min	35 min	1
Living charts	38 min	26 min	0
Tableau	3 hr 27 min	2 hr 1 min	3
D3.js	8 hr 14 min	2 hr 35 min	6
ArcGIS	4 hr 7 min	1 hr 28 min	2
Mapbox	1 hr 49 min	1 hr 2 min	1
Kepler.gl	1 hr 36 min	1 hr 13 min	0
Flowimmersive	3 hr 45 min	2 hr 40 min	2

This table shows the learning time, visualisation creation time and no of error encountered during the comparison of data storytelling tools.