StoryBlocks: Towards AI-Assisted Narrative Design for Data-Driven Storytelling

ANONYMOUS AUTHOR(S)

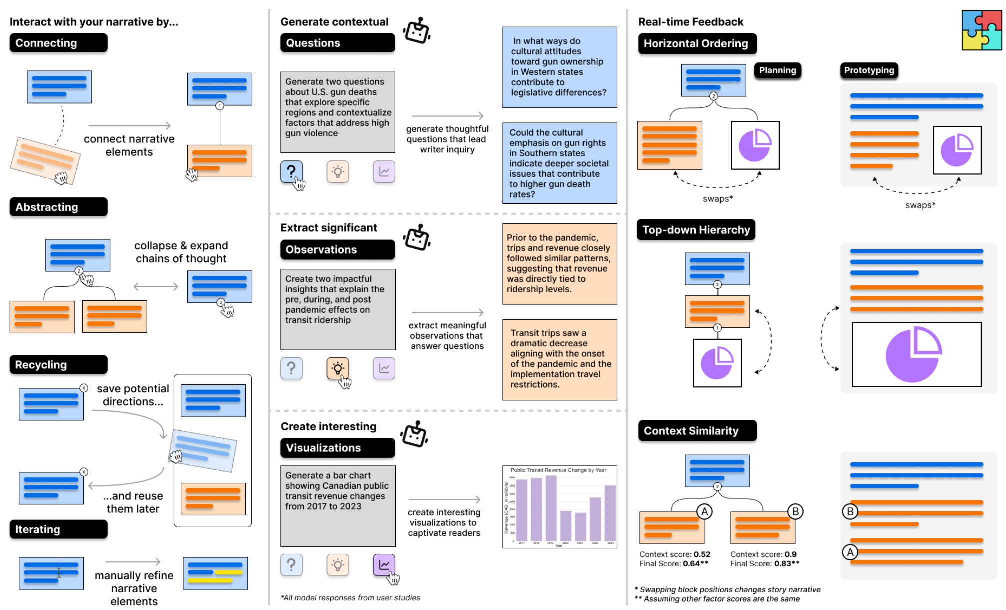


Fig. 1. StoryBlocks is an interactive environment supporting narrative design for data driven storytelling. The system is composed of three fundamental story elements: (1) questions that guide inquiries, (2) observations that provide supporting statistics, and (3) visualizations that capture dominant trends. It incorporates a multi-agent architecture to support connecting, abstracting, recycling and iterating various story elements for hierarchical organization of narrative.

Narrative design is central to data-driven storytelling (DDS), providing structure to data stories through sequenced observations, visualizations, and interpretations. Creating compelling DDS narratives is an iterative and collaborative process that requires both analytical reasoning and creative storytelling. While artificial intelligence (AI)–powered tools can assist in generating insights, visualizations, and descriptive summaries, they rarely support deeper reflection on narrative flow, organization, or alignment with story writer intent. We propose a dynamic, inquiry-guided narrative design approach that enables writers to establish writing objectives, structure ideas, reflect on their story, and refine stories. Our approach builds on *Socratic Questioning*—a technique widely adopted in learning technologies to help individuals explore complex ideas through thoughtful, open-ended questions. We embed this approach in StoryBlocks, an interactive system that combines hierarchical layout generation, non-linear structuring, real-time feedback, and multi-agentic interaction. Through systematic evaluations, we provide actionable insights for advancing Human–AI collaboration in narrative design.

Manuscript submitted to ACM

CCS Concepts: • **Human-centered computing → User studies**; *User interface programming*; *User centered design*; **Interface design prototyping**.

Additional Key Words and Phrases: Narrative Design, Data-Driven Storytelling, Human-AI Collaboration

ACM Reference Format:

Anonymous Author(s). 2018. Storyblocks: Towards AI-Assisted Narrative Design for Data-Driven Storytelling. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (Conference acronym 'XX).* ACM, New York, NY, USA, 33 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Narrative design in data-driven storytelling (DDS) refers to the process of *structuring* raw data, insights, and visualizations into a compelling, cohesive story to systematically communicate meaning and viewpoints [31, 49, 97]. Unlike simply presenting raw statistics or charts, narrative design emphasizes *story logic* - the sequencing, framing, and rhetorical strategies - that transforms data into a compelling narrative [91]. This process is not limited to mere structuring of information based on predetermined logic; it involves *ideating*, *exploring*, *probing*, and *refining* the very logic used to structure the information. Designing compelling narratives for DDS, therefore, is a cognitively demanding task that necessitates creative thinking, iterative refinement through feedback, critical reflection, and the ability to effectively explain, persuade, and engage target audiences [5]. Modern AI-powered writing technologies help non-expert writers by offloading various aspects of DDS workflows, such as automatically extracting insights from large datasets [79], building high-quality visualizations [32], translating statistical observations into written text [35, 84], and even automatically generating entire stories [49]. While outsourcing these tasks to AI systems help lower the entry barrier for non-expert data story writers[97], it often disengages the writers from these processes.

This disengagement risks deteriorating the narrative quality, as AI-generated outputs tend to exhibit *structural homogeneity* and *reduced diversity* in narrative flow [85, 103], despite resultant stories appearing novel or enjoyable at first [82]. Moreover, use of automated AI tools in writing workflows can have lasting negative impacts on an individual's writing style, limiting their expressivity and creativity and overall writing experience [102, 105]. Although prior research demonstrates that AI-powered writing assistants offer opportunities for more nuanced storytelling [63, 64], they can also diminish the critical reflection skills acquired during the creative process [59, 89, 114]. This is particularly problematic for non-expert writers engaging in DDS workflows, as it can lead to inaccurate data stories that present biased views of the world. In this research, we investigate the design space of AI-powered writing technologies that support non-expert writers to explore, ideate, reflect, and experiment with different data-driven story narratives while facilitating active engagement and reflection with the data. We argue that to effectively support the narrative design process, AI writing assistants should function as cognitive scaffolds that guide writers in sequencing their ideas, framing their interpretations, and helping them adopt compelling rhetorical strategies.

Narrative theorists define a narrative as a *representation of an event or a series of events* [1, 75, 81], or "*the 'telling' of a sequence of events* in a specific way so as to invite the readers to take a particular position towards the story" [51, 91]. In the context of DDS, narrative design is a method for synthesizing statistical observations from large complex datasets (existing as tables, spreadsheets, databases, or documents), and presenting them as a *meaningful* and *logical* narrative, referred to as a *data story* [4, 74]. Crafting coherent and effective data story narratives requires authors to engage deeply with the subject matter, reflect on their communicative goals, and select story elements that best represent those goals[5, 22]. One effective way to support this engagement is through a *question-guided inquiry* process, where writers explore complex ideas by posing a sequence of thought-provoking and deductive questions that build upon one another [15, 56, 97]. This approach, known as *Socratic questioning*, proposed by Socrates, is now widely used in educational and learning technologies to promote internal reflection [57, 71, 108]. *Socratic questioning* offers a structured way to help writers reflect on their narrative choices[57]. Instead of relying on fixed templates, it prompts individuals with targeted reflective questions that help them examine assumptions, clarify goals, and interpret observations more thoughtfully. Prior research shows that Socratic questioning based inquiry can be a powerful tool to encourage reflection and engagement with the subject matter [27, 71, 108]. In this research we investigate, *how can we design writing technologies that engage non-expert writers of data-driven stories in a Socratic dialogue?*

Rather than treating *Socratic questioning* as a purely reflective exercise, we operationalize it as a design strategy embedded within an AI-supported writing assistant. In our system, the AI poses thought-provoking questions on-demand to prompt critical engagement with data insights, visualizations, and narrative decisions. For instance, when a writer extracts an initial data insight, such as a trend capturing changes in public transit ridership, the system might first pose an *exploratory* question (e.g., "What broader trends or external factors might explain these changes?") to encourage contextual thinking. As an appropriate follow up to this, the system may generate a *clarifying* question (e.g., "By how much did ridership vary across different transit modes or time periods?") that can help refine the writer's analysis. Alternatively, a follow up question generated by the system that *challenges assumptions* (e.g., "Are these trends consistent across all demographic or geographic groups, or do disparities exist?") may provoke deeper inquiry and reflection. Through an iterative dialogue, our system scaffolds the author's reasoning process, helping them connect statistical evidence to narrative objectives [45, 95].

Our goal is to investigate how writers experience of Socratic-style inquiry-guided writing approach shapes the DDS narrative design process and their internal narratology. To examine this, we built Storyblocks, an interactive system combining hierarchical data story layout generation, non-linear structuring, and real-time feedback with a multi-agentic architecture. Storyblocks represents data stories as hierarchical assemblies of three constituent elements - *questions*, statistical *observations*, and data *visualizations*. These elements, metaphorically represented as building blocks, can be created, connected, refined, and recycled through a mixed initiative and direct manipulation paradigm. The system enables writers to ideate, explore, probe and refine various narrative flows, rather than automating the entire writing process. While Storyblocks can help writers of any expertise, we specifically focus our study on non-professional, novice data story writers who lack necessary skills in sensemaking and story telling and are more likely to benefit from an inquiry-guided approach. Through a user study with 21 participants using Storyblocks to craft narratives, we uncovered how writers experience this inquiry, what types of questions effectively stimulate reflection, and when the system overwhelmed or misaligned with their goals. We observed that the inquiry-guided approach helped promote ideation, probing, and uncovering unexpected insights which effectively shaped rhetorical framing and narrative refinement. However, the approach also posed challenges for narrative sequencing, especially when balancing data extraction with narrative framing and countering writer's internal or external biases. Our research has three key contributions:

(1) We introduce an inquiry-guided system that integrates AI supported data exploration, hierarchical narrative planning, and iterative story refinement.
(2) We propose a novel story rendering algorithm that transforms non-linear narrative plans into a linear story to provide real-time feedback.

(3) We present actionable insights that draw from a systematic evaluation, uncovering writer's experience of inquiry-guided writing and its impact on their writing process.

In the following sections, we first motivate our research through related work, followed by a description of our formative study, system design, architecture, and user evaluation findings. Finally, we synthesize practical implications that will inform future narrative design tools for DDS workflows.

2 RELATED WORK

Data stories help writers communicate nuanced insights about broad social issues, and when paired with anecdotal accounts, can amplify reader engagement [107]. Prior research has explored widely popular DDS formats, such as annotated charts [10], magazine styles [58], comic strips[101, 111, 112], as well as several novel formats, such as performance arts, games, and augmented reality [41]. In this research, we focus on the DDS workflow employed in computational journalism, where writers communicate social issues via narrative text and data visualizations [4, 91]. We describe below the related work done in this field.

2.1 Artificial Intelligence for Data-driven Storytelling

Designing narratives for DDS requires authors to engage in a series of cognitively demanding tasks. Authors must establish writing goals, gather analytical observations, articulate and structure these to form coherent sequences, and iteratively refine their story to align with target audiences and idiosyncratic objectives [43]. To support authors in these tasks, a wide body of research has investigated automated data transformation techniques for narrative generation [36, 47, 48], with early systems primarily relying on template-based approaches [37, 76, 106]. For instance, Voyager [106] helps authors select suitable visualization from a recommended set using predefined statistical perceptual measures, while IntroAssist [48] helps authors write introductory help requests via expert and peer-supported brainstorming. These tools primarily incorporate rule-based approaches that provide precision, but lack adaptability and expressiveness.

Recent advances in generative AI offer increased adaptability in writing technologies for helping authors extract visualizations [20, 21], observations [32], and textual descriptions [49] from a variety of datasets. Furthermore, it can help writers elaborate on existing ideas [19, 29, 99] and adjust the overall story tone [35]. Systems such as Data2Text [26] and models like GPT-x have demonstrated the ability to generate fluent and coherent textual narratives from structured data inputs. When integrated into storytelling pipelines, they help to improve narrative coherence, personalization, and engagement, along with increased story production speed [18, 49, 92]. For instance, multi-step agentic frameworks like DataNarrative [49] mimic the human process of data story generation and can assist users in automatically creating story outlines and narration. While these AI-based writing tools help writers *translate* their ideas into stories, they often limit critical reflection [33, 83], reduce expressivity [54], and fail to assist with other demanding aspects of the writing process, such as story *planning*, *evaluation*, and *narrative design* [34, 55]. Recent studies have shown that AI-generated stories frequently stabilize around predictable patterns, leading to stylistic uniformity [24], suppressed originality [25], and standardized and homogeneous narratives [2, 65, 85] In this research, we investigate the design of AI-powered systems that support writer creativity, while preserving personal writing styles in DDS workflows.

2.2 Narrative Design in Data-driven Storytelling

Narrative design has been widely studied through the lens of *narrative theory*, a body of literature which suggests that the ability for human storytelling is a natural phenomenon of our evolution and a cornerstone in how we interpret the world [17, 98]. Narrative theorists argue that narratives help people construct, express, and negotiate personal and collective identities [13, 68]. Recent research has investigated narrative design across domains such as education [23, 52], journalism [44], and interactive media [16, 44], highlighting its role in shaping audience engagement, comprehension, and persuasion. *Within DDS, researchers define narrative design as the practice of identifying the most effective organization of insightful observations and visualizations to communicate the relevant information to a target audience* [62, 87, 91]. To create compelling data stories, authors must design their narrative by carefully sequencing insights (e.g. statistical observations and trends in data), alongside devices of visual communication (e.g. charts and animations), to write stories that support information discovery [31, 69]. For instance, if the objective of a data story is to persuade an audience towards a particular position, the author(s) may adopt an *argumentation* narrative style, using insights and visuals to support or refute that stance [53, 60]. Conversely, if the story objective is to engage readers in a dialogue, the author(s) may choose to start with a *rhetorical question*, and select insights that seek to answer this question [94]. Very rarely do writers finalize their narrative without some degree of experimentation and reflection, thus making narrative design an iterative process [17]. It is important to note that narrative design is distinct from story planning. While story planning involves organizing story elements (i.e., questions, observations, visualizations), according to a given story logic, narrative design focuses on formulating that story logic and iteratively refining it through those elements to communicate a viewpoint. In this research, we investigate the design space of AI-powered writing assistants that support narrative design in DDS contexts.

2.3 Socratic Questioning for Inquiry-Guided Writing

Recent research has demonstrated that human-AI collaborative interfaces can leverage *Socratic questioning* techniques to support critical thinking, reflection, and subject matter engagement [29, 62, 94]. Systems such as Thinking Assistants [80] and ExploreSelf [96] use reflective questions to help writers organize their thoughts using conversational agents. Beyond academic research, several commercial systems also adopt this approach with promising results. For instance, OpenAI's ChatGPT Study Mode [78] encourages inquiry-based learning in students by prompting open-ended questions, rather than supplying direct answers. Similarly, Primer's Deep-Talk AI [9] generates reflective prompts to facilitate dialogue within teams. While these early efforts signal a broader shift towards designing AI systems that facilitate human inquiry rather than simply automating content generation, the conversational interaction paradigm that underpins these systems still presents notable limitations. By assigning the lead inquiry position to the AI agent, such tools often restrict writer agency and expression, reducing authors to passive respondents in a linear dialogue. This is particularly problematic for narrative design, which is inherently non-linear and exploratory, requiring writers to shift fluidly between generating ideas, organizing subplots, and refining arguments [30, 34]. Recent research with layered interface paradigms, such as Script&Shift [94], Jupybara [100], VISAR [110], and Idyll [21] suggest alternative, more efficient ways to support writers. By combining non-linear thinking with linear computation, these tools move beyond traditional text editors or chat-based systems and offer opportunities for refocused cognitive efforts[94, 100, 110]. Perhaps the closest to our research are Jupybara [100], which helps users generate coherent narratives using AI-agents embedded in Jupyter Notebook, and VISAR [110], which supports rapid prototyping in argumentative writing. However, our work differs from these approaches by directly addressing the challenges of narrative design. In this research, we build upon this body of work and leverage *Socratic questioning* to emphasize the the cognitive faculties of narrativization, such as experimentation and ideation. In the below section we outline our formative study and design methodology for Storyblocks.

3 DESIGN REQUIREMENTS

Socratic questioning positions knowledge as something actively constructed through dialogue with oneself or others. By prompting individuals to articulate why they believe something, how they arrived at a conclusion, and what alternative explanations might exist, the technique encourages critical reflection and deeper engagement with the knowledge produced [27]. Methods in *Socratic questioning* also align closely with popular theories in composition studies, which argue that effective storytelling emerges from critical reflection and engagement with subject matter [6, 17, 30]. Within DDS, Socratic-style inquiry may prompt writers to articulate why a given data interpretation is relevant to the story goal and how it can shape reader's perspectives. Such inquiry can significantly improve the writer's ability to connect claims with supporting evidence, leading to compelling data stories.

When constructing the narrative for data stories, *Socratic questioning* can support cognitive processes, such as sensemaking and hypothesis generation, that enable writers to move beyond surface-level observations [5, 88]. However, integrating such inquiry-guided approaches into narrative design tools poses unique challenges. The predominant friction is the interaction novelty introduced when exploring a fundamentally different mode of engagement compared to traditional writing assistants. Inquiry-guided writing requires authors to iteratively engage in reflective dialogue without losing sight of their writing goals or experiencing cognitive overload. To uncover such challenges for an inquiry-guided narrative design tool, we conducted a formative study with 9 participants.

The goal of our formative study was to understand the salient design features that can situate the users in an interactive dialogue with the system during DDS workflow. To achieve this, we developed a preliminary prototype that supported data-driven story writing using five publicly available datasets. These datasets covered post-pandemic transit ridership [70], global air pollution emissions [90], gun violence [40], federal election voting [61], and mortgage pricing [67] respectively. We manually curated a set of *observations*, *visualizations*, and guiding *questions* for each dataset. These elements served as the fundamental blocks to data stories and we embedded them in an interactive interface through a hierarchical graph layout metaphor. Prior research demonstrates that hierarchical story graphs can offer powerful interaction paradigm for narrative generation [86]. The interface enabled participants to create and organize these blocks, connect them through simple drag-and-link interactions, and render their hierarchically connected ideas into a linear story. To encourage inquiry, we embedded two restraints on building blocks: observations and visualizations could connect only to central sub-questions, while questions could branch to other questions.

We acknowledge that this initial prototype may have shaped how participants interacted with the system and eventually our findings. However, our formative study focused on identifying the most salient design features rather than curating an exhaustive list. Despite the interaction structure imposed by our prototype, the participants' behaviors and feedback reliably exposed critical design guidelines (**R1–R5** as described later in the section). While this technique can help writers with all levels of expertise, we focus our inquiry on unskilled writers - individuals with no formal training in professional writing or data-driven storytelling. We focus on these individuals for three reasons. Firstly, non-experts lack stable schema and routines for story planning and rhetorical framing, [14], which can make them more inclined to explore varied topic ideas and experiment with flexible narrative structures [7, 8] This makes them well suited for evaluating how effectively the system adapts to diverse and evolving writing styles. Secondly, observing changes in the reasoning and narrative development of unskilled writers enables us to extract a clear view of the system's impact without the confounding influence of prior expertise. Lastly, novice data story writers may benefit more from an inquiry-guided approach, as leading with questions supports their learning process and lowers the barrier to entry.

We recruited 9 participants between the ages of 21 - 55 years who were a mix of industry professionals and university students with limited prior experience in DDS. During this formative study, participants completed a brief tutorial that introduced story elements, several illustrative examples of data stories, and how to interact with the system. Participants were then prompted to design a narrative by asking questions, incorporating supporting observations and visualizations, linking ideas, and iterating on their prototype story. The free form interaction enabled them to explore any story angle and pursue a story direction of their choice using the pre-loaded observations and visualizations. We gathered qualitative data using contextual inquiry, a think-aloud protocol, and a post-usage reflection. Observation notes captured how participants used the interface to write their stories while think-aloud transcripts and post-usage reflection revealed participants' reasoning processes and strategies for constructing their narratives. We conducted a reflexive thematic analysis of this data to identify recurring patterns in participants' reasoning and extract actionable design requirements for the system We synthesize our findings from this preliminary study into as below:

Facilitating evolving story structures (R1): While many participants relied on our pre-existing library of story elements to assemble their initial drafts, these only served as starting points. As their ideas developed, participants expressed a need to manually modify and adapt their question and observations blocks to align more closely with evolving narrative goals. These edits frequently involved rephrasing questions to sharpen their focus, reframing observations to highlight relevance, and selectively curating which visualizations best supported their emerging storyline. The ability to filter and modify generated content provided a sense of agency in the writing process, suggesting that *systems should balance automation with opportunities for customization to support evolving story structures*.

Supporting non-linear story organization (R2): To better support the inherently non-linear ways in which individuals brainstorm, plan, and refine narratives, we encouraged a hierarchical structuring of story elements. Participants responded to this positively, noting that it *"allowed for easier flow of thought and felt more natural"* compared to traditional linear planning. However, this also revealed new design challenges. For example, when narratives grew to include a large number of blocks and dense interconnections, participants found difficulties managing, revisiting, and reflecting on individual story elements. To address this, they expressed the need for pan-and-zoom interactions and mechanisms for organizing and navigating large, complex trees of thought. Participants also frequently repositioned blocks as they experimented with aligning story elements, iterating on structures, and discarding connections that no longer served the narrative. This suggests that *systems supporting non-linear planning paradigms should incorporate interaction mechanisms for effective connection and navigation of ideas.*

Enabling multiple entry points (R3): Participants approached narrative construction from different starting points, reflecting the fluidity of their creative processes. Participants began by formulating questions to guide exploration, leading with compelling data-driven observations, or anchoring their stories in predominant visualizations. One participant explained that starting with observations made it easier to *"see patterns first, then figure out the story"*, while another preferred *"posing questions to get a sense of direction before diving deeper"*. This suggests that narrative design *systems should enable multiple entry points to support their sensemaking strategies and avoid enforcing structural assumptions.*

Promoting ideation (R4): Participants used a mixture of pre-existing and manually-created story elements to frame their narrative. In the early stages of narrative construction, they frequently engaged in open-ended exploration. This occurred when pre-existing elements and exploring multiple possibilities was more valuable than committing on a single storyline. To remedy this, participants expressed the need for features that encouraged ideation and experimentation. For example, the ability to add new question blocks that open alternate lines of inquiry, experiment with alternative visualizations, or edit the framing of observations to explore new perspectives were all mentioned. Additionally, participants often relied on questions as entry points into their narratives, but noted the difficulty of consistently generating meaningful or well-scoped question prompts on their own that were appropriately tailored towards the available data. These observations suggest that *systems should offer mechanisms to open multiple lines of inquiry and support rapid data exploration.*

Supporting feedback loops (R5): Several participants noted that their evolving prototype story clarified their thought process and provided useful reflection that helped them improve their narrative structure. However, they mentioned it was challenging to anticipate how the final output might read, which made it difficult to identify gaps and adjust their story structure accordingly. This suggests that supporting tight integration between non-linear planning and real-time story rendering is critical to foster effective feedback loops. As such, *systems that employ a story-prototyping mechanism should provide precise linearization heuristics within the interface and ensure a tight feedback loop with the graphical story.*

In this research, we incorporate the above design requirements in Storyblocks, an interactive system supported through a multi-agent architecture to assist narrative design in DDS workflow. In the following section, we describe our system design.

4 SYSTEM DESIGN

We developed Storyblocks to support writers with rapidly generating, experimenting and, evaluating different narratives through a building blocks metaphor. Our system is comprised of five essential design elements: a dynamic contextual question generator, an insight explorer, a narrative design playground, a block enrichment mechanism, and a story renderer.

4.1 Dynamic Contextual Question Generator

Writers continuously ask questions throughout the narrative planning process. These questions often act as pivots, guiding the flow of thought and linking facts to form coherent arguments[28, 39]. Therefore, posing thought provoking questions that encourage writers to challenge assumptions, explore new perspectives, and promote evidence-seeking are essential for crafting narratives. To foster this process, Storyblocks incorporates a dynamic question generator. The question generator creates engaging, thought provoking questions on-demand, which the writers can modify and queue as necessary. Writers may generate questions (and question queues) by entering a query into the query box (Figure 2-A) and invoking the *Question Agent* (QA) through a simple button click. For instance, a writer may enter an open-ended query, such as *"Generate a question about public transit usage"*, and then click on the generate question button (Figure 2-B). Upon receiving the user query, the QA first verifies it for robustness and correctness. It then retrieves the relevant dataset from the in-memory vector store, embeds both the query and contextual information into a prompt, and uses a generative model (GPT-4o-mini) [77] to produce candidate questions (Figure 3).

The QA generates questions based on specific entities or relationships between different entities **(R4)**. For instance, given the above user query, the QA might generate a central guiding question, such as *"How does fluctuation in employment rate impact public transit usage?"*. It may also generate follow-up questions, such as *"How has employment rate changed over the past years?"* or *"Which socio-economic factors link employment with public transit?"*, to help scaffold the emerging argument. These questions appear in the narrative design playground as mutable blocks that can be repositioned, and (re-)connected to other blocks (Figure 2-E). This free-form, natural language interaction offers high user expressivity and sufficient specificity to generate a diverse pool of question blocks. The user may refine the QA generated questions or create their own without the QA to guide their inquiry. They can accomplish this by directly

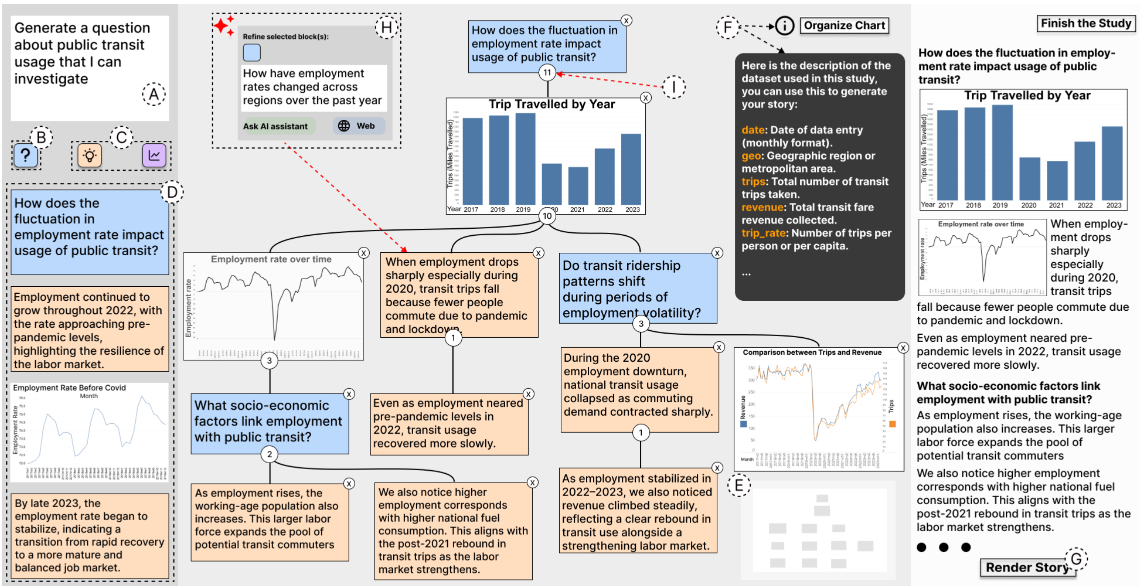


Fig. 2. The Storyblocks interface supports a node-based hierarchical structure for non-linear story planning. From the query box (A), users can query the question generator for relevant questions (B), as well as explore insights (i.e., observations and visualizations) (C). The scratch pad (D) stores unused elements for future use, while the real-time story renderer (G) updates dynamically as the narrative evolves. The narrative design playground (E) allows users to create and connect question, observation, and visualization blocks to build their story. An accompanying overview of the dataset (F) is available for reference as well. The block enrichment feature (H) allows writers to refine textual blocks using AI agents, while branches of blocks can be expanded or collapsed (I) to support navigation. This design enables writers to iteratively refine their ideas while maintaining a clear overview of the emerging narrative.

editing text in each question block or creating an empty question block through the same button click and adding a textual question to it **(R1)**. An overview of the dataset is provided as a tooltip that allows writer to familiarize themselves with all entities contained in the dataset (Figure 2-F).

4.2 Insight Explorer

While the inquiry is lead by questions, statistical insights ground it in data. We facilitated data exploration through natural language interactions, enabling users to query the dataset, search for evidence, and retrieve relevant observations to further their line of inquiry. Substantial improvements to LLM capabilities in recent years has made it possible to query LLM's directly through natural language, lowering the entry barrier for data exploration [3, 66, 73, 113]. Building upon this progress, Storyblocks incorporates an *Insight Explorer* (IE): an AI agent that can retrieve statistical findings and trends relevant to specific queries. Writers may extract insights by entering a natural-language query into the query box and selecting either textual descriptions or visualizations to be generated in response. (Figure 2-C). Continuing the previous example, after generating a question block, such as *"How does the fluctuation in employment rate impact public transit usage?"*, the writer might input a query in the query box, to retrieve supporting evidence. For instance, they might enter *"Show me public transit usage trends over the past five years."* in the query box, and select the preferred output format. They might then explore follow-up questions about employment rate changes or socioeconomic factors, and issuing additional queries such as *"Show me employment rate trends for the past five years"* or *"Provide a breakdown* *of demographic groups using public transit over the past five years.”*. The extracted insights may help the writer link employment and socio economic factors with public transit. At this stage, they may also invoke QA to generate a new question that leads to a different line of inquiry altogether. It should be noted that, while the symbiotic relationship between questions and insights may be mistaken for an equivalency, we argue that questions open up lines of thinking, whereas insights close them or lead to further inquiries. The distinction for both is necessary, as the respective structure that follows is fundamentally different **(R1)**.

We incorporated two modes of insight exploration - *textual* and *visual*. *Textual mode* delivers written descriptions of salient trends, patterns, outliers, and observations within the data. These outputs are descriptive, rather than predictive, and can support a range of analysis functions, including event diagnosis, argument construction, and hypothesis evaluation. *Visual mode* generates visualizations of queried variables and trends in the format of bar charts or scatter plots. The resulting visualizations are rendered as static charts embedded within movable blocks. The visual agent serves two complementary purposes: helping writers visualize key trends to identify potential story leads, and presenting insights to readers that are difficult to communicate through text alone **(R3)**. Insight blocks similarly appear as mutable blocks on the narrative design playground that can be repositioned and linked with other blocks to facilitate flexible narrative organization **(R2)**. Both textual and visual agents first validate the user’s query for logical correctness and contextual relevance. The textual agent then combines the user query with the retrieved context and prompts an LLM to generate statistical observations. The visual agent operates similarly, except it generates charts using d3.js [11] in the backend (Figure 3).

4.3 Narrative Design Playground

Storyblocks provides an interactive environment that emphasizes non-linear planning and the deconstruction of high-level ideas into smaller, mutable units of information. This helps writers intuitively decompose their ideas into concentrated directions and discourages them from circular thinking **(R2)**. Writers can begin narrative construction by generating contextual question blocks with the *Question Generator* and extracting relevant insights with the *Insight Explorer*. They can then organize these blocks into hierarchical story graphs that represent emerging lines of inquiry **(R3)**. For instance, a writer examining how fluctuations in employment rates influence public transit usage, might create question blocks such as *“How have employment rates changed across regions over the past year?"* and *“Do transit ridership patterns shift during periods of employment volatility?"*. Then, they might create insight blocks that contain observations about month-to-month employment changes and corresponding transit usage trends. Writers can spatially connect these blocks using *snapping connectors*, which automatically link two blocks placed in proximity and easily disconnect when dragged apart **(R2)**. This interaction helps writers form mental models of how their ideas can be clustered together, promoting structured arguments and deeper exploration into their story. For instance, the writer may connect a question block about short-term employment fluctuations to an insight block showing a sharp ridership decline during the same period. They may also use the connection to hypothesize a causal mechanism or identify a gap for further investigation.

In narrative design, nascent ideas often emerge that are not immediately relevant, but may be valuable later. This occurs even more frequently with collaborative exploration, e.g. generative agents, where unexpected responses can challenge or diverge from a writer’s initial assumptions. Storyblocks supports this by incorporating a scratch pad for storing work-in-progress ideas that writers can later refine and incorporate into their narrative (Figure 2-D). This serves two purposes: first, it allows writers to concentrate on individual lines of thinking without overloading their cognitive capacity; second, it promotes a diverse approach to data-driven exploration that limits biased writing **(R4)**. While the scratch pad can support cognitive processing, an inquiry grounded in *Socratic Questioning* can quickly produce story graphs that become quite large. This can lead writers to visually lose track of their story and can reduce mental clarity. To address this issue, Storyblocks allows writers to *reveal* and *hide* sub-plots within their narrative by expanding and collapsing sub-graphs of elements (**R2**) (Figure 2-I). Each sub-graph displays the number of blocks contained below them. This abstraction allows writers to focus on specific sections of their story and provides a clear mechanism for reflective observations on the macro-level objectives of their narrative. We also incorporated a a pan-and-zoom mechanism, to help writers navigate large hierarchical graphs **(R2)** (Figure 2-E).

4.4 Real-time Story Renderer

While non-linear thinking can improve cognitive clarity, it can be difficult for writers to visualize how their story will be viewed by readers. To address this challenge, Storyblocks includes a Real-time Story Renderer that converts a non-linear story graph into a linear narrative (Figure 2-G). We introduce a novel *proximity-based algorithm* that powers this renderer, transforming non-linear hierarchical graphs of story elements (questions, observations, and visualizations) into logically connected linear paragraphs. Our approach stands out for its ability to unify spatial structure (depth, and breadth of elements) with the semantic and contextual relationships among blocks. This produces layouts that emphasize clarity and preserves narrative coherence. We refer to the central questions, that guide the narrative flow, as *story blocks*, reflecting their important role in shaping the story. The algorithm then identifies sub-graphs of observations and visualization blocks that are (in-)directly connected to the central story block, while avoiding overlap with other story blocks. From this point, the algorithm proceeds in two nested stages: (1) an *intra-story block layout*, in which nodes represent individual elements (or blocks), such as questions, observations, and visualizations; and (2) an *inter-story block layout*, in which nodes represent the question-led clusters previously identified. In both stages, the layout is determined by iteratively expanding on a set of selected nodes according to proximity-based scoring functions, as defined below.

The process for both intra- and inter-story block arrangements begins with a designated root question node, r, with set $S = \{r\}$. To determine the next element in the sequence, all candidate nodes within a fixed radius (e.g., two depth levels) of the current set S are identified. Each candidate, b, is scored using three complementary metrics:

- **Position score** captures structural proximity by rewarding both depth alignment (top-down hierarchy) and sibling adjacency (horizontal ordering). Formally, we consider a candidate's depth and its sibling relationship when computing its structural score.
- **Flow score** encodes narrative intent by favoring nodes that are connected via user-defined edges. This step emphasizes user agency and ensures consistency with the intended story flow.
- **Context score** measures the semantic coherence of textual node (i.e., questions, observations) embeddings using BERT-similarity [109]. Because charts do not provide textual content, their context score is redistributed to the position and flow scores.

The final candidate score is a weighted combination of these metrics:

$$\text{final\_score}(S, b) = w_p \cdot \text{position}(S, b) \;+\; w_f \cdot \text{flow}(S, b) \;+\; w_c \cdot \text{context}(S, b) \qquad \text{where } w_p + w_f + w_c = 1$$

At each iteration, the highest-scoring candidate is added to S, and the reachable set for the next iteration are updated to include candidates within the new radius. This process repeats until all nodes are included, resulting in a linearized narrative sequence from a non-linear story graph. Our proposed algorithm produces layouts that not only preserve hierarchical continuity, but also highlight meaningful clusters of related ideas that better reflect authorial intent.

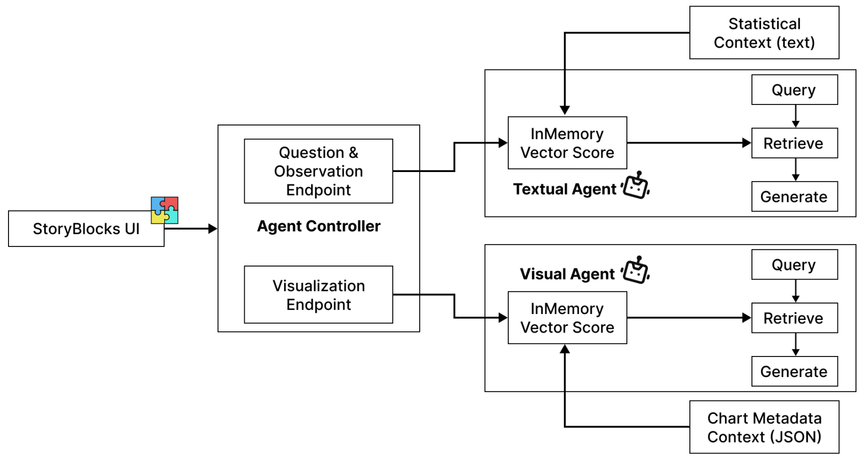


Fig. 3. The system architecture for StoryBlocks, demonstrating the three layers of abstraction: the user interface, the agent controller, and the respective textual and visual agents.

4.5 Block Enrichment Tool

StoryBlocks enables writers to formulate writing goals by creating questions to define the narrative flow and extracting relevant insights using embedded AI agents. However, as stories progress, writers often need to re-frame their arguments to better align with their evolving narrative. To support this refinement process, we embed a *Block Enrichment Tool* within StoryBlocks that allows writers to query an LLM (GPT-4o-mini) for real-time edits to individual or groups of blocks. It should be noted that this is separate from the agents used for creating questions and insights, as this model does not have access to the dataset embeddings. This feature helps writers rephrase questions, surface external interpretations that may have been originally overlooked, and optionally integrate relevant context from across the web **(R1)**. In the context of our employment–transit example, a writer might enrich a question block such as *"How have employment rates changed across regions over the past year?"* by incorporating recent contextual details - for instance, data from a new regional labor report or a policy change affecting transit subsidies. The writer may also incorporate additional facts and clarifications into this editing process (Figure 2-H).

The system architecture of StoryBlocks is organized into three core layers of abstraction. These include a frontend user interface, an intermediary API layer responsible for managing agent communication, and backend generative agents queried through user prompts (Figure 3). This lightweight, modular design allows for fast API calls with minimal overhead and guarantees responsive behavior, regardless of dataset or narrative size. We developed the system as a RESTful web application using React and React Flow for the sandbox environment. The dockerized image was hosted on a secure Linux server and OpenAI's GPT-4o-mini [77] was queried for the generative agents. In the next section, we describe the methodology for our subsequent user studies.

5 STORYBLOCKS: SYSTEM WALKTHROUGH

To illustrate how StoryBlocks supports narrative design in DDS workflow, we walk through a user scenario where Eva, an inexperienced writer, is working on a data story on employment rate fluctuations and public transit usage. Please refer to Figure 4 for key interactions supported by StoryBlocks.

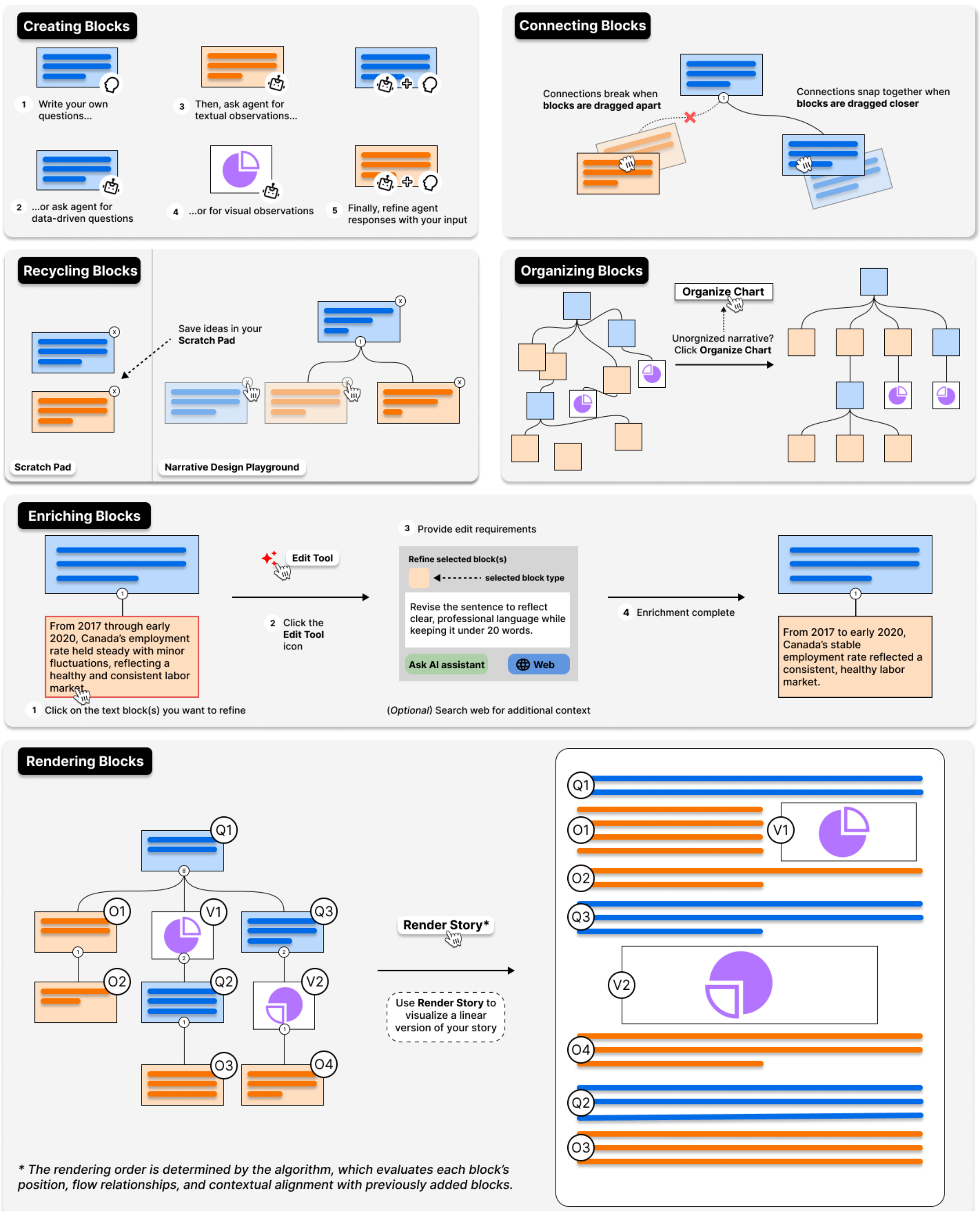


Fig. 4. The key interactions for StoryBlocks. Users can generate or write their own questions blocks and answer them with generated observation or visualization blocks. They can connect or disconnect blocks together by moving them closer or apart, recycle blocks for future use on the scratch pad, and enrich the content of textual blocks in real-time. To support planning, they can organize their non-linear story graph into structured readable graphs and render it as a linear prototype of the final story.

Eva, a novice data story writer, aims to write a story on whether employment rates correlate with public transit usage using a dataset she recently obtained. Although motivated, she lacks a clear narrative direction and finds the dataset difficult to explore due to her limited experience with data analysis tools. To initiate her investigation, she turns to Storyblocks. Eva begins by entering into the query box, "Create several questions involving employment volatility and public transit usage", and clicking the Generate Question button. The Question Agent generates several candidate *Exploratory*question blocks, which appear as interactive nodes within the Narrative Design Playground. One generated question "How have regional differences in employment rates changed over the past decade?", attracts her attention and she decides to pursue this line of inquiry. To maintain a clean workspace, she moves the remaining question blocks to Scratch Pad, via the Delete (X) control for possible later use. Eva then selects her chosen question block and opens the Block Enrichment panel, refines the wording for specificity and supplements the question with brief contextual notes sourced from recent regional labor reports across the web.

Eva extends her line of inquiry by selecting the question block and formulating several exploratory prompts. In the query field, she types, "Find several insights that build on regional differences in employment rate.". The Insight Explorer generates multiple candidate observations as modular blocks on the narrative design playground. One observation highlights a notable disparity between employment rates and public transit usage experienced in New York City. Eva clicks on this particular observation block and asks follow-up queries such as "What factors might explain this difference?", "How does this pattern compare to other regions?", and "What contextual variables might I be overlooking?" She generates both textual descriptions and visualizations for these observations, by entering queries like "Create a graph plotting employment rates in New York City."

Eva then decides to deepen the line of inquiry, and clicks on a observation block and writes a query, "Generate questions involving employment trends in 2005 and their link to recent city policies." The QA generates a series of *Probing* questions or questions that *Challenge Assumptions*, on of which attracts Eva's attention: "To what extent might policy changes in 2005 have indirectly shaped later fluctuations in public transit usage, even if the relationship is not immediately visible in the data?". Encountering this question, prompts Eva to pause and reassess the causal pathways she had been assuming, and reconsider her writing goals. As these new questions populate the Narrative Design Playground, Eva reviews them to determine which lines of inquiry meaningfully advance her story and which can branch out into parallel inquiries. Once Eva gathers sufficient blocks, she begins to weave the story by connecting relevant blocks together by dragging them closer and seeing how the story reads in real-time using the Renderer. Because Eva leads the inquiry, the questioning layer supports her exploration without steering it, naturally shaping the emerging narrative structure and enabling her to identify the most compelling observations with greater intentionality.

6 METHODOLOGY

In the previous sections, we described our motivation, our design, and an illustrative walkthrough of Storyblocks. In this section, we describe our user study design to understand how an inquiry-guided approach motivated by *Socratic Questioning* shapes writer's experiences and internal narratology during the narrative design process. Our goal with the user study was to (1) investigate the impact of inquiry-guided writing assistant on the data-driven writing process, (2) understand writers' experiences with questions during narrative design, and (3) examine how the approach shapes participants' internal narratology during tasks such as sequencing, framing, and developing rhetorical strategies. We describe our study methodology in detail below.

6.1 Task

Storyblocks supports writing data-driven stories with diverse tabular datasets. For our user study, we selected three datasets and randomly assigned them to each participant for their story-writing task. The datasets spanned U.S. gun violence statistics [40], Canadian political trends [61], and Canadian transit ridership rates following COVID-19 [70]. Each dataset contained a mix of nominal and ordinal qualitative variables, as well as interval and ratio quantitative variables. Participants could reference these variables, along with short descriptions of the data in each column, through a tooltip panel embedded in the interface (Figure 2-F). Each participant was provided with an open-ended prompt - *Write a story of your choice guided by questions and the dataset provided.* After a quick tutorial detailing how to use the system, participants could begin creating their stories. Participants were free to select any story angle and create a narrative of their choice. They were encouraged to experiment with system features — including the Question Generator, Insight Explorer, Block Enrichment Tool, and Real-time Story Renderer — as they developed their narrative in the Narrative Design Playground.

6.2 Study Design

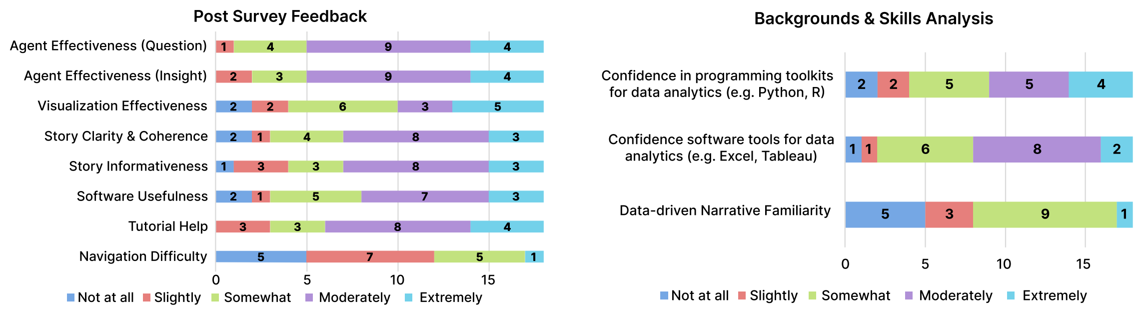
Each study session began with a verbal overview of the study objectives, followed by a pre-study questionnaire that assessed participant's experience with data-driven stories and general demographic information. The participants were then shown a tutorial that illustrated several data-driven story examples derived from popular mainstream news media. These examples were followed by a software tutorial that demonstrated how to use the system, e.g., interactions, features, controls, to write their own story with Storyblocks, and the writing prompt. Each study session lasted up to 60 minutes, although participants could conclude the study earlier if they felt their story was complete. Throughout the session, we employed a think-aloud protocol to elicit participants' real-time reflections on narrative structure, story planning, and the system's affordances. The study concluded with a post-study questionnaire that assessed usability, perceived utility, cognitive load, and participants' overall experience with an inquiry-guided story writing approach. All sessions were conducted individually, screen-recorded, and audio-recorded with consent from participants and the study was approved by the institutional review board / ethics committee.

6.3 Participant Recruitment

We recruited 21 participants using flyers and snowball sampling across the university campus. The participants were screened for prior skills in data-driven story telling, and we selected individuals with limited to no experience in data story writing. This participant pool represents the target audience most likely to benefit from an inquiry-guided approach. All participants were above 18 years of age and fluent in English. Each study session was designed to last an hour, and each participant was awarded a $20 gift card for their participation.

6.4 Data Collection

We gathered participants' data using pre-and post study questionnaires, think-aloud protocols, and interaction logs that captured participants' block creation, editing, and organization behavior. The pre-study questionnaire was administered before starting the study session and captured information about participants' (1) news reading habits, including frequency, genres, and preferred medium; (2) familiarity with writing, reading, and planning data stories; and (3) comfort with data analysis tools such as Python, R, Excel, and Tableau. The post-study questionnaire, administered at the end of the session, collected (1) demographic information such as age, education, and field of work, and (2)



(a) Post-survey feedback

(b) Backgrounds and Skills Analysis

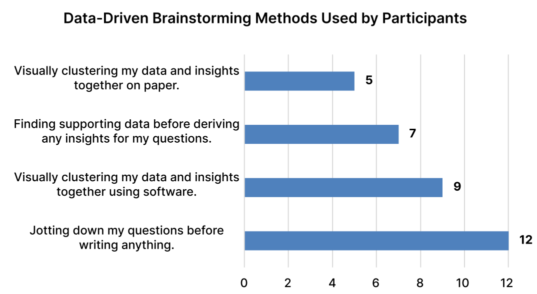
Fig. 5. Participants' post-survey responses regarding software and user experience(a) and their background and skill analysis(b)

participants' experiences using the system, including ease of interaction and navigation, perceived quality of the system's creative support, effectiveness in aiding narrative structuring, and open-ended reflections and comments. The interaction logging system gathered information about questions and insights generated, story organization mechanisms used, and other user interactions throughout the story writing process. During the study, we employed think-aloud protocols to encourage participants to articulate their reasoning as they planned and constructed their narratives. We transcribed the audio and screen recordings to capture participants' decision-making processes, moments of uncertainty or insight, and the specific system features they engaged with while shaping their story. Three researchers independently coded the qualitative data using a reflexive thematic analysis approach, resolving discrepancies through discussion and consensus-building. This mixed-method design allowed us to triangulate behavioral evidence, self-reported feedback, and interaction patterns to build a holistic understanding of writer's experience with StoryBlocks. In the next section we describe how we analyzed the qualitative and quantitative data and our findings from the user study.

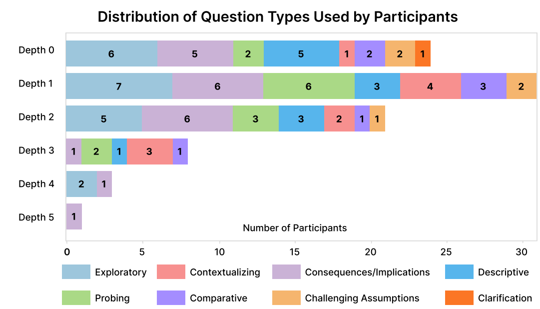
7 RESULTS

We recruited a total of 9 participants for our formative study (Section 3) and a different group of 21 participants for our evaluation study. Out of those 21 participants, 3 participants were unable to complete the study due to technical reasons. In this section, we present the results of the remaining 18 participants using StoryBlocks to write a data story using either of the three datasets. All 18 participants (16 male, 2 female) were aged 18–35 and had varied educational backgrounds - 10 held a high school diploma, 5 held a bachelor's degree, and 3 held a graduate degree. Most participants came from Engineering (7) and Science (7) with smaller representation from Humanities (2) and Social Sciences (1). Participants reported reading the news frequently (M = 3.61, SD = 0.78 on a 5-point scale) with Science (14), Politics (13), and Technology (12) as the most common interests. Participants were moderately familiar with interactive tools, such as Excel and Tableau (M = 3.6, SD = 0.99), and somewhat less familiar with programming toolkits, such as Python and R (M = 3.39, SD = 1.29) (Figure 5).

Participants preferred brainstorming strategies were jotting down questions before writing (12), visually clustering data and insights with software (9), identifying supporting data before deriving insights (7), and clustering on paper (5) (Figure 6). Participants reported reading data stories only once or twice a year on average (M = 2.39, SD = 1.24), and most had written more than one data story in the past (M = 2.78, SD = 2.07). Overall, they described themselves



(a) Brainstorm Methods



(b) Distribution of Question Types

Fig. 6. Participants' brainstorm methods when creating DDS (a), and the distribution of question types across depths of the hierarchical narrative graph (b). Each row in (b) corresponds to a depth level (Depth 0–5), and the colored segments represent the count of different question types within that depth.

as slightly familiar with generating story narratives from data (M = 2.39, SD = 1.09) and seldom engaged in explicit narrative planning tasks (M = 2.39, SD = 0.92). Participants spent an average of 59.20 minutes (SD = 7.97, max = 72.22, min = 45.5) completing the study session.

We analyzed both qualitative and quantitative data to understand how an inquiry-guided writing assistant shaped participants' data-driven storytelling practices. Specifically, we examined (1) how participants engaged with and produced questions during narrative design, (2) how inquiry-guided interactions influenced their narrative reasoning - particularly in sequencing, framing, and rhetorical development, and (3) how participants experienced the system's usability and creative support. We present our findings below accordingly.

7.1 Patterns of Inquiry in Participants' Questioning

To examine participants' narrative design reasoning process, we analyzed the questions they used in their stories using a qualitative coding methodology. Three researchers first independently conducted open coding of all questions, focusing on their style, purpose, and context, using categories derived from established *Socratic questioning* frameworks [27]. The initial codes were then consolidated and refined to produce a finalized codebook with clear inclusion criteria and definitions (Table 1). This codebook was applied to all participant-generated questions and inter-rater reliability was assessed on a subset of transcripts using Cohen's κ [104], yielding substantial agreement. We resolved remaining discrepancies through consensus meetings to ensure consistent interpretation. In total, we coded 88 questions that appeared in participants' final stories.

We observed that the majority of initial questions (at depth 0) were *Exploratory* (25%), *Descriptive* (21%), or *Consequences/Implications* (21%). However, subsequent questions (at depth 1) were slightly more diverse with *Exploratory* (22%), *Probing* (19%), *Consequences/Implications* (19%), and *Contextualizing* (13%) being the most common question types. At depth 2, questions were similarly concentrated in *Consequences/Implications* (29%) and *Exploratory* (24%). Beyond depth 3, the number of questions declined sharply (Figure 6); given the limited data, we did not analyze distributions at depths 4–6, as they do not reliably represent broader trends. This distribution of question types across depths reflects a dominant pattern, where exploratory sense-making gives way to deeper analytical probing. At depth 0, the dominance of *Exploratory* and *Descriptive* questions suggests that participants used questions to initially orient themselves and

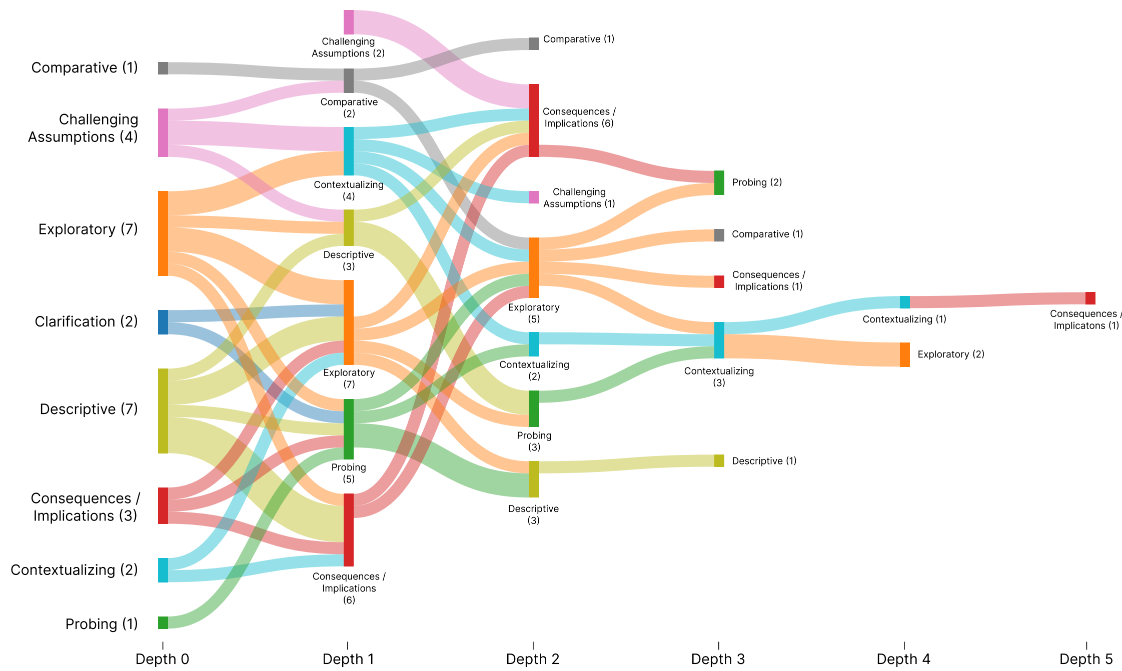


Fig. 7. Coding flow diagram of all participants' *Socratic question* types across story depth. Each block represents a coded question category (e.g., exploratory, descriptive, consequences/implications), and the flows illustrate how question types transitioned from one depth level to the next. Flow thickness corresponds to the frequency of each observed transition. Depth indicates the sequential level of questioning within a story, where depth 0 denotes the initial question and subsequent depths represent follow-up questions. The visualization shows that initial questions (depth 0) were dominated by exploratory, descriptive, and consequences/implications types, which then diversified into probing, contextualizing, and other categories at deeper levels. Beyond depth 3, the number of new questions declined substantially, reflecting limited continuation of deeper narrative chains.

establish a line of inquiry. By depth 1, the questions participants relied on became slightly more diverse, e.g., *Probing and Contextualizing*, suggesting that they were beginning to examine and challenging the evidence they had gathered. At depth 2, participants returned to *Consequences/Implications* and *Exploratory* questions, but these now carried greater nuance, indicating deeper reflection on their initial directions. While adopting an inquiry-guided approach helped writers systematically reflect during the early stages of narrative development (depths 0, 1 and 2), the sharp decline question usage at depth 3 may indicate cognitive limits in sustaining deeply nested inquiries without additional support. We elaborate on the implications of these findings in our discussion section.

We also examined block connectivity to understand how different question types connected to each other within participants' story graphs. We observed that participants who began with a *Descriptive* question often followed up with a *Consequences/Implications* question (Figure 7). Similarly, *Exploratory* questions were frequently followed by either another *Exploratory* question or a *Contextualizing* question. Additional patterns included *Challenging Assumptions* → *Contextualizing* and *Descriptive* → *Exploratory*. At depth 1, we often observed *Probing* → *Descriptive* and vice versa, while *Challenging Assumptions* → *Consequences/Implications* was similarly common. These patterns revealed how various question types shaped participant inquiry and their resulting narratives. For instance, frequent *Descriptive*

Question Type	Definition	Example
Challenging Assumptions	Question that investigates whether an event is a frequent occurrence.	Is U.S. gun violence a problem of gun control or a problem of increasing violence crime in general?
Comparative	Frames two or more entities against each other.	Considering both the New Democratic Party and the Liberal Party, which one serves as the main opposition party to the Conservative Party?
Consequences / Implications	The questions seek a call to action. The question asks how does the inference from dataset impact the outcomes. Implies next steps, actions, or solutions.	What strategies could Canadian transit authorities implement to attract riders back to public transit as they recover from the lingering impacts of the COVID-19 pandemic?
Descriptive	Relates to the descriptive stats (data) to seek evidence for a hypothesis.	What is the average age of transit riders in Canada, and how has this demographic changed over the period from 2017 to 2023 in relation to ridership trends?
Exploratory	Seeks open exploration without expecting one clear answer. Questions focus on hypothesizing how an event might unfold.	In what ways might regional economic conditions impact local businesses and their growth opportunities?
Probing	The question seeks exploration with a targeted event in context. It is different than an exploratory question as the context of the question is set in a prior observation.	What specific indicators should policymakers monitor to assess the likelihood of a full recovery in employment rates compared to pre-pandemic levels?
Clarification	Requests for additional statistical information on a specific event occurrence.	At which point in time did ticket sales for the transit systems begin a resurgence?
Contextualizing	Places the issue within a broader frame or background.	In what ways do the cultural attitudes towards gun rights in higher gun death states impact the effectiveness of gun control measures?

Table 1. Codebook used by researchers for coding different *Socratic question* types, including formal definitions and illustrative examples from real participant data.

→ *Consequences/Implications* transitions suggests that once participants established a factual understanding, their questions naturally guided them toward evaluating the significance or impact of those facts. Similarly, patterns such as *Exploratory* → *Exploratory* and *Exploratory* → *Contextualizing* suggest that the system encouraged the Socratic strategy of expanding the conceptual space before narrowing toward conclusions [15], and frequent pattern occurrences like *Challenging Assumptions* → *Contextualizing* indicate that once participants challenged an underlying premise, they often sought information to re-anchor their reasoning within broader contexts. Intuitively, when participants alternated between seeking detail and examining the meaning of those details, *Probing* questions were often followed by *Descriptive* questions, and vice versa. These connectivity patterns reveal that participants did not ask questions in isolation; rather, they constructed chained sequences resembling the structured reasoning of Socratic dialogue.

Our analysis also revealed that participants approached story narrative using distinct strategies depending on their types of questions. Some participants adopted a *question-centric* approach, framing their narratives primarily through successive, open-ended questions. For instance, they began with a broad question, such as *"What patterns exist in the data?"* and then layered follow-up questions, such as *"Why might this pattern occur?"* or *"What are the implications for different subgroups?"* to shape the storyline. Other participants adopted an *observation-centric* approach, emphasizing reasoning, interpretation, and the integration of data points to construct coherent arguments and concrete story structures. For instance, they began with a clear question, such as *"How does income vary across different demographic groups?"*, and rather than generating follow-up questions, they incorporated several observation blocks that answered their initial question. We also observed that several participants adapted a *visualization-centric* approach, depending on charts and visual representations to highlight patterns and enrich their narratives. For instance, participants often identified a trend in a visualization and then posed contextualizing or probing questions like *"What explains this spike?"* or *"How does this pattern change over time?"* to deepen their narrative. We demonstrate an example of each of these strategies in Figure 8. These varied approaches highlight the diversity of user strategies in data-driven storytelling and demonstrate the system's effectiveness in accommodating multiple approaches to narrative design. We describe how these different strategies and questioning patterns shaped participants' writing experience in the subsection below.

7.2 Inquiry-Guided Narratology and Design Experience

We transcribed and analyzed data from our think-aloud protocols and audio recordings to uncover participants' experience with an inquiry-guided approach during key tasks such as data exploration, narrative sequencing, rhetorical strategy development, and narrative framing. We conducted a qualitative thematic analysis using an iterative, bottom-up coding process. Three researchers independently performed open coding on all study data, then met to reconcile and refine the codes through discussion and comparison. The consolidated codes were then grouped into higher-level themes that captured recurring patterns in participant behaviors and reasoning. We present the salient themes that emerged from this analysis below.

7.2.1 Data exploration and analysis. Our system supported data exploration through natural language queries, enabling participants to interact with observations and visualizations without having to conduct independent exploratory data analysis. Participant think-aloud results revealed that data exploration in narrative design often unfolded as an open-ended dialogue with the dataset, frequently marked with uncertainty, curiosity, and cognitive overload. Most writers did not begin with a fixed story line, but instead explored the dataset – reading charts, relating existing questions, and scanning dataset columns to generate new inquiries. Participants viewed questions as anchors for these explorations, which sometimes guided their attention towards unexpected directions, as demonstrated when **P6** said, *"I like this question because it brings me into a certain state of mind. It makes me think about the labor market, which I hadn't considered."* However, for several participants, questioning introduced challenges during data exploration, as they struggled with divergent thinking (opening multiple branches of inquiry) and convergent structuring (extracting insights to form a coherent story point). For instance, **P19** mentioned, *"I definitely have an idea of where to take the story, but I'm lacking the data for that so far"*, in reference to their initial writing goal, while **P13** commented, *"...ideally*

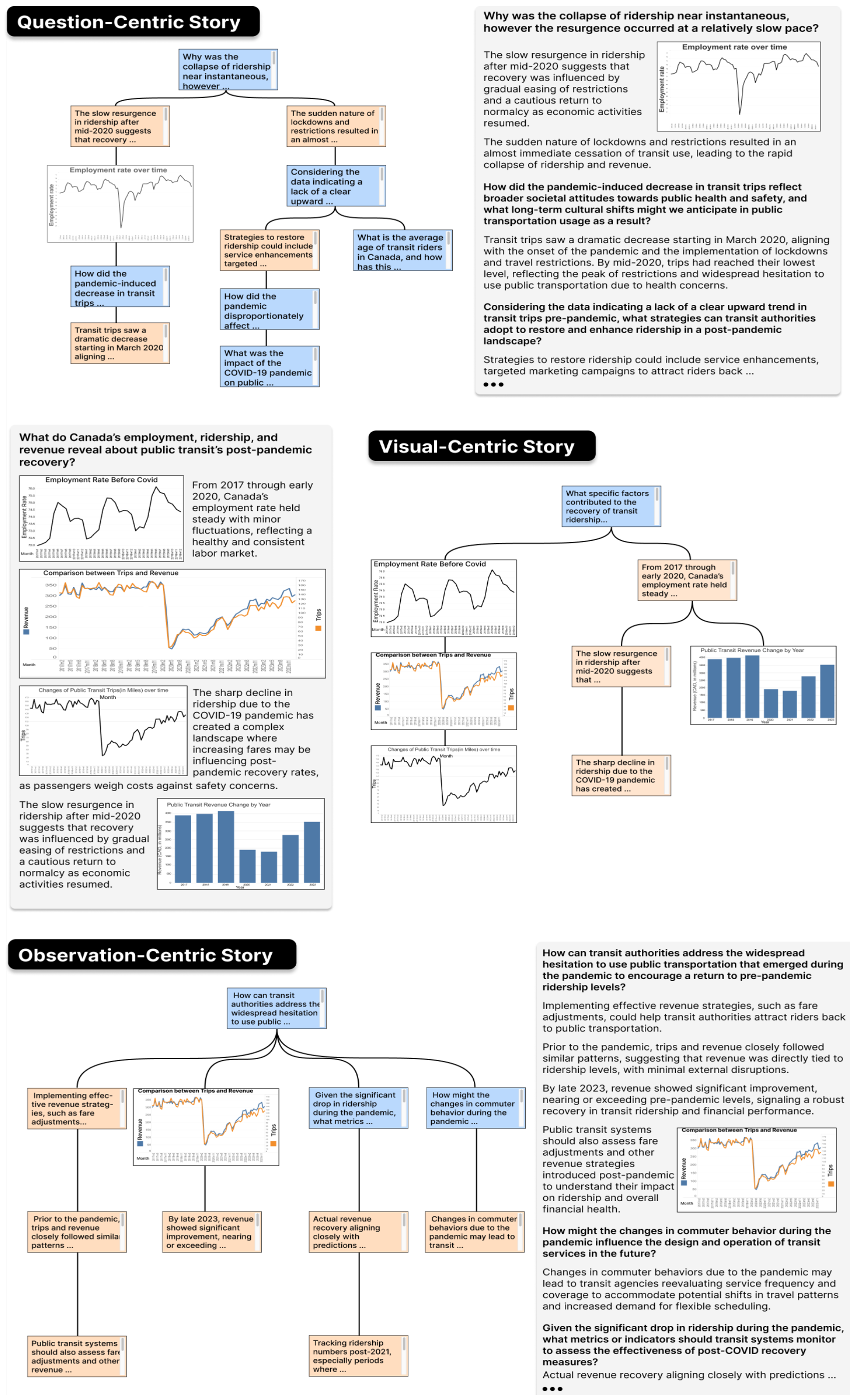


Fig. 8. The three story archetypes identified in our user studies—question-centric, observation-centric, and visual-centric—along with their linearized prototypes generated by our story renderer. All stories are taken from real user studies.

to answer this question, you would also want other data that's not gun violence data, like [violent] crime in general, but we don't have that." Some participants adopted a *data-centric approach*, allowing the data itself to shape their narrative rather than seeking evidence to support a presumed direction. As **P10** explained, *"I feel like the data I use should make the story, but I feel like sometimes I'm fighting against the data to create a narrative that I imagined."* This process surfaced scaling challenges, where some participants felt overwhelmed by excessive decision-making, incomplete or missing data, or determining which storylines to pursue. For instance, **P14** mentioned, *"I feel a bit overwhelmed by the choices and options I have. There's especially too much data to now be able to think– I'm not sure which direction to go in."*

7.2.2 Narrative Sequencing. Participants frequently noted that constructing event sequences through iterative questioning shaped how they conceptualized their own stories. They experimented with the order of data and context, with some preferring to introduce visuals early to capture attention, e.g., **P16**: *"I think it makes more sense to show the data first [with insights] and then explain the context later... I feel like this grabs the readers attention more"*. Others favored positioning explanatory elements first to orient readers (**P17**: *"I want the text before the graphics in the same way that you would put a title on a graph... The text gets you invested, then the graph supports that by explaining the actual data."*) While the ability to create several lines of inquiry was generally seen as useful for ideation, some participants found it challenging to connecting parallel lines of thought, e.g., **P10**: *"I'm not too sure how I can connect these two separate ideas [in reference to separate chains of thought]"*, **P21**: *"...I'm trying to find an insight that can connect these two [chains of thought]."* We anticipate this occurred because participants performed data extraction and narrative sequencing tasks simultaneously. One participant likened the sequencing process to puzzle solving, where insights could unify branches or shift narrative direction (**P19**: *"I feel like I'm putting a puzzle together."*) Simultaneously, participants were aware that overcrowding their story with too many branches could dilute its focus or alter its overall direction, e.g., **P21**: *"Adding too many branches, I think, would make the story too wide and not focused enough on my first question"*, **P19**: *"This question might be a good separate idea to run with my existing story. It'll definitely change how the story is perceived though."* Overall, their reflections suggest an ongoing negotiation between data extraction and narrative formulation to achieve a balanced story structure.

7.2.3 Rhetorical strategies. For many participants, story flow-the natural progression between ideas and insights in their narrative-was important for creating a persuasive rhetoric. **P17** commented, *"I think it's important to me for the story to flow well– Most articles I read follow a similar flow that I'm trying to copy"*, while **P10** mentioned, *"I don't love the flow of the story right now, but I think I can reorder it to get the final story to be how I want."* Interestingly, generated questions were seen as argumentative resources to constructing logical rhetoric, e.g., **P14**: *"I like this question because it seems to follow very logically from the graph I was just looking at [in reference to a generated question]"*, and served as potential pivots around which stories could be framed, e.g., **P21**: *"This question [one that was generated] seems to really be the crux of my story"*, **P17**: *"...these questions are definitely good, I just need to figure out how to build off of them to other data."* Participants frequently used questions to assess the strength of their rhetoric-whether it felt too obvious, too broad, or sufficiently challenging-while evaluating the story's persuasive capability. **P14** said, *"I feel like each question needs to flow right to make the following data easier to understand."* These strategies illustrate how participants employed both classical rhetorical principles, such as flow and audience consideration, with more exploratory tactics, leveraging questions as hooks or scaffolds to write engaging data stories.

7.2.4 Framing. Participants frequently sought to widen their scope and frame their story within a broader backdrop. They used questions to explore specific topics, building on anecdotal experience, e.g., **P16**: *"My university program* *is crime studies, so I'm naturally more interested in this topic"*, and to situate their story within a social context, e.g., **P21**: *"I'm curious about the stigma people have about riding on transit and how that affects the data."* This suggests that although participants sought to draw their own inductive conclusions from the inferences they made, it was challenging to do so without sufficient exploration, e.g., **P15**: *"The data in the [generated] responses is good, but it's not what I'm looking for here– I feel like I need another dataset to support this."* This line of thinking culminated in a conversation led by one participant about cherry-picking data and the susceptibilities of DDS workflows to external biases (**P17**: *"I think a big part of journalism in the media is obviously really biased– I just feel like I am cherry-picking my data to build the story that I want, rather than actually showing it."*) Concisely, the same participant said, *"The data isn't wrong, it's just picked to my liking."* These observations suggest that an inquiry-guided approach may unintentionally steer authors toward skewed or cherry-picked findings, highlighting the need for a more holistic approach that actively mitigates internal and external biases influencing the writer.

7.3 Perceptions of System Usability and Support

Our study design included pre- and post-survey questionnaires to capture user experience and relevant metrics. On a 5-point Likert scale, participants rated the software as moderately useful in supporting their data story planning ($M = 3.44$, $SD = 1.20$). The generated stories were perceived as moderately informative ($M = 3.5$, $SD = 1.15$), and coherent and clear ($M = 3.5$, $SD = 1.20$). Participants rated the AI-generated questions ($M = 3.89$, $SD = 0.83$) and insights ($M = 3.83$, $SD = 0.92$) as being slightly more effective than visualizations ($M = 3.39$, $SD = 1.33$) in supporting their writing and understanding. We summarize this post-study feedback on perceived system performance in Figure 5.

To understand participants' overall experience using the system, we define an *Overall Satisfaction* factor over six (6) user experience measures included in our survey - usefulness, informativeness, clarity, question support, insight support, and visualization support. We performed a regression analysis using PCA to evaluate whether factors like confidence in analytical tools, confidence in programming, tutorial helpfulness, and navigation difficulty (reverse-coded) impacted the overall user satisfaction (Table 2). The resulting model was significant ($F(4, 13) = 6.04$, $p = .006$, $R^2 = 0.65$) and suggested that navigation difficulty was the only significant predictor of satisfaction ($\beta = 1.23, p = .002$), while tutorial helpfulness showed a positive, but non-significant effect. Furthermore, we observed that participant confidence in analytical tools and programming did not significantly contribute to over all satisfaction.

We further explored whether system-generated questions and insights improved the perceived clarity of participants' data stories. The regression model evaluating participants' story clarity in relation to perceived question and insight effectiveness explained a relatively small portion of the variance in clarity (questions: $R^2 = 0.22$, adjusted $R^2 = 0.17$, $n = 18$; insights: $R^2 = 0.25$, adjusted $R^2 = 0.21$, $n = 18$). However, the coefficients for questions and insights were positive and statistically significant (questions: $\beta = 0.68, p = .049$, 95% CI = $[0.01, 1.35]$; insights: $\beta = 0.66, p = .033$, 95% CI = $[0.06, 1.25]$), indicating that participants who rated generated questions and insights as more effective also perceived the generated stories as clearer and more coherent (Table 3, Model 1 & 2). These findings suggest that both guiding questions and system-generated insights influenced participants' perceptions of story clarity and coherence.

Some participants noted the presence of minor inaccuracies and missing information in the generated stories ($M = 2.39$, $SD = 0.92$). We investigated whether these perceived inaccuracies in their stories affected their evaluation of the system's usefulness. A regression analysis on these ratings conclude that the relationship was not statistically significant ($R^2 = 0.11$, adjusted $R^2 = 0.05$, $n = 18$; Table 3, Model 3) and the coefficient for inaccuracies was negative, but non-significant ($\beta = -0.42, p = .185$, 95% CI = $[-1.08, 0.23]$), suggesting that participants generally found the system useful regardless of data inaccuracies.

Finally, we analyzed participants' interaction logs to identify the range of narrative design patterns employed. Most participants ($N = 14$) created and iterated over a single hierarchical graph, suggesting that many participants' preferred to refine and deepen one coherent structure rather than branch into alternative storylines. However, a smaller subset ($N = 3$) experimented with 2–3 graphs, with one (1) participant creating as many as seven (7), indicating that although the system enabled more exploratory, multi-path ideation, this approach was not the most commonly preferred strategy. Participants generated an average of 12 question blocks using the QA ($SD = 7.25, \max = 30, \min = 3$), and 18 observation blocks and 4 visualization blocks ($SD = 4.29, \max = 18, \min = 0$) using the IE, suggesting that the system may have encouraged inquisitive reasoning and hypothesis development. However, the large variance in question generation suggests that while some writers relied heavily on guided questioning for ideation, others preferred a more targeted or minimal approach at deeper levels. We observed that participants connected at least 2 observation/visualization blocks ($\max = 4, \min = 1$) with question blocks, suggesting that they generally grounded their inquiries in concrete evidence rather than posing questions in isolation. Surprisingly, the block enrichment tool was used sparingly ($M = 2.1,\ SD = 4.05, \min = 0, \max = 17,$), despite one participant using it 17 times. This suggests that most participants depended on their own drafting or generated content from the AI agents, and while questions prompted reflection, participants rarely modified them for their specific purposes. This observation has important implications on how AI over-reliance and under-reliance can impact DDS workflows.

7.4 Summary of Results

We observed that participants' use of questions in our inquiry-based approach shifted from broad, exploratory questions to analytical and reflective ones as the stories developed. However, question preference dropped sharply at deeper levels, suggesting limits to sustained Socratic-style inquiry. We found that certain question pairs frequently occurred in systematic sequences, shaping narrative structures that were question-, observation-, or visualization-centered.. Think-aloud analysis showed that inquiry-guided interactions supported data exploration, sequencing, framing, and rhetorical development, helping participants uncover unexpected insights and organize their ideas. However, they struggled to connect parallel lines of thought and manage bias. Despite these challenges, participants consistently valued questions as scaffolds for logical progression. Finally, system logs and survey responses indicated that participants found the tool useful regardless of prior skill, but often relied heavily on either their own writing or agent-generated content, rarely enriching blocks—highlighting risks of both over- and under-reliance on AI in data-driven storytelling.

8 DISCUSSION

Storyblocks is a novel attempt at synthesizing contemporary research in narrative theory, data-driven story telling, and AI-based agents into a singular, cohesive interface for data-driven narrative design. Through a systematic evaluation, we used Storyblocks to explore the structural archetypes exhibited across numerous data stories, and uncover questioning patterns between four different facets of narrative design. In this section, we articulate actionable design implications that inform the design space of inquiry-guided narrative design tools for data driven workflows.

8.1 Balancing Curiosity with Narrative Focus in Agent-Mediated Exploration

A key contribution of this work is demonstrating how *Socratic questioning* can help writers explore, ideate and create more objective and diverse stories. We embedded *Socratic questioning* within Storyblocks using an interaction paradigm that motivates inquiries through questions, and encourages writers to reflect on their interpretations of data. This inquiry-guided approach, combined with the inherit variability involved in AI-based agentic systems, naturally exposed writers to new ideas and perspectives in their data. This is especially relevant in data-driven journalism, as opposed to other less rigorous forms of journalism, e.g., opinion piece writing, as readers expect the available data to shape the story, not the writer's envisioned narrative to determine the selected data. However, the non-deterministic data exploratory process we introduced-where an external agent mediates between the writer and the data-also led writers to sometimes lose focus of the story, especially when bombarded with too many insights. This suggests the need to incorporate interaction techniques that can accommodate user's uncertainty and curiosity without overwhelming them. One way to approach this is to introduce *exploration maps* that document user's data exploration journey and appear on-demand alongside natural language queries [93]. These exploration maps can be extremely helpful when the *writers lead the Socratic-style inquiry*, as is the case with Storyblocks. Another approach may involve allowing the *AI agent to lead the inquiry* by enabling the question agent to prompt the writer to examine their reasoning at various times during writing. This involves embedding on-demand AI agents into the writing system to ask questions, e.g., "Why did you chose this story direction?", "Are your conclusions logical?", or "What is the overarching goal of your narrative?" Future research should continue exploring how to utilize *Socratic questioning* effectively and seamlessly integrate it as an interaction paradigm into AI systems .

8.2 Adapting to Writers' Shifting Modes of Inquiry

An important observation in our user studies was that participants alternated between data-driven and story-driven entry points, suggesting that narrative design support systems must be sufficiently adaptable to support both processes. Systems may incorporate a *story-first* mode, in which writers specify a potential question or argument and the system automatically suggests insights that could support or refute it. On the other hand, a *data-first* mode can help writers extract story logic entirely through data exploration. To help mitigate cognitive overload, inquiry-guided interactive tools may rank dataset variables based on relevance - when a user becomes overwhelmed, the system can recommend potential variables to write about. These design features would help narrative design tools mediate between divergent explorations and convergent focused stories, ensuring that richness in data does not hinder narrative creation.

Several participants focused on the story flow-akin to the narrative arc in storytelling-in which a narrative follows a systematic progression from setup to conflict and finally resolution [12]. However, participants often struggled to provide context and evidence for these stages of the narrative arc as it did not align with dataset capabilities. Future systems could help scaffold this process by presenting writers with questions and insights that are relevant to these stages and can be accomplished within the dataset. We also observed that participants used questions as anchors and built narratives that resembled the Toulmin model of argumentation [50], in which evidence and claims are connected together to form a persuasive argument. Tools that map questions and insights onto stages of the Toulmin model could help writers build stories that are both logically sound and rhetorically compelling. For instance, a question might function as a claim or as grounds in one instant, while an insight could serve as supporting evidence in another.

8.3 Supporting Socially Responsible AI-mediated Inquiry and Narrative Design

Data-driven journalism is becoming increasingly essential each year as our societies rely on large data and AI to understand the world. One of the advantages offered by data-driven journalism is its ability to mitigate bias and reduce faux journalism—a presently crucial endeavor—by strictly adhering to statistics [42]. Despite this, data-driven writing tools can still carry several forms of bias. Data story writers may assume that the underlying dataset is collected objectively–an assumption that is rarely true. Biases can arise from the individuals or organizations collecting and maintaining the dataset, errors in the equipment used to collect the data (e.g., faulty instruments), or sampling biases that disproportionally over-represent or under-represent certain demographics. This can lead writers to making misleading conclusions in their narratives and unintentionally propagating misinformation – the very problem data-driven journalism aims to address. A possible solution to this issue is verifying the provenance of the dataset and its creators, or contrasting results with similar datasets. However, this is not always feasible, especially for datasets that represent controversial or understudied experiments. Within Storyblocks, we intentionally selected a narrow scope of datasets that engaged with broad, often politically-polarized social issues to engage participants with their story. While the sourcing and validity of some of these datasets was exceptionally high, this could still have introduced biased reasoning or faulty outcomes of writers' stories. It is the responsibility of AI-supported narrative design assistants to support data provenance and implement measures that can mitigate writers' internal and external biases.

8.4 Designing for Transparency, Expressivity and Analytical Precision

Separate from data biases, language models also exhibit inherent bias due to their training data, despite engineers trying to mitigate and safeguard against this [38, 72]. This can especially amplify tensions when reporting on sensitive or politically-charged topics, where writers find themselves at a crossroads between reporting statistically meaningful information and protecting audiences from inappropriate or biased information. This was demonstrated in Storyblocks, where several model responses came close to overstepping this boundary with particularly divisive datasets, such as U.S. gun deaths. Because our inquiry-guided system introduced a non-deterministic data exploration paradigm, writers were naturally exposed to a broad range of interpretations. While this is valuable, it can also compromise analytical correctness. The solution to this problem is non-trivial and involves weighing arguments from both perspectives. One possible approach involves allowing users to control the amount of censorship and regulation prevalent in model responses through a multi-agentic approach. This would allow authors to gather several perspectives on particularly subversive topics before committing to a narrative [46]. Another approach involves using explainable AI (XAI) techniques that could help writers interrogate how and why certain responses were generated. Similar to the biases language models express through their training data, past human experiences bias the stories we tell. This can lead to writers cherry-picking data that supports their beliefs and projecting them onto the narrative they create. While Storyblocks encouraged exploring new directions through our mediator language model approach, the final decision is ultimately in the hands of writers. Nonetheless, we believe that our approach can at minimum offer greater depth of field to writers, reduce one-dimensional thinking, and transform writing from an isolated echo chamber of opinions into an open-minded exercise. Future work should focus on incorporating bias detection and claim verification tools for assessing social biases, such as DramatVis Personae [46].

9 LIMITATIONS AND FUTURE WORK

Our study has several limitations. First, the participant pool was relatively small and homogeneous, which offers limited insights into the broader narrative design practices in more diverse groups. Additionally, the study design duration was relatively short, and the system posed some learning curve due to its novelty. As a result, participants reported having limited time to fully develop their stories, and some relied on more straightforward or intuitive paths, rather than exploring unique pathways. Beyond these concerns, our system design also introduced some practical limitations. Although data visualization was introduced as a central element of the storytelling process, participants noted that some generated charts were unclear, inflexible, or overwhelming, which reduced their usefulness in narrative construction. Others expressed frustration with the abstract nature of data interaction, emphasizing the need for at least some direct engagement with the underlying dataset to create effective visualizations.

Building on these findings, in future we will focus on extending the scope and improving the usability of Storyblocks. Some approaches, such as expanding the available datasets and integrating improved mechanisms for incorporating contextual data, would help writers construct richer, more nuanced narratives. Enhancing the context provided to models, refining chart generation capabilities, and experimenting with prompting methods could yield more accurate and flexible agent responses. We also envision interfaces that allow writers to modulate the response contexts in real-time and offer greater control over how agents mediates between statistical detail and narrative accessibility. Most importantly, future directions should preserve Storyblocks' identity as a narrative design tool, rather than a fully generative writing assistant, to support writers' agency. Finally, studies with journalists and longer-term engagements will be critical to validating our results assessing the system's relevance for professional story writing.

10 CONCLUSION

We proposed an approach to data-driven storytelling that builds on narrative theory and recent advancements in agentic AI to improve narrative design for DDS workflow. By refining our system requirements through a formative study, we informed the design of Storyblocks, a DDS tool that incorporates a multi-agentic architecture and non-linear thinking paradigm to support narrative design. Studies with non-expert writers revealed diverse narrative design strategies, as well as variations in their reliance on AI agents when crafting data-driven stories. We also identified recurring themes in their internal narratology and how these shaped the narrative design process. We encourage future research to build upon Storyblocks as a sandbox environment for exploring the complex challenges of narrative design in data-driven storytelling.

References

[1] H. Porter Abbott. 2008. *The Cambridge Introduction to Narrative* (2 ed.). Cambridge University Press, Cambridge.

[2] Dhruv Agarwal, Mor Naaman, and Aditya Vashistha. 2025. AI Suggestions Homogenize Writing Toward Western Styles and Diminish Cultural Nuances. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 1117, 21 pages. doi:10.1145/3706598.3713564

[3] Sangzin Ahn. 2024. Data science through natural language with ChatGPT's Code Interpreter. *Translational and Clinical Pharmacology* 32, 2 (2024), 73.

[4] Morteza Asgari and Thomas Hurtut. 2024. A Design Language for Prototyping and Storyboarding Data-Driven Stories. *Applied Sciences* 14, 4 (2024), 25 pages. doi:10.3390/app14041387

[5] Benjamin Bach, Moritz Stefaner, Jeremy Boy, Steven Drucker, Lyn Bartram, Jo Wood, Paolo Ciuccarelli, Yuri Engelhardt, Ulrike Koeppen, and Barbara Tversky. 2018. Narrative design patterns for data-driven storytelling. In *Data-driven storytelling*. AK Peters/CRC Press, New York, 107–133.

[6] Carl Bereiter and Marlene Scardamalia. 2013. *The psychology of written composition*. Routledge, New York.

[7] Merim Bilalić, Peter McLeod, and Fernand Gobet. 2008. Inflexibility of experts—Reality or myth? Quantifying the Einstellung effect in chess masters. *Cognitive Psychology* 56, 2 (2008), 73–102. doi:10.1016/j.cogpsych.2007.02.001

[8] Merim Bilalić, Peter McLeod, and Fernand Gobet. 2008. Why good thoughts block better ones: The mechanism of the pernicious Einstellung (set) effect. *Cognition* 108, 3 (2008), 652–661. doi:10.1016/j.cognition.2008.05.005

[9] John Bohannon. 2021. Deep Talk: Using a machine to help humans be more human. https://primer.ai/research/deep-talk-using-a-machine-to-help-humans-be-more-human/. Accessed: 2025-08-[today's date].

[10] Mariana Borges, Claiton Marques Correa, and Milene Selbach Silveira. 2022. Fundamental elements and characteristics for telling stories using data. *Journal on Interactive Systems* 13, 1 (Jun. 2022), 77–86. doi:10.5753/jis.2022.2330

[11] Mike Bostock. 2012. D3.js - Data-Driven Documents. http://d3js.org/

[12] Ryan L. Boyd, Kate G. Blackburn, and James W. Pennebaker. 2020. The narrative arc: Revealing core narrative structures through text analysis. *Science Advances* 6, 32 (2020), eaba2196. doi:10.1126/sciadv.aba2196 arXiv:https://www.science.org/doi/pdf/10.1126/sciadv.aba2196

[13] Jerome Bruner. 1991. The narrative construction of reality. *Critical inquiry* 18, 1 (1991), 1–21.

[14] Linda Joan Carey, Linda S. Flower, J. R. Hayes, Karen A. Schriver, and Christina Haas. 1989. *Differences in Writers' Initial Task Representations*. Technical Report. Center for the Study of Writing. https://apps.dtic.mil/sti/tr/pdf/ADA210433.pdf

[15] Joyce Armstrong Carroll and Edward E. Wilson. 2014. *The Critical Writer: Inquiry and the Writing Process*. Libraries Unlimited, an imprint of ABC-CLIO, Santa Barbara, California.

[16] Elin Carstensdottir, Erica Kleinman, and Magy Seif El-Nasr. 2019. Player interaction in narrative games: structure and narrative progression mechanics. In *Proceedings of the 14th International Conference on the Foundations of Digital Games* (San Luis Obispo, California, USA) *(FDG '19)*. Association for Computing Machinery, New York, NY, USA, Article 23, 9 pages. doi:10.1145/3337722.3337730

[17] L. B. Cebik. 1986. Understanding Narrative Theory. *History and Theory* 25, 4 (1986), 58–81. http://www.jstor.org/stable/2505132

[18] John Joon Young Chung, Melissa Roemmele, and Max Kreminski. 2025. Toyteller: AI-powered Visual Storytelling Through Toy-Playing with Character Symbols. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 331, 23 pages. doi:10.1145/3706598.3713435

[19] Elizabeth Clark, Anne Spencer Ross, Chenhao Tan, Yangfeng Ji, and Noah A. Smith. 2018. Creative Writing with a Machine in the Loop: Case Studies on Slogans and Stories. In *Proceedings of the 23rd International Conference on Intelligent User Interfaces* (Tokyo, Japan) *(IUI '18)*. Association for Computing Machinery, New York, NY, USA, 329–340. doi:10.1145/3172944.3172983

[20] Matthew Conlen, Jeffrey Heer, Hillary Mushkin, and Scott Davidoff. 2023. Cinematic Techniques in Narrative Visualization. arXiv:2301.03109 [cs.HC] https://arxiv.org/abs/2301.03109

[21] Matthew Conlen, Megan Vo, Alan Tan, and Jeffrey Heer. 2021. Idyll Studio: A Structured Editor for Authoring Interactive & Data-Driven Articles. In *The 34th Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) *(UIST '21)*. Association for Computing Machinery, New York, NY, USA, 1–12. doi:10.1145/3472749.3474731

[22] Keshav Dasu, Yun-Hsin Kuo, and Kwan-Liu Ma. 2024. Character-Oriented Design for Visual Data Storytelling. *IEEE Transactions on Visualization and Computer Graphics* 30, 1 (2024), 98–108. doi:10.1109/TVCG.2023.3326578

[23] Michele D Dickey. 2011. Murder on Grimm Isle: The impact of game narrative design in an educational game-based learning environment. *British journal of educational technology* 42, 3 (2011), 456–469.

[24] Anil R. Doshi and Oliver P. Hauser. 2024. Generative AI enhances individual creativity but reduces the collective diversity of novel content. *Science Advances* 10, 28 (2024), eadn5290. doi:10.1126/sciadv.adn5290 arXiv:https://www.science.org/doi/pdf/10.1126/sciadv.adn5290

[25] Anil R. Doshi and Oliver P. Hauser. 2024. Generative AI enhances individual creativity but reduces the collective diversity of novel content. *Science Advances* 10, 28 (2024), eadn5290. doi:10.1126/sciadv.adn5290 arXiv:https://www.science.org/doi/pdf/10.1126/sciadv.adn5290

[26] Longxu Dou, Guanghui Qin, Jinpeng Wang, Jin-Ge Yao, and Chin-Yew Lin. 2018. Data2Text Studio: Automated Text Generation from Structured Data. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, Eduardo Blanco and Wei Lu (Eds.). Association for Computational Linguistics, Brussels, Belgium, 13–18. doi:10.18653/v1/D18-2003

[27] Linda Elder and Richard Paul. 1998. The role of Socratic questioning in thinking, teaching, and learning. *The Clearing House* 71, 5 (1998), 297–301.

[28] Rachelyn Farrell, Scott Robertson, and Stephen G. Ware. 2016. Asking Hypothetical Questions About Stories Using QUEST. In *Interactive Storytelling*, Frank Nack and Andrew S. Gordon (Eds.). Springer International Publishing, Cham, 136–146.

[29] K. J. Kevin Feng, Kevin Pu, Matt Latzke, Tal August, Pao Siangliulue, Jonathan Bragg, Daniel S. Weld, Amy X. Zhang, and Joseph Chee Chang. 2025. Cocoa: Co-Planning and Co-Execution with AI Agents. arXiv:2412.10999 [cs.HC] https://arxiv.org/abs/2412.10999

[30] Linda Flower and John R. Hayes. 1981. A Cognitive Process Theory of Writing. *College Composition and Communication* 32, 4 (1981), 365–387. http://www.jstor.org/stable/356600

[31] Jacob Garbe, Max Kreminski, Ben Samuel, Noah Wardrip-Fruin, and Michael Mateas. 2019. StoryAssembler: an engine for generating dynamic choice-driven narratives. In *Proceedings of the 14th International Conference on the Foundations of Digital Games* (San Luis Obispo, California, USA) *(FDG '19)*. Association for Computing Machinery, New York, NY, USA, Article 24, 10 pages. doi:10.1145/3337722.3337732

[32] Sneha Gathani, Anamaria Crisan, Vidya Setlur, and Arjun Srinivasan. 2024. Groot: A System for Editing and Configuring Automated Data Insights . In *2024 IEEE Visualization and Visual Analytics (VIS)*. IEEE Computer Society, Los Alamitos, CA, USA, 36–40. doi:10.1109/VIS55277.2024.00015

[33] Michael Gerlich. 2025. AI Tools in Society: Impacts on Cognitive Offloading and the Future of Critical Thinking. *Societies* 15, 1 (2025), 28 pages. doi:10.3390/soc15010006

[34] Katy Gero, Alex Calderwood, Charlotte Li, and Lydia Chilton. 2022. A Design Space for Writing Support Tools Using a Cognitive Process Model of Writing. In *Proceedings of the First Workshop on Intelligent and Interactive Writing Assistants (In2Writing 2022)*, Ting-Hao 'Kenneth' Huang, Vipul Raheja, Dongyeop Kang, John Joon Young Chung, Daniel Gissin, Mina Lee, and Katy Ilonka Gero (Eds.). Association for Computational Linguistics, Dublin, Ireland, 11–24. doi:10.18653/v1/2022.in2writing-1.2

[35] Katy Ilonka Gero and Lydia B. Chilton. 2019. How a Stylistic, Machine-Generated Thesaurus Impacts a Writer's Process. In *Proceedings of the 2019 Conference on Creativity and Cognition* (San Diego, CA, USA) *(Creativity and Cognition '19)*. Association for Computing Machinery, New York, NY, USA, 597–603. doi:10.1145/3325480.3326573

[36] Katy Ilonka Gero and Lydia B. Chilton. 2019. Metaphoria: An Algorithmic Companion for Metaphor Creation. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. doi:10.1145/3290605.3300526

[37] P. Gervás, E. Concepción, C. León, G. Méndez, and P. Delatorre. 2019. The long path to narrative generation. *IBM Journal of Research and Development* 63, 1 (2019), 8:1–8:10. doi:10.1147/JRD.2019.2896157

[38] Michael Gira, Ruisu Zhang, and Kangwook Lee. 2022. Debiasing Pre-Trained Language Models via Efficient Fine-Tuning. In *Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion*, Bharathi Raja Chakravarthi, B Bharathi, John P McCrae, Manel Zarrouk, Kalika Bali, and Paul Buitelaar (Eds.). Association for Computational Linguistics, Dublin, Ireland, 59–69. doi:10.18653/v1/2022.ltedi-1.8

[39] Arthur C Graesser, William Baggett, and Kent Williams. 1996. Question-driven explanatory reasoning. *Applied Cognitive Psychology* 10, 7 (1996), 17–31.

[40] John Gramlich. 2025. What the data says about gun deaths in the U.S. https://www.pewresearch.org/short-reads/2025/03/05/what-the-data-says-about-gun-deaths-in-the-us/

[41] Andrea Guarneri, Laura A. Ripamonti, Francesco Tissoni, Marco Trubian, Dario Maggiorini, and Davide Gadia. 2017. GHOST: a GHOst STory-writer. In *Proceedings of the 12th Biannual Conference on Italian SIGCHI Chapter* (Cagliari, Italy) *(CHItaly '17)*. Association for Computing Machinery, New York, NY, USA, Article 24, 9 pages. doi:10.1145/3125571.3125580

[42] Mahmoud-Mohamed-Abdel Haleem and Hagar Talaat-Alnajjar. 2025. The Power of Data Journalism: The Effects of Data-Driven News Reports in Correcting Climate Change Misinformation. *Communication & Society* 38, 1 (Jun. 2025), 263–280. doi:10.15581/003.38.1.020

[43] Shan Hao, Zezhong Wang, Benjamin Bach, and Larissa Pschetz. 2024. Design Patterns for Data-Driven News Articles. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 231, 16 pages. doi:10.1145/3613904.3641916

[44] Gary M. Hardee. 2016. Immersive Journalism in VR: Four Theoretical Domains for Researching a Narrative Design Framework. In *Virtual, Augmented and Mixed Reality*, Stephanie Lackey and Randall Shumaker (Eds.). Springer International Publishing, Cham, 679–690.

[45] David Herman. 2002. *Story Logic: Problems and Possibilities of Narrative*. University of Nebraska Press, Lincoln, Nebraska.

[46] Md Naimul Hoque, Bhavya Ghai, and Niklas Elmqvist. 2022. DramatVis Personae: Visual Text Analytics for Identifying Social Biases in Creative Writing. In *Proceedings of the 2022 ACM Designing Interactive Systems Conference* (Virtual Event, Australia) *(DIS '22)*. Association for Computing Machinery, New York, NY, USA, 1260–1276. doi:10.1145/3532106.3533526

[47] Chieh-Yang Huang, Shih-Hong Huang, and Ting-Hao Kenneth Huang. 2020. Heteroglossia: In-Situ Story Ideation with the Crowd. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–12. doi:10.1145/3313831.3376715

[48] Julie S. Hui, Darren Gergle, and Elizabeth M. Gerber. 2018. IntroAssist: A Tool to Support Writing Introductory Help Requests. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3173574.3173596

[49] Mohammed Saidul Islam, Md Tahmid Rahman Laskar, Md Rizwan Parvez, Enamul Hoque, and Shafiq Joty. 2024. DataNarrative: Automated Data-Driven Storytelling with Visualizations and Texts. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, Yaser Al-Onaizan, Mohit Bansal, and Yun-Nung Chen (Eds.). Association for Computational Linguistics, Miami, Florida, USA, 19253–19286. doi:10.18653/v1/2024.emnlp-main.1073

[50] Joan Karbach. 1987. Using Toulmin's model of argumentation. *Journal of Teaching Writing* 6, 1 (1987), 81–92.

[51] Brian Felipe Keith Norambuena and Tanushree Mitra. 2021. Narrative maps: An algorithmic approach to represent and extract information narratives. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW3 (2021), 1–33.

[52] Hartmut Koenitz, Teun Dubbelman, and Christian Roth. 2019. An Educational Program in Interactive Narrative Design. In *Interactive Storytelling*, Rogelio E. Cardona-Rivera, Anne Sullivan, and R. Michael Young (Eds.). Springer International Publishing, Cham, 22–25.

[53] Robert Kosara. 2017. An argument structure for data stories. In *Proceedings of the Eurographics/IEEE VGTC Conference on Visualization: Short Papers* (Barcelona, Spain) *(EuroVis '17)*. Eurographics Association, Goslar, DEU, 31–35. doi:10.2312/eurovisshort.20171129

[54] Max Kreminski. 2024. The Dearth of the Author in AI-Supported Writing. In *Proceedings of the Third Workshop on Intelligent and Interactive Writing Assistants* (Honolulu, HI, USA) *(In2Writing '24)*. Association for Computing Machinery, New York, NY, USA, 48–50. doi:10.1145/3690712.3690725

[55] Max Kreminski and Chris Martens. 2022. Unmet Creativity Support Needs in Computationally Supported Creative Writing. In *Proceedings of the First Workshop on Intelligent and Interactive Writing Assistants (In2Writing 2022)*, Ting-Hao 'Kenneth' Huang, Vipul Raheja, Dongyeop Kang, John Joon Young Chung, Daniel Gissin, Mina Lee, and Katy Ilonka Gero (Eds.). Association for Computational Linguistics, Dublin, Ireland, 74–82. doi:10.18653/v1/2022.in2writing-1.11

[56] Carol C. Kuhlthau, Leslie K. Maniotes, and Ann K. Caspari. 2012. *Guided Inquiry Design®: A Framework for Inquiry in Your School*. Libraries Unlimited, an imprint of ABC-CLIO, Santa Barbara, California.

[57] Woro Kusmaryani. 2021. The Effect of Socratics Questioning Method in Improving Students' Speaking Skill and Critical Thinking in English as a Foreign Language Learning. In *Proceedings of the 2nd International Conference on Innovation in Education and Pedagogy (ICIEP 2020)*. Atlantis Press, Paris, France, 289–295. doi:10.2991/assehr.k.211219.051

[58] Sébastien Lallé, Dereck Toker, and Cristina Conati. 2021. Gaze-Driven Adaptive Interventions for Magazine-Style Narrative Visualizations. *IEEE Transactions on Visualization and Computer Graphics* 27, 6 (2021), 2941–2952. doi:10.1109/TVCG.2019.2958540

[59] Siti Latifah, Abdul Muth'im, and Nasrullah Nasrullah. 2024. The Use of QuillBot in Academic Writing: A Systematic Literature Review. *Journey: Journal of English Language and Pedagogy* 7, 1 (Mar. 2024), 110–121. doi:10.33503/journey.v7i1.872

[60] Bongshin Lee, Nathalie Henry Riche, Petra Isenberg, and Sheelagh Carpendale. 2015. More Than Telling a Story: Transforming Data into Visually Shared Stories. *IEEE Computer Graphics and Applications* 35, 5 (2015), 84–90. doi:10.1109/MCG.2015.99

[61] Stephanie Levitz. 2023. Justin Trudeau's Liberals head into summer with lowest share of committed voters since they were elected in 2015. https://www.thestar.com/politics/federal/justin-trudeau-s-liberals-see-lowest-approval-rating-since-they-formed-government-poll-shows/article_d7b8db95-23df-531c-ac4f-a089947cc916.html

[62] Haotian Li, Yun Wang, and Huamin Qu. 2024. Where Are We So Far? Understanding Data Storytelling Tools from the Perspective of Human-AI Collaboration. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 845, 19 pages. doi:10.1145/3613904.3642726

[63] Ruth Li. 2024. A "Dance of storytelling": Dissonances between substance and style in collaborative storytelling with AI. *Computers and Composition* 71 (2024), 102825. doi:10.1016/j.compcom.2024.102825

[64] Zhuoyan Li, Chen Liang, Jing Peng, and Ming Yin. 2024. The Value, Benefits, and Concerns of Generative AI-Powered Assistance in Writing. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 1048, 25 pages. doi:10.1145/3613904.3642625

[65] Oscar Daniel Lucas-Moreira and Jenniffer Núñez-Díaz. 2025. Narratives in the Age of AI: Reflections on Literature and Communication. *YUYAY: Estrategias, Metodologías & Didácticas Educativas* 4, 2 (2025), 77–93.

[66] Pingchuan Ma, Rui Ding, Shuai Wang, Shi Han, and Dongmei Zhang. 2023. Demonstration of InsightPilot: An LLM-Empowered Automated Data Exploration System. arXiv:2304.00477 [cs.DB] https://arxiv.org/abs/2304.00477

[67] Andre Mayer. 2024. The dirty secret of the housing crisis? Homeowners like high prices. https://www.cbc.ca/news/business/housing-prices-affordability-real-estate-1.7170775

[68] Dan P McAdams, Ruthellen Ed Josselson, and Amia Ed Lieblich. 2006. *Identity and story: Creating self in narrative.* American Psychological Association, Washington, DC.

[69] S. McKenna, N. Henry Riche, B. Lee, J. Boy, and M. Meyer. 2017. Visual Narrative Flow: Exploring Factors Shaping Data Visualization Story Reading Experiences. *Comput. Graph. Forum* 36, 3 (June 2017), 377–387. doi:10.1111/cgf.13195

[70] Sködt McNalty. 2024. This kind of public transportation is growing fast. https://www.cbc.ca/news/science/what-on-earth-bike-sharing-1.7260301

[71] Norul Hidayah Mamat Muhammad and Jamilah Amin Niazi. 2022. Socratic questioning: a philosophical approach in developing critical thinking skills. *Al-Hikmah: International Journal of Islamic Studies and Human Sciences* 5, 4 (2022), 143–161.

[72] Roberto Navigli, Simone Conia, and Björn Ross. 2023. Biases in Large Language Models: Origins, Inventory, and Discussion. *J. Data and Information Quality* 15, 2, Article 10 (June 2023), 21 pages. doi:10.1145/3597307

[73] Farhad Nooralahzadeh, Yi Zhang, Jonathan Furst, and Kurt Stockinger. 2024. Explainable Multi-Modal Data Exploration in Natural Language via LLM Agent. arXiv:2412.18428 [cs.AI] https://arxiv.org/abs/2412.18428

[74] Lisa Oberascher, Christian Ploder, Johannes Spiess, Reinhard Bernsteiner, and Willemijn Van Kooten. 2023. Data storytelling to communicate big data internally-a guide for practical usage. *European Journal of Management Issues* 31, 1 (2023), 27–39.

[75] Elinor Ochs and Lisa Capps. 1996. Narrating the self. *Annual review of anthropology* 25, 1 (1996), 19–43.

[76] Judith S. Olson, Dakuo Wang, Gary M. Olson, and Jingwen Zhang. 2017. How People Write Together Now: Beginning the Investigation with Advanced Undergraduates in a Project Course. *ACM Trans. Comput.-Hum. Interact.* 24, 1, Article 4 (March 2017), 40 pages. doi:10.1145/3038919

[77] OpenAI. 2024. GPT-4o System Card. arXiv:2410.21276 [cs.CL] https://arxiv.org/abs/2410.21276

[78] OpenAI. 2025. *Introducing Study Mode in ChatGPT*. OpenAI. https://openai.com/index/chatgpt-study-mode/ Accessed: 2025-08-[today's date].

[79] Hiroyuki Osone, Jun-Li Lu, and Yoichi Ochiai. 2021. BunCho: AI Supported Story Co-Creation via Unsupervised Multitask Learning to Increase Writers' Creativity in Japanese. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI EA '21)*. Association for Computing Machinery, New York, NY, USA, Article 19, 10 pages. doi:10.1145/3411763.3450391

[80] Soya Park and Chinmay Kulkarni. 2023. Thinking assistants: Llm-based conversational assistants that help users think by asking rather than answering. 29 pages.

[81] Molly Patterson and Kristen Renwick Monroe. 1998. Narrative in political science. *Annual review of political science* 1, 1 (1998), 315–331.

[82] Zhenhui Peng, Qingyu Guo, Ka Wing Tsang, and Xiaojuan Ma. 2020. Exploring the Effects of Technological Writing Assistance for Support Providers in Online Mental Health Community. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–15. doi:10.1145/3313831.3376695

[83] Hardiyanti Pratiwi, Suherman, Hasruddin, and Muhammad Ridha. 2025. Between Shortcut and Ethics: Navigating the Use of Artificial Intelligence in Academic Writing Among Indonesian Doctoral Students. *European Journal of Education* 60, 2 (2025), e70083.

[84] Olaf Resch and Aglika Yankova. 2019. Open knowledge interface: a digital assistant to support students in writing academic assignments. In *Proceedings of the 1st ACM SIGSOFT International Workshop on Education through Advanced Software Engineering and Artificial Intelligence* (Tallinn, Estonia) *(EASEAI 2019)*. Association for Computing Machinery, New York, NY, USA, 13–16. doi:10.1145/3340435.3342723

[85] Jill Walker Rettberg and Hermann Wigers. 2025. AI-generated stories favour stability over change: homogeneity and cultural stereotyping in narratives generated by gpt-4o-mini. *ArXiv* abs/2507.22445 (2025), 17 pages. https://api.semanticscholar.org/CorpusID:280391425

[86] M.O. Riedl and R.M. Young. 2006. From linear story generation to branching story graphs. *IEEE Computer Graphics and Applications* 26, 3 (2006), 23–31. doi:10.1109/MCG.2006.56

[87] Melissa Roemmele and Andrew S. Gordon. 2018. Automated Assistance for Creative Writing with an RNN Language Model. In *Companion Proceedings of the 23rd International Conference on Intelligent User Interfaces* (Tokyo, Japan) *(IUI '18 Companion)*. Association for Computing Machinery, New York, NY, USA, Article 21, 2 pages. doi:10.1145/3180308.3180329

[88] Daniel M. Russell, Mark J. Stefik, Peter Pirolli, and Stuart K. Card. 1993. The cost structure of sensemaking. In *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems* (Amsterdam, The Netherlands) *(CHI '93)*. Association for Computing Machinery, New York, NY, USA, 269–276. doi:10.1145/169059.169209

[89] Cereneo Santiago Jr, Steve Embang, Ricky Acanto, Kem Warren Ambojia, Maico Demi Aperocho, Benedicto Balilo Jr, Erwin Cahapin, Marjohn Conlu, Samson Lausa, Ester Laput, Beverly Malabag, Jerby Paderes, and Joan Romasanta. 2023. Utilization of Writing Assistance Tools in Research in Selected Higher Learning Institutions in the Philippines: A Text Mining Analysis. *International Journal of Learning, Teaching and Educational Research* 22 (11 2023), 259–284.

[90] Kira Schacht. 2024. How megacities are tackling air pollution. https://www.dw.com/en/4-ways-megacities-are-tackling-air-pollution/a-67645710

[91] Edward Segel and Jeffrey Heer. 2010. Narrative Visualization: Telling Stories with Data. *IEEE Transactions on Visualization and Computer Graphics* 16, 6 (2010), 1139–1148. doi:10.1109/TVCG.2010.179

[92] Vidya Setlur, Sarah E. Battersby, Melanie Tory, Rich Gossweiler, and Angel X. Chang. 2016. Eviza: A Natural Language Interface for Visual Analysis. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology* (Tokyo, Japan) *(UIST '16)*. Association for Computing Machinery, New York, NY, USA, 365–377. doi:10.1145/2984511.2984588

[93] Vidya Setlur, Sarah E. Battersby, Melanie Tory, Rich Gossweiler, and Angel X. Chang. 2016. Eviza: A Natural Language Interface for Visual Analysis. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology* (Tokyo, Japan) *(UIST '16)*. Association for Computing Machinery, New York, NY, USA, 365–377. doi:10.1145/2984511.2984588

[94] Momin N Siddiqui, Roy D Pea, and Hari Subramonyam. 2025. Script&Shift: A Layered Interface Paradigm for Integrating Content Development and Rhetorical Strategy with LLM Writing Assistants. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 532, 19 pages. doi:10.1145/3706598.3714119

[95] Macarena Silva and Kate Cain. 2019. The use of questions to scaffold narrative coherence and cohesion. *Journal of Research in Reading* 42, 1 (2019), 1–17.

[96] Inhwa Song, SoHyun Park, Sachin R Pendse, Jessica Lee Schleider, Munmun De Choudhury, and Young-Ho Kim. 2025. ExploreSelf: Fostering User-driven Exploration and Reflection on Personal Challenges with Adaptive Guidance by Large Language Models. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 306, 22 pages. doi:10.1145/3706598.3713883

[97] Charles D. Stolper, Bongshin Lee, Nathalie Henry Riche, and John Stasko. 2016. *Emerging and Recurring Data-Driven Storytelling Techniques: Analysis of a Curated Collection of Recent Stories*. Technical Report MSR-TR-2016-14. Microsoft Research. https://www.microsoft.com/en-us/research/publication/emerging-and-recurring-data-driven-storytelling-techniques-analysis-of-a-curated-collection-of-recent-stories/

[98] Michelle Scalise Sugiyama. 2001. Narrative theory and function: Why evolution matters. *Philosophy and literature* 25, 2 (2001), 233–250.

[99] Thiemo Wambsganss, Christina Niklaus, Matthias Cetto, Matthias Söllner, Siegfried Handschuh, and Jan Marco Leimeister. 2020. AL: An Adaptive Learning Support System for Argumentation Skills. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14. doi:10.1145/3313831.3376732

[100] Huichen Will Wang, Larry Birnbaum, and Vidya Setlur. 2025. Jupybara: Operationalizing a Design Space for Actionable Data Analysis and Storytelling with LLMs. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 1005, 24 pages. doi:10.1145/3706598.3713913

[101] Zezhong Wang, Shunming Wang, Matteo Farinella, Dave Murray-Rust, Nathalie Henry Riche, and Benjamin Bach. 2019. Comparing Effectiveness and Engagement of Data Comics and Infographics. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. doi:10.1145/3290605.3300483

[102] Azmine Toushik Wasi, Raima Islam, and Mst Rafia Islam. 2024. Ink and Individuality: Crafting a Personalised Narrative in the Age of LLMs. In *Proceedings of the Third Workshop on Intelligent and Interactive Writing Assistants* (Honolulu, HI, USA) *(In2Writing '24)*. Association for Computing Machinery, New York, NY, USA, 43–47. doi:10.1145/3690712.3690724

[103] Emily Wenger and Yoed N. Kenett. 2025. We're Different, We're the Same: Creative Homogeneity Across LLMs. *ArXiv* abs/2501.19361 (2025), 18 pages. https://api.semanticscholar.org/CorpusID:276079902

[104] Wikipedia contributors. 2025. Cohen's kappa — Wikipedia, The Free Encyclopedia. https://en.wikipedia.org/wiki/Cohen%27s_kappa [Online; accessed 11-September-2025].

[105] Matt Willis and Mohammad Hossein Jarrahi. 2019. Automating Documentation: A Critical Perspective into the Role of Artificial Intelligence in Clinical Documentation. In *Information in Contemporary Society*, Natalie Greene Taylor, Caitlin Christian-Lamb, Michelle H. Martin, and Bonnie Nardi (Eds.). Springer International Publishing, Cham, 200–209.

[106] Kanit Wongsuphasawat, Dominik Moritz, Anushka Anand, Jock Mackinlay, Bill Howe, and Jeffrey Heer. 2016. Voyager: Exploratory Analysis via Faceted Browsing of Visualization Recommendations. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (Jan. 2016), 649–658. doi:10.1109/TVCG.2015.2467191

[107] Shaomei Wu, Lindsay Reynolds, Xian Li, and Francisco Guzmán. 2019. Design and Evaluation of a Social Media Writing Support Tool for People with Dyslexia. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14. doi:10.1145/3290605.3300746

[108] Pezhman Zare and Jayakaran Mukundan. 2015. The use of Socratic method as a teaching/learning tool to develop students' critical thinking: A review of Literature. *Language in India* 15, 6 (2015), 256–265.

[109] Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q. Weinberger, and Yoav Artzi. 2020. BERTScore: Evaluating Text Generation with BERT. arXiv:1904.09675 [cs.CL] https://arxiv.org/abs/1904.09675

[110] Zheng Zhang, Jie Gao, Ranjodh Singh Dhaliwal, and Toby Jia-Jun Li. 2023. VISAR: A Human-AI Argumentative Writing Assistant with Visual Programming and Rapid Draft Prototyping. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (San Francisco, CA, USA) *(UIST '23)*. Association for Computing Machinery, New York, NY, USA, Article 5, 30 pages. doi:10.1145/3586183.3606800

[111] Jian Zhao, Shenyu Xu, Senthil Chandrasegaran, Chris Bryan, Fan Du, Aditi Mishra, Xin Qian, Yiran Li, and Kwan-Liu Ma. 2023. ChartStory: Automated Partitioning, Layout, and Captioning of Charts into Comic-Style Narratives. *IEEE Transactions on Visualization and Computer Graphics* 29, 2 (2023), 1384–1399. doi:10.1109/TVCG.2021.3114211

[112] Zhenpeng Zhao, Rachael Marr, Jason Shaffer, and Niklas Elmqvist. 2019. Understanding Partitioning and Sequence in Data-Driven Storytelling. In *Information in Contemporary Society*, Natalie Greene Taylor, Caitlin Christian-Lamb, Michelle H. Martin, and Bonnie Nardi (Eds.). Springer International Publishing, Cham, 327–338.

[113] Jun-Peng Zhu, Peng Cai, Boyan Niu, Zheming Ni, Kai Xu, Jiajun Huang, Jianwei Wan, Shengbo Ma, Bing Wang, Donghui Zhang, Liu Tang, and Qi Liu. 2024. Chat2Query: A Zero-Shot Automatic Exploratory Data Analysis System with Large Language Models. In *2024 IEEE 40th International Conference on Data Engineering (ICDE)*. IEEE, Piscataway, NJ, USA, 5429–5432. doi:10.1109/ICDE60146.2024.00420

[114] Pohanyar Qyamudin Ziar. 2025. Impact of AI-Based Tools on Writing Skills. *Journal of Research in Education* 1, 1 (May 2025), 97–112. doi:10.3126/jore.v1i1.78725

A Appendix

Panel A: Explained Variance						
Component	PC1	PC2	PC3	PC4	PC5	PC6
Explained variance ratio (%)	57.16	19.22	10.95	7.47	3.15	2.05
Panel B: Loadings						
Usefulness	0.484	-0.116	-0.036	-0.487	-0.428	-0.576
Informativeness	0.437	0.383	-0.337	0.001	0.708	-0.217
Clarity	0.503	0.027	-0.070	-0.329	-0.117	0.787
Visualization	0.335	-0.502	-0.466	0.627	-0.159	-0.000
Insights	0.335	-0.448	0.733	0.102	0.372	-0.036
Questions	0.311	0.622	0.355	0.501	-0.371	-0.035

Table 2. PCA explained variance (Panel A) and six loadings (Panel B). PC1 is used as the Satisfaction factor.

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009

Predictor	Coef.	Std. Err.	t	p	95% CI Lower	95% CI Upper
Model 1: Clarity ∼ Questions						
Intercept	0.858	1.264	0.679	0.507	-1.820	3.537
Questions	0.679	0.318	2.135	0.049	0.005	1.354
Model fit: R^2 = 0.222, Adjusted R^2 = 0.173, n = 18						
Model 2: Clarity ∼ Insights						
Intercept	0.989	1.105	0.895	0.384	-1.354	3.339
Insights	0.656	0.281	2.334	0.033	0.060	1.250
Model fit: R^2 = 0.254, Adjusted R^2 = 0.207, n = 18						
Model 3: Usefulness ∼ Inaccuracies						
Intercept	4.467	0.788	5.668	< 0.001	2.796	6.138
Inaccuracies	-0.420	0.390	-1.384	0.185	-1.083	0.227
Model fit: R^2 = 0.107, Adjusted R^2 = 0.051, n = 18						

Table 3. OLS regressions predicting clarity from system-generated questions and insights, and predicting usefulness from perceived inaccuracies.