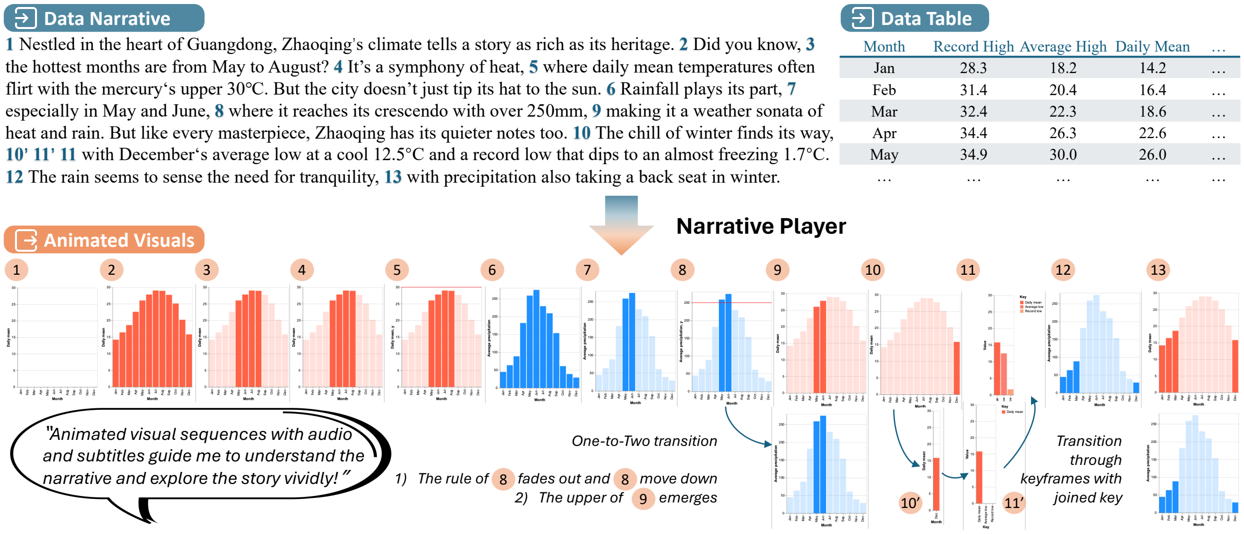
Narrative Player: Reviving Data Narratives with Visuals

Zekai Shao, Leixian Shen, Haotian Li, Yi Shan, Huamin Qu, Yun Wang, and Siming Chen



1 Nestled in the heart of Guangdong, Zhaoqing's climate tells a story as rich as its heritage. **2** Did you know, **3** the hottest months are from May to August? **4** It's a symphony of heat, **5** where daily mean temperatures often flirt with the mercury‘s upper 30°C. But the city doesn't just tip its hat to the sun. **6** Rainfall plays its part, **7** especially in May and June, **8** where it reaches its crescendo with over 250mm, **9** making it a weather sonata of heat and rain. But like every masterpiece, Zhaoqing has its quieter notes too. **10** The chill of winter finds its way, **10' 11' 11** with December‘s average low at a cool 12.5°C and a record low that dips to an almost freezing 1.7°C. **12** The rain seems to sense the need for tranquility, **13** with precipitation also taking a back seat in winter.

Month	Record High	Average High	Daily Mean	…
Jan	28.3	18.2	14.2	…
Feb	31.4	20.4	16.4	…
Mar	32.4	22.3	18.6	…
Apr	34.4	26.3	22.6	…
May	34.9	30.0	26.0	…
…	…	…	…	…

Fig. 1: An animated visual sequence example with seamless transitions, audio and subtitles automatically generated by Narrative Player from data narrative and data table for engaging reading experience, where ⓧ means the visuals will transition to the next when the narration moves forward to the corresponding x-th segment.

Abstract— Data-rich documents are commonly found across various fields such as business, finance, and science. However, a general limitation of these documents for reading is their reliance on text to convey data and facts. Visual representation of text aids in providing a satisfactory reading experience in comprehension and engagement. However, existing work emphasizes presenting the insights of local text context, rather than fully conveying data stories within the whole paragraphs and engaging readers. To provide readers with satisfactory data stories, this paper presents Narrative Player, a novel method that automatically revives data narratives with consistent and contextualized visuals. Specifically, it accepts a paragraph and corresponding data table as input and leverages LLMs to characterize the clauses and extract contextualized data facts. Subsequently, the facts are transformed into a coherent visualization sequence with a carefully designed optimization-based approach. Animations are also assigned between adjacent visualizations to enable seamless transitions. Finally, the visualization sequence, transition animations, and audio narration generated by text-to-speech technologies are rendered into a data video. The evaluation results showed that the automatic-generated data videos were well-received by participants and experts for enhancing reading.

Index Terms—Narrative, data facts, large language model, visual sequence, storytelling, reading experience

1 Introduction

Data-rich documents, scientific reports, and a wide range of reading materials encompassing fields such as business, finance, and science, are commonly encountered in daily life. These data narrations serve as a means for readers to comprehend the underlying data stories, thereby facilitating the knowledge acquisition and decision-making processes.

However, the conventional text-based format of data communication has several limitations when it comes to data comprehension. It can be difficult to interpret, fragmented, and prone to misinterpretation [39, 40, 71]. Moreover, lengthy text often leads to a monotonous reading experience, causing boredom and a lack of focus, which negatively impacts user engagement [23]. In contrast, visual representations of text, such as illustrations and visualizations, aid in constructing mental models of the information and enhance comprehension [18, 22, 23]. Furthermore, visualizations have played a significant role in distant reading [41] by transforming text into an abstract view that highlights global features. Dynamic visualizations and videos take this a step further, actively engaging readers and enhancing data communication through live storytelling [60, 77]. Considering these factors, we are motivated to generate live data stories from data-rich documents, aiming to enrich the reading experience and enhance user engagement.

Given the potential benefits of integrating text with visuals, multiple studies have proposed solutions to enhance the reading experience of archival documents from diverse perspectives. One line of research employs annotated visualizations primarily for navigation [21, 25], offering brief annotations to situate a document within a larger corpus instead of exploring the document's specific story or context. This approach's superficial grasp of the context limits its ability to generate visuals that enrich the reading experience by deepening comprehension of the data story. Others focus on reading within individual small data-rich subsets of documents, but ignore the broader data narratives

- Z. Shao, Y. Shan, and S. Chen are with Fudan University. S. Chen is also with Shanghai Key Laboratory of Data Science. E-mail: zkshao23@m.fudan.edu.cn; {yshan20, simingchen}@fudan.edu.cn.
- L. Shen, H. Li, and H. Qu are with The Hong Kong University of Science and Technology. E-mail: {lshenaj, haotian.li}@connect.ust.hk; huamin@cse.ust.hk.
- Y. Wang is with Microsoft. E-mail: wangyun@microsoft.com.
- S. Chen and Y. Wang are the corresponding authors.

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

[7,40]. For instance, Elastic Documents [7] establishes cross-references between tables and text through keyword matching and generates on-demand visualizations for user-selected sentences. However, these methods focus on enhancing comprehension via visualizing local data insights, rather than the global text context and data stories of overall paragraphs, and they do not emphasize reading enjoyment.

In this paper, we propose a novel method named Narrative Player, which aims for the enhancing the reading experience by automatic revival of data narratives with visuals. It generates animated visual sequences with audio and subtitles to provide users with a multi-channel experience. For instance, for the data narratives presented in Fig. 1, Narrative Player will generate the animated visuals with transitions and convey textual information through audio narration and subtitles, thereby enhancing the user experience. Narrative Player consists of two major modules, i.e., narrative analysis and visual generation. Given a data narrative and its corresponding data table as input, the narrative analysis module utilizes large language models (LLMs) to extract contextualized data facts (i.e., numerical or statistical results of data mining [74]) from the narrative, representing the underlying story of the narrative. The visual generation module then organizes the extracted contextualized data facts into a cohesive and contextually meaningful visualization sequence with a carefully designed optimization function. Additionally, the module incorporates seamless transition animation effects between adjacent visualizations to strengthen user engagement. Finally, the visualization sequence, transition animations, and audio narration generated by text-to-speech technologies are combined to produce a data video with narration-animation interplay.

To evaluate Narrative Player comprehensively, we conducted a user study and expert study. In the user study, we examined the data videos generated by Narrative Player against a set of baselines by user-perceived satisfaction. These baselines included plain text to assess the enhancement of user experience, videos crafted by professionals to evaluate video quality, and videos from two ablation studies to measure the effectiveness of our two modules. The results revealed that our generated data videos effectively enhance the user experience through two essential modules. Furthermore, these videos were well-received and displayed comparable quality to human-composed ones. We also presented the videos generated by Narrative Player to experts for their feedback from a professional perspective, incorporating multi-dimensional ratings and interviews. The experts agreed that Narrative Player shows satisfactory performance in generating coherent visuals and maintaining consistency and contextualization, contributing to enhancing reading experience in both comprehension and enjoyment.

In summary, our contributions are concluded as follows:

- Narrative Player, an approach for automatically generating consistent, contextualized, and animated visual sequences for data narratives to enhance the reading experience.
- A narrative analysis module powered by LLM and embedding models for handling semantics and extracting data facts from long narratives, and a visual generation module powered by optimization considering side-by-side visualizations, visual focus, and primary visualizations to select a contextualized and consistent visual sequence.
- An evaluation combining a user study for the analysis of user-perceived satisfaction, and an expert study with ratings and interviews to demonstrate the effectiveness of Narrative Player.

2 Related Work

In this section, we discuss related works from three perspectives, i.e., natural language-based data visualization generation, visual generation, and text-driven storytelling.

2.1 Natural Language-Based Data Visualization Generation

Recently, natural language-based data visualization generation technologies have gained increasing attention, allowing users to express their intents conveniently using familiar natural language [57]. For example, commercial software such as Microsoft Power BI and Tableau both have their natural language interface (NLI) services. Techniques range from translating natural language queries into visualization specifications [16, 38, 43] to enhancing user interaction through natural language and multimodal inputs [64, 66].

In addition to one-shot visualization generation, recent studies have explored conversational interaction experiences [54, 57, 70]. Some studies enhanced the efficiency and effectiveness of NLIs. They have explored applying pragmatics principles to analytical dialogues [24, 65], simplifying query processes [17], and providing explanations for understanding visualization outcomes [20]. Others expanded the functionality and interactivity of NLIs. They have developed interactive dialogue systems for direct visualization interaction [53], complex request and simultaneous conversations handling [19], and visualization authoring [72], alongside exploring design principles for NLI-based data exploration [70].

Overall, the aforementioned systems mostly generate visualizations based on simple natural language queries (e.g., "show me how horsepower varies each year by origin") for data analysis purposes. However, there is a lack of work that generates a sequence of visualizations for visual data storytelling purposes based on the context of data narratives. Our work extends the scope of natural language-based data visualization generation technologies by moving towards augmenting data narratives with dynamic data charts for storytelling.

2.2 Visualization Sequence Generation

Narrative sequencing plays a vital role in visual data storytelling, which can affect the audience's comprehension. Hullman et al. [26] conducted two studies to identify a relative ranking of visualization transition types by viewer perspective. GraphScape [30] further formulates a directed graph-based model for measuring visualization similarity and ranking visualization sequences. Based on the model, TaskVis [59] proposes a task-driven strategy to recommend a set of combined visualizations to give an overview of the dataset. Dziban [35] supports both partial specification visualization and an anchoring mechanism for conveying the desired context. Shi et al. [61] leverage reinforcement learning to generate chart sequencing, capturing the relationships between charts.

Data facts, derived from raw data, underpin analysis and understanding, while visual data stories enlive these facts with sequential visualizations for easier interpretation. For instance, DataShot [74] can automatically generate fact sheets from tabular data, consisting of three parts, i.e., fact extraction, fact composition, and visual synthesis. Calliope [63] accepts a spreadsheet as input, progressively generates story pieces, and organizes the facts into a visual data story based on the fact importance. AutoClips [62] can automatically generate a data video by configuring a sequence of data facts with a fact-driven animation clip library. Erato [69] supports human-computer cooperation to create visual data stories, where the user needs to provide a set of keyframes and describe the story theme and structure. ChartStory [78] is designed to help users craft comic-style data stories from a set of user-created charts, suggesting the underlying sequence of data-driven narratives.

Coherent visualization sequences can enhance readers' understanding and memory of data stories. However, most work focuses on data insights and facts, neglecting the importance of narrative text with context . This paper aims to generate vivid data stories, specifically required consistency and contextualization, by understanding data narratives' context and generating coherent visuals controlled by context, rather than directly configuring visuals or clips in data-driven ways.

2.3 Text-Driven Storytelling with Visuals

Narrative text plays an essential role in data-centric storytelling [52], often in conjunction with visualizations to promote information dissemination. For example, a set of works focuses on the efficient writing of data documents enriched in visualizations to present data insights [10, 12, 15, 32, 45, 55, 67, 68]. Aiming at enlivening news articles, Contextifier [25] and NewsViews [21] automatically generate annotated visualizations based on the news text to aid storytelling. Elastic document [7] is an interactive system for generating on-demand visuals based on readers' selection of narratives and tables. Most recently, Charagraph [40] helps paper readers to dynamically create interactive charts from data-rich paragraphs to obtain a better sense of data in text.

These works primarily employ keyword-matching techniques to select data for visuals and are only aware of the local context as the input limits to individual sentences or a few sentences as snippets. Advanced NLP technologies are not utilized to deal with extensive contextualization , while we try to employ LLMs to infer the semantics of vague clauses in long narratives and fully convey the data story.

Moreover, by enhancing static materials with compelling dynamics, data video has also been gaining increasing popularity in recent years. Cheng et al. [13] investigated the roles and interplay of narrations and animations in data videos, highlighting their combined role in storytelling. Following the study, WonderFlow [73] is an interactive authoring tool that allows users to easily create data videos with narration-animation interplay from static visualizations and text. Data Player [60] further automates the process. It first leverages LLMs to create text-visual links and then recommends an animation sequence based on constraint programming. These works focus on animate single given visual under the guidance of narration, instead of generating visual sequences from solely data narratives.

Despite the abundance of research on text-driven storytelling techniques, most of these methods require users to invest a significant amount of time and effort into creating stories and undergoing trial-and-error processes [10, 12, 15, 40, 73] or ask users to provide complete narrative texts and data visualizations [14, 32, 60, 67]. Our work is in line with the research of text-driven storytelling and attempts to automatically generate consistent and contextualized visual sequences based solely on data narratives to enhance the vividness of data stories.

3 Narrative Player

In this section, we first give an overview of Narrative Player, and discuss the two key modules, i.e., narrative analysis and visual generation.

3.1 Overview

3.1.1 Design Requirements

Previous studies have enhanced reading materials with the integration of individual visuals [7, 25, 40]. However, these enhancements partially address the complexities of comprehension, as data narratives encompass coherent insights that necessitate diverse visuals for a complete understanding [31]. Additionally, they have not exploited enjoyment. The importance of animated visual sequences for experience enhancement lies in their ability to present a memorable and complete story [27, 44]. Such sequences link individual data insights and contexts within the narrative through transitions and audio narration. This approach, however, poses challenges associated with ensuring consistency and contextualization, surpassing the complexities of generating single visuals with limited context. Acknowledging this, we have derived a set of design requirements based on the previous work and our motivation to inform the design of Narrative Player.

R1 Understand the narrative and extract textual intents: To present the data stories, the first step should be to analyze the input narrative to understand its underlying explicit and implicit textual intents [70] and extract the key data facts for visual representation. The extracted intent information acts as the abstraction layer between the data narrative and generated visuals, ensuring that the visuals effectively communicate the intended message.

R2 Handle ambiguous expressions within context: Due to the ambiguous nature of natural language, data narratives may contain expressions from which data facts cannot be extracted explicitly [57], but they are important to the coherence of the contextual story. Ignoring such sentences can lead to confusion for the viewer and inversely harm comprehension and enjoyment. To handle this, the system should analyze the context in which an ambiguous sentence appears and determine the most likely interpretation based on surrounding information.

R3 Align textual and visual intents: A seamless binding between the narrative and visuals is crucial for user experience, necessitating a close alignment between the textual and visual intents [10, 13, 60]. This may involve using similar visual elements or themes in the visualizations to reinforce the key ideas or concepts presented in the text. Additionally, the system should be able to adapt the visuals to better match the intended message, such as changing the color or shape of a visualization to emphasize a particular point in the narrative [70].

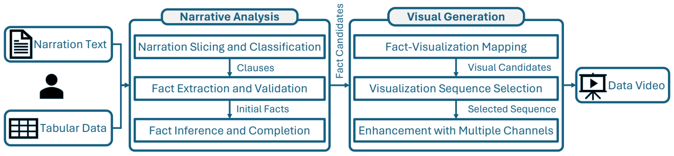

Narration Text
Tabular Data
Narrative Analysis
Narration Slicing and Classification
Clauses
Fact Extraction and Validation
Initial Facts
Fact Inference and Completion
Fact Candidates
Visual Generation
Fact-Visualization Mapping
Visual Candidates
Visualization Sequence Selection
Selected Sequence
Enhancement with Multiple Channels
Data Video

Fig. 2: Narrative Player system overview and processing pipeline.

R4 Ensure visual consistency and contextualization: Consistency in style, color, and layout is crucial for cohesive visual representations that match the narrative's theme and tone. In addition, visual sequences should balance visualization diversity and user cognitive load to help maintain the viewer's understanding of context. Frequent changes in the sequences can impose a huge cognitive load on the viewer [60]. Real-world data videos typically center on one or two main visualizations, supplemented by other embellished information [13].

R5 Enable multi-channel perception with transitions, audio and subtitles: Static rendering of visual sequences can bore viewers and obscure differences between visuals. However, dynamic transitions between different visualizations can effectively engage viewers and clarify narrative progression [29]. The system should enable smooth transitions between visualizations by incorporating transitional animation effects (e.g., fade-ins, wipes, or slides) that enhance the flow of the narrative. Audio adds dimension to the display, with narration acting as subtitles to guarantee that the final video accurately conveys the original textual information.

3.1.2 Workflow

Based on the aforementioned design requirements, we propose Narrative Player to automatically revive data narratives with visuals. It works sequentially, beginning with the extraction of contextualized data facts, which serve as elementary building blocks of a data story. This is followed by the generation of visuals, resulting in a well-organized visual sequence and multi-channel experience via videos. Fig. 2 illustrated the two key modules used in this pipeline:

Narrative Analysis. The narrative analysis module (Sec. 3.2) divides the narrative text into clauses[1] by the LLM, which is viewed as the basic units that contain data facts. Then it uses the LLM with well-designed prompts and the Sentence Embedding model to extract and validate data facts from clauses (R1). Initially, this yields validated facts that are semantically aligned with the narrative, which are a subset of all fact candidates. The module then employs LLM and rule-based heuristics to infer and complete fact sets (R2), ultimately producing enriched, contextualized data facts.

Visual Generation. The visual generation module (Sec. 3.3) maps the contextualized facts candidates into visualization candidates (R3). Leveraging an optimization function emphasizing both contextualization and consistency (R4), it determines a specific visualization for each clause, crafting a coherent sequence. Furthermore, to enable a more engaging storytelling experience, the module enhances the visualization sequence with transition animations and audio narration and finally renders them into a data video (R5).

3.2 Narrative Analysis

To generate corresponding visual aids from data narratives, we need to understand the narrative intents by extracting data facts (R1) and handle ambiguity within the context (R2). The task is complex with technical challenges (TCs). Specifically, data facts must be contextualized to align with the narrative's semantics (TC1) and its broader context (TC2) and must ensure completeness to enrich visual generation (TC3).

Several conversational NLI systems consider the context of the preceding NL utterance by keywords-parsing [70] or LSTM [72], but

[1] According to English grammar scholars [75], a clause is identified as a unit of a sentence that forms recognizable syntactic constituents: a subject and predicate, with or without adjuncts, or a predicate with or without adjuncts.

they can't address the wider context of the whole narration. Prior text-driven storytelling works also use keywords-parsing [7, 40, 68] and fall short in structuring data fact representations from lengthy narratives. They mostly operate on individual words or phrases without adequately addressing their relations or handling ambiguities. Besides, the lack of sufficient datasets for training in data storytelling has long been acknowledged [37]. LLMs address these challenges through few-shot learning and human-like reasoning, offering superior performance in context-rich tasks and scenarios with sparse resources for their emergent ability [34]. Hence we design the automatic narrative analysis process based on LLMs. We use the OpenAI GPT-4-turbo model with 128k context and temperature 0.3.

We will detail the preparation and definitions in Sec. 3.2.1 and Sec. 3.2.2, describe the methods for fact extraction and validation in Sec. 3.2.3 (TC1), and address issues of context-based inference (TC2) and fact completion (TC3) in Sec. 3.2.4.

3.2.1 Narration Slicing and Classification

Data narratives consist of two main components: factual sentences that present data facts and story or background sentences that provide context [13]. These two segments may be interspersed with visualizations to enhance understanding. To understand the structure of the data story, narratives are first segmented into sentences using punctuation marks and LLMs assess each sentence for data facts presence. Factual sentences are further segmented into clauses for subsequent processing.

3.2.2 Data Fact Formulation

One sentence can have multiple clauses that correspond to various data facts [13]. We consider clauses as the basic units of the narrative for data fact extraction. We formalize the data fact as a 6-tuple structure, adapted from DataShot [74], but exclude the importance score as we derive facts matched to narrative semantics rather than mining them from tabular input, while we discuss the future consideration of fact importance at the end of Sec. 5.

$$fact := \{type, parameters, measure(s), context, breakdown(s), focus\} \quad (1)$$

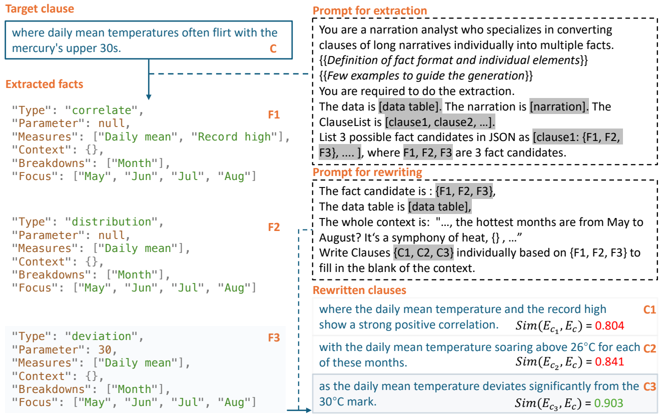
where *type* denotes the data fact type (e.g., trend, comparison, deviation), as adopted from TaskVis [58], with *parameter* specifying its details like deviation value; *context* defines the data subspace, *breakdowns* segment this subspace into groups, and *measure* assess each group's value, with *focus* highlighting specific data, as shown in Fig. 3. On this basis, we define the whole narrative structure that reflects the underlying story as a dictionary, i.e., $story : \{clauses[i] : facts_i \mid i = 1, 2, \ldots, n\}$, where $clauses[i]$ represents the i-th clause in the data narrative, $facts_i$ is the set of all facts related to $clauses[i]$, and n is the total number of clauses. Each $facts_i$ can be defined as: $facts_i = \{facts[i,1], facts[i,2], \ldots, facts[i,m_i]\}$, where $facts[i,j]$ signifies the j-th *fact* conveyed in the i-th clause and m_i is the number of facts for the i-th clause.

3.2.3 Data Fact Extraction and Validation

By providing the model with well-composed prompts and examples as shown in the lower left of Fig. 3, we effectively guide models in extracting data facts behind each clause (TC1). When prompting LLM, considering its hallucination nature [28], we first try to achieve a comprehensive data fact set for each clause by utilizing three sessions, each using the same prompt with temperature 0.3 and generating three fact candidates per clause. This approach yields nine potential data facts for each clause.

From analyzing clause-data fact relationships, we identify that generated candidates may overlap or diverge from the text's semantics (F1, F2, F3 in Fig. 3), necessitating validation and filtration to ensure the data facts' accuracy and reliability.

We proposed a novel method that leverages LLMs and sentence embedding for data fact candidate validation. Specifically, as shown in Fig. 3, for a clause (C) and a specific data fact candidate (F3), we dig the clause out of the narrative as context and fill in new clause (C3) based on this context and the data fact (F3). The rewritten clause tends to be more direct in its representation of data facts, omitting the vague expressions and rhetorical language often found in vague clauses.



where daily mean temperatures often flirt with the mercury's upper 30s.

```
"Type": "correlate",
"Parameter": null,
"Measures": ["Daily mean", "Record high"],
"Context": {},
"Breakdowns": ["Month"],
"Focus": ["May", "Jun", "Jul", "Aug"]
```

```
"Type": "distribution",
"Parameter": null,
"Measures": ["Daily mean"],
"Context": {},
"Breakdowns": ["Month"],
"Focus": ["May", "Jun", "Jul", "Aug"]
```

```
"Type": "deviation",
"Parameter": 30,
"Measures": ["Daily mean"],
"Context": {},
"Breakdowns": ["Month"],
"Focus": ["May", "Jun", "Jul", "Aug"]
```

You are a narration analyst who specializes in converting clauses of long narratives individually into multiple facts.
{{*Definition of fact format and individual elements*}}
{{*Few examples to guide the generation*}}
You are required to do the extraction.
The data is [data table]. The narration is [narration]. The ClauseList is [clause1, clause2, …].
List 3 possible fact candidates in JSON as [clause1: {F1, F2, F3}, ...], where F1, F2, F3 are 3 fact candidates.

The fact candidate is : {F1, F2, F3},
The data table is [data table],
The whole context is: "…, the hottest months are from May to August? It's a symphony of heat, {} , …"
Write Clauses {C1, C2, C3} individually based on {F1, F2, F3} to fill in the blank of the context.

where the daily mean temperature and the record high show a strong positive correlation. $Sim(E_{c_1}, E_c)$ = 0.804

with the daily mean temperature soaring above 26°C for each of these months. $Sim(E_{c_2}, E_c)$ = 0.841

as the daily mean temperature deviates significantly from the 30°C mark. $Sim(E_{c_3}, E_c)$ = 0.903

Fig. 3: An illustrating case describing how to extract, validate, and select data fact candidates. In one LLM session for preliminary extraction, three facts (F1, F2, F3) have been extracted for one clause (C). Another LLM session with appropriate prompts rewrites the target clause based on the three facts and context into three rewritten clauses (C1, C2, C3). The sentence embedding model generates embeddings (E_{c_i} and E_c) for all the clauses. Finally, F3 was ranked first based on the cosine similarity.

Then we use rankcse [36], the SOTA model for unsupervised sentence representation learning via learning to rank, to obtain the embeddings of both the original and rewritten clauses. Finally, we calculate the cosine similarity between these two embeddings. As shown in Fig. 3, C3 and C exhibit a high degree of similarity in their sentence embeddings, and thus F3 is re-ranked as the most qualified fact among these three. This approach avoids using LLMs for self-correction, acknowledging their limitations in semantic understanding and sentence embedding [42].

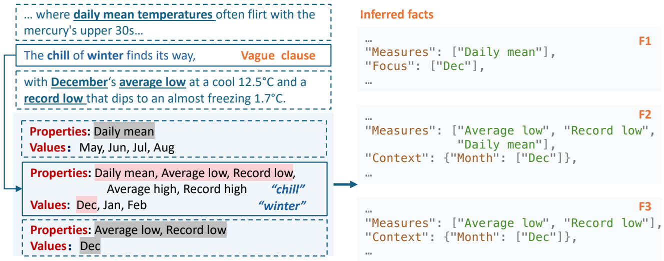
In our observation, clauses with clear data properties or easily inferred values often closely match several rewritten clauses. Thus "clear clause" is defined as having at least 6 out of 9 candidates with a similarity score above 0.85, a threshold determined from empirical testing across narratives. We remove duplicates and select the top three candidates by similarity as qualified data facts. Conversely, a "vague clause" produces fewer than 6 facts with scores above 0.85.

3.2.4 Data Fact Inference and Completion within Context

After prioritizing explicit intent [70] by data fact validation and clause characterization, two issues still exist for us to address on implicit intents: 1) Vague clauses ambiguously reference data properties, values, or table subspaces. For example, "the chill of winter finds its way" might refer to various temperature metrics and specific months in the data table, without clarity. 2) Clear clauses yield three facts, yet these represent just a subset of all possible facts. For instance, "especially in May and June" might refer solely to these months or highlight their distinctiveness, suggesting varied interpretations as either a complete *context* or specific *focus* within a broader narrative.

The challenge with vague clauses is determining relevant data properties or values that match the clause semantics and context (TC2). To tackle this, we use another LLM session as outlined in Fig. 4:

Firstly, keywords are identified within each vague clause and mapped to potential data properties or values in the table. This process generates a candidate set of *measure*, *context*, or *focus* for fact candidates. For instance, "chill" and "winter" lead to inferring five properties about temperature and three context-appropriate values about month in the Northern Hemisphere, based on the location mentioned in the narration. Then we infer reference data properties or values within context. Clear clauses that are semantically related and adjacent are selected as reference clauses. The intersection of the candidate set and the qualified candidate facts for reference clauses forms a set of reference data properties or values. In Fig. 4, two adjacent clear clauses serve as references. The intersection of data properties or values—one property from the first and two properties plus one value from the second—are selected to match with the target vague clause's inferred properties or


… where daily mean temperatures often flirt with the mercury's upper 30s…
The chill of winter finds its way,
Vague clause
with December's average low at a cool 12.5°C and a record low that dips to an almost freezing 1.7°C.
Properties: Daily mean
Values: May, Jun, Jul, Aug
Properties: Daily mean, Average low, Record low, Average high, Record high
"chill"
"winter"
Values: Dec, Jan, Feb
Properties: Average low, Record low
Values: Dec
Inferred facts
F1
F2
F3

```
…
"Measures": ["Daily mean"],
"Focus": ["Dec"],
…
```

```
…
"Measures": ["Average low", "Record low",
             "Daily mean"],
"Context": {"Month": ["Dec"]},
…
```

```
…
"Measures": ["Average low", "Record low"],
"Context": {"Month": ["Dec"]},
…
```

Fig. 4: An illustrating case of inferring vague clause within context. The narrative analysis module first detects keywords, "chill" and "winter", relating them to five candidate properties about temperature and three candidate values for months based on the narration and data. Then the module selects two clear clauses around as references, inferring two sets of referred properties and values, as shown in dashed boxes and arrows. The module further merges the intersection of them as filtered properties and values and extracts three fact candidates (F1, F2, F3).

values. After that, the union of the intersection of each reference set and the candidate set establishes the filtered data properties or values. In Fig. 4, one value and three properties from the two references are merged as the results. Finally, the fact candidates are generated. If no candidates are initially identified, fact candidates from the nearest clause are used. Otherwise, three fact candidates are generated using a subset of the filtered properties or values as *measure*, *context*, or *focus*. Each candidate aligns semantically with the first reference clause (F1), the second reference clause (F3), or a combination of both (F2), as shown in Fig. 4.

The process of obtaining a complete fact set (TC3) involves addressing three types of incompleteness: single facts, multiple facts within a single clause, and multiple facts across clauses. These are tackled by applying the following heuristics, and use an Example Clause (EC: "with December's average low at a cool 12.5°C and a record low that dips to an almost freezing 1.7°C.") to illustrate the process..

For a single fact, since keywords represent either an entire subspace as *context* or a *focus* within a larger space, the fact candidates are expanded by interchanging these two elements, provided that the *focus* corresponds to a smaller data range than the *context*. For example, the initially generated fact (noted as F) of EC includes the attributes: {*type*: *extreme*, *measures*: {*Average low*, *Record low*}, *context*: {}, *focus*: *Dec*}. By swapping *context* and *focus*, F_1 is obtained, which includes the attributes: {*focus*: {}, *context*:{*Month*:[*Dec*]}}.

Also, fact candidates with the same *measure* and *breakdowns* but different *type* and *parameter* may be generated, suggesting similar semantics but arbitrary distinctions. In such cases, *type* and *parameter* are cross-combined between the two facts. Among the initial facts of EC, one has the same *measures* and *breakdown* as F_1, but its *type* is *distribution*. Replacing the *type* of F_1 with *distribution* results in F_2.

Besides, keywords in clear clauses may capture part of the context but miss the context suggested by adjacent clauses. If a fact matches an adjacent clause's fact in either *measure* or *context* and is a subset of that fact, we expand the *measure* or *context* to match. The *measures* of a fact in the previous sentence of EC, "The chill of winter finds its way", also include *Daily mean* (see Fig. 4). Adding it to the *measures* of F_2 results in F_3. F_3 has the attributes: {*type*: *distribution*, *measures*: {*Average low*, *Record low*, *Daily mean*}, *context*:{*Month*:[*Dec*]}, *focus*: {}}. F_3 differs significantly from F but aligns closely with the context.

We observe that the rewritten text of fact candidates after completion closely matches the original. The generated and refined facts indicate the intent comprehension and supports the further visual generation.

3.3 Visual Generation

Contextualized data facts, bridging narratives and visuals, outline the story structure within data narratives. To deliver a vivid experience, Narrative Player then maps data facts to visuals, organizes them into sequences, and synthesizes videos that incorporate animated transitions, audio, and subtitles across multiple channels.

3.3.1 Data Fact-Visualization Mapping

Inspired by existing mapping relations between data, task, and visual marks [59, 74], we map each data fact candidate to the Vega-Lite visualizations [50]. based on their *type*, *breakdowns*, and *measure*. Narrative Player emphasizes *focus* of data facts by adjusting the opacity and stroke of relevant visual elements. For example, the data table of "Grades" in Fig. 5 includes four attributes: $\{gender, first\_test, second\_test, desk\}$. The partial fact of 2 is $\{type: comparison, measures: \{first\_test\}, breakdowns: \{gender\}\}$, while fact of 3 differs by having $\{measures: \{first\_test, second\_test\}\}$. This distinction influences the x and y-axis settings in visuals 2 and 3, where the selection of visualization data strictly follows the fact, assuming pre-processed input tables. Currently, we support basic types such as (stacked, grouped) bars, (multi-series) lines, points, and ticks. In line with real-world data video practices and the prior empirical study [70], we incorporate side-by-side visualizations for the "compare" and "correlate" fact types, using *breakdowns* as the aligned axis, displayed either vertically or horizontally.

Narrative Player enhances visualization sets for consistency between text and visuals, drawing on Qu and Hullman's [48] consistency principles for Multiple Views visualization. Narrative Player standardizes the color and scale across charts in these sets. It standardizes color and scale across charts by extracting fields from visuals and using LLM to match comparable fields. LLM recommends color palettes that align with the semantics [70], and uniformity in color, scale, and order is ensured based on whether fields are the same or not. As illustrated by visual 10 and visual 11 in Fig. 1, they share the same y-axis range and use varying shades of red to differentiate the three comparable temperature fields semantically. Moreover, it applies consistent visual effects, like stroke or opacity, to highlight story patterns (Fig. 5).

Additionally, for vague clauses, Narrative Player adjusts visualizations to better align with the narrative (R3). For vague clauses initiating a sequence, visual emphasis is removed to prevent premature data insights, as shown in visual 2 of Fig. 1; otherwise, it mimics the previous visualization. For clauses lacking data facts, mid or end clauses maintain the prior visualization, while starting sentences adopt features from the next one, keeping only the axes and title.

3.3.2 Visualization Sequence Selection

After mapping data facts to visuals for each clause, we select one visualization per clause from the candidates to create a sequence that represents the data story. Narrative Player selects the optimal visualization sequence $V = (V_0, V_1, V_2, ..., V_n)$ from a large set, where V_0 is a *null* spec for specific-to-general selection [30], and other V_i are clause-specific visualizations. Similar to prior works [62, 78], we adopt a global optimization of the visual sequence, instead of incremental methods adopted in the conversational NLI system [70], to ensure dynamic consistency and contextualization (R4). Inspired by literature [4, 26, 30, 62], Narrative Player considers three heuristic features: 1) the cost of dynamic transitions of side-by-side visualizations and fixed elements for maintaining local similarity and consistency, 2) the emphasis on visual focus for insight understanding, and 3) the activation of the primary visualization for global contextualization.

Minimize the dynamic transition cost by considering side-by-side visualizations and fixed elements. We use the transition cost model from Graphscape [30] as a basic operator, which has been widely used in prior studies [35, 58, 63, 78] for measuring visualization similarity and ranking visualization sequences. Narrative Player further considers 1) side-by-side visualizations and 2) fixed elements in dynamic visuals to ensure local visual similarity. By minimizing the sum of transition costs $\mathscr{T}$ as follows, adjacent clauses that align with similar visualizations within their semantic scope are ensured, thereby addressing the challenge of verifying facts for vague clauses.

$$\mathscr{T} = \sum_{i=1}^{|V|} T''(V_{i-1}, V_i) \tag{2}$$

where $T''(V_{i-1}, V_i)$ is the transition cost variants combination between two adjacent visualizations. Given that a clause may link to side-by-side visualizations when the fact *type* is either "compare" or "correlate", we denote $V_{i-1} = \{s_1, s_2\}$ and $V_i = \{e_1, e_2\}$, where s_i and e_j are individual visualizations extracted from V_{i-1} and V_i, respectively. This results in four possible transition scenarios: one-to-one, one-to-two, two-to-one, and two-to-two:

$$T''(s,e) = \begin{cases} T'(s_1,e_1) & s_2 = null, e_2 = null \\ T'(s_1,e_1) + T'(s_1,e_2) & s_2 = null, e_2 \neq null \\ T'(s_1,e_1) + T'(s_2,e_1) & s_2 \neq null, e_2 = null \\ \min\left(T'(s_1,e_1) + T'(s_2,e_2), T'(s_1,e_2) + T'(s_1,e_1)\right) & s_2 \neq null, e_2 \neq null \end{cases} \tag{3}$$

where $T'(s_i, e_j)$ is the variant of static visualization transition cost for one-to-one transition. For a one-to-two or two-to-one transition, we sum the costs of two pairs of visualizations. For a two-to-two transition, the minimum sum of the intersecting transition costs is chosen, as each visualization from the initial clause moves independently to the next, simplifying the need for comprehensive combinations (see visuals 8, 9, and 10 in Fig. 1). The one-to-one transition computation differs from the standard static visualization transition cost $T(s_i, e_j)$:

$$T'(s_i,e_j) = \begin{cases} T(s_i,s_i') + T(e_j',e_j) & \text{if } s_i \text{ and } e_j \text{ have varied fields with join keys} \\ T(s_i,e_j) & \text{otherwise} \end{cases} \tag{4}$$

The interim states s_i' and e_j' of s_i and e_j, respectively, limit the data range to their shared data, as all other visual aspects stay the same. This method arises from noting that in dynamic visuals, maintaining specific elements like bars in joined visualizations minimizes inconsistencies due to data alterations (see visuals 10, 10', 11', and 11 in Fig. 1).

Emphasize the visual focus. Heuristic analysis suggests that a clear clause typically seeks to draw users' attention to key insights, often achieved through specific visual cues like opacity, stroke, or annotation. Hence, we design a visual focus bonus. $\mathscr{B}$:

$$\mathscr{B} = \sum_{i=1}^{|V|} B(V_i) \tag{5}$$

If V_i corresponds to clear clause and its *focus* $\neq$ null, we set $B(V_i)$ as 1; otherwise, it's 0. This ensures Narrative Player doesn't always favor visual sequences with a narrow data range lacking visual focus. Whereas, vague clauses don't receive this consideration since they typically function as transitions between insights rather than highlighting specific visual focuses.

Activate the primary visualization. Data videos often feature a primary visualization that either stays visible for long stretches or frequently returns, anchoring the topic and maintaining context. In computational psychology, models are proposed to measure the activation and retrieval probabilities of items in working memory. These models consider each item's exposure duration along with mechanisms of forgetting and interference, capturing how cognitive load and time influence memory dynamics [4, 8, 33]. Drawing from these models, Narrative Player encodes the activation level of individual visualizations as $A = \{A_1, A_2, ..., A_n\}$, and encodes the retrieving probability of primary visualization as $\mathscr{P}$:

$$\mathscr{P} = \max_i \frac{e^{A_i}}{\sum_j e^{A_j}} \tag{6}$$

$$A_i = \sum_k (\alpha + \beta n_{i,k}) \tag{7}$$

where $n_{i,k}$ refers to the number of continuous clauses when the i^{th} visualization appear for the k^{th} time, while α and β are two parameters for linear activation. Due to the complexity of a fully realized computational model for working memory [8, 33], we use a linear model, bypassing factors like forgetting and interference, to determine the primary visualization's prominence. Setting α at 1 and β at 0.5 has proven effective in our context.

We aim to identify a visualization sequence that optimally satisfies all discussed factors. Thus, the overall optimization function $\mathscr{F}$ is defined as the weighted sum of these three factors.

$$\mathscr{F} = \omega_1 \cdot \mathscr{T} + \omega_2 \cdot \mathscr{B} + \omega_3 \cdot \mathscr{P} \tag{8}$$

Directly solving the optimization problem can result in extended computation times, especially with a large number of clauses generating numerous potential facts (e.g., exceeding 15 with each clause yielding between 3 to 8 potential facts). To mitigate this, we employed pruning strategies focused on maintaining visual consistency and insightfulness (R4), specifically when: 1) the same data field is represented in multiple visualization types, and 2) the sequence lacks a distinct visual focus. These strategies significantly cut down the generation time of visualization sequences to minutes, efficiently preserving performance.

3.3.3 Enhance Visual Sequence with Multiple Channels

Narrative Player provides a multi-channel enhanced experience by adding transitions to illustrate the vivid visual story and maintain textual information through audio narration and subtitles.

To achieve seamless transitions between narrative segments, we further apply transitional animation effects on the visualization sequence to guide the viewer through the sequence (R5). We employ Gemini [29], a system that combines declarative grammar with recommendations for animated transitions, to create the transition effects between individual visualizations. The overall transitions are determined by the transition cost calculation logic discussed in Eq. 3 and Eq. 4:

- **No transition.** For adjacent visualizations V_{i-1} and V_i that don't share data fields and lack joined data, apply no transition effects.
- **One-to-One Transition.** We enable smooth transitions between s_i and e_j recommended by Gemini. Especially, when they involve varied fields with joined keys as described in Eq. 3, we interpolate s_i' and e_j' with joined data between s_i and e_j, as the transition between visual 10 and 11 in Fig. 1.
- **One-to-Two and Two-to-One Transitions.** The visualization pair with a smaller transition cost is the primary one to transition in Eq. 4. In a one-to-two scenario, after this primary transition, the secondary visualization emerges in the scene, as shown in Fig. 1 when visual 8 transitions to 9. Conversely, in a two-to-one scenario, one visualization vanishes first, followed by the primary transition of the visualization pair.
- **Two-to-Two Transition.** The two selected visualization pairs, with the minimal transition cost combination as illustrated in Eq. 4, transition simultaneously.

Furthermore, narration segments are converted into audio narration using text-to-speech technologies [3]. By aligning the start and end times of each clause in the audio narration, Narrative Player ultimately renders the visualization sequence, transition animations, and audio narration into a data video with narration-animation interplay. Subtitles are also added and serve as another channel to present original text information for the reading experience.

4 Evaluation

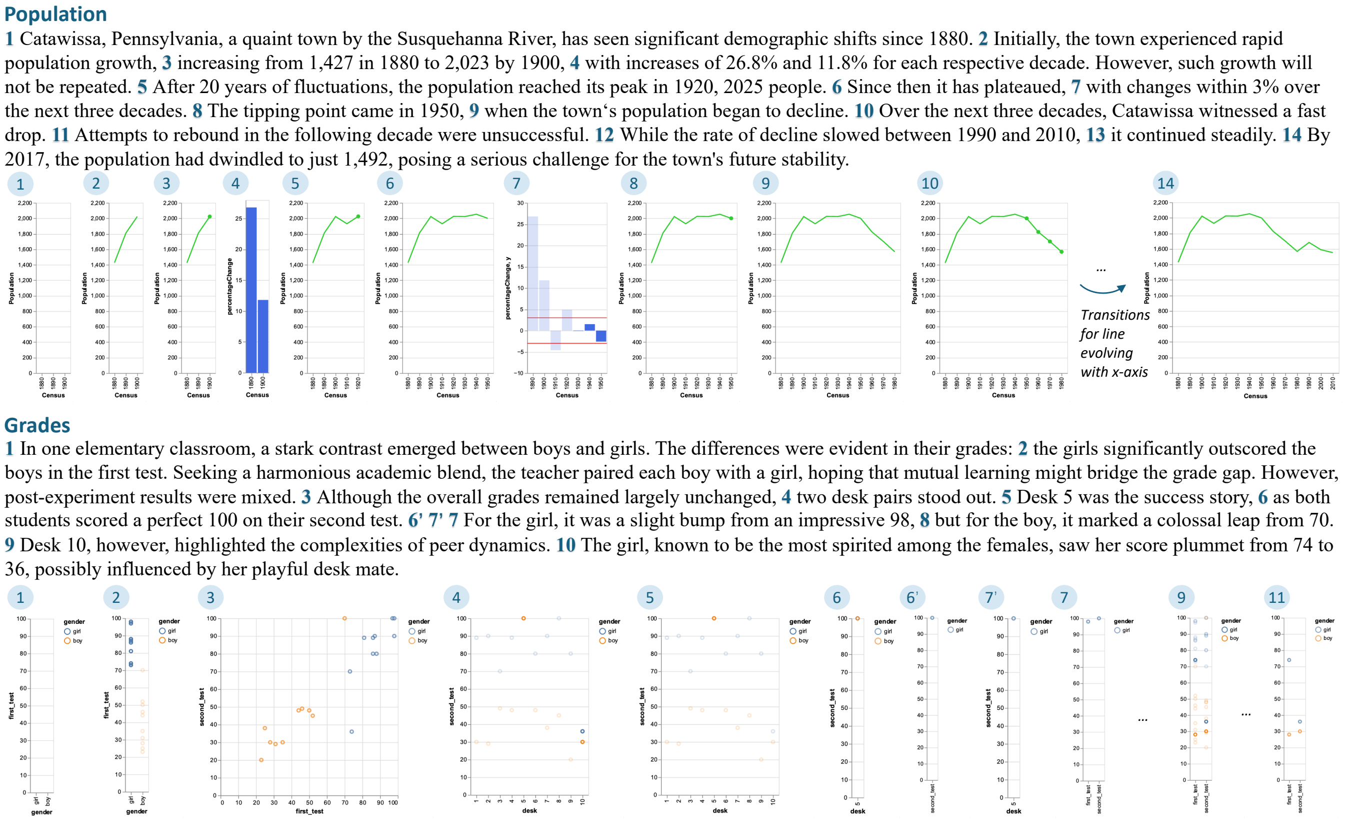
We evaluated Narrative Player's usability through three approaches. First, real-world datasets were used to produce Data Videos, forming a gallery for illustration. Next, a user study illustrated the experience enhancement, video quality, and technical effectiveness of Narrative Player by comparing automatic-generated videos with plain text and those from ablation studies and human-made ones. Lastly, expert feedback gathered through detailed multidimensional ratings and interviews provided in-depth insights and implications.

4.1 Example Gallery

We collected diverse datasets and texts from public websites and generated an example gallery with Narrative Player, encompassing various visualization types (e.g., bar, line, point) and narrative themes (e.g., population, weather, COVID-19). Beyond Fig. 1 and Fig. 5, additional examples can be found in `https://datavideos.github.io/Narrative_Player/`.

4.2 User Study

This study 1) analyzes the experience boost by videos against plain text, 2) assesses the quality of videos automatically generated against human-composed videos, and 3) evaluates the effectiveness of two main modules by ablation studies.



Population

1 Catawissa, Pennsylvania, a quaint town by the Susquehanna River, has seen significant demographic shifts since 1880. **2** Initially, the town experienced rapid population growth, **3** increasing from 1,427 in 1880 to 2,023 by 1900, **4** with increases of 26.8% and 11.8% for each respective decade. However, such growth will not be repeated. **5** After 20 years of fluctuations, the population reached its peak in 1920, 2025 people. **6** Since then it has plateaued, **7** with changes within 3% over the next three decades. **8** The tipping point came in 1950, **9** when the town's population began to decline. **10** Over the next three decades, Catawissa witnessed a fast drop. **11** Attempts to rebound in the following decade were unsuccessful. **12** While the rate of decline slowed between 1990 and 2010, **13** it continued steadily. **14** By 2017, the population had dwindled to just 1,492, posing a serious challenge for the town's future stability.

Grades

1 In one elementary classroom, a stark contrast emerged between boys and girls. The differences were evident in their grades: **2** the girls significantly outscored the boys in the first test. Seeking a harmonious academic blend, the teacher paired each boy with a girl, hoping that mutual learning might bridge the grade gap. However, post-experiment results were mixed. **3** Although the overall grades remained largely unchanged, **4** two desk pairs stood out. **5** Desk 5 was the success story, **6** as both students scored a perfect 100 on their second test. **6' 7' 7** For the girl, it was a slight bump from an impressive 98, **8** but for the boy, it marked a colossal leap from 70. **9** Desk 10, however, highlighted the complexities of peer dynamics. **10** The girl, known to be the most spirited among the females, saw her score plummet from 74 to 36, possibly influenced by her playful desk mate.

Fig. 5: Two examples automatically generated by Narrative Player, each featuring a visualization sequence with corresponding transition animations. Transition animations are activated when the audio narration hits the relevant segments, with representative transitions highlighted in the figure.

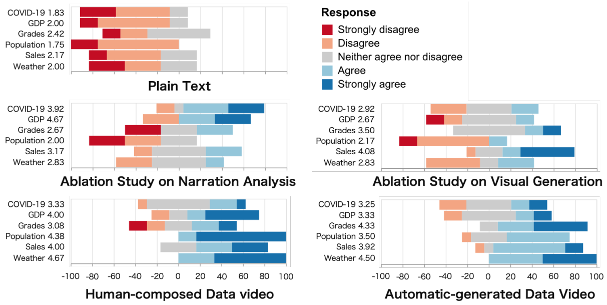


Fig. 6: User ratings for general satisfaction with the five materials on a 5-point Likert scale.

4.2.1 Dataset

Our study uses datasets based on real-world narratives, including weather in Zhaoqing ("Weather"), the population in Catavissa ("Population"), COVID-19 regional patterns ("COVID-19"), GDP of top economies ("GDP"), course grades ("Grades"), and corporate sales ("Sales"). The "Weather" and "Population" data come from the ToTTo dataset [46] tailored for table-to-text tasks and are enhanced by an expert for richer storytelling. "COVID-19" data is from the WTO and also features in the CrossData [12], "GDP" from the World Bank and its affiliated blogs, and "Grades" and "Sales" are inspired by online data stories [1, 2], all refined to align with narrative styles.

4.2.2 Preparation

To evaluate the effectiveness of Narrative Player and its two main modules, we prepared five storytelling versions for each narrative: an automatically generated data video by Narrative Player, plain text for studying experience boost, a human-crafted data video by visualization researchers for assessing quality, and two data videos from ablation studies for learning technical nuances. Since it's a new and challenging task, there is no previous baseline available for comparison, and other rule- or template-based alternatives cannot be quickly designed to satisfy the design requirements in Sec. 3.1.1.

Manual-composed by human. We enlisted two experienced visualization researchers, each with over three years in the field and publications in significant conferences like IEEE VIS and ACM CHI, to craft visual sequences and videos. Their selection was based on their proficiency with Vega-Lite for visualization and their background in animation and data video creation, ensuring that the designs were practical and grounded, thus avoiding comparisons with unfeasible, overly imaginative concepts. To guarantee a fair comparison, they were initially briefed on Narrative Player's core functionalities and design space. They then outlined their visual designs for specific narrative segments, either verbally or through sketches, which our programmer implemented. Through iterative refinement with these researchers, we ensured the final visual sequences and videos aligned with their visions.

Ablation study on narration analysis. This ablation study, termed "Ablation1," assessed the narrative analysis module's effectiveness in extracting contextualized data facts. We applied the module to focus solely on clear clauses, as described in Sec. 3.2.3, using these facts as anchors. Facts from adjacent vague clauses were treated as replications of these clear anchor facts, leading to visuals generated without the detailed inference and completion of vague facts within context.

Ablation study on visual generation. In this ablation study, termed "Ablation2," we aim to validate the advantages of including visual focus and primary visualization in the optimization function for generating visualization sequences. We focused solely on the first term of Eq. 8, which involves minimizing the transition cost and has been empirically proven to be reasonable and necessary for creating data videos [62].

4.2.3 Qualitative Analysis of Perceived Satisfaction

We enlisted 12 participants (7 males, 5 females), labeled P1-P12, with backgrounds in reading data-rich documents, including data analysts and graduate students in fields like visualization (VIS), machine learning, data science, software engineering, human-computer interaction (HCI), and design. The study was conducted with some participants attending in-person and others virtually. Each participant received six sets of materials comprising data tables and five versions of storytelling versions (data narratives and four videos per narrative), as listed in Sec. 4.2.2. The order of narratives and videos was randomized. Participants reviewed the narratives and tables before watching the videos on their own computers, rating them on a 5-point Likert scale for general satisfaction via Google Forms. They could re-watch videos and were encouraged to share their thoughts openly throughout the 60-70 minute study. Participants received a $15 for their time.

Videos engage readers more effectively than text, with quality determining their superiority. As indicated by the rating results, all the videos generated by Narrative Player and those crafted by humans consistently outperformed text. P2 remarked, "*Videos with visuals and animations are way more attractive. They tell the story more vividly and engagingly than plain text.*" P9 added, "*As the visuals shifted, I could tell there is a new insight coming. The animations helped me saw the links between insights by showing what stayed the same and what changed.*" However, the text did not invariably score lower than the other four videos. For example, the "Population" story in both ablation studies received a lot of "strong disagreement" feedback, which scored critically low with mean scores of 2.00 and 2.17. P3, while viewing a video from the ablation study on visual generation, noted, "*With visuals constantly toggling and axes lacking clear relation, the conveyed information seems too muddled. I prefer plain text under this condition.*" Similarly, P5 remarked on videos from the ablation study on narrative analysis, "*The years mentioned in the voiceover seem randomly shown as the whole x-axis or just highlighted points. The x-axis range keeps changing and I can't follow the context.*" Ensuring the consistency and contextualization of generated visualization sequences and videos is paramount for users to perceive their superiority over text.

Automatically generated videos are generally well received by users. The user ratings in the study are shown in Fig. 6. All six narratives scored an average of above 3, with none receiving a "Strong Disagree" rating, a unique achievement among the five renditions. "Population" and "GDP" underperformed compared to human-composed videos, whereas the other four narratives matched or surpassed them, notably "Grades" (Mean Score: 4.33 vs. 3.08). Against ablation on Narration Analysis, Narrative Player excelled in all but "GDP" and "COVID-19" narratives, showing no significant advantage in "GDP" (Mean Score: 3.33 vs. 3.67) and underperforming in "COVID-19" (Mean Score: 3.25 vs. 3.92). For ablation on visual generation, it excelled except for "Sales" with similar scores (Mean Score: 3.92 vs. 4.08). All other ratings are as expected, except for the "Grades" of the human-composed version, which will be illustrated in the expert study. During experiments, participants occasionally expressed surprise at Narrative Player's capability to capture vague details and their consequent visual representations. P1 commented, "*The continuously rising line chart in 'Population' was striking. Though the text didn't specify the entire range of the x-axis, the algorithm's ability to track the timeline's progression was impressive.*" While reviewing "Weather" (Fig. 1), P4 observed, "*Presenting temperature and precipitation together was unexpected, yet it clearly illustrated the interplay between sunshine and rainfall.*" Such feelings attributed to the analysis of clauses in data fact inference and completion within context.

The role of both modules is positively related to narrative complexity. Despite the general lower scores received by videos in both ablation studies, two cases illustrate the impact of narrative complexity on the significance of modules. "GDP" and "COVID-19" showed no significant score difference between the ablation of narrative analysis and the automatic videos, with both narratives explicitly conveying the data facts. This indicates that when narratives predominantly contain clear facts, or when a narrative's information is sparse but mostly comprises vague clauses with minimal insights, simple replication using clear data facts as anchors is effective. For "Sales", the absence of significant differences in the visual generation ablation study indicates that if a narrative consistently involves the same data field, resulting in facts with similar *context* and *type*, the impact of the visual generation module's ablation is minimized.

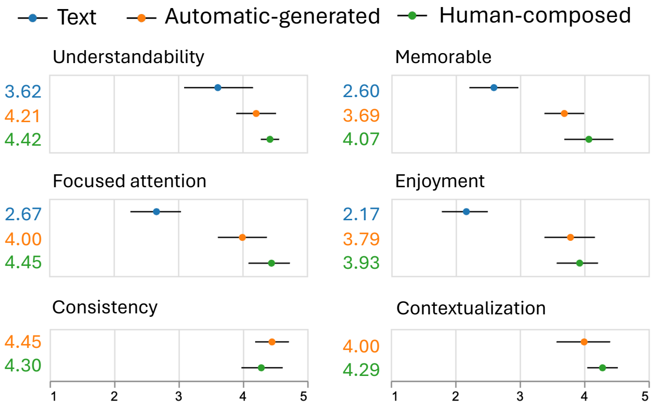


Fig. 7: Expert ratings on six dimensions for text and two kinds of videos with different color legends. Means are represented by shapes, while their 95% CIs are illustrated with error bars.

4.3 Expert Study

To provide an in-depth analysis and identify the differences between automatic-generated videos and human-composed ones [60], we invited four designers (D1-D4), who daily crafted videos with professional editing tools (e,g., After Effects), and three visualization researchers (V1-V3), who have publications in top-tier conferences and journals in VIS and HCI. After acquainting them with the background described in Sec. 4.2.2 and without disclosing the video versions, they watched and rated six sets of narratives, automatic-generated videos, and human-composed ones, used in the user study. They gave ratings across various dimensions: understandability, memorability, focused attention, enjoyment, consistency, and contextualization. The first four dimensions aim to boost user experience, with a focus on the last two due to their technical importance and the challenge of quantifying them without ground truth. All three versions were rated based on the first four dimensions, but only two videos were assessed on the last two, given their visual relevance. After the ratings, we revealed the video versions and conducted interviews where experts explained their ratings and analyzed the differences between automatic and manual production.

4.3.1 Effectiveness Across Dimensions and Audiences

Overall, as shown in Fig. 7, both videos significantly outperform the text across all dimensions, with the human-composed ones being rated marginally higher than the automatic-generated ones. All experts concurred that visual sequences generated by Narrative Player align reasonably well with the text and effectively capture and showcase contextual relationships. Thereby, it mitigates the limitations imposed by gaps in textual information, as evidenced by the high understandability rating. These sequences liberate viewers from the confines of information-dense text, enabling a more vivid understanding of the data story and providing new possibilities for presentations, as highlighted by significant differences in the three dimensions for engagement . They praised the alignment between text and visuals in the generated videos and the consistency across visuals, with ratings on par with those of human-composed videos for the two technical dimensions . Under the same design space, human-composed versions take several hours to iterate, while Narrative Player generates comparable videos in minutes.

Experts agreed on effectiveness in most cases, but their ratings differed in a few, possibly due to audience background. Like some participants in the user study, D2 felt that the "Grades" of the human-composed version were not satisfactory because it was hard to quickly understand insight (the change in grades) from a scatter plot using two test scores as the x and y axes, and the highlight in a single chart is a bit boring. In contrast, V1 and V2 disagreed, noting that visual researchers have a stronger grasp of such visuals, while designers may favor more engaging animations. This highlights the preferences of different audiences.

4.3.2 Differences between automatic and manual production

In our interviews, experts analyzed differences between automatic and manual versions.

Annotations. The distinction was most evident in the utilization of annotations. D1 noted, "*Human-created videos typically featured detailed annotations, such as the bolder red line and clearer text labels in 'Weather', which were missing in the automatically generated videos.*" In fact, automating decisions on when to annotate, and determining its optimal size and position—considering potential occlusions and placements of visual elements—is inherently challenging.

Detailed control of visual elements. V1 appreciated the absence of a legend at the beginning of the human-created "Sales", which was introduced only when the data dimensions escalated, while Narrative Player currently didn't support it. "*Gradually introducing visual elements by the textual content helps me focus on the key points,*" V1 commented. It suggests that the elements of text-visual mapping in narrative visualization need to be fine-tuned at the level of individual visual elements.

Flexible introduction of video. Moreover, human-created videos frequently incorporated refined treatments for the introduction of video. For instance, in cases like "Population" and "GDP", they began with a complete visualization without a visual focus to set the background and subsequently transitioned or highlighted specific data points. Such approaches were not uniformly adopted across cases, which depended on whether the introductory sentence had a clear link with the subsequent content. Participants found such flexible treatment engaging, whereas Narrative Player's consistent choice to begin with a blank visualization appeared somewhat formulaic.

Flexible alignment between visuals and narration. Moreover, D1 noted that visuals sometimes should convey information not explicitly stated in the narration. We partially achieved this through Narrative Player's data fact inference and completion within context. Besides, this requirement involves employing visual patterns that transcend local text to describe the global narrative structure, which Narrative Player does not support. For instance, D1 highly praised the human-composed version of "Population", as it concurrently maintains both the line chart for population data and the bar chart for changing rates even if the latter part of the narrative doesn't contain information about percentage change. Correspondingly, D1 critiqued the automatic-generated version, which is consistent with the ratings in the user study: "*These visualizations do adhere to the textual descriptions, but the strict correspondence occasionally disrupts the viewer's thought process by sometimes displaying population growth rates.*" Similarly, some visuals corresponding to unimportant data facts for vague clauses can be intentionally skipped. Otherwise, although we set the minimal animation duration, frequent transitioning between multiple visualizations and rapidly changing axes may disorient the viewer.

5 Discussion

Mitigating potential side-effects of LLMs. Our current system significantly reduces factual inaccuracies through a verification mechanism using sentence embedding and a multi-session approach, ensuring a reliable set of data facts aligns with the narrative. We recognize the potential limitations in the generalizability and accuracy of our method due to the subjective nature of clause classification and parameter setting. Moreover, challenges such as unstable performance and slower output speed persist. We believe that these issues can be addressed by the rapid advancement of LLM research. Interactive tools enabling users to adjust LLM outputs and workflows [76], such as modifying thresholds or examining clauses, and design of multi-agent systems [56] could enhance narrative analysis.

Consideration of more complex writing structure and longer narrative. Narrative Player currently analyzes paragraphs of the narrative by examining clauses and their neighbors for local context and introduces primary visualization for a global perspective. Yet, long narratives often display structured relationships—parallel, progressive, echoing—that demand a nuanced narrative analysis. Narrative structure directly influences visualization selections, with implicit design guidelines suggesting, for instance, consistent visualization styles for parallel or echoing data facts. These patterns have been noticed in the visual storytelling community [10, 62], but how visuals are selected and arranged when the narratives comprise in-depth writing strategy and styles remains to be studied. Moreover, in data videos, insights can be discontinuous and scattered across long segments. For extended narratives, identifying less critical insights or segments lacking data and employing computer vision to create engaging video content is crucial to retain viewer interest.

Exploring more reliable evaluation methods for narrative-driven visual storytelling. The NLP community has begun preliminary studies on evaluating LLMs for understanding and generating coherent stories [9], while quantitative evaluation of the technical validity of visual storytelling remains a significant challenge. Due to the diverse preferences and lack of ground truth in storytelling, we currently rely on user studies instead of verifying "accuracy." To make the study paradigm more reliable, we require broader input data narratives, increased participation from a diverse range of designers and researchers to provide varied human-composed versions, and extensive user crowdsourcing experiments. These help understand and differentiate storytelling preferences among people for various types of data narratives, which contributes to more appropriate experiments and analysis.

Advanced considerations beyond consistency and contextualization. Visuals and videos as educational tools suggest integrating cognitive and educational theories into visual storytelling, especially those related to working memory and cognition. Our use of primary visualization and working memory theory-based modeling in the second module is a step toward incorporating cognitive principles. Indeed, literature on cognitive theory often employs visualization or video as illustrative examples [6, 11, 51] and studies in infovis [5, 47] have integrated cognitive theory to examine the efficacy and complexity of individual visualizations. Extending such approaches to animated visualization sequences and videos could prove beneficial. Additionally, addressing personal and emotional expression, as highlighted by user feedback, is crucial. Narrative Player lacks automated emphasis detection in narratives and adjustable visual and audio cues to highlight key points. To meet user needs for customization and emotional expression, future enhancements could include more design options and interactive settings for users to personalize their experience, alongside employing contextualized icons or glyphs for more engaging videos.

Extension of our pipeline. The narrative analysis module's handling of ambiguous semantics provides a reference for generating contextualized data facts based on long narratives, and it can be expanded to generate or assist in creating data comics, scrollytelling, and other narrative visualizations related to data facts [37, 74]. Similarly, the consideration of multiple visualizations, visual focus, and primary visualizations in the visual generation module can be applied to other works for sequencing and ordering. Our pipeline could also be generalized to broader authoring with visuals as input: imagine a data analyst creating a data video by initially conducting simple data analysis and writing a data narrative, with the process-generated visuals serving as additional visual input, which could be used as fixing visuals for certain key clauses or primary visualizations in the sequence. In addition, there are also some scenarios where authors do not create visuals. For example, using LLM to interpret data tables and seamlessly craft narratives is a viable creative approach. The Narrative Player, as it stands, is effectively suited for those creators without requiring modifications.

Limitations and future work. Currently an end-to-end automated tool, we aim to expand it into a full-scale authoring platform for visual storytelling. It autonomously generates and refines narratives based on data input enabled by human-LLM co-writing experience, with options for human-in-the-loop customization [34, 49]. For example, users may specify color palettes, the primary visualization, and determine the visuals corresponding to certain clauses as anchors. This streamlines the creation of personalized, high-quality data videos while reducing computational load. Additionally, narration may mention complex data with high dimensions or incorporate multiple facts with varied insights. To manage this complexity, two approaches could be considered: selecting the facts with the most important insight to avoid overly complex visuals, and enhancing visual expressiveness to cover more facts and comprehensive semantics. Moreover, the visual types and transitions available in Narrative Player are limited to those supported by GraphScape [30] and Gemini [29]. We will explore incorporating a broader variety of visualizations and associated visual cues to enhance visual expressiveness and tell more diverse stories. Also, the default Vega-Lite settings currently used to map facts to visualizations can result in issues like narrow layouts or small visual marks. In the future, we should consider more aesthetic and functional factors in visual design, such as responsiveness for layouts and visual elements.

6 Conclusion

This paper introduces Narrative Player, a tool that generates animated visualization sequences from data narratives and tables, emphasizing consistency and contextualization. Leveraging LLMs and sentence embedding, it interprets narratives to extract contextualized data facts. After mapping facts to visualizations, it optimize them into a coherent sequence enriched with transition animations. The resultant data video, combining animated visuals and audio narration, received positive feedback in the user study and expert study, affirming its comparability to human-created videos and the effectiveness of its core modules. Encouraged by the feedback, we aim to further refine Narrative Player to more closely match the quality of real-world data videos.

References

[1] A complete guide to stacked bar charts, 2023. https://chartio.com/learn/charts/stacked-bar-chart-complete-guide/. 7

[2] Data visualization in matplotlib, 2023. https://blog.adnansiddiqi.me/data-visualization-in-python-scatter-plots-in-matplotlib/. 7

[3] Microsoft azure text-to-speech service., 2023. https://azure.microsoft.com/zh-cn/products/cognitive-services/text-to-speech/. 6

[4] E. M. Altmann and C. D. Schunn. Decay versus interference: A new look at an old interaction. *Psychological Science*, 23(11):1435–1437, 2012. 5, 6

[5] E. W. Anderson, K. C. Potter, L. E. Matzen, J. F. Shepherd, G. A. Preston, and C. T. Silva. A user study of visualization effectiveness using eeg and cognitive load. In *Computer graphics forum*, vol. 30, pp. 791–800. Wiley Online Library, 2011. 9

[6] P. Ayres and F. Paas. Making instructional animations more effective: A cognitive load approach. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, 21(6):695–700, 2007. 9

[7] S. K. Badam, Z. Liu, and N. Elmqvist. Elastic documents: Coupling text and tables through contextual visualizations for enhanced document reading. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):661–671, 2019. 2, 3, 4

[8] P. Barrouillet, S. Bernardin, S. Portrat, E. Vergauwe, and V. Camos. Time and cognitive load in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(3):570, 2007. 6

[9] D. Callan and J. Foster. How interesting and coherent are the stories generated by a large-scale neural language model? comparing human and automatic evaluations of machine-generated text. *Expert Systems*, p. e13292, 2023. 9

[10] Y. Cao, J. L. E, Z. Chen, and H. Xia. DataParticles: Block-based and Language-oriented Authoring of Animated Unit Visualizations. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 1–15. ACM, 2023. 2, 3, 9

[11] J. C. Castro-Alonso, P. Ayres, and J. Sweller. Instructional visualizations, cognitive load theory, and visuospatial processing. *Visuospatial processing for education in health and natural sciences*, pp. 111–143, 2019. 9

[12] Z. Chen and H. Xia. CrossData: Leveraging Text-Data Connections for Authoring Data Documents. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 1–15. ACM, 2022. 2, 3, 7

[13] H. Cheng, J. Wang, Y. Wang, B. Lee, H. Zhang, and D. Zhang. Investigating the Role and Interplay of Narrations and Animations in Data Videos. *Computer Graphics Forum*, 41(3):527–539, 2022. 3, 4

[14] P. Chi, Z. Sun, K. Panovich, and I. Essa. Automatic Video Creation From a Web Page. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*, pp. 279–292. ACM, 2020. 3

[15] M. Conlen and J. Heer. IdylL: A markup language for authoring and publishing interactive articles on the web. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*, pp. 977–989. ACM, 2018. 2, 3

[16] W. Cui, X. Zhang, Y. Wang, H. Huang, B. Chen, L. Fang, H. Zhang, J. G. Lou, and D. Zhang. Text-to-Viz: Automatic Generation of Infographics from Proportion-Related Natural Language Statements. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):906–916, 2020. 2

[17] K. Dhamdhere, K. S. McCurley, R. Nahmias, M. Sundararajan, and Q. Yan. Analyza: Exploring data with conversation. In *Proceedings of the 22th International Conference on Intelligent User Interfaces*, pp. 493–504. ACM, 2017. 2

[18] N. K. Duke and P. D. Pearson. Effective practices for developing reading comprehension. *Journal of education*, 189(1-2):107–122, 2009. 1

[19] E. Fast, B. Chen, J. Mendelsohn, J. Bassen, and M. S. Bernstein. Iris: A conversational agent for complex tasks. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 1–12. ACM, 2018. 2

[20] Y. Feng, X. Wang, B. Pan, K. K. Wong, Y. Ren, S. Liu, Z. Yan, Y. Ma, H. Qu, and W. Chen. XNLI: Explaining and Diagnosing NLI-based Visual Data Analysis. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–14, 2023. 2

[21] T. Gao, J. Hullman, E. Adar, B. Hecht, and N. Diakopoulos. NewsViews: An automated pipeline for creating custom geovisualizations for news. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 3005–3014. ACM, 2014. 1, 2

[22] A. M. Glenberg and W. E. Langston. Comprehension of illustrated text: Pictures help to build mental models. *Journal of memory and language*, 31(2):129–151, 1992. 1

[23] V. Gyselinck and H. Tardieu. The role of illustrations in text comprehension: What, when, for whom, and why? 1999. 1

[24] E. Hoque, V. Setlur, M. Tory, and I. Dykeman. Applying Pragmatics Principles for Interaction with Visual Analytics. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):309–318, 2018. 2

[25] J. Hullman, N. Diakopoulos, and E. Adar. Contextifier: Automatic generation of annotated stock visualizations. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 2707–2716. ACM, 2013. 1, 2, 3

[26] J. Hullman, S. Drucker, N. Henry Riche, B. Lee, D. Fisher, and E. Adar. A deeper understanding of sequence in narrative visualization. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2406–2415, 2013. 2, 5

[27] A. A. Jaffar. Youtube: An emerging tool in anatomy education. *Anatomical sciences education*, 5(3):158–164, 2012. 3

[28] Z. Ji, N. Lee, R. Frieske, T. Yu, D. Su, Y. Xu, E. Ishii, Y. J. Bang, A. Madotto, and P. Fung. Survey of hallucination in natural language generation. *ACM Computing Surveys*, 55(12):1–38, 2023. 4

[29] Y. Kim and J. Heer. Gemini: A grammar and recommender system for animated transitions in statistical graphics. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):485–494, 2020. 3, 6, 10

[30] Y. Kim, K. Wongsuphasawat, J. Hullman, and J. Heer. GraphScape: A model for automated reasoning about visualization similarity and sequencing. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 2628–2638. ACM, 2017. 2, 5, 10

[31] S. Latif, S. Chen, and F. Beck. A deeper understanding of visualization-text interplay in geographic data-driven stories. In *Computer Graphics Forum*, vol. 40, pp. 311–322. Wiley Online Library, 2021. 3

[32] S. Latif, Z. Zhou, Y. Kim, F. Beck, and N. W. Kim. Kori: Interactive Synthesis of Text and Charts in Data Documents. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):184–194, 2022. 2, 3

[33] B. Lemaire and S. Portrat. A computational model of working memory integrating time-based decay and interference. *Frontiers in psychology*, 9:416, 2018. 6

[34] H. Li, Y. Wang, and H. Qu. Where Are We So Far? Understanding Data Storytelling Tools from the Perspective of Human-AI Collaboration. In *Proceedings of CHI Conference on Human Factors in Computing Systems*, pp. 1–28, 2024. 4, 9

[35] H. Lin, D. Moritz, and J. Heer. Dziban: Balancing agency & automation in visualization design via anchored recommendations. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 1–12, 2020. 2, 5

[36] J. Liu, J. Liu, Q. Wang, J. Wang, W. Wu, Y. Xian, D. Zhao, K. Chen, and R. Yan. RankCSE: Unsupervised sentence representations learning

via learning to rank. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 13785–13802. Association for Computational Linguistics, 2023. 4

[37] J. Lu, W. Chen, H. Ye, J. Wang, H. Mei, Y. Gu, Y. Wu, X. L. Zhang, and K.-L. Ma. Automatic generation of unit visualization-based scrollytelling for impromptu data facts delivery. In *IEEE 14th Pacific visualization symposium*, pp. 21–30. IEEE, 2021. 4, 9

[38] Y. Luo, N. Tang, G. Li, J. Tang, C. Chai, and X. Qin. Natural Language to Visualization by Neural Machine Translation. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):217–226, 2022. 2

[39] M. Macdonald-Ross. How numbers are shown: A review of research on the presentation of quantitative data in texts. *AV communication review*, 25(4):359–409, 1977. 1

[40] D. Masson, S. Malacria, G. Casiez, and D. Vogel. Charagraph: Interactive Generation of Charts for Realtime Annotation of Data-Rich Paragraphs. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 1–18. ACM, 2023. 1, 2, 3, 4

[41] F. Moretti. *Distant reading*. Verso Books, 2013. 1

[42] N. Muennighoff. Sgpt: Gpt sentence embeddings for semantic search. *arXiv preprint arXiv:2202.08904*, 2022. 4

[43] A. Narechania, A. Srinivasan, and J. Stasko. NL4DV: A Toolkit for Generating Analytic Specifications for Data Visualization from Natural Language Queries. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):369–379, 2021. 2

[44] S. Niknejad and B. Rahbar. Comprehension through visualization: The case of reading comprehension of multimedia-based text. *International Journal of Educational Investigation*, 5(2):144–151, 2015. 3

[45] Y. Ouyang, L. Shen, Y. Wang, and Q. Li. NotePlayer: Engaging Jupyter Notebooks for Dynamic Presentation of Analytical Processes. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology, UIST'24*, pp. 1–15. ACM, 2024. 2

[46] A. P. Parikh, X. Wang, S. Gehrmann, M. Faruqui, B. Dhingra, D. Yang, and D. Das. Totto: A controlled table-to-text generation dataset. *arXiv preprint arXiv:2004.14373*, 2020. 7

[47] R. E. Patterson, L. M. Blaha, G. G. Grinstein, K. K. Liggett, D. E. Kaveney, K. C. Sheldon, P. R. Havig, and J. A. Moore. A human cognition framework for information visualization. *Computers & Graphics*, 42:42–58, 2014. 9

[48] Z. Qu and J. Hullman. Keeping multiple views consistent: Constraints, validations, and exceptions in visualization authoring. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):468–477, 2017. 5

[49] P. Ren, Y. Wang, and F. Zhao. Re-understanding of data storytelling tools from a narrative perspective. *Visual Intelligence*, 1(1):11, 2023. 9

[50] A. Satyanarayan, D. Moritz, K. Wongsuphasawat, and J. Heer. Vega-lite: A grammar of interactive graphics. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):341–350, 2016. 5

[51] W. Schnotz and C. Kürschner. A reconsideration of cognitive load theory. *Educational psychology review*, 19:469–508, 2007. 9

[52] E. Segel and J. Heer. Narrative visualization: Telling stories with data. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1139–1148, 2010. 2

[53] V. Setlur, S. E. Battersby, M. Tory, R. Gossweiler, and A. X. Chang. Eviza: A natural language interface for visual analysis. In *Proceedings of the 29th Annual ACM Symposium on User Interface Software and Technology*, pp. 365–377. ACM, 2016. 2

[54] V. Setlur and M. Tory. How do you Converse with an Analytical Chatbot? Revisiting Gricean Maxims for Designing Analytical Conversational Behavior. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. ACM, 2022. 2

[55] L. Shen, H. Li, Y. Wang, T. Luo, Y. Luo, and H. Qu. Data Playwright: Authoring Data Videos with Annotated Narration. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–11, 2024. 2

[56] L. Shen, H. Li, Y. Wang, and H. Qu. From Data to Story: Towards Automatic Animated Data Video Creation with LLM-based Multi-Agent Systems. *arXiv: 2408.03876*, pp. 1–8, 2024. 9

[57] L. Shen, E. Shen, Y. Luo, X. Yang, X. Hu, X. Zhang, Z. Tai, and J. Wang. Towards Natural Language Interfaces for Data Visualization: A Survey. *IEEE Transactions on Visualization and Computer Graphics*, 29(6):3121–3144, 2023. 2, 3

[58] L. Shen, E. Shen, Z. Tai, Y. Song, and J. Wang. TaskVis: Task-oriented Visualization Recommendation. In *Proceedings of the 23th Eurographics Conference on Visualization (Short Papers)*, pp. 91–95. Eurographics, 2021. 4, 5

[59] L. Shen, E. Shen, Z. Tai, Y. Xu, J. Dong, and J. Wang. Visual Data Analysis with Task-Based Recommendations. *Data Science and Engineering*, 7(4):354–369, 2022. 2, 5

[60] L. Shen, Y. Zhang, H. Zhang, and Y. Wang. Data Player: Automatic Generation of Data Videos with Narration-Animation Interplay. *IEEE Transactions on Visualization and Computer Graphics*, 30(1):109–119, 2024. 1, 3, 8

[61] D. Shi, Y. Shi, X. Xu, N. Chen, S. Fu, H. Wu, and N. Cao. Task-Oriented Optimal Sequencing of Visualization Charts. In *IEEE Visualization in Data Science*, pp. 58–66. IEEE, 2019. 2

[62] D. Shi, F. Sun, X. Xu, X. Lan, D. Gotz, and N. Cao. AutoClips: An Automatic Approach to Video Generation from Data Facts. *Computer Graphics Forum*, 40(3):495–505, 2021. 2, 5, 7, 9

[63] D. Shi, X. Xu, F. Sun, Y. Shi, and N. Cao. Calliope: Automatic visual data story generation from a spreadsheet. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):453–463, 2020. 2, 5

[64] A. Srinivasan, B. Lee, N. Henry Riche, S. M. Drucker, and K. Hinckley. InChorus: Designing Consistent Multimodal Interactions for Data Visualization on Tablet Devices. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 1–13. ACM, 2020. 2

[65] A. Srinivasan and V. Setlur. Snowy:Recommending Utterances for Conversational Visual Analysis. In *Proceedings of the 34th Annual ACM Symposium on User Interface Software and Technology*, pp. 1–17. ACM, 2021. 2

[66] A. Srinivasan and J. Stasko. Orko: Facilitating Multimodal Interaction for Visual Exploration and Analysis of Networks. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):511–521, 2018. 2

[67] N. Sultanum, F. Chevalier, Z. Bylinskii, and Z. Liu. Leveraging Text-Chart Links to Support Authoring of Data-Driven Articles with VizFlow. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 1–17. ACM, 2021. 2, 3

[68] N. Sultanum and A. Srinivasan. Datatales: Investigating the use of large language models for authoring data-driven articles. In *2023 IEEE Visualization and Visual Analytics*, pp. 231–235. IEEE, 2023. 2, 4

[69] M. Sun, L. Cai, W. Cui, Y. Wu, Y. Shi, and N. Cao. Erato: Cooperative Data Story Editing via Fact Interpolation. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–11, 2022. 2

[70] M. Tory and V. Setlur. Do What I Mean, Not What I Say! Design Considerations for Supporting Intent and Context in Analytical Conversation. In *Proceedings of the 12th IEEE Conference on Visual Analytics Science and Technology*, pp. 93–103. IEEE, 2019. 2, 3, 4, 5

[71] E. R. Tufte. *The visual display of quantitative information*, vol. 2. Graphics press Cheshire, CT, 2001. 1

[72] Y. Wang, Z. Hou, L. Shen, T. Wu, J. Wang, H. Huang, H. Zhang, and D. Zhang. Towards Natural Language-Based Visualization Authoring. *IEEE Transactions on Visualization and Computer Graphics*, 29(1):1222 – 1232, 2023. 2, 3

[73] Y. Wang, L. Shen, Z. You, X. Shu, B. Lee, J. Thompson, H. Zhang, and D. Zhang. WonderFlow: Narration-Centric Design of Animated Data Videos. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–15, 2024. 3

[74] Y. Wang, Z. Sun, H. Zhang, W. Cui, K. Xu, X. Ma, and D. Zhang. Datashot: Automatic generation of fact sheets from tabular data. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):895–905, 2019. 2, 4, 5, 9

[75] E. Winter. *Towards a Contextual Grammar of English: The Clause and its Place in the Definition of Sentence*. Taylor & Francis, 2020. 3

[76] T. Wu, M. Terry, and C. J. Cai. Ai chains: Transparent and controllable human-ai interaction by chaining large language model prompts. In *Proceedings of the CHI conference on human factors in computing systems*, pp. 1–22, 2022. 9

[77] L. Ying, Y. Wang, H. Li, S. Dou, H. Zhang, X. Jiang, H. Qu, and Y. Wu. Reviving Static Charts into Live Charts. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–16, 2024. 1

[78] J. Zhao, S. Xu, S. Chandrasegaran, C. Bryan, F. Du, A. Mishra, X. Qian, Y. Li, and K. L. Ma. ChartStory: Automated Partitioning, Layout, and Captioning of Charts into Comic-Style Narratives. *IEEE Transactions on Visualization and Computer Graphics*, 29(2):1384–1399, 2023. 2, 5