The Role of Expertise on Insight Generation from Visualization Sequences

Stephanie Rosenthal
Computer Science Department
Carnegie Mellon University
Pittsburgh, USA
srosenth@andrew.cmu.edu

Tingting (Rachel) Chung
Mason School of Business
College of William & Mary
Williamsburg, USA
rachel.chung@mason.wm.edu

***Abstract*—Data analysts often tediously create visualization sequences to derive insights about what they see. While recent AI-driven approaches generate sequences to optimize visualization appeal and individual user preferences, extended cognitive fit theory suggests that expertise and insight type will affect the visualizations that analysts prefer. To investigate the role of expertise on insight generation from visualization sequences, we asked data scientists and accountants to report their insights as they investigated two business datasets. We found that both groups frequently followed the visualization sequences in order. However, expertise played a role in predicting the types of visualizations that each group chose to visit when they had finished the sequence but had time remaining. We also found significant interaction effects of visualization type, insight type, and expertise when assessing the numbers of insights generated per participant. Based on these results, we recommend that AI-driven data visualization tools should incorporate expertise as a feature for predicting new visualizations to produce.**

***Index Terms*—visualization sequences, expertise, cognitive fit, insight generation, data analysis**

I. Introduction

Visualization is an important aspect of data analysis that helps analysts quickly identify patterns and find errors and outliers in data [1]–[3]. However, the analyst must first overcome the challenge of determining what information is interesting and how to visualize the relevant data in order to be able to generate the best insights (analyses or discoveries of information) about their data in an efficient manner [4]. The cognitive fit theory [5] suggests that different visualizations are useful for identifying different insights. For example, tables are critical for finding symbolic insights, such as missing data or negative values, while graphs are more helpful for finding spatial insights, such as data trends [4]. Cognitive fit theory has been extended to demonstrate that expertise also affects the types of visualizations that different people find most helpful for analyzing data. Expertise in a particular domain (e.g., biology versus statistics) and in terms of the number of years studying that domain have both been examined for their impact on the interpretation of visualizations [5]. This is because expertise or domain knowledge can critically influence mental representations, and the extent of the fit between mental representations versus visual representations of the task can alter task performance.

Recently, there has been interest in automatic generation of visualizations [6]–[8] because it could be more efficient to present an analyst with candidate visualizations than for the analyst to develop those visualizations on their own. For example, computers can test more types of visualizations and more combinations of data than humans. Taking automated visualization generation a step further includes developing sequences or stories or dashboards of charts. Analysts rarely explore only a single visualization. These series can be created to drill down on a particularly interesting aspect of the data or to compare similar data features across several topics [8], [9]. Visualization generation research has used machine learning techniques, such as active learning, to learn visualizations that are “interesting” to the analyst in the hopes of improving a visualization sequence over time. However, it has not focused on personalizing the visualization sequences based on domain expertise, despite strong evidence that it does impact analysts’ interests and interpretation of data.

In this work, we conducted a study to understand how domain expertise can drive visualization sequence selection and insight generation. We asked data scientists and accountants to make insights about two different datasets, each presented as a sequence of 10 visualizations that were numbered but could be selected for viewing in any order. For each set of visualizations, participants spent 30 minutes logging patterns, outliers, and other insights that they observed along the way. We compared the sequences of visualizations that participants visited, the visualization types that they visited and in what order, and the number and types of insights that they created.

We found that data scientists and accountants both visited visualizations in numbered order to begin. However, when they reached decision points (e.g., the end of the sequence), our results mirrored the extended cognitive fit theory prediction that different domain experts gravitate towards different visualizations. Specifically, our data scientists revisited 2D visualizations more frequently than accountants, who preferred tabular views. We also compared the number of insights generated based on insight type (e.g., finding outliers, patterns, time trends, business common sense, etc.), visualization type (table, 1D plot, 2D plot, line plot), and expertise. Both groups

978-1-6654-4214-5/22/$31.00 ©2022 IEEE

generated similar numbers of insights in total but focused those insights on different plots and insight types.

Based on these findings, we conclude that visualization sequences significantly impact how data analysts view data. Domain expertise was not found to be a main effect but significantly interacted with other variables to produce differences in sequence visit order and number of insights generated. We recommend that creators of dashboards and automated visualization tools should incorporate expertise into design considerations when determining which visualizations people may want to view and the types of data analysis tasks with which they are most comfortable.

II. Related Work

Interactive data visualization technologies, such as Tableau and Microsoft Power BI, enable users to display different representations and arrangements of data, show how related data items are connected, select data items of interest, show or hide details of a dataset, filter data based on specified conditions, and explore data in open-ended manners [3]. Users of these tools identify trends, relationships, and anomalies in datasets [1]–[3]. Consequently, visualization tools are quickly becoming an essential part of exploratory data analysis and validation critical for machine learning and AI work [10]–[12] as well as business operations [13]–[16]. One of the most valuable benefits of visualization is for the user to derive insights for making decisions. Insights require analysts to understand their data beyond simply perceiving values on a graph or table or answering factual questions. North and colleagues [17]–[19] demonstrate that data visualization tools are effective in eliciting user insights and hypotheses in an unguided environment. In this work, we are interested in how people use visualizations to make insights about their data.

1) Visualization Sequences: Visualization generation tools also allow users to explore the same dataset using different graphics sequentially [9]. Nowadays, a single visualization is rarely the end-product of many data analysis tasks. Modern programs for interactive data visualization allow users to explore multiple visualizations of the same dataset simultaneously and to create dashboards for others to make insights as well.

Recent work in visualization sequences focuses on automating and predicting both the analyst's desired visualizations and the order for viewing those visualizations. Some tools present visualizations in a sequence by optimizing statistical properties of the dataset [6]–[8]. For example, GraphScape [7] sequences visualizations based on the transition costs of editing operations between chart specifications and a global cost function that rewards consistent chart specifications. VizPilot [8] creates a sequence of visualizations by prioritizing charts that are distinct from each other. These approaches use a sequencing method based on an algorithm that maximizes some assumed utility function, and does not incorporate user feedback. More recently, Cao et al. [9] designed a system that allowed the user to label visualizations with personal preferences; then, the system would generate a sequence of visualizations by predicting user preferences based on those labels. User interactions with these visualization systems can be captured using reVISit, a platform that logs navigation behaviors for usability studies [20]. While these approaches are innovative in their own right, they shed little light on how expertise may influence effectiveness of visualization types of sequences. In this work, we are interested in how people with different domains of expertise explore visualization sequences to generate insights.

2) Matching Visualization Type to Insight Tasks: In parallel to the work on advancing algorithms for generating visualizations and visualization sequences, there has also been work in understanding what kinds of visualizations would help people be most efficient at particular data insight tasks. Cognitive fit theory [5] is a subset of cognitive cost-benefit theory that specifies the trade-off between cognitive effort versus accuracy when making decisions [21], [22]. The theory was first introduced to explain performance differences in the use of different visualizations (e.g., graphs versus tables). Cognitive fit occurs when problem representation (graph vs. tables) matches the problem task [5]. For example, a spatial task is one in which the user must find a trend in data, while a symbolic task is one in which the user must find a relevant number. Finding an outlier in data may be easier to do spatially rather than looking through a large table of values. Conversely, finding missing values may be easier to do within a table. In this work, we will assess our data visualization for business applications within the context of cognitive fit to understand how visualization use is impacted by the different types of insights that people want to make during data exploration and the order in which the visualizations are presented.

It has long been recognized that properties of mental structures can greatly affect the benefits of graphical representations [23], [24]. As such, the cognitive fit theory has been extended to consider mental representation, task problem representation, and the task itself [25]. Individual characteristics are widely believed to affect mental representation, and domain knowledge or expertise is one key characteristic. Expertise affects the user's cognition, including their ability to chunk data and their working memory [26], which in turn affects how users process visualizations [27]. Expertise may train users to search for more information in particular formats; consequently, such users may have the tendency to prefer those formats rather than the formats that cognitive fit theory might otherwise infer [25], [28]. Therefore, expertise has become an important user characteristic to consider in the design of adaptive visualizations [29]. Here, we consider the expertise of having a particular educational background: accounting or data science. Both of these backgrounds require significant quantitative data analysis skills, and people from both areas are increasingly being hired to perform data analysis on a variety of business datasets. While accountants work in Microsoft Excel and other tabular formats for much of their training, data scientists are frequently trained in a larger variety of media, including tables and many plotting tools.

3) Visualization and Insight Study Design: Classic studies of cognitive fit theory and visualizations typically were created with benchmark tasks where the insight types and insights themselves were pre-determined by the experimenter [e.g., [20]]. Such classic studies often reported better performance in terms of accuracy or reaction time when the problem was presented in a visual format that fit the specific task. For example, when the task is spatial and requires finding a trend, performance is better when the participant uses a graph rather than a table [30]. Studies of data scientists and how they analyze data and make insights have also been largely focused on using visualizations for particular tasks (e.g., [12], [31]–[34]). For example, Gestalt displayed multiple visualizations to help analysts relate their data, features, and results, and he demonstrated that these interactive tools helped them find significantly more errors than non-visual tools [31]. A study of dashboard visualizations to help data scientists understand their data distributions demonstrated that different visualizations were useful for finding both spatial insights and errors [12].

These benchmark tasks, also commonly used in user studies of visualizations [20], [25], allow straightforward quantitative analyses, but they additionally prescribe results and threaten ecological validity because realistic data validation scenarios are more open-ended exploration and rarely presented as predetermined tasks [8]. They are also frequently tested on a single visualization rather than sequences of visuals. In contrast, our study asked participants to work towards an open-ended goal to derive meaningful insights from their data exploration. By tracking their decision-making process, we aim to understand the potential differences in their choices about what insights they made and what visualizations they used.

III. Study Design

In order to understand the impact of expertise on the analysis of visualization sequences, we designed a study in which participants were asked to validate two datasets, each presented as a sequence of ten visualizations. We were interested in business data analysis tasks due to the breadth of expertise needed to identify patterns and outliers in these datasets. Participants in our study had either data science or accounting expertise. Both groups are trained in quantitative data analysis. However, accountants have more training on the rules needed to understand particular business data, while data scientists receive more training on identifying statistical trends in a variety of data types. Here, we describe our experimental design and procedure along with our participants' backgrounds.

A. Data Validation Task

We asked participants to take on the role of a Financial Data Analyst who had been asked to make data-specific insights and observations as they explored two 10-visualization sequences that we had created for two datasets in advance of the experiment and held constant across all participants (Figure 1 (Left), described later). We instructed them that the data was not perfect and that their job was to explore the data carefully and report integrity issues. Participants were explicitly instructed to make only data insights, not business insights. For example, we did not want the participants to comment on the organizational structure of the different departments in the business (e.g., the CEO must be in Department 1), but we did want them to comment on whether the structure led to a pattern in the financial data (e.g., one department had a higher average salary than others).

They were given an online interface in which to log insights about errors or inconsistencies that they found in the data as well as any patterns that they observed (Figure 1 (Right)). The logging interface included checkboxes to generally categorize insights (i.e., outliers, specific data points, patterns, trends, shapes of data, violations of business common sense, or other), space to write a specific insight (e.g., "Joe Smith spent \$550 in cash on restaurants"), checkboxes to indicate which visualization(s) was (were) used to make the insight, a 3-point scale to provide a confidence rating about the insight, and boxes to indicate whether they developed a hypothesis and searched for this information or whether it was found unexpectedly. The categories of insights were generated based on feedback from a pilot study.

Unlike prior studies on cognitive fit and extended cognitive fit that asked participants to find or confirm specific pieces of information in a visualization using benchmark questions [35], this study had an open-ended design. Participants in our study needed to formulate their own model of the data based on their domain expertise to find patterns and detect outliers. While the benchmark tests typically studied visualization comprehension and application [36], our insight-based study asked participants to analyze relationships and trends, which we believe is more reflective of analysis tasks. The visualization sequences also provided opportunities to synthesize data to form insights from multiple visualizations, which represents a deeper understanding of the data [18].

B. Financial Datasets

In order to be fair to both participant groups, we chose a domain that both data scientists and accountants should be able to analyze: corporate financial data. There are many types of financial datasets that we could have chosen for the participants to explore. We chose to focus on the tasks of auditing (1) a Human Resource Department's salary data and (2) a Finance Department's expense reports.

1) Salaries: We generated three months of salary and bonus data for a fictitious company, Williams and Pitt, LLC. Each of the 200 employees at the company had a department, a start date, an annual salary, monthly salary payments, and monthly bonuses. Participants were given the company rule that those employees paid under \$40,000 per year were not to be given bonuses. Nothing was said about whether all employees above that salary would necessarily receive bonuses.

Inspired by Nigrini et al. [37], we injected inconsistencies into the data. These inconsistencies required some domain

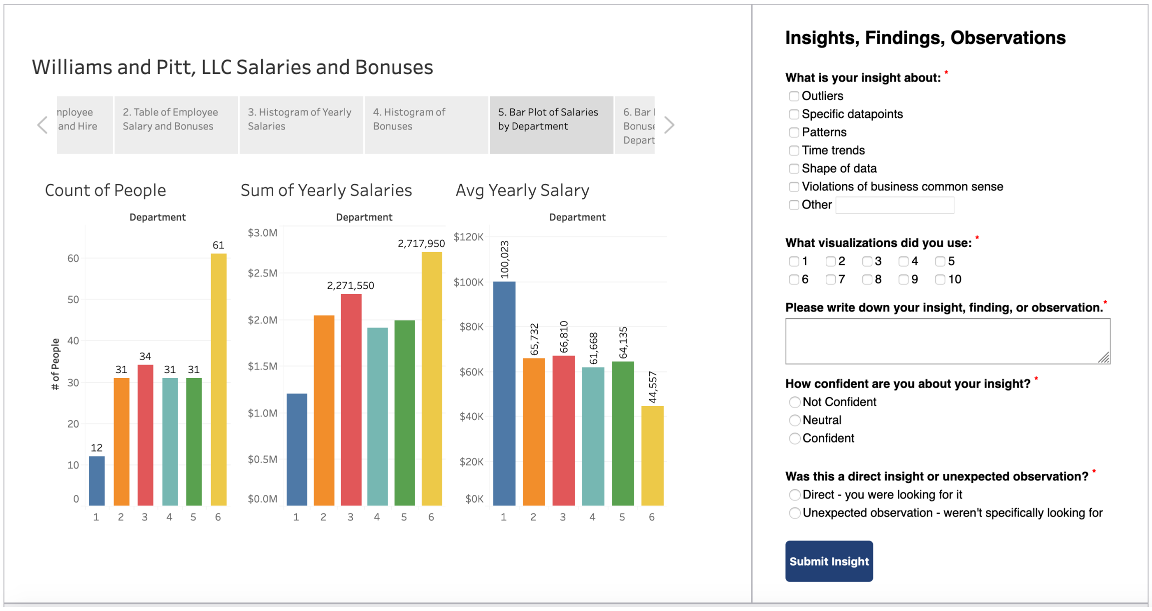


Fig. 1. (Left) The Tableau online interface with the 10-visualization sequence navigation at the top. (Right) The Insight logging interface was presented to the right of the Tableau interface so that participants could see a visualization as they logged.

or common sense knowledge to find. The inconsistencies included negative salaries and bonuses, salaries given before employment start dates, bonuses for employees making under \$40k, no bonus for people with higher salaries, and bonuses much higher or lower than other colleagues with the same salary. In addition to these inconsistencies, there were other patterns. Common patterns cited include one department having more employees than the others, one department having a higher average salary than the other departments, one department having much lower than average salaries, and three employees with much higher salaries than everyone else. In all, there were 19 patterns and inconsistencies that we expected participants to find. They could, of course, find more.

2) Expenses: We generated about 1,000 expense reports for another fictitious company with 70 employees called Burg and Burgh Associates. Each of the expenses was submitted by a particular employee in a particular department. The expenses were labeled with a category (air tickets, rental car, lunch, dinner, etc.), a payment type (cash, credit), and a submission date. Again, based on prior work by Nigrini et al. [37], we injected inconsistencies into this data, including negative expenses, high cost of a single item, high number of expenses by a single person, and high percentage of charges by cash. We also included other patterns. For example, two departments had a majority of the expenses and the rest had very few. Eighteen inconsistencies were injected into the expense dataset.

C. Visualization with Tableau

Because financial datasets (such as those used in this study) tend to be large, it is typical for both data scientists and accountants to generate graphs and charts to summarize the data and make it easier to spot inconsistencies and patterns in addition to looking at tabular views. We chose to visualize the data using Tableau because of its elegant graphs, use by recent graduates of both accounting and data science degrees (i.e., the interface did not favor one group of participants over the other), and the ability to host the visualizations online so that we could collect data remotely. An example visualization is shown within our data collection interface in Figure 1.

We generated 10 visualizations for each dataset, including summary tables (tables); histograms and bubble charts (one-dimensional (1D) visualizations); box plots, scatter plots, and bar charts (two-dimensional (2D) visualizations); and line graphs (time series). Each visualization told a story about one aspect of the data, such as expenses by department or the correlation between salary and bonus. Some of the visualizations included multiple side-by-side graphs of the same or similar data (e.g., the count of people in a department and the sum of the salaries in those departments). Visualizations were color-coordinated so that the same categories on each plot used the same colors. All plots were labeled appropriately and often included mouse rollovers with additional information. Some plots included interactivity that enabled participants to click on categories to reduce the data used in the plot. A complete list of the visualization types is in Table I.

It was hard to make the visualization sequences exactly equivalent because the datasets had different qualities. The 10 visualizations were ordered such that the first visualizations were a low-level view of the data (e.g., tabular views and 1D plots), the next visualizations focused on dependent variables (salaries or expenses) broken down by important category features (department, year, number of people, etc.), and the last visualizations focused on time or the order of events. Multiple plots about the same feature, like departments, were presented sequentially with their impact on different dependent variables

TABLE I
THE VISUALIZATION TYPES FOR EACH OF OUR TWO DATASETS.

	Salary Data (William & Pitt)	Expense Data (Burg & Burgh)
1	Table	1D Bubble Plot
2	Table	Table
3	1D Histogram	2D Bar Plot
4	1D Histogram	1D Histogram
5	2D Bar Plot	Table
6	2D Bar Plot	Table
7	2D Scatterplot	2D Box Plot
8	2D Box Plot	1D Histogram
9	2D Box Plot	Line Plot
10	Line Plot	Line Plot

(e.g., a visualization of cash vs. credit expenses by department occurred just before a visualization showing the categories of expenses by department). The visualizations were numbered and titled by the chart type and variable names, and they were presented in such a way as to imply that the sequence was important. However, the interface allowed participants to scroll through the titles and click on any visualization in any order. Figure 1 shows a scrollable grey bar of buttons above the bar charts. Each button was independently clickable, but they were presented in numeric order.

D. Experimental Procedure

To collect the workflows of both data scientists and accountants as they validated financial data, we designed an open-ended, think-aloud study in which participants spoke about their thought processes while they examined the data visualizations for the datasets described above and made insights about them.

After the study was approved by our Institutional Review Board, emails were sent to recruit data science and accounting participants at several major universities. When students replied to the email and confirmed that they had taken a sufficient amount of relevant coursework, they were asked to attend a Zoom session with a researcher for the study. After consenting to be a part of the experiment, the participant began the study using a website link provided by the experimenter, and they shared their screen before the researcher started recording both the audio and the screen of the Zoom session.

Participants completed a demographics survey asking about their background as either a data scientist or an accountant (if both, they were told to choose the area about which they are more knowledgeable). Then, participants read the instructions for the task and were given a practice task for a small dataset that used Tableau Online, the insight interface, and both simultaneously. They also practiced thinking aloud (i.e., speaking what they were thinking). We chose to use a think-aloud protocol so that we could hear what they were thinking while reading the visualizations and making insights.

After completing the practice task, participants were randomly assigned to complete one of the datasets first. They were given 30 minutes to explore, observe, and validate the data using the ten visualizations provided. The researcher encouraged the participants to think aloud if they were silent for one minute. They also encouraged the participant to log insights using the log form if they were speaking insights but not recording them. At the end of 30 minutes, the website automatically presented a survey for them to assess their own task performance and whether the task was similar to any that they had seen before. After completing the survey, they then repeated the task for 30 minutes with the second dataset and took another survey. At the end of the study (approximately 1.5 hours), participants provided their email addresses in order to be compensated $30 in Amazon gift cards for their time.

E. Measures

We captured a variety of measures to evaluate differences between data scientists and accountants for the financial data analysis tasks. From the screen recordings, we captured:

- **Visit Count**: The count of the number of distinct times that a participant spent on each visualization;
- **Skipping Around Count**: The count of distinct times participants spent navigating between several visualizations but not analyzing any particular visualization (less than 15 seconds Visit Time per visualization);
- **Visualization Order**: The sequence of visualizations viewed.

We also used the insight logging interface to assess:

- **Insight Type**: General category of insight (outliers, specific data points, patterns, time trends, shape of data, violations of business common sense, other);
- **Visualization(s) Used**: The visualization number(s) and visualization type(s) used for the insight (table, 1D, 2D, line plot);
- **Visualization Count per Insight**: The number of visualizations used for the insight (a proxy for insight depth);
- **Insight Count**: The number of insights generated.

Finally, we assessed the relationship between visualizations in sequence. We computed the following measures for analysis:

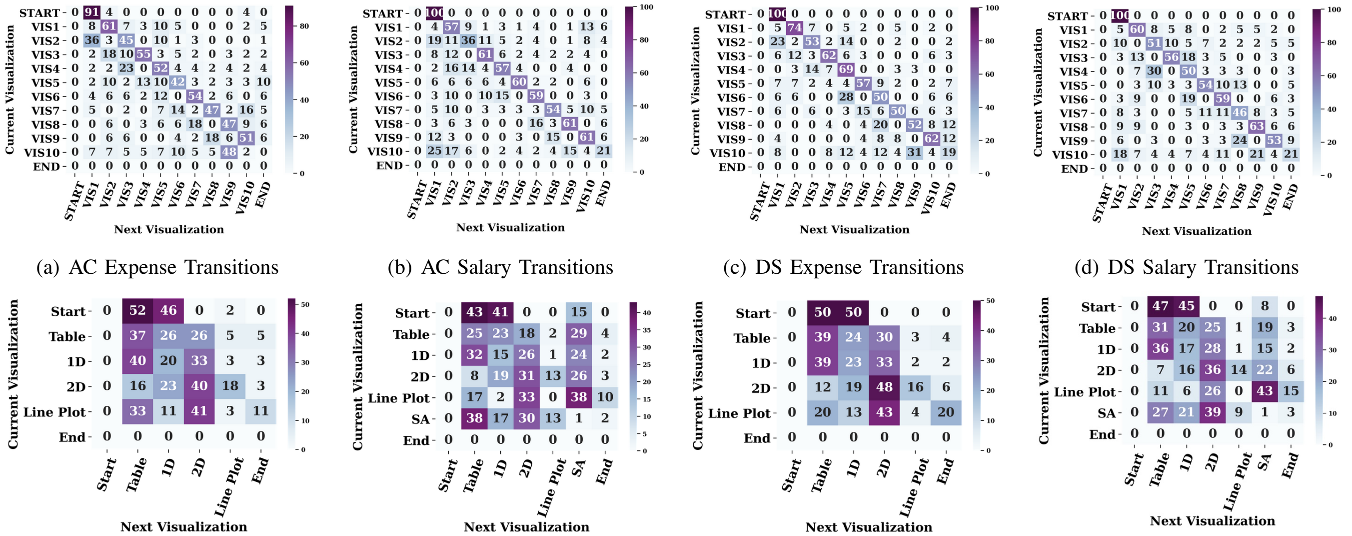
- **Visualization Number Transition**: The percentage of transitions from a given visualization number to each other visualization (Skipping Around was not an option);
- **Visualization Type Transition**: The percentage of transitions from a given visualization type (table, 1D, 2D, line plot) to all others, computed with and without Skipping Around as a transition option.

We compared these dependent measures across expertise (data scientist versus accountant) to understand the potential impact of domain expertise on data exploration and insights.

F. Participants

Twenty-three participants with an accounting (AC) background and 20 participants with a data science (DS) background completed the 1.5-hour study. They were recruited from Master's degree programs at several prominent business schools in the United States.

In a demographics questionnaire, we asked AC participants about their experience with Excel and with auditing tools on a scale of 1 to 5. They rated their Excel experience (μ = 3.6; SD



(a) AC Expense Transitions (b) AC Salary Transitions (c) DS Expense Transitions (d) DS Salary Transitions

(e) AC w/o SA Transitions (f) AC w/ SA Transitions (g) DS w/o SA Transitions (h) DS w/ SA Transitions

Fig. 2. Each figure represents the percentage (out of 100) of time that participants started at each row visualization and transitioned to each of the column visualizations. Rows that do not sum to 100% are due to rounding errors. (Top) When modeling the transitions between visualization numbers in the sequence, both AC and DS participants transitioned in numerical order either forwards or backwards by one visualization. (Bottom) When modeling the transitions between visualization types, participants spent on average 20% of their time Skipping Around (SA). When we take those SA times into account, it is clearer that AC participants chose to navigate back to tables after skipping around, and DS participants frequently chose to navigate back to 2D plots. Similarly, AC participants skipped around after tables rather than transitioning to other plots like those with DS expertise did.

= .77) slightly higher than their experience with auditing tools (μ = 3.20; SD = 1.36), but the difference was not statistically significant ($t(19) = 1.79$, $p = .09$). We asked DS participants about their experience with Python and other tools for data science. They rated their Python experience as fairly high (μ = 4.05; SD = .69) and their experience of other tools slightly higher (μ = 4.25; SD = .64), but the difference was not not statistically significant ($t(19) = -1.29$, $p = .21$). Other tools that they identified included R, Matlab, Stata, SPSS, SAS, Mathematica, Rapid Miner, PowerBI, and Excel.

We asked both groups about their knowledge of Tableau. The DS group rated their experience creating Tableau visualizations as significantly higher (μ = 3.30; SD = .92) than the AC group did (μ = 2.48; SD = 1.04), ($t(41) = 2.72$, $p = .01$). The participants did not have to create visualizations using Tableau in our study; they only viewed them. None of the participants expressed difficulty reading, using or comprehending the Tableau visualizations. Therefore, we do not believe this difference meaningfully affected our findings.

IV. Results

In order to assess the differences between the Data Scientists (DS) and the Accountants (AC) on the data analysis tasks, we performed t-tests and ReML mixed effects tests for our dependent measures. For most of the results, we combined the data from both tasks. When data from the tasks was analyzed separately, we note it below.

A. Visualization Sequences

We first analyzed the order in which the visualizations were viewed. We computed the percentage of time that our two groups transitioned from one visualization number to another (e.g., V1→V2) or from one visualization type to another (e.g., 2D plot → table).

1) Visualization Number Transitions: The tables representing the percentage of transitions taken from each "current visualization" row to "next visualization" column are presented in Figures 2 a-d. Rows that do not sum to 100% are due to rounding errors. "Start" represents the starting state of the session and "End" represents the decision to stop analyzing the dataset or when 30 minutes had passed.

There was an apparent effect of visualization number on the order in which participants visited each visualization, shown as the bright diagonal line of high percentages. Nearly all participants visited the visualizations sequentially at first (0% of the DS group and less than 9% of the AC group chose to visit them out of order). Altogether, these results indicate that the participants visited the visualizations in sequence 59.3% of the time. The second most popular transition was a backward transition (e.g., from visualization Viz2 to Viz1 or from Viz5 to Viz4).

The transitions that happened after reaching Viz10 were less straightforward. The accountants transitioned back to Viz9 48% of the time on the Expense data, but only 15% of the time on the Salary data. Similarly, the data scientists transitioned to Viz9 31% of the time on the Expense data and 21% of the time on the Salary data. This indicates that there is some other effect besides visualization number order that may be causing them to make different decisions about where to transition after finishing the sequence.

2) Visualization Type Transitions: We hypothesized that some of the remaining variation in visit order may be due to participants gravitating towards particular visualization types. Additionally, the visualization number analysis did not account for the times when the participants would stop analyzing one visualization and skim through other visualizations at particular periods of time. Due to the sparse data, we combined visit data across the two datasets to focus our analyses on visualization type. We included an extra state called "Skipping Around" (SA) that represented a significant period of time that participants were clicking on the visualizations but not interpreting them. This state represented 20% of all transitions.

We present the results with and without SA for comparison (Figures 2e-h). The transition tables with SA provide a clearer picture. AC participants skipped around after most visualization types more frequently than DS participants did. When they were done skipping around, AC participants most frequently transitioned back to tables (38%), while DS participants transitioned back to 2D plots (39%). We performed χ^2 analyses to compare the frequencies of transitions between each current visualization type and the next types, and we evaluated statistical significance at the $p < 0.05/6$ level (Bonferoni-corrected for performing 6 tests). We found no statistically significant differences in the transition frequencies overall, and $p > 0.05$ for all tests performed.

B. Insights Generated

In total, our 43 participants reported 845 insights that included 41 unique insights about the data. All 41 insights (including 4 that we had not previously identified) were correct given the data visualizations. We analyzed the insight counts to understand how they were impacted by our dependent variables – expertise, visualization number, visualization type, and insight type.

1) Insight Count by Visualization Number and Expertise: We analyzed the effects of Visualization Number and Expertise on the number of insights made by performing an ReML analysis of mixed effects, with Participant nested in Expertise as a random effect. We found a significant effect of Visualization Number (F(9,369) = 6.67, $p < 0.001$) on the insight count. There was no significant effect of Expertise alone (F(1,41)=1.98, $p = 0.16$), but there was an interaction effect between Visualization Number and Expertise (F(9,369) = 1.95, $p = 0.05$). Figure 3 shows the effect of Expertise on the insight count by Visualization Number. A Tukey HSD post-hoc ($\alpha = 0.05$ for all Tukey tests) analysis showed that AC participants made more insights on Viz1 (4.09) than they did on Viz3, 4, 5, 6, 7, 8, 9, and 10 and more than DS did on Viz4, 6, and 8 (all under 3 insights). AC also made more insights on Viz2 compared to Viz8. In contrast, DS made more insights on Viz7 compared to AC's insights on Viz4 and Viz8, and they made more insights on Viz1 and Viz5 compared to AC on Viz8. No other differences were detected. This indicates that AC participants spent more time making insights on the tables at the beginning of the visualization sequences and less time at the end, while the DS participants' insights were more spread out through the sequences.

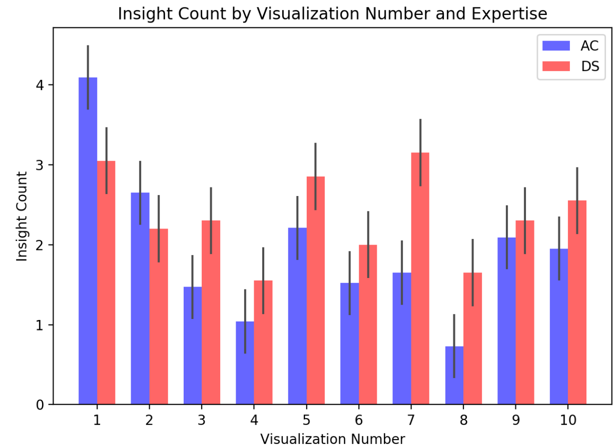


Fig. 3. The participants made insights throughout the visualization sequence. (Error bars represent standard deviations.) AC participants made the most insights on Visualization 1 (Table and 1D plot), while DS participants made the most insights on Visualization 7 (2D plots).

2) Insight Count by Visualization Type, Insight Type, and Expertise: Additionally, we investigated the impact of Expertise, Insight Type, and Visualization Type on Insight Count. We performed an ReML analysis with Expertise, Visualization Type (table, 1D, 2D, line plot), Insight Type (participants indicated whether their insights were about outliers, specific data points, patterns, data distributions, business common sense, and other), and all of their interactions as fixed effects. We modeled Participant nested in Expertise as a random effect. If a single insight arose from multiple visualizations, it was counted multiple times - once per Visualization Type.

There were statistically significant main effects of Visualization Type (F(3,1107) = 17.71, $p < 0.001$) and Insight Type (F(3,1107) = 45.62, $p < 0.001$) on the Insight Count. The main effect of Expertise was not significant (F(1,41) = 3.57, $p = 0.066$). Additionally, there were statistically significant interaction effects between Insight Type and Expertise (F(6,1107) = 4.68, $p < 0.001$), and Insight Type and Visualization Type (F(18,1107) = 6.33, $p < 0.001$). The interaction effect between Expertise and Visualization Type was marginally significant (F(3,1107) = 2.49, $p = 0.059$). There was no interaction effect across all three variables.

Visualization Type Main Effect. A post-hoc Tukey HSD test on Visualization Type showed that participants generated more insights about the tables (μ=1.43, SD = 0.10) and 2D plots (1.23) compared to 1D plots (0.86) and line plots (0.66). No significant differences were observed between tables and 2D plots nor between 1D plots and line plots.

Insight Type Main Effect. A post-hoc Tukey HSD test on Insight Type showed that participants generated significantly more insights about outliers (μ=2.11, SD = 0.12) and data patterns (1.95) compared to data points (1.20), data distributions (0.73), trends (0.66), business sense (0.42), and "other" topics

(0.22). There were also significantly more insights about data points than the four remaining categories, and more insights about data distributions than "other".

Insight Type and Expertise Interaction Effect. A post-hoc Tukey HSD test was performed. The DS group created more insights about outliers (μ =2.43, s.d. 0.17) and patterns (2.44) compared to all other combinations of expertise and insight types except AC for patterns (1.79). AC created more insights about patterns (1.79) and outliers (1.47) than AC and DS for trends (0.73 and 0.60 respectively), AC and DS insights about business sense (0.41 and 0.43 respectively), AC and DS insights about other topics (0.32 and 0.11), and AC insights about distributions (0.42). AC created more insights about data points (1.26) compared with AC and DS insights on business sense, AC and DS insights on other topics, and AC insights on distributions. DS insights on data points (1.14) and DS insights on distributions (1.05) were significantly different from DS insights on other topics (0.41). Overall, both groups generated more insights about patterns and outliers than other insight types.

Insight Type and Visualization Type Interaction Effect. A post-hoc Tukey HSD test was performed again. Participants most frequently logged insights about outliers on 2D plots (3.0 per person) and tables (2.92) as well as insights about data patterns on 2D plots (2.58) and tables (2.49). These four combinations occurred at a statistically significantly higher rate compared to all other combinations of visualizations and insight types except for insights logged about specific data points on tables (2.28) and specific patterns on 1D plots (2.1). Participants made fewer than 1.3 insights for each of the other combinations of visualization type and insight type. Additionally, participants logged insights about patterns on line plots (1.29) significantly more than "other" on line plots (0.09).

V. Discussion

We performed an experiment in which we asked both data scientists (DS) and accountants (AC) to work on an open-ended analysis task for two business datasets and log their insights. Our results show that there is a meaningful effect of Visualization Sequence and Visualization Type on the order in which people analyze their data. In particular, DS participants chose to view 2D plots more frequently after skipping around while AC participants chose tables. Insight Count was also significantly affected by both Visualization Type and Insight Type as main effects, which follows from the cognitive fit theory prediction that people use different visualizations to make different insights. While Expertise did not have a main effect on the viewing order nor on the number of insights generated, it did interact with Insight Type. Participants with different expertise made different types of insights while viewing the same data. Overall, these results suggest that sequences of visualizations should be tailored to the expertise of the viewers and to the types of information that viewers might seek (e.g., outlier detection versus pattern summarization).

1) Visualization Sequence Transitions and Prediction: We analyzed the viewing order using the transitions from one visualization to another. We were surprised that the participants so frequently visited the visualizations in order on both datasets. This indicates that the sequencing (i.e., the order and numbering of the visualizations) likely primed all of the participants in both conditions to transition in order. We would also like to note that, while not statistically significant, the differences in the frequencies of different transitions between the two expertise groups are meaningful if we use the transitions as an AI model of visualization order. The AC model would transition from Skipping Around to Tables, while the DS model would transition from Skipping Around to 2D plots. Similarly, the AC model would transition from Tables to Skipping Around, and the DS model would transition from Tables to another Table. An AI tool that predicted sequences to show would present different visualizations to these two groups.

It was not possible to differentiate the effects of the visualization types from the effects of the visualization order due to our design decisions about the visualization sequences. In particular, while we did aim to be as consistent as possible, we were not able to ensure that the visualization type for each visualization number was the same across the two datasets because the datasets warranted different visualizations. We also chose not to randomize the order of the sequences because the order would affect the users' understanding of the datasets. Future work is needed to tease apart the roles of visualization sequence position and visualization type to determine their effects on the order in which people visit the visualizations.

2) Expertise and Visualization Type: We were able to analyze the transitions between different visualization types. The transitions indicate that AC and DS expertise may predispose the participants to choose visualization types with which they are more comfortable. Accountants are trained to look at data in tabular form (e.g., Microsoft Excel), while Data Scientists are frequently trained to view tables and also create 2D plots to find correlations. Considering these differences in education, the accountants' preferences for viewing tables in our study and their use of tables to create significantly more of their insights makes sense. Data scientists used tables and 2D plots more evenly. Another qualitative difference that we found between the expert groups is that the accountants were skimming the visualizations for more business information that could help them understand the tables while data scientists were more willing to read the plots at face value. In order to make fair assessments of our data, accountants asked much more frequently for additional information about the businesses (which we did not provide). More work is needed to understand how professionals with other areas of expertise compare to accountants and data scientists.

Additionally, we found that 'skipping around' states were decision points where more differences between our expertise groups occurred. More work is needed to understand users' goals when skipping around, whether the skipping around is a result of having a limited number of visualizations, and whether the process of skipping helps users make insights in some way. Additionally, more work is needed to understand whether and how automated visualization sequence generators could use the skipping around state as an indication of how to tailor sequences or present new visualizations.

3) Expertise and Insight Count: The differences observed in insight type and visualization type did not affect the overall insight counts across expertise groups, nor did they affect any other measure of effectiveness at the task. Accountants and data scientists may approach the problem of quantitative analysis differently and use different information to make insights, but it does not mean that either groups' insights are of lower quality. This is an important consideration for issues like team diversity, as people with different expertise may be able to assess the quality of data differently and more completely than if only one type of expertise is employed. Future work should test other types of expertise (e.g., scientific backgrounds) and other types of open-ended tasks as well.

4) Cognitive Fit Theory: Extended cognitive fit theory argues that both domain expertise and analysis type (symbolic versus spatial) affect which visualizations people find most effective for analyses. To the best of our knowledge, nearly all prior work on extended cognitive fit focused on prescriptive tasks to find specific information in specific plots. The open-endedness of our experiment is unique and was paramount to our ability to study the choices that our participants made about which visualizations they made insights about.

Our findings support the cognitive fit theory that people use different kinds of plots to make different kinds of insights. Interestingly, our results showed that outliers and patterns were discovered at a much higher rate than the other types of insights across all participants. This could be a result of the skew of our inconsistencies towards these findings, because they stood out the most in our visualizations (especially the 2D plots and tables), or because people tend to find those inconsistencies easiest or fastest. Because we tailored our visualizations to explain the data rather than randomly generating plots that may or may not be insightful, more work could be done to understand the relationship between insight types and visualization types and whether cognitive fit theory is upheld in these open-ended explorations.

We also found that expertise had a statistically significant interaction effect with insight type, which indicates that the two groups generated different insight types at different rates. While we did not find a main effect of expertise, the interaction effects do indicate the importance of expertise in visualization and data analysis tasks. More work at the intersection of extended cognitive fit theory and visualization is needed to understand how expertise impacts the use of different visualizations to make insights.

VI. Conclusion

Data analysis is a challenging process in which analysts must determine what data to visualize and then make meaningful insights about their visualizations. In this work, we presented an open-ended study in which we asked two groups of participants - accountants and data scientists - to make insights about two different datasets presented as visualization sequences. We analyzed the order in which they viewed the visualizations as well as the insights that they generated.

Our results showed that most participants viewed the visualization sequence in order first, indicating a strong main effect of visualization number. However, there were times when they did not view the sequence in order. At these decision points, participants often skipped around before settling back on a particular visualization to analyze. Accountants more frequently chose tables, and data scientists chose 2D plots, likely based on domain familiarity. Additionally, we analyzed the impact of insight type, visualization type, and expertise on the number of insights made. Expertise had an interaction effect with insight type, indicating that the different groups generated different types of insights. However, there were no overall differences in the number of insights generated between the two groups.

Based on these results, we conclude that data scientists and accountants could be used together on teams to generate different but equally important insights about the same datasets. As we move to automate more of the process of visualization to help streamline analysts' work, our study findings indicate that expertise may be another feature besides informativeness and personal preferences with which to assess the selection visualizations and visualization sequences.

Acknowledgments

We would like to thank the Institute for Fraud Prevention (now a division of the ACFE Foundation) for funding this research, Juliette Wong and Elizabeth Finneran for their tireless efforts collecting and coding the study data, and Moshe Mash for his help visualizing the data analysis.

References

[1] B. Lee, N. H. Riche, P. Isenberg, and S. Carpendale, "More Than Telling a Story: Transforming Data into Visually Shared Stories," *IEEE Computer Graphics and Applications*, vol. 35, no. 5, Sep. 2015.

[2] E. Segel and J. Heer, "Narrative Visualization: Telling Stories with Data," *IEEE Transactions on Visualization and Computer Graphics*, vol. 16, no. 6, pp. 1139–1148, Nov. 2010.

[3] J. S. Yi, Y. a. Kang, J. Stasko, and J. Jacko, "Toward a Deeper Understanding of the Role of Interaction in Information Visualization," *IEEE Transactions on Visualization and Computer Graphics*, vol. 13, no. 6, pp. 1224–1231, Nov. 2007.

[4] P. Saraiya, C. North, and K. Duca, "An insight-based methodology for evaluating bioinformatics visualizations," *IEEE transactions on visualization and computer graphics*, vol. 11, no. 4, pp. 443–456, 2005.

[5] I. Vessey, *The theory of cognitive fit*, 2006, vol. 5, pp. 141–183.

[6] J. Hullman, R. Kosara, and H. Lam, "Finding a Clear Path: Structuring Strategies for Visualization Sequences," *Computer Graphics Forum*, vol. 36, no. 3, pp. 365–375, 2017.

[7] Y. Kim, K. Wongsuphasawat, J. Hullman, and J. Heer, "GraphScape: A Model for Automated Reasoning about Visualization Similarity and Sequencing," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, May 2017, pp. 2628–2638.

[8] D. J.-L. Lee, H. Dev, H. Hu, H. Elmeleegy, and A. Parameswaran, "Avoiding drill-down fallacies with *VisPilot*: assisted exploration of data subsets," in *Proceedings of the 24th International Conference on Intelligent User Interfaces*, Mar. 2019, pp. 186–196.

[9] Y. R. Cao, J. Y. Pan, and W.-C. Lin, "User-oriented generation of contextual visualization sequences," in *CHI EA 2020 - Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, Apr. 2020.

[10] A. Ng, "AI Doesn't Have to Be Too Complicated or Expensive for Your Business," *Harvard Business Review*, Jul. 2021, section: Data. [Online]. Available: https://hbr.org/2021/07/ai-doesnt-have-to-be-too-complicated-or-expensive-for-your-business

[11] N. Sambasivan, S. Kapania, H. Highfill, D. Akrong, P. Paritosh, and L. M. Aroyo, "Everyone wants to do the model work, not the data work: Data Cascades in High-Stakes AI," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, New York, NY, USA, May 2021, no. 39, pp. 1–15.

[12] F. Hohman, K. Wongsuphasawat, M. B. Kery, and K. Patel, "Understanding and Visualizing Data Iteration in Machine Learning," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, Apr. 2020, pp. 1–13.

[13] R. Chang, A. Lee, M. Ghoniem, R. Kosara, W. Ribarsky, J. Yang, E. Suma, C. Ziemkiewicz, D. Kern, and A. Sudjianto, "Scalable and Interactive Visual Analysis of Financial Wire Transactions for Fraud Detection," *Information Visualization*, vol. 7, no. 1, pp. 63–76, Mar. 2008.

[14] W. N. Dilla and R. L. Raschke, "Data visualization for fraud detection: Practice implications and a call for future research," *International Journal of Accounting Information Systems*, vol. 16, pp. 1–22, Mar. 2015.

[15] R. A. Leite, T. Gschwandtner, S. Miksch, S. Kriglstein, M. Pohl, E. Gstrein, and J. Kuntner, "EVA: Visual Analytics to Identify Fraudulent Events," *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 1, pp. 330–339, Jan. 2018.

[16] J. M. Teets, D. P. Tegarden, and R. S. Russell, "Using Cognitive Fit Theory to Evaluate the Effectiveness of Information Visualizations: An Example Using Quality Assurance Data," *IEEE Transactions on Visualization and Computer Graphics*, vol. 16, no. 5, pp. 841–853, Sep. 2010.

[17] C. North, P. Saraiya, and K. Duca, "A comparison of benchmark task and insight evaluation methods for information visualization," *Information Visualization*, vol. 10, no. 3, pp. 162–181, Jul. 2011.

[18] C. North, "Toward measuring visualization insight," *IEEE Computer Graphics and Applications*, vol. 26, 3, pp. 6–9.

[19] P. Saraiya, C. North, and K. Duca, "An insight-based methodology for evaluating bioinformatics visualizations," *IEEE Transactions on Visualization and Computer Graphics*, vol. 11, no. 4, pp. 443–456, Jul. 2005.

[20] C. Nobre, D. Wootton, Z. Cutler, L. Harrison, H. Pfister, and A. Lex, "reVISit: Looking Under the Hood of Interactive Visualization Studies," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, May 2021, pp. 1–13.

[21] L. R. Beach and T. R. Mitchell, "A contingency model for the selection of decision strategies," *The Academy of Management Review*, vol. 3, no. 3, pp. 439–449, 1978.

[22] J. W. Payne, "Contingent decision behavior," *Psychological Bulletin*, vol. 92, no. 2, pp. 382–402, 1982, place: US Publisher: American Psychological Association.

[23] Z. Liu and J. Stasko, "Mental Models, Visual Reasoning and Interaction in Information Visualization: A Top-down Perspective," *IEEE Transactions on Visualization and Computer Graphics*, vol. 16, no. 6, pp. 999–1008, Nov. 2010.

[24] M. Scaife and Y. Rogers, "External cognition: how do graphical representations work?" *International Journal of Human-Computer Studies*, vol. 45, no. 2, pp. 185–213, Aug. 1996.

[25] C. Nobre, D. Wootton, L. Harrison, and A. Lex, "Evaluating Multivariate Network Visualization Techniques Using a Validated Design and Crowdsourcing Approach," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, Apr. 2020, pp. 1–12.

[26] K. Ericsson, *The Cambridge Handbook of Expertise and Expert Performance*. Cambridge University Press, 2018.

[27] C. Conati, G. Carenini, D. Toker, and S. Lallé, "Towards User-Adaptive Information Visualization," in *Twenty-Ninth AAAI Conference on Artificial Intelligence*, Mar. 2015.

[28] R. E. Patterson, L. M. Blaha, G. G. Grinstein, K. K. Liggett, D. E. Kaveney, K. C. Sheldon, P. R. Havig, and J. A. Moore, "A human cognition framework for information visualization," *Computers & Graphics*, vol. 42, pp. 42–58, Aug. 2014.

[29] C. Amyrotos, "Adaptive Visualizations for Enhanced Data Understanding and Interpretation," in *Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization*, Jun. 2021, pp. 291–297.

[30] I. Vessey and D. Galletta, "Cognitive Fit: An Empirical Study of Information Acquisition," *Information Systems Research*, vol. 2, no. 1, pp. 63–84, 1991.

[31] K. Patel, N. Bancroft, S. M. Drucker, J. Fogarty, A. J. Ko, and J. Landay, "Gestalt: Integrated support for implementation and analysis in machine learning," in *Proceedings of the 23nd Annual ACM Symposium on User Interface Software and Technology*, ser. UIST '10, 2010, p. 37–46.

[32] D. Ren, S. Amershi, B. Lee, J. Suh, and J. D. Williams, "Squares: Supporting interactive performance analysis for multiclass classifiers," *IEEE transactions on visualization and computer graphics*, vol. 23, no. 1, pp. 61–70, 2016.

[33] M. Kahng, P. Y. Andrews, A. Kalro, and D. H. Chau, "A cti v is: Visual exploration of industry-scale deep neural network models," *IEEE transactions on visualization and computer graphics*, vol. 24, no. 1, pp. 88–97, 2017.

[34] Y. Ming, H. Qu, and E. Bertini, "Rulematrix: Visualizing and understanding classifiers with rules," *IEEE transactions on visualization and computer graphics*, vol. 25, no. 1, pp. 342–352, 2018.

[35] J. M. Teets, D. P. Tegarden, and R. S. Russell, "Using cognitive fit theory to evaluate the effectiveness of information visualizations: An example using quality assurance data," *IEEE transactions on visualization and computer graphics*, vol. 16, no. 5, pp. 841–853, 2010.

[36] A. Burns, C. Xiong, S. Franconeri, A. Cairo, and N. Mahyar, "How to evaluate data visualizations across different levels of understanding," in *2020 IEEE Workshop on Evaluation and Beyond - Methodological Approaches to Visualization (BELIV)*, 2020, pp. 19–28.

[37] M. J. Nigrini, *Forensic analytics: Methods and techniques for forensic accounting investigations*. John Wiley & Sons, 2020.