Improving Visual Comparison Across Multiple Views with Shadow Marks

Adam Baker[1], Carl Gutwin[1], Justin Matejka[2], and Ian Stavness[1]

[1] Computer Science, University of Saskatchewan, `{first.last}@usask.ca`
[2] Autodesk Research, `Justin.Matejka@autodesk.com`

Abstract. Making spatial comparisons in visualizations with multiple views is often difficult (e.g., comparing locations is difficult across small multiples, and overlay techniques quickly become cluttered). To improve multi-view spatial comparisons, we developed Shadow Marks – visual marks that are replicated across all views – that provide a common spatial reference frame at the location where a comparison needs to be made. We evaluated Shadow Marks through three crowdsourced studies that explored different visual-analytics tasks. Study results show that Shadow Marks consistently led to more accurate answers than either a small-multiples grid or an overlay. Shadow Marks also required less effort and were strongly preferred, showing that user-controlled spatial reference frames can improve multiple-view comparison.

Keywords: Shadow Marks · Visual Comparison · Small Multiples

1 Introduction

Many visual analytics tasks involve making spatial comparisons across multiple similar views – for example, comparing movement paths in videos of novice and expert golfers, looking for outliers in related scatterplots, or assessing visual features in time-lapse images of crop trials. A common approach to showing multiple views involves *juxtaposition*, including techniques such as small-multiples displays [7], scatterplot matrices [22], trellis views [5], or side-by-side views [21, 42]. However, it can be difficult to compare specific locations across several views: for example, in a grid of videos showing plant growth under different conditions, the user might need to determine which plants is tallest, and even if all the videos have the same size and perspective, height comparisons can be difficult because there is no common reference frame.

A few techniques have been considered for this problem, but none provide a comprehensive solution. For example, videos sometimes include a physical ruler or grid sheet – but fine-grained comparisons are still difficult if the grid is not in the exact location of the comparison. Multiple views can be placed in the same spatial reference frame (e.g., overlays [13], split-screen views [1], or 'shine-through' representations [35, 63]), but overlay approaches quickly become cluttered with more than two views. Brushing, where a highlighted data point is marked in all views, only allows comparison of the same data point across views; tasks that involve different data points for each view (e.g., finding the largest outlier in a scatterplot matrix) cannot be solved with brushing.

To improve spatial comparisons across multiple views, we developed *Shadow Marks*, which let the user add a common spatial reference frame to a set of views. Shadow marks are visual features that are duplicated in all views; there are different types of marks that can be used for different kinds of comparisons, including points, lines, shapes, and a replicated-cursor tool. The user can place a Shadow Mark at the location that is most important for making a fine-grained comparison: e.g., to compare outliers in a set of scatterplots, the user can place a horizontal reference line at one plot's outlier; the line appears in all the videos, and so can be used to compare other outliers to the mark.

Although techniques similar to Shadow Marks have been seen in a few limited contexts [66, 29, 8, 37], they are not implemented in the vast majority of multi-view visualization systems, and the basic idea of user-controlled local reference frames for comparison tasks has never been evaluated. Therefore, we carried out crowdsourced studies (N=101) with three different spatial comparison tasks: plant growth extent, baseball video registration, and scatterplot outliers. The studies compared Shadow Marks to two other methods that are the current standard for multiple-view visualization: a small-multiples grid, and an interactive overlay. The study accuracy and completion time as well as subjective measures about perceived effort, user experience, and preference.

The studies provide strong evidence that Shadow Marks can improve performance and experience in spatial visual-comparison tasks. First, participants were more accurate with Shadow Marks than the other techniques: Shadow Marks had substantially fewer errors than both the small multiples and overlay techniques; this is important because in many analytics tasks, there is no clear right answer, and so early accuracy is critical. Second, trials were completed faster with Shadow Marks in two of the three tasks (and in the other, small multiples was faster but at the cost of very low accuracy). Third, participants consistently rated Shadow Marks as better in terms of perceived accuracy and task difficulty, and strongly preferred Shadow Marks to both small multiples and overlays.

Our research makes two main contributions: we provide the Shadow Marks technique as a way to augment views with a reference frame that is specialized to the needs of the comparison task; and we provide empirical evidence demonstrating the effectiveness of Shadow Marks and reinforcing the importance of a common reference frame in three realistic analytics scenarios. Overall, Shadow Marks are an easy-to-understand and easy-to-implement tool that can substantially improve multiple-view comparison.

2 Related Work

2.1 Comparison in Information Visualization

Gleicher's taxonomy of comparison for information visualization divides the design space based on how the visual elements are presented for comparison: juxtaposition, superposition, and explicit encoding [20, 19]. *Juxtaposition* interfaces lay out views side-by-side or in a grid so that all views can be seen at once. Common visualization techniques that employ juxtaposition include Trellis displays [5] and the small-multiples method [7]. Small multiples in particular has become popular in the information-visualization community, and previous work has often incorporated this approach in the design of visualization systems [22, 49, 34, 50]. However, the complexity of comparisons in small-multiples displays can quickly increase under certain conditions. First, it is important that all visualizations are viewed in the same frame of reference – otherwise, visual differences from the framing of the data may be misinterpreted as comparative differences between the data. Second, the number of views displayed in a small-multiples arrangements affects comparison: a study by Hosseinpour et al. found that increasing the number of views in a small-multiples setup decreases accuracy linearly [24]. Previous work has studied ways to support the comparison of high-dimensional categorical data [28, 60], but juxtaposed comparison is always be constrained by the number of views. Designers also need to consider the size of the workspace, since showing all views can limit view resolution [40].

Superposition involves stacking two or more visualizations in the same space so that differences between the data can be observed within the same frame of reference. The most common form of superposition is an overlay, where transparency levels are controlled so layers are visible through one another. This technique has been employed in sports visualizations to show differences in ball or player movements – e.g., a baseball overlay system that used a trained model to draw the path of different pitches, superimposed for an easier comparison [13]. However, there is a limit to the number of views that can be stacked: filtering techniques and transparency control have been used to reduce occlusion [36], but overlays can quickly become cluttered with more than a few images [35]. Other superposition techniques provide interactive access to the bottom layer of a two-layer visualization, such as a 'peel back' interaction [63], a 'window-blind' slider [12], or a 'shine-through' method [63, 35].

Explicit encoding techniques take in datasets as input and produce new visualizations that directly represent the relationships between the data. This frees the user from maintaining a mental model of the differences between each dataset. Previous work has demonstrated how different relationships can be visualized, including similarities in multidimensional tabular data [32, 31, 33] or difference relationships between tables [44], versioning "diffs" in documents [43, 62, 57, 23, 10], or pairwise differences between line graphs in a "diff matrix" [59].

Researchers have also used animation as a support mechanism for comparing data over time (e.g., [56, 55, 54, 58]). Animations have also been used to communicate uncertainty visualizations, which are often difficult to interpret accurately. Hypothetical Outcome Plots (HOPs) [25, 27, 48] and Network Hypothetical Outcome Plots (NetHOPs) [67] show uncertainty measurements in data by flashing different 'draws' from the distribution in an animated style.

2.2 Linked and Coordinated Views

Data exploration regularly involves comparison of different features within a dataset. A common method for supporting exploration is presenting data in multiple linked views – often called *Multiple Coordinated Views* [53]. The main difference between small-multiples representations and multiple coordinated views is that small multiples shows different datasets using a repeated visual presentation (e.g., same axes and colors), whereas multiple coordinated views typically show one dataset multiple times in multiple ways – and often with visual connections between linked data points [40]. For example, coordinated-view applications have used various visualization formats to support exploration of time-series data [68], behaviour evolution in multiplayer online games [11], tree graphs [41], tabular data [51], and other visualization types that can be linked algorithmically [15]. Formal models have also been created to organize design and implementation of these systems [9, 46, 29, 26].

Linked highlighting – also known as *brushing* [40] – was one of the first techniques to support data exploration across multiple views [4, 6]. With this technique, users select (brush) a set of data points in one view, and the corresponding data points in another view are highlighted. This forms a visual link between the two views that can be used to identify associations between different attributes in the data. Progressively brushing through an entire dataset allows the user to view and compare the interdependencies between different variables. Brushing was originally used to analyze multivariate data visualized in scatterplot matrices – e.g., comparing the radiation level, temperature, and wind speed at various altitude levels in the ozone layer [4]. Subsequent studies have looked at brushing other forms of data such as node graphs [61], parallel coordinate plots [2, 22], scatterplot and permutation matrices, and Andrews curves [22].

The closest previous work to Shadow Marks includes a few techniques for maintaining spatial positioning across multiple views using a line [37] or crosshair [8, 29, 66]. Similar to the way that linked highlighting supports visual analysis by emphasizing dimensional correlations, these linked 2D elements support comparison by providing a visual anchor for the user to focus on as their eye moves between views that do not maintain a common spatial reference frame. However, the effectiveness of these visual elements was not assessed in this previous work – to our knowledge there are no prior studies that measure the effects of localized spatial annotations on multiple-view comparison accuracy or task completion time. We report on such a study later in the paper, after first identifying design issues for the Shadow Marks technique.

3 Shadow Marks Design

Multi-view spatial comparison tasks – such as determining whether an object's position in one video is the same in other videos – are difficult when the views do not contain consistent spatial references that can be used to anchor the comparison. In a small-multiples layout, often the only reference frame that is common to all of the views is the border of the view itself; but the edges and corners in this border are typically too far from the objects of interest to enable accurate comparisons. Overlay approaches solve this problem by putting all datasets into the same reference frame, but overlays quickly become cluttered.

The goal of Shadow Marks is to provide a spatial reference frame that appears in all views of a separated-view layout, in order to support spatial comparison tasks. There are two main characteristics of a Shadow Mark. First, they are *tem-*

Fig. 1: Shadow Mark types. Each mark is replicated in other views.

Table 1: Summary of Shadow Mark type properties and applications.

Mark	**Interaction Properties**	**Task Applicability**
Point	- Small and precise - Fixed size - Add/remove on press	Compare specific locations (e.g., compare the release location of a baseball for different pitchers)
Line	- Can be drawn at any angle - Bounded to view dimensions - Click and drag to resize	Compare slopes and angles (e.g, compare a novice golfer's club angle to an expert golfer)
Cursor	- Persistent and fast - Fixed line orientation - Visible while mouse cursor is hovering over view	Quickly compare horizontal and vertical extents (e.g., find the largest outlier in a scatterplot matrix)
Freehand line	- Flexible for different use cases - Unbounded - Click and drag to draw line	Compare arbitrary contours / paths (e.g., compare paths of different baseball pitches)
Oval / Rectangle	- Flexible for different areas - Bounded to view dimensions - Click and drag to resize	Compare sizes of different areas (e.g., compare the bounded areas of various parts of plants)

porary – intended to be used for a specific comparison and then removed. Second, they are *localized* – placed in the specific area of the view where the comparison takes place, and precisely located to enable specific spatial comparisons.

Different types of Shadow Marks are derived from basic comparison tasks previously identified in previous work [19, 52, 3] – for example, the task of comparing locations leads to the *Point* mark; similarly, comparing extents leads to *Line* marks and *Cursor* marks, comparing areas leads to *Oval and Rectangle* marks, and comparing contours leads to *Freehand line* marks. Other shape-based comparisons can easily be approximated with these core mark types. A summary of the interaction properties and applicability of each core mark type is shown in Table 1 and Figure 1.

Three other requirements for Shadow Marks influenced their design. First, the marks must have *high visibility* in order to be obvious while not obscuring the content that needs to be compared; we controlled the transparency of each mark and provided a customizable colour palette to prevent background blending. Second, Shadow Marks need to *minimize clutter* in localized comparisons where overlapping objects quickly become indistinguishable; we implemented several types of Shadow Mark so that many comparison tasks can be completed with a single mark. Third, adding and removing Shadow Marks must be *lightweight*, since they are temporary augmentations to a visualization; we developed a drawing-tool-like interface to allow easy creation of different mark types, and allow fast removal by hovering over a mark and pressing the D key.

4 Study Methods – Three Linked Studies

We tested Shadow Marks in three different visual-analytics tasks that involve spatial comparisons – each task was run as a separate study, and each task compared Shadow Marks to a standard small-multiples layout and an overlay mechanism. Tasks and interfaces for the three comparison techniques are shown in Figures 2 - 3. Ethical approval was received from the University of Saskatchewan.

4.1 Study Tasks

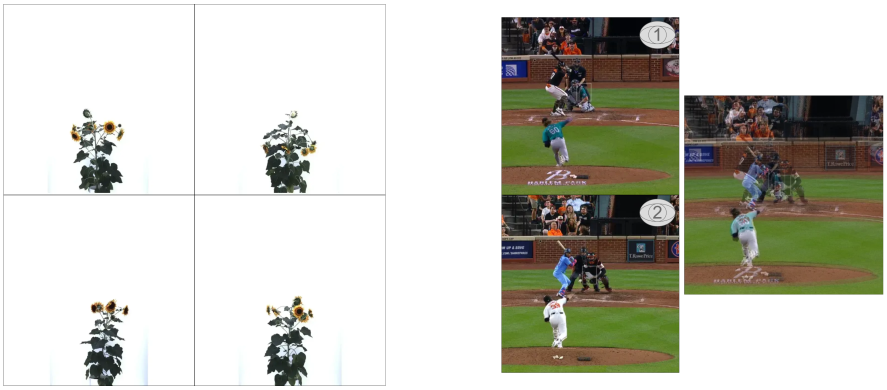
A custom web-based application (using HTML, CSS, and P5JS) implemented the study tasks and comparison conditions, and gathered performance and questionnaire data. The three tasks (run as separate studies) were:

Task/Study 1: Plant Growth Extent. Plant scientists often observe the growth cycles of plants and assess the extent to which they grow [17, 39, 14, 65, 45, 18]. The first study was built to emulate this process by instructing participants to analyze nine timelapse sunflower growth videos [16] that were all taken from the same distance and camera angle. Participants were asked to find the tallest sunflower at any time throughout its growth cycle (Figure 4a).

Task/Study 2: Baseball Video Registration. Comparing multiple videos requires that the objects of interest are *registered* (i.e., spatially aligned with one another); analysts often need to determine which videos are correctly registered.

The goal of the task in the second study was to find and select two baseball pitching videos, from a set of eight, that were correctly registered with a reference video. Baseball pitching videos were manually collected online from Baseball Savant [38]. Participants were asked to use the ball release point as the reference for assessing correct alignment (Figure 4b).

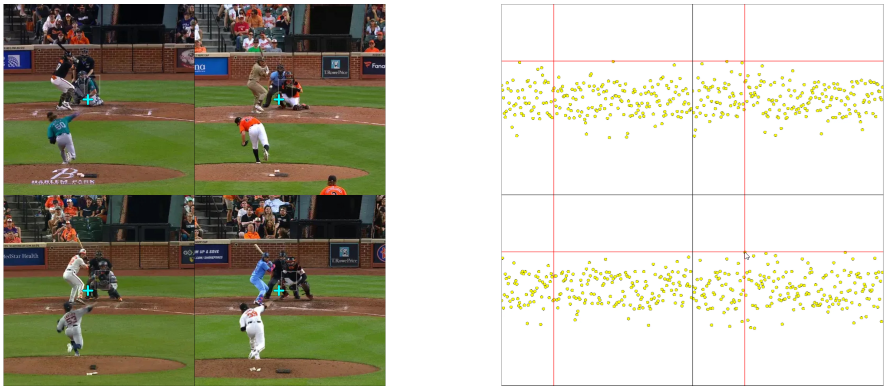
Task/Study 3: Identifying Scatterplot Outliers. In addition to videos, we wanted to test the Shadow Marks technique on static visualizations. Scatterplots were chosen because previous work has found that small multiples supports the comparison of these charts [54]. Sets of nine scatterplots were generated using a pseudo-random process (all plots had the same scale), and participants were asked to identify the plot containing the largest outlier (Figure 4c).



(a) S1: Plants with Small Multiples

(b) S2: Baseball with Overlays

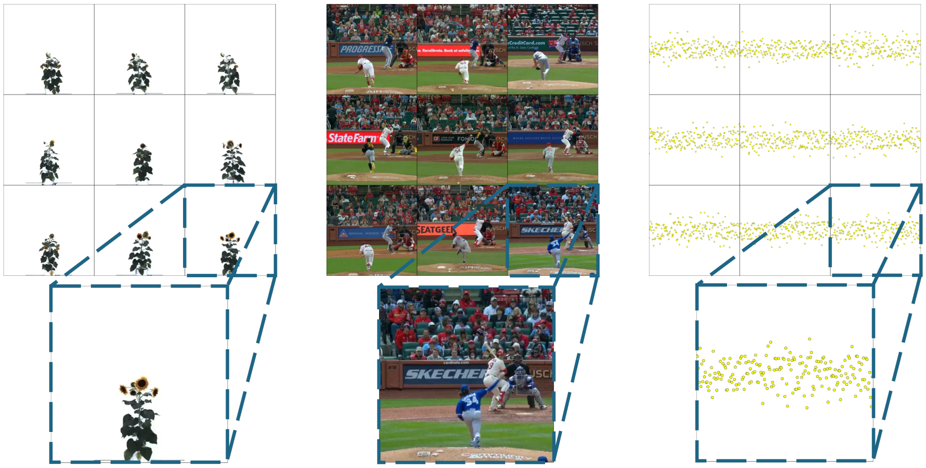
Fig. 2: Small-Multiples condition (left) and Overlay condition (right).



(a) S2: Baseball with Shadow Marks

(b) S3: Scatterplots with Shadow Marks

Fig. 3: Shadow Marks condition (Mark thickness and opacity are enhanced).



(a) S1: Plant Task (b) S2: Baseball Task (c) S3: Scatterplot Task

Fig. 4: Comparison tasks used in the three studies.

In all tasks, participants selected their answers by Control-clicking on the views and pressing the Enter key to submit. Users were informed of incorrect responses and were then allowed to try again until they found the correct answer (or a time limit was reached, as discussed below). The tasks were designed to require a high degree of precision; incorrect responses often deviated from the right answer by only 5-10 pixels.

4.2 Study Conditions

For each trial, nine datasets (either videos or images) were loaded into the system and placed in a 3x3 grid that was scaled to fill the screen (see Figures 4a - 4c). Users had the option to adjust the zoom level of the workspace once the data had loaded. Video displays were controlled by a global scrollbar at the bottom of the application that allowed the user to play through all videos at the same time. The system provided three different comparison techniques for each task:

Small Multiples. This technique places image or video displays in the default 3x3 grid. Display dimensions are manipulated to guarantee that participants can see all nine at once. Zoom controls are available if the user requires a detailed view of each display. Global video playback is controlled by a scrollbar.

Overlays. This technique allows views from the 3x3 grid to be copied and layered in a separate display area on the right side of the screen. Clicking on an image or video adds/removes it from the overlay stack. Once added, views in the 3x3 grid show a number in their top-right corner indicating the position on the stack. Display opacity is set automatically using a harmonic series (i.e., the bottom displays are fully opaque, and the opacities of successive layers are

1/2, 1/3, 1/4...) to ensure that all layers have equal visibility [36]. The global scrollbar controls overlaid videos in addition to the other views.

Shadow Marks. As described above, this technique allows the user to add visual marks to any image or video display in the 3x3 grid; marks are duplicated in all other displays at the same spatial location. We limited the number of available shadow mark types to one per study task to avoid overwhelming participants with multiple options; the mark type was chosen based on the type of comparison used for each task. The Shadow Mark used in S2 was the point mark, as this task required comparison of precise locations. The Shadow Mark used in S1 and S3 was the cursor mark, as this task required lightweight comparison of vertical extent. The mark types were implemented in the study systems as described above – for example, hovering over a Shadow Mark highlighted the display where it was created, and pressing D removed it from all displays.

4.3 Study Procedure

In all studies, participants completed an informed consent form and a demographics questionnaire, and were then randomly assigned to an order condition for the comparison techniques (fully counterbalanced). Participants then worked with each comparison technique individually: after reading through a tutorial page describing how to best use the technique for the task, they trained with each technique in a practice task (data discarded), and then carried out two trials for the task type used in that study (the three tasks were run as three separate studies). After each comparison technique, participants completed a questionnaire asking about effort (using the NASA TLX) and user experience (four questions related to whether the user had to guess using the technique, the accuracy of the technique, the overall task difficulty, and the difficulty of the technique itself). After all comparison techniques were complete, participants completed a final preference questionnaire that asked which technique was fastest, which was most accurate, and which was preferred overall.

Users were instructed to remain in fullscreen mode throughout the study to maximize the size of the workspace and limit window size variation. If the user moved their cursor outside the system window for more than 5 seconds, minimized the window, opened another window over the study window, or exited fullscreen mode, the system would ask that they return to their task.

4.4 Study Participants

For Study 1 (Plant growth), 35 participants completed the study on the Prolific crowdsourcing platform (17 women, 18 men; mean age 32.5 years). All participants were experienced with desktop PCs, but had variable experience with visual comparison (mean less than 1.5 hours/week). Participants were paid £2.50, and the study took ∼15 minutes. For Study 2 (Baseball), 30 participants completed the study on Prolific (15 women, 15 men; mean age of 33.3 years). All participants were experienced with PCs, but less so with visual comparisons (∼1 hour per week). Participants were paid £5.00, and the study took ∼30 minutes.

For Study 3 (Scatterplots), 36 participants completed the study on Prolific (16 women, 17 men, 3 Non-binary; mean age 35.1). All participants were experienced with PCs, but less so with comparing visualizations ($<$ 2 hours per week). Participants were paid £2.50, and the study took $\sim$15 minutes.

4.5 Study Design

All three studies used the same within-participants factorial design, with factor **Technique** (Small Multiples, Overlays, Shadow Marks). Dependent variables included performance measures (error count per trial, error distance per trial, and completion time per trial), as well as subjective ratings (effort questions, user experience questions, and preferences). The order of Technique was fully counterbalanced. With two trials per task and three comparison techniques, each participant completed six trials. In Study 1 (35 participants) 210 selection trials were recorded; in Study 2 (30 participants), 180 trials were recorded; in Study 3 (36 participants), 216 trials were recorded.

5 Results

Analyses are organized below by dependent variable (errors, completion time, perceived effort, and preferences); we present results for all three studies in each section. Effect sizes for significant ANOVA results are reported as generalized eta squared (η^2) [47], with $< .01$ considered small, .06 medium, and $> .14$ large [30]. Follow-up t-tests were corrected using the Holm-Bonferroni method. In all charts below, error bars show 95% confidence intervals. Outliers were removed based on pilot testing to determine reasonable upper bounds for completion of the tasks, both in terms of errors and completion time. For errors in all studies, any trial with more than 12 errors was removed (S1: 3 trials; S2: 3 trials; S3: 1 trial). For completion time, trials in Study 1 (Plants) and Study 2 (Baseball) were capped at 180 seconds; however, because the Baseball task was more difficult, we allowed trials to go beyond the 180-second threshold if the participant's guess after the threshold was correct. For Study 3 (Scatterplots), we reduced the time cap to 90 seconds because the task was easier due to the datasets being static visualizations. Following these rules, we capped 20 S1 trials, 37 S2 trials, and 28 S3 trials. Furthermore, if the participant's first guess exceeded the time cap, the entire trial was removed (S1: 9 trials removed, S2: 12 trials, S3: 5 trials). After removing outliers, one participant in S1 and two participants in S2 were removed because they exceeded time limits on their first guess in multiple trials. ANOVA results are shown in Table 2.

5.1 Error Counts and Error Distance

Our primary criteria for judging task performance is accuracy – finding the correct answer with fewer mistakes (even if it takes longer) is more important in a comparison task than being fast but incorrect. We took two measures of accuracy: the number of errors before the participant got the correct answer; and the error distance between their first guess and the correct answer.

Error counts ranged from about 1 to 4 errors per trial (Figure 5). Shadow Marks had the fewest errors (1.25 errors/trial for Plants, 1.79 for Baseball, and 0.88 for Scatterplots), followed by Overlays (1.8, 3.13, and 1.62), and Small Multiples (2.97, 3.42, and 2.97). These differences were significant: as shown in Table 2, Shadow Marks were more accurate than Small Multiples in all three studies, and were also more accurate than Overlays for the Baseball task (Overlays were also more accurate than Small Multiples).

Table 2: ANOVA results for all tasks (factor = Technique), for each DV.

DV	**F(DF),p**	η^2	**Pairwise Contrasts (mean), t-test result**
Plant Task			
Error Count	F(2,68)=9.76, p=**1.87**e-04	0.15	Sh.Marks (1.25) < Sm.Multiples (2.97), p=5.5e-05 Overlays (1.8) < Sm.Multiples (2.97), p=.0075
Error Distance	F(2,68)=4.09, p=**.021**	0.08	Sh.Marks (1.5) < Sm.Multiples (3.28), p=.0063
Time	F(2,68)=12.88, p=**1.81e**-05	0.12	Sm.Multiples (59.26) < Overlays (90.6), p=2.4e-04
Baseball Task			
Error Count	F(2,58)=7.32, p=**.0015**	0.12	Sh.Marks (1.79) < Sm.Multiples (3.42), p=9.6e-04 Sh.Marks (1.79) < Overlays (3.13), p=.0065
Error Distance	F(2,58)=8.95, p=**4.1e-04**	0.15	Sh.Marks (15.64) < Sm.Multiples (22.95), p=8.8e-05 Sh.Marks (15.64) < Overlays (21.80), p=8.7e-04
Time	F(2,58)=7.78, p=**.001**	0.1	Sh.Marks (82.95) < Sm.Multiples (106.18), p=.034 Sh.Marks (82.95) < Overlays (114.09), p=.0049
Scatterplot Task			
Error Count	F(2,70)=14.33, p=**6.1e-06**	0.17	Sh.Marks (0.88) < Sm.Multiples (2.97), p=1.8e-06 Overlays (1.62) < Sm.Multiples (2.97), p=.0021
Error Distance	F(2,70)=9.73, p=**1.87e-04**	0.08	Sh.Marks (5.49) < Sm.Multiples (10.3), p=.003 Overlays (6.36) < Sm.Multiples (10.3), p=.013
Time	F(2,70)=3.5, p=**.035**	0.05	(No pairwise differences)

Error distance was calculated by identifying the ‘critical location’ for each image or video (i.e., the maximum extent of each plant video, the initial position of the ball in the baseball videos, and the location of the largest outlier in each scatterplot), and then taking the difference between the critical location in the participant’s first guess and the location in the correct answer.

As shown in Figure 5, error distances followed a similar pattern to the error-count results. Shadow Marks had the smallest mean error distances, for all studies (1.5 pixels for Plants, 15.64 pixels for Baseball, and 5.49 pixels for Scatterplots), again followed by Overlays (2.64 pixels, 21.80 pixels, and 6.36 pixels), and Small Multiples (3.28 pixels, 22.95 pixels, and 10.3 pixels). Similar to the error-count analysis, the error distances were significantly different: as shown in Table 2, Shadow Marks had lower error distance than Small Multiples in all three studies, and had lower distances than Overlays for the Baseball task. These results indicate that even when a user does not get the absolute correct answer in a comparison task, Shadow Marks at least leads them to a more accurate guess than the other techniques.

5.2 Completion Time

We also measured completion time across all attempts to observe if there were trade-offs between speed and accuracy. As shown in Figure 5, Small Multiples required a large number of attempts but had the shortest average completion time in the Plant study (59.3s), followed by Shadow Marks (75.8s) and Overlays (90.6s). However, Shadow Marks was the fastest technique in the other two studies (82.9s for Baseball and 37.9s for Scatterplots), followed by Small Multiples (106.2s and 41.3s), and Overlays (114.0s and 47.4s). The completion time analysis revealed significant differences (Table 2): Small Multiples was faster than Overlays in the Plant study, and Shadow Marks was fastest in the Baseball study. In the Scatterplot study, there was a significant effect of Technique on completion time, and although Shadow Marks had the lowest mean time, our follow-up tests do not have sufficient power to identify further differences.

5.3 Subjective Measures: NASA TLX, UX, and Preference

After all trials with a technique, participants completed a NASA-TLX survey. After applying the Aligned Rank Transform (ART) [64], one-way ANOVAs were performed on each TLX question for factor Technique. Mean responses are shown in Figure 6 (complete ANOVA results are included in the appendices). Overall, participants reported that Shadow Marks required significantly less effort than Small Multiples and Overlays to complete their tasks (significant differences in favour of Shadow Marks for all three tasks on TLX items involving Effort, Performance, and Mental Demand, all $p<.05$).

Participants also rated their experience with each technique using four questions (7-point scales). We carried out one-way ANOVAs (ART-transformed) for factor Technique. Mean responses are shown in Figure 6 (see appendices for complete ANOVA results). Participants reported that they were more accurate while using Shadow Marks, which resulted in less guesswork during each trial (significant differences in favour of Shadow Marks on all tasks for UX questions involving Perceived Accuracy, Guessing, and Task Difficulty, all $p<.05$).

We also found that even though our participants had minimal experience in comparing videos or visualizations, there was no effect of Technique on questions about technique difficulty. Furthermore, participants found that the tasks were easier to complete while using Shadow Marks.

At the end of the study, participants answered questions about which technique was fastest, most accurate, and preferred overall. Mean responses are shown in Figure 6. In each study, participants preferred Shadow Marks on all three criteria (χ-squared tests confirmed significance, $p<.05$). Participants also

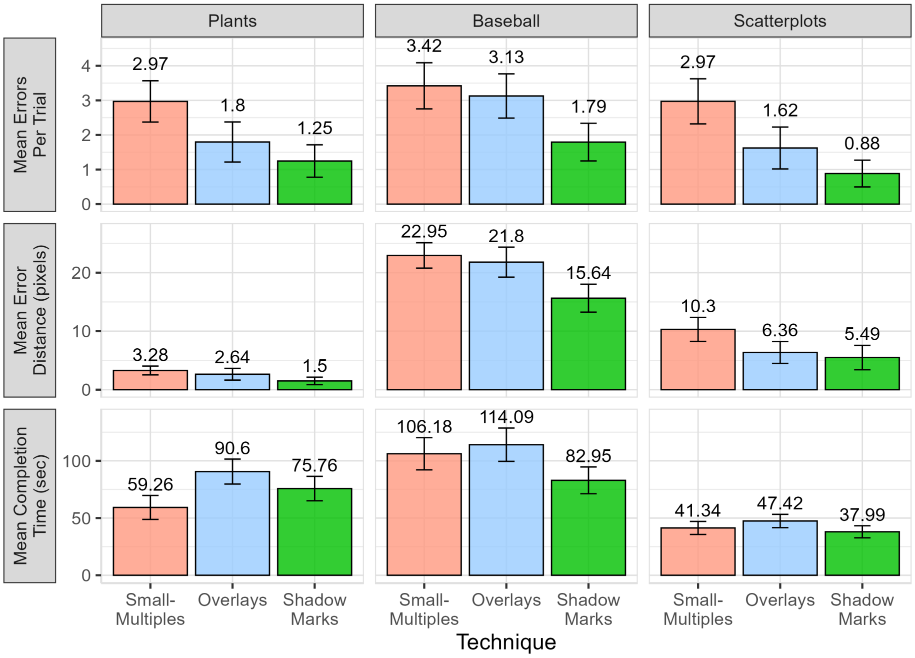


Fig. 5: S1-S3: Mean errors per trial, mean trial error distance, and mean trial completion time, by Technique

provided comments to explain their preferences. In the plant task, several participants found that Shadow Marks were the easiest due to the precision of the horizontal lines (e.g., "Once I [...] realized the flowers had different peaks, it was easiest to compare the sunflowers with these guided lines that did all the distance measuring for me"). Shadow Marks were also preferred for their natural connection to real-world tools (e.g., "you have a direct comparison of each of the nine boxes at their same relative squares like you do with a ruler").

In the baseball task, participants reported similar ease of use with Shadow Marks (e.g., "much easier because you had a clear point of reference to compare and you could watch frame by frame."). Another participant criticized Small Multiples, saying "multiples made me do a lot of guess work", and stated that "Overlays were hard to see". Shadow Marks were strongly preferred for being the "faster, clearer, and more intuitive method" for this type of task.

Several participants in the Scatterplot task discussed the different techniques' precision. One participant said that Small Multiples and Overlays "required a lot of squinting and double-checking"; in contrast, Shadow Marks were praised for how "the precision of the red line on all of the plots at once made it easier to compare and identify if there were any higher outliers at a glance".

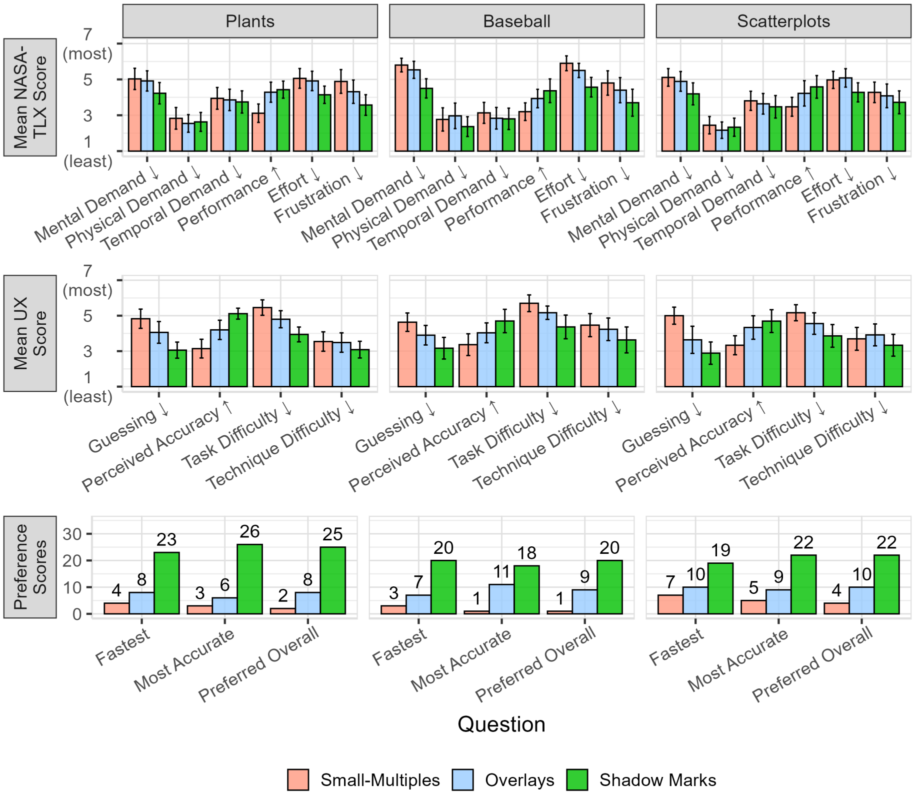


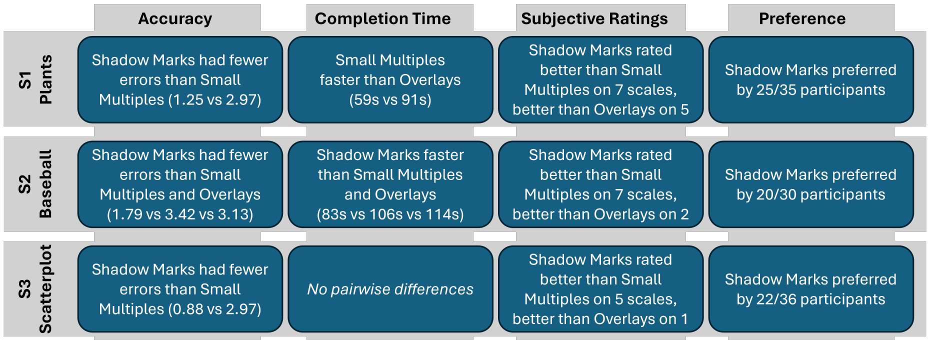
Fig. 6: S1-S3: NASA-TLX, UX, and Preference Questionnaire responses

6 Discussion

Our studies assessed the value of Shadow Marks – user-controlled spatial reference frames that are localized to a specific comparison task – in relation to two common multiple-view methods (small multiples and overlays). Two shadow mark types were evaluated in the studies, and the results provided a range of empirical evidence for the success of the approach, summarized in Figure 7.

6.1 Explanations for Main Results

Performance of Shadow Marks. The strong performance results, subjective results, and participant comments suggest that Shadow Marks worked as hypothesized (see Section 3) – that is, they provided a localized reference frame that made spatial comparisons easier and more accurate. In addition, it appeared that the advantages provided by the Shadow Marks were not outweighed by the overhead of setting up the marks, or by the need to learn how to use the marks (since none of our participants had used the technique before). It is



	Accuracy	Completion Time	Subjective Ratings	Preference
S1 Plants	Shadow Marks had fewer errors than Small Multiples (1.25 vs 2.97)	Small Multiples faster than Overlays (59s vs 91s)	Shadow Marks rated better than Small Multiples on 7 scales, better than Overlays on 5	Shadow Marks preferred by 25/35 participants
S2 Baseball	Shadow Marks had fewer errors than Small Multiples and Overlays (1.79 vs 3.42 vs 3.13)	Shadow Marks faster than Small Multiples and Overlays (83s vs 106s vs 114s)	Shadow Marks rated better than Small Multiples on 7 scales, better than Overlays on 2	Shadow Marks preferred by 20/30 participants
S3 Scatterplot	Shadow Marks had fewer errors than Small Multiples (0.88 vs 2.97)	*No pairwise differences*	Shadow Marks rated better than Small Multiples on 5 scales, better than Overlays on 1	Shadow Marks preferred by 22/36 participants

Fig. 7: Summary of results from all studies.

not surprising that providing a spatial reference simplifies comparisons that are relative to that reference, but our studies show that this benefit (a) was in fact realized in the Shadow Marks technique, and (b) was realized without causing other problems such as confusion, clutter, or extensive learning time.

The main advantage that Shadow Marks appeared to have over Small Multiples was in the provision of a reference frame in views where there is no common reference. The main benefit compared to Overlays – since Overlays also provide a common reference frame – was the reduction in clutter. Overlay users were forced into a more time-consuming strategy of comparing two images at a time; and in the Baseball task, the large amount of visual information in each video magnified the clutter problem, leading to reduced accuracy.

Although the value of adding a visual marker at the comparison location may seem obvious in hindsight, this idea has not been implemented in the vast majority of multi-view visualization systems. Our results clearly indicate that the point and cursor marks would be valuable additions to systems that compare precise locations and horizontal or vertical extents. For other use cases, designers can refer to Table 1 to choose which marks should be used.

Performance of Overlays. The Overlays technique performed well on two of the three tasks (Plants and Scatterplots), and many participants found successful strategies for making comparisons with this method. For example, a common strategy was to load one dataset into the overlay, and then flip a second dataset in and out of the overlay to look for visual changes. There are definite advantages to having data in exactly the same reference frame – but with clutter as the limiting factor (and the added time needed to deal with this issue).

The problem of visual clutter could possibly be addressed by altering videos and images to be more differentiable while in an overlay – for example, image-processing techniques such as an edge-detection filter, or making each layer a different colour, could be used to reduce occlusion in the Baseball task. However, clutter will always be a problem for this technique, and it is likely that making comparisons with small subsets of the views will always be a requirement.

However, for some tasks, Overlay's single location has advantages – it compares all views simultaneously, so the user can inspect many different parts of the scene, which could allow broader comparisons with complex data. In contrast, using Shadow Marks for multiple elements would require multiple marks, which could increase occlusion.

Performance of Small Multiples. Small Multiples was the least accurate technique in all studies, was rated as being worse on subjective measures, and was preferred by very few people (only 7 of 101). The poor performance points to the underlying problem with this technique: spatial comparisons are more difficult in small-multiples layouts, because after the user identifies a location in one view, it is difficult for them to find the same location in other separated views, particularly if there is no common reference frame in the images themselves, and the location is not near the edges or corners of the view.

Small multiples are often proposed for tasks in which people look for one dataset that is different from the others (e.g., a scatterplot trend line that is different than the majority); however, when differences between views are smaller and so do not visually pop out, comparisons become much more difficult.

Small Multiples were fastest in the Plants task, and this result does suggest that the technique can be used for quick assessment of several views. In the Plants task, users often scrubbed the videos back and forth, looking at the plants in relation to the top of the view, until they formed an impression of which plant had the highest extent. Even though Small Multiples had higher error counts, this strategy was fast and reasonably effective. People are highly capable of seeing differences in similar views – which is one of the motivations behind the technique [40]. However, this strategy worked less well when views were divided into multiple rows, requiring comparisons at different vertical positions. making it more difficult to compare videos in one row to those in another.

Overall, participant comments showed that even though people could get through the tasks with Small Multiples, this technique was frustrating and difficult. We note that since Shadow Marks build on the basic layout of Small Multiples, it seems likely that users will be able to make use of both strategies: making larger-scale comparisons with the regular Small-Multiples tool, and adding a Shadow Mark for finer-grained spatial comparisons.

6.2 Generalizing the Results

Our evaluation of Shadow Marks covered three different spatial comparison tasks that were based on realistic analytics scenarios. Individually, the studies were narrow in scope, but together, they highlight the flexibility of Shadow Marks in different application areas – i.e., plant science, sports analytics, and data visualization. We believe that our results are likely to generalize to a wide variety of real-world tasks that involve spatial comparisons across multiple views.

Shadow Marks were designed to be lightweight visualization augmentations that can be easily implemented. Marks are simple objects that hold a position, colour and a type, and the logic required for Shadow Marks is limited to drawing, spatial detection, and cursor tracking which are already nearly universal concepts in system design. For example, the cursor mark can be easily integrated into applications with juxtaposed views because it does not require any on-screen controls. Developers simply draw a horizontal and vertical line at the same spatial location as the cursor in every view. Persistent mark positions can be stored in the system model for retrieval at any time. Other attributes – such as line size, transparency, and colour – can be customized to address visualization-specific challenges (e.g., visibility issues that may arise in highly varied or colourful images may require shadows under lines or local dimming to make the marks pop out). This means that designers should be able to easily implement Shadow Marks into many types of visualization systems.

Only two of the six mentioned shadow mark types were tested in our evaluation. However, each mark type is related to a specific kind of visual comparison, and so we believe the observed benefits of providing a common spatial reference frame will extend to the other four mark types, provided that they are used in tasks that are appropriate for the mark (summarized in Table 1). Overall, our results provide designers with an easy-to-implement extension for precise visualization exploration and comparison. However, designers should also consider the complexity of the routine interactions in their system and the data being displayed. Shadow Marks work best on a large number of registered views where a common frame of reference is absent from the views but is needed for precise comparisons. Unregistered views require a non-trivial solution that is beyond the capability of regular Shadow Marks.

Finally, with the addition of minor interaction features, the value of Shadow Marks can also generalize to images of different resolutions, and thus to tasks with different precision requirements. In general, precision in a visual comparison task is related to image resolution (e.g., in our study tasks, the smallest comparisons that could be made were at the size of a pixel). Tasks that require greater precision can be supported with Shadow Marks by adding zoom features to the multiple-view setup: since a Shadow Mark has an absolute size in display space (e.g., the cursor mark draws a one-pixel line), zooming the image will change the ratio of the size of the mark to the size of the object to be compared. This allows for greater precision, as long as the source image is available at higher resolution. Zoom features need to be coordinated across the multiple views: the user must be able to zoom all views simultaneously (e.g., to see the immediate region of a point mark in all views), but must also be able to manipulate individual views (e.g., when comparing the height of objects that may appear in different parts of the image). We are currently adding these features to our implementation.

6.3 Limitations and Future Work

Expertise of participants. Our studies recruited participants who were not experts in video comparison or information visualization; and although the underlying reasons for the performance benefits of Shadow Marks are likely to apply to experts as much as to novices, we plan to carry out additional studies with experts to examine strategy use and performance when people have more experience with video, visualizations, and comparison tasks.

Number of trials. Because our studies used tasks with a high degree of complexity, we had participants complete only one practice trial followed by two recorded trials with each condition. We will carry out future studies on the effects of increasing experience with Shadow Marks.

Number of views to compare. Our study tasks asked participants to compare nine views, but many real-world scenarios involve many more views. We believe that the relative benefits of Shadow Marks will increase as the set size increases, and we are planning studies to investigate larger-scale comparison tasks.

Precision required in spatial comparisons. Our study tasks were appropriate for the image resolution, but we plan to test a wider range of precision requirements in future studies. As discussed above, these studies will be carried out with additional zoom capabilities added to the Shadow Marks technique.

In addition, we are planning extensions and improvements to the Shadow Marks technique – in particular, we will add capabilities for image processing and segmentation as new types of mark. For example, a tool that automatically draws a tight-fitting outline around an object can be used as an automated version of our freehand line tool, and an interactive segmentation system could be used to automatically create path marks (e.g., the path of the baseball in our pitching videos). It may also be possible to apply registration algorithms between views using Shadow Marks to amplify the strength of the technique.

7 Conclusion

Multi-view visualizations such as small multiples are often used to show several datasets in a single area – and although this approach shows a great deal of data at once, it can be difficult to make spatial comparisons across the multiple views because the datasets do not share a common reference frame. Overlays can address this problem by placing multiple datasets in the same view, but overlays quickly become cluttered if more than two datasets are added. Inspired by examples in visualization research, we developed Shadow Marks as a way to improve multi-view comparisons. Shadow Marks let users add visual marks that are replicated across all views – this can provide a common spatial reference frame at the exact location where a comparison needs to be made. We evaluated Shadow Marks against Small Multiples and Overlays in three crowdsourced studies that explored three different comparison tasks: comparing the extent of plant growth, checking the registration of baseball videos, and finding the largest outlier in a set of scatterplots. In all three studies, Shadow Marks was the most accurate of the three techniques, was rated best on subjective measures of effort, accuracy, and performance, and was strongly preferred over the other techniques. Our results indicate that tools for adding localized and temporary spatial reference frames can improve performance and user experience for comparison tasks in multiple-view visualization systems.

References

1. Adobe: Adobe lightroom (2024), https://lightroom.adobe.com/, accessed on September 9, 2024
2. Andrienko, G., Andrienko, N.: Exploring spatial data with dominant attribute map and parallel coordinates. Computers, Environment and Urban Systems **25**(1), 5–15 (2001)
3. Baker, A., Gutwin, C., Matejka, J., Stavness, I.: Interaction techniques for comparing video. In: Proceedings of the 50th Graphics Interface Conference. pp. 1–13 (2024)
4. Becker, R.A., Cleveland, W.S.: Brushing scatterplots. Technometrics **29**(2), 127–142 (1987)
5. Becker, R.A., Cleveland, W.S., Shyu, M.J.: The visual design and control of trellis display. Journal of computational and Graphical Statistics **5**(2), 123–155 (1996)
6. Becker, R.A., Cleveland, W.S., Wilks, A.R.: Dynamic graphics for data analysis. Statistical science **2**(4), 355–383 (1987)
7. Bertin, J.: Graphics and graphic information processing. de Gruyter (1981)
8. Booshehrian, M., Möller, T., Peterman, R.M., Munzner, T.: Vismon: Facilitating analysis of trade-offs, uncertainty, and sensitivity in fisheries management decision making. In: Computer Graphics Forum. vol. 31, pp. 1235–1244. Wiley Online Library (2012)
9. Boukhelifa, N., Roberts, J.C., Rodgers, P.J.: A coordination model for exploratory multiview visualization. In: Proceedings International Conference on Coordinated and Multiple Views in Exploratory Visualization-CMV 2003-. pp. 76–85. IEEE (2003)
10. Bremm, S., von Landesberger, T., Heß, M., Schreck, T., Weil, P., Hamacherk, K.: Interactive visual comparison of multiple trees. In: 2011 IEEE Conference on Visual Analytics Science and Technology (VAST). pp. 31–40. IEEE (2011)
11. Chen, W., Lu, J., Kong, D., Liu, Z., Shen, Y., Chen, Y., He, J., Liu, S., Qi, Y., Wu, Y.: Gamelifevis: visual analysis of behavior evolutions in multiplayer online games. Journal of Visualization **20**, 651–665 (2017)
12. Cherukuru, N.W., Scheitlin, T.: Visual comparator: An interactive tool for dynamic spatiotemporal comparative visualizations. Bulletin of the American Meteorological Society **101**(10), E1861–E1869 (2020)
13. Chou, T.H., Chu, W.T.: Automatic baseball pitch overlay. In: Proceedings of the 2021 International Conference on Multimedia Retrieval. pp. 599–602 (2021)
14. Coley, P.: Effects of plant growth rate and leaf lifetime on the amount and type of anti-herbivore defense. Oecologia **74**, 531–536 (1988)
15. Collins, C., Carpendale, S.: Vislink: Revealing relationships amongst visualizations. IEEE Transactions on Visualization and Computer Graphics **13**(6), 1192–1199 (2007)
16. Das Choudhury, S., Guha, S., Das, A., Das, A.K., Samal, A., Awada, T.: Flowerphenonet: Automated flower detection from multi-view image sequences using deep neural networks for temporal plant phenotyping analysis. Remote Sensing **14**(24), 6252 (2022)
17. Demir, S., Çelikel, F.G.: Effects of plant growth regulators on the plant height and quantitative properties of narcissus tazetta. Turkish journal of agriculture and forestry **43**(1), 105–114 (2019)
18. Fernandez, M.G.S., Becraft, P.W., Yin, Y., Lübberstedt, T.: From dwarves to giants? plant height manipulation for biomass yield. Trends in plant science **14**(8), 454–461 (2009)

19. Gleicher, M.: Considerations for visualizing comparison. IEEE transactions on visualization and computer graphics **24**(1), 413–423 (2017)
20. Gleicher, M., Albers, D., Walker, R., Jusufi, I., Hansen, C.D., Roberts, J.C.: Visual comparison for information visualization. Information Visualization **10**(4), 289–309 (2011)
21. Henry, N., Fekete, J.D.: Matrixexplorer: a dual-representation system to explore social networks. IEEE transactions on visualization and computer graphics **12**(5), 677–684 (2006)
22. Hinterberger, H.: The visulab: An instrument for interactive, comparative visualization. Technical Report/ETH Zurich, Department of Computer Science **682** (2010)
23. Hofmann, C., Boettcher, U., Fellner, D.W.: Change awareness for collaborative video annotation. In: Proceedings of COOP 2010: Proceedings of the 9th International Conference on Designing Cooperative Systems, May, 18-21, 2010, Aix-en-Provence. pp. 101–117. Springer (2010)
24. Hosseinpour, H., Matzen, L.E., Divis, K.M., Castro, S.C., Padilla, L.: Examining limits of small multiples: Frame quantity impacts judgments with line graphs. IEEE Transactions on Visualization and Computer Graphics (2024)
25. Hullman, J., Resnick, P., Adar, E.: Hypothetical outcome plots outperform error bars and violin plots for inferences about reliability of variable ordering. PloS one **10**(11), e0142444 (2015)
26. Javed, W., Elmqvist, N.: Exploring the design space of composite visualization. In: 2012 ieee pacific visualization symposium. pp. 1–8. IEEE (2012)
27. Kale, A., Nguyen, F., Kay, M., Hullman, J.: Hypothetical outcome plots help untrained observers judge trends in ambiguous data. IEEE transactions on visualization and computer graphics **25**(1), 892–902 (2018)
28. Kehrer, J., Piringer, H., Berger, W., Gröller, M.E.: A model for structure-based comparison of many categories in small-multiple displays. IEEE transactions on visualization and computer graphics **19**(12), 2287–2296 (2013)
29. Keller, M.S., Manz, T., Gehlenborg, N.: Use-coordination: Model, grammar, and library for implementation of coordinated multiple views. In: 2024 IEEE Visualization and Visual Analytics (VIS). pp. 166–170. IEEE (2024)
30. Levine, T.R., Hullett, C.R.: Eta squared, partial eta squared, and misreporting of effect size in communication research. Human Communication Research **28**(4), 612–625 (2002)
31. Lex, A., Schulz, H.J., Streit, M., Partl, C., Schmalstieg, D.: Visbricks: multiform visualization of large, inhomogeneous data. IEEE Transactions on Visualization and Computer Graphics **17**(12), 2291–2300 (2011)
32. Lex, A., Streit, M., Partl, C., Kashofer, K., Schmalstieg, D.: Comparative analysis of multidimensional, quantitative data. IEEE Transactions on Visualization and Computer Graphics **16**(6), 1027–1035 (2010)
33. Lex, A., Streit, M., Schulz, H.J., Partl, C., Schmalstieg, D., Park, P.J., Gehlenborg, N.: Stratomex: visual analysis of large-scale heterogeneous genomics data for cancer subtype characterization. In: Computer graphics forum. vol. 31, pp. 1175–1184. Wiley Online Library (2012)
34. Liu, J., Prouzeau, A., Ens, B., Dwyer, T.: Design and evaluation of interactive small multiples data visualisation in immersive spaces. In: 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR). pp. 588–597. IEEE (2020)
35. Lobo, M.J., Pietriga, E., Appert, C.: An evaluation of interactive map comparison techniques. In: Proceedings of the 33rd annual ACM conference on human factors in computing systems. pp. 3573–3582 (2015)

36. Matejka, J., Glueck, M., Bradner, E., Hashemi, A., Grossman, T., Fitzmaurice, G.: Dream lens: Exploration and visualization of large-scale generative design datasets. In: Proceedings of the 2018 CHI conference on human factors in computing systems. pp. 1–12 (2018)
37. McLachlan, P., Munzner, T., Koutsofios, E., North, S.: Liverac: interactive visual exploration of system management time-series data. In: Proceedings of the SIGCHI conference on human factors in computing systems. pp. 1483–1492 (2008)
38. MLB Advanced Media, L.: Baseball savant: Statcast, trending mlb players and visualizations (2024), baseballsavant.mlb.com, accessed on August 8, 2024
39. Moles, A.T., Warton, D.I., Warman, L., Swenson, N.G., Laffan, S.W., Zanne, A.E., Pitman, A., Hemmings, F.A., Leishman, M.R.: Global patterns in plant height. Journal of ecology **97**(5), 923–932 (2009)
40. Munzner, T.: Visualization analysis and design. CRC press (2014)
41. Munzner, T., Guimbretiere, F., Tasiran, S., Zhang, L., Zhou, Y.: Treejuxtaposer: scalable tree comparison using focus+ context with guaranteed visibility. In: ACM SIGGRAPH 2003 Papers, pp. 453–462 (2003)
42. Namata, G.M., Staats, B., Getoor, L., Shneiderman, B.: A dual-view approach to interactive network visualization. In: Proceedings of the sixteenth ACM conference on Conference on information and knowledge management. pp. 939–942 (2007)
43. Neuwirth, C.M., Chandhok, R., Kaufer, D.S., Erion, P., Morris, J., Miller, D.: Flexible diff-ing in a collaborative writing system. In: Proceedings of the 1992 ACM conference on Computer-supported cooperative work. pp. 147–154 (1992)
44. Niederer, C., Stitz, H., Hourieh, R., Grassinger, F., Aigner, W., Streit, M.: Taco: visualizing changes in tables over time. IEEE transactions on visualization and computer graphics **24**(1), 677–686 (2017)
45. Niklas, K.J.: Maximum plant height and the biophysical factors that limit it. Tree physiology **27**(3), 433–440 (2007)
46. North, C., Shneiderman, B.: Snap-together visualization: a user interface for coordinating visualizations via relational schemata. In: Proceedings of the working conference on Advanced visual interfaces. pp. 128–135 (2000)
47. Olejnik, S., Algina, J.: Generalized eta and omega squared statistics: measures of effect size for some common research designs. Psychological methods **8**(4), 434 (2003)
48. Padilla, L., Kay, M., Hullman, J.: Uncertainty visualization (2020)
49. Peña-Araya, V., Bezerianos, A., Pietriga, E.: A comparison of geographical propagation visualizations. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. pp. 1–14 (2020)
50. Preston, A., Ma, K.L.: Communicating uncertainty and risk in air quality maps. IEEE Transactions on Visualization and Computer Graphics **29**(9), 3746–3757 (2022)
51. Pretorius, A.J., Zhou, Y., Ruddle, R.A.: Visual parameter optimisation for biomedical image processing. BMC bioinformatics **16**, 1–13 (2015)
52. Quadri, G.J., Rosen, P.: A survey of perception-based visualization studies by task. IEEE transactions on visualization and computer graphics **28**(12), 5026–5048 (2021)
53. Roberts, J.C.: State of the art: Coordinated & multiple views in exploratory visualization. In: Fifth international conference on coordinated and multiple views in exploratory visualization (CMV 2007). pp. 61–71. IEEE (2007)
54. Robertson, G., Fernandez, R., Fisher, D., Lee, B., Stasko, J.: Effectiveness of animation in trend visualization. IEEE transactions on visualization and computer graphics **14**(6), 1325–1332 (2008)

55. Rosling, H.: The seemingly impossible is possible. In: TED (Technology, Entertainment, Design) conference presentation (2007)
56. Rosling, H.: Debunking myths about the "third world". ted (technology, entertainment, design) conference presentation (2006)
57. Schipper, A., Fuhrmann, H., von Hanxleden, R.: Visual comparison of graphical models. In: 2009 14th IEEE International Conference on Engineering of Complex Computer Systems. pp. 335–340. IEEE (2009)
58. Shin, M., Kim, J., Han, Y., Xie, L., Whitelaw, M., Kwon, B.C., Ko, S., Elmqvist, N.: Roslingifier: Semi-automated storytelling for animated scatterplots. IEEE Transactions on Visualization and Computer Graphics **29**(6), 2980–2995 (2022)
59. Song, H., Lee, B., Kim, B.H., Seo, J.: Diffmatrix: Matrix-based interactive visualization for comparing temporal trends. In: EuroVis (Short Papers) (2012)
60. Spechtenhauser, F., Piringer, H.: Diffpin: Interactive specification of references for comparative small-multiple displays. In: EuroVis (Short Papers). pp. 107–111 (2018)
61. Swayne, D.F., Buja, A.: Exploratory visual analysis of graphs in ggobi. In: COMPSTAT 2004—Proceedings in Computational Statistics: 16th Symposium Held in Prague, Czech Republic, 2004. pp. 477–488. Springer (2004)
62. Tam, J., Greenberg, S.: A framework for asynchronous change awareness in collaboratively-constructed documents. In: Groupware: Design, Implementation, and Use: 10th International Workshop, CRIWG 2004, San Carlos, Costa Rica, September 5-9, 2004. Proceedings 10. pp. 67–83. Springer (2004)
63. Tominski, C., Forsell, C., Johansson, J.: Interaction support for visual comparison inspired by natural behavior. IEEE Transactions on visualization and computer graphics **18**(12), 2719–2728 (2012)
64. Wobbrock, J.O., Findlater, L., Gergle, D., Higgins, J.J.: The aligned rank transform for nonparametric factorial analyses using only anova procedures. In: Proceedings of the SIGCHI conference on human factors in computing systems. pp. 143–146 (2011)
65. Yin, X., McClure, M.A., Jaja, N., Tyler, D.D., Hayes, R.M.: In-season prediction of corn yield using plant height under major production systems. Agronomy Journal **103**(3), 923–929 (2011)
66. Yushkevich, P.A., Gao, Y., Gerig, G.: Itk-snap: An interactive tool for semi-automatic segmentation of multi-modality biomedical images. In: 2016 38th annual international conference of the IEEE engineering in medicine and biology society (EMBC). pp. 3342–3345. IEEE (2016)
67. Zhang, D., Adar, E., Hullman, J.: Visualizing uncertainty in probabilistic graphs with network hypothetical outcome plots (nethops). IEEE Transactions on Visualization and Computer Graphics **28**(1), 443–453 (2021)
68. Zhao, J., Chevalier, F., Balakrishnan, R.: Kronominer: using multi-foci navigation for the visual exploration of time-series data. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 1737–1746 (2011)