

# Learning Objectives, Insights, and Assessments: How Specification Formats Impact Design

Elsie Lee-Robbins, Shiqing He, and Eytan Adar

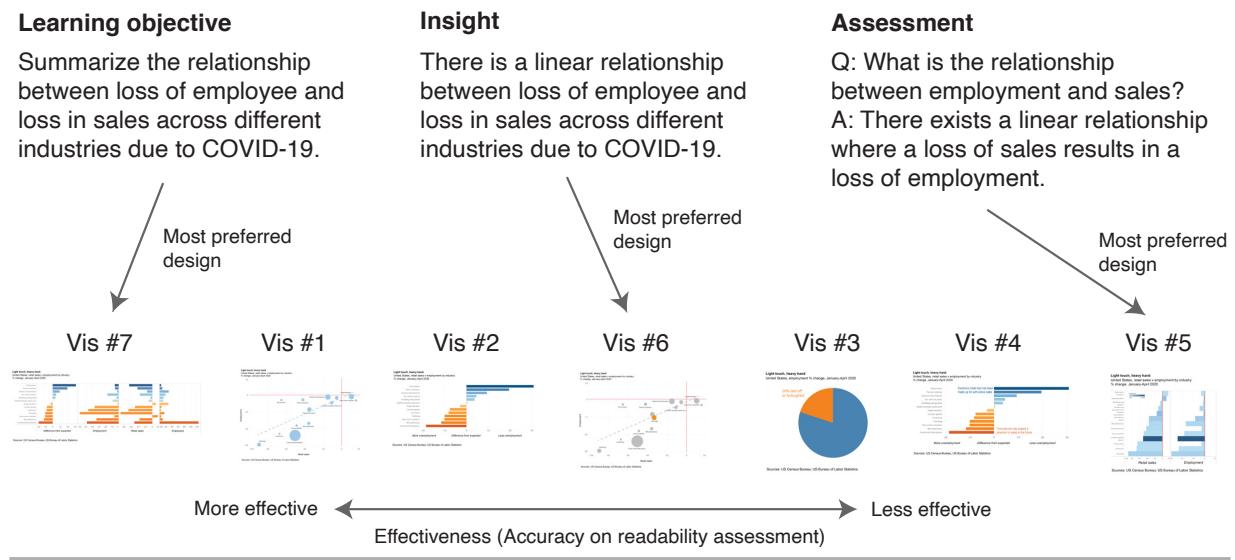


Fig. 1. Example of three specifications (learning objective, insight, and assessment) with the most preferred visualization design identified for each specification. The seven visualization designs are ordered from most effective to least effective.

**Abstract**— Despite the ubiquity of communicative visualizations, specifying communicative intent during design is ad hoc. Whether we are selecting from a set of visualizations, commissioning someone to produce them, or creating them ourselves, an effective way of specifying intent can help guide this process. Ideally, we would have a concise and shared specification language. In previous work, we have argued that communicative intents can be viewed as a learning/assessment problem (i.e., what should the reader learn and what test should they do well on). Learning-based specification formats are linked (e.g., assessments are derived from objectives) but some may more effectively specify communicative intent. Through a large-scale experiment, we studied three specification types: learning objectives, insights, and assessments. Participants, guided by one of these specifications, rated their preferences for a set of visualization designs. Then, we evaluated the set of visualization designs to assess which specification led participants to prefer the most effective visualizations. We find that while all specification types have benefits over no-specification, each format has its own advantages. Our results show that learning objective-based specifications helped participants the most in visualization selection. We also identify situations in which specifications may be insufficient and assessments are vital.

**Index Terms**— Communicative visualization, evaluation, visualization specification.

## 1 INTRODUCTION

Communicative visualizations are omnipresent. They exist in everything from news articles to scientific papers and from Web pages to television broadcasts. The people involved in the visualization design process must make design decisions based on their intents or goals. Unfortunately, most design guidelines do not connect communicative intent to the actual design. Instead, there is a significant amount of information on what makes communicative visualizations effective perceptually, focusing on making the visualization readable. However,

this low-level information often fails to account for what kind of impact people would ultimately like to make: to influence the viewer cognitively (i.e., to have them learn something) or affectively (e.g., to have them believe something) [1, 23].

Current design literature is often biased towards low-level cognitive efficiency [7, 17, 36, 37, 40]. This literature is built on significant academic foundations [38] that have allowed us to identify which encodings can be read accurately and quickly. However, if designers only rely on cognitive efficiency evaluation methods, such as how fast a viewer can decode the visualization, they might fail to consider higher-level communication goals for what they want their reader to be able to know or do. In most instances, being able to read the visualization is only the first step and may not even be the most crucial [12, 25]. A designer may also want the viewer to remember the message, take personal action, or generate hypotheses. If the reader only reads the visualization but does not take any of these next steps, it would not be considered a good design. In addition to focusing on low-level cognitive efficiency, identifying higher-level communicative goals would benefit everyone involved in the design process.

• Elsie Lee-Robbins, Shiqing He, and Eytan Adar are with University of Michigan, School of Information. E-mail: elsielee,heslicia,eadar@umich.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

There may be several stakeholders who are responsible for creating a visualization, ranging from those who have an initial idea for it, to those who create it, to those who choose the final version to publish. It's possible that there are different people for each of these steps, in the case of a client, a contractor, and an executive editor, or it's possible that one person carries out all steps of this process from start to finish. At every step, each stakeholder would benefit from a formal language for communicative intent. For creating a visualization, a clear communicative intent specification can help the designer identify their own goals and evaluate whether their visualization achieves them. Professional designers may already know how to identify a communicative intent. Even so, they could still benefit from a framework to communicate with other stakeholders. For a client commissioning a designer, a shared language of communicative intent would be helpful for the client to be able to succinctly express their goals to the designer. Sometimes, there might be several visualizations, either created by a designer or automatically generated by a computer, that a client would need to choose from in order to best fulfill their communicative goals. In this scenario, a formal structure for communicative intent can help the client analyze the different visualization designs to choose the best one. While there might be several stakeholders involved in creating, choosing, and publishing a visualization, all of them would benefit from a formal language for communicative intent to guide their design decisions, communicate with each other, and decide between multiple visualization designs.

The framing of communicative visualization design as a learning design problem [1] is a formal structure that prioritizes higher-level intents. The stakeholder takes on the role of a teacher, explicitly defining their goals of what they want the viewer (student) to be able to do or remember after viewing the visualization. For example, *the viewer will be able to recall the change in unemployment last year, or the viewer will be able to determine the optimal treatment*. These learning objective statements are based on Bloom's Taxonomy [3], though many other variants are possible [5, 39]. Critically, a *specification* of a communicative visualization as a learning problem can lead to appropriate *assessments*. A stakeholder can implement an actual test (e.g., a multiple-choice question "post-test") to determine if the visualization meets the specification criteria and satisfies the communicative intents.

Even in the context of learning, there are many ways that a specification can be crafted for a communicative visualization. For example, *learning objective statements* (e.g., *the viewer will be able to describe the reasons why Norway is a Winter Olympics powerhouse*) are broad. This type of specification is high-level; it includes both the cognitive ability (e.g., the verb *describe*) and knowledge (e.g., the 'noun' *Norway's success in the Olympics*). To achieve this, the viewer may need to read many facts, or *insights*, that are in one or more visualizations. An alternative specification format might emphasize one or more of these lower-level insights: *the high number of Olympic medals relative to population* or *the many skating rinks that Norway has*. While insights may not be as broad as learning objectives, they may be easier to map to 'graphical facts' or annotations and might be more comfortable for a designer to work with. A third specification format might privilege the assessments themselves—the tests we could use to validate learning: *In the last ten Winter Olympics, how many times did Norway win the most medals?* While specific, this format may help the designer select a visualization that will do best in testing. Conversely, an over-focus on specific test questions or insights may lead to visualization choices that are not effective broadly. This is the design equivalent of 'teaching to the test.'

The formats of these specifications are often transformable. For example, we could create one or more assessments for each insight or learning objective. Similarly, we can identify the insights that need to be communicated and learned to achieve a higher-level learning objective. We hypothesize that different specification forms emphasize different ways of thinking. Given these different specifications, designers or editors might make different choices. More critically, some of these choices might be better for learning outcomes.

In this paper, we measure the impact of different specifications on visualization design choices. We contrast the visualization preferences

when participants are exposed to different specifications in the form of learning objectives, insights, and assessments, and contrast these to baseline preferences. Additionally, we run a set of assessments against the possible visualizations to determine which produce the most accurate answers. Using both the preference data (i.e., the most preferred visualization given the specification) and accuracy (i.e., the best *performing* visualizations), we identify when different specifications lead to better learning outcomes. Specifically, we find:

1. Adding specification information changes preferences over baseline judgments.
2. Learning objectives and insight specifications guide designers to prefer better-performing visualizations.
3. Evaluations of assessments with user testing provide critical feedback of what designs are effective.

## 2 RELATED WORK

We briefly examine various ways to describe visualization intent. We focus on summarizing the pros and cons of different approaches.

### 2.1 Cognitive Effectiveness

A common approach in understanding communicative visualizations is a focus on the outcome rather than intent. Specifically, evaluations provide a mechanism to judge whether a visualization achieves a designated communication goal. However, these approaches often emphasize cognitive effectiveness metrics. For example, one common way to evaluate a visualization is to test how fast and accurately a user can perform a task while viewing the visualization. Because designers can use low-level perceptual task taxonomies (e.g., as described by [2]) at the beginning of their design process, these evaluation mechanisms can encompass designers' intent. Using a combination of tasks, designers can identify the visualization that is most effective for each task [35].

However, while benchmark usability tests are standard, these tasks are elementary and not representative of more complicated interactions. Furthermore, these measures may involve trade-offs between speed and accuracy [31]. These generic and simplified tasks cannot evaluate more domain-specific and higher-level goals. This limitation inspires us to identify a better mechanism for goal specification and evaluation.

### 2.2 Memorability

There are various memory evaluations for visualizations. For example, we could simply ask: "if shown a visualization twice, would a viewer recall seeing it before?" An alternative test may target a subgoal of the visualization by asking whether the viewer remembers a particular takeaway after the viewer has stopped looking at the visualization. Memorability is an essential part of communicative intent; designers should not be happy if the information went in one ear and out the other. Works in semiology such as [4, 29, 30] offer insights into how meaning is formed and communicated through images and that they should be 'memorable.' Although this group of works has limited analyses for communicative visualizations, they argue that designers should consider long-term engagement and retention as valuable goals.

Memorability tests shape designers' design decisions and direction. For example, visual images such as chart junk may increase the memorability of a visualization [6]. Designers might also choose to add visual difficulties to a visualization to increase engagement, higher processing, and long-term retention [25]. Our experimental design targets memorability as a key feature.

### 2.3 Usability Tests and Qualitative Evaluations

Designers can evaluate visualizations with various usability tests. Specific techniques range from usability inspection metrics (as described by [24]), running interface-focused user studies [24], and running qualitative measures such as interviews and observations [27]. Evaluations of visualizations can be based on common design guidelines and heuristics, such as "this visualization makes important information visually salient" [18, 21]. For narrative visualizations, [22] and [32] discuss the potentials of using the Elicitation Interview technique to evaluate static and narrative visualizations. For visual analytic systems, process

data could be collected as the user explores the data and can be used to evaluate the system [10].

Compared to generic task taxonomies, usability tests are specific to the visualization systems and their communicative intent. Therefore, they have the potential to target more complex cognitive levels. Nevertheless, they can be costly and time-consuming. Based on heuristics used in the tests, these evaluations might require trained experts.

## 2.4 Storytelling, Draw Your Own, and Games

While most of the evaluation mechanisms target generic visualizations, some strategies help designers achieve communicative visualization objectives. For example, using visualization rhetoric as an analytical framework for narrative visualization can help designers make engaging, layered visualizations that have clear interpretation priorities [26]. Draw-your-own style visualizations ask viewers to chart out their expectations before seeing the data, an action that supports learning [28]. Furthermore, bringing gamification to visualization has the potential to improve engagement [13]. Nevertheless, these strategies describe *how* to achieve an objective rather than offering a language to specify those objectives.

## 2.5 Specifications

Overall, existing evaluation methods rarely focus on the specific higher-level communicative goals. While there are strategies to improve communicative visualizations, designers might still find it challenging to describe their intent due to a lack of resources and guidelines. As a result, designers often need to evaluate communicative visualizations from the cognitive efficiency perspective. A different approach is to frame the evaluation of visualizations as a learning-based problem, both for communicative visualizations [1] and visual analytic systems [10]. Specifically, this framework connects studies in learning science and communication theory to construct an insight-to-learning-objective taxonomy that models designers' intent.

In the visualization community, insights are commonly thought of as the goal of visualizations [8,41]. A widely accepted definition of insight is a unit of information or knowledge that a visualization communicates [9]. An insight could be an observation of a fact, pattern, or relationship in the data [11]. Insights can range from simple to complex on several characteristics: complex, deep, qualitative, unexpected, and relevant [31]. From a communicative viewpoint, designers could consider insights as the conclusion or the primary takeaway from the graph.

With insights as communicative intent, designers aim to guide the viewer to understand potentially multiple data insights. Before or during the design process, designers will know what specific insight(s) they want to communicate. There are several limitations to this type of specification. Using insights to formulate communicative intent focuses on the data, therefore, often does not capture how the *viewer* will interact with the data. Also, insights might not capture all of the different kinds of communicative goals, such as outcomes or actions that designers want the viewer to do after seeing the visualization.

An alternative to insight is learning objectives. We adapt learning objectives from Bloom's Taxonomy for describing communicative intent for visualizations [1]. Through the lens of a teaching problem, designers take on a teacher's role, explicitly defining their goals of what they want the viewer (student) to do or remember after viewing the visualization. Learning objectives have a structure of *The viewer will [verb] [noun]*, where verbs specify cognitive change and the noun is a piece of knowledge. Bloom's taxonomy [3] provides a structure for designers to articulate what they want their audience to do effectively and at varying complexity levels.

## 3 VISUALIZATION GOAL SPECIFICATION

In our study, we used three types of specifications to frame communicative intent: learning objectives, insights, and assessments. Each type reflects a different way of considering how the viewer might learn.

To briefly introduce these types of specifications, we discuss a simple example of a dataset with three values of fruit sales: \$500 in apple sales, \$200 in pear sales, and \$100 in kiwi sales. This data can be visualized as a bar chart with three groups (Figure 2, left). For this

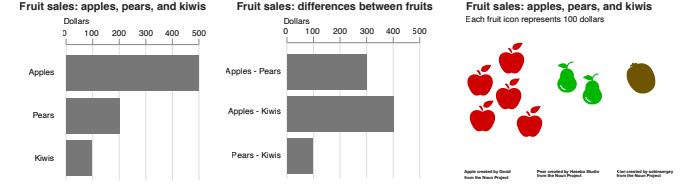


Fig. 2. Example visualizations for our fruit dataset.

dataset, we can imagine that the person creating or choosing a visualization has learning objectives:  $LO_1$ : *the viewer will recall the amount sold of each fruit* and/or  $LO_2$ : *the viewer will recall the differences between fruit sales*. Depending on the learning objective(s) and their relative importance (and factoring in the properties of the audience), we might make different choices when selecting one of the visualizations in Figure 2. Learning objectives can build off each other (e.g., we have to recall something before we can use it for synthesis tasks). Mathematically, learning objectives may also be equivalent or overlapping. In our simple example, one can calculate the differences between sales from the values. Within our two learning objective, there are six different potential *insights* that arise:

- $I_1$ : There was \$500 in apple sales [ $LO_1$ ]
- $I_2$ : There was \$200 in pear sales [ $LO_1$ ]
- $I_3$ : There was \$100 in kiwi sales [ $LO_1$ ]
- $I_4$ : Apple sales were \$300 more than pear sales [ $LO_2$ ]
- $I_5$ : Apple sales were \$400 more than kiwi sales [ $LO_2$ ]
- $I_6$ : Pears sales were \$100 more than kiwi sales [ $LO_2$ ]

Depending on our learning goal, we may focus on the  $LO_1$  or  $LO_2$  insights. Note that our definition of insight centers on facts or statements about data. These need not be the 'a-ha' style insights experienced in analytical visualization forms [9]. Rather, these are communicative insights—those chosen by the visualization designer—that they would like the viewer to recognize as 'facts' both during and after reading a visualization. We specifically relate insights to the learning objectives in that they are statistical statements about the data that emerge from the noun portion of the learning objective. In most situations, we expect that the viewer would be able to validate these insights as true or false. In our example above, there may be multiple insights that arise from one learning objective. Thus, we can specify our intent using these statements (or some representative sub-sample).

As we generated insights, we can also create *assessments*. For  $LO_2$ , or equivalently, the second set of insights ( $I_4, I_5, I_6$ ), we could write a 'test' to evaluate whether the viewer had recalled the differences between fruit sales. There are multiple possible assessments that we could use. For example, we could ask the viewer to answer "What's the difference between apple sales and kiwi sales?" Alternatively, we could ask a true or false question, "True or False: apple sales were \$200 more than kiwi sales?" There are also easier, but insufficient assessments one could use, such as ranking fruits by sales. This question would assess whether the viewer knows the ordinal rank of each fruit, which would be a first step in knowing the exact differences. Another test would be to ask a question to assess  $I_1$  "What was the total value of apple sales?" This assessment would give the stakeholder information on what fundamental information may be learned or lacking from the viewer. This assessment provides more detailed information on what parts of the visualization are falling short, allowing the designer to address any shortcomings. Just as we could create a specification of our intent using learning objectives, or insight statements, we could use a set of assessments instead. Critically, each form of specification has positive and negative features.

### 3.1 Learning Objectives

Our simple learning objective above took the form of, "the viewer will recall the amount sold of each fruit. ( $LO_1$ )" The template for this statement is derived from the application of the Bloom's Taxonomy [1, 3]. Learning objectives have a structure of *the viewer will [verb] [noun]*. Verbs correspond to one of the cognitive dimensions:

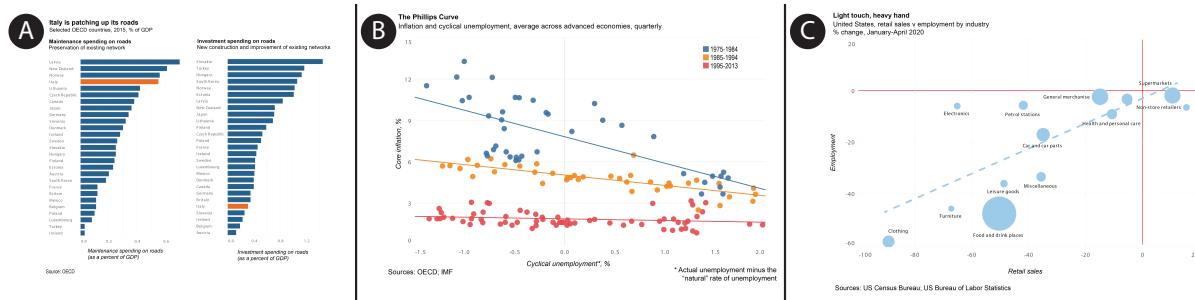


Fig. 3. Recreations of the three Economist Graphic Detail visualizations: (A) Italian road quality [15], (B) Phillips curve [14], and (C) COVID and employment [16].

*remember, understand, apply, analyze, evaluate, and create* (each with sub-categories and synonyms). These categories are largely hierarchical, with lower-level tasks (e.g., *recall*) contrasting to high-level objectives (e.g., *generate*). Nouns, in our formulation, are specific. For example, “unemployment levels in California” or “the best algorithm.” Nouns also fall broadly into *factual, conceptual, procedural, and metacognitive* knowledge. Taken together, a more sophisticated learning objective might be:

*LO<sub>3</sub>*: The viewer will summarize the relationship between loss of employee and loss in sales across different industries due to COVID-19.

We have previously demonstrated how communicative visualization intents can be mapped to this taxonomy [1]. Our work demonstrated that designers can map many of their intentions to this framework. The benefit of this approach is that learning objectives can encapsulate a wide range of lower-level goals (e.g., remembering a set of facts) through succinct high-level descriptions. However, creating objectives is difficult even with training [19, 34]. Thus, constructing objectives as specifications may not always be practical or convenient.

## 3.2 Insights

An alternative way of considering learning is to ignore the cognitive dimension and focus on key data “facts” or “insights.” Though there are numerous ways of modeling insight [9], we are concerned with a narrower definition. In the context of communicative visualization, insights are statements about data that the designer would like the viewer to recognize and internalize. In our fruit sales example, we suggested an insight (*I<sub>1</sub>*) of, “There was \$500 in apple sales.” An insight may be the “noun” portion of a learning objective but is often more atomic and more specific. For example, if we consider the objective (*LO<sub>3</sub>*) above, an insight form might be:

*I<sub>7</sub>*: There is a linear relationship between loss of employees and loss in sales across different industries due to COVID-19.

Communicative insights are often easy to identify and articulate. Many visualization annotations are reflections of insights in graphical form. Thus, a visualization goal specification may consist of one or more insights. The designer can focus on this set when designing or selecting a visualization. However, because insights are low-level and speak directly to data facts, we may require many insights in our specification to cover one learning objective. Too many insights in a specification may overwhelm the designer. Too few, and they might fixate on a narrower objective than intended. Additionally, because of their data orientation, insights do not always support other types of learning. For example, we may want a communicative visualization that teaches the viewer how to apply an algorithm (e.g., a triage flowchart at a hospital). It is more difficult to formulate that type of goal as an insight statement.

## 3.3 Assessments

Assessments are an alternative, outcome-focused way of describing learning. A feature of learning objectives is an extensive mechanism

for mapping objectives to concrete assessments [20, 33]. For example, an assessment for the objective/insight (*LO<sub>3</sub>*/*I<sub>7</sub>*) above might be:

Q: What is the relationship between employment and sales?

A: There exists a linear relationship where a loss of sales results in a loss of employment.

Even multiple-choice style questions can be combined to identify if learning has been achieved. When these are insufficient, open-ended questions with a rubric may also work. In our context, assessments of this type are used to measure “program effects.” That is, we are not concerned with a single viewer’s performance, but rather the effect of the visualization on all the viewers. By specifying visualization goals through a list of questions, the designer may be forced to reckon with their choice relative to a very specific set of outcomes. The designer should create or select the visualization that will lead to the best assessment results. An assessment-based formulation may have both the benefit and disadvantages of focusing the designer on an outcome metric. Instructors will recognize the danger here of, “teaching to the test” (or in this case, “designing to the test.”). As with insights, we may need many assessment questions to describe one learning objective or ‘triangulate’ if learning has happened.

To summarize, we note that these three specification formats are connected to each other, but vary in systematic ways. Learning objectives, which are explicit in specifying the cognitive impact, are often broader and more abstract than insights or assessments. Assessments are more specific than insights, focusing designers on only one possible question they could ask of a viewer. We expect that these differences in specification format will lead to differences in designer preferences. It is also ultimately possible that the best specification will be a mix of these approaches. The different forms of specification have different levels of granularity, ambiguity, generalizable properties, communicative efficiency, and other features that would make their use more or less desirable for describing intent. While these facets are all potentially interesting, we begin with a more high-level question: does communicative intent specified in one of these forms lead to better learning outcomes?

## 4 STUDY DESIGN

In our study, we conduct two experiments to investigate whether the type of specification had an impact on preferences for designs and, ultimately, their effectiveness. More specifically, our main research questions are:

1. Do different specifications (learning objectives, insights, assessments) affect the choice of visualization designs?
2. Are the chosen designs more effective than the alternatives?

## 4.1 Materials

To create materials for this experiment, we identified a set of three starting articles from which we generated visualization designs, learning objectives, insights, and assessments. We used three articles from The Economist’s Graphic Detail series: “American retailers have laid off or furloughed one-fifth of their workers” (COVID) [16], “The Phillips

## Stimuli Generation

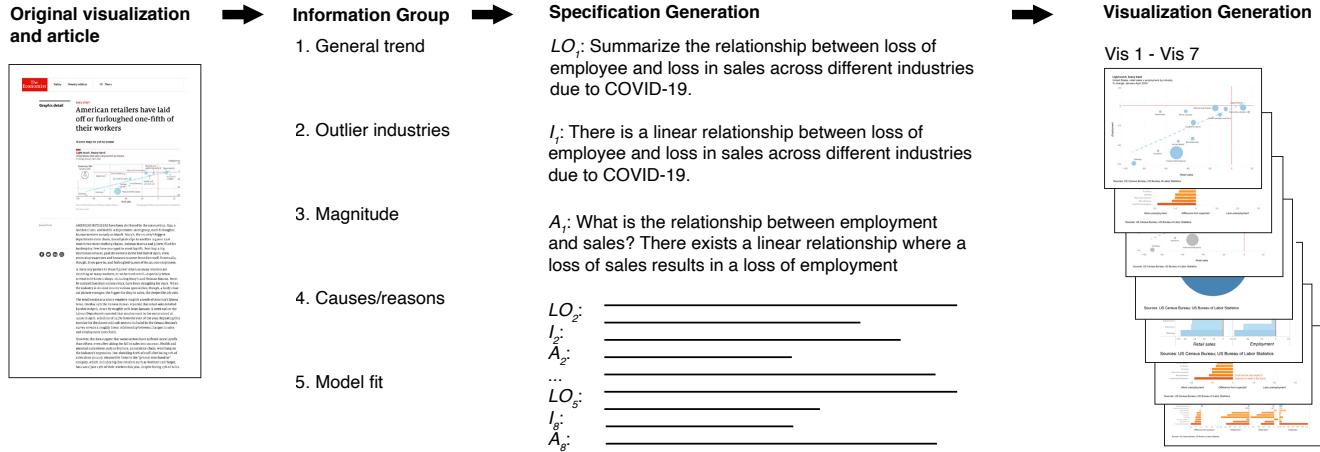


Fig. 4. A diagram of the specification and visualization generation. The three specifications shown stem from information group 1. The top visualization was created specifically for information group 1. A full set of the specifications and visualizations are in Supplemental Materials.

curve may be broken for good” (Phillips Curve) [14], and “Italy spends lots fixing old roads, not enough building new ones” (Road Quality) [15]. Each of the articles is centered on one or two visualizations with a relatively short article that describes key details (see Figure 3). Because the text is mostly anchored to elements of the visualization, it becomes easier to infer the likely communicative intents and reverse engineer plausible specifications.

For each key fact discussed in the article, we developed a learning objective, the same information expressed as an insight(s), and the assessment(s). These linked *information groups* of specifications all focus on communicating the same element of the visualization. By having roughly equivalent specifications, we can meaningfully compare specification types to each other in our study. Because of the lack of resources on how to formulate learning objectives and how they relate to insights and assessments, we explored possible ways to make linked groups of specifications. We ultimately decided to create specification groups by having learning objectives encompassing one or more insights, and each insight has one corresponding assessment question. Because of this structure, we ended up with fewer learning objectives than insights and assessments—one learning objective could comprise of more than one insight and assessment.

For the COVID retail article [16], a broad learning objective is “Summarize outlier industries/ differences in sectors.” The three insights in this information group are: “Health and Personal care have an expected amount of layoffs”; “Food service has more layoffs than expected”; and “Electronic stores have fewer layoffs than expected.” However, in more ‘basic’ (i.e., simpler) learning objectives, the information between the two specifications are more similar. For example, one of our learning objectives was “Recall that American’s retailers have laid off of furloughed a fifth of employees because of COVID-19,” and the single corresponding insight was “American’s retailers have laid off of furloughed a fifth of employees because of COVID-19”. To constrain our experiment, we created one assessment per insight. In reality, one could construct assessments that cover multiple insights simultaneously or, conversely, multiple assessments for the same insight.

After creating the set of linked information groups of specifications, we created several different visualization designs. The original visualizations from the Economist are professionally constructed (our reconstructed versions are shown in Figure 3). These visualizations provided us with a starting point for alternatives. With learning in mind, we generated multiple visualizations of varying expressiveness [30]. We created designs that we thought would be the most effective for a particular specification. For example, we designed a simple pie chart

Dataset	Vis.	Learn. Obj.	Insights	Questions
COVID	7	5	8	8
Phillips Curve	8	7	9	9
Road Quality	4	3	4	4

Table 1. Materials produced for our experiments. Each learning objective include one or more insights and each insight is associated with one multiple choice assessment question.

highlighting 20% for the specification “American’s retailers have laid off or furloughed a fifth of employees because of COVID-19.”

Figure 4 illustrates the process that we used to generate the materials. Table 1 summarizes the materials generated for the three articles. The complete instrument is available in our supplemental materials.

## 4.2 Preference Experiment

In our first experiment, we collected participant preferences for multiple visualization designs. Though we only had our participants choose between different visualization designs, our study may be more broadly applicable to designers that actually create different designs, weighing different design decisions based on communicative goals.

Participants in our study began by reading an informed consent document and completing two qualification questions. These questions tested the participants’ abilities to correctly extract information from a bar graph and a scatterplot. This ensured a level of visual literacy. After passing the qualification questions, participants read instructions and completed three practice questions with feedback. In this experiment, the visualizations were shown within-subjects and the specifications were shown between-subjects. Participants previewed all of the visualizations from one dataset on the next page. This preview was to encourage participants to compare the visualizations to each other in their ratings. On the following page, we showed each visualization with a specification and elicited a rating on a five-point Likert scale from “Extremely good” to “Extremely bad.” Each participant rated each visualization once based on a single specification.

In previous pilot iterations, we experimented with having participants rank graphs either in a pair or in a group. The advantage of rating each graph in a Likert scale instead of ranking is that we could evaluate the magnitude of differences between the graphs. Additionally, rating all of the visualizations meant that preferences could be relative to all graphs in the set.

## Accuracy Evaluation Conditions

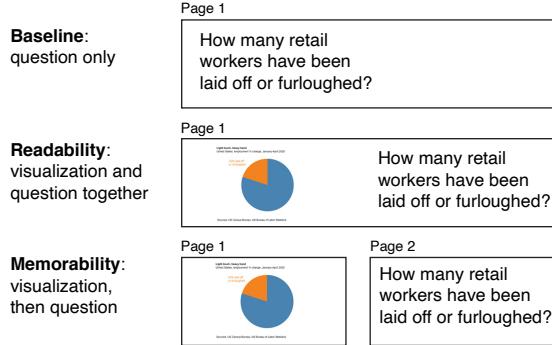


Fig. 5. Three conditions: baseline, readability, and memorability.

To obtain baseline ratings of visualizations, we conducted the same experiment *without* providing any specification. In this condition, we simply asked “Please evaluate each visualization.” Participants were free to interpret this as they chose. With the same procedure as before, participants rated the visualization on its own without a specification. With this measure, we can adjust each preference based on how much it deviates from its baseline rating. This allows us to compare the visualizations to each other by adjusting for each graph’s varying aesthetics and designs.

For the preference experiment, we hosted our experiment on Qualtrics and recruited 667 participants from Amazon Mechanical Turk. We compensated participants with \$1.00, and the median completion time was 6.2 minutes. For the baseline experiment of evaluating the visualizations without any specifications, we recruited 69 participants. We compensated participants with \$0.45, and the median completion time was 3.4 minutes.

### 4.3 Accuracy Experiment

In contrast to the participant who decides which visualization is best (the message sender), we refer to the recipient as the *viewer*. To compare the participant preferences to an empirical evaluation of the effectiveness of the design, we conducted another experiment to evaluate *viewer* accuracy on the assessments. In order to evaluate the visualizations, we implemented three versions of the accuracy experiment. In this experiment, the visualizations were shown between-subjects. We asked participants to answer the question in one of three conditions: before they saw the visualization (no-visualization baseline), while they were viewing the visualization (readability), or after they had seen the visualization and it was taken away (memorability). Each participant only viewed one test question and one visualization. The no-visualization baseline condition allows us to determine the question’s baseline difficulty based on guessing and prior knowledge. Then, we evaluated readability accuracy for answering the assessment while viewing the visualization. Finally, we evaluated the memorability accuracy on how effective the visualizations were for information that was to be remembered.

For the accuracy experiment we again used Qualtrics and recruited participants from Amazon Mechanical Turk. We recruited 203 participants for the no-visualization baseline assessment (\$0.10, median completion time = 1.1 minutes), 1658 participants for the readability assessment during the visualization (\$0.10, median completion time = 1.4 minutes), and 1126 participants for the memorability assessment (\$0.20, median completion time = 2.1 minutes).

## 5 DATA ANALYSIS

After collecting the experimental data, we are able to return to our two main research questions. First, we investigate how different visualization specifications (learning objectives, insights, assessments) impact

### Baseline preferences are significantly higher compared to specification preferences.

Insight preferences are significantly higher compared to both learning objectives and assessments.

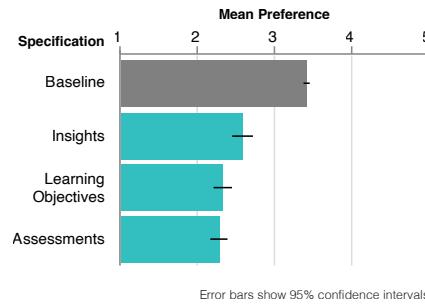


Fig. 6. The average preference for each of the specification types: no specification (baseline), learning objectives, insights, and assessments. Error bars show 95% confidence intervals. Participants preferred visualizations more when rating them with no information.

the choice of visualization designs. Then, we determine if the chosen visualizations are more effective than the alternatives.

### 5.1 Preferences

Our analysis begins by solely looking at preferences. That is, we explore whether the different types of specifications changed preference ratings of visualizations. Recall that participants rated their preference of visualizations on a five-point Likert scale from “Extremely good” (5) to “Extremely bad” (1). In our analysis, we look at both the raw averages of the preference scores and the rank of visualizations in relation to each other.

First, we examined whether there were differences in average preference for the type of specification (learning objectives, insights, assessments, or no information). A Kruskal-Wallis rank sum test showed that preferences significantly differed by specification type,  $H(3) = 74.35, p < 0.001$ . Post-hoc Dunn’s test with a Bonferroni correction showed that preferences for the three specifications were significantly lower than the baseline of no information,  $p < 0.001$ . On average, the participants rated the visualization lower when evaluating based on a specification compared to no specification (see Figure 6). A possible explanation for this is that the constraints imposed by specifications lead participants to be more critical of the visualizations. Additionally, preference scores from the insight specification were significantly higher than the learning objectives specifications ( $p = 0.03$ ) and the assessment specifications ( $p = 0.02$ ).

Next, we looked at the similarity of the ranks of preferences. Even though the baseline preferences are significantly higher than the specification preferences, the visualizations may all be ranked the same. We conducted a Spearman correlation between the *ranks* of the visualizations for each specification type. We found that all specification types were correlated with each other. Table 2 shows correlations of preference ranks for the specification types. Thus, preference ranks are positively correlated with each other, though clearly this is imperfect.

	Learning Obj.	Insight	Assessment
Insight	0.49	-	
Assessment	0.49	0.48	-
Baseline	0.57	0.67	0.48

Table 2. Spearman correlation coefficients of preference ranks for each of the specification types: learning objectives, insights, assessments, and no specification (baseline). All correlations are significant ( $p < 0.001$ ). Preference ranks are positively correlated with each other.

We investigated whether different specification types lead participants to have different ‘most-preferred’ visualizations. For each specification group, we looked at the most-preferred visualization for each

specification type. Out of the 21 groups, four groups had different most-preferred visualizations for all three specifications. Eleven groups had one specification lead to a different most-preferred visualization, while the other two specifications lead to the same most-preferred visualization. Six groups had all three specifications lead to the same most-preferred visualization. See Figure 1 for an example of an information group where each specification type leads to a different most-preferred visualization.

Additionally, we can contrast the baseline most-preferred visualization (i.e., the one chosen based on considerations without specifications) against the specification’s most-preferred visualization. We find in six cases the baseline matched no specifications (i.e., given a specification, the visualization was never most-preferred), seven times it matched one specification, five times it matched two specifications, and three times it matched all three specifications. This demonstrates the potential pitfall in a completely specification-free choice.

Just as we study the most-preferred visualization, we can also consider the least-preferred one. Out of the 21 groups, five groups had different least-preferred visualizations for all three specifications. Seven groups had one specification lead to a different least-preferred visualization, while the other two specifications lead to the same least-preferred visualization. Nine groups had all three specifications lead to the same least-preferred visualization. This indicates more agreement on the least-preferred choice given specifications. When considering the least-preferred visualization from the no-specification baseline, we find a similar pattern. In five cases the least-preferred no-specification visualization did not match any of the least-preferred visualizations given a specification. That is, the specification caused participants to reassess the ‘worst’ visualizations. In six cases the least-preferred no-specification matched one specification, five times it matched two specifications. In five cases it matched all three specifications. In those few cases, the initial ‘gut’ preference on the worst visualization corresponded to the preference after observing the specification.

To summarize, specifications affected visualization preferences. When given no information about the communicative goal, participants rated the visualizations higher than when they were given a specification. Out of the three specification formats, participants rated visualizations higher when viewing insights compared to either learning objectives or assessments. This suggests that insights may have more of a direct mapping onto graphical features of the visualization. Even though there was a difference between preferences, we found that the ranks of designs among specification types were correlated with each other, indicating some similarities between formats.

## 5.2 Accuracy

While understanding preferences is interesting, we would ultimately like to know if specification formats lead participants to select more effective visualizations. We do so by analyzing the viewer accuracy on the assessments. We investigate three evaluative conditions: baseline, readability, and memorability (see Figure 5). Recall that baseline is the performance *before* seeing the visualization, readability is the performance *while* seeing the visualization, and recall is *after* seeing it (as described in section 4.3).

We start by examining if there were differences in average accuracy between the three conditions (baseline, readability, and memorability). An ANOVA showed that accuracy significantly differed by condition,  $F(2, 429) = 10.57, p < 0.001$ . Post-hoc tests with a Bonferroni correction showed that baseline accuracy is significantly lower than the readability accuracy ( $p < 0.001$ ) and memorability accuracy ( $p = 0.005$ ). Memorability accuracy is not significantly different from readability accuracy ( $p = 0.6$ ). Overall, the baseline accuracy for questions was approximately chance ( $M = 0.37, SD = 0.18$ ). As expected, readability accuracy ( $M = 0.49, SD = 0.24$ ) increased from the baseline of answering questions without the visualization. The memorability accuracy was between those two conditions ( $M = 0.45, SD = 0.25$ ), though not significantly different from readability accuracy. These results are consistent with expectations (see Figure 7). Without a visualization viewers perform the worst. While looking at the visualization viewers perform the best (in many cases they can decode the answer from the vi-

**Accuracy is lower in the baseline evaluation compared to visualization evaluations**

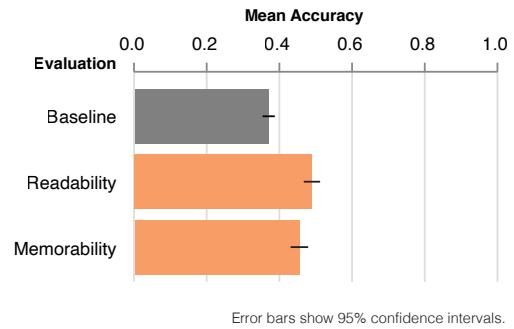


Fig. 7. The average accuracy for each of the conditions: no-visualization baseline, readability, and memorability. Error bars show 95% confidence intervals. Participants performed equally as good on the memorability assessments as they did on the readability assessments.

ualization). When we take the visualization away, viewer performance is slightly lower as viewers have to rely on their memory to answer the question (but not significantly different from when viewing the visualization). We note that performance in the memorability variant may decline as additional time or distractions are introduced.

While the general baseline-readability-memory pattern was consistent, we found that the baseline difficulty of the questions without a visualization varied. The baseline difficulty ranged from an accuracy of 0.11 (more likely to answer incorrectly than chance) to 0.80 (more likely to answer correctly than chance). We use the baseline difficulty to adjust questions to their underlying difficulty. We calculate the increase (or decrease) in accuracy from the baseline to the readability or memorability accuracy. This allows us to study the specific effect that the visualization has on accuracy.

## 5.3 Preference and Accuracy

We now compare the preferences to the accuracy to understand if the specification type influences participants to prefer better or worse visualizations. In this section, accuracy is defined as the increase (or decrease) in accuracy from the no-visualization baseline to the readability or memorability accuracy, as described above.

We computed Spearman’s correlation coefficient for each specification between preference and change in accuracy from the baseline for readability accuracy and memorability accuracy. There is a significant relationship between preferences for learning objectives and change in accuracy for both the readability accuracy and the memorability accuracy (see Table 3 and Figure 8). For insights, there is a significant relationship between preferences and change in accuracy for readability accuracy, but not memorability accuracy. For assessments, there was not a significant relationship between preferences and change in accuracy for neither readability accuracy nor memorability accuracy. Additionally, there is not a significant relationship between baseline preferences and accuracy. When looking at the difference between the two accuracy conditions, we find the same trends for accuracy in the memorability condition as the readability condition, but not as strong. These correlations suggest that learning objective specifications lead people to prefer better visualization designs.

A multiple regression model predicting the change in accuracy (baseline to readability) from baseline preference, learning objective preference, insight preference, and assessment preference was significant,  $F(4, 139) = 7.80, R^2 = 0.18, p < 0.001$ . Learning objective preferences and insight preferences significantly predict accuracy (see Table 4). Neither the assessment preferences nor the baseline preferences were significant predictors. Consistent with the correlation tests, we found similar results for a multiple regression model predicting change in accuracy for baseline accuracy to memorability accuracy,  $F(4, 139) =$

Specification	Readability		Memorability	
	r	p	r	p
Learning Objectives	0.32	< 0.001*	0.24	0.003*
Insight	0.28	< 0.001*	0.18	0.03
Assessment	0.11	0.19	0.04	0.65
Baseline	-0.02	0.85	0.07	0.41

Table 3. Spearman's correlation coefficients between preferences and the adjusted accuracy for each specification. Adjusted accuracy is defined as change from the no-visualization baseline to either the readability assessments or the memorability assessments. P-values are judged at the bonferroni-adjusted significance level of  $p = 0.0125$ .

Readability	b	Std. Err	p
Learning Objectives	0.1	0.03	< 0.001*
Insight	0.06	0.03	0.04*
Assessment	-0.02	0.03	0.5
Baseline	-0.07	0.05	0.1

Memorability	b	Std. Err	p
Learning Objectives	0.07	0.03	0.02*
Insight	0.07	0.03	0.02*
Assessment	-0.05	0.03	0.2
Baseline	-0.008	0.05	0.9

Table 4. Multiple regression models predicting accuracy from baseline preference, learning objective preference, insight preference, and assessment preference. The first model is predicting change from baseline accuracy to *readability* accuracy; the second model is predicting change from baseline accuracy to *memorability* accuracy.

4.91,  $R^2 = 0.12$ ,  $p < 0.001$ .

Furthermore, we found that designer preferences for graphs do not always line up with assessments. For each question, we identified the design that had the highest accuracy from the baseline and looked at preferences for that design. In a few cases, participants identified the graph with the highest accuracy as the *least* preferred. For example, question 11 - "Q: What is the Phillips Curve? A: The trade-off between unemployment and inflation" is most correctly answered (in the readability condition) with a very minimal line chart. However, participants rated that chart as least preferred for all specifications (see Figure 9). This shows that articulating a specification is important but sometimes insufficient. Thus, while assessments may not be the best specification format, they may be important in evaluating designer choices.

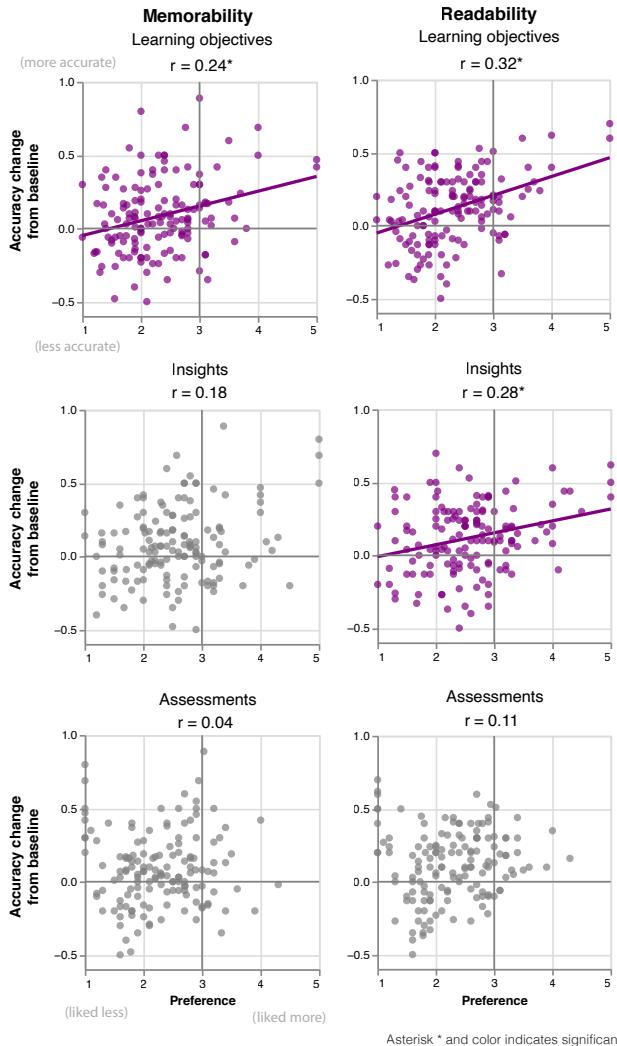
## 6 DISCUSSION

Our project investigated the effects of different specification types on selecting 'better' visualizations. We found that just having a specification changed preferences over no information at all. More importantly, specifications led participants to prefer more effective visualizations, while baseline preferences did not. Our result shows that it is important to define a communicative intent as part of the design process. Specifying a communicative intent can guide the designer to focus on a goal, and help them make better design choices towards this goal.

In our analysis of the specifications, we found evidence that the three specification forms are similar, but may lead to significant differences in some situations. The informational content in each is similar, so it is not surprising that they were correlated with each other (Table 3). However, the specification format can emphasize a different way of thinking about the goal. We found that preferences from learning objectives (and to a lesser extent, insights) were the most correlated with design effectiveness (Figure 8), suggesting they may be best overall.

While our results conclude that specifications help people choose better visualizations, it also reveals that they may not be sufficient to choose the *best* visualization. In some cases, the most effective designs for a specification may be intuitive. However, sometimes evaluations may reveal effective design options that would not have otherwise been considered. The analysis of the most effective design option illustrates this point (Figure 9). Participants generally preferred the most effective

Accuracy change from baseline and preferences



Asterisk \* and color indicates significance

Fig. 8. Scatterplots of preferences and accuracy change from baseline for the three specifications.

graph. However, occasionally they were very wrong, rating the most effective graph as their least preferred graph. Incorporating this evaluative feedback into the design process could help designers choose better graphs than preferences alone.

Our study also shows that having viewers complete the assessments after viewing the visualization helps evaluate the effectiveness of the design relative to communicative intent. Through the assessments, we evaluated both readability accuracy (answering the test question while looking at the visualization) and memorability accuracy (answering the test question after the visualization has been taken away). Our results demonstrate that while they are useful as a specification, assessments may be even more valuable in the design process as they help identify less effective choices. Because assessments are derived directly from the objectives and insights they can better ensure that the design choices made are good ones. Additionally, running multiple assessments may allow the designer to understand the trade-offs between different goals (i.e., when one visualization cannot 'maximize' effectiveness across multiple intents).

Finally, it is worth noting that in many situations designers are working in collaborative settings (e.g., with editors, journalists, clients, publishers, etc.). While an expert designer may create or select good visualizations on an independent project, a formal specification 'language'

## Preference ranks of the most effective visualization design

Designers generally prefer the most effective graph, but occasionally are very wrong. Light purple boxes indicates that the most effective design was the *least* preferred.

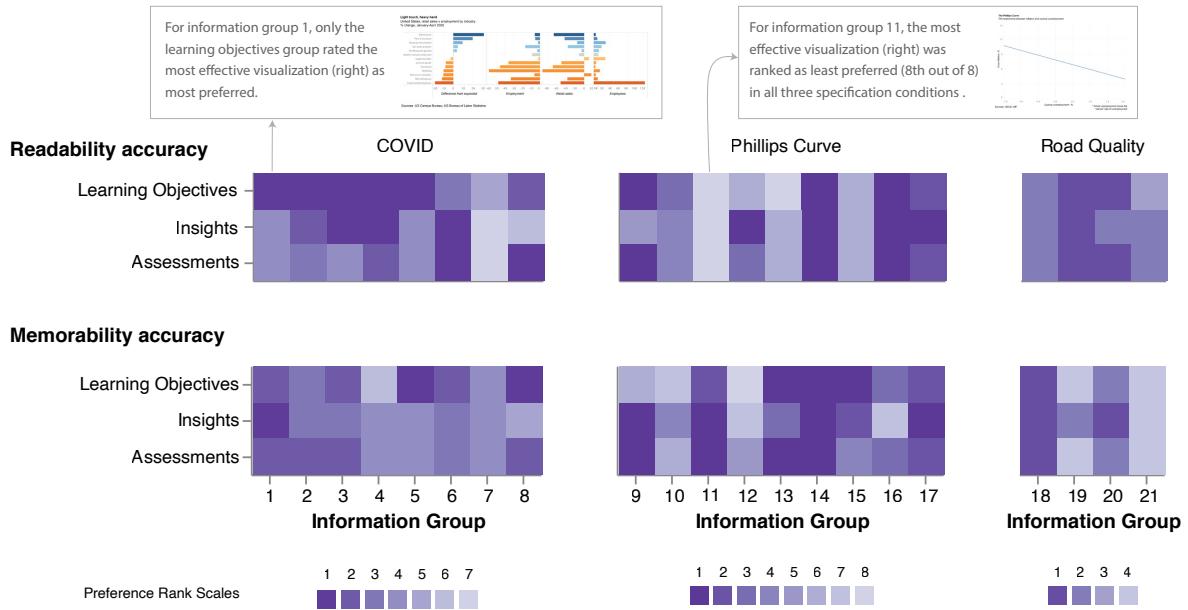


Fig. 9. The visualization with the highest accuracy (for both readability assessments and memorability assessments) was identified for each information group. We show the rank of that visualization for each of the specifications (Learning objectives, Insights, Assessments).

may enable better collaborations.

### 6.1 Limitations and Future Directions

Our study also demonstrates that creating learning objectives for visualization design is a possible way to formulate communicative intent. Although the participants in our study did not create the learning objectives themselves, they could meaningfully use them to choose better designs. As we demonstrate, it is feasible to create these learning objectives. However, while feasible, generating specifications is not easy. During the process of creating these learning objectives we ourselves needed multiple iterations to create the learning objectives. Previous work in this area has also noted that using learning objectives is difficult and needs to be learned and practiced [1]. We suggest that more resources be developed to help designers—or those working with them—to create learning objectives for visualizations (just as there are resources in the original domain to help educators create learning objectives in the context of classrooms). It is also worth acknowledging that many individuals work as visualization designers without any specific training (e.g., a scientist writing an article with figures). A formal specification may help guide these many ‘novice designers’ in building better visualizations. Even though there is a learning curve to learning objectives, we have shown an advantage to this specification form over the other types that we have studied.

Communicative visualizations are often embedded in a broader context: articles, presentations, reports. Our goal, in this first study, was to understand the effectiveness of visualizations independently of this additional context. While our specifications were extracted from the original materials, these intents were emphasized differently in the text in ways that could not be easily controlled for. Future studies can clarify the interaction between specification, context, and visualization(s).

In this study, we recruited crowd workers to rate different visualizations. The reality is that ‘designers’ in our broad sense, are common. Many people must create or pick visualizations without significant training. Thus, helping them specify their intent may lead to better outcomes. However, among this population, we observed a relatively low qualification pass rate. In some sense this is encouraging as it

ensured a high level of visual literacy among our participants. However, we hope to validate our work with other populations. For example, we would like to know if those with more training or experience in design similarly benefit from specifications (and which ones).

To allow for broad participation, we utilized pre-designed graphs instead of focusing on the visualization creation process. In the future, we hope to examine how different types of specifications will affect the design process. We hope to conduct in-depth qualitative design sessions to reveal more about how designers think about communicative intent and how they use it to make design choices. Finally, we hope to study further how the combination of learning objectives and assessments can not only guide communicative intent, but can be utilized as part of the design process to inform better design choices.

## 7 CONCLUSIONS

Communicative intent is an essential part of the visualization design process. Without clear intent, choosing what data to show and how to show it, may lead to worse choices. There are multiple ways to construct a statement of communicative intent (learning objective, insights, assessments). In this work, we identified which specification format was most likely to change baseline preferences for the better. While we found that learning objectives may be the best specification format, we saw improvements over the baseline for the other specification forms as well. Moreover, establishing a communicative intent makes it possible to evaluate the graphs based on the goal. If designers do not have a clear idea of what the visualization should be achieving, then it is hard to meaningfully evaluate the graphs other than just by aesthetics and generic designer guidelines. We encourage designers to articulate their communicative intent to guide more effective design choices.

## ACKNOWLEDGMENTS

We thank our anonymous reviewers for their feedback. We are grateful to the NSF for their support of this work through NSF IIS-1815760.

## REFERENCES

[1] E. Adar and E. Lee. Communicative visualizations as a learning problem. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):946–956, 2020.

[2] R. Amar, J. Eagan, and J. Stasko. Low-level components of analytic activity in information visualization. In *IEEE Symposium on Information Visualization, 2005.*, pp. 111–117. IEEE, 2005. doi: 10.1109/INFVIS.2005.1532136

[3] L. W. Anderson and D. R. Krathwohl, eds. *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. Allyn & Bacon, New York, 2 ed., December 2001.

[4] J. Bertin. *Semiology of Graphics: Diagrams, Networks, Maps*. University of Wisconsin Press, Madison, WI, 1983.

[5] J. B. Biggs and K. F. Collis. *Evaluating the quality of learning: The SOLO taxonomy (Structure of the Observed Learning Outcome)*. Academic Press, 1982.

[6] M. A. Borkin, Z. Bylinskii, N. W. Kim, C. M. Bainbridge, C. S. Yeh, D. Borkin, H. Pfister, and A. Oliva. Beyond memorability: Visualization recognition and recall. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):519–528, 2016. doi: 10.1109/TVCG.2015.2467732

[7] A. Cairo. *The Functional Art: An introduction to information graphics and visualization*. New Riders, 2012.

[8] S. K. Card, J. D. Mackinlay, and B. Schneiderman, eds. *Readings in Information Visualization: Using Vision to Think*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1999.

[9] R. Chang, C. Ziemkiewicz, T. M. Green, and W. Ribarsky. Defining insight for visual analytics. *IEEE Computer Graphics and Applications*, 29(2):14–17, 2009. doi: 10.1109/MCG.2009.22

[10] R. Chang, C. Ziemkiewicz, R. Pyzh, J. Kielman, and W. Ribarsky. Learning-based evaluation of visual analytic systems. In *Proceedings of the 3rd BELIV'10 Workshop: BEyond time and errors: novel evaLuation methods for Information Visualization*, pp. 29–34, 2010.

[11] Y. Chen, J. Yang, and W. Ribarsky. Toward effective insight management in visual analytics systems. In *2009 IEEE Pacific Visualization Symposium*, pp. 49–56. IEEE, 2009.

[12] R. C. Clark and C. Lyons. *Graphics for learning: Proven guidelines for planning, designing, and evaluating visuals in training materials*. John Wiley & Sons, 2010.

[13] N. Diakopoulos, F. Kivran-Swaine, and M. Naaman. Playable data: Characterizing the design space of game-y infographics. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, p. 17171726. ACM, 2011. doi: 10.1145/1978942.1979193

[14] Economist. The Phillips curve may be broken for good, 2017. <https://www.economist.com/graphic-detail/2017/11/01/the-phillips-curve-may-be-broken-for-good>.

[15] Economist. Italy spends lots fixing old roads, not enough building new ones, 2018. <https://www.economist.com/graphic-detail/2018/08/15/italy-spends-lots-fixing-old-roads-not-enough-building-new-ones>.

[16] Economist. American retailers have laid off or furloughed one-fifth of their workers, 2020. <https://www.economist.com/graphic-detail/2020/06/03/american-retailers-have-laid-off-or-furloughed-one-fifth-of-their-workers>.

[17] S. Few. *Show Me the Numbers: Designing Tables and Graphs to Enlighten*. Analytics Press, 2004.

[18] C. Forsell and J. Johansson. An heuristic set for evaluation in information visualization. In *Proceedings of the International Conference on Advanced Visual Interfaces*, pp. 199–206. ACM, 2010. doi: 10.1145/1842993.1843029

[19] N. E. Gronlund and S. M. Brookhart. *Gronlund's writing instructional objectives, 8th ed.* Pearson, 2008.

[20] T. M. Haladyna. *Developing and validating multiple-choice test items, 3rd ed.* Routledge, 2015.

[21] M. A. Hearst, P. Laskowski, and L. Silva. Evaluating information visualization via the interplay of heuristic evaluation and question-based scoring. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 5028–5033. ACM, 2016. doi: 10.1145/2858036.2858280

[22] T. Hogan, U. Hinrichs, and E. Hornecker. The elicitation interview technique: Capturing people's experiences of data representations. *IEEE Transactions on Visualization and Computer Graphics*, 22(12):2579–2593, 2016. doi: 10.1109/TVCG.2015.2511718

[23] F. Hohman, M. Conlen, J. Heer, and D. H. P. Chau. Communicating with interactive articles. *Distill*, 2020. <https://distill.pub/2020/communicating-with-interactive-articles>. doi: 10.23915/distill.00028

[24] T. Hollingsed and D. G. Novick. Usability inspection methods after 15 years of research and practice. In *Proceedings of the 25th annual ACM international conference on Design of communication - SIGDOC '07*, p. 249. ACM, 2007. doi: 10.1145/1297144.1297200

[25] J. Hullman, E. Adar, and P. Shah. Benefiting InfoVis with visual difficulties. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2213–2222, 2011. doi: 10.1109/TVCG.2011.175

[26] J. Hullman and N. Diakopoulos. Visualization rhetoric: Framing effects in narrative visualization. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2231–2240, 2011. doi: 10.1109/TVCG.2011.255

[27] P. Isenberg, T. Zuk, C. Collins, and S. Carpendale. Grounded evaluation of information visualizations. In *Proceedings of the 2008 conference on BEyond time and errors novel evaLuation methods for Information Visualization - BELIV '08*, pp. 1–8. ACM Press, 2008. doi: 10.1145/1377966.1377974

[28] Y.-S. Kim, K. Reinecke, and J. Hullman. Explaining the gap: Visualizing one's predictions improves recall and comprehension of data. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 1375–1386, 2017.

[29] G. Kindlmann and C. Scheidegger. An algebraic process for visualization design. *IEEE Transactions on Visualization and Computer Graphics (Proceedings VIS 2014)*, 20(12):2181–2190, 2014. doi: 10.1109/TVCG.2014.2346325

[30] J. Mackinlay. Automating the design of graphical presentations of relational information. *ACM Transactions on Graphics (TOG)*, 5(2):110–141, 1986. doi: 10.1145/22949.22950

[31] C. North. Toward measuring visualization insight. *IEEE Computer Graphics and Applications*, 26(3):6–9, 2006. doi: 10.1109/MCG.2006.70

[32] S. Nowak, L. Bartram, and T. Schiphorst. A micro-phenomenological lens for evaluating narrative visualization. In *2018 IEEE Evaluation and Beyond-Methodological Approaches for Visualization (BELIV)*, pp. 11–18. IEEE, 2018.

[33] W. J. Popham. *Test better, teach better: The instructional role of assessment*. Association for Supervision and Curriculum Development, 2003.

[34] J. Raible, L. Bennett, and K. Bastedo. Writing measurable learning objectives to aid successful online course development. *Int. J. for Scholarship of Technology Enhanced Learning*, 1(1), 2016.

[35] B. Saket, A. Endert, and Ç. Demiralp. Task-based effectiveness of basic visualizations. *IEEE transactions on visualization and computer graphics*, 25(7):2505–2512, 2018.

[36] E. Tufte. *The Visual Display of Quantitative Information*. Graphics Press, 2001.

[37] E. R. Tufte. *Beautiful evidence*. Graphics Press Cheshire, CT, 2006.

[38] C. Ware. *Information visualization: perception for design*. Morgan Kaufmann, 2019.

[39] G. Wiggins and J. McTighe. *Understanding by Design*. Association for Supervision and Curriculum Development, 2005.

[40] D. Wong. *The Wall Street Journal Guide to Information Graphics: The Dos and Dont's of Presenting Data, Facts, and Figures*. W. W. Norton, Incorporated, 2013.

[41] J. S. Yi, Y.-a. Kang, J. T. Stasko, and J. A. Jacko. Understanding and characterizing insights: How do people gain insights using information visualization? In *Proceedings of the 2008 Workshop on BEyond Time and Errors: Novel evaLuation Methods for Information Visualization, BELIV '08*. ACM, 2008. doi: 10.1145/1377966.1377971