Data Player: Automatic Generation of Data Videos with Narration-Animation Interplay

Leixian Shen, Yizhi Zhang, Haidong Zhang, and Yun Wang

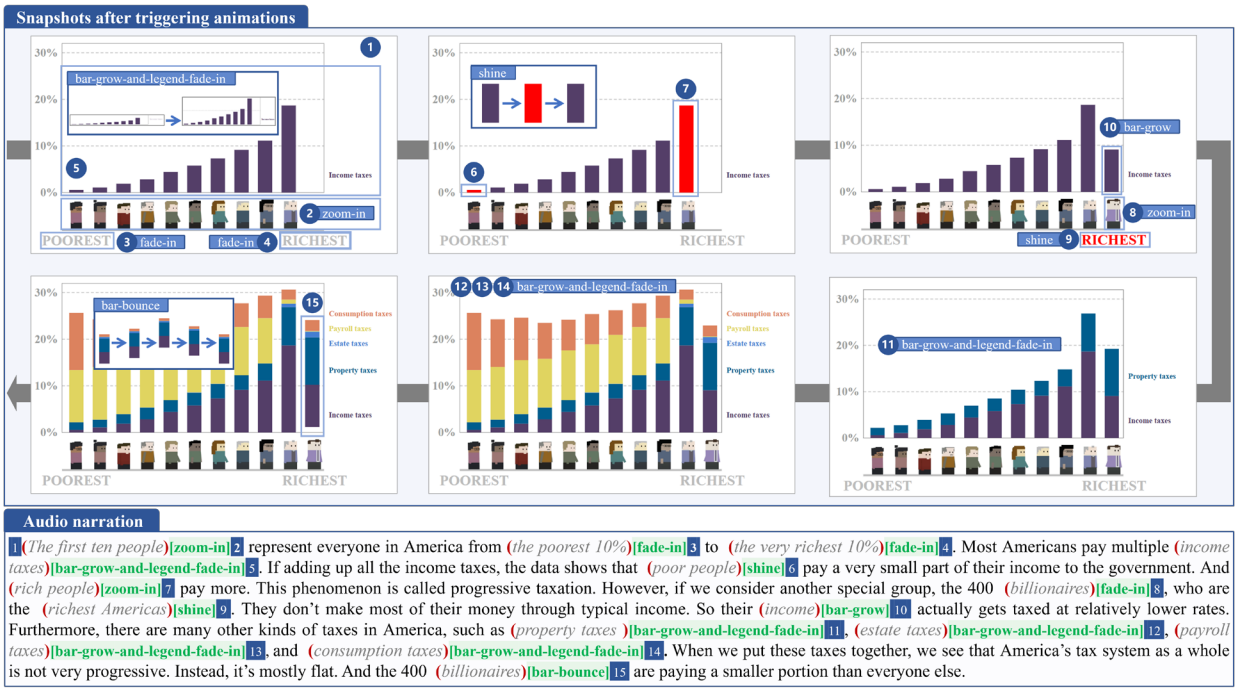


Figure 1: A data video example with narration-animation interplay automatically generated by Data Player from a static visualization and descriptive text, where (*narration segment*) animation-effect x means the animation effect will be triggered when the audio reaches the corresponding narration segment and lasts for the duration of the segment span in the audio, resulting in the x-th keyframe above.

Abstract— Data visualizations and narratives are often integrated to convey data stories effectively. Among various data storytelling formats, data videos have been garnering increasing attention. These videos provide an intuitive interpretation of data charts while vividly articulating the underlying data insights. However, the production of data videos demands a diverse set of professional skills and considerable manual labor, including understanding narratives, linking visual elements with narration segments, designing and crafting animations, recording audio narrations, and synchronizing audio with visual animations. To simplify this process, our paper introduces a novel method, referred to as Data Player, capable of automatically generating dynamic data videos with narration-animation interplay. This approach lowers the technical barriers associated with creating data videos rich in narration. To enable narration-animation interplay, Data Player constructs references between visualizations and text input. Specifically, it first extracts data into tables from the visualizations. Subsequently, it utilizes large language models to form semantic connections between text and visuals. Finally, Data Player encodes animation design knowledge as computational low-level constraints, allowing for the recommendation of suitable animation presets that align with the audio narration produced by text-to-speech technologies. We assessed Data Player's efficacy through an example gallery, a user study, and expert interviews. The evaluation results demonstrated that Data Player can generate high-quality data videos that are comparable to human-composed ones.

Index Terms—Visualization, Narration-animation interplay, Data video, Human-AI collaboration

1 Introduction

Visualization and corresponding descriptions often work together for data storytelling. Combining data visualization and narratives, data videos have become popular among practitioners as a visual storytelling form in fields such as journalism, marketing, and education [5,43]. Over more than a decade of research, data videos have demonstrated their ability to deliver condensed information, increase audience engagement, and support comprehension and memorization of data facts in data communication [6, 12, 52].

In a data video, rich information is usually packed compactly and delivered through the coordination of audio narrations and animated graphics. As indicated by the dual-coding theory, human cognition can process verbal and visual objects simultaneously, and both of them play

- *L. Shen is with The Hong Kong University of Science and Technology. E-mail: lshenaj@connect.ust.hk.*
- *Y. Zhang is with Cornell University. E-mail: yz2668@cornell.edu.*
- *H. Zhang and Y. Wang are with Microsoft Research Asia (MSRA). E-mail: {haizhang, wangyun}@microsoft.com. Y. Wang is the corresponding author.*
- *Work done during L. Shen and Y. Zhang's internship at MSRA.*

Manuscript received 31 March 2023; revised 1 July 2023; accepted 8 August 2023. Date of publication 3 November 2023; date of current version 21 December 2023. Digital Object Identifier no. 10.1109/TVCG.2023.3327197

an essential role [15]. However, creating such a data video requires a variety of skills in multiple areas, including understanding narratives, visual animation design, narration scripting, and time alignment of audio and animations, which are usually difficult to perform for novices without instruction.

To help users overcome these barriers, various technologies have been developed for different aspects of data video creation. For example, the visualization community has developed data visualization animation-specific technologies to facilitate the animation creation process, such as declarative specification grammars [19, 24, 68], authoring systems [7, 18, 57], and automated algorithms [25, 51, 59]. However, they neglect the importance of narration-animation interplay. In a recent study, Cheng et al. [12] investigated the role of narrations and animations in data videos. They found that users usually have static visual designs and text descriptions at hand for storytelling, and narration-animation interplay can effectively enhance liveliness compared to static forms. There are also a set of works that link static text and visualization together, using text-visual interplay to enhance readability in the form of interactive documents [29, 36], visualization annotations [27], etc. However, they do not fully exploit the potential of data animation to model how the data changes over time or space. In conclusion, existing authoring tools lack features for integrating narration with data animations in data videos for engaging storytelling.

To address this gap, this paper targets to design an intuitive and powerful approach that enables the automatic creation of informative data videos with narration-animation interplay from static visualizations and accompanying descriptive text. To achieve this, we first conducted a formative study to understand users' process of crafting data videos and explore the key design considerations of data video creation in their prior experience. From the study, we derived a set of high-level design constraints. The interviewees also expressed the need for support in understanding narratives, linking text and visuals, generating animations, and aligning the timeline of audio and animations.

In response to the feedback, we take the first step towards automating the generation of data videos with narration-animation interplay and design Data Player, which automates the four stages above, lowering the technical barriers of creating data videos, especially for novices. To enable narration-animation interplay, Data Player constructs references between visualizations and text input. We first extract data from input visualizations so that we convert the text-visual linking challenge into a matching problem between data table rows and narration segments. Subsequently, Data Player leverages the powerful natural language understanding ability of Large Language Models (LLMs) to associate narrative words with related data table rows, thereby establishing links between textual and visual elements. Data Player further produces a sequence of animation by modeling animation design as a Constraint-Satisfaction Problem (CSP). In detail, text-to-speech technologies are adopted to automatically generate audio narrations with timestamps of each word. Data Player then encodes design knowledge learned from the formative study and existing literature into computational low-level constraints, which are further fed into the constraint solver to generate suitable animation sequence with a pre-defined animation library. Finally, the audio and animations are rendered into a data video with narration-animation interplay.

To evaluate the liveliness of data videos produced by Data Player, we curated an example gallery and conducted a user study and an expert interview. The results showed that the automatic-generated data videos are comparable to the human-composed ones, suggesting that Data Player can effectively produce data videos with narration-animation interplay, conveying the intended information while engaging the audience. The main contributions of this paper are as follows:

- A formative study to understand users' processes and key design considerations of data video creation, leading to a set of high-level design constraints for the automatic coordination of narration and animation in data videos.
- Data Player, an innovative approach that takes the first step towards the automatic generation of vivid data videos with narration-animation interplay from a static visualization and its description. Data Player leverages LLMs to understand narratives and establish text-visual links. It further uses constraint programming to recommend suitable animation sequences.
- An example gallery, a user study, and an expert interview to evaluate the effectiveness of Data Player. The results demonstrated that Data Player can automatically produce human-comparable data videos with narration-animation interplay.

2 Related Work

Data Player draws upon prior efforts in data video creation, visualization-text interplay, and constraint-based generation approach.

2.1 Data Videos Creation

Data video is one of creative data presentation genres [43], which uses animations and audio in addition to static data to provide additional channels of communication for information transformation [6].

Prior studies have contributed insights into the comprehension and creation of animated data visualizations. For example, Tversky et al. [58] first suggested two design principles, i.e., Congruence and Apprehension, which was followed by Heer and Robertson [20], who proposed ten specific guidelines that focus on animated transitions in statistical data visualizations based on the two initial overarching principles. Amini et al. [5] conducted a systematic analysis of 50 data videos, identifying the most commonly used visualization types and attention cues, as well as high-level narrative structures. Their findings confirmed that animation in data videos has a positive effect on viewer engagement [6]. Thompson et al. [56] analyzed design primitives of animated data-driven graphics from four perspectives: object, graphics, data, and timing. Further examining the animated visual narratives in data videos, Shi et al. [52] developed a design space for motion design.

Furthermore, numerous authoring and programming tools have been created and are being continually developed to facilitate the production of animations and data videos. These tools are intended to enable creators to bring their ideas to life in a more efficient and effective manner [10]. General video creation tools (e.g., Adobe After Effects and Premier) provide fine-grained control of videos, but they require a high level of expertise and manual effort and are not tailor-made for data videos. In the visualization community, animation-specific grammars (e.g., Canis [19], Gemini [24], and Animated Vega-Lite [68]) have been developed to provide high-level specification languages for implementing keyframe-based animated transitions in data graphics. However, it requires programming skills and can be laborious for the authors. To ease the process, interactive user interfaces have emerged to enable novices to create their own data videos. DataClips [7] allows novices to create data videos by selecting and concatenating pre-defined animations from a comprehensive library, which includes clips that are suitable for visualizing different types of data. Based on the library, Kineticharts [28] enhances users' emotional engagement by improving the storytelling aspect of data presentation without compromising users' comprehension of the data. CAST [18] and Data Animator [57] support recommendations for auto-completion so that users only need to provide keyframes. Researchers have also developed automatic approaches to further reduce time-consuming manual operations. Gemini2 [25] improves Gemini [24] by providing keyframe suggestions to help users create visually appealing animations. InfoMotion [59] enables the recommendation of animations of infographics based on their graphical properties and information structures. AutoClips [51] allows users to easily input a sequence of data facts, which are then automatically transformed into a polished data video.

However, these works have neglected an important channel, narration, when applying these techniques to data videos. Cheng *et al.* [12] recently investigated the role and interplay of narrations and animations and identified close links between the two perspectives. Following the study, our work serves as the first step towards the automatic generation of data videos with narration-animation interplay.

2.2 Visualization-Text Interplay

The interplay between visualization and text plays an important role in data storytelling [46]. Recent studies have shown that the separation of text and charts may cause a split-attention effect and introduce

cognitive burden for users [29]. By contrast, linking visualization and text can promote the communication of data facts [53], support the interpretation of machine learning models [21], and enhance readers' comprehension and engagement [29, 67].

Given these benefits, researchers have actively integrated visualizations and text for interactive purposes in data presentations. For example, Vis-Annotator [27] automatically presents annotated charts according to the text description. Kong et al. [26] proposed a crowdsourcing method to collect high-quality annotations for the references of charts and text. Subsequently, automatic techniques are proposed to link text and charts with rule-based algorithms [38]. Latif et al. [29] further proposed a framework to construct references between text and charts in data documents by explicitly declaring the links. The study from Kim et al. [23] found that text-table linking in documents can support readers to pursue content better with highlighted cells. And the interactive data articles enhanced with widgets such as "stepper" and "scroller" also enable the control for users to be more autonomous during their reading [37, 67]. In addition, CrossData [11] leverages text-data links to interactively author data documents. To further ease the process of creating text-chart connections and support chart highlighting, the following studies have developed programming language [16], authoring tools [54], and interactive approaches [9, 36].

Different from static charts and documents, the dynamic changes with time progressing in videos grant it its own narrative structures [43]. Hence, the visualization-text linking in static data stories needs to be extended to narration-animation interplay in data videos [12]. To further unleash the power of integrating oral narration and visual animation in data videos, our work steps towards the automatic transformation of static text and visualizations into engaging data videos with narration-animation interplay.

2.3 Constraint-Based Generation Approach

Constraint-based approaches have been widely applied to generate visualizations [39, 49], interface alternatives [55, 65], and short videos [13, 14]. For example, URL2Video [14] captures quality materials and design styles extracted from a web page and converts them into a short video given temporal and visual constraints. While not taking narratives into consideration and are not data-oriented, it inspires us to extract design elements from static visualizations and arrange them in the animation timeline based on pre-defined constraints. However, the scenario of data video introduces new challenges for the design of narration-animation interplay [12]. On the other hand, Moritz et al. [39] demonstrated that theoretical design knowledge can be expressed in a concrete, extensible, and testable form by modeling them as a collection of constraints. Therefore, we adopt a constraint-based method to model our derived design knowledge about narration-animation interplay and incorporate them into an automatic creation workflow. The resulting approach can recommend data video designs satisfying different aspects of guidelines and further facilitate designers' crafting.

3 Formative Study

The goals of the formative study are to (1) understand practitioners' process of video crafting, and (2) explore the key design considerations of narration-animation interplay in their previous design experiences.

3.1 Participants

To achieve the above goals, we recruited 10 participants from both academia and industry with diverse backgrounds, including professional designers of video, motion graphics, animation, film post-producer, and visualization researchers. They have all acquired professional training or degrees, including three Ph.D.s, five M.S.s, and two B.S.s. All of them have experience in data video crafting through professional tools (e,g., Adobe After Effects and Premiere) or other simplified video creation tools (e.g., Microsoft PowerPoint and iMovie), with a self-reporting level of familiarity with this area (M = 4.12, SD = 0.83, range = 3-5 with 1 = "No Experience" and 5 = "Expert"). They were aged 24–32 years (5 females and 5 males, M = 27, SD = 3.10). We recruited them through online advertising and word-of-mouth.

3.2 Study Setup

The study procedure consists of two sessions with retrospective analysis and semi-structured interviews, respectively. First, we conducted a retrospective analysis, which has been proven to be an effective method for reconstructing participants' behaviors, rationales, and emotions for previous events [42]. Participants were asked to provide and show 2-3 examples from their prior data storytelling works. To promote reflection, they were required to demonstrate the creation process and explain the rationale for their design decisions. Finally, 25 videos were presented, covering 8 common chart types (e.g., maps, bars, lines, etc.).

After the retrospective analysis, we held one-on-one semi-structured interviews with the participants. The questions focused on concrete examples of narration-animation interplay in the works shown, allowing participants to recall more details of their designs and provide more useful information. We also explored the participants' views on design principles of narration-animation interplay by asking about the role of narratives and motions in their projects and the relationship between them. Finally, participants shared their difficulties in crafting data videos, particularly in aligning narrations and animations, providing insights for automatic workflows. The entire process lasted about 90 minutes and was recorded for subsequent analysis. The participants were compensated $15 for their time.

3.3 Data Video Creation Process

Participants normally prepare materials including visualization vector graphics and narrative scripts before crafting data videos in professional software such as Adobe After Effects, Cinema 4D, and Blender. During this preparation phase, they focus on the aesthetic design of their graphics and descriptive narrative writing, with little attention to the dynamic interplay between them and the motion effects of output videos. In some cases, graphics and texts may also be given to the creator by other collaborators such as graphic participants or screenwriters. After that, our findings identified four distinct design stages in terms of video crafting, which is of interest to our research.

Stage 1: Refining Narration Text. At this stage, participants try to collect the text descriptions for the charts and compose the narration text of the video to produce audio narration. If the text description is not created by them, they need to understand the intents that the text authors or storytellers would like to convey to the audience. In this stage, data video creators usually decompose the messages in the text narration and formulate the messages to convey to prepare for their further design of animation in the videos.

Stage 2: Building Visual References. Based on the formulated messages, participants match them to the visual references in the visualizations. Visual references are graphic elements in charts, such as the specific line in a line chart or the related rectangle in a bar chart. Participants build visual references associated with the messages in the narratives. Sometimes they may group them if there are some relationships between the visual elements. For example, one interviewee told us that he usually grouped two comparative data elements so that he designs animations for them in the subsequent motion design phase.

Stage 3: Animation Design with Semantic Metaphors. After preparing the text narration and visualizations, participants design animations with semantic metaphors in line with the intent of storytelling. For example, if the storyteller wants to express a rising trend for a line chart, participants would make a dynamic growth curve from the lowest point to the highest point. When emphasizing the visual elements, participants may modify transparency, saturation, contours, or other visual properties as animation cues to highlight information. Additionally, three types of animation effects are most commonly adopted: entering, exiting, and emphasizing.

Stage 4: Coordination of Audio Narration and Animations. Finally, participants align the created series of animations with the audio created by the voiceover artist on the timeline. They often take the approach of manually adjusting the keyframes, setting the start frame and the end frame of the animation at the corresponding time points in the narrative audio. Most interviewees (8/10) reported that this process was very time-consuming and laborious because they needed to listen to the audio and watch the animation repeatedly to check for out-sync.

3.4 Design Constraints

Through retrospective analysis and interviews, we found that participants were concerned about the echoes of narratives and animations in four dimensions: visual structure, data facts, semantics, and temporality. Further, we derived a set of design constraints from interviewees' considerations collected from the study, as well as existing literature [6, 12, 20, 52, 58].

First, the participants emphasized the importance of visual constraints in creating an organized and logical presentation. They suggested grouping relevant visual elements associated with the same data fact, establishing a sense of hierarchy and facilitating audience comprehension. Additionally, they recommended introducing unrelated background elements, such as titles and axes, at the beginning of the video to provide context. Maintaining consistent animation effects within groups of similar elements further enhances clarity.

The participants mentioned the importance of data interplay in enhancing audience understanding. They advocated for echoing narrations and animations by reiterating conveyed information, as well as selecting and animating visual elements relevant to the data facts in the narrative script. This approach helps emphasize key points and maintain the audience's attention. Moreover, incorporating animations into data facts, rather than other narratives, can improve comprehension.

The participants also highlighted the role of semantic rules in conveying the intended message accurately, which refers to the implicit interactions that arise from the meanings and intentions of visual elements and their narrations in voiceover. They underscored the need to align the semantic intents of narrations and animations to avoid confusion. Furthermore, they suggested supplementing missing narrative information with annotations, such as labels, explanations, and context, to provide a more comprehensive representation of data.

Finally, temporal interplay emerged as another critical aspect. The participants stressed the importance of coordinating animation sequences with narrative structures to preserve the meaning of the information, synchronizing narrations and animations for consistency, and adapting the timing of animations to match narrations, ensuring that the information remains manageable for the audience.

We regard these design constraints as high-level because they cannot be directly translated into actionable computational programs, but lay the foundation for the subsequent formalization of low-level design constraints, which will be discussed in detail in Section 4.4.2.

4 Data Player

In this section, we first introduce the conceptual model of data video with a set of design variables. Then we give an overview of Data Player, and further discuss the two important modules: text-visual linking and animation sequence generation.

4.1 Data Video Modeling

Data videos consist of three design elements: visualizations, narrations, and animations.

$$video := (visualization, narration, animation) \tag{1}$$

4.1.1 Visual Elements

Overall, visualizations are composed of a set of visual elements or element groups in a given static graphic.

$$visualization := visual\ element | groups \tag{2}$$

$$group := visual\ elements \tag{3}$$

The visual elements can be marks, axis, legends, annotations, etc. Each is defined as a tuple:

$$visual\ element := (id, type, data) \tag{4}$$

where *id* is the unique identifier, *type* indicates the graphic shapes and visualization structure of the element, and *data* is embedded in each visual element like dSVG in Canis [19]. For each element, *data* can be null and can also correspond to multiple data items.

4.1.2 Narration Entities

Static narration text will be converted into audio speech, and each entity will be an audio unit with time; thus, a conceptual model of the narration text is:

$$narration := narration\ entities \tag{5}$$

$$narration\ entity := (audio, time) \tag{6}$$

$$time := (start, duration) \tag{7}$$

where *start* refers to the start timestamp of an audio unit and *duration* refers to the time span that an audio unit lasts.

4.1.3 Animation Elements

The animation sequence applied in data videos is a series of animation *units*. The animation is defined as:

$$animation := animation\ units \tag{8}$$

$$animation\ unit := (visual\ elements, time, action, effect) \tag{9}$$

where each animation unit targets a visual element or a group of elements and declares a start time and a duration. Additionally, the *action* specifies which kind of animation *effect* can be applied.

4.2 Overview

Prior research has pinpointed specific features of narration-animation interplay in high-quality data videos [12]. Additionally, the formative study identified four common stages in the process of creating data videos, each of which requires considerable time and manual effort from users. We aim to automate the process of creating data videos with narration-animation interplay, making them more accessible and user-friendly for novice users. To this end, we design Data Player that consists of two modules: (1) *Understand input narration text and visualization, and semantically link them.* Narration text frequently captures the central messages of data stories, incorporating data facts and insights associated with the visualization. The module automatically parses the text to extract the data facts presented in the narration. Further, it establishes connections between narration segments and visual elements within the visualization, which can be further leveraged to create audio and animations in the data video. (2) *Recommend animation sequence and synchronize audio narration and visual animations.* Using visual cues corresponding to the words spoken in the narration is crucial in data video creation, and animations can make data more engaging and memorable for the viewer. To avoid any confusion or misleading the audience, the module automatically generates an appropriate animation sequence that serves the same purpose and intent as the storytelling, and synchronizes the audio narration with the visual animation, ensuring that the viewer receives the information clearly and coherently.

We propose an automatic pipeline to guide the design of Data Player, as shown in Figure 2. First, a static visualization and corresponding narration text are inputted in the form of Scalable Vector Graphics (SVGs) and plain text (a), respectively, so that they can be decoupled into multiple visual elements and narration segments (b). Then, Text-To-Speech (TTS) techniques are used to generate audio voiceovers and return timestamps of each word, which also act as the timeline of the data video (c). Furthermore, the Large Language Model (LLM) is adopted to establish the semantic links between visual components (one or a group of graphic elements) and narrative entities (one or more words) based on the data facts to be told (d). To be specific, the linking module identifies the visual elements of the visualization inputs that can be animated, extracts data facts from them into tables, and links the table rows with semantic entities in the narration. After that, the animation generation module encodes collected design knowledge about narration-animation interplay into computational constraint programs and leverages the constraint solver to generate a suitable animation sequence with pre-designed animation presets based on the established text-visual links (e). Moreover, the module seeks to automatically organize animation sets in alignment with the generated audio timeline. It

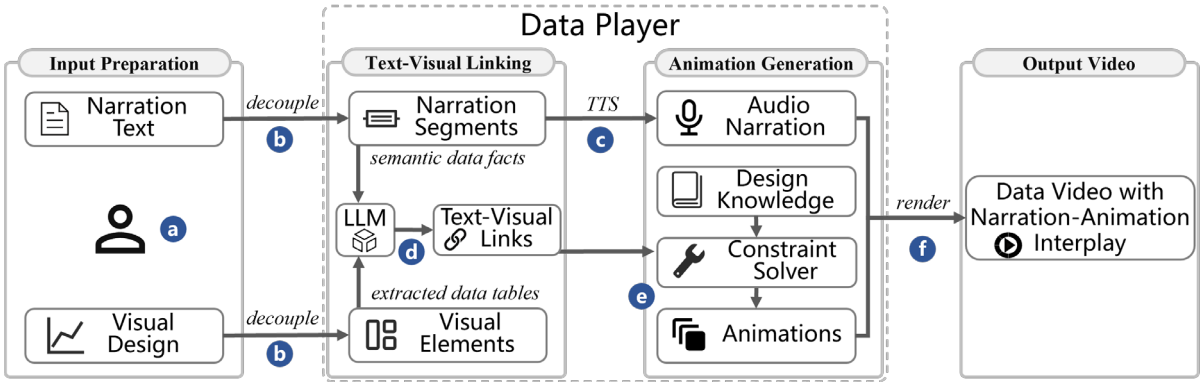


Figure 2: The pipeline of automatic generation of data videos with narration-animation interplay.

makes temporal decisions to allocate using constraint-based programming. As a result, a sequence of audio-animation packs is specified, which are further rendered into the data video (f).

4.3 Text-Visual Linking

To generate data videos with narration-animation interplay, it is crucial to understand the narration text and its relations with the visual elements. We propose an LLM-based approach, shown in Figure 3, to generate these semantic links for animation. By extracting visual candidates that can be semantically linked in the visualization, LLM is then used to match these candidates with relevant narration segments. We illustrate this below with an example of a 15-day PM2.5 value in Beijing.

4.3.1 Data Extraction

As described in Section 4.1.1, each visual element has a *data* property that contains semantic information. To effectively organize and utilize this information, we transfer the semantics into data tables and group elements based on the data items they contain.

Visual candidates in the visualization can be divided into two categories: basic graphical representations (e.g., marks, axes, legends, etc.) and annotations [40] (Figure 3-a). First, as demonstrated in Figure 3-b, our method consolidates the data information in the SVG into a basic data table, which includes all values represented by graphical marks and axes. We also maintain a map that correlates each data table row to the corresponding visual elements. For example, the first row of data (Day: 1st, PM2.5 Value: 54.8) corresponds to both a bar mark (id is bar-0) and an x-label (id is x-label-0). While the data table captures most of the information present in the static chart, there may be missing information, particularly in regard to annotations, which play an important role in information communication. Referring to the annotation design space proposed by Ren et al. [40], we divide annotations into text, graphics (including shapes and images), and their combinations. These elements contain valuable semantic information, such as the text ("hazardous") and the red rule annotation in Figure 3-a, both of which express information about the hazardous threshold (300). Therefore, we extract a separate data table and group the corresponding elements. Overall, each input data visualization will derive one data table for the chart marks and optionally one or more data tables for the annotations.

4.3.2 Linking by an LLM

We have extracted semantics from static visualization graphics, as well as the mapping relationship between semantics and visual elements. To link narration segments (one or more words) and visual components (one or a group of graphical elements), the next step is to detect the occurrences of similar entities of these semantics in narration text.

We further model this linking problem as a matching problem, semantically matching narration segments (Figure 3-c) with data table rows (Figure 3-b), and then mapping the data table rows to visual elements. Specifically, we leverage the powerful natural language understanding ability of LLM (We use the OpenAI gpt-3.5-turbo model in our work.) to link the two perspectives, as shown in Figure 3-d. The prompt engineering aims to ask the LLM to accept data tables and narration words as input, and output semantic links as "(*narration segments*)[table x: R_i,...]", where x is the table index and R_i is the data table row index (see Figure 3-b). Inspired by existing successful prompt engineering experiences [1, 34], the prompt includes few-shot pre-defined examples for a better illustration of our task. The LLM output is shown in Figure 3-e. Finally, having links between table rows and narration segments in hand, we further obtain text-visual links with the help of mapping relationship between table rows and visual elements. In addition, we also fine-tune and deduplicate the text-visual links to avoid unnecessary animations in the subsequent steps.

Our approach has several advantages over existing works about establishing references between text and charts [8, 27, 29]. First, we formulate the problem of text-visual linking into a problem of matching data table rows and narration words, which allows us to capture the semantic relationships between text and visuals more effectively. For instance, interpreting the phrase "the following day" requires an understanding of the temporal context of the data table, which is difficult to achieve by traditional rule-based linking methods. Second, we leverage LLMs, state-of-the-art language models, to perform similarity matching between data table rows and narration words with high accuracy due to larger knowledge support and better natural language understanding ability compared to prior NLP packages. However, traditional methods outperform the LLM-based method in terms of real-time performance.

4.4 Animation Sequence Generation

In this subsection, we introduce the animation recommendation module, which encodes collected high-level design knowledge into low-level constraints to automatically generate a suitable animation sequence.

4.4.1 Animation Modeling

According to the formative study and existing literature [12], designers concern with three types of semantics to implement appropriate animation effects. Specifically, they distinguish the semantic beginning and end of the narrative description of one data fact, as well as the emphasis intent in it and the information complement to the chart. Therefore, based on the definition in Section 4.1.3, we specify three animation actions: "enter" animations are applied when an object appears on the canvas and "exit" animations are applied when an object disappears from the canvas, while "emphasis" animations are applied to draw attention to an object that is already on the canvas. Each action includes several commonly seen changes in visual channels (e.g., "fade", "grow", "zoom", etc.). The timing and duration of the animations can also be adjusted to suit specific needs.

4.4.2 Constraint Encoding

Prior work has demonstrated methods to generate designs from a set of design constraints [14, 39, 55], which motivates us to formulate narration-animation interplay design into a Constraint-Satisfaction Problem (CSP). In detail, We model the design elements discussed

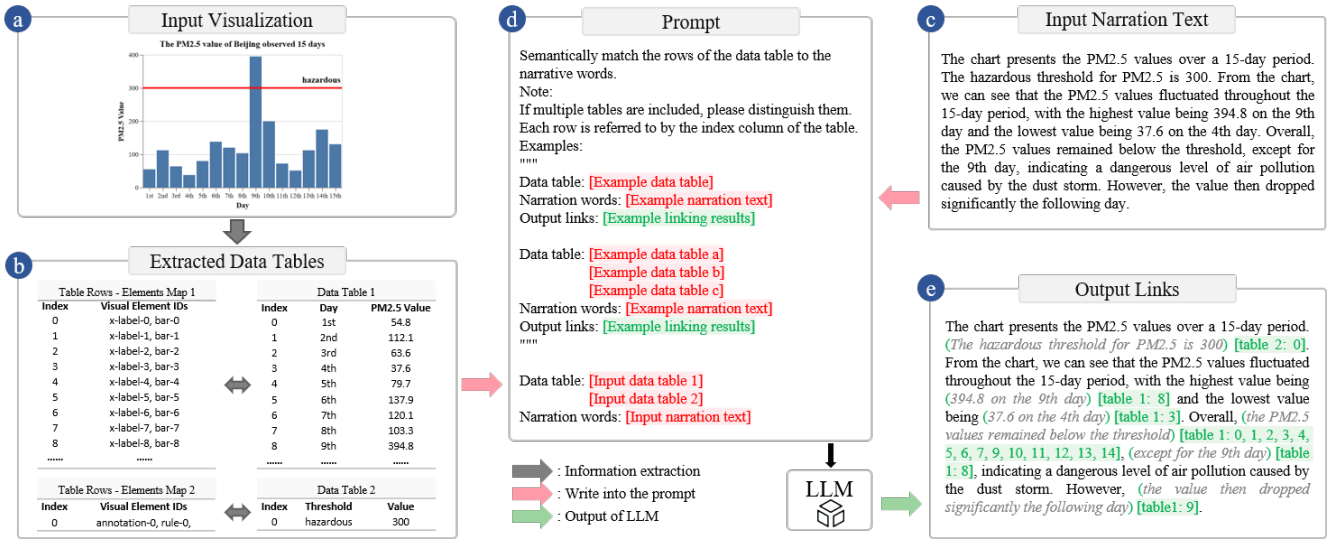


Figure 3: A walkthrough example illustrating the text-visual linking workflow. Data information behind the visualization (a) is extracted into data tables (b). The LLM with appropriate prompts (d) accepts the narration text (c) and data tables (b) as input and outputs semantic links between data table rows and narration segments (e).

in Section 4.1 (i.e., visualization, narration, and animation) with encoded variables. Each variable has a domain. For example, animation actions include "enter", "exit", and "emphasize" (Section 4.4.1), and each action corresponds to a set of animation effects. The variable domains of visual elements and narration entities are derived from the generated text-visual links (Section 4.3.2). To generate the animation sequence, Data Player assigns concrete values to specific variables and leverages the CSP solver to explore numerous combination alternatives in the large search space. Specifically, we encode high-level design knowledge summarized from the formative study and existing literature [6, 12, 20, 52, 58] as computational low-level constraints. All constraints are formalized as equations and fed into the Z3 [17] CSP solver. The solver outputs suitable animations. For instance, the animation sequence specified for the example in Figure 3 is shown in Figure 4 (top). Ultimately, the audio narration and visual animations are rendered into a dynamic data video (.mp4 format) with narration-animation interplay with the FFmpeg multimedia framework [2]. The low-level constraint encoding is detailed as follows:

First, to ensure basic visual design quality, we use the established text-visual links to generate visual structure and data facts constraints, matching textual and visual entities and grouping visual elements. Specifically, we design `linking` constraints that allow only the visual elements that are linked to specific narration segments to be animated. The `integrity` constraints ensure that all elements involved in text-visual links need to be animated. We design `group` constraints to group the visual elements that are related to data facts in the text-visual links. Meanwhile, we design the `association` constraints to ensure that if one element is linked to narration, itself and other elements in the same data group can be animated. In addition, our `consistency` constraints specify that elements from different groups that are visually consistent should apply the same animation.

Second, we encode sets of temporal constraints to time-align animations and narration. Each group of constraints specifies how different elements of the data video should be timed and arranged on the timeline according to their type, effect, and relation to the narration. First, narration text inherently contains a chronological relationship between words. We further use Microsoft Azure Text-to-Speech services [4] to automatically generate audio narration and obtain the timestamps of each word in the audio, which also acts as the timeline to arrange animation effects applied to the visual elements. On this basis, we encode a `duration` constraint to determine that the animation effects are triggered by the onset of the first word in each linked narration segment, and last for the duration of the corresponding text span in the audio. The last frame of the previous animation will be retained for the time period when no animation is applied. The `conflict` constraints enforce the inherent logical order of animation actions. For example, visual elements can only be emphasized or disappeared after they appear, and elements cannot be emphasized after they disappear. And `on screen` constraints determine when an element appears or disappears from the canvas based on the "*on_screen*" variable assigned to each element. If "*on_screen*" is true at time t, then the corresponding element is visible at that time. Otherwise, it is hidden. By assigning different values of "*on_screen*" to each element at different timestamps, we can create a table that shows which elements are on the canvas at any given moment. The table can help us control the animation actions to avoid overlapping or conflicting movements. For instance, visual elements that have an enter animation applied will not appear on the canvas until the animation is triggered. Elements that have an exit animation applied will disappear from the canvas after the animation. Elements that do not have any animation applied will appear on the canvas by default. In addition, a set of `order` constraints defines an optional logical sequence of elements such as background, title, axis and data items and the `synchronization` constraints ensure that elements in the same data group activate together.

Third, different animations can produce different effects and serve different purposes. To align the semantic intents between the linked narration segments and animated visual elements, we encode a variety of constraints to assign appropriate animations from the pre-defined library to visual elements based on the data facts being presented, the visual structure, and the desired audience engagement. We also specify constraints on animated annotations to avoid messing up the canvas. Furthermore, we define a series of implicit mappings with priorities between the involved visualization structures and the appropriate animation combinations based on long-term practical experience. These mappings are effective for defining the animations of the elements within a group. For instance, in a pie chart, the sector and its corresponding legend elements (e.g., symbol and label) are usually bound into a group. So we define a new animation called "pie-wheel-and-legend-fly-in", which means that the pie chart's sector will wheel clockwise and the legend-related elements will fly in at the same time, as shown in Figure 4 (middle). As a result, we can apply only one animation to multiple elements, avoiding specifying animations for each element individually. On this basis, we define an objective function to minimize the number of animations used: $\min \sum_{i=1}^{m} A_i$, where A_i is the number of animations applied to the i-th text-visual link and m is the number of text-visual links. This function ensures that the module uses

Figure 4: Automatic-generated data videos by Data Player. Each example includes a sequence of animation. The animation will be triggered when the audio narration reaches the corresponding segment. The snapshot images show the effect after the animation has been triggered.

our predefined animation combinations as much as possible to maintain narrative coherence.

4.4.3 Animation Presets

A comprehensive library of animation effects for each action and mappings between semantics and effects can enable a wide range of designs. However, constructing such a large-scale library requires significant development costs. Thus, we utilize a small set of pre-designed animation effects based on the GSAP animation platform [3] for different actions as a technology probe and proof-of-concept to explore our main research concern [61]. For instance, "fade-in" and "wipe" for entrance, "zoom-out" and "fade-out" for exiting, and "shine" and "change-color" for emphasis, etc. Additionally, depending on different chart types and element orientations, the configurations of one animation effect are adjusted, such as the "grow" effect for bar marks, "wipe" for lines, and "wheel" for circular marks. In the future, we will explore more vivid animations to enrich the library.

5 Evaluation

To evaluate the liveliness of data videos generated by Data Player, we (1) built an example gallery from real-world data storytelling practices, (2) conducted a user study to compare automatic-generated videos with those created by novices and designers, and (3) performed expert interviews to further understand the difference between automatic-generated data videos and human-composed ones.

5.1 Example Gallery

To demonstrate the expressiveness of the automatic approach, we generate a variety of example data videos based on a set of public design files. These examples cover a wide range of visualisation types (e.g., bar, pie, line, etc.) and narrative themes (e.g., PM2.5, tourism, stock price, tax payment, etc.). Figure 1 and Figure 4 show a subset of cases, more data video examples can be found in https://datavideos.github.io/Data_Player/.

5.2 User Study

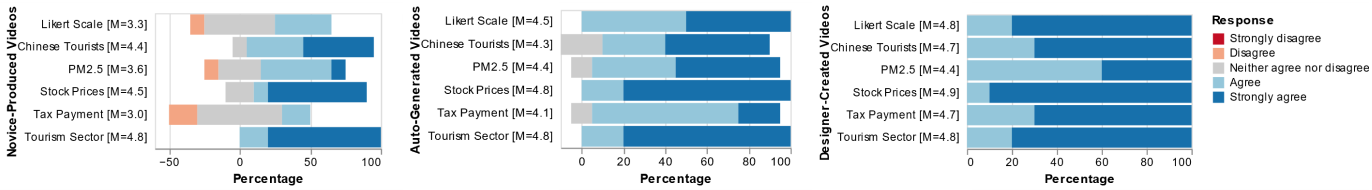
In this study, we aim to understand the quality of data videos produced by Data Player, by comparing them with data videos produced by novices and designers.

5.2.1 Dataset

To prepare data videos for the user study, we collected six sets of static charts and their descriptions from real-world data storytelling practices, including a line chart with stroked point markers that shows spending on outbound tourism of Chinese tourists (short as "Chinese Tourists", Figure 4 (bottom)), a pie chart that describes the future outlook of the tourism sector over the next year (short as "Tourism Sector", Figure 4 (middle)), an annotated bar chart that depicts the PM 2.5 value of Beijing observed in 15 days (short as "PM 2.5", Figure 4 (top)), a stacked bar chart that describes America's tax system (short as "Tax Payment", Figure 1), a diverging stacked bar chart for sentiments towards a set of eight questions with a 5-point Likert scale (short as "Likert Scale"), and a multi-series line chart that shows the stock prices for five high-tech companies (short as "Stock Prices").

5.2.2 Procedure

Pre-experimental preparation. We first invited four participants to manually create videos, including two novices and two professional designers. The novices were unfamiliar with video creation and only had experience using visualization to present data insights and using MS PowerPoint to create animations. The two designers' daily work involved using professional video creation tools like Adobe After Effects. The two designers also participated in our prior formative study (Section 3). Based on the static visualizations and descriptions collected from the internet (Section 5.2.1), each participant was asked to create three videos. In order to control the conditions of comparison, we implemented an interactive authoring tool [61] that allows participants to manually specify the links between narration segments and visual elements, apply appropriate animations, generate audio from text,



(a) Novice-Produced (b) Automatic-Generated (c) Designer-Created

Figure 5: User study results with a 5-point Likert scale. From left to right are novice-produced, automatic-generated, and designer-created videos.

fine-tune the timeline, and preview the data video. They can ask us any questions they have during the manual creation process to ensure they can master the tools without compromising the quality of their creative data videos. Finally, we confirmed with them that the data videos they produced were representative of their level of design. For each piece of material, we also obtained an automatically generated version using Data Player. In total, we collected six novice-composed data videos, six designer-created data videos, and six automatic-generated videos.

During experiment. We recruited another 10 participants (denoted as P1 to P10) who have data analysis and visualization needs in their daily work, including data analysts, ML researchers, software engineers, and visualization researchers. Each participant viewed six sets of videos, each set including three versions: novice-created, automatic-generated, and designer-composed. The sequence of the three videos and the six sets was shuffled. We also provided the textual descriptions that were used to create the data videos. The participants can watch the video repeatedly as they like and were asked to rate the overall quality of each video in terms of expressiveness and liveliness using a 5-point Likert scale. They were encouraged to leave reasons for the decisions and speak of any comments on any aspect of the videos. After all data videos have been viewed and evaluated, we asked participants to identify the automatic-generated one in each set. The experiment lasted about 60 minutes.

5.2.3 Results

Subjective Satisfaction: The results of the participants' ratings are shown in Figure 5. From left to right are novice-produced, automatic-generated, and designer-created data videos. Overall, participants had a positive sentiment towards all the data videos. The majority of participants rated the videos as "agree" or "strongly agree", with mean values all greater than or equal to 3.0. In detail, the results showed that the videos automatically produced by Data Player were generally well-received by the participants, with a mean score of 4.4 (*SD*=0.70) for "PM2.5", 4.3 (*SD*=0.82) for "Chinese Tourists", 4.1 (*SD*=0.57) for "Tax Payment", 4.8 (*SD*=0.42) for "Stock Prices", 4.5 (*SD*=0.53) for "Likert Scale", and 4.8 (*SD*=0.42) for "Tourism Sector". These scores were higher than the mean ratings given to the novice-produced data videos and comparable to the mean ratings given to the designer-created videos. In terms of the specific topics of the data videos, the automated data videos received the highest mean ratings for "PM2.5" and "Tourism Sector", with means of 4.4 and 4.8, respectively. The designer-composed videos had slightly higher mean ratings than automatic-generated ones for other topics. However, the paired t-test results showed that there is no significant difference in ratings between automatic-generated and designer-created data videos for all topics.

We also found that for visualizations with relatively simple structures and patterns, there is little variation in the user ratings of different types of videos. For example, all three types of videos received the same ratings for "Tourism Sector" (both 4.8) and the paired t-test results show that the difference is not significant between ratings for "Chinese Tourists" and "Stock Prices". However, when the chart structure is more complex (e.g., "Tax Payment" and "Likert Scale"), especially when the narrative expresses some in-depth content, novices often encounter difficulties to tell the story more reasonably, and the designers' experience can help them deal with these situations better. The automatic algorithm was able to effectively communicate the information in a clear and engaging way, at a level between novices and designers.

Participants can identify the automatic-generated videos from the three versions with an average accuracy of only 31.7%. This means that the automatic-generated videos were close to the human-composed ones. Moreover, we found that 45% of the misjudged videos were actually composed by novices, and 23.3% were actually made by designers. This suggests that the level of skill and creativity of the human composers also affected the perception of the evaluators.

Feedback: All participants agreed that narration-animation interplay can enhance the efficiency and vividness of data insights communication compared to static forms. They also praised the data videos' expressiveness, liveliness, and overall quality, and noted that they would consider using the automatic method in their own work. P2 said, "*the videos dynamically present information while ensuring completeness. I was so surprised that the wonderful data videos were generated automatically.*" They also expressed a preference for some aspects that were focused on in our design knowledge. For example, all participants appreciated the use of animation in the right context. P4 also commented, "*the videos were visually consistent overall, as the same animations were used for similar visual structures.*"

We also learned some lessons about users' preferences. Some participants criticized the visual effects and style of the videos and hoped that the algorithm could better meet individualized needs. For example, P1 did not like the "Bounce" animation effect of bar marks, and P3 did not expect the "Change Color" effect to always be red. For legend, P3 preferred to have the graphical elements and their corresponding legends presented together in sequence so that the audience can get an immediate understanding of the visual elements. While P1 preferred to see the legend first to get a general impression of the context, and then the marks and the narration appear in sequence. This indicated that a user interface is needed to extend the existing automation pipeline and incorporate humans into the workflow.

5.3 Expert Interview

To further compare the difference between automatic-generated videos and human-composed ones, we invited six experts to provide feedback through interviews, including four designers (denoted as D1 - D4), of whom D1 and D2 helped us create designer-composed videos in the user study (Section 5.2.2), and two visualization researchers (denoted as V1 and V2), who have more than five years of experience conducting visualization research and publishing visualization papers in major conferences (e.g., IEEE VIS and ACM CHI). Moreover, all of them joined our prior formative study (Section 3) and provided valuable feedback. They were asked to watch the six sets of data videos used in the user study, and they were informed of the respective versions in advance. Then they need to provide specific feedback on comparing different versions. In addition, they were also asked to comment on the method's strengths and weaknesses, its possible application scenarios, and the future outlook.

Overall, all the participants agreed that all data videos were able to effectively interact between narration and animation. When comparing the three data video versions, the participants generally thought that there was no very obvious difference between them overall. However, D1 and D2 found that humans (especially designers) did a better job fine-tuning the timeline. In the creation of data videos, the animation duration is dependent on the length of the narration segment in the text-visual link. This may result in animations that are too short, such as those corresponding to only one word, which can lead to user confusion, particularly for animations intended to emphasize certain points. They (D1 and D2) also noted that they spent a certain time previewing

and iteratively refining the animations for the designer-created videos, including adjusting the trigger timestamp and duration of animations, so they can mitigate this issue. Participants (D1, D3, D4, V1, and V2) also pointed out the animation-intensive nature of Data Player. D1 stated, "*the automatic algorithm feels like it wants to add animations to every sentence,*" while V2 agreed, "*this level of animation density can be tiring for viewers.*" and V1 complemented, "*I often found it hard to keep track of multiple animated visual elements at the same time, especially when they move or change in different ways.*"

The participants also provided valuable insights into the potential application scenarios and future improvement directions. All of them agreed that Data Player is an effective tool for empowering novices to create data videos. V1 and V2 also suggested that Data Player can be used as a module integrated into other large systems to prototype more complex videos, automate slide design, etc. D3 commented that our automatic technique can be further extended to enable other creative ways of storytelling, such as scrollytelling [44] and interactive data videos [22], which requires more interactive and exploratory experience. D1 and D4 expected that professional video production software (e.g., Adobe After Effects) can integrate the animation recommendation module. Even though all participants praised the convenience of Data Player, some of them (V1, V2, D2, D4) also suggested further automatically generating static visualizations and narration text for users. More importantly, during the process, the system should allow users to fine-tune the generated results at each stage of data video generation to provide an interactive human-in-the-loop experience.

6 Discussion

Automation vs. Personalization. Creating narration-enriched data videos is a highly specialized and time-consuming task. Data Player can help users automate this process based on the input static visualization and description. It enables rapid exploration of design alternatives, thus increasing efficiency for data insight presentation. However, in our user study, we found that participants' personal preferences (e.g., animation effects and visual styles) may affect their ratings of data videos. These highly personalized needs are difficult to be thoroughly satisfied in a full-automatic algorithm [62]. Subsequently, we plan to design a user interface and develop Human-AI collaboration methods [30]. First, the automated method can prototype data videos (including automatic generation of visualizations and corresponding narration text), and then users can further modify them with fine-grained control on the interface. Next, the system can allow users to input their preferences at different stages of data video generation, and progressively generate data videos. Moreover, the system can provide various animation examples for users to select and adapt to their own designs [48], and the system can also automatically learn personalized needs from the user's interaction history through multi-modal interactive task learning [41]. Another interesting approach would be to help users maintain a personalized knowledge base. Based on existing professional design knowledge, users can continuously expand and update their personalized rules.

Application of LLMs. We applied LLMs to match narration segments and visual elements. Before this, we tried traditional NLP packages and BERT-based n-gram similarity matching schemes, which were somewhat mechanical and rigid. Recently, LLMs (e.g., chatGPT and GPT 4) have demonstrated remarkable capabilities in generating and understanding natural language. After using LLMs, the matching module achieved better accuracy and flexibility. However, existing LLMs still have some inherent limitations, such as being not good at complex computational tasks, producing inconsistent outputs in different rounds, taking a long time for generation, and hallucination problems, etc. These factors affect the accuracy, timeliness, and practicality of our methods to some extent. But we believe that future research will address these issues. In the future, we will mainly explore three directions based on LLMs: The first is how to better improve the accuracy of text-visual linking with LLMs, such as exploring more accurate prompts and integrating with other interactive tools [63]. The second is to further expand the existing automatic pipeline based on LLMs, such as helping users interactively generate and modify narrations [45] and automatically create and update visualizations [60] based on the data table. The third is to fit the use of LLMs within human-in-the-loop scenarios. For example, LLMs can extract insights from data. Users can also choose the insights of interest based on their analytic tasks [47], and further leverage LLMs to automatically generate targeted narration text, visualizations, and chart annotations. Furthermore, given some specific material (e.g., a visualization or a narration segment), LLMs can be used for similarity searching to obtain more material to aid storytelling [31, 35].

The completeness of design knowledge. Design knowledge plays an important role in the narration-animation interplay design. In this paper, we have explored several key design constraints from the formative study and existing literature. However, it is important to note that these constraints are not exhaustive and we only consider them as a minimal set to assist data video creators and researchers in crafting narration-animation interplay for data videos. There may be other factors that have an impact on the narration-animation interplay in different contexts and domains (e.g., spatial alignment, visual complexity, and cognitive load). Therefore, we encourage further research to investigate more design guidelines to assist data video creators in creating more effective and engaging narration-animation interplay. Additionally, the target function in our current animation recommendation module can be further improved. In the future, we can also consider setting hard (must be satisfied) and soft (will be penalized if not satisfied) constraints [39] to enable more flexible recommendations.

Understanding emotions in the narration. Our automatic pipeline primarily focuses on the semantic matching of narration segments and visual elements, as well as the rationality of animations. We also use text-to-speech technology to automatically generate audio narration. This allows us to generate reasonable and vivid data videos, but we do not further understand the emotions from the narration. In the future, it would be interesting to investigate the emotional design space in narration-animation interplay and express emotions in the narration with appropriate emotional animations, tones, and visual cues [28, 64]. Additionally, in order to draw attention to significant events and changes in animation, it is important to use sound effects, which are also not considered in our pipeline. Furthermore, having background elements appear at the beginning can also serve to set the tone and mood of the video. For example, if the video is presenting visualization related to a serious topic, such as a disease outbreak, having a somber and serious title card at the beginning can help to establish the tone of the video.

Limitations and Future Work. In our user study, we found that the quality of the input visualization and narration itself can influence users' judgments. In the future, we can use users' own materials to create videos and then ask them to compare the videos with their previous data presentation forms. Another limitation of our study was the relatively small number of participants. It would be interesting to conduct a larger crowdsourced user study to further verify the quality of the automated data videos and to investigate the factors that contribute to their perceived quality. Additionally, Data Player currently only supports data video generation with a single chart and structured data. In the future, it can be extended to more visualization scenarios (e.g., multiple charts or dashboard [33], glyph [66], and infographics [59]) and data types (e.g., geographic [32], graph [50], and word cloud [64]) to tell more complex stories.

7 Conclusion

To streamline the complex process of crafting narration-animation interplay for data videos, this paper proposes Data Player, which enables the automatic transformation of static text and visualizations into engaging data videos, enabling more novices to share their insights and research findings using data videos. Data Player leverages advanced LLMs to extract data facts and establish text-visual links, and uses constraint programming to recommend animations for the links and time-align audio narrations and visual animations. The results of the user study indicated that the data videos automatically produced by Data Player were well-received by participants and comparable in quality to human-composed ones. We hope that the approach can help people with little video production experience quickly create high-quality data videos, and inspire future research about narration-animation interplay in data storytelling.

REFERENCES

[1] Chatgpt prompt engineering for developers. https://learn.deeplearning.ai/chatgpt-prompt-eng/. 5

[2] Ffmpeg multimedia framework. https://ffmpeg.org/. 6

[3] Gsap animation platform. https://greensock.com/gsap/. 7

[4] Microsoft azure text-to-speech service. https://azure.microsoft.com/services/cognitive-services/text-to-speech. 6

[5] F. Amini, N. Henry Riche, B. Lee, C. Hurter, and P. Irani. Understanding Data Videos: Looking at Narrative Visualization through the Cinematography Lens. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI'15*, pp. 1459–1468. ACM, 2015. 1, 2

[6] F. Amini, N. H. Riche, B. Lee, J. Leboe-McGowan, and P. Irani. Hooked on data videos: assessing the effect of animation and pictographs on viewer engagement. In *Proceedings of the 2018 International Conference on Advanced Visual Interfaces*, AVI '18, pp. 1–9. ACM, 2018. 1, 2, 4, 6

[7] F. Amini, N. H. Riche, B. Lee, A. Monroy-Hernandez, and P. Irani. Authoring Data-Driven Videos with DataClips. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):501–510, 2017. 2

[8] 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. 5

[9] 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 2023 CHI Conference on Human Factors in Computing Systems, CHI '23*, pp. 1–15. ACM, 2023. 3

[10] Q. Chen, S. Cao, J. Wang, and N. Cao. How Does Automation Shape the Process of Narrative Visualization: A Survey of Tools. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–20, 2023. 2

[11] Z. Chen and H. Xia. CrossData: Leveraging Text-Data Connections for Authoring Data Documents. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, CHI'22*, pp. 1–15. ACM, 2022. 3

[12] 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. 1, 2, 3, 4, 5, 6

[13] P. Chi, N. Frey, K. Panovich, and I. Essa. Automatic Instructional Video Creation from a Markdown-Formatted Tutorial. In *The 34th Annual ACM Symposium on User Interface Software and Technology, UIST'21*, pp. 677–690. ACM, 2021. 3

[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, UIST'20*, pp. 279–292. ACM, 2020. 3, 5

[15] J. M. Clark and A. Paivio. Dual coding theory and education. *Educational Psychology Review*, 3(3):149–210, 1991. 2

[16] 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*, UIST '18, pp. 977–989. ACM, 2018. 3

[17] L. de Moura and N. Bjørner. Z3: An Efficient SMT Solver. In *Tools and Algorithms for the Construction and Analysis of Systems.*, pp. 337–340. Springer, 2008. 6

[18] T. Ge, B. Lee, and Y. Wang. CAST: Authoring Data-Driven Chart Animations. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, pp. 1–15. ACM, 2021. 2

[19] T. Ge, Y. Zhao, B. Lee, D. Ren, B. Chen, and Y. Wang. Canis: A High-Level Language for Data-Driven Chart Animations. *Computer Graphics Forum*, 39(3):607–617, 2020. 2, 4

[20] J. Heer and G. Robertson. Animated Transitions in Statistical Data Graphics. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1240–1247, 2007. 2, 4, 6

[21] F. Hohman, A. Srinivasan, and S. M. Drucker. TeleGam: Combining Visualization and Verbalization for Interpretable Machine Learning. In *Proceedings of the 2019 IEEE Visualization Conference, VIS'19*, pp. 151–155, 2019. 3

[22] J. Hook. 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, pp. 43–55. ACM, 2018. 9

[23] D. H. Kim, E. Hoque, J. Kim, and M. Agrawala. Facilitating Document Reading by Linking Text and Tables. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*, UIST '18, pp. 423–434. ACM, 2018. 3

[24] 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, 2021. 2

[25] Y. Kim and J. Heer. Gemini2: Generating Keyframe-Oriented Animated Transitions Between Statistical Graphics. In *Proceedings of the 2021 IEEE Visualization Conference, VIS'21*, pp. 201–205, 2021. 2

[26] N. Kong, M. A. Hearst, and M. Agrawala. Extracting references between text and charts via crowdsourcing. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems, CHI'14*, pp. 31–40, 2014. 3

[27] C. Lai, Z. Lin, R. Jiang, Y. Han, C. Liu, and X. Yuan. Automatic Annotation Synchronizing with Textual Description for Visualization, chi'20. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–13. ACM, 2020. 2, 3, 5

[28] X. Lan, Y. Shi, Y. Wu, X. Jiao, and N. Cao. Kineticharts: Augmenting affective expressiveness of charts in data stories with animation design. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):933–943, 2021. 2, 9

[29] 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*, 2021. 2, 3, 5

[30] H. Li, Y. Wang, Q. V. Liao, and H. Qu. Why is AI not a Panacea for Data Workers? An Interview Study on Human-AI Collaboration in Data Storytelling. *arXiv: 2304.08366*, 2023. 9

[31] H. Li, Y. Wang, A. Wu, H. Wei, and H. Qu. Structure-aware Visualization Retrieval. In *Proceedings of the CHI Conference on Human Factors in Computing Systems, CHI'22*, vol. 1, pp. 1–14. ACM, apr 2022. 9

[32] W. Li, Z. Wang, Y. Wang, D. Weng, L. Xie, S. Chen, H. Zhang, and H. Qu. GeoCamera: Telling Stories in Geographic Visualizations with Camera Movements. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, CHI'23*, vol. 1, pp. 1–15. ACM, 2023. 9

[33] Y. Lin, H. Li, A. Wu, Y. Wang, and H. Qu. DMiner: Dashboard Design Mining and Recommendation. *IEEE Transactions on Visualization and Computer Graphics*, 14(8):1–15, 2023. 9

[34] V. Liu and L. B. Chilton. Design Guidelines for Prompt Engineering Text-to-Image Generative Models. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, CHI'22*, pp. 1–23. ACM, 2022. 5

[35] Y. Luo, Y. Zhou, N. Tang, G. Li, C. Chai, and L. Shen. Learned Data-aware Image Representations of Line Charts for Similarity Search. *Proceedings of the ACM on Management of Data, SIGMOD'23*, 1(1):1–29, 2023. 9

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

[37] S. McKenna, N. Henry Riche, B. Lee, J. Boy, and M. Meyer. Visual Narrative Flow: Exploring Factors Shaping Data Visualization Story Reading Experiences. *Computer Graphics Forum*, 36(3):377–387, 2017. 3

[38] R. Metoyer, Q. Zhi, B. Janczuk, and W. Scheirer. Coupling Story to Visualization: Using Textual Analysis as a Bridge Between Data and Interpretation. In *23rd International Conference on Intelligent User Interfaces*, IUI '18, pp. 503–507. ACM, 2018. 3

[39] D. Moritz, C. Wang, G. L. Nelson, H. Lin, A. M. Smith, B. Howe, and J. Heer. Formalizing Visualization Design Knowledge as Constraints: Actionable and Extensible Models in Draco. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):438–448, 2019. 3, 5, 9

[40] D. Ren, M. Brehmer, Bongshin Lee, T. Hollerer, and E. K. Choe. ChartAccent: Annotation for data-driven storytelling. In *Proceedings of the IEEE Pacific Visualization Symposium, PacificVis'17*, pp. 230–239. IEEE, 2017. 5

[41] M. Ruoff, B. A. Myers, and A. Maedche. ONYX: Assisting Users in Teaching Natural Language Interfaces Through Multi-Modal Interactive Task Learning. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, CHI'23*, pp. 1–16. ACM, apr 2023. 9

[42] D. M. Russell and E. H. Chi. Looking Back: Retrospective Study Methods for HCI. In J. S. Olson and W. A. Kellogg, eds., *Ways of Knowing in HCI*, pp. 373–393. Springer, 2014. 3

[43] E. Segel and J. Heer. Narrative visualization: Telling stories with data. *IEEE transactions on visualization and computer graphics*, 16(6):1139–1148, 2010. 1, 2, 3

[44] D. Seyser and M. Zeiller. Scrollytelling – An Analysis of Visual Story-

telling in Online Journalism. In *Proceedings of the 2018 22nd International Conference Information Visualisation, IV'18*, pp. 401–406, 2018. 9

[45] H. Shen, C.-Y. Huang, T. Wu, and T.-H. K. Huang. ConvXAI: Delivering Heterogeneous AI Explanations via Conversations to Support Human-AI Scientific Writing. *arXiv: 2305.09770*, 2023. 9

[46] 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

[47] 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), EuroVis'21*, pp. 91–95. Eurographics, 2021. 9

[48] L. Shen, E. Shen, Z. Tai, Y. Wang, Y. Luo, and J. Wang. GALVIS: Visualization Construction through Example-Powered Declarative Programming. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management, CIKM'22*, pp. 4975–4979. ACM, 2022. 9

[49] 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. 3

[50] L. Shen, Z. Tai, E. Shen, and J. Wang. Graph Exploration with Embedding-Guided Layouts. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–15, 2023. 9

[51] 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

[52] Y. Shi, X. Lan, J. Li, Z. Li, and N. Cao. Communicating with motion: A design space for animated visual narratives in data videos. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, CHI'21*, pp. 1–13, 2021. 1, 2, 4, 6

[53] A. Srinivasan, S. M. Drucker, A. Endert, and J. Stasko. Augmenting Visualizations with Interactive Data Facts to Facilitate Interpretation and Communication. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):672–681, 2019. 3

[54] 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 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, pp. 1–17. ACM, 2021. 3

[55] A. Swearngin, C. Wang, A. Oleson, J. Fogarty, and A. J. Ko. Scout: Rapid Exploration of Interface Layout Alternatives through High-Level Design Constraints. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, CHI'20*, pp. 1–13. ACM, 2020. 3, 5

[56] J. Thompson, Z. Liu, W. Li, and J. Stasko. Understanding the Design Space and Authoring Paradigms for Animated Data Graphics. *Computer Graphics Forum*, 39(3):207–218, 2020. 2

[57] J. Thompson, Z. Liu, and J. Stasko. Data Animator: Authoring Expressive Animated Data Graphics. In *Proceedings of the ACM Conference on Human Factors in Computing Systems, CHI'21*, 2021. 2

[58] B. Tversky, J. B. Morrison, and M. Betrancourt. Animation: can it facilitate? *International Journal of Human-Computer Studies*, 57(4):247–262, 2002. 2, 4, 6

[59] Y. Wang, Y. Gao, R. Huang, W. Cui, H. Zhang, and D. Zhang. Animated Presentation of Static Infographics with InfoMotion. *Computer Graphics Forum*, 40(3):507–518, 2021. 2, 9

[60] 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. 9

[61] 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. *arXiv: 2308.04040*, pp. 1–11, 2023. 7

[62] A. Wu, Y. Wang, X. Shu, D. Moritz, W. Cui, H. Zhang, D. Zhang, and H. Qu. AI4VIS: Survey on Artificial Intelligence Approaches for Data Visualization. *IEEE Transactions on Visualization and Computer Graphics*, 28(12):5049–5070, 2022. 9

[63] 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 2022 CHI Conference on Human Factors in Computing Systems, CHI'22*, pp. 1–22. ACM, 2022. 9

[64] L. Xie, X. Shu, J. C. Su, Y. Wang, S. Chen, and H. Qu. Creating Emordle: Animating Word Cloud for Emotion Expression. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–14, 2023. 9

[65] P. Xu, H. Fu, T. Igarashi, and C.-L. Tai. Global beautification of layouts with interactive ambiguity resolution. In *Proceedings of the 27th annual ACM symposium on User interface software and technology, UIST'14*, pp. 243–252. ACM, 2014. 3

[66] L. Ying, X. Shu, D. Deng, Y. Yang, T. Tang, L. Yu, and Y. Wu. MetaGlyph: Automatic Generation of Metaphoric Glyph-based Visualization. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–11, 2022. 9

[67] Q. Zhi, A. Ottley, and R. Metoyer. Linking and Layout: Exploring the Integration of Text and Visualization in Storytelling. *Computer Graphics Forum*, 38(3):675–685, 2019. 3

[68] J. Zong, J. Pollock, D. Wootton, and A. Satyanarayan. Animated VegaLite: Unifying Animation with a Grammar of Interactive Graphics. *IEEE Transactions on Visualization and Computer Graphics*, 29(1):149–159, 2023. 2