How Personality and Visual Channels Affect Insight Generation

Tomás Alves,* Carlota Dias,† Daniel Gonçalves,‡ and Sandra Gama§
HUMAN Lab, Graphics and Interaction, INESC-ID, and Instituto Superior Técnico, University of Lisbon, Lisbon, Portugal

Joana Henriques-Calado¶
CICPSI, Faculdade de Psicologia, Universidade de Lisboa, Lisboa, Portugal

Abstract

Gaining insight is considered one of the relevant purposes of visual data exploration, yet studies that categorize insights are rare. This paper reports on a study to understand if the categorization model used to describe insights and personality factors affect insight-based evaluations' findings. Participants completed a set of tasks with three hierarchical visualizations and then reported what insights they could gather from them. Results show that the insight categorization taxonomies produce different descriptions of insights based on the same corpus of responses. In addition, our findings suggest that the openness to experience trait positively influences the number of reported insights. Both these factors may create obstacles to the design of insight-based evaluations and, consequently, should be controlled in the experimental design. We discuss the study implications, lessons learned, and future work opportunities.

Index Terms: Human-centered computing—Visualization—Visualization theory, concepts and paradigms; Human-centered computing—Visualization—Empirical studies in visualization

1 Introduction

Gaining insights has been considered one of the major purposes of information visualization (InfoVis) [10]. In particular, several researchers advocate diversifying evaluation measures beyond traditional speed and accuracy measures (see BELIV conference). However, there is little empirical data to provide robust guidelines for practitioners focused on the process of gaining insight into data. Firstly, prior work has contributed with specific insight types such as correlations [50], outliers [17], or peaks [41]. Nevertheless, few works have converged on a taxonomy to categorize insights. To our knowledge, only Chen et al. [12] and Moere et al. [35] focused on creating a categorization to be broadly applied. As the research field continues to grow and new taxonomies may arise, it is valuable to consider if different insights models produce distinct descriptions of insights based on the same corpus of responses. In case there is indeed a difference, researchers should focus on validating the existing frameworks and, at the same time, assess if this modeling discrepancy hinders the effects caused by visualization design.

Inspired by these questions, we conducted an experiment to understand whether visualizations with the same data but distinct visual encodings affect the insight type distribution that participants gain between and within insight categorization models. Besides the effect of the graphical disposition, we are also interested in observing whether individual differences play a role in how people gain insights. Recent research studied how individual differences affect search performance across hierarchical [56], time series [46], and item comparison [9] visualization designs, visualization use [56], and behavioural patterns [40]. As such, typical measures to understand the effect of individual characteristics are speed, accuracy, or subjective feedback [32]. Considering such a large multitude of personality manifestations in visualization, we question whether personality factors affect the type of insights that participants report. Further, researchers need to understand whether personality plays a significant role in insight-based evaluations or if the design does not need to control it in the experiment.

Only Green and Fisher [21] studied how personality constructs affect insight generation in a visualization setting. In particular, the authors leveraged the neuroticism and extraversion traits from the Five-Factor Model (FFM) [15], and the Locus of Control (LoC) [43]. Therefore, there is an untapped set of individual characteristics that may provide further knowledge regarding the effect of personality on insights attainment. This research gap led us to include and control for the effect of personality in insight reporting. In particular, we try to replicate and extend the findings of Green and Fisher [21]. Regarding insights generation, the term *insight* carries several definitions in the visualization community depending on the context it is studied [10]. In our study, we follow the insight definition of Green and Fisher [21] and study hierarchical visualizations. Insight refers to "anything unexpected or novel learned by the participants while completing the tasks". Further, hierarchical visualizations are one of the most common and relevant information structures in computing [48] and, more specifically, evaluating personality effects on visualization settings (e.g., [21, 54, 56]). Participants performed task sets in three distinct hierarchical layouts: sunburst, treemap, and Sankey diagram. After performing the tasks for a specific layout, we asked each subject to tell us which insights they could gain from that visualization. These insights are then analyzed using three taxonomies: Chen et al.'s [12], Moere et al.'s [35], and Insight Valence.

Our key contributions are as follows: First, we identify how the categorization model affects the description of insights based on the same corpus of responses. Second, we show that openness to experience affects insight reporting, hindering the credibility of the insight-based evaluation without controlling for this factor. Finally, the dataset from our experiment is available for further insight- and personality-related research. These findings provide practical implications for obstacles that may arise in the design of insight-based evaluations.

2 Related Work

This section presents a literature review regarding insight generation and the effect of personality in visualization settings.

2.1 Insight-Based Evaluation

Traditional evaluations of visualization systems often rely on task-based metrics such as task response time or accuracy, hindering the assessment of exploratory strategies [34, 37]. As a means to counter the limitations of these approaches, research has leveraged insight extraction to measure the degree to which visualizations amplify analytical reasoning [36, 42]. Although there has been some work ranging from classifications [10, 12, 35] to insight-acquiring processes [53], the community has yet to converge on a formal definition of *insight*. For instance, while Saraiya et al. [45] define insight as "an

*e-mail: tomas.alves@tecnico.ulisboa.pt
†e-mail: carlota.lopes.dias@tecnico.ulisboa.pt
‡e-mail: daniel.goncalves@inesc-id.pt
§e-mail: sandra.gama@tecnico.ulisboa.pt
¶e-mail: joanahenriquescalado@gmail.com

individual observation about the data by the participant, a unit of discovery", North [36] characterizes it as "a non-trivial discovery about the data or, as a complex, deep, qualitative, unexpected, and relevant assertion". Researchers can study insights based on quantity [21] as well as quality [23], e.g., through understanding levels [8]. Insights can be evaluated through viewer takeaways [8, 21, 35] or annotations on the visualization [1, 13, 33]. Further, Yi et al. [53] focused on *how* people gain insights rather than *what* insights are, providing four distinctive processes of gaining insight: *provide overview*, *adjust*, *detect pattern* and *match mental model*.

We believe that designers can develop information visualization systems to foster insights [2]. In particular, both North [36] and Yi et al. [53] highlight the importance of considering design aspects in insight acquisition, as the measuring of insights may "enable the direct comparison of visualization design alternatives" [36]. Moreover, several other studies showcase the potential of using combinations of visual tasks, visualization types, and comparison arrangement designs to investigate how people arrive at their insights (e.g., [25, 39]). Alas, to our knowledge, few studies (e.g. [13, 21, 35]) benchmark visualization alternatives against each other to study how the visual channels model insight acquisition. Our work builds on the mentioned research by studying how the visual representation of hierarchical data can affect insight generation. Moreover, we complement our analysis by including personality constructs to improve our understanding of the user profile.

2.2 Personality Factors

Several studies have studied how personality predicts goal-setting behaviors, how a person rates importance and meaningfulness, and how individuals interpret information [24]. In particular, the two most used personality models are the locus of control (LoC) [31] and the Five-Factor Model (FFM) [15, 19, 44]. Among the most common metrics to evaluate the effect of personality in visualization, there is a body of research showing how personality dimensions affect performance across hierarchical [21, 55, 56], time series [46], and item comparison [9] visualization designs. Another relevant approach leverages visualization use [54, 56] and behavioral patterns [40]. Both eye-tracking [29, 30] and mouse data [7, 40] have predicted how LoC, extraversion, and neuroticism lead users to interact differently with data representations to facilitate information processing. Other studies evaluate user experience through subjective feedback such as design preferences [4, 56]. Although the mentioned research collectively shows that personality factors provide an opportunity to expand the user profile further, we can observe that the research field leans heavily towards studying performance metrics of search tasks in hierarchical layouts (see Liu et al. [32]).

We believe that personality traits may hold promising results regarding insight generation. Green and Fisher [21] already studied the impact of individual differences on the insights that users gain from information visualization. In particular, Green and Fisher [21] found that external LoC, introversion, and emotional stability lead subjects to report more insights with their hierarchical-based interactive data visualizations of genomic information. However, more personality traits can offer a deeper understanding of this relationship. For instance, we expect that openness to experience will directly affect the number of generated insights since this trait influences whether people tend to devise novel ideas [15]. Further, Ziemkiewicz et al. [55] found that high openness to experience led individuals to be faster while solving problems related to hierarchical visualizations with conflicting visual and verbal metaphors. The agreeableness trait plays a significant role in the trust process through the general tendency to be trusting and cooperative with others [15]. If the different graphical dispositions produce variations of visualization trustworthiness, it may lead to individuals trusting less in a specific visualization and reporting fewer insights, for instance. Finally, conscientiousness predicts how one organizes and plans ahead [22]. Organizational research reports the effect of conscientiousness on insight orientation [20]. However, researchers only evaluated the role of neuroticism and extraversion in visualization-supported insight generation. Our work tries to replicate the findings of previous work [21] and extends them with other personality traits to enhance insight-based evaluations.

3 METHODOLOGY

Our main goal is to understand *if obstacles exist towards the practical use of insight-based methods for evaluating visualizations.*

3.1 Research Questions

We address this goal by trying to answer two main research questions. First, we want to understand if and to what extent the insight categorization model affects the descriptions of insights based on the same corpus of responses. Previous studies have shown that insight generation depends on the presentation and organization of visualizations [21, 35]. In particular, Moere et al. [35] studied how people arrive at their insight with visualizations varying in embellishment features. The authors started by categorizing insights based on Chen et al.'s taxonomy [12]. However, Moere et al. suggest that Chen et al.'s taxonomy is more appropriate to differentiate analytical insights. Consequently, Moere et al. developed another taxonomy with fewer but broader categories that cover non-analytical aspects of insight generation. However, further research is necessary to understand if the different taxonomies affect the effects fostered by visualization design. Therefore, our first research question is:

RQ1 *Do categorization models create obstacles to insight-based evaluations?*

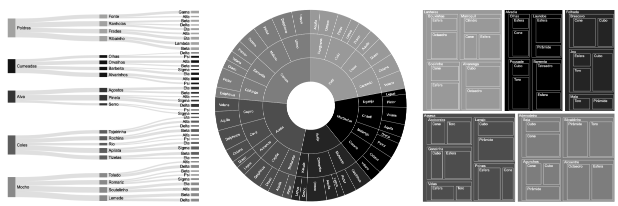
Second, prior work by Green and Fisher [21] shows that neuroticism and extraversion affect insight reporting with hierarchical visualizations. To the best of our knowledge, this is the only work that tries to understand if personality factors affect the number of insights in insight-based evaluations. Although Green and Fisher [21] provide the initial steps to study the manifestation of personality in this setting type, more personality traits may show measurable results and, consequently, enhance the user profile characterization in insight-based evaluations. We decided to try to replicate the findings of Green and Fisher [21] and extend them by including the remaining traits of the FFM: openness to experience, agreeableness, and conscientiousness. In particular, we want to observe if gaining insights varies between the visualization layouts while controlling for the personality traits. Our second research question is:

RQ2 *Does personality affect the insight generation process?*

3.2 Visualizations

Following our prior methodology described in Alves et al. [3] and state-of-the-art research [4, 55], we decided to use a set of three hierarchical layouts composed of a Sankey diagram, a sunburst, and a treemap (see Figure 1). We believe that this composition offers enough complexity since each of the chosen graphs has a different approach to encoding a quantitative variable: (i) a **Sankey** diagram where the length of a bar encodes quantitative values and the hierarchical order unfolds from left to right; (ii) a **sunburst** chart that uses an angle channel to describe quantitative measures, and a segment of the inner circle has a hierarchical relationship with those segments of the outer circle which lie within the angular sweep of the parent segment; and (iii) a **treemap** which uses nested rectangles whose area is proportional to a quantitative variable to depict a tree-structure. Additionally, visualization literature reports that leveraging different visual encodings such as length (Sankey), angle (sunburst), and area (treemaps) may influence how relationships are perceived (e.g., [25, 39]).

The main application of these visual idioms is to support reasoning about the prevalence of specific quantities while following a hierarchical structure [47, 49]. In contrast with the remaining charts, designers usually use Sankey diagrams to represent the flow count between categorical variables with similar hierarchical levels. Nevertheless, we believe that Sankey charts can provide fresh insights into our study since these graph types use a visual channel based on the length and can also incorporate hierarchy characteristics effectively [18]. Further, we fixed the hierarchical depth at a factor of three so that it is possible to ask the subjects to conduct complex examinations through multiple non-sequential hierarchy levels. By fostering more complex exploration patterns, we believe that participants may generate insights that vary in how deep they bring about new knowledge or create further engaging questions. Finally, hovering an item triggers the appearance of a tooltip showing the exact value of the item.



(a) Sankey. (b) Sunburst. (c) Treemap.

Figure 1: Three charts of the same hierarchical dataset, but with different graphical configurations.

Regarding the databases we base the visualizations on, we developed three versions of a structurally identical dataset where the labels differ across the datasets to diminish the knowledge and preference biases that some datasets inherently carry. Further, using datasets with the same structure but different domains provides a means for tasks to focus on the same target element independently of the visualization and, consequently, diminish noise in data collection. We chose the labels of items to avoid any learning bias or potential perceptual and semantic confounds. Above each visualization was a title corresponding to the context of the chart. The three domains were: (i) *Fans*, which presented the number of fans per band acting at music festivals hosted in different cities; (ii) *Sand*, which shows the number of grains of sand of a particular shape present in deserts from different cities; and (iii) *Students*, which includes the number of students per faculty of a university from different cities.

3.3 Tasks

We devised a set of five types of tasks that participants performed to interact with each visualization (from a related study [3]). Our choices cover a wide range of task types achievable using each chart type. In particular, these low-level goal tasks are part of the most primitive analysis task types in visual analytics (VA) [5]. We ran a pilot test (N=3) beforehand to ensure that the wording and difficulty allowed participants to accomplish the tasks without misunderstanding. The set is composed of tasks that focus on (i) hierarchical fragmentation (*"Which category has more subcategories?"*), (ii) between-levels analysis (*"Which category in the highest hierarchic level has the largest quantity of a specific category in the lowest hierarchic level?"*), (iii) maxima identification (*"For a specific category, which of its subcategories has the largest quantity?"*), (iv) sum estimation (*"What is the quantity of a specific category in the highest hierarchic level?"*), and (v) value retrieval (*"What is the quantity of a specific category in the lowest level?"*). Moreover, participants had no time limit to complete the tasks, and we accepted only one response to each question. For each task type, there was only one instance for a domain. It results in a total of 15 tasks (5 types × 3 domains).

3.4 Insight Evaluation Models

Chen et al.'s [12] Model This taxonomy is based fundamentally on categorizing analytical insights. Chen et al. use a taxonomy consisting of twelve categories. However, in our analysis, we also include the Meaning category, which Moere et al. [35] developed and added to this taxonomy. We opted to use this model since it is among the first general taxonomies that categorize insights [12].

Moere et al.'s [35] Model Another relevant model was defined by Moere et al. [35]. We decided to include this model as it provides more breadth to analyze insights not based on facts, which is a limitation of the Chen et al.'s [12] taxonomy. Following an open coding strategy, Moere et al. [35] grouped similar insights based on their "type". In particular, an insight can be of one of these types: rational, technical, emotional, plain, analytical, or interface.

Insight Valence After studying previous work on insight characterization [12, 35], we verified that the existent models focus mostly on the content of the insight. Therefore, the models often neglect *how* individuals to report the content of the insight. We believe this analysis may also allow us to understand the effect of the visual encodings of hierarchical data on insight generation. In particular, we want to expand the study of this effect to the *valence* of the insight, i.e., whether the insight includes negative, positive, or no sentiment words. For instance, *"It is easy to compare categories with this chart"* has positive sentiment, while *"I dislike the squared encoding"* has a negative valence. Other insights such as *"The most frequent category is cube"* are neutral since they do not contain any words related to valence. As such, an insight can be classified as **negative**, **neutral**, or **positive**.

3.5 Measures

Insights Insight recording can be applied through several methods such as the think-aloud protocol [21, 38] or annotations [23, 35]. In our study, we use the think-aloud method. We collected insights after the participant performed the five tasks for visualization. We asked each subject to report "any interesting findings or observations they could collect from the information visualization". This way, user insight is characterized by the following dimensions: the visualization that the user interacted with, the domain that the user observed, and a category for each model presented in Section 3.4.

Demographics We recorded the gender, age, self-reported visual acuity, and whether the participant was color-blind. We also monitored familiarity factors related to the visual idioms; first, we presented an instance of each of the charts we leverage with an exemplary but different from the other domain (see Section 3.2). Then, we asked participants to (i) assess their familiarity with that visual representation in a five-point Likert scale ranging from *not familiar* (1) to *very familiar* (5), (ii) report the name of the chart, and (iii) perform an analysis task similar to the ones used in the study (see Section 3.3) to assure whether participants could understand the information the visualization conveyed independently of their self-assessed familiarity.

Personality factors The FFM [15] has been shown to subsume most known personality traits, and researchers claim that this model represents the "basic structure" underlying the variation in human behavior and preferences [26]. This model is a hierarchical organization of personality traits in five dimensions: neuroticism, extraversion, openness to experience, agreeableness, and conscientiousness. We collected personality data with the Revised NEO Personality Inventory (NEO PI-R) [16]. The questionnaire identifies the intensity of each personality trait using high-score and low-score items. In particular, both openness to experience and conscientiousness are measured with a total of 48 items (eight items for each six facets), where each item is based on assertions semantically connected to behaviors and five possible alternatives of agreement: *strongly agree*, *agree*, *undecided*, *disagree*, and *strongly disagree*. As such, the score for a trait is a continuous variable encoded by the sum of each Likert scale value of the subset of items corresponding with a specific personality trait.

3.6 Procedure

We recruited subjects in university settings through direct contact and word of mouth. Subjects included any participant at least 18 years old. Our study comprises a total of 51 participants (26 females, 25 males) aged 20 – 61 ($M = 28.31, SD = 11.02$). We attempt to balance the educational background by recruiting 22 (43.1%) participants with a computer sciences background, 12 (23.5%) in other engineering areas, and the remaining 17 (33.3%) in additional research areas. Moreover, our sample is composed of undergraduates (64.7%), masters (29.4%), and doctorates (5.9%). Firstly, participants signed a consent form. Then, they completed the NEO PI-R and the demographic and familiarity questionnaire. We then presented each visualization layout in random order and asked the participant to complete a set of five tasks for each chart ($3 \times 5 = 15$ tasks total). We randomized the order of the tasks and datasets to reduce any potential bias. In particular, we randomized each pair (visualization, domain) to produce mutually exclusive instances of visualization and dataset, e.g., one experiment consisted of the order {(sunburst,fans), (Sankey,students), (treemap,sand)}. After performing all tasks in a layout, we asked subjects to report all insights they gained from the visualization. The assistant asked the participant if there were any more insights they could generate independently of the number of insights that subjects reported per visualization. After interacting with the three visual idioms, participants were eligible for a raffle of three 20€gift cards. The total time for each session was between 18 and 39 minutes. This study obtained ethics approval from the Ethics Committee of the university.

3.7 Study Design and Data Analysis

We conducted a mixed-measures study where each participant interacts with all hierarchical layout conditions, one at a time in random order. We first analyze the reported insights with descriptive statistics to provide an overview of the collected data. Two researchers classified each insight based on Chen et al.'s [12], Moere et al.'s [35], and Insight Valence taxonomies. We did not include any more experts as this approach led to an inter-coder agreement of 31.55%, 40.06%, and 90.22%, respectively. We expected the low inter-coder agreement since Moere et al. [35] also had a value of 34.4% when applying Chen et al.'s [12] taxonomy. To counter the low inter-coder agreement, the first author discussed with the experts to revisit all insights and consolidate the classifications in mutual agreement. Consequently, all researchers agree with the insight characterizations of the final dataset. In addition to the descriptive statistics, we ran a chi-square test of independence (r×c) with hierarchical layout (3 levels) and insight categories as factors. The number of levels in the insight categories depends on the categorization model.

We ran an apriori power analysis using the `pwr` R library[1] to find the minimum sample size required. The required sample size to achieve 80% power for detecting a medium effect (0.3) with a significance criterion of $\alpha = .05$ was N = 26 for multiple regression methods. We believe the current sample size (N=51) is adequate for our statistical analysis. We investigate insight generation based on the number of insights each user reported. We ran a one-way repeated-measures ANCOVA with a hierarchical layout (3 levels) as a factor. We include the FFM personality traits and self-reported familiarity as covariates. Alves et al. [3] also used these personality data. We tested for sphericity (Mauchly's test) and used the Greenhouse-Geisser correction when the assumption was not met. We complement our analysis through Spearman's rank-order correlation tests and include LOESS (locally estimated scatterplot smoothing) lines to help analyze the correlations.

[1] `https://cran.r-project.org/web/packages/pwr/vignettes/pwr-vignette.html` (Last access: September 23, 2022).

4 Results

This section covers the results of our study. We present data as mean $\pm$ standard deviation unless otherwise stated.

4.1 Insight Generation

We collected 313 valid insights with each participant reporting between 0 and 16 insights (6.14 ± 3.54). The Sankey diagram generated 114 insights (36.42%), followed by the sunburst with 104 (33.23%), and then the treemap with 95 (30.35%). We analyzed the reported insights through the models presented in Section 3.4.

4.1.1 Chen et al.'s Model

Table 1 depicts the insights distribution according to Chen et al.'s [12] taxonomy. At first glance, the large portion of Meta Facts shows that we captured the major limitation of Chen et al.'s [12] taxonomy. Our results are similar to Moere et al. [35] since this taxonomy cannot categorize insights that do not have an analytical nature. In our study, the Meta Fact type makes up 44.19% of the insights sample. However, insights such as *"It is hard to compare categories that are far away."* or *"The angle is the most important feature in this graph."* are essential to show characteristics of clearness, intuitiveness, and usability. The other categories with considerable portions are Distribution (e.g. *"The majority of the universities has two faculties."*), Extreme (e.g. *"Coles has the highest number of universities."*), and Trend (e.g. *"Each desert has more than one shape of sand grains."*). The remaining categories showed small percentages. However, we were able to find some interesting findings. The Distribution and Extreme insights percentages in the Sankey are more than double compared to the treemap. In contrast, the treemap made users report four times more Difference and two times more Outliers insights than the other two charts. Finally, only the sunburst led users to report Categories insights.

Although Chen et al.'s [12] taxonomy suffer from a robust limitation in insight analysis, we can understand that the encoding of hierarchical data fosters insights from different categories. Notably, the data structure is the same independently of the hierarchical idiom, reinforcing the significance of the effect. Afterward, we ran a chi-square test of independence between hierarchical layout and insight category according to Chen et al.'s [35] taxonomy. 15 cells (50.0%) have an expected count of less than five. There was not a statistically significant association between layout and category, $\chi^2(18) = 28.688, p = .052$. In particular, the association was moderate [14], Cramer's V $= 0.214$. This result suggests that **the graphical disposition of the hierarchical data may be relevant** but the large percentage of cells with a count of less than five hinders the statistical analysis.

4.1.2 Moere et al.'s Model

Table 1 reports on the insight generation according to the Moere et al.'s [35] taxonomy. Many insights fell under the Analytical (e.g., *"The majority of the universities has two faculties."*) or Emotional (e.g., *"It is quite easy to see the distribution in subcategories."*) categories. All layouts showed similar percentages suggesting that the graphical disposition of the information does not interfere with the analytic reasoning of the participants. Both Plain (e.g., *"The donut shape is not present in Lanhelas."*) and Interface (e.g., *"The letters are hard to read."*) held around 10% of the insights each. However, it is noteworthy that the Sankey diagram received almost 50% of the Interface insights. The scarcer categories were Technical (e.g., *"Corvos and Arada have half the fans of all concerts."*) and Rational (e.g., *"Corvos appeals to more people because it has all*

Table 1: Insights by each taxonomy, in absolute and relative numbers. Lightness encodes frequency per column.

		Sankey	*Sunburst*	*Treemap*	**Total**
Chen et al.'s	*Association*	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
	Categories	0 (0.0%)	2 (1.92%)	0 (0.0%)	2 (0.64%)
	Cluster	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
	Compound	6 (5.26%)	9 (8.65%)	6 (6.32%)	21 (6.74%)
	Difference	2 (1.75%)	2 (1.92%)	8 (8.42%)	12 (4.03%)
	Distribution	17 (14.91%)	13 (12.50%)	6 (6.32%)	36 (11.24%)
	Extreme	25 (21.9%)	14 (13.46%)	10 (10.53%)	49 (15.31%)
	Meaning	1 (0.88%)	1 (0.96%)	1 (1.05%)	3 (0.96%)
	Meta Fact	48 (42.11%)	47 (45.19%)	43 (45.26%)	138 (44.19%)
	Outliers	3 (2.63%)	2 (1.92%)	7 (7.37%)	12 (3.97%)
	Rank	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
	Trend	9 (7.89%)	10 (9.62%)	13 (13.68%)	32 (10.40%)
	Value	3 (2.63%)	4 (3.85%)	1 (1.05%)	8 (2.51%)
	Total	114 (36.42%)	104 (33.23%)	95 (30.35%)	313
Moere et al.'s	*Analytical*	55 (48.25%)	41 (39.42%)	46 (48.42%)	142 (45.36%)
	Emotional	35 (30.70%)	38 (36.54%)	30 (31.58%)	103 (32.94%)
	Interface	13 (11.40%)	9 (8.65%)	5 (5.26%)	27 (8.44%)
	Plain	7 (6.14%)	13 (12.50%)	12 (12.63%)	32 (10.42%)
	Rational	1 (0.88%)	1 (0.96%)	1 (1.05%)	3 (0.96%)
	Technical	3 (2.63%)	2 (1.92%)	1 (1.05%)	6 (1.87%)
	Total	114 (36.42%)	104 (33.23%)	95 (30.35%)	313
Insight Valence	*Negative*	19 (16.67%)	20 (19.23%)	22 (23.16%)	61 (19.69%)
	Neutral	78 (68.42%)	63 (60.58%)	66 (69.47%)	207 (66.16%)
	Positive	17 (14.91%)	21 (20.19%)	7 (7.37%)	45 (14.16%)
	Total	114 (36.42%)	104 (33.23%)	95 (30.35%)	313

bands from the other cities."). These distributions suggest that the hierarchical layouts pushed the insight generation to visual patterns rather than filters or reasoning.

We conducted a chi-square test of independence between hierarchical layout and insight category according to Moere et al.'s [35] taxonomy. Six cells (33.3%) have an expected count of less than five. Similar to the previous test, there was a statistically nonsignificant association between layout and category, $\chi^2(10) = 7.730, p = .655$. The association was small [14], Cramer's V = 0.111. Overall, it appears that **the graphical disposition does not affect how users generate insights** according to Moere et al.'s [35] taxonomy.

4.1.3 Insight Valence

Table 1 presents the insight distribution in terms of valence. Nearly two-thirds of the insights have a neutral valence (e.g., *"The donut shape is not present in Lanhelas."*). Negative valence insights (e.g., *"It is hard to compare categories so far away."*) were slightly more frequent than positive ones (e.g., *"It is easy to see the hierarchy structure."*). However, the most appealing aspect is how the chart type affected the report of positive valence insights. This type of insight was much more frequent in the Sankey and sunburst compared to the treemap. Finally, we conducted a chi-square test of independence between reported a nonsignificant association between layout and insight valence, $\chi^2(4) = 7.576, p = .108$. The association was small [14], Cramer's V = 0.110. These findings suggest that **the graphical disposition usually does not affect the valence of insights**. However, positive insights are less frequent in the treemap.

4.2 Personality Factors

An ANCOVA showed no statistically significant differences across the different layouts on the reported number of insights while controlling for personality, $F(1.658, 74.598) = 3.117, p = .059$, partial $\eta^2 = .065$. The small p-value and a medium effect size led us to explore in-depth the role of personality (Figure 2). We found a statistically significant main effect of openness to experience on the number of reported insights between visualizations, $F(1,45) = 5.261, p = .027$, partial $\eta^2 = .105$. In particular, it appears that the effect is more evident in the Sankey, $r_s(51) = .426, p = .002$. The nonsignificant effects were weak to moderate in the sunburst, $r_s(51) = .163, p = .254$, and in the treemap, $r_s(51) = .206, p = .148$. Findings suggest that the openness to experience scores positively influences the number of reported insights.

The remaining traits did not show measurable main effects. Contrary to Green and Fisher [21], we did not find evidence for the manifestation of neuroticism, $F(1,45) = 0.059, p = .809$, partial $\eta^2 = .001$, or extraversion, $F(1,45) = 0.470, p = .496$, partial $\eta^2 = .010$. Neuroticism consistently showed nonsignificant weak correlations in the Sankey, $r_s(51) = .126, p = .377$, sunburst, $r_s(51) = .096, p = .505$, and treemap, $r_s(51) = .174, p = .222$. Further, extraversion showed a nonsignificant moderate correlation in the Sankey, $r_s(51) = .231, p = .102$ and weak effects on the sunburst, $r_s(51) = -.160, p = .261$, and the treemap, $r_s(51) = .114, p = .425$. Regarding the remaining traits, agreeableness did not significantly affected the number of reported insights, $F(1,45) = 1.867, p = .179$, partial $\eta^2 = .040$, although we observe a small to medium effect. Results show nonsignificant weak correlations in the Sankey, $r_s(51) = .009, p = .950$, sunburst, $r_s(51) = -.072, p = .616$, and treemap, $r_s(51) = -.069, p = .631$. Finally, it appears that conscientiousness does not play a role in gaining insights as well, $F(1,45) = 0.233, p = .632$, partial $\eta^2 = .005$. We found nonsignificant weak correlations in the Sankey, $r_s(51) = -.025, p = .862$, and treemap, $r_s(51) = -.012, p = .935$. However, there were a nonsignificant moderate correlation with in the sunburst, $r_s(51) = -.255, p = .070$.

4.3 Additional Findings

We complement our analysis by understanding the effect of the visualization domain and self-reported familiarity in the measurements. Regarding the domain of the visualizations, we found a balanced distribution between the *Fans* (32.59%), *Sand* (34.82%), and *Students* (32.59%) domains. An ANCOVA showed no statistically significant differences across the different domains on the reported number of insights while controlling for personality, $F(1.507, 67.814) = 0.297, p = .681$, partial $\eta^2 = .007$. Next, we ran Spearman correlations for each chart type to find if the self-reported literacy with said chart affected how many insights the participant reported. It appears that there is no significant correlation between the reported insights and the familiarity with the Sankey, $r_s(51) = -.135, p = .344$, sunburst, $r_s(51) = .050, p = .726$, and treemap, $r_s(51) = .054, p = .709$, charts.

4.4 Discussion

Findings show that the visual encoding of hierarchical data does not have a measurable effect on the type of insights users report according to the taxonomies. In addition, results predict that the openness to experience trait predicts insight reporting.

4.4.1 Insight Categorization

We used three taxonomies to analyze the insights. First, Chen et al.'s [12] taxonomy is one of the first categorization models dedicated to insights. On the one hand, this categorization clearly supports the visualization system evaluation through analytical insights. Analytical tasks and the insights they derive are part of the hallmarks of visualization [51]. In particular, these lower-level insights may play a significant role in higher-level insights generation and, consequently, knowledge [6]. For instance, researchers first have to understand what some clusters or trends (lower-level insights) are before being able to predict future states or discover causality effects (higher-level insights). On the other hand, the focus on analytical tasks may also prove to be an obstacle in insight-based evaluations. This task type is usually used as benchmark tasks to foster user interaction [11]. However, they suffer from being often too narrow and having simple answers to allow for unexpected insights. Moreover, they require a short completion time with a definitive answer. All these factors usually constrain the thought process when moving on to the open-ended portion [36]. Consequently, the research agenda pushes for qualitative analysis based on user exploration. Chen et al. created the Meta Fact category to account for other aspects of insight generation besides analytical tasks. Additionally, we expanded Chen

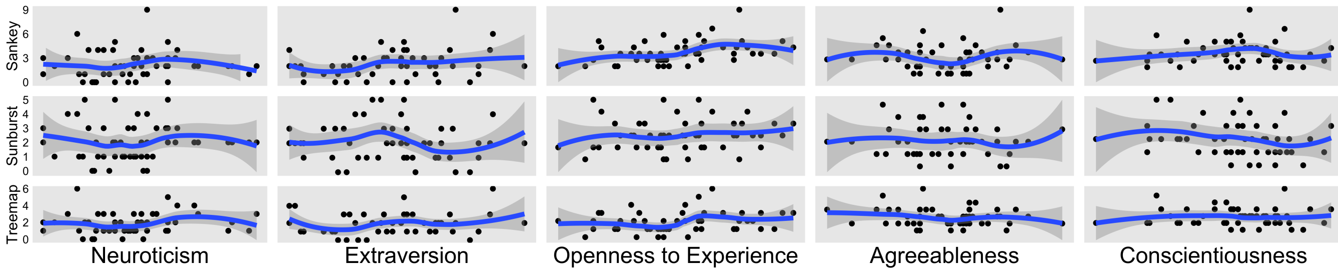


Figure 2: Correlation of the reported insights per visualization and personality trait.

et al.'s taxonomy to contain Meaning facts based on Moere et al. [35]. However, we also found that combining both categories could not account for Chen et al.'s taxonomy unless they asked participants to report only on analytical facts they gained from the data.

Moere et al. created a single category to account for analytical insights and subdivided the remaining insights into five other types. If we compare the taxonomies, Chen et al.'s showed 45.15% non-analytical and 54.85% analytical insights, while Moere et al.'s led to 54.64% non-analytical and 45.36% analytical insights, respectively. This discrepancy between the models appears to derive from the remaining categories from Moere et al.'s taxonomy such as Plain and Rationale. For instance, insights such as "Each city can have more than one desert." or "There are many shapes of grain sands." can be classified as Plain insights according to Moere et al. since they report "broad, general observation(s) with no reasoning and few filters" [35]. However, Chen et al. could categorize them as Distribution insights since they characterize distributions [12]. Another clear example is "It seems that the Ulono concert was the biggest of them all." Moere et al. would classify it as an Emotional insight since it is an observation that contains some subjective interpretation or judgment from the participant. In contrast, Chen et al. would see it as an Extreme insight. Results also allow us to observe that Moere et al.'s and the Insight Valence models may provide a similar description of the insights. We found that Emotional insights accounted for 32.94% of the total number of reported insights according to Moere et al., which is similar to the sum of the number of insights with a negative or positive valence from the Insight Valence model (33.84%). However, if we assume that analytical insights are of a neutral insight valence since they are only factual observations, we can verify another discrepancy. In particular, the Insight Valence model reported that 66.16% of the insights fulfill this requirement, while Chen et al.'s taxonomy shows that 54.85% of the reported insights are analytical. Consequently, our findings lead us to believe that the categorization model may produce different descriptions of insights based on the same corpus of responses.

4.4.2 Personality Factors

A vast body of knowledge leverages personality factors to enhance the user profile in information visualization. Concerning insight generation, Green and Fisher [21] showed that LoC, extraversion, and neuroticism affect the number of insights generated by users. Our work tries to replicate the findings of Green and Fisher [21] related to neuroticism and extraversion. However, we did not find measurable effects of both traits in the number of reported insights. These traits showed only nonsignificant weak to moderate effects across visualization layouts. Several possible explanations exist for our failure to validate the work of Green and Fisher [21]. First, social and demographic contexts strongly influence personality by defining one's decision-making process [27]. Since the study samples were from different continents, some differences may have introduced some noise in the results. Second, both studies leverage hierarchical visualizations. However, Green and Fisher use high-fidelity visualizations built to display genomic information. In our case, the visualizations have a plain look and complexity. Moreover, the visualization layouts are different. Although both studies focus on a hierarchical context, these differences may have contributed to the insight generation process. Third, Green and Fisher used open tasks and let participants answer until they were correct. In contrast, we ask straightforward questions and only accept the first answer. These different approaches may affect how much time users take interpreting the visualization and, consequently, how many insights they could generate from it.

However, other personality factors from well-established Personality Psychology research, more specifically the FFM, may hold valuable insights. We extend the current state-of-the-art by studying how the remaining personality traits from the FFM affect insight reporting. We found a positive influence of openness to experience on the number of reported insights. This dissimilarity was more noticeable in the Sankey diagram. We believe that it bases on how comfortable individuals with high openness to experience scores are with abstract and imaginative thinking [55]. Ziemkiewicz and Kosara [55] found that high openness to experience led individuals to be faster while solving problems related to hierarchical visualizations that include conflicting visual and verbal metaphors. Paired with our findings, high openness to experience may lead individuals to learn a novel interface faster and report more pertinent information than their counterparts. Finally, the agreeableness and conscientiousness traits did not significantly affect insight generation. To the best of our knowledge, our results are similar to past research since these traits did not manifest measurable effects in user interaction with visualization [7,55]. However, we believe that the nonsignificant small to medium size effects that both traits show a need for more work to explore this relationship more in-depth in visualization contexts.

4.4.3 Lessons Learned

Inspired by our findings, we formulated the following lessons learned focused on the design of insight-based evaluations:

Insight categorization models introduce noise: We used two models from state-of-the-art research and complemented our analysis with a self-developed model. As we have shown, the categories from Chen et al.'s [12] taxonomy are more appropriate for factual insights. Our study demonstrates that taxonomies that leverage the analytical features of insights and their meaning [35] coupled with *how* users frame their insights (Insight Valence model) provide an expanded understanding of this process. However, we found discrepancies between models when investigating the insight descriptions based on the same corpus of responses. It suggests a challenge for insight-based evaluations. In particular, researchers should consider that their findings may derive from the used insight taxonomy rather than the design conditions, e.g., differences between visualizations. The visualization community should also pay attention to the low inter-code agreement that both Chen et al.'s [12] and Moere et al.'s [12] taxonomies revealed. Further, researchers may focus on assessing insight quality (e.g., through interaction logs [23]) rather than quantity or use dimension reduction techniques to find key insight categories. We advise future studies to consider creating and validating a model that converges on a more resilient taxonomy for insights-based evaluations.

Personality synergies affect insight reporting: Prior research has already created a substantial body of knowledge focused on the effect of personality factors on user interaction with information visualization systems (e.g., [21, 54, 55]). In particular, neuroticism, extraversion, and LoC are the only personality traits considered in insight-based evaluations [21]. We were not able to validate the findings of Green and Fisher [21] regarding the neuroticism and extraversion traits. However, our findings show that high scores in the openness to experience trait lead individuals to report more insights and vice-versa. In particular, this effect has a medium to large size. We believe that researchers should consider these effects in the experimental design. In particular, these findings suggest that the credibility of insight-based evaluations may be suspect without controlling for personality effects. We believe enhancing the user profile with individual characteristics and breaking the cycle of one-size-fits-all design is a significant step for future research.

4.4.4 Limitations and Future Work

Although the above recommendations provide preliminary steps into insight generation and its susceptibility to personality factors, there are some limitations to the results of this study. First, the sample size was adequate to study the insight reporting across the different visualization layouts. However, including latent variables such as personality traits often requires hundreds of participants to achieve stable estimations [28]. Future studies should increase the sample size to verify whether our results hold for larger samples. Second, our study and Green and Fisher's [21] show that personality plays a significant role in the insight generation process. Future studies should replicate the experiments and validate the findings to provide a more robust body of knowledge. For instance, researchers can leverage a guidance system to support individuals with average or low scores in openness to experience. In particular, visualizations can integrate notes and help features to foster user exploration. This approach type makes visualizations more cognitively in line with the personality characteristics of users and, consequently, may allow them to find a higher number of insights.

Third, we designed a set of tasks to foster within and between hierarchical levels comparisons. Contrary to past work [55], the tasks were not yes/no questions but required the participant to report an exact number or label. Furthermore, the study involved a set of fixed tasks before probing for insights. This study design decision comes from our attempt to diminish bias. We chose the datasets to de-emphasize the role of domain knowledge and to try to isolate the effect of the different graphical dispositions in insight generation. Consequently, we believe it was necessary to prompt the users to interact with the visualizations before asking them to report any insights. Both these factors may have introduced some noise and primed the participants to report insights from a specific type more frequently [36]. Future work may also consider how our results might differ from a more open-ended data exploration (e.g., [35]). Finally, our results apply only to Sankey, sunburst, and treemap charts with three hierarchical levels. Future research can vary the visual idiom to present hierarchical data and non-hierarchical data, its features (chart size, scale, or color palette), and the number of hierarchy levels to further explore how the graphical disposition of the data affects how people gain insights. This would help to tease apart what exactly visualizations produce differences in insight reporting, e.g., the actual data aggregation or the visual encodings [52].

5 Conclusions

This study reports on findings regarding obstacles in the design of insight-based evaluations. We conducted a user study where participants interacted with a Sankey, a sunburst, and treemap charts and then reported insights about these visualizations. We continue prior work on insight categorization by applying Chen et al.'s [12] and Moere et al.'s [35] taxonomies. We also propose a new Insight Valence model to study the insight reporting delivery. Results suggest that the insight taxonomy used to analyze the data may introduce noise in the insights description. Further, we show that openness to experience affects the insight generation process. These findings suggest that personality traits may create crucial tensions in the process of gaining insights and, consequently, researchers should control or account for it in this type of experiment. We advise future studies to use insight-based evaluations to control for the obstacles that may undermine their results.

Acknowledgments

This work was supported by national funds through Fundação para a Ciência e a Tecnologia with references SFRH/BD/144798/2019 and UIDB/50021/2020. The authors would also like to thank the reviewers for their essential feedback.

References

[1] E. Adar and E. Lee. Communicative visualizations as a learning problem. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):946–956, 2021. doi: 10.1109/TVCG.2020.3030375

[2] I. Alexandre. Promoting insight: A case study of how to incorporate interaction in existing data visualizations. In *2016 20th International Conference Information Visualisation (IV)*, pp. 203–208. IEEE, 2016.

[3] T. Alves, C. Dias, J. Henriques-Calado, D. Gonçalves, and S. Gama. Exploring how visual channels and personality affect trust in hierarchical visualizations. Under review.

[4] T. Alves, B. Ramalho, J. Henriques-Calado, D. Gonçalves, and S. Gama. Exploring how personality models information visualization preferences. In *2020 IEEE Visualization Conference (VIS)*, pp. 201–205. IEEE, 2020.

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

[6] R. Amar and J. Stasko. Knowledge precepts for design and evaluation of information visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 11(4):432–442, 2005. doi: 10.1109/TVCG.2005.63

[7] E. T. Brown, A. Ottley, H. Zhao, Q. Lin, R. Souvenir, A. Endert, and R. Chang. Finding waldo: Learning about users from their interactions. *IEEE Transactions on visualization and computer graphics*, 20(12):1663–1672, 2014.

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

[9] D. Cashman, Y. Wu, R. Chang, and A. Ottley. Inferential tasks as a data-rich evaluation method for visualization. In *EVIVA-ML: IEEE VIS Workshop on EValuation of Interactive VisuAl Machine Learning systems*, vol. 7, 2019.

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

[11] C. Chen and Y. Yu. Empirical studies of information visualization: a meta-analysis. *International Journal of Human-Computer Studies*, 53(5):851–866, 2000. doi: 10.1006/ijhc.2000.0422

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

[13] E. K. Choe, B. Lee, et al. Characterizing visualization insights from quantified selfers' personal data presentations. *IEEE computer graphics and applications*, 35(4):28–37, 2015.

[14] J. Cohen. *Statistical power analysis for the behavioral sciences*. Academic press, 2013.

[15] P. Costa and R. R. McCrae. The revised neo personality inventory (neo-pi-r). *The SAGE Handbook of Personality Theory and Assessment*, 2:179–198, 01 2008. doi: 10.4135/9781849200479.n9

[16] P. T. Costa Jr and R. R. McCrae. *The Revised NEO Personality Inventory (NEO-PI-R).* Sage Publications, Inc, 2008.

[17] M. B. Duggimpudi, S. Abbady, J. Chen, and V. V. Raghavan. Spatio-temporal outlier detection algorithms based on computing behavioral outlierness factor. *Data & Knowledge Engineering*, 122:1–24, 2019.

[18] S. G. Eick and G. J. Wills. Navigating large networks with hierarchies. In *Proceedings Visualization'93*, pp. 204–210. IEEE, 1993.

[19] L. Goldberg. The structure of phenotypic personality traits. *American Psychologist*, 48:26–34, 02 1993. doi: 10.1037/0003-066X.48.1.26

[20] A. Gori, A. Arcioni, E. Topino, L. Palazzeschi, and A. Di Fabio. Constructing well-being in organizations: First empirical results on job crafting, personality traits, and insight. *International Journal of Environmental Research and Public Health*, 18(12):6661, 2021.

[21] T. M. Green and B. Fisher. Towards the personal equation of interaction: The impact of personality factors on visual analytics interface interaction. In *2010 IEEE Symposium on Visual Analytics Science and Technology*, pp. 203–210. IEEE, 2010.

[22] S. Halko and J. A. Kientz. Personality and persuasive technology: an exploratory study on health-promoting mobile applications. In *International conference on persuasive technology*, pp. 150–161. Springer, 06 2010. doi: 10.1007/978-3-642-13226-1_16

[23] C. He, L. Micallef, L. He, G. Peddinti, T. Aittokallio, and G. Jacucci. Characterizing the quality of insight by interactions: A case study. *IEEE transactions on visualization and computer graphics*, 2020.

[24] K. Hooker and D. P. McAdams. Personality reconsidered: A new agenda for aging research. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 58(6):P296–P304, 2003.

[25] N. Jardine, B. D. Ondov, N. Elmqvist, and S. Franconeri. The perceptual proxies of visual comparison. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):1012–1021, 2020. doi: 10.1109/TVCG.2019.2934786

[26] M. Kosinski, Y. Bachrach, P. Kohli, D. Stillwell, and T. Graepel. Manifestations of user personality in website choice and behaviour on online social networks. *Machine Learning*, 95(3):357–380, 2014. doi: 10.1007/s10994-013-5415-y

[27] B. Krahé. *Personality and social psychology: Towards a synthesis.* Sage Publications, Inc, 1992.

[28] A. Kretzschmar and G. E. Gignac. At what sample size do latent variable correlations stabilize? *Journal of Research in Personality*, 80:17–22, 2019. doi: 10.1016/j.jrp.2019.03.007

[29] S. Lallé and C. Conati. The role of user differences in customization: a case study in personalization for infovis-based content. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*, pp. 329–339, 2019.

[30] S. Lallé, D. Toker, C. Conati, and G. Carenini. Prediction of users' learning curves for adaptation while using an information visualization. In *Proceedings of the 20th International Conference on Intelligent User Interfaces*, pp. 357–368, 2015.

[31] H. M. Lefcourt. *Locus of control: Current trends in theory & research*. Psychology Press, 2014.

[32] Z. Liu, R. J. Crouser, and A. Ottley. Survey on individual differences in visualization. In *Computer Graphics Forum*, vol. 39, pp. 693–712. Wiley Online Library, 2020.

[33] A. Lundgard and A. Satyanarayan. Accessible visualization via natural language descriptions: A four-level model of semantic content. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):1073–1083, 2022. doi: 10.1109/TVCG.2021.3114770

[34] E. Mayr, M. Smuc, and H. Risku. Many roads lead to rome: Mapping users' problem-solving strategies. *Information Visualization*, 10(3):232–247, 2011.

[35] A. V. Moere, M. Tomitsch, C. Wimmer, B. Christoph, and T. Grechenig. Evaluating the effect of style in information visualization. *IEEE transactions on visualization and computer graphics*, 18(12):2739–2748, 2012.

[36] C. North. Toward measuring visualization insight. *IEEE computer graphics and applications*, 26(3):6–9, 2006.

[37] C. North, R. Chang, A. Endert, W. Dou, R. May, B. Pike, and G. Fink. Analytic provenance: process+ interaction+ insight. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems*, pp. 33–36. 2011.

[38] C. North, P. Saraiya, and K. Duca. A comparison of benchmark task and insight evaluation methods for information visualization. *Information Visualization*, 10(3):162–181, 2011.

[39] B. Ondov, N. Jardine, N. Elmqvist, and S. Franconeri. Face to face: Evaluating visual comparison. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):861–871, 2019. doi: 10.1109/TVCG.2018.2864884

[40] A. Ottley, H. Yang, and R. Chang. Personality as a predictor of user strategy: How locus of control affects search strategies on tree visualizations. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pp. 3251–3254, 2015.

[41] G. Palshikar et al. Simple algorithms for peak detection in time-series. In *Proc. 1st Int. Conf. Advanced Data Analysis, Business Analytics and Intelligence*, vol. 122, 2009.

[42] M. Rester, M. Pohl, S. Wiltner, K. Hinum, S. Miksch, C. Popow, and S. Ohmann. Evaluating an infovis technique using insight reports. In *2007 11th International Conference Information Visualization (IV'07)*, pp. 693–700. IEEE, 2007.

[43] J. B. Rotter. General expectancies for internal versus external control of reinforcement. *Psychological monographs*, 80:1–28, 02 1966. doi: 10.1037/h0092976

[44] M. T. Russell, R. B. Cattell, A. Cattell, H. E. Cattell, and D. L. Karol. *16PF fifth edition administrator's manual*. Institute for Personality and Ability Testing, Incorporated, 1994.

[45] P. Saraiya, C. North, and K. Duca. An insight-based methodology for evaluating bioinformatics visualizations. *IEEE transactions on visualization and computer graphics*, 11(4):443–456, 2005.

[46] J. Sheidin, J. Lanir, C. Conati, D. Toker, and T. Kuflik. The effect of user characteristics in time series visualizations. In *Proceedings of the 25th International Conference on Intelligent User Interfaces*, pp. 380–389, 2020.

[47] B. Shneiderman and M. Wattenberg. Ordered treemap layouts. In *IEEE Symposium on Information Visualization, 2001. INFOVIS 2001.*, pp. 73–78. IEEE, 2001.

[48] J. Stasko, R. Catrambone, M. Guzdial, and K. McDonald. An evaluation of space-filling information visualizations for depicting hierarchical structures. *International Journal of Human-Computer Studies*, 53(5):663–694, 2000. doi: 10.1006/ijhc.2000.0420

[49] J. Stasko and E. Zhang. Focus+ context display and navigation techniques for enhancing radial, space-filling hierarchy visualizations. In *IEEE Symposium on Information Visualization 2000. INFOVIS 2000. Proceedings*, pp. 57–65. IEEE, 2000.

[50] Y. Wang, S. Romano, V. Nguyen, J. Bailey, X. Ma, and S.-T. Xia. Unbiased multivariate correlation analysis. In *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 31, 2017.

[51] S. Wehrend and C. Lewis. A problem-oriented classification of visualization techniques. In *Proceedings of the First IEEE Conference on Visualization: Visualization '90*, pp. 139–143, 1990. doi: 10.1109/VISUAL.1990.146375

[52] C. Xiong, J. Shapiro, J. Hullman, and S. Franconeri. Illusion of causality in visualized data. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):853–862, 2020. doi: 10.1109/TVCG.2019.2934399

[53] J. S. Yi, Y.-a. Kang, J. T. Stasko, and J. A. Jacko. Understanding and characterizing insights: how do people gain insights using information visualization? In *Proceedings of the 2008 Workshop on BEyond time and errors: novel evaLuation methods for Information Visualization*, pp. 1–6, 2008.

[54] C. Ziemkiewicz, R. J. Crouser, A. R. Yauilla, S. L. Su, W. Ribarsky, and R. Chang. How locus of control influences compatibility with visualization style. In *2011 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pp. 81–90. IEEE, 2011.

[55] C. Ziemkiewicz and R. Kosara. Preconceptions and individual differences in understanding visual metaphors. In *Computer Graphics Forum*, vol. 28, pp. 911–918. Wiley Online Library, 2009.

[56] C. Ziemkiewicz, A. Ottley, R. J. Crouser, A. R. Yauilla, S. L. Su, W. Ribarsky, and R. Chang. How visualization layout relates to lo-

cus of control and other personality factors. *IEEE transactions on visualization and computer graphics*, 19(7):1109–1121, 2012.