Augmenting Visualizations with Predictive and Investigative Insights to Facilitate Decision Making

Md Main Uddin Rony
University of Maryland

Fan Du
Adobe Research

Ryan A. Rossi
Adobe Research

Jane Hoffswell
Adobe Research

Niyati Chhaya
Adobe Research

Iftikhar A. Burhanuddin
Adobe Research

Eunyee Koh
Adobe Research

ABSTRACT

Many people find it difficult to comprehend basic charts on the web, let alone make effective decisions from them. To address this gap, several ML models aim to automatically detect useful insights from charts and narrate them in a simpler textual format. However most of these solutions can only detect basic factual insights (a.k.a. *descriptive* insights) that are already present in the chart, which may help with chart comprehension, but not decision-making. In this work, we study whether more advanced *predictive* and *investigative* insights can help users understand what will happen next and what actions they should take. These advanced insights can help decision makers better understand the reasons behind anomaly events, predict future unfolding trends, and recommend possible actions for optimizing business outcomes. Through a study with 18 participants, we found that predictive and investigative insights lead to more insights recorded by users on average and better effectiveness ratings.

KEYWORDS

Visualization, data insights, descriptive insights, predictive insights

1 INTRODUCTION

Visualizations are powerful tools for communicating information. By using effective visualizations, readers can gain valuable insights from vast amounts of data and notice interesting features, such as trends and outliers. However, some visualizations can be too complicated for novice users to fully understand. For example, Romano et al. found that only 41% of respondents could correctly answer questions about log-scaled graphs [4]. To address this challenge, several recent research efforts [1–3, 5–8] have focused on automatically generating insight-centric captions, visualizations, and stories on the web. These automatically generated narratives often focus on describing facts present in the charts to make the information easier for users to consume. However, these facts are in a descriptive format that simply explains what is present in the chart. For example, the descriptive insight in Figure 1 summarizes the basic fact in the chart that *"The period from Apr. 25th to May. had the most significant continuous drop in the amount of revenue."* While these basic factual descriptions can help readers understand *what* the visualization shows, the responsibility for interpreting what the visualizations *mean* or how the user should proceed next remains up to the user.

WWW '23, May 1–5, 2023, Austin, TX, US
2023. ACM ISBN 978-1-4503-xxxx-x/18/06.
https://doi.org/10.1145/3366424.xxxxxxx

In this work, we extract basic statistical facts from the data, with the goal of ultimately providing more advanced predictive and investigative insights to support decision-making. To this end, we use trained machine learning models for prediction and further identify the differences between the predicted and actual values. Finally, we generate insights using predefined insight templates with both the basic facts and the ML-predicted values. To understand the effectiveness of these predictive and investigative insights, we conducted a user study with 18 participants comparing three configurations: NO-INSIGHTS, DESCRIPTIVE-ONLY, and ADVANCED-INSIGHTS.

2 FRAMEWORK FOR INSIGHT GENERATION

In this work, we focus on three insight categories: **descriptive, predictive**, and **investigative**.

2.1 Descriptive Insights

Descriptive insights report facts related to the metric present in the main visualization in a simple textual format; in other words, descriptive insights highlight information that can be directly extracted from the visualization, such as the min or max value for a given period. For example, *"The lowest amount of revenue happened in May. 3rd as 537.38. It was 20% less than average"*. This descriptive insight only shows the lowest revenue for a particular period. It also outlines the average trend (to compare). To generate descriptive insights, we focus on three basic types of data facts: **extremum (min, max)**, **cyclic pattern**, and **dramatic change**.

Extremum: For the given dataset D_s where $d_{a\in[1,n]} = \{t_a, v_a\}$, with a time unit t_a and a metric value v_a, the algorithm finds the index, i, where $v_i = max(v_1...v_n)$ or $v_i = min(v_1...v_n)$.

Cyclic Pattern: To detect notable cyclic patterns, we leverage a statistical test that first computes autocorrelation acf between a series of data D_s and a delayed copy over a particular window interval w, such that $acf = \textsc{AutoCorr}(D_s, w)$. A cyclic pattern is considered significant if the acf is higher than a threshold t_f. We then generate an insight describing the cyclic pattern.

Dramatic Change: We define dramatic change as a sudden increase or decrease in a metric. To do this, we compute the difference between the highest and lowest value for the given period, and if the dramatic change is equal to or smaller than a certain percentage (defined as the change ratio C) of that difference, the fact is considered significant. More formally, for a given dataset D_s where $d_{a\in[1,n]} = \{t_a, v_a\}$ consists of a time unit t_a and a metric value v_a, our algorithm identifies a range of index $[i, j]$ where $1 \leq i < j \leq n$ and $v_i > v_{i+1} > v_{i+2}... > v_j$ or $v_i < v_{i+1} < v_{i+2}... < v_j$, and $|v_j - v_i| \leq (max(v_1, ..v_n) - min(v_1, ...v_n)) * C/100$.

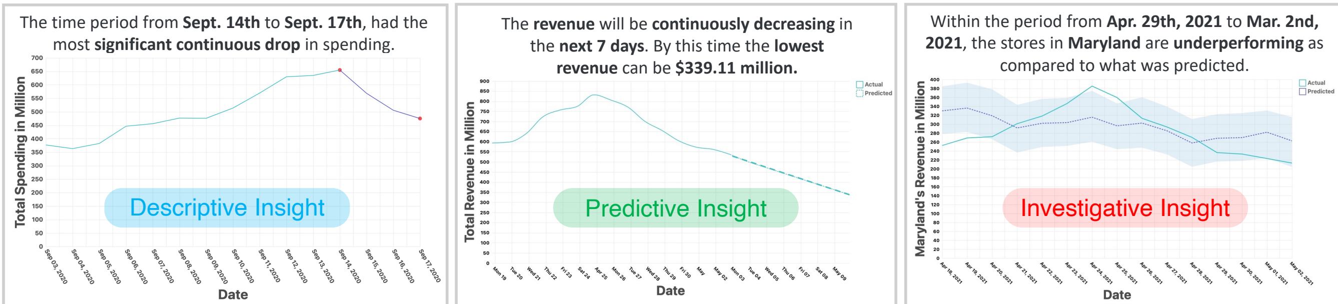

Figure 1: Overview of the different types of insights.

2.2 Predictive Insights

Based on the trends represented in the data, predictive insights state probable outcomes of future events. For example, *"The revenue will be continuously decreasing in 7 days. By this time the lowest revenue can be 339.11."* To generate predictions, we train a regression model M_r on the whole dataset of user uploads $D_f = [d_1, d_2, ..., d_m]$ to capture seasonality and trends where each $d_{a \in [1,m]} = \{t_a, v_a\}$ consists of a time unit t_a and a metric unit v_a. Then, the model predicts the potential outcomes for the future time unit (e.g., days). Suppose the user wants to forecast the next l time units, the model predicts the output on $t_o = \{t_{m+1}, t_{m+2}, ..., t_{m+l}\}$ as $v_o = M_r(t_o)$, where $v_o = \{v_{m+1}, v_{m+2}, ..., v_{m+l}\}$ is a set of predicted metric units. Then, we generate insights with similar data facts to $d_o = \{t_o, v_o\}$.

2.3 Investigative Insights

Since users may prefer insights that explain the reasons behind any event, we introduce investigative insights which explore the reasons behind dramatic changes in the data. For example, *"In Georgia, revenue decreased significantly from 2021-04-29 to 2021-05-03 as compared to what was expected."* To generate investigative insights, the algorithm explores one or more user-specified dimensions. For example, the user selected the 'location' dimension in the example above. More formally, suppose the user selected z dimensions $[dm_1, dm_2, ..., dm_z]$ in order to explore the metric v on time frame tm. The algorithm filters the data (D_s) based on each dimension (dm) and time range ($tm = \{t_1, t_2, ..., t_w\}$), and then applies regression analysis as described in Section 2.2 on the filtered data ($D_r = [t_m, v_a]$) to calculate the expected metric value (v_e). Here, $v_a = \{va_1, va_2, ...va_w\}$ is the set of original values and $v_e = \{ve_1, ve_2, ..., ve_w\}$ is the set of expected metric values. A subsequent step involves identifying *dramatic changes* (described in Sec. 2.1) from the original values and comparing them with the expected values in the time slot when they are detected. If there is a significant deviation between the actual and predicted value, the algorithm identifies this difference and generates an insight. When users do not select a dimension to explore, the heuristic first applies correlation analysis to the metric with the available dimensions to identify the most suitable one and then applies the same procedure.

3 EVALUATION

To evaluate the utility of predictive and investigative insights compared to basic visualization systems with and without descriptive insights, we conducted a within-subjects user study with 18 participants. We, therefore, focus on three configurations:

NO-INSIGHTS: This configuration contains only a static timeline visualization and does not include any natural language insights. This condition includes many existing visualization systems that lack natural language captions and insights.

DESCRIPTIVE-ONLY: This condition includes only *descriptive* insights (Sec 2.1), similar to systems that use automated captions [5, 6].

ADVANCED-INSIGHTS: This configuration includes *descriptive* insights as well as the more advanced *predictive* and *investigative* insights from Figure 1.

3.1 Hypotheses

The goal of this study was to understand how the different interface configurations affect users' engagement with the web-based visual analytics system. Our hypotheses are as follows:

- **H1:** Users will find ADVANCED-INSIGHTS more effective than either NO-INSIGHTS or DESCRIPTIVE-ONLY.
- **H2:** Users will record more insights with ADVANCED-INSIGHTS and DESCRIPTIVE-ONLY than with NO-INSIGHTS.
- **H3:** Users will prefer ADVANCED-INSIGHTS more than either NO-INSIGHTS or DESCRIPTIVE-ONLY.

We hypothesized that users would interact more with ADVANCED-INSIGHTS and write down more takeaways (**H2**) due to the larger number and variety in the insights that are provided, thus resulting in more user engagement with the system overall. Furthermore, given that ADVANCED-INSIGHTS provides insights beyond the basic (descriptive) level, we hypothesized that users would prefer (**H3**) ADVANCED-INSIGHTS over the other two configurations as it provides more in-depth information to enhance the data analysis process and can thus make the system more effective (**H1**).

3.2 Participants and Procedure

Participants: In this study, we recruited 18 graduate students (7 males, 9 females, 2 prefer not to say, age range 18-39 years) by email. All but two of the participants had prior data analysis experience (ranging from 0 - 12 years, M=4.26, SD=3.03).

Procedure: The study was conducted as an online survey using Qualtrics[1]. Participants were asked to complete the study in one sitting. After collecting the demographic information, participants were shown one of the three configurations mentioned earlier and asked to write down the main takeaways: "*Suppose, you are now working with the revenue data of company 'X'. Under the*

[1]https://www.qualtrics.com/

following configuration, you can see a figure showing the revenue over time for the last 14 days. Please write down your main takeaways/findings/understanding from this figure. (Write them down in bullet points)." With the DESCRIPTIVE-ONLY and ADVANCED-INSIGHTS configurations, an additional statement was shown: *"You can find some automatically generated insights in the following section to help you write your takeaways. Feel free to copy, paste, or edit them."* After using the configuration to identify the key insights, participants were asked to rate the effectiveness of the configuration they explored on a 7-point Likert scale and to explain their ratings.

Participants then repeated this task and rated the effectiveness of the other two configurations using different datasets. The order of the configurations was counterbalanced. The system automatically recorded the task completion time. Following the use of all three configurations, participants were asked to rank them and explain their preferences. The study took between 30 and 35 minutes to complete and each participant received $10 as compensation.

Datasets: Using realistic but simplified features, we synthesized three datasets for use in the evaluation. Each dataset contained 1020 records of archived business data. The datasets were temporal in nature and contained two categorical attributes: product (Shirt, Jacket, Pant, Watch, or Shoe) and location (Georgia, Maryland, or New York). Furthermore, each dataset contained a different numerical value (Dataset 1: Sales, Dataset 2: Spending, Dataset 3: Revenue). We included similar patterns (e.g., continuous increase or decrease) with random variations to maintain similarity between the datasets.

3.3 Quantitative Results

For each configuration, we calculated the average effectiveness rating and the average number of insights recorded by the participants. In addition, we performed post-hoc comparisons using paired t-tests with a 0.05 significance level. Participants were also asked to provide their preferences by ranking the three configurations.

3.3.1 Effectiveness. As reported in Figure 2A, participants found ADVANCED-INSIGHTS (M=4.61, SD=1.34) more effective than both the NO-INSIGHTS (M=4.22, SD=1.69) and DESCRIPTIVE-ONLY (M=4.33, SD=1.64) configurations. While ADVANCED-INSIGHTS has a higher average rating than the DESCRIPTIVE-ONLY configuration, the difference is not significant. However, a significant difference ($T(34) = -2.23, p < 0.05$) was found when comparing NO-INSIGHTS and ADVANCED-INSIGHTS, so **H1** is partially supported. We further discuss the participants' qualitative explanations in Section 3.4.1.

3.3.2 Preference. Out of 18 participants, 11 chose ADVANCED-INSIGHTS as their favorite configuration, while 6 chose DESCRIPTIVE-ONLY and 1 participant chose NO-INSIGHTS over the other configurations (Figure 2B). 12 participants chose DESCRIPTIVE-ONLY as their second favorite, and 3 participants each chose NO-INSIGHTS and ADVANCED-INSIGHTS as their second favorite. As for selecting the least favorite configuration, 14 participants selected NO-INSIGHTS, and the rest chose ADVANCED-INSIGHTS. Thus, **H3** appears to be true, though we further unpack the results in Section 3.4.2.

3.3.3 Insights Count. We counted the insights participants wrote down while exploring each configuration (Figure 3). In ADVANCED-INSIGHTS, users wrote an average of 4.56 insights (SD=1.83), higher than DESCRIPTIVE-ONLY (M=3.0, SD=1.0) and NO-INSIGHTS (M=2.83, SD=1.07) configurations. The pairwise comparison found significant differences between NO-INSIGHTS - ADVANCED-INSIGHTS ($T(34) = 3.35, p < 0.05$) and DESCRIPTIVE-ONLY - ADVANCED-INSIGHTS ($T(34) = -3.07, p < 0.05$), which confirms **H2**.

3.4 Qualitative Results

3.4.1 Effectiveness. Most participants rated ADVANCED-INSIGHTS as either "Extremely Effective" (3), "Effective" (8), or "Somewhat Effective" (3), thereby expressing a generally positive attitude toward the effectiveness of the more advanced types of insights compared to the other configurations (Figure 2A). In particular, ADVANCED-INSIGHTS was found effective for various reasons, including providing an in-depth understanding of the data and prompting analysis from multiple perspectives. For example, *P15* explained that *"I found [ADVANCED-INSIGHTS] extremely effective because it helped to go beyond my initial thinking. It gave me the idea of looking at the graph from other angles as well." P4* also explained how the predictive insights could prompt more careful analysis of the data, stating that *"since it was predicted to decrease, I went to examine the graph further to speculate the reason behind this drop."* Other participants particularly emphasized the analytical aspects of the takeaways and the utility of the computer-generated insights; *P1* noted that *"it seems like the automated suggestions were far more analytic than what humans could draw out from the graph."* Despite most participants finding ADVANCED-INSIGHTS effective, some participants noted that trust could be improved by providing more explanations for how the predictions were generated.

The DESCRIPTIVE-ONLY configuration was similarly found to be "Extremely Effective" (3), "Effective" (5), or "Somewhat Effective" (2) though to a lesser extent than ADVANCED-INSIGHTS, as this configuration only provided basic takeaways for participants to view without having to come up with them on their own (Figure 2A). Although participants liked the basic insights presented by the DESCRIPTIVE-ONLY configuration, some participants looked for more advanced insights; *P15* explained that *"These were relatively simpler insights. It would be interesting to see the results for more difficult trends, ones that require reference to different data ranges or visualizations."* Participants were more divided about the NO-INSIGHTS configuration (Figure 2A). The participants that liked it felt it was sometimes *"good enough"* (*P12*) or appreciated the simplicity (*P6*). However, many participants mentioned that it was difficult to identify insights without any support from the system; for example, *P1* noted that *"I think the basic stat analysis that could be drawn out from the above figure is pretty difficult. "* In particular, many participants noted that they were required to approximate the insights and could not otherwise describe the *"exact numbers"* (*P11*, *P12*, *P13*). Overall, while it is possible to extract insights from the visualization alone, providing more advanced insights can facilitate analysis by relieving some of the burdens on the user.

3.4.2 Configuration Preference. Most participants rated ADVANCED-INSIGHTS as either their "Most Favorite" (11) or "Second Favorite" (4) configuration (Figure 2B). Those who preferred ADVANCED-INSIGHTS were interested in capturing information beyond the basics provided in the DESCRIPTIVE-ONLY configuration. *P2* mentioned that *"I found [ADVANCED-INSIGHTS] to be most useful. For data analysis, I like to get as much information as I can from the*

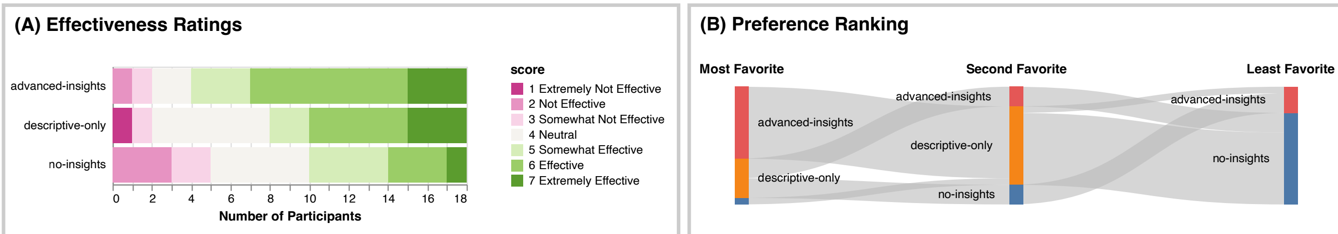

Figure 2: (A) Participants' effectiveness ratings for each configuration on a 7-point Likert scale (1=Extremely NOT Effective to 7=Extremely Effective). Participants generally found ADVANCED-INSIGHTS to be more effective than DESCRIPTIVE-ONLY or NO-INSIGHTS configurations. (B) Participants ranked the three configurations from their "Most" to "Least" favorite. 11 of 18 participants chose ADVANCED-INSIGHTS as their favorite, thus showcasing the effectiveness of the more advanced insights.

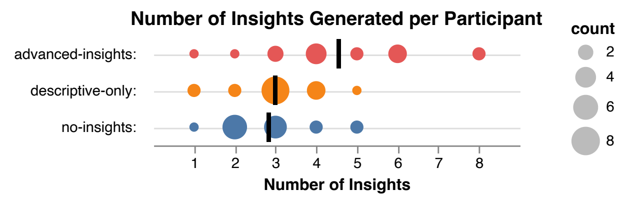

Figure 3: We recorded the number of insights written down by each participant for each configuration. Here, the circle size indicates the number of participants that wrote down the given number of insights, thereby showing the distribution of insights generated per configuration. The black line indicates the average number of insights generated per participant. Overall, participants generated significantly more insights with the ADVANCED-INSIGHTS configuration than either the NO-INSIGHTS or DESCRIPTIVE-ONLY configurations.

data." Also, *P18* preferred ADVANCED-INSIGHTS because it provided a more comprehensive overview of the analysis, which other configurations failed to offer: *"[ADVANCED-INSIGHTS] gave a more holistic perspective of the data."* For the DESCRIPTIVE-ONLY configuration, most participants (12) chose it as their "Second Favorite", and the rest (6) chose it as their "Most Favorite" configuration. Participants appreciated the DESCRIPTIVE-ONLY configuration because it offered the *"right amount of information"* (*P13*) and *"alleviated [the analysis] workload"* (*P6*) by capturing the major points quickly (*P7, P15*). For example, *P13* mentioned that *"I found [the DESCRIPTIVE-ONLY configuration] to provide the right amount of information, from which I could extract better insights."* In spite of its usefulness, some participants found the insights *"too basic"* (*P3, P16, P18*), therefore ranking it as their "Second Favorite" configuration. The majority of participants rated the NO-INSIGHTS configuration as their "Least Favorite" (14) due to it being *"too basic"* (*P3*), difficult to interpret (*P1, P5, P14, P17*), and time-consuming (*P15*). Overall, these results emphasize the importance of including insights in visualization systems to improve understanding of the visualizations, particularly for novice users; though more advanced insights take additional time to understand, such insights can improve the analysis process by providing a deeper look into the data and suggesting new avenues to explore that may otherwise be overlooked.

3.4.3 Insight Count and Timing. We hypothesized users would record more insights when given ADVANCED-INSIGHTS in addition to descriptive insights. Our statistical analysis confirmed this hypothesis (**H2**), as discussed in Section 3.3.3. Based on the users' explanations, we found that ADVANCED-INSIGHTS enabled them to speculate more about the data, which made it easier for them to gather more insights. For example, one participant pointed out that: *"Descriptive insights help grasp the main points faster (e.g. min, max, average). ADVANCED-INSIGHTS was effective in reflecting on the data – since it was predicted to decrease, I went to examine the graph further to speculate the reason behind this drop"* (*P4*). As shown in Figure 3, the average number of insights identified by participants in the NO-INSIGHTS (M=2.83, SD=1.07) and DESCRIPTIVE-ONLY (M=3.0, SD=1.0) configurations were almost the same, but participants took more time in the DESCRIPTIVE-ONLY configuration (7.94 minutes) than in the NO-INSIGHTS configuration (4.51 minutes). Similarly, providing ADVANCED-INSIGHTS helps users discover even more insights (M=4.56, SD=1.83) than the other configurations, but again requires participants to spend more time (9.48 minutes) understanding the results. For example, *P5* commented on the efficiency of ADVANCED-INSIGHTS in identifying a variety of new and more complex insights: *"The descriptive insights are helpful, but it gives only surface ideas. But [ADVANCED-INSIGHTS] gives more insights about the graph which I found effective."* While it may take additional time and effort for users to fully understand these more complicated results, *P12* felt that *"the additional [ADVANCED-INSIGHTS] are probably worth spending some time."*

REFERENCES

[1] Rui Ding, Shi Han, Yong Xu, et al. 2019. Quickinsights: Quick and automatic discovery of insights from multi-dimensional data. In *SIGMOD*. 317–332.
[2] Can Liu, Liwenhan Xie, Yun Han, Datong Wei, and Xiaoru Yuan. 2020. AutoCaption: An approach to generate natural language description from visualization automatically. In *PacificVis*. IEEE, 191–195.
[3] Xin Qian, Eunyee Koh, Fan Du, Sungchul Kim, Joel Chan, Ryan A Rossi, Sana Malik, and Tak Yeon Lee. 2021. Generating accurate caption units for figure captioning. In *WWW*. 2792–2804.
[4] Alessandro Romano et al. 2020. The public do not understand logarithmic graphs used to portray COVID-19. *LSE COVID-19 Blog* (2020).
[5] Danqing Shi, Xinyue Xu, Fuling Sun, Yang Shi, and Nan Cao. 2020. Calliope: Automatic visual data story generation from a spreadsheet. *TVCG* 27, 2 (2020).
[6] Arjun Srinivasan, Steven M Drucker, Alex Endert, and John Stasko. 2018. Augmenting visualizations with interactive data facts to facilitate interpretation and communication. *TVCG* 25, 1 (2018), 672–681.
[7] Bo Tang, Shi Han, Man Lung Yiu, Rui Ding, and Dongmei Zhang. 2017. Extracting top-k insights from multi-dimensional data. In *Proceedings of the 2017 ACM International Conference on Management of Data*. 1509–1524.
[8] Kanit Wongsuphasawat, Dominik Moritz, Anushka Anand, Jock Mackinlay, Bill Howe, and Jeffrey Heer. 2015. Voyager: Exploratory analysis via faceted browsing of visualization recommendations. *TVCG* 22, 1 (2015), 649–658.