Unveiling How Examples Shape Visualization Design Outcomes

Hannah K. Bako, Xinyi Liu, Grace Ko, Hyemi Song, Leilani Battle, and Zhicheng Liu

Abstract—Visualization designers (e.g., journalists or data analysts) often rely on examples to explore the space of possible designs, yet we have little insight into how examples shape data visualization design outcomes. While the effects of examples have been studied in other disciplines, such as web design or engineering, the results are not readily applicable to visualization due to inconsistencies in findings and challenges unique to visualization design. Towards bridging this gap, we conduct an exploratory experiment involving 32 data visualization designers focusing on the influence of five factors (*timing*, *quantity*, *diversity*, data *topic similarity*, and data *schema similarity*) on objectively measurable design outcomes (e.g., *numbers of designs* and *idea transfers*). Our quantitative analysis shows that when examples are introduced after initial brainstorming, designers curate examples with topics less similar to the dataset they are working on and produce more designs with a high variation in visualization components. Also, designers copy more ideas from examples with higher data schema similarities. Our qualitative analysis of participants' thought processes provides insights into why designers incorporate examples into their designs, revealing potential factors that have not been previously investigated. Finally, we discuss how our results inform how designers may use examples during design ideation as well as future research on quantifying designs and supporting example-based visualization design. All supplemental materials are available in our OSF repo.

Index Terms—data visualization, design, examples

1 Introduction

Data visualization designers often seek inspiration by recalling interesting designs they have encountered in the past or actively searching for new visualization examples [9, 11, 53]. In this work, we use the term "designer" to broadly refer to individuals who create data visualizations for various analytical and communicative purposes, including but not limited to journalists, business managers, and scientists [65, 72]. Little work has studied the influence of examples on data visualization design outcomes even though the influence of examples have been studied in other domains such as graphic or web design [35, 55] and engineering [5,21,34,59]. These works suggest that while examples can improve the number, quality, and novelty of design outcomes [51, 76], there is also a risk that designers may prematurely fixate on inappropriate or irrelevant examples [60, 76] leading to ineffective designs [9].

In particular, previous studies have identified several factors that modulate the influence of examples on design outcomes, including the *quantity and quality* of examples [76], *commonness* of an example (i.e., if the example is a familiar solution to a problem) [21, 61, 76], and the *timing* of example introduction during design ideation [51, 56]. However, the results from these studies do not readily transfer to data visualization design. First, these prior studies have conflicting results. For instance, there is a lack of consensus on when to introduce examples [22,41,75] during the design process. Furthermore, visualization design presents unique challenges and complexities that are not shared by other disciplines. Visualization designs are constrained by the data to be presented, and designers struggle to anticipate how real data may impact visual forms, constrain design choices, and lead to unexpected edge cases [79]. Thus, there is a need for empirical data that sheds light on the influence of examples in the context of data visualization. Specifically, we need to assess *how factors identified in prior studies, alongside those unique to visualization, modulate the influence of examples on design outcomes*. Additionally, we need to understand *how designers decide which aspects of examples to incorporate into their designs.*

- *Hannah K. Bako, Hyemi Song, Grace Ko, and Zhicheng Liu are with the University of Maryland. E-mail: {hbako, gko1, hsong02, leozcliu}@cs.umd.edu.*
- *Xinyi Liu is with the University of Texas. E-mail: xinyi.liu@utexas.edu.*
- *Leilani Battle is with the University of Washington. E-mail: leibatt@cs.washington.edu.*

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

Based on these considerations, **we contribute an exploratory study with 32 visualization designers** to understand the influence of examples in data visualization design. The study loosely follows the protocol of Smith et al.'s seminal work evaluating the influence of examples [77], where designers are given a design task with accompanying examples and asked to brainstorm design ideas. However, we make some strategic adaptations to our study design. First we use a faceted browsing interface for participants to explore and curate relevant examples, allowing us to examine multiple factors (e.g., quantity, commonness, etc.) in the same study. Secondly, we include a think-aloud protocol to capture deep contextual information about designers' thought processes during design ideation (details on our study design are discussed in Sec. 3). Our analysis focuses on quantifying the features of an example and using this information to generate metrics that describe the collective properties of example sets curated by participants (Sec. 4). We use quantitative methods to measure how properties of curated example sets influence objectively measurable design outcomes like the number and variety of designs and ideas copied from examples (Sec. 5). Finally, we perform qualitative coding to gain insights into designers' rationale for selecting examples (Sec. 6).

Our results show that 1) the timing of example introduction potentially influences the types of examples designers curate and the number and variety of designs produced. 2) a higher similarity between the data schema in examples and the data schema a designer wants to present, leads to more ideas transferred from examples into designs. These results are supported by our qualitative evaluations highlighting *two forms of idea transfer, i.e., partial transfer and design replication*. Finally, we identify salient factors that attract designers to use examples or deter them from doing so, such as the underlying visual tasks (e.g., comparison, distribution), visual composition, design complexity, and the similarity between examples and prior design ideas. Our findings shed light on results from prior work and introduce new factors that influence design outcomes, giving rise to guidance on how designers may incorporate examples into the design process and avenues for further research on using examples for visualization design.

2 Background and Related Work

Extensive research in fields such as psychology, design, and mechanical engineering have highlighted the challenges of finding and extracting ideas from relevant examples [6, 32, 35, 55] and explored how examples affect the ideas designers create [28, 32, 37, 55, 64, 76, 77]. Collectively, these works show that examples significantly affect creativity: some report examples as a source of inspiration [9, 28, 32, 55, 60], while others report constraining effects of examples [26, 37, 37, 44, 50, 63, 76]. Together, these works have identified key factors that modulate the effect that an example has on the ideas designers come up with. To aid understanding, we review important factors and aspects of design outcomes identified in prior work below.

2.1 Factors that Modulate an Example's Influence

Timing of Example Introduction. Much research has explored *when* examples should be introduced in a design process. Smith et al. show that introducing a delay between exposure to examples and design ideation does not reduce conformity (the number of features copied from examples) within designs [77]. These findings support long-term retention of examples [74] and suggest that design conformity may be a result of the unintentional use of memory from prior exposure to examples [50, 52]. However, there is no consensus on when examples should be introduced. Kulkarni et al. suggest early exposure to examples [41], while others advocate for on-demand provision of examples [22, 75]. Yet, another thread of research shows that when designers have had time to brainstorm potential solutions before exposure to examples, they can better integrate examples into their solutions instead of conforming their solutions to examples [56,76,78]. Our work extends these emerging findings on the timing of examples introduction by investigating how the time of exposure to examples may lead to meaningful differences in design outcomes.

Diversity of Examples. Evaluations of the examples used in prior work do not explicitly explore the diversity of features in a set of examples. Instead, they focus on evaluating the "commonness" of an example, which refers to how typical an example is as a solution for a given problem [21, 76]. For instance, in data visualization, bar charts are a *common* solution for comparison tasks where one seeks to observe similarities or differences in values across different data groups. Research finds that focusing on common examples, closely related examples, or examples related to the problem domain limits the space of ideas explored, which may lead to fixation [37]. Expanding the space of explored examples to include solutions slightly distant from the problem domain [24,31] or even analogical examples [14,20,21,34] has proven more beneficial for idea generation [63], leading to higher quality and novelty of design solutions [76]. Within visualization design, the heterogeneity of the features within a *collection* of examples is a more meaningful measure compared to the commonness of a single example. Therefore, we focus on the ***diversity*** of examples in this study.

Quantity of Examples. Compared to other properties, there has been less investigation into how the number of examples examined by designers influences designs. Marsh et al. investigated how exposure to different numbers of examples (n = [1, 3, 6, 9]) may increase conformity [51]. Their results show that more examples lead to more critical features copied from examples to designs. Sio and Kotovsky also find a negative correlation between the number of examples and the novelty of design ideas [76]. However, there is no consensus on the appropriate number of examples to provide as experiment stimuli. Recent studies tend to arbitrarily choose the number of examples from an exploration of the solution space [21, 24, 31], while others allow participants to independently curate examples from the web [55].

Quality of Examples. Literature shows that examples induce conformity for designers [37, 51, 77]. However, the quality of examples determines the degree to which conformity enhances or impairs design outcomes. Janson and Smith found that individuals do not only copy ideas from examples that solve the provided problem but also from examples with obvious design flaws [37]. Chrysikou and Weisberg found that individuals copy inappropriate ideas from examples even when instructions explicitly ask them to avoid these inappropriate ideas [26]. Fu et al. also show that introducing poor examples leads to a decrease in the ability to converge on solutions and the quality of solutions [30]. In these works, "poor" examples are designs with problematic features that do not adequately solve the defined design problem. Assessing the quality and effectiveness of visualization designs is highly subjective [19, 62], dependent on many dimensions beyond just encoding choices, such as the context of use and the intended audience. In this work, we focus on objective measures of example properties. As such, we do not measure the quality of examples.

2.2 Metrics for Evaluating Design Outcomes

In previous work design outcomes are often measured by quantitative metrics such as the ***quantity, and conformity*** of designs and subjective evaluation of the ***quality, variety***, and ***novelty*** of designs. The ***conformity*** of designs, as described by Smith et al., is the tendency for designers to copy features of examples into their design ideas [77]. In most studies conformity is measured by assessing the number of example-related ideas duplicated in final designs [76] and is believed to be an indication of the constraining influence of examples [51, 76, 77].

Subjective measures such as the ***quality*** of design outcomes, measure the degree to which a design solves the specified problem [18, 21]. In general, the evaluation of design quality is based entirely on metrics decided upon by researchers, such as ease of use [18], cost feasibility, and build time [21, 31]. Similarly, the ***novelty*** of designs is based on how unique a generated design is relative to design ideas generated by other participants [4, 70, 76, 78]. ***Variety***, on the other hand, measures the explored space of solution ideas, i.e., the range of ideas produced by a single participant [4, 18, 78].

2.3 Factors and Outcomes of Interest in This Study

Previous research identified four key factors that influence design outcomes; our study focuses on three of them. As data informs the design choices a visualization designer makes, we introduced two additional factors relevant to data visualization: the similarities between examples' underlying datasets and the dataset used in the design task in terms of data *topic* and *schema* (i.e., number and types of attributes). These new factors explore whether or not designers curate examples based on closeness in topics or the alignment between the working dataset and an example's data schema. For instance, if a designer is working on the Boston weather dataset [54], which contains temporal and numerical attributes, is the designer more interested in examples visualizing similar topics (e.g., average temperatures or melting ice caps), or do they consider a broader set of examples visualizing temporal and numerical data? Additionally, does focusing on topic or schema similarity influence their designs? This expands the scope of example-related factors we analyze in this study to include the *timing* of example introduction, *diversity*, and *quantity* of examples, and the data topic (*topic_sim.*) and schema (*schema_sim.*) similarities.

Regarding design outcomes, we focus on the more obvious quantitative metrics: the *quantity* and *variety* of designs, as well as the conformity of examples, which we refer to as *idea_transfer*. We do not assess subjective factors such as the quality and novelty of examples, nor do we evaluate the effectiveness or creativity of design outcomes. Assessing visualization designs across these axes remains open research for future work.

3 Study Design

Our work aims to understand how preexisting and visualization-specific factors influence design outcomes and how designers decide what aspects of examples should be incorporated into their designs. In this section, we discuss our experiment design rationale and procedure.

3.1 Experiment Overview

From our survey of prior work, we observed that the study design in most of these works follows the protocol of the foundational study by Smith et al. [77], which examines examples' constraining effects on creativity. In these studies, participants are presented with a design task and various examples (or no examples) and asked to develop designs that fulfill the task. We considered replicating Smith et al.'s experiment [77] with a data visualization design context. However, a straightforward replication is not feasible as the design space for data visualization is much broader than previously studied fields. For instance, in Smith et al.'s study, examples are characterized by simple binary features, (e.g., if a tail is present in the example). In contrast, the features of a visualization example can be analyzed along multiple dimensions, including but not limited to mark choices, visual encodings, layout, and graphical styles. Each dimension has numerous values; for instance, a mark choice includes but is not limited to rectangle, circle, polyline, and area. It is unclear which dimensions and values

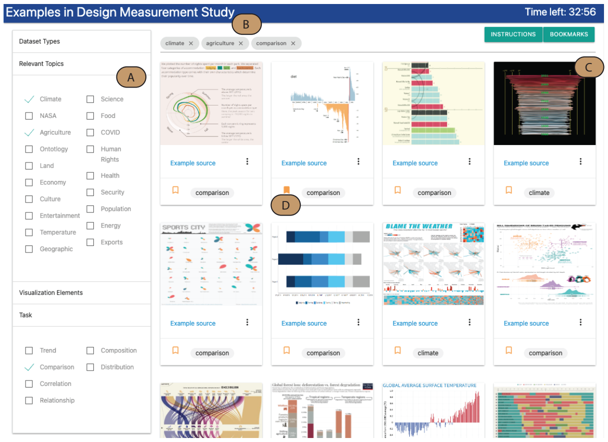


Figure 1: A snapshot of the experiment interface provided to participants during the experiment. (A) select tags to filter examples, (B) search for tags with auto-complete suggestions, (C) browse examples in the main panel, and (D) bookmark relevant examples.

Table 1: Example selection criteria and definitions

Criteria	Description
Knowledge	How well does the visualization express the basic facts about its underlying dataset (i.e. expressiveness)?
Comprehension	How easy is it for a user to understand the information that is being conveyed in the visualization (i.e. effectiveness)?
Relevance	(A) How valuable is the information in the visualization in relation to the experiment task? (B) How closely related is the visualization to the tasks participants will be asked to perform?
Visual Encoding	How appropriate are the selected data encodings in this visualization?
Aesthetics	How attractive is the visual design for the visualization (e.g., bespoke vs common visualization designs, color schemes used)?
Variety	How does this example improve the variety of designs (i.e. number of similar visualization types) in the entire example set?

are crucial in selecting and characterizing the features of examples to present to designers. Additionally, there is no clear guidance from prior work on how many examples should be provided in the experiment.

Consequently, we made strategic changes to our study design based on the following considerations. We have five example-related factors to be examined. Controlling for all these factors in a single experiment would be impractical as this would lead to an exponential combination of experiment conditions. Giving participants the freedom to curate examples can naturally give rise to variations in both the example features (e.g., marks, layouts, encodings, etc.) and factors `quantity`, `diversity`, data `schema_sim.`, and `topic_sim.` This approach is more similar to designers' real-world example curation behavior than providing them with a small set of fixed examples. Moreover, letting designers curate examples from the web with complete freedom can be impractical. In a pilot study with four participants, we discovered that participants were often distracted by information irrelevant to the design task when exploring examples directly from source pages. Thus, we constructed a corpus of 50 diverse examples for the participants to use in the experiment (Sec. 3.3).

We decided on a between-subjects experiment with *two* `timing` *conditions*. We included *two design scenarios* to capture how our results may generalize across multiple datasets and design goals . This results in **four experimental conditions** (Sec. 3.2). We provided an experiment interface for the participants to browse and bookmark relevant examples from the corpus. Participants were asked to explore a given dataset, brainstorm and sketch design ideas for their assigned scenario. These sketches represent the design outcomes of each experiment and were evaluated to identify the `quantiy` and `variety` of the designs and instances of `idea_transfer` from examples.

After brainstorming, participants selected two sketches that they perceived satisfied the design goal, which the experimenter implemented to evaluate the feasibility of the designs. In the pilot study, we found that the time needed to assess designs, perform data transformations, and implement sketched visualizations exceeded the allocated time for the experiment. As a result, the experimenter implemented the designs independently after the experiment. The participants then review the implemented version of their designs in a 15-minute follow-up session. Sec. 3.5 describes the experiment procedure in detail.

3.2 Experiment Conditions

3.2.1 Timing Conditions.

Motivated by a lack of consensus in prior work on when examples should be introduced into the design process, we evaluate two `timing` conditions in our experiment. Particularly, we investigate if a delayed exposure to examples leads to a meaningful difference in the types of examples curated by participants and the number of ideas transferred from examples into designs. Hence, we include a baseline condition (T_{srt}), where participants were provided with the example corpus at the start of the brainstorming process. In the treatment condition, the introduction of the examples was delayed (T_{del}); participants were first asked to brainstorm design ideas for 15 minutes, then presented with the example corpus, and then asked to continue brainstorming.

3.2.2 Design Scenarios.

We included two design scenarios to capture how our results may generalize across design goals and datasets. In selecting datasets, we opt for those not commonly used in data analysis or visualization training to ensure that participants are not inadvertently primed toward certain visualization designs from prior exposure to datasets.

Blog Post Design Scenario (S_1): The first scenario is to design a visualization as the central figure for a blog post discussing the effects of average temperature warming on the US climate [1]. Participants were provided with a dataset on the yearly average weather recordings from thousands of weather stations around the US from 1921 to 2022 [29]. Since our focus in this study is on the design process and not data analysis, we also included two insights on the rate of warming across the decades. Participants were then asked to come up with designs that communicate these insights using the provided dataset.

Bespoke Design Scenario (S_2): The second scenario focuses on the design of highly customized visualizations, which we describe as bespoke visualization designs. Participants assigned to this condition were given a dataset on vegetable crop production land use, deforestation drivers in Indonesia, and expected yield for vegetable crops [71]. Similar to participants in S_1, participants were given insights on the activities that are drivers of deforestation in Indonesia and palm oil land use to yield ratio. Participants were then asked to create a visualization highlighting these insights. For this condition, we varied the design task by prompting participants to think creatively instead of focusing on communication as in S_1. Participants were informed that their designs would be entered into the Iron Viz competition and would be judged on both the novelty and efficacy of their designs.

3.3 Examples Corpus and Interface

3.3.1 Corpus Construction

We searched for visualization designs shared online on popular blogs and resources such as Our World in Data [2], Storytelling with Data [3], and Information is Beautiful [1]. Informed by a subset of the evaluation criteria proposed by Burns et al. [17], we evaluated candidate examples across six key criteria: expressed knowledge, comprehension, relevance, encoding choices, aesthetics, and the variety of designs in the entire corpus (see Tab. 1). Two of the authors sampled a total of

[1] Due to space considerations, the prompts, datasets, and stimuli examples are provided in the supplementary materials.

50 visualizations. Each selected example was evaluated to ensure it was easy to understand (comprehension), expressive, and had appropriate visual encodings. To allow for serendipity, we optimized our selection process to include standard (e.g., bar charts, bubble plots) and customized (e.g., parallel coordinates plot) visualizations.

Tagging Stimuli Examples. To support easy filtering, two authors manually assigned a set of tags that describe each example's features based on the four search criteria used by visualization designers identified by Bako et al [9]. These tags capture metadata that describe: 1) the underlying dataset characteristics (e.g., time series, categories), 2) the visual elements in each example (e.g., points, annotation), 3) the visualization task (e.g., comparison, distribution), and 4) the topic (e.g., climate, human rights).

3.3.2 Experiment Interface

We developed a web-based tool to instrument our experiment, shown in Fig. 1. Our tool consists of a faceted search interface that allows users to filter and browse visualization examples. Users can either choose tags from the side panel (Fig. 1A) or type in a tag in the search box (Fig. 1B). The main view of the interface (Fig. 1C) presents users with examples matching the selected tags. Each example is contained in a single card with a thumbnail, a short description, and a link to see an enlarged view of the visualization. Users could bookmark examples by toggling the bookmark icon next to each example (Fig. 1D). We logged users' interactions (e.g., clicks, hovers).

3.4 Participants

We recruited participants via forums such as Reddit's r/dataisbeautiful, academic and professional mailing lists, and posts on X (previously Twitter). We selected 32 participants (Female=10, Male=22) between the ages of 18 and 44. Participants had between 3 months to 5+ years of experience creating visualizations. Participants reported creating visualizations for a variety of purposes: clients (n=12), blog posts and media graphics (n=7), personal portfolios and projects (n=13), data analysis (n=22), and design studies (n=12). 12 of our participants created visualizations weekly while others created visualizations less than once a week (n=9) or less than once a month (n=11)

To ensure that participants had experience with a wide range of visualization types, we collected information on the designs participants had created in the past, the tools used to implement visualizations, and how often they had created or used them to design visualizations. We selected participants with experience creating visualizations outside of the three most popular visualization types (i.e., bar charts, line charts, and scatterplots) [10] and had used different types of visualization design tools such as direct manipulation tools (e.g., Tableau, PoweBI), programming tools (e.g., ggplot, D3.js) or design focused tools (e.g., Adobe Illustrator, Canva). Selected participants created a wide array of charts ranging from simple charts like bar charts, line charts, and scatterplots (n=32 for all three) to more complex charts like Chord Diagrams (n=9), Sunbursts (n=13) and Parallel Coordinate charts (n=17). Participants also used a range of tools to create visualizations, such as D3 (n=16), Matplotlib (n=22), Adobe Illustrator (n=17), and Microsoft Excel (n=27). The complete participant demographic information is included in our supplementary materials.

3.5 Experiment Procedure

Participants were first given a brief introduction to the purpose and objectives of our study, the nature of the tasks they would be asked to complete, and an overview of their rights. Participants were then asked to complete the consent form, after which the experimenter provided participants with a link to the experiment tool and a Google Jamboard for their sketches. Participants were randomly assigned to one of our four conditions ($T_{srt} \times S_1$, $T_{del} \times S_1$, $T_{srt} \times S_2$, and $T_{del} \times S_2$). Each participant was presented with their assigned design scenario and was given 5 minutes to explore the given dataset in a linked Google sheet. Once the 5 minutes were over, the participants were prompted to return to the interface and asked to brainstorm and sketch design ideas for 35 minutes. Participants in the baseline timing (T_{srt}) condition were allowed access to the experiment interface immediately after the dataset

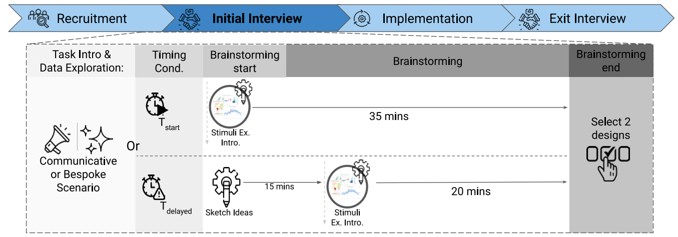


Figure 2: A depiction of our experiment protocol focusing on the initial interview. After familiarizing themselves with the scenario description and dataset, participants begin the brainstorming process. Depending on the `timing` condition, participants are either provided with examples at the start of the process (T_{srt}) or asked to brainstorm ideas before introducing examples(T_{del}).

exploration session. Participants in the delayed timing condition (T_{del}) were asked to begin the brainstorming activity immediately after the dataset exploration session. After 15 minutes of brainstorming, participants in the delayed condition were given access to the experiment interface. Once participants were in the experiment interface, they could explore and interact with examples at their discretion. We asked the participants to bookmark relevant examples and to think aloud.

Once the brainstorming session was complete, the experimenter asked the participants to select their top two designs. The experimenter then implemented these two designs asynchronously using Tableau, Data Illustrator [46], D3 [16], Vega-Lite [73], Charticulator [69], or Microsoft Excel at the experimenter's discretion. A follow-up meeting was scheduled to collect feedback from the participants on the implemented designs and to allow them to complete the exit interview. Both the initial interview and the follow-up interview took place remotely on Zoom, and each interview was audio and video recorded. Participants were compensated with a $20 gift card.

4 Quantifying Curated Examples and Design Outcomes

In this work, our intention is not only to understand how our five factors influence visualization design but also to explore what aspects of the designs are measurable and how to measure the influence of examples on these aspects. Similar to prior work [51], we code the features of the examples in our corpus and the sketches produced by participants (Sec. 4.1). Subsequently, we use these codes to derive metrics that describe different *properties of the example sets* curated by participants (Sec. 4.2). Below, we discuss our coding process, derived metrics, and results of preliminary data evaluations.

4.1 Coding Visualization Design Components

We first performed qualitative coding to identify each visualization's data types, marks, visual encodings, layouts, compositions, and annotations. One of the authors created an initial codebook based on surveys of existing visualization taxonomies [10, 23, 33, 47, 57, 66, 68]. A subset of 5 random examples was selected from the example corpus, and three authors independently coded the examples and met to discuss the generated codes and refine the code book. Once the codebook was finalized, the three coders independently coded the remaining 45 examples and design sketches produced by participants. The coders achieved Krippendorff's Alpha inter-rater reliability score of 0.85 [40]. Our codebook is included in the supplementary materials.

Data Types. We capture the type of data attributes used in each visualization using the data classifications provided by Mackinlay et al. [47] to code the quantitative-independent (`qi`), quantitative-dependent (`qd`), categorical (`c`), and categorical-date (`cd`) data attributes encoded in a design. We assign the index for each identified attribute as a concatenation of a suffix and unique identifier (e.g., $qi1, qi2, c1$).

Marks. To understand the basic graphical marks used to represent data items, we coded the individual marks based on the taxonomy of graphical marks described by Heer [33] and Munzner [57]. This

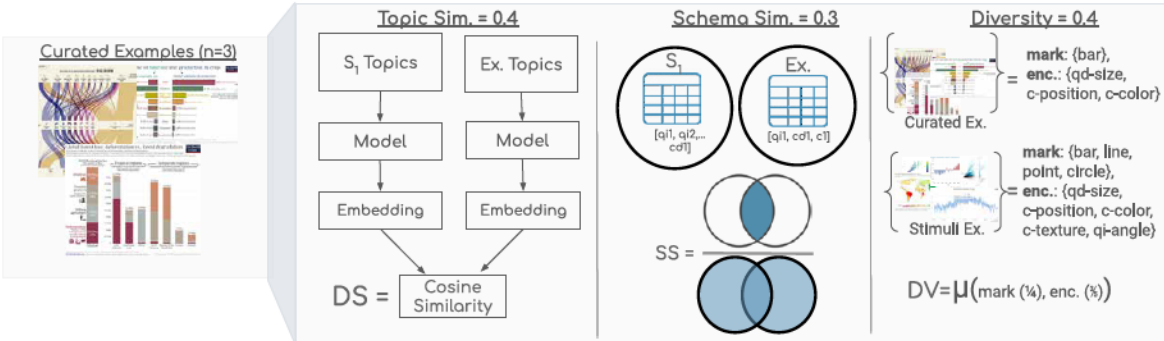


Figure 3: An illustration of how example properties are calculated for a set of curated examples. Please read Sec. 4.2 for more details.

includes marks such as `points`, `rects`, `bars`, `lines`, etc. We treat all graphical items with more than one mark as `glyphs`.

Visual Encodings. We use Munzner's taxonomy on visualization channels [57] to capture visual encodings. We capture each encoding as a pair of the data attribute and the visual channel, e.g., if a `quantitative-dependent` variable is represented as `position`, we represent it as $[qd1 - position]$.

Layouts. Certain visualizations present multiple marks together as a group and rely on the positions of these individual marks to encode data for that group. For instance, stacked bar charts are composed of individual rect marks of varying lengths stacked together to form a single bar. At the time of performing this analysis, there is no comprehensive taxonomy on visualization layouts. As a result, the primary author surveyed papers [7, 23, 27] and documentation of popular visualization languages [45, 73] and used this to generate a set of codes to represent the layout of marks within an example which includes `stacked, layered, packing, treemap, radial, grouped, network, and branched`.

Annotations. Annotations are often used in visualizations to add context [25] or draw readers' attention to specific parts of a visualization [39, 42]. To code the annotations present in each example, we use Ren et al.'s design space exploration of annotated charts [68], which we augment with a newer taxonomy presented by Rahman et al. [66].

Composition. Composite visualizations consist of multiple views. For each composite visualization, we code individual views separately and rely on Javed and Elmqvist's taxonomy [38] to capture the composition strategy used to combine the views into a single visualization.

Visualization Type. Finally, we rely on Battle et.al.'s taxonomy [10] to capture the types of visualization[s].

4.2 Deriving Properties of Curated Example Set

The codes generated from our analysis in Sec. 4.1 provide descriptive data on the features present in individual examples. However, we need quantitative metrics to describe the *collective properties* for the entire set of curated examples. We rely on heuristics provided in past literature on the factors that modulate examples influence to derive these metrics (see Sec. 2). Specifically, we focus on the `quantity` of curated examples and `diversity` of the entire example set. We also measure the similarity between the examples and the provided dataset captured in the `schema_sim.` and `topic_sim.`.

Quantity of Curated Examples: Participants were asked to bookmark examples they felt were relevant or inspiring during the example exploration stage. We also collected data on the examples that participants interacted with (i.e., expanded to inspect additional details closely) to capture unintentional influences of examples they were exposed to [50, 52]. Together, these bookmarked and expanded examples represent the set of *curated examples* produced by participants.

Topic Similarity: We calculate how close the dataset topic matches the topic[s] of selected examples, which we call the `topic_sim.` of a curated example set. For each curated example, we extracted short descriptions of the visualization from their source website (these descriptions were also presented to participants during the study). We use Open AI's ChatGPT 3.5 [58] to generate topic summaries for each description, which were manually evaluated for correctness by the lead author. Our scenarios had the following data topics `S_1:{global warming, temperature trends, data analysis}` and `S_2:{palm oil plantations, deforestation, land use}`. The extracted topics for all examples and the scenario prompt provided to participants were embedded using sentence transformers [67]. We then compute the cosine similarity between the two embeddings for each example. The similarity scores are then normalized. For instance, in Fig. 3, a participant in the S_1 scenario bookmarked examples *img-003, img-010, and img-008*, comparing the topic for each of these examples to the scenario topic produces a `topic_sim.` score of ≈ 0.4. Participants with scores closer to 1 curated examples that are closer to their assigned scenario topic.

Schema Similarity: We were also interested in examining if participants selected examples with similar data schema to the dataset provided to them. For the two datasets used in the scenarios provided to participants, S_1 had a data schema: $\{qi: 4, cd: 1\}$. While S_2 had a data schema: $\{qi: 8, cd: 2, c: 4\}$. We computed a `schema_sim.` score between the participants' assigned dataset and their curated example set. Related work on schema matching has explored the use of data types and distributions [12], linguistic similarity within attribute names [43, 49], and meta-data [48] to identify similarities between dataset schema. While we cannot access the underlying data for each example, we borrow techniques from past work and rely on the codes generated for data types in Sec. 4.1 to compute this metric. We represent examples and scenario datasets as sets of unique data attributes x and y, respectively. We calculated the Jaccard Index [36] $\{J(x,y) = |x \cap y| / |x \cup y|\}$ between the data attributes in each example and scenario dataset. In Fig. 3, we can compare the data attributes for our scenario dataset for S_1 to the attributes in our example (ex)= $\{qi_1, cd_1, c_2\}$. The overlapping attributes between these two would be $S_1 \cap \texttt{ex} = \{qi_1, cd_1\}$, the Jaccard Index of both sets of attributes, gives a `schema_sim.` score of 0.25. We normalized the scores to generate a single score between 0 and 1 per participant. Participants with scores closer to 0 selected examples with less `schema_sim.` to the provided dataset and vice versa.

Diversity: We compute the `diversity` of examples, which captures the explored solution space investigated by designers. Our analysis computes diversity by comparing the proportion of unique visualization components in curated example sets to the unique components in the entire example corpus. Our intuition is that curated example sets with a higher proportion of unique elements per component are more diverse than those with a lower proportion of coverage.

For each set of curated examples, we itemize the unique elements for each visual component as identified in Sec. 4.1. This process is repeated for the entire example corpus provided to par-

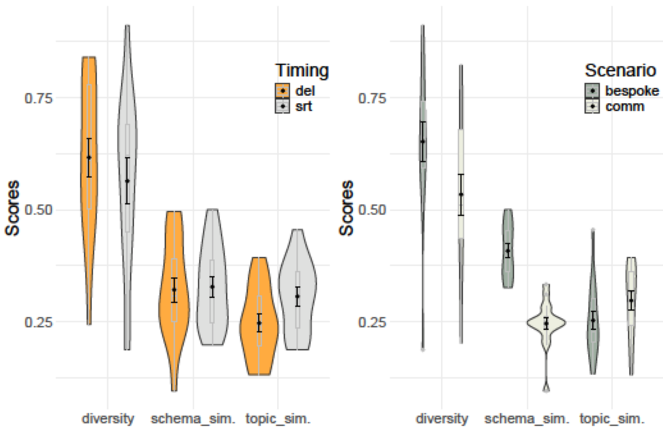
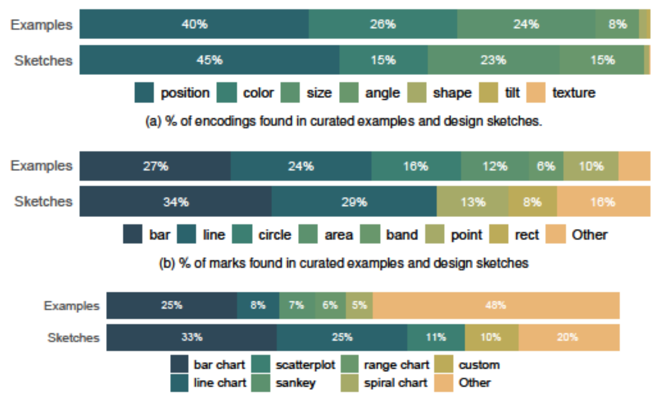


Figure 4: Distribution of scores for properties of curated examples based on the different experiment conditions in our study.



(a) % of encodings found in curated examples and design sketches.

(b) % of marks found in curated examples and design sketches

(c) % of Top Visualization types found in curated examples and design sketches

Figure 5: Overview of components found in curated examples and design sketches produced by participants. "Other" contains the sum of all components with less than 5% of each group.

ticipants. Fig. 3 shows a simplified version of this calculation: the curated examples contain $mark : \{bar\}, encodings : \{qd-size, c-position, c-color\}$ and the entire corpus of stimuli examples contains $mark : \{bar, line, point, circle\}, encodings : \{qd-size, c-position, c-color, c-texture, qi-angle\}$. For each component, we calculate the ratio of elements in the curated example set to those in the example corpus. We then compute the `diversity` score by averaging the ratios, which is equivalent to assigning the same weights to each component ratio so that the maximum possible score is 1.

4.2.1 Influence of Timing on Curated Examples Properties

Participants curated an average of 7.7 examples ($sd = 4.4$), with average scores of ≈ 0.3 for `topic_sim.` ($sd = 0.04$) and `schema_sim.` ($sd = 0.1$), and ≈ 0.6 for `diversity` ($sd = 0.18$). Fig. 4 shows the distributions of scores across the three derived metrics by timing and design scenario. One participant did not bookmark or interact with any example during the experiment; data for this participant was excluded from the analysis. During our analysis, we found a high correlation between the `quantity` of curated examples and `diversity` ($r(29) = 0.86, p < .001$). As a result, we exclude the quantity of curated examples for the rest of our analysis as the `diversity` of examples is a more meaningful measure.

Timing influences curated examples' data topic similarity To evaluate how the `timing` of example introduction may have influenced the properties of curated examples, we fit linear mixed effects models with the three calculated metrics (`topic_sim.`, `schema_sim.`, and `diversity`) as dependent variables, `timing` as the fixed effect and scenario as the random effect. Our analysis presents a statistically significant effect of `timing` on the data `topic_sim.` ($\chi^2(1) = 4.35, p = .037$). *Participants who saw examples at the start of the experiment are likely to curate examples whose data topics are 6% ($\beta = .06$) more similar to the dataset they are working on.* We find no statistically significant result for `timing` on `diversity` and `schema_sim.`

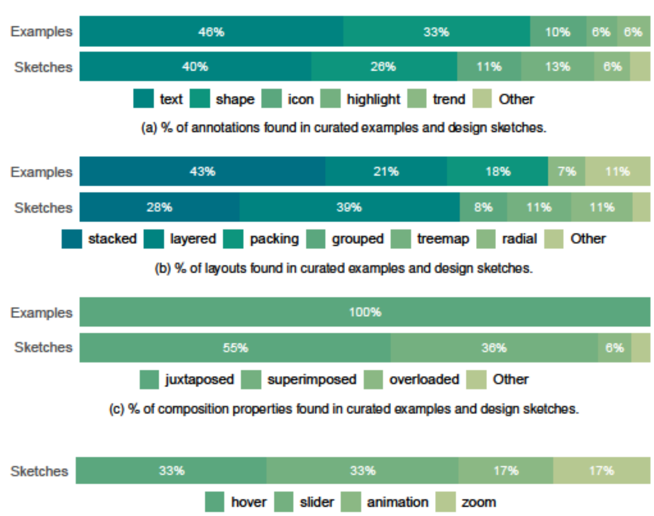
5 Measuring Design Outcomes and its Modulators

We conducted a series of quantitative analyses to identify what factors modulate the number of design outcomes produced by participants, the variety of design outcomes, and the number of ideas transferred from examples into designs. In this section, we discuss the data collected and the results of our analyses.

5.1 Data Description

A total of 128 sketches were produced by participants ($\mu = 4, sd = 1.8$). Similar to our coding process described in Sec. 4.1, we labeled the components of each sketch created by participants. We observe that bar charts account for 25% of all examples bookmarked by participants, making them the most frequently curated examples. Similarly, we see that 33% of the sketches were bar charts as shown in Fig. 5b. We also find that position was the most commonly used visual channel, making up 45% of all encodings used in the sketches and 40% in the curated examples (see figure Fig. 5a). Bar charts are commonly used by designers [8,10], and their popularity may be a confounding factor for the types of designs participants produced during our experiment. We also considered that the designer's experience could be a confounding factor in the designs produced across the experiment conditions. We account for this by including the number of years designers have spent creating visualizations in our analyses described in Sec. 5.2.

We find significant overlap in components used in curated examples and design sketches; nonetheless, a few sketches deviate from this trend as they contain components not found in curated examples. For instance, we found sketches that use custom glyphs and icons as the primary marks and texture as an encoding, as seen in Fig. 5b. The sketches that participants produced also had more annotations and layouts than the curated examples and included compositions and interactions not found in curated examples, as shown in Fig. 6.



(a) % of annotations found in curated examples and design sketches.

(b) % of layouts found in curated examples and design sketches.

(c) % of composition properties found in curated examples and design sketches.

(d) % of interactions found in curated examples and design sketches.

Figure 6: An overview of the style-related components found in curated examples and design sketches created by participants.

5.2 Influence of Timing and Curated Examples On the Quantity and Variety of Designs

We sought to explore the influence of `timing` and curated example properties (`diversity`, `schema_sim.`, and `topic_sim.`) on the `quantity` and `variety` of design outcomes. For example, does curating diverse examples lead to a higher number of designs and more design variety? Or might early exposure to examples lead to a decrease in the number and variety of designs? Here, `quantity` simply refers to the total sketches produced by participants. While `variety` measures the number of unique components (i.e., encodings, marks, etc., identified in Sec. 4) included in design sketches. We fit generalized linear mixed effects (GLME) models using `timing` as the fixed effect with participant's design experience (i.e., years creating visualizations) as the random effect for each dependent variable: `quantity` and `variety`. An analysis of the intercepts of the GLME model which includes the scenario as a random effect, yields very small coefficients (≤ 0) for the scenario. This indicates no systematic effect from including the design scenario in the model. As such, a simpler model without the scenario is preferred. We report the results of these analyses below.

Delay in introducing examples increases the quantity of designs created by participants. We observe a significant effect of the `timing` on the `quantity` of designs produced ($\chi^2(1) = 5.48, p = .019$). *Participants who were provided with examples at the start of the experiment (T_{srt}) produced (35%) fewer designs on average compared to participants who were provided with examples after brainstorming (T_{del}) ideas ($\beta = -.43$).* These findings are consistent with prior work that finds a delay in introducing examples is more effective at improving the number of ideas produced while problem solving [56, 76].

Delay in introducing examples increases the variety of designs created by participants. On average, participants used 9 unique components in their designs ($sd = 3.1$). Likelihood ratio tests show a significant effect of `timing` on the `variety` of design ideas used by participants ($\chi^2(1) = 4.54, p = .033$). *Participants who were provided with examples at the start of the experiment (T_{srt}) produced designs with 26% fewer unique components compared to participants who saw examples after brainstorming ideas.*

Example Properties do not influence the number or variety of design outcomes. To assess if curated examples influence the `quantity` and `variety` of design outcomes, we fit generalized linear models for the dependent variables with example properties as independent variables. We observe no significant effect of `diversity`, `topic_sim.`, or `schema_sim.` on either `quantity` or `variety` of design outcomes.

5.3 Influence of Curated Examples on Idea Transfer

Part of our investigation includes exploring objective ways to measure the transfer of ideas from examples into final designs. In prior work, idea transfer was measured by simply counting overlapping features between examples and designs. However, this technique cannot be directly applied here as participants may have underlying biases towards certain designs based on their prior experience, evident in the prevalence of bar charts in the sketches produced by participants. This introduces a confounding factor that needs to be mitigated.

Consequently, we made two accommodations in our analysis. First, there is no way to measure the pure influence of examples on designs for participants exposed to examples at the start of the brainstorming process. Hence, we do not include any of the sketches produced by participants in the T_{srt} condition. This brings our total number of sketches examined to 77, i.e., $\approx 60\%$ of all sketches produced during the experiment. Second, to account for participants' prior knowledge, *we assume that `idea_transfer` occurs only if a design component used in a sketch is also present in at least one of the examples bookmarked by a participant but not present in any of the sketches they created before exposure to examples*. For instance, if a participant bookmarks example `img-001`, which has a quantitative dependent variable encoded as size (`qd-size`) and the same encoding is used in a new sketch(sk_3) but not in any of the sketches they had previously created. We record `qd-size` as a component that has been transferred from `img-001` into sk_3. We found 41 instances of `idea_transfer` among 13 participants.

Higher schema similarity leads to more transferred ideas. We considered how the properties of examples might influence the number and types of `idea_transfer` by participants. For instance, do higher `diversity`, `schema_sim.` and `topic_sim.` scores lead to a decrease in `idea_transfer`? We fit a generalized linear model to evaluate how the number of ideas transferred (dependent variable) is affected by the curated example set's properties (independent variables).

Our results show a significant effect between `schema_sim.` and the number of transferred ideas ($d = 6.38$, $p = .02$). These results indicate that it is likely that *a unit increase in the `schema_sim.` of curated examples will result in an increase of 4.08 in the number of transferred ideas.* We find no significant results for `diversity` or `topic_sim.`

Section 5 Summary.
We explored what and how factors such as the timing of example introduction and the properties of curated examples influenced design outcomes. Our findings show that:

- When examples are introduced after initial brainstorming of ideas, designers are likely to produce more designs and use a wider variety of visualization components in their designs.
- Designers who curated more examples resembling their target data schema tend to copy more ideas from examples.

We acknowledge that due to the small population of designers in this exploratory study, further research is needed to validate the generalizability of these results. We discuss the implications of these findings for future work in Sec. 7.

6 Why and How Designers Incorporate Examples

In this section, we present the results of our qualitative analysis to understand designers' behavior during the experiment and their rationale for using examples. Two coders independently coded video recordings of our interviews, capturing participants' utterances related to the design process. One of the coders consolidated these codes, organizing them into clusters and then refining these clusters to identify emergent themes. Due to space considerations, we focus on two key themes from our analyses: the features influencing example usage and the forms of idea transfer. We also present a review of infeasible designs and participants' reflections on implemented designs.

6.1 What Example Features Influence Designers?

6.1.1 Features that attract designers to examples

Relatedness and uniqueness of the data presented in the example. The primary reason expressed by participants for choosing examples was the relationship between the data they wanted to present and that which is present in the examples. Participants often chose examples that were compatible with their data (n=6) or the examples presented the same data in a unique (n=3) or different way (n=6).

Similarity in visualization tasks. Participants also selected examples that support similar *task[s]* (e.g., comparison, trend) to the task they were trying to present (n=7). P11, when looking at an example of a ridge-line plot (`img-041`) remarked *"...this is something which we could leverage to try and distinguish the worst 5 years in terms of the temperature...I like this. This is like a combination of exactly comparison and annotation."*

Similarity of design to prior ideas. Certain participants were interested in how close the visualization design was to designs they had already created. All but one of the 6 participants who expressed this sentiment had seen examples only *after* brainstorming ideas. These participants were attracted to examples not just for design similarities but also to see if the examples had additional design elements that could improve their original ideas. For instance, when P17 encountered `img-10`, they said, *"Oh, this is like exactly what I did. Did they do anything better? They have annotations, ... instead of trying to put them on the Y-axis, they put a percentage of the total. That's kind of useful."*. P17 then proceeded to sketch out a new design that incorporates *img-010*. This revised sketch can be seen in Fig. 7b.

Visual Composition. Participants also expressed that they were influenced by the *composition* of design elements used in the examples (n=23). These include the overlay of multiple visualizations in a single chart (n=6) or encoding choices used in the design (n=2). Participants were also drawn to the use of annotations (n=6), layouts of marks (n=4), color schemes (n=3), and general design aesthetics (n=2).

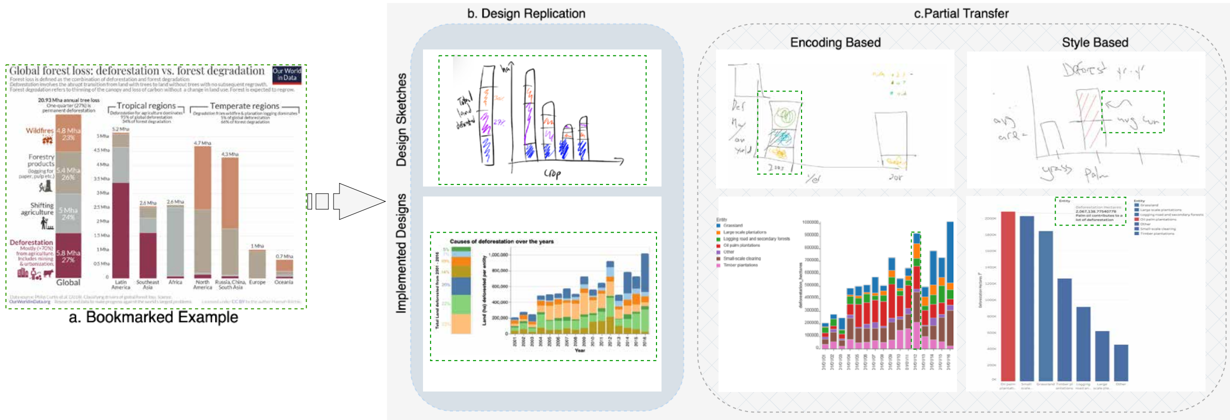


Figure 7: A depiction of the forms of idea transfer identified in sketches produced by three participants who bookmarked the same example. In design replication, the entire example is copied, whereas, in partial transfer, participants copy either encodings or styles from the example.

6.1.2 Features that deter designers from examples

Confusing designs. Participants reject examples if they have difficulty understanding what is being presented in the visualization (n=9). Participants described these examples as being *"too complex"-P28*, *"confusing"-P10, P9, P28*, or *"unreadable"-P25*. Sometimes, the confusion may stem from a lack of clarity on how to apply the example to their data. For instance, P20 remarks on `img-001` *"I'm just worried that since temperature anomaly goes negative. That would confuse people a little bit. so I would have to find another way to show that..."*

Incompatibility with data. We observed that participants would reject example designs if the example is incompatible with the data they are working on. For instance, participants rejected examples if they had different data types (n=5) or if they felt the design would not represent temporal data properly (n=3). Participants would also reject examples if they could not mentally map the data they had to the data encodings (n=6). This could be because some data transformations must be performed to format the data or introduce new variables.

Lack of inspiration. Finally, designers may reject the examples if they find the designs uninspiring (n=10). Participants often describe these designs as *"not helpful"*. Participants expressed that they found designs to be uninspiring when they were too similar to the ideas they had already brainstormed or were designs they had encountered a lot. One participant said, *"I've seen this visualization or similar visualization before. I don't tend to see as much value in this as the visualization we just did [referring to their sketch]" -P15*

6.2 Forms of Idea Transfer

We were interested in understanding ***how*** designers transferred ideas from examples into their designs. We evaluated the 41 instances of idea transfer we identified in Sec. 5.3 and reviewed notes taken during the experiment. *Our observations reveal two forms of idea transfer.*

Partial Idea Transfer. We observed that participants selected individual features from one or more examples to incorporate into their designs. This behavior has been observed in earlier work where visualization designers describe selecting and merging ideas into their designs [9]. Here, we distinguish between features that specify the parameters for binding data to graphical components i.e., *encodings* (visual encodings and marks), and those that focus on the presentation of these features within the sketch, i.e., *styles* (layout, composition, and annotations). We found 23 instances of idea transfer, which involved participants copying encodings from examples into their sketches. We call this *encoding level transfer*, and an example of this can be seen in Fig. 7c where a bookmarked example (`img-010`) has encoded the total deforested land as the size of the bars (`qd-size`) which participant P18 copied to denote the total deforested hectares in their sketch. We found 18 instances of *style-level idea transfer* in our participants' sketches. The most common form of style-level transfer was copying text and shape annotations as shown in Fig. 7c.

Design Replication. We observed situations where participants copied entire designs from examples. In this form of idea transfer, participants replicated examples, only substituting the data used. An instance of this is shown in Fig. 7b where P17 copied example `img-010`. 13 participants engaged in this behavior, producing 21 sketches where entire designs were replicated. Participants often remarked interest in these designs because the examples presented data in clever, useful, and appealing forms they hadn't considered. For instance, P22, who copied example `img-012` remarked: *"It's an interesting way of visualizing color, making it appealing to the eye and interesting to look at... That's pretty visually interesting. It almost looks like mountains and sunsets"*. Two participants recalled examples they had seen in the past which were not part of the stimuli corpus and replicated these examples. This behavior of long-term recollection of examples has been reported in prior work [9, 50].

6.3 Reflecting on Implemented Designs

As part of our experiment, we implemented a subset of the design sketches created by participants. We present our observations on the feasibility of these designs as well as participants' reflections.

Infeasible Designs. A total of 64 designs were selected by participants, for implementation. However, not all designs selected by participants were successfully implemented; we classify these as *infeasible* designs (n=7) and successfully implemented designs as *feasible* designs (n=57). We find that ***a mismatch between the intended visualization form and the data primarily caused infeasible designs***. An example of this type of mismatch is when P2 wanted a Sankey diagram that showed the flow of crop oil production to the total land used for the specific crop and the total deforested land, faceted by a grouping of years (before and after 1996). Given the dataset provided to this participant, the data transformation needed to prepare the data for this design was impractical as the data provided to them for the causes of deforested land did not include association with the crops. We also observed 3 cases of infeasible designs resulting from participants using data not provided in the dataset to create their designs. For example, P21 wanted to contrast temperature change influenced by human activity with expected temperature change without human influence. Overall, these infeasible designs reflect a lack of understanding of data constraints and abstractions, echoing observations made by Bigelow et al. [13]. All infeasible design sketches are included in our supplementary materials.

Reflections on implemented Designs. In follow-up interviews, participants were surprised to find that the implemented versions of their designs differed from their original sketches. For some participants, this unforeseen difference was positive (n=6) because the perceived data distributions or changes were not as drastic as they thought. For others, these differences highlighted design issues they had not considered, such as plotting two attributes with different domains on the same axis. Some participants described their designs as distracting (n=4) or too complex (n=4) or inappropriate for the given task (n=3).

Section 6 Summary.
We sought to understand designers' rationale for selecting relevant examples and their thought processes during the design process. Our findings confirm our quantitative results that designers select examples based on the similarity between the example data and their own data. However, we find other *features that were not previously considered, such as the similarity in tasks, the visual composition, and the complexity of designs*. Furthermore, we identify two forms of idea transfer, i.e., the partial transfer of ideas and design replication. Finally, participants' reflections on implemented designs highlight the intricacies involved in the iterative evaluation and refinement of visualization designs.

7 Discussion and Future Work

This paper presents an exploratory study addressing critical questions about the influence of examples on design outcomes. Our investigation reveals that the data schema similarity and timing of example introduction significantly impact how designers curate and extract ideas from examples. We also uncover insights into designers' rationale for selecting examples and how ideas are incorporated into designs. We present discussions on implications for visualization designers, future research directions, and the limitations of our study.

7.1 Implications for Visualization Designers

Enhancing design diversity with timed example introduction. A major finding of our work is that delaying the introduction of examples increases both the number and variety of designs. Designers may benefit more from using examples after initial brainstorming rather than to jumpstart the design process. This approach expands the space of explored ideas, allowing them to merge new concepts and reduce fixation. Furthermore, data visualization educators can adopt this strategy by delaying the use of examples during class exercises, allowing students to brainstorm first before incorporating variations from examples.

Considerations for incorporating example properties into designs. Participants often noted a mismatch between their initial design sketches and final implementations due to data-task incompatibility and overly complex or confusing designs. This issue was especially pronounced for those who fixated on certain examples and tried to force their data to fit those examples. It is important that when using examples, designers first evaluate example property dimensions such as data compatibility, visualization components, and design complexity. Understanding the interactions of these properties and their impact on design feasibility and comprehension could lead to more appropriate idea transference and better design outcomes.

7.2 Implications for Future Research

Investigating emergent factors. Based on our analysis of participants' rationale for choosing examples, we find 3 new factors that researchers had not previously evaluated: the similarity of visual tasks, prior design ideas and preferences, and the visual composition of designs. Future work is needed to investigate how to capture these properties and measure their potential influence on design outcomes. This is particularly important for eliciting design preferences, as we observed during our study that designers struggle to verbalize their design intent.

Developing metrics to quantify example properties. As part of our analysis, we needed to develop metrics to quantify the properties of selected examples in terms of their topic and schematic relatedness to datasets, as well as the diversity of designs. We did not find sufficient guidelines in the visualization literature on developing such metrics, especially for the diversity of designs. In this work, we rely on the underlying visual and data components to generate scores and opt for simple methods for metric calculation. However, these techniques are exploratory and based on intuition. More work is needed to evaluate new methods for measuring design diversity.

While developing the code book for identifying the components of designs, we also found gaps in the literature on taxonomies for describing the layout of marks within a visualization. As a result, our codes for evaluating the layout of visual designs may not be comprehensive as we had to survey a limited set of literature to extract these codes. We encourage the visualization community to look into how we can develop more comprehensive taxonomies for visualization layouts.

Analogical Inspiration for Visualization Design. In this study, we focus on understanding how existing visualization examples influence design outcomes. However, inspiration for designs is not restricted to visualization examples but can also come from artifacts found in nature, art, or physical objects around us. Prior research has shown that examples from fields not closely related to the design task influence the novelty and quality of design outcomes [14, 15, 21]. These works show the potential of augmenting ideation with analogical examples to promote the creation of new and unique concepts. Untangling the influences of analogical inspiration on visualization design outcomes is an interesting direction for future research.

Studying design iterations. During the experiment, we asked participants to sketch the design ideas they generated as they explored the dataset and stimuli examples. These design sketches are the by-product of the brainstorming exercises and represent the first iteration of visualization designs. Prior work has identified that visualization design requires multiple iterations before a final design is produced [79]. Our work only captures the first iteration of the design process; future work could consider a long-term study design where the experimenter and participants could implement and iterate on multiple designs. Nonetheless, our results present novel perspectives on using examples during the design process. Since we have identified that a delay in introducing examples significantly increases design ideas, future work can explore using examples to support the different design stages. For instance, as a designer creates a visualization design, can we provide suggestions of alternative designs that may be more suited for their selected data attributes and assumed design task?

7.3 Limitations of Our Study

We acknowledge that our choice to limit participants to only 50 curated examples introduces limitations to our study design as it does not accurately reflect the real-world example search process for designers. However, we had to make concessions to have our study design be as close to Smith et al.'s protocol as possible. Additionally, both design scenarios in our work are about communicating insights, and we did not study the influence of examples on the design of visualizations for data exploration or analysis. Visualizations often need to be interactive to support analytic tasks. Quantifying and evaluating interaction design is beyond the scope of this work, and future research is needed to investigate this interesting direction. Finally, we considered the participants' design experience based on their years of experience creating data visualizations. However, this is not a comprehensive measure of design experience. Our future work will examine the development of more robust measures of design experience.

8 Conclusion

This paper presents an exploratory study addressing critical questions about the influence of examples on design outcomes. Our findings show that a delayed introduction of examples leads to increased design ideas and the number of ideas transferred from examples is influenced by the similarity between datasets. We identify two forms of idea transfer: the partial transfer of encoding or style-based ideas and visualization replication. We shed light on salient features that influence the selection of examples, such as the underlying visual tasks and compositions of designs. Finally, we present how this work informs opportunities for future research on example-based visualization design.

Acknowledgments

Thanks to the Human-Data Interaction Group for their feedback and support. This work was supported by NSF grants IIS-2239130, IIS-2141506, and IIS-1850115.

REFERENCES

[1] Information is beautiful. https://informationisbeautiful.net/. Accessed: March 28, 2024. 3

[2] Our world in data. https://ourworldindata.org/. Accessed: March 28, 2024. 3

[3] Storytelling with data. https://www.storytellingwithdata.com/. Accessed: March 28, 2024. 3

[4] M. Agogué, A. Kazakçi, A. Hatchuel, P. Le Masson, B. Weil, N. Poirel, and M. Cassotti. The impact of type of examples on originality: Explaining fixation and stimulation effects. *The Journal of Creative Behavior*, 48(1):1–12, 2014. doi: 10.1002/jocb.37 2

[5] M. Agogué, P. Le Masson, C. Dalmasso, O. Houdé, and M. Cassotti. Resisting classical solutions: The creative mind of industrial designers and engineers. *Psychology of Aesthetics, Creativity, and the Arts*, 9(3):313, 2015. doi: 10.1037/a0039414 1

[6] S. Amershi and M. R. Morris. Cosearch: a system for co-located collaborative web search. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 1647–1656, 2008. doi: 10.1145/1357054.1357311 1

[7] B. Bach, Z. Wang, M. Farinella, D. Murray-Rust, and N. Henry Riche. Design patterns for data comics. In *Proceedings of the 2018 chi conference on human factors in computing systems*, pp. 1–12, 2018. doi: 10.1145/3173574.3173612 5

[8] H. Bako, A. Varma, A. Faboro, M. Haider, F. Nerrise, B. Kenah, and L. Battle. Streamlining visualization authoring in d3 through user-driven templates. In *2022 IEEE Visualization and Visual Analytics (VIS)*, pp. 16–20. IEEE, 2022. doi: 10.1109/VIS54862.2022.00012 6

[9] H. K. Bako, X. Liu, L. Battle, and Z. Liu. Understanding how designers find and use data visualization examples. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–11, 2022. doi: 10.1109/TVCG.2022.3209490 1, 4, 8

[10] L. Battle, P. Duan, Z. Miranda, D. Mukusheva, R. Chang, and M. Stonebraker. Beagle: Automated extraction and interpretation of visualizations from the web. In *Proceedings of the 2018 CHI conference on human factors in computing systems*, pp. 1–8, 2018. doi: 10.1145/3173574.3174168 4, 5, 6

[11] L. Battle, D. Feng, and K. Webber. Exploring d3 implementation challenges on stack overflow. In *2022 IEEE Visualization and Visual Analytics (VIS)*, pp. 1–5. IEEE, 2022. doi: 10.1109/VIS54862.2022.00009 1

[12] J. Berlin and A. Motro. Database schema matching using machine learning with feature selection. In *Advanced Information Systems Engineering: 14th International Conference, CAiSE 2002 Toronto, Canada, May 27–31, 2002 Proceedings 14*. Springer, 2002. doi: 10.1007/3-540-47961-9_32 5

[13] A. Bigelow, S. Drucker, D. Fisher, and M. Meyer. Reflections on how designers design with data. In *Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces*, AVI '14, 8 pages, p. 17–24. Association for Computing Machinery, New York, NY, USA, 2014. doi: 10.1145/2598153.2598175 8

[14] N. Bonnardel. Creativity in design activities: The role of analogies in a constrained cognitive environment. In *Proceedings of the 3rd conference on Creativity & cognition*, pp. 158–165, 1999. doi: 10.1145/317561.317589 2, 9

[15] N. Bonnardel and E. Marmèche. Evocation processes by novice and expert designers: Towards stimulating analogical thinking. *Creativity and Innovation Management*, 13(3):176–186, 2004. doi: 10.1111/j.0963-1690.2004.00307.x 9

[16] M. Bostock, V. Ogievetsky, and J. Heer. D^3 data-driven documents. *IEEE transactions on visualization and computer graphics*, 17(12):2301–2309, 2011. doi: 10.1109/TVCG.2011.185 4

[17] A. Burns, C. Xiong, S. Franconeri, A. Cairo, and N. Mahyar. How to evaluate data visualizations across different levels of understanding. In *2020 IEEE Workshop on Evaluation and Beyond - Methodological Approaches to Visualization (BELIV)*, pp. 19–28, 2020. doi: 10.1109/BELIV51497.2020.00010 3

[18] C. Cardoso and P. Badke-Schaub. The influence of different pictorial representations during idea generation. *The Journal of Creative Behavior*, 45(2):130–146, 2011. doi: 10.1002/j.2162-6057.2011.tb01092.x 2

[19] S. Carpendale. Evaluating information visualizations. In *Information visualization: Human-centered issues and perspectives*, pp. 19–45. Springer, 2008. doi: 10.1007/978-3-540-70956-5_2 2

[20] J. Chan, S. P. Dow, and C. D. Schunn. Do the best design ideas (really) come from conceptually distant sources of inspiration? *Engineering a Better Future: Interplay between Engineering, Social Sciences, and Innovation*, pp. 111–139, 2018. doi: 10.1016/j.destud.2014.08.001 2

[21] J. Chan, K. Fu, C. Schunn, J. Cagan, K. Wood, and K. Kotovsky. On the benefits and pitfalls of analogies for innovative design: Ideation performance based on analogical distance, commonness, and modality of examples. 2011. doi: 10.1115/1.4004396 1, 2, 9

[22] J. Chan, P. Siangliulue, D. Qori McDonald, R. Liu, R. Moradinezhad, S. Aman, E. T. Solovey, K. Z. Gajos, and S. P. Dow. Semantically far inspirations considered harmful? accounting for cognitive states in collaborative ideation. In *Proceedings of the 2017 ACM SIGCHI Conference on Creativity and Cognition*, C&C '17, 13 pages, p. 93–105. Association for Computing Machinery, New York, NY, USA, 2017. doi: 10.1145/3059454.3059455 1, 2

[23] X. Chen, W. Zeng, Y. Lin, H. M. AI-maneea, J. Roberts, and R. Chang. Composition and configuration patterns in multiple-view visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1514–1524, 2021. doi: 10.1109/TVCG.2020.3030338 4, 5

[24] I. Chiu and L. H. Shu. Investigating effects of oppositely related semantic stimuli on design concept creativity. *Journal of Engineering Design*, 23(4):271–296, 2012. doi: 10.1080/09544828.2011.603298 2

[25] E. K. Choe, B. Lee, and m. schraefel. Characterizing visualization insights from quantified selfers' personal data presentations. *IEEE Computer Graphics and Applications*, 35(4):28–37, 2015. doi: 10.1109/MCG.2015.51 5

[26] E. G. Chrysikou and R. W. Weisberg. Following the wrong footsteps: fixation effects of pictorial examples in a design problem-solving task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(5):1134, 2005. doi: doi/10.1037/0278-7393.31.5.1134 1, 2

[27] W. Cui, X. Zhang, Y. Wang, H. Huang, B. Chen, L. Fang, H. Zhang, J.-G. Lou, and D. Zhang. Text-to-viz: Automatic generation of infographics from proportion-related natural language statements. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):906–916, 2020. doi: 10.1109/TVCG.2019.2934785 5

[28] C. Eckert and M. Stacey. Sources of inspiration: a language of design. *Design Studies*, 21(5), 2000. doi: 10.1016/S0142-694X(00)00022-3 1

[29] N. C. for Environmental Information. National time series: Climate at a glance. https://www.ncei.noaa.gov/access/monitoring/climate-at-a-glance/national/time-series, 2022. Accessed: Dec 12, 2022. 3

[30] K. Fu, J. Cagan, and K. Kotovsky. Design Team Convergence: The Influence of Example Solution Quality. *Journal of Mechanical Design*, 132(11):111005, 11 2010. doi: 10.1115/1.4002202 2

[31] K. Fu, J. Chan, J. Cagan, K. Kotovsky, C. Schunn, and K. Wood. The meaning of "near" and "far": the impact of structuring design databases and the effect of distance of analogy on design output. *Journal of Mechanical Design*, 135(2):021007, 2013. doi: 10.1115/1.4023158 2

[32] M. Gonçalves, C. Cardoso, and P. Badke-Schaub. What inspires designers? preferences on inspirational approaches during idea generation. *Design studies*, 35(1):29–53, 2014. doi: 10.1016/j.destud.2013.09.001 1

[33] J. Heer. Data types, graphical marks, and visual encoding channels. https://observablehq.com/@uwdata/data-types-graphical-marks-and-visual-encoding-channels, Sep 2021. Accessed: Sept 28, 2023. 4

[34] M. Helms, S. S. Vattam, and A. K. Goel. Biologically inspired design: process and products. *Design studies*, 30(5):606–622, 2009. doi: 10.1016/j.destud.2009.04.003 1, 2

[35] S. R. Herring, C.-C. Chang, J. Krantzler, and B. P. Bailey. Getting inspired! understanding how and why examples are used in creative design practice. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 87–96, 2009. doi: 10.1145/1518701.1518717 1

[36] P. Jaccard. The distribution of the flora in the alpine zone. 1. *New phytologist*, 11(2):37–50, 1912. 5

[37] D. G. Jansson and S. M. Smith. Design fixation. *Design Studies*, 12(1):3–11, 1991. doi: 10.1016/0142-694X(91)90003-F 1, 2

[38] W. Javed and N. Elmqvist. Exploring the design space of composite visualization. In *2012 ieee pacific visualization symposium*, pp. 1–8. IEEE, 2012. doi: 10.1109/PacificVis.2012.6183556 5

[39] R. Kosara and J. Mackinlay. Storytelling: The next step for visualization. *Computer*, 46(5):44–50, 2013. doi: 10.1109/MC.2013.36 5

[40] K. Krippendorff. Estimating the reliability, systematic error and random error of interval data. *Educational and psychological measurement*, 30(1):61–70, 1970. doi: 10.1177/001316447003000105 4

[41] C. Kulkarni, S. P. Dow, and S. R. Klemmer. Early and repeated exposure to

examples improves creative work. In *Design thinking research: Building innovation eco-systems*, pp. 49–62. Springer, 2013. doi: 10.1007/978-3-319-01303-9_4 1, 2

[42] B. Lee, R. H. Kazi, and G. Smith. Sketchstory: Telling more engaging stories with data through freeform sketching. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2416–2425, 2013. doi: 10.1109/TVCG.2013.191 5

[43] W.-S. Li and C. Clifton. Semantic integration in heterogeneous databases using neural networks. In *Proceedings of the 20th International Conference on Very Large Data Bases*, VLDB '94, 12 pages, p. 1–12. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1994. doi: 10.5555/645920.672826 5

[44] J. S. Linsey, I. Tseng, K. Fu, J. Cagan, K. L. Wood, and C. Schunn. A Study of Design Fixation, Its Mitigation and Perception in Engineering Design Faculty. *Journal of Mechanical Design*, 132(4):041003, 04 2010. doi: 10.1115/1.4001110 1

[45] Z. Liu, C. Chen, F. Morales, and Y. Zhao. Atlas: Grammar-based procedural generation of data visualizations. In *2021 IEEE Visualization Conference (VIS)*. IEEE, 2021. doi: 10.1109/VIS49827.2021.9623315 5

[46] Z. Liu, J. Thompson, A. Wilson, M. Dontcheva, J. Delorey, S. Grigg, B. Kerr, and J. Stasko. Data illustrator: Augmenting vector design tools with lazy data binding for expressive visualization authoring. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–13, 2018. doi: 10.1145/3173574.3173697 4

[47] J. Mackinlay, P. Hanrahan, and C. Stolte. Show me: Automatic presentation for visual analysis. *IEEE transactions on visualization and computer graphics*, 13(6):1137–1144, 2007. doi: 10.1109/TVCG.2007.70594 4

[48] J. Madhavan, P. Bernstein, A. Doan, and A. Halevy. Corpus-based schema matching. In *21st International Conference on Data Engineering (ICDE'05)*, pp. 57–68, 2005. doi: 10.1109/ICDE.2005.39 5

[49] J. Madhavan, P. A. Bernstein, and E. Rahm. Generic schema matching with cupid. In *Proceedings of the 27th International Conference on Very Large Data Bases*, VLDB '01, 10 pages, p. 49–58. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2001. doi: 10.5555/645927.672191 5

[50] R. L. Marsh and J. D. Landau. Availability in cryptomnesia: Assessing its role in two paradigms of unconscious plagiarism. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(6):1568, 1995. doi: doi/10.1037/0278-7393.21.6.1568 1, 2, 5, 8

[51] R. L. Marsh, J. D. Landau, and J. L. Hicks. How examples may (and may not) constrain creativity. *Memory & cognition*, 24(5):669–680, 1996. doi: 10.3758/BF03201091 1, 2, 4

[52] R. L. Marsh, T. B. Ward, and J. D. Landau. The inadvertent use of prior knowledge in a generative cognitive task. *Memory & cognition*, 27(1):94–105, 1999. doi: 10.3758/bf03201216 2, 5

[53] A. M. McNutt and R. Chugh. Integrated visualization editing via parameterized declarative templates. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–14, 2021. doi: 10.1145/3411764.3445356 1

[54] S. MEHER. Boston weather 2013-2023. `https://www.kaggle.com/datasets/swaroopmeher/boston-weather-2013-2023?select=boston_weather_data.csv`, March 2023. Accessed: June 24, 2024. 2

[55] S. R. Miller and B. P. Bailey. Searching for inspiration: An in-depth look at designers example finding practices. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, vol. 46407, p. V007T07A035. American Society of Mechanical Engineers, 2014. doi: 10.1115/DETC2014-35450 1, 2

[56] J. Moss, K. Kotovsky, and J. Cagan. The effect of incidental hints when problems are suspended before, during, or after an impasse. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(1):140, 2011. doi: 10.1037/a0021206 1, 2, 7

[57] T. Munzner. *Visualization analysis and design*. CRC press, 2014. doi: 10.1201/b17511 4, 5

[58] OpenAI. Gpt-3.5. `https://openai.com/gpt-3`, 2021. Accessed: March 4, 2024. 5

[59] O. Ozkan and F. Dogan. Cognitive strategies of analogical reasoning in design: Differences between expert and novice designers. *Design Studies*, 34(2):161–192, 2013. doi: 10.1016/j.destud.2012.11.006 1

[60] P. Parsons, P. Shukla, and C. Park. Fixation and creativity in data visualization design: Experiences and perspectives of practitioners. In *2021 IEEE Visualization Conference (VIS)*, pp. 76–80, 2021. doi: 10.1109/VIS49827.2021.9623297 1

[61] M. Perttula and P. Sipilä. The idea exposure paradigm in design idea generation. *Journal of Engineering Design*, 18(1):93–102, 2007. doi: 10.1080/09544820600679679 1

[62] C. Plaisant. The challenge of information visualization evaluation. In *Proceedings of the working conference on Advanced visual interfaces*, pp. 109–116, 2004. doi: 10.1145/989863.989880 2

[63] A. Purcell and J. S. Gero. Design and other types of fixation. *Design Studies*, 17(4):363–383, 1996. Special Issue: Design Cognition and Computation. doi: 10.1016/S0142-694X(96)00023-3 1, 2

[64] A. Purcell, P. Williams, J. Gero, and B. Colbron. Fixation effects: Do they exist in design problem solving? *Environment and Planning B: Planning and Design*, 20(3):333–345, 1993. doi: 10.1068/b200333 1

[65] G. J. Quadri, A. Z. Wang, Z. Wang, J. Adorno, P. Rosen, and D. A. Szafir. Do you see what i see? a qualitative study eliciting high-level visualization comprehension. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, CHI '24, article no. 204, 26 pages. Association for Computing Machinery, New York, NY, USA, 2024. doi: 10.1145/3613904.3642813 1

[66] M. D. Rahman, G. J. Quadri, B. Doppalapudi, D. A. Szafir, and P. Rosen. A qualitative analysis of common practices in annotations: A taxonomy and design space, 2023. 4, 5

[67] N. Reimers and I. Gurevych. Sentence-bert: Sentence embeddings using siamese bert-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 11 2019. doi: 10.18653/v1/D19-1410 5

[68] D. Ren, M. Brehmer, B. Lee, T. Höllerer, and E. K. Choe. Chartaccent: Annotation for data-driven storytelling. In *2017 IEEE Pacific Visualization Symposium (PacificVis)*, pp. 230–239, 2017. doi: 10.1109/PACIFICVIS.2017.8031599 4, 5

[69] D. Ren, B. Lee, and M. Brehmer. Charticulator: Interactive construction of bespoke chart layouts. *IEEE transactions on visualization and computer graphics*, 25(1):789–799, 2018. doi: 10.1109/TVCG.2018.2865158 4

[70] E. F. Rietzschel, B. A. Nijstad, and W. Stroebe. Relative accessibility of domain knowledge and creativity: The effects of knowledge activation on the quantity and originality of generated ideas. *Journal of experimental social psychology*, 43(6), 2007. doi: 10.1016/j.jesp.2006.10.014 2

[71] H. Ritchie and M. Roser. Forests and deforestation. `https://ourworldindata.org/forests-and-deforestation`, 2021. 3

[72] B. Saket, S. Huron, C. Perin, and A. Endert. Investigating direct manipulation of graphical encodings as a method for user interaction. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):482–491, 2020. doi: 10.1109/TVCG.2019.2934534 1

[73] A. Satyanarayan, D. Moritz, K. Wongsuphasawat, and J. Heer. Vega-lite: A grammar of interactive graphics. *IEEE transactions on visualization and computer graphics*, 23(1):341–350, 2016. doi: 10.1109/TVCG.2016.2599030 4, 5

[74] R. N. Shepard. Recognition memory for words, sentences, and pictures. *Journal of verbal Learning and verbal Behavior*, 6(1):156–163, 1967. doi: 10.1016/S0022-5371(67)80067-7 2

[75] P. Siangliulue, J. Chan, K. Z. Gajos, and S. P. Dow. Providing timely examples improves the quantity and quality of generated ideas. In *Proceedings of the 2015 ACM SIGCHI Conference on Creativity and Cognition*, pp. 83–92, 2015. doi: 10.1145/2757226.2757230 1, 2

[76] U. N. Sio, K. Kotovsky, and J. Cagan. Fixation or inspiration? a meta-analytic review of the role of examples on design processes. *Design Studies*, 39:70–99, 2015. doi: 10.1016/j.destud.2015.04.004 1, 2, 7

[77] S. M. Smith, T. B. Ward, and J. S. Schumacher. Constraining effects of examples in a creative generation task. *Memory & cognition*, 21(6):837–845, 1993. doi: 10.3758/BF03202751 1, 2

[78] I. Tseng, J. Moss, J. Cagan, and K. Kotovsky. The role of timing and analogical similarity in the stimulation of idea generation in design. *Design studies*, 29(3):203–221, 2008. doi: 10.1016/j.destud.2008.01.003 2

[79] J. Walny, C. Frisson, M. West, D. Kosminsky, S. Knudsen, S. Carpendale, and W. Willett. Data changes everything: Challenges and opportunities in data visualization design handoff. *IEEE Transactions on Visualization & Computer Graphics*, 26(01):12–22, jan 2020. doi: 10.1109/TVCG.2019.2934538 1, 9