

Data Formulator 2: Iteratively Creating Rich Visualizations with AI

CHENGLONG WANG, Microsoft Research, USA
 BONGSHIN LEE, Yonsei University, Korea
 STEVEN DRUCKER, Microsoft Research, USA
 DAN MARSHALL, Microsoft Research, USA
 JIANFENG GAO, Microsoft Research, USA

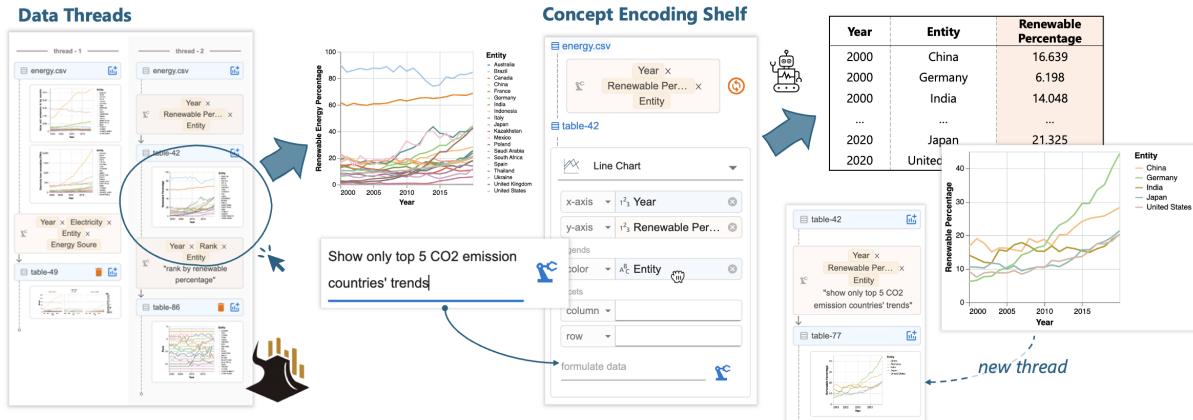


Fig. 1. With Data Formulator 2, analysts can navigate the iteration history in Data Threads and select previous designs to be reused towards new ones; then, using Concept Encoding Shelf, analysts specify their chart design using blended UI and natural language inputs, delegating data transformation effort to AI. When new charts are created, data threads are updated for future reference. Data Formulator 2 is available at <https://github.com/microsoft/data-formulator>.

To create rich visualizations, data analysts often need to iterate back and forth among data processing and chart specification to achieve their goals. To achieve this, analysts need not only proficiency in data transformation and visualization tools but also efforts to manage the branching history consisting of many different versions of data and charts. Recent LLM-powered AI systems have greatly improved visualization authoring experiences, for example by mitigating manual data transformation barriers via LLMs' code generation ability. However, these systems do not work well for iterative visualization authoring, because they often require analysts to provide, in a single turn, a text-only prompt that fully describes the complex visualization task to be performed, which is unrealistic to both users and models in many cases. In this paper, we present Data Formulator 2, an LLM-powered visualization system to address these challenges. With Data Formulator 2, users describe their visualization intent with blended UI and natural language inputs, and data transformation are delegated to AI. To support iteration, Data Formulator 2 lets users navigate their iteration history and reuse previous designs towards new ones so that they don't need to start from scratch every time. In a user study with eight participants, we observed that Data Formulator 2 allows participants to develop their own iteration strategies to complete challenging data exploration sessions.

Authors' Contact Information: Chenglong Wang, chenglong.wang@microsoft.com, Microsoft Research, Redmond, Washington, USA; Bongshin Lee, b.lee@yonsei.ac.kr, Yonsei University, Seoul, Korea; Steven Drucker, sdrucker@microsoft.com, Microsoft Research, Redmond, Washington, USA; Dan Marshall, danmar@microsoft.com, Microsoft Research, Redmond, Washington, USA; Jianfeng Gao, jfgao@microsoft.com, Microsoft Research, Redmond, Washington, USA.

1 Introduction

From an initial design idea, data analysts often need to go back and forth on a variety of charts before reaching their goals. Throughout this iterative process, besides updating the chart specifications, analysts face the challenges to transform and manage different data formats to support these visualization designs. Iterative chart authoring is prevalent in exploratory data analysis [42], where analysts often discover new directions from initial charts. For example, after noticing that the line chart about renewable energy percentage in Figure 1 are quite dense for comparing different countries’ trends, the analysts may want to filter it to show only top 5 CO₂ emitter’s trends, or visualize ranks of these countries each year instead. To achieve these, the analysts need different data transformations: the former requires filtering the data with each country’s aggregated CO₂ emission values, and the latter requires partitioning the data by year to compute each country’s ranking. Similar challenges are also relevant in the data-driven storytelling context [40, 41], where authors needs to derive new data to refine chart designs (e.g., annotation). For example, to highlight which countries are leading in renewable energy adoption, the author would superimpose a trend line of global median adoption rates over the line chart; the author may later convert the chart into small multiples to tell a story about most sustainable countries from each continent. Again, these new designs require data transformation from the current results.

Managing different data and chart designs together in these iterative authoring processes is challenging. As the analyst comes up with new chart designs, they need not only to understand the data format expected by the chart and tool, but also need to know how to use diverse transformation operators (e.g., reshaping, aggregation, window functions, string processing) in data transformation tools or libraries to prepare the data. Many AI-powered tools have been developed to tackle these visualization challenges (e.g., [2, 9, 26, 31, 54, 55]). These tools let users describe their goals using natural language, and they leverage the underlying AI models’ code generation ability [1, 5] to automatically write code to transform the data and create the visualization. Despite their success, current tools do not work well in the *iterative* visualization authoring context. Most of them require analysts to provide, in a single turn, a text-only prompt that fully describes the complex visualization authoring task to be performed, which is usually unrealistic to both users and models.

- First, despite free-form text prompts provide unbounded expressiveness for users to describe their visualization intent, they miss UI interactions’ precision and affordance, making it difficult for users to clearly describe complex designs they come up in later iteration stages. For example, the user needs a verbose prompt to clearly elaborate which fields they would like to use in each visual channel to create a faceted bar chart; without it, AI models could misinterpret the intent and create undesired charts, requiring further disambiguation efforts from the user. and the system can provide immediate visual feedback to the user. In fact, writing high-quality prompts requires skill and efforts. Even with clear goals in mind, it is challenging for inexperienced users to clearly describe their intent [48, 64].
- Second, existing AI-powered tools support only either single-turn or linear interactions with AI models, and therefore do not accommodate branching and backtracking that commonly occur in the iterative authoring process. To use single-turn text-to-vis tools in an iterative manner, users need to re-specify their intent from scratch each time they create a new design, even though the design update is minor. This not only is time consuming for the user, but also increases the chance of the AI model to fail the task, since the model needs to solve a complex task in one shot. While chat-based tools [26, 34, 66] support multi-turn interactions by reusing previous outputs in subsequent turns, they do not work well for branching contexts. When non-linear contexts are merged into a linear history, it is not only challenging for users to communicate which designs should be used towards next iterations, but also challenging for AI model to correctly retrieve relevant content from the long conversation history [14, 21, 65].

To overcome these limitations, we design a new interaction approach for iteratively chart authoring. Our key idea is to blend GUI and natural language (NL) inputs so that users can specify charts both precisely and flexibly, and we

design an interface for users to control the contexts, so that users can navigate and reuse previous design towards new ones, as opposed to starting from scratch each time. We realize the design with **the concept encoding shelf** for specifying charts beyond data format constraints and **data threads** for managing the user’s non-linear authoring history (Figure 1).

Chart specification with blended UI and NL inputs. Resembling shelf-configuration UIs [40, 55], the concept encoding shelf allows users to drag existing data fields they wish to visualize and drop them to visual channels to specify chart designs. Differently, with concept encoding shelf, users can also input new data field names in the chart configuration to express their intent to visualize fields that they want from a transformed data. Then, they can provide a supplemental NL instruction to explain the new fields and ask the AI to transform data and instantiate the chart. This blended UI and NL approach for chart specification makes user inputs both precise and flexible. Since Data Formulator 2 can precisely extract chart specification from the encoding shelf, the user doesn’t need verbose prompt to explain the design. By conveying data semantics using NL inputs, the user delegates data transformation to AI, and thus they doesn’t need to worry about data preparation. This approach also improves the task success rate of AI models. Because Data Formulator 2 can infer the visualization script directly from UI input, the AI model only needs to generate data transformation code. With the chart design provided as contexts to the AI model, the model has more information to ground the user’s instruction for better code generation. **Managing**

and leveraging iteration contexts with data threads. Data Formulator 2 presents the user’s non-linear iteration history as data threads and lets them manage data and charts created throughout the process. With data threads, users can easily navigate to an earlier result, fork a new branch, and reuse its context to create new charts. This way, users only need to inform the model how to update the previous result (e.g., “show only top 5 CO2 emission countries’ trends”, Figure 1) as opposed to re-describing the whole chart from scratch. When the user decides to reuse, the Data Formulator 2 tailors the conversation history to include only contexts relevant to that data to derive new result, allowing the AI to generate code with clear contextual information free from (irrelevant) messages from other threads. Besides general navigation and branching supports, data threads also provide shortcut for users to quickly backtrack and revise prompts to update recently created charts, which can be useful for analysts to explore alternative designs or correct errors made by AI.

Based on these two key designs, we developed Data Formulator 2, an AI-powered visualization tool for iterative visualization authoring. Data Formulator 2 supports diverse visualizations provided by Vega-Lite marks and encodings, and the AI can transform data flexibly to accommodate different designs, supporting operators like reshaping, filtering, aggregation, window functions, and column derivation. Like other AI tools [9, 55], Data Formulator 2 also provides users panels to view generated data, transformation code and code explanations to inspect and verify AI outputs.

To understand how Data Formulator 2’s multi-modal interaction benefits analysts in solving challenging data visualizations tasks, we conducted a user study consisting of eight participants with varying data science expertise. They were asked to reproduce two professional data scientists’ analysis sessions to create a total of 16 visualizations, 12 of which require non-trivial data transformations (e.g., rank categories by a criterion and combine low-ranked ones into one category with the label, “Others”). The study shows that participants can quickly learn to use Data Formulator 2 to solve these complex tasks, and that Data Formulator 2’s flexibility and expressiveness allow participants to develop their own verification, error correction, and iteration strategies to complete the tasks. Our inductive analysis of study sessions reveals interesting patterns of how users’ experiences and expectations about the AI system affected their work styles.

In summary, the main contributions of this paper are as follows:

- We design a multi-modal UI, composed of concept encoding shelf and data threads, to blend UI and NL interactions for users to specify their intent for iterative chart authoring.

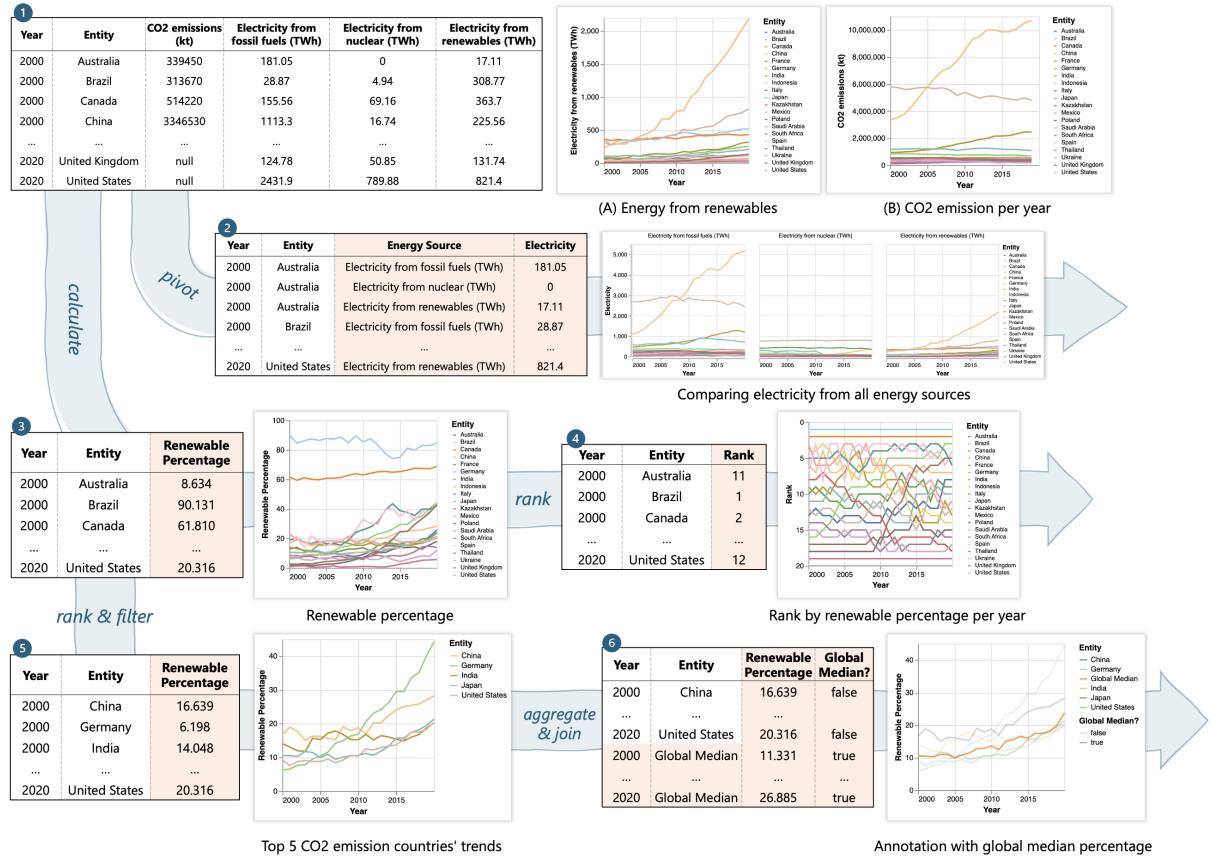


Fig. 2. A data analysis session where the analyst explores energy from different sources, renewable percentage trends, and ranks of countries by their renewable percentages from a dataset about CO₂ and electricity of 20 countries between 2000 and 2020 (table 1). The analyst has to create five versions of the data to support different chart designs in three branches. Data Formulator 2 lets users manage the iteration contexts and create rich visualizations beyond the initial data with blended UI and natural language inputs.

- We implement our design with an interactive visualization tool, Data Formulator 2, which enables users to iteratively create rich visualizations that requires multiple rounds of data transformations along the way.
- We conducted a user study to learn how the blended UI and NL interaction benefits analysts in data exploration sessions. We observed that analysts can easily develop their own strategies to work the AI system to perform data analysis and visualization tasks that best reflect their personal experience and expectation with the AI model.

2 Illustrative Scenarios: Exploring Renewable Energy Trends

In this section, we describe scenarios to illustrate users' experiences of creating a series of visualizations to explore global sustainability from a dataset of 20 countries' energy from 2000 to 2020. The initial dataset, shown in Figure 2-①, includes each country's energy produced from three sources (fossil fuel, renewables, and nuclear)

each year and annual CO₂ emission value (the CO₂ emission data only ranges from 2000 to 2019). We compare a professional data analyst's experience with computational notebooks and a journalist's experience using Data Formulator 2 to complete the analysis session shown in Figure 2.

2.1 Exploration with computational notebooks

Heather is an analyst who is proficient with a computational notebook and R libraries, ggplot2 and tidyverse. Because ggplot2 expects all data fields to be visualized on visual channels (e.g., x, y-axes, color, facet) are columns in the input data, Header uses tidyverse for data transformation.

Basic charts. To start, Heather wants to visualize the amount of electricity produced from renewables per country over the years with a line chart to see “if our planet is sustainable.” Since the input data (table-①) includes all required fields, Heather creates the line chart with ease, by mapping columns Year → x, Electricity from renewables (TWh) → y and Entity → color (chart ①-A). She then creates another line chart for CO₂ emission trends, mapping CO₂ emissions (kt) to the y axis (chart ①-B). Heather is puzzled that China, the country with considerable increased use of renewable energy, also has the biggest increase in CO₂ emissions. This is counterintuitive because renewables themselves would not cause CO₂ emission increase. Thus, Heather decides to dive deeper.

Renewable energy versus other sources. Heather suspects the CO₂ emission increase is caused by a surge of fossil fuel consumption. To compare fossil fuel usage against renewables, she wants a faceted line chart that shows electricity from each energy source side by side (chart ②). To create the chart, Heather needs to have a data table with columns—Year, Electricity, Entity, and Energy Source—and map the columns to x, y, color, and facet, respectively so that the chart is divided into subplots based on values from the Energy Source column. Because table ① stores electricity values across three columns in the wide format, Heather unpivots table ① into the long format, to fold specified column names into values in the Energy Source field and corresponding values into the Electricity field. She then creates the desired chart ② with the transformed data ②, and verifies her assumption: despite the increase of renewables usage, the usage of fossil fuel also grows significantly, leading to CO₂ emission increase. This motivates Heather to explore renewable trends by visualizing trends of the *percentage* of electricity from renewables over all three resources.

Renewable energy percentage and ranks. To visualize renewable energy percentage, Heather goes back to table ① to derive a new column Renewable Percentage, by dividing Electricity from renewables (TWh) from the total produced electricity for each country per year. With the new data ③, Heather visualizes the renewable percentage trends in chart ③, which shows that the percentage increase is slower than their absolute value increase (as shown in chart ①).

Because many countries share similar renewable percentage, it is quite difficult to compare different countries' trends. Heather thus decides to create a visualization of countries' renewable percentage ranks to complement existing charts. To calculate ranks of each country among others per year, Heather uses a window function on table ③ to partition the table based on Year, and apply the rank() function to Renewable Percentage to derive a new column Rank. With Rank mapped to y-axis, chart ④ allows Heather to clearly examine how different countries' ranks change in the last two decades; for example, Germany and UK are the two top ranked countries emerge from the bottom pack in 2000.

Renewable trends from top CO₂ emitters. Finally, Heather wants to focus on renewable percentage trends from top CO₂ emission countries, which make most influences to global sustainability. Despite table ③ contains all columns to be visualized, Heather needs to filter it based on the countries' CO₂ emission. To do so, Heather goes back to table ① to aggregate each country's total CO₂ emission, sort it and find top five. Heather then uses this intermediate result to filter table ③ to obtain renewable percentage from top five CO₂ emitters (shown as table ⑤) and creates chart ⑤.

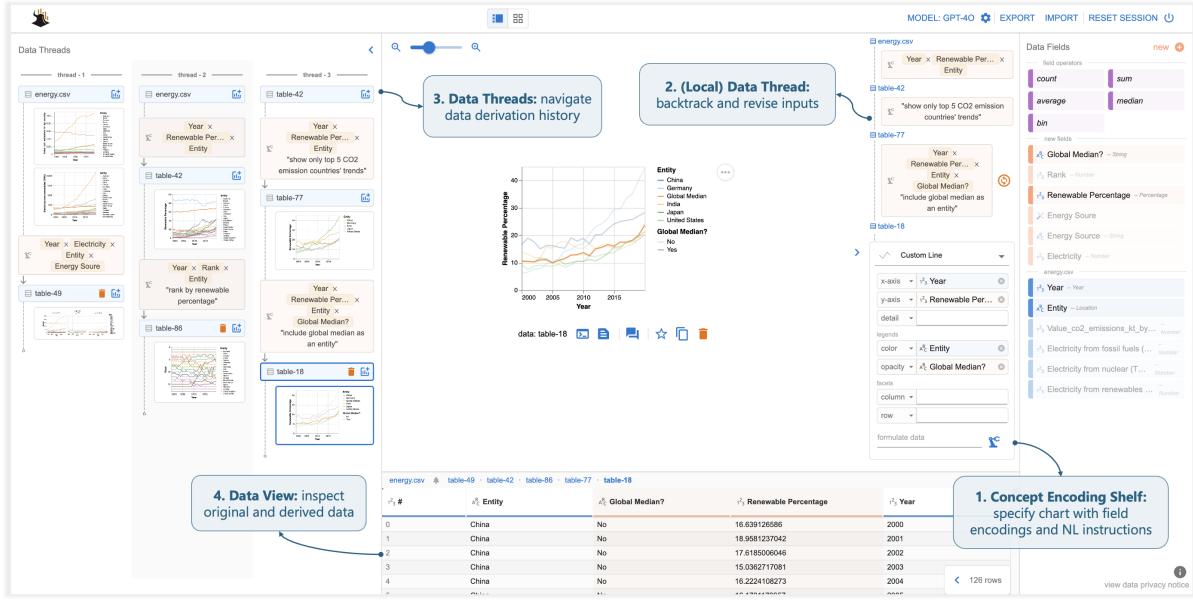


Fig. 3. Data Formulator 2 overview. The user creates visualizations by providing fields (drag-and-drop existing fields or type in new ones) and NL instructions to **Concept Encoding Shelf** and delegates data transformation to AI. **Data View** shows the derived data. The user can navigate data derivation history using **Data Threads**. They can then locate the desired point to refine or create new charts by providing follow-up instructions in **Concept Encoding Shelf**.

From this chart, it is clear that top CO₂ emitters are indeed heading in the right direction towards sustainability, despite total CO₂ emissions are still increasing with total energy produced also increasing each year. To publish this visualization, Heather decides to add an annotation to the plot with the median global renewable percentage. On top of table ⑤, Heather appends the median renewable percentage each year calculated from table ③ and includes a new column Global Median?, used as a flag to assist plotting so that global median can be colored in a different opacity. Chart ⑥ shows the final result, by including Global Median as an Entity and mapping Global Median? → opacity, median renewable percentage is visualized along other countries in a different opacity. Heather is satisfied with the results and concludes the session.

2.2 Exploration with Data Formulator 2

Megan is a journalist who has a solid understanding about data visualization. She utilizes visualizations effectively in her work but she doesn't program. Megan can create and refine rich visualizations iteratively with Data Formulator 2 (Figure 3), which inherits the basic experience of shelf-configuration style tools. She can specify charts by mapping data fields to visual channels of the selected chart and provide additional contexts using natural language.

Basic charts. Megan starts with line charts to visualize trends of electricity from renewables (Figure 2-①A). Since all three required fields are available from the input data, Megan simply selects chart type “line chart” in the encoding shelf and drags-and-drops fields to their corresponding visual channels (Figure 4-①). Data Formulator 2 then generates the desired visualization. To visualize the CO₂ emission trends, Megan swaps y-axis encoding with CO₂ emissions (kt) → y.

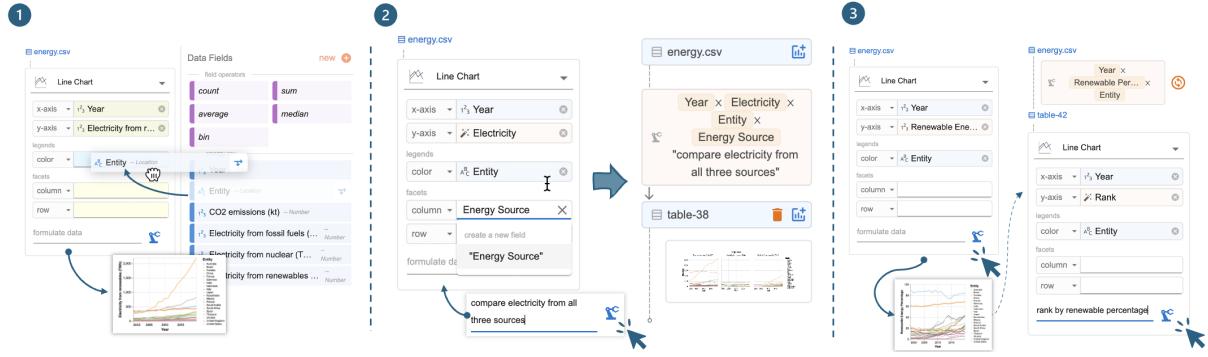


Fig. 4. Experiences with Data Formulator 2: (1) creating the basic renewable energy chart using drag-and-drop to encode fields; (2 and 3) creating charts that requiring new fields by providing field names and optional natural language instructions to derive new data.

Renewable energy vs other sources. Megan now needs to create the faceted line chart to compare electricity from all energy sources, which requires new fields Electricity and Energy Source. With Data Formulator 2, Megan can specify the chart using *future fields* and *NL instructions* in Concept Encoding Shelf (Figure 3-2) and delegate data transformation to AI.

As Figure 4-② shows, Megan first drags-and-drops existing fields Year and Entity to x-axis and color, respectively; then, she types in names of new fields Electricity and Energy Source in y-axis and column, respectively, to tell the AI agent that she expects two new fields to be derived for these properties; finally, Megan provides an instruction “compare electricity from all three sources” to further clarify the intent and clicks the formulate button. To create the chart, Data Formulator 2 first generates a Vega-Lite spec skeleton from the encoding (to be completed based on information from the transformed data); it then summarizes the data, encodings, and NL instructions into a prompt to ask an LLM to generate a data transformation code to prepare the data that fulfills all necessary fields, which is then used to instantiate the chart skeleton. After reviewing the generated chart and data, Megan is satisfied and moves to the next task.

Data Formulator 2 also updates data threads (Figure 3-⑤) with the newly derived data and chart so that Megan can manage and leverage data provenance. For example, Megan can delete the chart, create/fork new charts from either the original or new data, or select an existing chart to iterate from.

Renewable energy percentage and ranks. Megan proceeds to visualizing renewable energy percentage. Despite it required a different data transformation, Megan enjoys the same experience as the previous task: Megan drags-and-drops Year and Entity to x-axis and color (Figure 4-③), and enters the name of the new field “Renewable Energy Percentage” to y-axis; then, since Megan believes the field names are self-explanatory, she proceeds to formulate the new data without an additional NL instruction. Data Formulator 2 generates the desired visualization (Figure 5-①).

To visualize the ranks of countries based on their renewable percentage, Megan decides to continue it from the previous chart, which already computes renewable percentage. To do so, Megan duplicates the renewable percentage chart and update the y-axis field to another new field Rank and clicks “derive.” As shown in Figure 4-③, Megan’s interaction positions the Concept Encoding Shelf in the contexts prior result as opposed to the original data, which conveys her intent of reusing the data towards the new one. With the context information, the AI model successfully derives the desired chart (Figure 2-④) even only with Megan’s simple inputs.

Renewable trends from top CO₂ emitters. Next, Megan decides to visualize top-5 CO₂ countries’ renewable percentage trends. Megan again decides to iterate on the previous chart, otherwise she needs more effort to create a

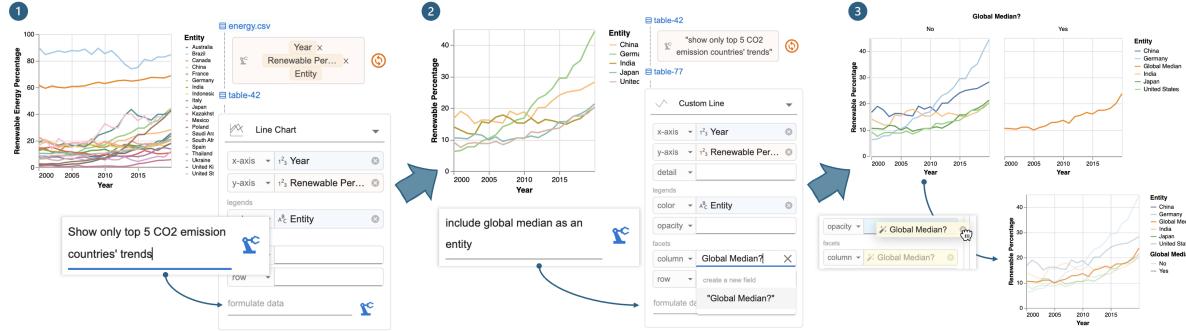


Fig. 5. Iteration with Data Formulator 2: (1) provide a new instruction on top of the renewable energy percentage chart to filter by only top CO₂ countries, (2) update the chart with a new field Global Median? and instruct Data Formulator 2 to add global median besides top 5 CO₂ countries' trends, and (3) move Global Median? from column to opacity to update chart design without deriving new data.

longer prompt to specify the task at once. Megan first use data threads (Figure 3-⑤) to locate renewable percentage chart. On top of that, Megan provides a new instruction below the local data thread, “*show only top 5 CO₂ emission countries’ trends*,” and clicks the “derive” button (Figure 5-①). Data Formulator 2 updates the previous code to include a filter clause to produce the new data and visualization (Figure 5-②).

Megan continues on the iteration process, to include global median trends besides top 5 CO₂ countries’ trends. Since this new chart requires different encodings and she wants to keep both visualizations around, Megan forks a branch by copying the previous chart. Then, she updates the Concept Encoding Shelf by (1) adding a new encoding Global Median?→column and (2) providing the edit instruction “include global median as an entity” (Figure 5-②). Once she clicks the derive button, Data Formulator 2 generates the new chart (Figure 5-③). Upon inspection, Megan prefers to combine two views in one, with global average rendered in a different opacity. Since these two charts require the same data fields, she simply selects a new chart type “custom line” (which exposes more chart properties than the basic line chart) and moves Global Median? to the opacity channel. Since it requires no data transformation, Data Formulator 2 doesn’t need to invoke AI and directly renders the chart. With all desired chart created, Megan concludes the analysis session. Figure 3-⑤ shows all three threads created by Megan that lead to the final designs.

2.3 Comparison of Exploration Experiences

The experience of Heather and Megan exploring global sustainability (using data ①) demonstrates an inherently iterative process. Both of them started with a high-level goal without concrete designs in mind and gradually formed the design from explorations in various branches. This iterative exploration process required a series of data transformation and the management of provenance, and thus is challenging for people not proficient in data transformation and programming. Here, we compare their exploration experiences to highlight how Data Formulator 2 bridges Megan’s skill gap, enabling her to achieve the analysis Heather, an experienced data analysis, performed.

2.3.1 Data transformation and chart creation. When new designs are considered, Heather needs to prepare new data to accommodate the design, even when some designs are seemingly close (e.g., charts ③ and ⑤). This requires her to understand the data shape expected by the charts, choose the right transformation idiom (e.g., unpivot for table ②, join and union for table ⑥), and implement them with proper operators. Once the data is prepared, Heather

can easily specify chart by mappings data columns to visual channels of the selected chart type. Her proficiency in data transformation is essential for her to create rich visualizations beyond the initial dataset.

To bridge Megan’s skill gap in data transformation, Data Formulator 2 lets Megan specify her intents in a unified interaction that combines chart encodings and natural language, regardless of the types of data transformation required behind the scene, and data transformation is delegated to AI. Because the Concept Encoding Shelf resembles the shelf-configuration UI, Megan’s experience from Power BI translates well into Data Formulator 2. Furthermore, since Megan communicates the chart design using concept encodings, she only needs to provide a short supplementary NL instruction to clarify her intents; based on these inputs, Data Formulator 2 assembles a detailed prompt to communicate with the AI model. If Megan were to use text-only interface to interact with AI, she needs more detailed prompt to explain her intent to avoid ambiguity, including explaining chart encodings she created from drag-and-drop interactions easily.

2.3.2 Managing branching contexts. During the exploration, Heather backtracks several times to reuse previous results toward new designs (e.g., chart ④ → data ③ → chart ⑤), creating three branches along the way. Because Heather programs in a notebook, she can either copy and adapt previous code snippets or reuse variables computed in previous iterations for new designs. This way, Heather lowers her specification efforts despite new designs are more complex. Heather’s programming expertise is essential for her to manage the branching contexts in a linear programming environment.

For Megan, managing branching contexts with different version of data could be challenging without Data Formulator 2. Should Megan use a chat-based AI interface, she would need to prepare a verbose prompt to explain contexts and data transformation goals in detail each turn to avoid extra disambiguation efforts, especially when multiple branches are mixed in the chat history and the task becomes more complicated later on. Data Formulator 2’s data threads address this challenge. Data threads not only provide a visualization for Megan to review history, but also let her visit previous states and reuse them towards new branches as Heather did. This way, Megan only needs to specify updates to be applied as opposed to describing the full design from scratch in one shot, and the AI model can generate results more reliably leveraging the contexts Megan provided. Shall Megan spot undesired results, she could also use data threads (Figure 3-③) to rerun or backtrack one step to revise instructions, as opposed to restarting from the scratch.

3 The Data Formulator 2 System Design

As described earlier, Data Formulator 2 combines UI and NL interactions in a multi-modal UI to reduce analysts’ visualization authoring efforts, and it provides data threads for users to navigate iteration history and specify new designs on top of previous ones. Data Formulator 2 employs the following system designs to support such interactions:

- First, to allow users to specify chart design and data transformation goals from different paradigms (shelf-configuration UI versus NL inputs), Data Formulator 2 **decouples chart specification and data transformation** and solve them with different techniques (template instantiation versus AI code generation).
- Second, to support reusing, Data Formulator 2 **organizes the iteration history as data threads, treating data as first class objects**. Data Formulator 2 enables users either to locate a chart from a different branch and follow up, or to quickly revise and rerun the most recent instruction leading to the current chart.

We next detail how Data Formulator 2 realizes these designs, and additional features designed to assist users to understand AI-generated results.

3.1 Multi-modal UI: Decoupling chart specification and data transformation

Data Formulator 2 decouples chart specification and data transformation so that users can benefit from both the precision of UI interaction to configure chart designs and the expressiveness of NL descriptions to specify

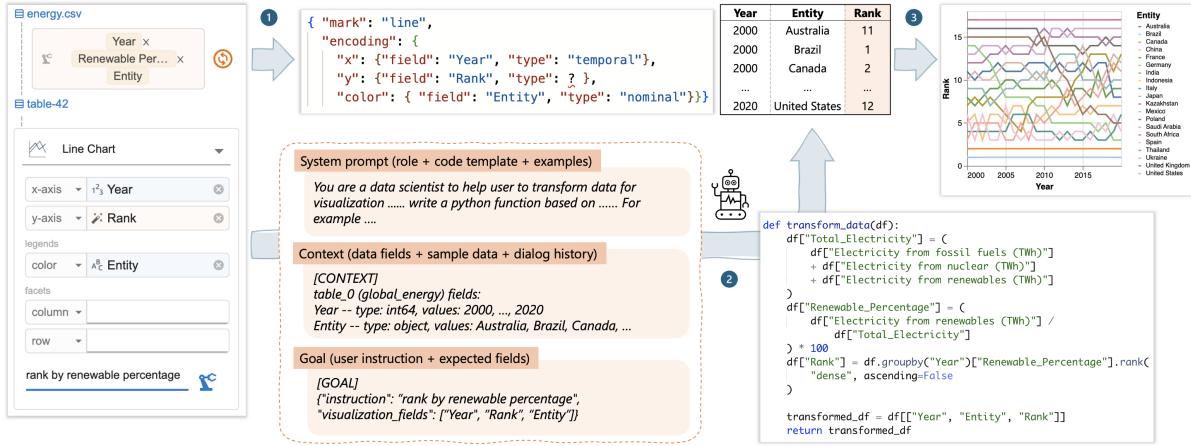


Fig. 6. Data Formulator 2’s workflow. (1) Given the user specification in the concept encoding shelf, Data Formulator 2 first generates a Vega-Lite spec skeleton from the selected chart type. (2) When the chart requires new fields (e.g., Rank), Data Formulator 2 compiles a prompt and delegate data transformation to AI, and (3) Upon completion, the Vega-Lite skeleton is instantiated with the new data to produce the desired chart.

data transformation goals. As shown in Figure 6, given a user specification in the concept encoding shelf, Data Formulator 2 generates the desired chart in three steps: (1) generating a Vega-Lite script from the selected chart type, (2) compiling a prompt and delegate data transformation to AI, and (3) using the generated data to instantiate Vega-Lite script to render the desired chart.

Chart specification generation. Data Formulator 2 adopts a chart type-based approach to represent visualizations, supporting five categories of charts: scatter (scatter plot, ranged dot plot), line (line chart, dotted line chart), bar (bar chart, stacked bar chart, grouped bar chart), statistics (histogram, heatmap, linear regression, boxplot) and custom (custom scatter, line, bar area, rectangle where all available visual channels are exposed for advanced users). Each chart type is represented as a Vega-Lite template with a set of predefined visual channels, including position channels (*x*, *y*), legends (color, size, shape, opacity), and facet channels (column, row) shown to the user in the concept encoding shelf. For example, a line chart is represented as a Vega-Lite template `{ "mark": "line", "encoding": { "x": "null", "y": "null", "color": "null", "column": "null", "row": "null" }}`, and when the user selects line chart, channels *x*, *y*, color, column, and row are displayed in the concept encoding shelf. Using chart type-based design, Data Formulator 2 supports predefined layered chart (e.g., ranged dot plot and linear regression plot that are composed from line and scatter, Figure 7-right). Additional chart types (e.g., bullet chart) can be supported by adding new Vega-Lite templates with respected channels to the library.

As the user inputs fields into the concept encoding shelf, either by dragging and dropping it from existing data fields or by typing in new fields they wish to visualize, Data Formulator 2 instantiates the Vega-Lite template with provided fields. For example, as shown in Figure 6-①, when the user drags Year → *x*, Entity → *y* and types Rank in *y*, the line chart template mentioned above is instantiated with provided fields: if the field is available in the current data table, both field name and encoding type are instantiated (e.g., Year with type “temporal”), otherwise the encoding type is left as a “<placeholder>” to be instantiated later when data transformation completes.

The shelf-configuration design provides users with simple yet precise interaction. The concept encoding shelf saves users efforts from writing prompts to explain the chart design. Figure 7 further illustrates how the specification

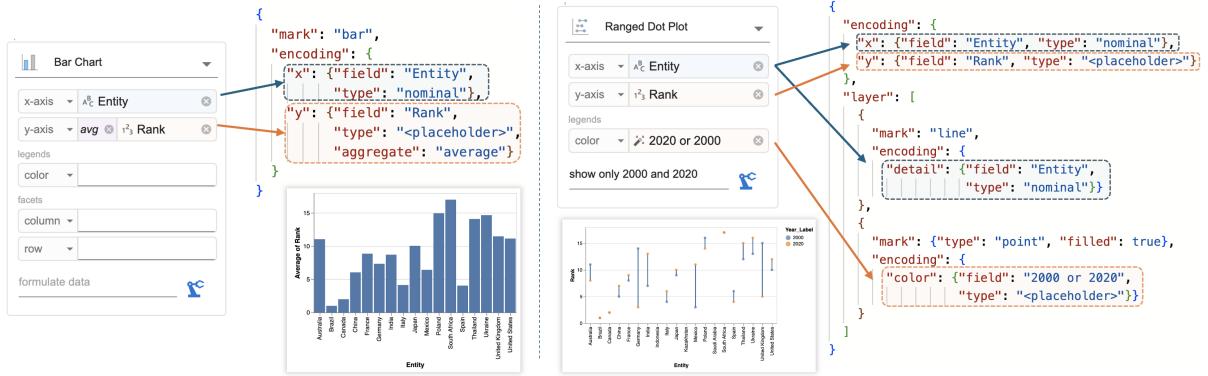


Fig. 7. Concept encoding shelf instantiates users' encodings as a Vega-Lite specification. The user creates a bar chart showing average rank of countries with an “avg” operator on y-axis, and a ranged dot plot to compare ranks of each country in 2000 and 2020 (the chart template routes users’ x-axis encoding “Entity” to both x and detail channels).

in the concept encoding shelf interacts with the underlying Vega-Lite scripts. In Figure 7-left, the user can specify an ‘avg’ operator the y-axis to transform the axis and the operator is instantiated as the “aggregate” property of y-axis in the script. In addition, Figure 7-right shows another example of the user working with a layered chart (ranged dot plot): as the user fills fields in the UI, Data Formulator 2 populates corresponding fields to different parameters in the predefined chart template.

Data transformation with AI. From the concept encoding shelf, Data Formulator 2 assembles a prompt and queries LLM to generate a python code to transform data. The data transformation prompt contains three segments: the system prompt, the data transformation context and the goal (illustrated Figure 6-②, full prompt is shown in the Appendix):

- The **system prompt** describes the role of the LLM and output format. Besides generic role descriptions (i.e., LLM as a data scientist to help with data transformation), the system prompt instructs LLM to solve the data transformation task in two steps: (1) refine the user’s goal and output as a json object, recapitulating intermediate fields and final fields to be computed from the original data and (2) generate a python code following a provided template. The system prompt ends with an input output example (following few-shot prompting strategy) demonstrating the process. The design rationale behind the “goal refinement” step is to allow LLM to reason about potential discrepancy between users’ provided fields and their instruction (e.g., the user may ask about color by energy type but didn’t put “energy type” on the color encoding) and determine the final list of fields to be computed.
- Data Formulator 2 then assembles a section of **context prompts** that describe the data transformation contexts to be performed. When the chart is created from scratch, the context prompts describe the input data to be transformed, including its data fields (data type and example values) as well as first five rows of the data. The data context not only helps the LLM understand semantics of new fields specified by the user, but also provides essential information related to data formats (especially data type, string formats and whether columns contains null values) to ensure the generated transformation code is executable on the given data. When the chart is specified on top of previous results, the dialog history between Data Formulator 2 and the LLM leading to that data (including both instructions from the user and previous code generated by the LLM) is also appended in context. This way, despite users’ followup prompts can be short, the grounded contexts help the model to understand user intent and reuse previously generated code.

- Finally, Data Formulator 2 assembles a **goal prompt** section, combining the NL instruction provided in the text box and field names used in the encodings. When user skips NL instruction (as shown in Figure 4-③), the instruction part is simply left blank. This goal will be refined by the LLM as instruction by the system prompt before attempting to generate the data transformation code.

With the full input, Data Formulator 2 prompts the LLM to generate a response, consisting of the refined objective and the code. Below shows the LLM’s refined objective for the task in Figure 6, and the generated code is shown in Figure 6-②.

```
{
  "detailed_instruction": "Calculate the percentage of electricity generated from renewables for each country
    ↳ per year. Then, rank the countries by their renewable percentage for each year.",
  "output_fields": ["Year", "Entity", "Renewable_Percentage", "Rank"],
  "visualization_fields": ["Year", "Rank", "Entity"],
  "reason": "To achieve the goal of ranking countries by their renewable percentage, we need to calculate the
    ↳ renewable percentage for each country per year and then determine the rank based on this percentage." }
```

Data Formulator 2 then runs the code on the input data. If the code executes without errors, the output data is used to instantiate the Vega-Lite script generated in the previous step, by first inferring semantic types of newly generated columns (to determine their encoding type), and then assembling the data with the script to render the visualization (Figure 6-③). Occasionally, the generated code may cause runtime errors, either due to attempting to use libraries that are not imported, references to invalid column names, or incorrectly handling of undefined or NaN values. When errors occur, before asking users to retry, Data Formulator 2 tries to correct the errors, by querying the LLM with the error message and a follow-up instruction to repair its mistakes [8, 33]. When repair completes, the visualization is similarly generated. Either way, Data Formulator 2 updates the data threads and presents the results to the user.

3.2 Data threads: navigating the iteration history

During the iterative visualization process, the analyst needs to navigate their authoring history to locate relevant artifacts (data or charts) to take actions (delete, duplicate or followup). Data Formulator 2 introduces data threads to represent the tree-structured iteration history to support navigation tasks. In data threads, we treat data as the first class objects (nodes in data threads) that are connected according to the user’s instruction provided to the AI model (edges), and visualizations are attached to the version of data they are created from. Centering the iteration history around data benefits user navigation because it directly reflects the sequence of user actions in creating these new data. This design also benefits the AI model: when user issues a follow-up instruction, Data Formulator 2 automatically retrieves its conversation history with the AI towards the current data and then instruct the AI model to rewrite the code towards new goals based on the retrieved history. This way, the AI model does not pose risk of incorrectly using conversation history from other branches to make incorrect data transformation. As shown in Figure 8, the code and the conversation history is attached to each data nodes. Each turn when the user provides a follow-up instruction, the AI model generates new code by updating the previous code (could be deletion, addition or both) to achieve the user’s goal; this way, the code always takes the original data as the input with all information accessible. Comparing to an alternative design where we only pass current data to the AI model and asks it to write a new code to further transform it (i.e., reusing the data as opposed to reusing the computation leading to the data), our design has more flexibility to accommodate different styles of followup instructions — either the user wants to further update the data (e.g., “now, calculate average rank for each country”), revise previous the computation (e.g., “also consider nuclear as renewable energy”) or creating an alternatives (e.g., “rank by CO₂ instead”) — since the AI has access to the full dialog history and the full dataset. In contrast, the data-only reuse approach restricts the AI model’s access to only the current data, limiting its ability to support “backtracking” or “alternative design” styles instructions.

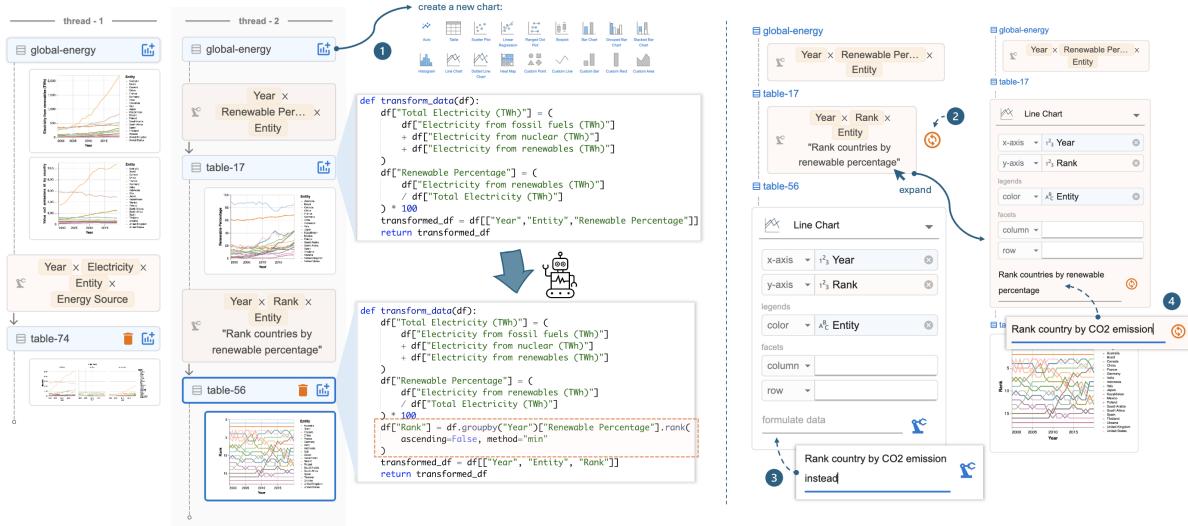


Fig. 8. Data threads and local data threads (right). In data threads, the user can create new charts from previous versions of data, and open previous charts in the main panel to create new branches; when creating new data, the AI model is instructed to revise previous code based on user instructions. In local data threads, the user can easily (1) rerun the previous instruction, (2) issue a follow-up instruction or (3) expand the previous card to revise and rerun the instruction.

During iteration, the analyst needs both (1) locating a data or a chart further from the current one to create new branch in derivation tree and (2) performing quick follow-up/revisions of latest instruction from the latest data. To accommodate these different needs, Data Formulator 2 presents both (global) data threads and local data threads. For navigation, the key challenge is assist user to distinguish the desired content from others, and thus data threads are located in a separate panel with previews of data, instruction and charts to assist navigation (Figure 3). This support users different navigation styles, either they want to navigate by provenance (i.e., using instruction cards to locate desired data) or navigate by artifacts (i.e., using visualization snapshots to recall data semantics). Once the user locates the desired data, they can click and open a previous chart to display it in the main panel for further updates as well as create a new chart directly from the data Figure 8-①. To support quick updates from the current result, Data Formulator 2 aims to minimize users' interaction overhead. Thus, the local data thread is designed as part of the main authoring panel Figure 3. The local data thread visualizes only the history leading from the initial data to the current one and omits chart snapshots to minimize distraction (the full history is still available in the global data thread). By integrating the local data thread with the concept encoding shelf, Data Formulator 2 helps the user understand the authoring contexts and enables them to perform quick local updates. As shown in Figure 8, the user can rerun the previous instruction (e.g., when the AI produces an incorrect result and they would like to quickly retry before updating instructions, ②), provide a follow-up instruction to refine the data (③), as well as quickly open the previous instruction to modify and rerun the command (④).

With data threads, analysts can manage and navigate the history and perform iterative updates from previous results, similar to how data analysts reuse code and data in computation notebooks. Otherwise, if the analyst needs to start from scratch and ask the AI to achieve the goal at once, they need efforts to prepare rather detailed prompts to reduce ambiguity, especially for describing more complex charts they need in later analysis stages.

3.3 Miscellaneous: inspecting results and styling charts

As an AI-powered tool, Data Formulator 2 lets the user verify AI-generated results and resolve mistakes made by AI. It displays transformed data, visualization, the code, and an explanation of the code in the main panel. This design accommodates various user verification styles identified by prior work [12, 54]: e.g., viewing high-level correctness from chart, inspecting corner cases in data, inspecting the transformation output, as well as understanding the transformation process from code. Data Formulator 2 utilizes a code explanation module to query the AI model to translate code into step-by-step explanations assist users to understand the process. Furthermore, despite data transformations generated in the later iteration stages can be complex, users only need to verify its correctness against its predecessor because Data Formulator 2 users create visualizations incrementally. This considerably lowers users' verification efforts, as we discovered in our study in Section 4. As previously mentioned in Figure 8, when the user discovers errors, they can take advantages of the data thread's iterative mechanism to rerun, follow up or revise instructions to correct results.

Benefiting from the decoupled chart specification and data transformation processes, when the user wants to update visualization styles (e.g., change color scheme, change sort order of an axis, or swap encodings) that do not require additional data transformation, they can directly perform edits in the concept encoding shelf, by expanding the channel property and update parameters or swapping encoded fields. These updates are directly reflected in the Vega-Lite script and rendered in the main panel. Unlike interactions with AI which has a slightly delayed response time, this approach allows the user to achieve quick and precise edits with immediate visual feedback to refine the design.

3.4 Implementation

Data Formulator 2 is implemented as a React web application, with a backend Python server running on a Dv2-series CPU with 3.5 GiB RAM. Data Formulator 2 has been tested with different versions with OpenAI models, including GPT-3.5-turbo, GPT-4, GPT-4o and GPT-4o-mini (we used GPT-3.5-turbo in our user study) all of which except GPT-4 can generally response within 10 seconds. Since the LLM generates code to manipulate data as opposed to directly consume data, data size does not affect its response time. Data Formulator 2 can sometimes be slow due to Vega-Lite rendering overhead (e.g., large dataset with > 20,000 rows, long data threads with > 20 charts), we envision that on-demand re-rendering of charts can improve its performance in deployment.

4 Evaluation: Iterative Exploratory Analysis

We conducted a user study to understand potential benefits and usability issues of Data Formulator 2, as well as strategies developed by users when iteratively creating visualizations in an exploratory data analysis session.

4.1 Study Design

Participants. After piloting and refining the design of user study and Data Formulator 2 with three volunteers, we recruited eight participants from a large company. Participants self rated their skills (Figure 9) on a scale (“Novice,” “Intermediate,” “Proficient,” and “Expert”) in the following aspects: (1) chart creation – experience with chart authoring tools or libraries, (2) data transformation – experience with data transformation tools and library expertise, (3) programming, and (4) AI assistants – experience with large language models (e.g., ChatGPT [1]) and prompting.

Setup and procedure. Each study session, conducted remotely with screen sharing, consisted of four sections within a 2-hour slot. After a brief introduction of the study goal, participants were asked to follow step-by-step instructions in a tutorial presented in slides (~25 minutes). To make sure that they understood the tool and process, practice tasks (~15 minutes) were presented during and after the tutorial, where participants could ask questions as they worked through the tasks. Then, participants were given two study tasks to complete, where only clarification

ID	Role	Chart	Data	Programming	AI assistants	Dataset 1	Dataset 2	Hints
P1	Developer	Proficient	Expert	Expert	Intermediate	1047s	1666s	1
P2	Data Scientist	Proficient	Expert	Expert	Expert	1636s	1886s	0
P3	Data Architect	Proficient	Expert	Expert	Expert	715s	2207s	0
P4	Developer	Novice	Intermediate	Proficient	Intermediate	1036s	1521s	1
P5	Developer	Intermediate	Intermediate	Proficient	Novice	1251s	2937s	1
P6	Data Scientist	Intermediate	Expert	Intermediate	Proficient	856s	1148s	3
P7	Data Scientist	Proficient	Expert	Proficient	Proficient	1638s	2372s	1
P8	Developer	Proficient	Proficient	Expert	Novice	1043s	1987s	2

Fig. 9. Participants' self-reported role and expertise on chart creation, data transformation, programming, and AI assistants experiences, their task completion time and hints needed during the study tasks.

questions were allowed – we recorded hints participants requested about the tool when they got stuck. The two study tasks involved 16 visualizations to be created in total, with 12 of them requiring data transformation. Participants were encouraged to think aloud as they performed the tasks. We concluded the session with a debriefing with participants to (1) compare their experiences with Data Formulator 2 with the tools they have been using for data analysis, (2) understand strategies behind the way they used Data Formulator 2, and (3) gather impressions and suggestions for improvements to the tool. Participants were encouraged to take breaks between phases.

Tutorial and practice tasks. We use the global energy dataset (described in Section 2) for tutorial and practice tasks. In the tutorial, participants follow detailed instructions to recreate the six visualizations, all but one (chart ④) in Figure 2. Besides, participants also learned to inspect results and work with AI's mistakes. In the practice tasks, participants were asked to do similar analysis but focusing on the electricity generated from nuclear, with an additional task of creating a bar chart to compare the difference of nuclear between 2000 and 2020, which requires both table pivoting and calculation.

Study tasks. To focus on peoples' iterative processes rather than their ability either to create a single chart or to gain insights in exploring data, we decided to employ an *exploration session reproduction* approach, which asks participants to reproduce two data exploration sessions conducted by an experienced data scientist. We wanted to see if participants could iteratively create charts with Data Formulator 2, without requiring them to come up with exploration objectives on the fly (otherwise we would have limit our participants to only highly skilled data scientists). We took two exploration sessions from David Robinson's live stream analysis of Tidy Tuesday datasets.

Figure 10-① shows the first data exploration session: given a dataset on college majors and income data (173 rows × 7 columns), participants were asked to create seven visualizations (2 basic charts + 5 charts requiring data transformation) that progressively explore top earning majors and the relationship between women ratio and major salary. The exploration process required participants to derive new fields (e.g., women ratio), filter data (filter by top 20 earning majors), derive new data (aggregate to obtain major categories with top earnings) and perform conditional formatting (color by top 4 categories and "others"). In our task presentation, we provided the description of the task and reference chart (similar to the chart reproduction study [39, 41]) for all but the last two visualizations. We hid the reference charts for the final two visualizations and asked participants to verify the correctness, so that we could use them to probe participants' verification strategies. We did not provide the iteration direction (i.e., which charts should they base on to create a new one) in the task description, which let participants develop a variety of iteration techniques with only the high-level task guidance.

Figure 10-② shows the second data exploration session: given a dataset of movies with their budget and gross (3281 rows × 8 columns), participants were asked to explore movies and genres with highest return-on-investment values, comparing profit and profit ratio as metrics with 9 visualizations created along the way. Besides two basic box plots to show budget and worldwide gross distribution, the other seven charts require data transformation, including calculation and aggregation (average profit / profit ratio for each genre), string processing (extract year



Fig. 10. Study tasks. Dataset 1: Understanding top earning majors and the relation between salary and women percentage. Dataset 2: Exploring movies genres with best return-on-investment values (profit vs. profit ratio) and top movies. The branching directions here are only for illustration and are not provided to participants; participants developed different iteration strategies themselves.

for trends), filtering ($\text{year} > 2000$), and partitioning and ranking (top 20 movies for each metric). We again hide references of the final two charts to probe participants' verification process.

4.2 Results

Task completion. All participants successfully completed all 16 visualizations (Figure 9): participants took less than 20 mins on average to finish the seven charts in task 1, and about 33 mins for the nine charts in task 2. Since we let participants deviate from the main exploration task (e.g., in task 2, P4 asked to sort the bar chart for top profitable movies are based on their profits, even though it was not required), the recorded completion time is an overestimate of the actual task time. During the study, six participants asked for hints to get unstuck during tasks; we categorize them as follows:

- Task clarification: P1 didn't realize top movies are restricted to movies after 2000; P4 and P6 required hints about the difference between profit and profit ratio in task 2; P6 also asked about whether *x*-axis should be Year or Date when plotting movie profit trends.
- Data clarification: P6 and P8 were hinted to notice the difference between fields Major and Major Category in task 1.
- System performance: P5 encountered a performance issue, as they created large sized charts: in task 2, they created multiple bar charts with Movie mapped to the *x*-axis, resulting bar charts containing 1300 categorical values causing rendering issues. They were suggested to reset the exploration session.
- Chart encoding: P7 and P8 required hints on “why the chart didn't render color legends” when they didn't put a field in the color encoding; they expected to specify it only in NL input but not in the concept encoding shelf.¹

During the debriefing, participants commented that these tasks were much more difficult to complete with tools they are familiar with. P1 mentioned that they were “*obviously much faster*” with Data Formulator 2 as it helped with data transformation despite being an programming expert. When asked about their experience comparing against chat-based AI assistants, participants noted (1) the iteration support makes it easy to create more charts and (2) the UI + NL approach in Data Formulator 2 is more effective for communicating and constraining intent. For example, P2 mentioned “*with ChatGPT, I would have to put a bit more effort to specify the instructions to get what I want, iterations here is much faster with UI*”, and P4 mentioned that “*with ChatGPT, you need to much more contexts, I need to describe in detail about what x,y-axes should be, but here I can just provide with UI*.”

Iteration styles. Data Formulator 2 lets users develop their own iteration strategies. We observed three major distinct styles of iteration, in terms of which tables or charts participants chose to derive a new chart.

The first type of users preferred to achieve a particular chart through small, incremental changes from an existing chart that shared either similar data fields or similar chart configuration. For example, P2 and P3 chose to create the line chart showing profit ratio trends overtime on top of the bar chart showing the average profit ratio per genre, and next visualized movies with highest profit ratio further on top, since they share the same derived field profit ratio. P2 mentioned “*I definitely like to be able to just work on top of that and like going forward by just giving a new prompt, because it remembers the context prior to the last one, it ends up generating the right data and visualization.*” P2 further commented that they did not like too much branching: “*...felt that it would be harder to go back to the source and fix every single time.*” P7 also preferred incremental changes, but with a focus on visual similarity as opposed to data similarity.

In contrast, the second type of users preferred to go back and re-issue a prompt to achieve all the changes from the initial data as succinctly as possible. For example, P1 mentioned that “[I] like keeping it as terse as possible that will get me the right result.” P4 also felt that sometimes it was more productive to just start over from the original dataset throwing out all iterations, especially when they failed to produce a desired outcome: “*when we had all of those failures, I went back to the original base dataset and then frame my question there.*”

The third type of users primarily think about the iterations in terms for adding (or retrieving) columns from the dataset. P5 preferred to first instruct Data Formulator 2 to add/remove columns from an existing data (e.g., bring back fields that might have been dropped in previous iterations as needed, or add a new field required for the desired chart), and then create visualization from the right data.

Organization of iteration history. When asked about their rationale behind branching strategies, all participants agreed data threads are essential for managing iteration histories. Regarding their preferred organization style, P1 mentioned “*I don't like to pollute my workspace*” and “*I'd like to keep my workspace as clean as possible*” and thus they always chose to backtrack and fix previous instruction when encountering undesired results. P2, who mentioned “*going back created too much branching*” instead preferred follow through. P4 used prompts to help navigate iterations to find the one they were looking for: “*I was using the prompts as my anchor to figure out where*

¹In the study version, Data Formulator 2 didn't include the feature of refining users' goals when there were conflicts between NL and encoding shelf specification (described in Section 3 system prompt). This feature is introduced later to address this particular issue.

I wanted to go.” P8 found it sometimes difficult to iterate in Data Formulator 2 because data threads were “*linear instead of hierarchical*”: they preferred a tree-view data thread organization, where they could scan quickly through the entire branching tree for a dataset, its transformations and visualizations and then collapse branches that were not of interest for the current goals.

Verification. To proceed through iterative exploration, or repeat/correct a step, participants needed to verify that the chart or transformation was performed correctly. Some used the explanations of the code, some (even non-python programmers) used the actual code, and some used the result tables to validate the impact of the transformations. P3 mentioned “*as an expert, I like to see the prompt to the model, and then the code generated; but as a business user, I would imagine using more data, chart, and explanations.*” P4 mentioned “[explanation] steps were really, really helpful in terms of figuring out whether it is doing the right thing as to what I’m asking it to do. That and also the data chart underneath.” Interestingly, P7 stated that they preferred to use code rather than explanations of the code, but in the study, they used almost exclusively the explanations. They stated that they felt some pressure from the study environment not to spend too much time understanding code for which they were not familiar with, but they would trust code more. We also observed participants who developed trust in a workflow (by examining code and data tables) when it was straightforward, and then, they assumed the more complicated transformations built on top of these steps worked.

Miscellaneous. Several users noted potential improvements of Data Formulator 2 for iterative chart authoring. P1 commented on how small interface variations might give different affordances. For instance, “*if there was a large view for data threads, it would encourage me to do more transformations and do more branching.*” P3 mentioned that they prefer AI to ask the user to disambiguate when the intent is unclear rather than trying to solve the task with unclear specification. P7 used instructions that were very detailed and sometimes incorrect, which in turn, made iteration more difficult, since it was difficult to incrementally modify these instructions. We discussed the potential of having templates or AI feedback for instruction crafting to reduce errors.

5 Related Work

Data Formulator 2 builds on top of existing work on data transformation, chart authoring, and AI-powered visualization tools.

LLM-powered visualization tools. Large language models’ code generation ability [1, 5, 24, 50] motivates the designs of new AI-powered visualizations tools [10, 26, 49, 55] that allows users to create visualization using high-level natural language descriptions. For example, given a dataset and a visualization prompt, LIDA [9] automatically generates a data summary and prompts the LLM to generate python code to transform data and generate visualizations. Because LLMs can struggle in understanding complex chart logics, ChartGPT [49] decomposes visualization tasks into fine-grained reasoning pipelines (e.g., data column selection, filtering logic, chart type, visual encoding), using chain-of-thoughts prompting [56] to guide LLMs to generate code step by step. Data Formulator [55] leverages LLMs to derive new data columns that can be used in traditional shelf-configuration UI. Because these tools focuses on single-turn user interaction with abstract NL descriptions, they are not suitable for iterative analysis where the analyst may branch or revise designs throughout. For multi-turn interactions, users can directly have conversation with LLMs in Code Interpreter [1] or Chat2Vis [26]: Code Interpreter equips the LLM with a Python interpreter so that the model can generate and execute code to transform data and create visualizations with the user interactively; Chat2Vis further includes visualization-specific prompts to help the model generate visualizations more reliably. Since these tools organize the dialog linearly, when the context contains branches, the user needs extract efforts to explain the task so that the model can retrieve the correct context, otherwise the model are more likely to produce undesired results [14, 21, 65]. Besides, since these tools are based on NL inputs, when the user has concrete designs in mind, they need additional efforts to elaborate the design clearly (especially when the design is complex) so that the model can produce their desired results.

Data Formulator 2 is also LLM-powered tool that shares similar prompt designs like LIDA and Chat2Vis (e.g., using data summary to explain the authoring context) and supports NL interaction. Instead of using only NL inputs, Data Formulator 2 blends UI and NL inputs for chart specification so that users can communicate their intent both precisely and flexibly. This design is different from Data Formulator, whose UI and NL inputs work independently: Data Formulator’s NL interface restricts data transformation to column-wise computation, and the user needs additional efforts to complete other transformations like reshaping and aggregation separately using UI with a different paradigm (programming by example); with unified UI and NL interaction in Data Formulator 2, the user can achieve more expressive data transformation with less input efforts. Data Formulator 2’s data threads generalize linear contexts used in existing dialog systems, it allows users to navigate branching contexts and reuse previous results to better support iterative visualization authoring.

Other AI and synthesis-powered tools. Besides LLM-powered tools above, neural semantic parsing [6, 29, 31], and program synthesis-based tools [54] have also been developed to address the visualization challenge. For example, NL4DV [31] and NcNet [25] are natural language interfaces (NLIs) based on recurrent neural networks trained from parallel NL and chart specification corpus that can generate charts from NL queries. NL2Vis [62] and Graphy [6] use semantic parser to extract entities from the user’s NL query and apply program synthesis techniques to compose chart specifications. Unlike LLM-based tools that can generate general purpose python programs to support expressive data transformation and visualization from abstract instructions, semantic parsing based NLIs are less expressive, requiring more concrete descriptions from the user and supporting only limited data transformation. In particular, these tools require tidy data inputs [58], and they do not support transformations like string processing, column derivation and reshaping. While programming-by-examples (PBE) techniques are developed to tackle data reshaping challenges in chart authoring (e.g., Falx [54] and Data Formulator [55]’s reshaping module), these tools require users to prepare low-level examples to demonstrate the transformation intent, which can be difficult for new users as it deviates from the high-level visualization workflow. Unlike LLM-based tools where the user can directly have conversation with the model to disambiguate inputs, semantic parsing and PBE-based tools develop special techniques for resolving ambiguous user intent. For example, DataTone [11] introduces disambiguation widgets to allow users to select alternative extracted entities in the generated queries to resolve ambiguity, and it paraphrases the generated query in NL to explain the result. Falx [54] renders charts from multiple versions of data consistent with user examples for user inspection.

Benefit from LLMs, Data Formulator 2 supports a much wider range of data transformation and does not limit inputs to tidy data. Inspired by how prior work displays candidate results and explains code to help users understand system outputs [11, 12, 55], Data Formulator 2 displays generated code, data, chart and code explanation to assist user inspection. To resolve ambiguous outputs, the user can use data threads to follow up or backtrack and revise their instructions.

Visualization grammars and interactive tools. The grammar of graphics [60] inspired many modern visualization grammars (e.g., ggplot2 [57], Vega-Lite [45], Altair [53]), where visualizations are built from mapping data columns to visual channels and lower-level chart properties. Comparing to more flexible and expressive languages like D3 [3] and Atlas [22], high-level grammars hide the computation process of linking data items to visual objects to reduce visualization efforts. Powered by these high-level grammars, interactive tools like Lyra [44], Data Illustrator [23], Charticulator [40], Tableau [47]) are introduced. With the shelf-configuration interface, users of these tools specify chart designs by mapping data columns to visual encoding “shelves” using the intuitive drag-and-drop interaction. Despite these tools make chart specification easy, they require tidy input data [58]: every variable to be visualized should be organized in a column in the input data, and every mark should come from a row. Thus, users need skills and additional efforts to prepare data in the right format with data transformation tools [7, 15–18, 35–37, 59].

Data Formulator 2 internally represents charts using Vega-Lite, and it benefits from Vega-Lite’s expressiveness to support rich visualization designs. Data Formulator 2 inherits the shelf-configuration design from existing

interactive tools and blends it with NL inputs for chart specification. This way, users can specify chart designs easily with UI in Data Formulator 2, yet they do not need to worry about data transformation, as Data Formulator 2 delegates data transformation to AI.

Exploration history. Graphical history [13] and data provenance [4] are essential in visualization authoring, especially in exploration tasks where branching and iterations are common. In computation notebooks, the exploration history is organized based on code blocks [28, 32]. Data transformation tools like somnus [63] and Tableau Prep visualize data provenance based on transformation operators. Directed-graph models [19, 46] based on visual similarity are also used for visualization organization. Data Formulator 2’s data threads draws inspirations from these systems. The key difference is that Data Formulator 2 organizes history around high-level user interactions with AI and hides operator-level details to enhance navigation and reuse. In future, Data Formulator 2 could render data threads as hierarchical trees [19] to support navigation of large data threads in multiple granularity.

Multi-modal interaction. Despite natural language provides flexible and expressive interaction between human and AI, NL-only interaction is not always optimal for the users to clearly convey their intent, especially for conveying designs they pictured in their mind. To address this limitation, multi-modal models like ChatGPT [1] and Gemini [38] are introduced, allowing users to provide audios and images in their conversation with AI. New interactive tools are also developed to support multi-modal interaction. For example, DirectGPT [27] allows users to direct point and click on a canvas to specify contexts or objects that NL instruction is based on to reduce prompting efforts, DynaVis [52] generates UI widgets dynamically based on user’s NL inputs for chart editing so that they can explore and repeat edits and see instant visual feedback from edits. Data Formulator 2’s concept encoding shelf bridges the precision and affordance of GUI interaction with flexibility of NL inputs, and it contributes a sample of multi-modal UI design for visualization authoring.

Others. Data Formulator 2 focuses on visualization authoring, where AI completes tasks planned by the user. There are potential to combine Data Formulator 2 with exploration and recommendation systems like Voyager [61], Draco [30], and Lux [20] for suggesting visualization goals to assist users “cold-start” their analysis. Data Formulator 2 currently focuses on grammar-of-graphics-based charts (provided by Vega-Lite), and provides limited supports of custom chart designs (e.g., new layouts, interaction, animation, or annotation). Data Formulator 2’s data transformation and history management could also assist animation designs [68] and interactive visualization authoring [67]. Data Formulator 2 in future can incorporate canvases for editing layouts [43, 51] and marks [40] to expand its design space.

6 Conclusion

Visualization authors often create visualizations in an iterative fashion, going back and forth between data transformation and visualization steps. To achieve such iterative analysis process, authors not only needs to be proficient with data transformation and visualization tools, but also needs to spend considerable efforts managing the branching history consisting of many different versions of data and charts. Despite AI-powered tools have been developed to reduce users efforts, they do not work well for iterative analysis, because they often expect users to specify their intent all at once with only NL inputs. We presented Data Formulator 2, an interactive system for iterative creation of rich visualizations. Data Formulator 2 features a multi-modal UI that lets users to specify visualization with blended UI and NL inputs. Benefiting from both the precision of UI interaction and expressiveness of NL descriptions, users can more precisely convey complex designs without verbose prompts. To support management of the iteration history, Data Formulator 2 introduces data threads, where users can navigate, branch and reuse previous designs towards new ones as opposed to creating everything from scratch. In the user study, we invited eight participants to reproduce two challenging data exploration sessions consisting of 16

visualizations. We observed that Data Formulator 2 let participants develop their own iteration and verification strategies to solve the task with confidence with minimal hints.

7 Acknowledgement

We would like to thank John Thompson, Jeevana Priya Inala, Dave Brown, Gonzalo Ramos, and Kori Inkpen's contributions and feedback to this work.

References

- [1] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774* (2023).
- [2] Shraddha Barke, Michael B James, and Nadia Polikarpova. 2023. Grounded copilot: How programmers interact with code-generating models. *Proceedings of the ACM on Programming Languages* 7, OOPSLA1 (2023), 85–111.
- [3] Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. 2011. D³ Data-Driven Documents. *IEEE Trans. Vis. Comput. Graph.* 17, 12 (2011), 2301–2309. <https://doi.org/10.1109/TVCG.2011.185>
- [4] Peter Buneman, Sanjeev Khanna, and Tan Wang-Chiew. 2001. Why and where: A characterization of data provenance. In *Database Theory—ICDT 2001: 8th International Conference London, UK, January 4–6, 2001 Proceedings* 8. Springer, 316–330.
- [5] Mark Chen, Jerry Tworek, Heewoo Jun, Qiming Yuan, Henrique Pondé de Oliveira Pinto, Jared Kaplan, Harrison Edwards, Yuri Burda, Nicholas Joseph, Greg Brockman, et al. 2021. Evaluating Large Language Models Trained on Code. *CoRR* abs/2107.03374 (2021). arXiv:2107.03374 <https://arxiv.org/abs/2107.03374>
- [6] Qiaochu Chen, Shankara Pailoor, Celeste Barnaby, Abby Criswell, Chenglong Wang, Greg Durrett, and Isil Dillig. 2022. Type-directed synthesis of visualizations from natural language queries. *Proceedings of the ACM on Programming Languages* 6, OOPSLA2 (2022), 532–559.
- [7] Qiaochu Chen, Xinyu Wang, Xi Ye, Greg Durrett, and Isil Dillig. 2020. Multi-modal synthesis of regular expressions. In *Proceedings of the 41st ACM SIGPLAN conference on programming language design and implementation*. 487–502.
- [8] Xinyun Chen, Maxwell Lin, Nathanael Schärlí, and Denny Zhou. 2023. Teaching large language models to self-debug. *arXiv preprint arXiv:2304.05128* (2023).
- [9] Victor Dibia. 2023. LIDA: A Tool for Automatic Generation of Grammar-Agnostic Visualizations and Infographics using Large Language Models. *arXiv preprint arXiv:2303.02927* (2023).
- [10] Victor Dibia and Çağatay Demiralp. 2019. Data2vis: Automatic generation of data visualizations using sequence-to-sequence recurrent neural networks. *IEEE computer graphics and applications* 39, 5 (2019), 33–46.
- [11] Tong Gao, Mira Dontcheva, Eytan Adar, Zhicheng Liu, and Karrie G. Karahalios. 2015. DataTone: Managing Ambiguity in Natural Language Interfaces for Data Visualization. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology, UIST 2015, Charlotte, NC, USA, November 8–11, 2015*, Celine Latulipe, Bjoern Hartmann, and Tovi Grossman (Eds.). ACM, 489–500. <https://doi.org/10.1145/2807442.2807478>
- [12] Ken Gu, Ruoxi Shang, Tim Althoff, Chenglong Wang, and Steven M Drucker. 2023. How Do Analysts Understand and Verify AI-Assisted Data Analyses? *arXiv preprint arXiv:2309.10947* (2023).
- [13] Jeffrey Heer, Jock Mackinlay, Chris Stolte, and Maneesh Agrawala. 2008. Graphical histories for visualization: Supporting analysis, communication, and evaluation. *IEEE transactions on visualization and computer graphics* 14, 6 (2008), 1189–1196.
- [14] Cheng-Ping Hsieh, Simeng Sun, Samuel Kriman, Shantanu Acharya, Dima Rekesh, Fei Jia, and Boris Ginsburg. 2024. RULER: What's the Real Context Size of Your Long-Context Language Models? *arXiv preprint arXiv:2404.06654* (2024).
- [15] Yanwei Huang, Yunfan Zhou, Ran Chen, Changhao Pan, Xinhuan Shu, Di Weng, and Yingcai Wu. 2023. Interactive table synthesis with natural language. *IEEE Transactions on Visualization and Computer Graphics* (2023).
- [16] Zhongjun Jin, Michael R. Anderson, Michael J. Cafarella, and H. V. Jagadish. 2017. Foofah: Transforming Data By Example. In *SIGMOD Conference 2017, Chicago, IL, USA, May 14–19, 2017*, Semih Salihoglu, Wenchao Zhou, Rada Chirkova, Jun Yang, and Dan Suciu (Eds.). ACM, 683–698. <https://doi.org/10.1145/3035918.3064034>
- [17] Sean Kandel, Andreas Paepcke, Joseph Hellerstein, and Jeffrey Heer. 2011. Wrangler: Interactive visual specification of data transformation scripts. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*. 3363–3372. <https://doi.org/10.1145/1978942.1979444>
- [18] Mary Beth Kery, Donghao Ren, Fred Hohman, Dominik Moritz, Kanit Wongsuphasawat, and Kayur Patel. 2020. Mage: Fluid Moves Between Code and Graphical Work in Computational Notebooks. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology (Virtual Event, USA) (UIST '20)*. Association for Computing Machinery, New York, NY, USA, 140–151. <https://doi.org/10.1145/3379337.3415842>

- [19] Younghoon Kim, Kanit Wongsuphasawat, Jessica Hullman, and Jeffrey Heer. 2017. Graphscape: A model for automated reasoning about visualization similarity and sequencing. In *Proceedings of the 2017 CHI conference on human factors in computing systems*. 2628–2638.
- [20] Doris Jung Lin Lee, Dixin Tang, Kunal Agarwal, Thyne Boonmark, Caitlyn Chen, Jake Kang, Ujjaini Mukhopadhyay, Jerry Song, Micah Yong, Marti A. Hearst, and Aditya G. Parameswaran. 2021. Lux: Always-on Visualization Recommendations for Exploratory Dataframe Workflows. *Proc. VLDB Endow.* 15, 3 (2021), 727–738. <https://doi.org/10.14778/3494124.3494151>
- [21] Nelson F Liu, Kevin Lin, John Hewitt, Ashwin Paranjape, Michele Bevilacqua, Fabio Petroni, and Percy Liang. 2024. Lost in the middle: How language models use long contexts. *Transactions of the Association for Computational Linguistics* 12 (2024), 157–173.
- [22] Zhicheng Liu, Chen Chen, Francisco Morales, and Yishan Zhao. 2021. Atlas: Grammar-based Procedural Generation of Data Visualizations. In *2021 IEEE Visualization Conference, 2021 - Short Papers, New Orleans, LA, USA, October 24-29, 2021*. IEEE, 171–175. <https://doi.org/10.1109/VIS49827.2021.9623315>
- [23] Zhicheng Liu, John Thompson, Alan Wilson, Mira Dontcheva, James Delorey, Sam Grigg, Bernard Kerr, and John Stasko. 2018. Data Illustrator: Augmenting Vector Design Tools with Lazy Data Binding for Expressive Visualization Authoring. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*. 123:1–13. <https://doi.org/10.1145/3173574.3173697>
- [24] Anton Lozhkov, Raymond Li, Loubna Ben Allal, Federico Cassano, Joel Lamy-Poirier, Nouamane Tazi, Ao Tang, Dmytro Pykhtiar, Jiawei Liu, Yuxiang Wei, et al. 2024. Starcoder 2 and the stack v2: The next generation. *arXiv preprint arXiv:2402.19173* (2024).
- [25] Yuyu Luo, Nan Tang, Guoliang Li, Jiawei Tang, Chengliang Chai, and Xuedi Qin. 2022. Natural Language to Visualization by Neural Machine Translation. *IEEE Trans. Vis. Comput. Graph.* 28, 1 (2022), 217–226. <https://doi.org/10.1109/TVCG.2021.3114848>
- [26] Paula Maddigan and Teo Susnjak. 2023. Chat2vis: Generating data visualisations via natural language using chatgpt, codex and gpt-3 large language models. *Ieee Access* (2023).
- [27] Damien Masson, Sylvain Malacria, Géry Casiez, and Daniel Vogel. 2024. Directgpt: A direct manipulation interface to interact with large language models. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–16.
- [28] Andrew M McNutt, Chenglong Wang, Robert A Deline, and Steven M Drucker. 2023. On the design of ai-powered code assistants for notebooks. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–16.
- [29] Rishab Mitra, Arpit Narechania, Alex Endert, and John Stasko. 2022. Facilitating conversational interaction in natural language interfaces for visualization. In *2022 IEEE Visualization and Visual Analytics (VIS)*. IEEE, 6–10.
- [30] Dominik Moritz, Chenglong Wang, Greg L. Nelson, Halden Lin, Adam M. Smith, Bill Howe, and Jeffrey Heer. 2019. Formalizing Visualization Design Knowledge as Constraints: Actionable and Extensible Models in Draco. *IEEE Trans. Vis. Comput. Graph.* 25, 1 (2019), 438–448. <https://doi.org/10.1109/TVCG.2018.2865240>
- [31] Arpit Narechania, Arjun Srinivasan, and John T. Stasko. 2021. NL4DV: A Toolkit for Generating Analytic Specifications for Data Visualization from Natural Language Queries. *IEEE Trans. Vis. Comput. Graph.* 27, 2 (2021), 369–379. <https://doi.org/10.1109/TVCG.2020.3030378>
- [32] Observable. [n. d.]. <https://observablehq.com/>
- [33] Theo X Olausson, Jeevana Priya Inala, Chenglong Wang, Jianfeng Gao, and Armando Solar-Lezama. 2023. Is Self-Repair a Silver Bullet for Code Generation?. In *The Twelfth International Conference on Learning Representations*.
- [34] OpenAI, Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, and Sam Altman et al. 2024. GPT-4 Technical Report. *arXiv:2303.08774 [cs.CL]* <https://arxiv.org/abs/2303.08774>
- [35] The pandas development team. 2023. *pandas-dev/pandas: Pandas*. <https://doi.org/10.5281/zenodo.7741580>
- [36] Oleksandr Polozov and Sumit Gulwani. 2015. FlashMeta: a framework for inductive program synthesis. In *Object-Oriented Programming, Systems, Languages, and Applications, OOPSLA 2015, Pittsburgh, PA, USA, October 25-30, 2015*, Jonathan Aldrich and Patrick Eugster (Eds.). ACM, 107–126. <https://doi.org/10.1145/2814270.2814310>
- [37] Vijayshankar Raman and Joseph M. Hellerstein. 2001. Potter's Wheel: An Interactive Data Cleaning System. In *VLDB 2001, Proceedings of 27th International Conference on Very Large Data Bases, September 11-14, 2001, Roma, Italy*, Peter M. G. Apers, Paolo Atzeni, Stefano Ceri, Stefano Paraboschi, Kotagiri Ramamohanarao, and Richard T. Snodgrass (Eds.). Morgan Kaufmann, 381–390. <http://www.vldb.org/conf/2001/P381.pdf>
- [38] Machel Reid, Nikolay Savinov, Denis Teplyashin, Dmitry Lepikhin, Timothy Lillicrap, Jean-baptiste Alayrac, Radu Soricu, Angeliki Lazaridou, Orhan Firat, Julian Schrittwieser, et al. 2024. Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context. *arXiv preprint arXiv:2403.05530* (2024).
- [39] Donghao Ren, Matthew Brehmer, Bongshin Lee, Tobias Höllerer, and Eun Kyung Choe. 2017. Chartaccent: Annotation for data-driven storytelling. In *2017 IEEE Pacific Visualization Symposium (PacificVis)*. Ieee, 230–239.
- [40] Donghao Ren, Bongshin Lee, and Matthew Brehmer. 2019. Charticulator: Interactive Construction of Bespoke Chart Layouts. *IEEE Trans. Vis. Comput. Graph. (Proceedings of InfoVis)* 25, 1 (2019). <https://doi.org/10.1109/TVCG.2018.2865158>
- [41] Donghao Ren, Bongshin Lee, Matthew Brehmer, and Nathalie Henry Riche. 2018. Reflecting on the Evaluation of Visualization Authoring Systems : Position Paper. In *2018 IEEE Evaluation and Beyond - Methodological Approaches for Visualization, BELIV 2018, Berlin, Germany, October 21, 2018*, Michael Sedlmair, Petra Isenberg, Miriah Meyer, and Tobias Isenberg (Eds.). IEEE Computer Society, 86–92. <https://doi.org/10.1109/BELIV.2018.8634297>

- [42] Adam Rule, Aurélien Tabard, and James D Hollan. 2018. Exploration and explanation in computational notebooks. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–12.
- [43] Bahador Saket, Hannah Kim, Eli T. Brown, and Alex Endert. 2017. Visualization by Demonstration: An Interaction Paradigm for Visual Data Exploration. *IEEE Trans. Vis. Comput. Graph.* 23, 1 (2017), 331–340. <https://doi.org/10.1109/TVCG.2016.2598839>
- [44] A. Satyanarayan and J. Heer. 2014. Lyra: An interactive visualization design environment. *Computer Graphics Forum (Proceedings of EuroVis)* 33, 3 (2014). <https://doi.org/10.1111/cgf.12391>
- [45] Arvind Satyanarayan, Dominik Moritz, Kanit Wongsuphasawat, and Jeffrey Heer. 2017. Vega-Lite: A grammar of interactive graphics. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis)* 23, 1 (2017), 341–350. <https://doi.org/10.1109/TVCG.2016.2599030>
- [46] Danqing Shi, Yang Shi, Xinyue Xu, Nan Chen, Siwei Fu, Hongjin Wu, and Nan Cao. 2019. Task-oriented optimal sequencing of visualization charts. In *2019 IEEE Visualization in Data Science (VDS)*. IEEE, 58–66.
- [47] Chris Stolte, Diane Tang, and Pat Hanrahan. 2002. Query, analysis, and visualization of hierarchically structured data using Polaris. In *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, July 23–26, 2002, Edmonton, Alberta, Canada*. ACM, 112–122. <https://doi.org/10.1145/775047.775064>
- [48] Lev Tankelevitch, Viktor Kewenig, Auste Simkute, Ava Elizabeth Scott, Advait Sarkar, Abigail Sellen, and Sean Rintel. 2024. The Metacognitive Demands and Opportunities of Generative AI. In *Proceedings of the CHI Conference on Human Factors in Computing Systems, CHI 2024, Honolulu, HI, USA, May 11–16, 2024*, Florian 'Floyd' Mueller, Penny Kyburz, Julie R. Williamson, Corina Sas, Max L. Wilson, Phoebe O. Toups Dugas, and Irina Shklovski (Eds.). ACM, 680:1–680:24. <https://doi.org/10.1145/3613904.3642902>
- [49] Yuan Tian, Weiwei Cui, Dazhen Deng, Xinjing Yi, Yurun Yang, Haidong Zhang, and Yingcai Wu. 2024. Chartgpt: Leveraging llms to generate charts from abstract natural language. *IEEE Transactions on Visualization and Computer Graphics* (2024).
- [50] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. 2023. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971* (2023).
- [51] Theophanis Tsandilas. 2021. StructGraphics: Flexible Visualization Design through Data-Agnostic and Reusable Graphical Structures. *IEEE Trans. Vis. Comput. Graph.* 27, 2, 315–325. <https://doi.org/10.1109/TVCG.2020.3030476>
- [52] Priyan Vaithilingam, Elena L Glassman, Jeevana Priya Inala, and Chenglong Wang. 2024. DynaVis: Dynamically Synthesized UI Widgets for Visualization Editing. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–17.
- [53] Jacob VanderPlas, Brian E. Granger, Jeffrey Heer, Dominik Moritz, Kanit Wongsuphasawat, Arvind Satyanarayan, Eitan Lees, Ilia Timofeev, Ben Welsh, and Scott Sievert. 2018. Altair: Interactive Statistical Visualizations for Python. *J. Open Source Softw.* 3, 32, 1057. <https://doi.org/10.21105/joss.01057>
- [54] Chenglong Wang, Yu Feng, Rastislav Bodík, Isil Dillig, Alvin Cheung, and Amy J. Ko. 2021. Falx: Synthesis-Powered Visualization Authoring. In *CHI '21: CHI Conference on Human Factors in Computing Systems, Virtual Event / Yokohama, Japan, May 8–13, 2021*, Yoshifumi Kitamura, Aaron Quigley, Katherine Isbister, Takeo Igarashi, Pernille Bjørn, and Steven Mark Drucker (Eds.). ACM, 106:1–106:15. <https://doi.org/10.1145/3411764.3445249>
- [55] Chenglong Wang, John Thompson, and Bongshin Lee. 2023. Data Formulator: Ai-powered concept-driven visualization authoring. *IEEE Transactions on Visualization and Computer Graphics* (2023).
- [56] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems* 35 (2022), 24824–24837.
- [57] Hadley Wickham. 2009. *ggplot2 - Elegant Graphics for Data Analysis*. Springer. <https://doi.org/10.1007/978-0-387-98141-3>
- [58] Hadley Wickham. 2014. Tidy data. *The Journal of Statistical Software* 59 (2014). Issue 10. <http://www.jstatsoft.org/v59/i10/>
- [59] Hadley Wickham, Mara Averick, Jennifer Bryan, Winston Chang, Lucy McGowan, Romain François, Garrett Grolemund, Alex Hayes, Lionel Henry, Jim Hester, Max Kuhn, Thomas Pedersen, Evan Miller, Stephan Bache, Kirill Müller, Jeroen Ooms, David Robinson, Dana Seidel, Vitalie Spinu, Kohske Takahashi, Davis Vaughan, Claus Wilke, Kara Woo, and Hiroaki Yutani. 2019. Welcome to the tidyverse. *J. Open Source Softw.* 4, 43 (Nov. 2019), 1686. <https://doi.org/10.21105/joss.01686>
- [60] Leland Wilkinson. 2005. *The Grammar of Graphics, Second Edition*. Springer.
- [61] Kanit Wongsuphasawat, Zening Qu, Dominik Moritz, Riley Chang, Felix Ouk, Anushka Anand, Jock Mackinlay, Bill Howe, and Jeffrey Heer. 2017. Voyager 2: Augmenting Visual Analysis with Partial View Specifications. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) (CHI '17). Association for Computing Machinery, New York, NY, USA, 2648–2659. <https://doi.org/10.1145/3025453.3025768>
- [62] Zhengkai Wu, Vu Le, Ashish Tiwari, Sumit Gulwani, Arjun Radhakrishna, Ivan Radiček, Gustavo Soares, Xinyu Wang, Zhenwen Li, and Tao Xie. 2022. NL2Viz: natural language to visualization via constrained syntax-guided synthesis. In *Proceedings of the 30th ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering*. 972–983.
- [63] Kai Xiong, Siwei Fu, Guoming Ding, Zhongsu Luo, Rong Yu, Wei Chen, Hujun Bao, and Yingcai Wu. 2022. Visualizing the scripts of data wrangling with SOMNUS. *IEEE Transactions on Visualization and Computer Graphics* (2022).

- [64] JD Zamfirescu-Pereira, Richmond Y Wong, Bjoern Hartmann, and Qian Yang. 2023. Why Johnny can't prompt: how non-AI experts try (and fail) to design LLM prompts. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–21.
- [65] Qingru Zhang, Chandan Singh, Liyuan Liu, Xiaodong Liu, Bin Yu, Jianfeng Gao, and Tuo Zhao. 2023. Tell your model where to attend: Post-hoc attention steering for llms. *arXiv preprint arXiv:2311.02262* (2023).
- [66] Tianyu Zheng, Ge Zhang, Tianhao Shen, Xueling Liu, Bill Yuchen Lin, Jie Fu, Wenhui Chen, and Xiang Yue. 2024. OpenCodeInterpreter: Integrating Code Generation with Execution and Refinement. *arXiv preprint arXiv:2402.14658* (2024).
- [67] Jonathan Zong, Dhiraj Barnwal, Rupayan Neogy, and Arvind Satyanarayan. 2021. Lyra 2: Designing Interactive Visualizations by Demonstration. *IEEE Trans. Vis. Comput. Graph.* 27, 2 (2021), 304–314. <https://doi.org/10.1109/TVCG.2020.3030367>
- [68] Jonathan Zong, Josh Pollock, Dylan Wootton, and Arvind Satyanarayan. 2022. Animated Vega-Lite: Unifying animation with a grammar of interactive graphics. *IEEE Transactions on Visualization and Computer Graphics* 29, 1 (2022), 149–159.