FAVis: Visual Analytics of Factor Analysis for Psychological Research

Yikai Lu*   Chaoli Wang†

University of Notre Dame

Figure 1: FAVis (https://luyikei.github.io/favis/) visualizes a three-factor common factor model of the Motivational State Questionnaire dataset, selecting variables with the absolute values of their factor loadings larger than 0.3. (A) Matrix view shows the most direct visualization of a factor loadings matrix; (B) Network view visualizes cross-loadings most effectively and reduces redundant information; (C) Parallel-coordinates view shows factor loadings for each variable/factor without information loss and allows for selecting variables/factors in a range of values; (D) Tag view shows the summary of the relevance of tags for each factor by counting the frequency of tags annotated for variables based on a theory; (E) Word cloud view helps interpret factors by correlating fonts with the values of factor loadings; (F) Threshold view controls the number of factor loadings shown in the matrix and network views; (G) Factor correlation view shows the network of factor correlations; (H) Top bar allows users to filter variables by their prefix and limit the number of variables/factors shown in the application.

ABSTRACT

Psychological research often involves understanding psychological constructs through conducting factor analysis on data collected by a questionnaire, which can comprise hundreds of questions. Without interactive systems for interpreting factor models, researchers are frequently exposed to subjectivity, potentially leading to misinterpretations or overlooked crucial information. This paper introduces FAVis, a novel interactive visualization tool designed to aid researchers in interpreting and evaluating factor analysis results. FAVis enhances the understanding of relationships between variables and factors by supporting multiple views for visualizing factor loadings and correlations, allowing users to analyze information from various perspectives. The primary feature of FAVis is to enable users to set optimal thresholds for factor loadings to balance clarity and information retention. FAVis also allows users to assign tags to variables, enhancing the understanding of factors by linking them to their associated psychological constructs. Our user study demonstrates the utility of FAVis in various tasks.

1 INTRODUCTION

Psychology studies human characteristics and behavior. Unlike disciplines such as physics, where phenomena can be directly measured and observed, psychology often faces challenges in directly measuring aspects like personality, emotion, or intelligence [15, 25, 27, 29, 30]. These qualities are termed *psychological constructs* [5]. The primary goal of research in psychology is often to study these constructs. Although psychological constructs are unobservable, they can be inferred from observed variables. In statistical terms, psychological constructs are construed as *latent variables* [19]. Latent variable models, including the *common factor model* (CFM) [19, 28], are employed for statistical inference.

Mathematically, let $\mathbf{x} = (x_1, \ldots, x_p)$ be a column vector representing p observed variables, and $\mathbf{y} = (y_1, \ldots, y_q)$ be a column vector denoting q common factors (or latent variables in broader terms), where typically $q \ll p$. The CFM is described as follows

$$\mathbf{y} \sim N_q(0, \mathbf{\Phi}), \qquad (1)$$

*e-mail: ylu22@nd.edu
†e-mail: chaoli.wang@nd.edu

where $\mathbf{\Phi}$ is the $q \times q$ factor correlation matrix. Additionally,

$$\mathbf{x} = \mathbf{\Lambda}\mathbf{y} + \boldsymbol{\mu} + \boldsymbol{\epsilon}, \qquad (2)$$

where $\mathbf{\Lambda}$ is a $p \times q$ matrix of factor loadings, $\boldsymbol{\mu}$ accounts for a non-zero mean of $\mathbf{x}$, and $\boldsymbol{\epsilon}$, the noise of the model, follows $\boldsymbol{\epsilon} \sim N_p(0, \mathbf{\Psi})$, with $\mathbf{\Psi}$ being a $p \times p$ diagonal matrix representing specific variances. The CFM is viewed as a method for dimensionality reduction, with $\mathbf{y}$ explaining most of the covariance in $\mathbf{x}$. From Equations 1 and 2, it follows that

$$\mathbf{x} \sim N_p(\boldsymbol{\mu}, \mathbf{\Lambda}\mathbf{\Phi}\mathbf{\Lambda}^T + \mathbf{\Psi}). \qquad (3)$$

To enhance the interpretability of the parameters, including $\mathbf{\Lambda}$, $\mathbf{\Phi}$, and $\mathbf{\Psi}$, they can be refined through factor rotation by putting some sparsity criteria on $\mathbf{\Lambda}$ without altering the model.

Factor analysis (FA) is the method used to estimate the CFM. This paper focuses on *exploratory factor analysis* (EFA) [19], a technique where no predefined structure is imposed on the factor loadings. There are variants of FA or CFM. For instance, probabilistic principal component analysis (PCA) [28] assumes equal specific variances ($\psi_i = \sigma^2$). In this scenario, the maximum likelihood estimators $\mathbf{\Lambda}_{\text{ML}}$ and σ_{ML}^2 for the isotropic error model correspond to PCA. Consequently, our visualization tool can also assist in interpreting PCA results.

Psychologists utilize the CFM to identify psychological constructs (i.e., estimating $\mathbf{\Lambda}$ from data). Typically, researchers create observed variables like test questions in a relevant domain and apply FA to the data. They often standardize $\mathbf{x}$ to ensure uniform scaling. Then, they examine the factor loadings matrix $\mathbf{\Lambda}$ to determine which dimensions of $\mathbf{y}$ correspond to the psychological constructs.

For instance, a study measured two soldiers' psychological constructs: fear response and optimism [19]. The researchers developed 14 seven-point adjectival rating scales for measurement. After data collection, standardization, and FA, they obtained an initial factor loadings matrix $\mathbf{\Lambda}_1$ with four common factors and an identity matrix for $\mathbf{\Phi}$. However, after they attempted to interpret $\mathbf{\Lambda}_1$ by reading its values, the last two factors in $\mathbf{\Lambda}_1$ were labeled as "difficult to interpret," leading them to apply varimax rotation to the first two factors, resulting in $\mathbf{\Lambda}_2$. This process can be subjective and challenging, especially with a large p. The example illustrates the existence of "difficult to interpret" factors, where simply reading numbers for interpretation is demanding.

Evaluating the quality of a CFM based on its results is crucial. In psychology, an ideal factor loading structure is known as a "simple structure" [19], characterized by minimal cross-loadings. *Cross-loadings* refer to instances where a variable has multiple large (more than two) factor loadings [4, 26]. Suppose λ_{ik} is the factor loading for variable i and factor k. λ_{ik} and λ_{il} are considered as cross-loadings if both absolute values are large. Additionally, this paper introduces a concept deemed more problematic than cross-loadings: redundant-loadings. *Redundant-loadings* occur when multiple variables have large loadings on multiple factors. Specifically, λ_{ik}, λ_{il}, λ_{jk}, and λ_{jl} are considered redundant-loadings if all of them are large. This makes interpreting factors more difficult, as variables i and j do not contribute to interpreting factors k and l due to a lack of differentiation. Then, how do we decide if a factor loading is *large*? Traditionally, a threshold is used to interpret factor loadings, and various standards or methods are proposed to choose an optimal value [26, 11, 17]. The subjective nature of this process influences the interpretation of the CFM.

In this paper, we design FAVis, a visualization tool to aid in interpreting and evaluating FA results. FAVis features multiple interactive views, including matrix, network, and parallel coordinates, to help users understand different aspects of the model parameters. Our tool could reveal insights from "difficult to interpret" factors, avoiding omitting important clues in the model and data. This tool was implemented using D3 [1] and Vue.js [2].

2 RELATED WORKS

Visualizing high-dimensional data. Our work intersects with the visualization of high-dimensional data, mainly because FA is a dimensionality reduction method that extracts latent variables that explain observed variables. This aligns with numerous studies in visual analytics research focusing on high-dimensional data [16, 23]. Various visualization tools and methods, especially those utilizing PCA, have been developed, such as iPCA [13] and BaVA [12]. These approaches, however, differ in their treatment of PCA and dimensionality reduction. For instance, iPCA employs PCA primarily as a data preprocessing technique, with its main focus being the visualization of the original data points in a projection view. This contrasts with our FAVis, where the emphasis is on interpreting the model and factor loadings themselves, treated as input data. While some methods aim to improve model parameters through visual analytics [12, 9], they do not typically focus on the visualization of these parameters for interpretative purposes. FAVis, therefore, occupies a unique space in dimensionality reduction by concentrating on the interpretation of model parameters.

Visualizing psychological data. Visualization in psychological data, particularly in cognitive psychology and neuroscience, is well-developed. Existing research primarily focuses on EEG [14, 24] and fMRI data [21, 8]. For instance, Ji et al. [14] developed an interactive visualization for dynamic EEG coherence networks. Similarly, a combination of visual analytics and scientific visualization has been applied in fMRI research [7, 6].

However, psychological data in other domains, such as personality, educational, and developmental psychology, often rely on behavioral measures like tests, psychological scales, and surveys. FA is a standard tool for analyzing such data. One of these few examples is qgraph [10], which is an R package to visualize the CFM using graphs. However, qgraph only provides R functions to generate such visualization, thus lacking interactivity. Our FAVis aims to bridge this gap in the visual analytics literature, providing an interactive tool for these other psychology domains.

3 DESIGN GOALS

We discussed with two experts in quantitative psychology and conducted a literature review (including Section 2 and interactive visual analytics tools [3, 20]) to derive the following design goals.

- **G1 – Facilitate understanding of the associations between variables and factors.** The factor loadings matrix $\mathbf{\Lambda}$ describes how much a factor is related to a variable (as in Equation 2), which can be hundreds or thousands of numbers depending on the number of variables and factors. We designed several means to visualize the matrix to reduce subjectivity in interpreting the factor loadings matrix by only highlighting relevant elements according to the threshold.

- **G2 – Provide useful between-factor information that affects the quality of a CFM.** As it is desirable that a CFM has a simple structure, the tool should support highlighting important cross-loadings. Also, an oblique rotation could cause factors to be correlated (i.e., the factor correlation matrix $\mathbf{\Phi}$ is not identical). Thus, the design should ensure that any crucial between-factor association will not be overlooked.

- **G3 – Identify a threshold for factor loadings that balances interpretability and information loss.** Traditionally, a threshold is often used to facilitate the interpretation of the factor loadings matrix. With visualization, the design should lead users to determine an optimal threshold specific to a mode or dataset.

- **G4 – Filter and sort variables and factors based on desirable characteristics.** When interpreting factors, a researcher might want to know which variables have large loadings on a specific factor and vice versa.

- **G5 – Provide a mechanism to relate a theory to a CFM.** Researchers often have a theory of how variables are related to psychological constructs before they conduct a study to collect data and construct a model. The design should be able to connect variables to presumed psychological constructs and effectively visualize the degree to which factors represent the psychological constructs to reduce the subjectivity of interpreting factors in terms of a theory.

4 FAVIS SYSTEM DESIGN

As shown in Figure 1, FAVis has multiple views: matrix, network, parallel-coordinates views for variables and factors, as well as tag, word cloud, threshold, and factor correlation views. Scaling and dragging are supported for the objects in matrix, network, and factor correlation views.

Matrix view (Figure 1A) shows the factor loadings matrix Λ as a heatmap and as a matrix. Each rectangle represents a factor loading corresponding to its variable and factor (**G1**, **G2**). Each axis represents variables or factors depending on its *transpose* setting. The color of rectangles is based on the values of factor loadings. Only the absolute values of factor loadings larger than the threshold value will be shown here. Displaying absolute or actual values is determined by its *absolute values* settings. If users decide to show absolute values of factor loadings, only the degree of the associations between variables and factors will be displayed. On the other hand, if actual values are shown, positive values are displayed as light blue, and negative values are displayed as light red, signifying the direction of the correlation between its variable and factor. Variables and factors can be sorted by clicking a variable or factor of interest, which helps users identify the most relevant variables/factors to their corresponding factors/variables. The sorting will also be reflected on the parallel-coordinates view.

Network view (Figure 1B) displays a graph of variables based on factor loadings (**G1**, **G2**) using the force-directed layout. Each node represents a variable, and an edge is formed if two variables have factor loadings in any factor larger than the threshold. This visualization is generated from the factor loadings matrix to highlight cross-loadings and eliminate redundant-loadings. Suppose λ_{ik} is the factor loading for variable i and factor k, and α is the threshold value. An edge between variable i and variable j is formed if $|\lambda_{ik}| > \alpha$ and $|\lambda_{jk}| > \alpha$ for any k. Multiple factors could satisfy this condition for variables i and j. In such a case, these factor loadings are considered as redundant-loadings. Depending on a goal (**G1** or **G2**), users can choose between two modes of visualization in the network view. For **G1**, the color of nodes and edges are determined based on the largest factor loading. Specifically, the color of nodes will be determined by choosing the factor with the largest factor loading, and the color of edges will be determined using the factor with the largest average factor loadings of variables i and j. Since the network view can further reduce the amount of information by eliminating redundant-loadings, it offers a more summarized visualization than the matrix view. For **G2**, both nodes and edges are colored based on the number of cross-loadings (which includes redundant-loadings). This mode can easily find an optimal threshold value as the number of cross-loadings is vividly visualized.

Parallel-coordinates views (Figure 1C) show the line plots of factor loadings. One displays variables on the *x*-axis and factors on the *y*-axis, and the other does the opposite. These views select variables/factors with the largest factor loadings on a specific factor/variable (**G4**). Users can create filters by dragging a mouse on a *y*-axis, highlighting a subset of variables/factors in all the other views. Since these views are not affected by the threshold, they are suitable for examining a specific variable or factor without information loss. To reduce visual clutter while maintaining task effectiveness, we only show one in every five *y*-axes when there are more than 20. The hidden ones will be shown when hovering over them.

Tag view (Figure 1D) links a theory to the resultant model (**G5**) to enhance understanding of a CFM. Users can assign a tag representing a psychological construct of interest to a presumed related variable. They can add tags within the "Edit Tag" interface (Figure 1Da). This is done by entering the desired tag name into the input box and clicking the "Add Tag" button. To facilitate this process, users can automatically add tags using a codebook file (a dictionary specifying variables and their associated list of tags). Tags associated with a selected variable are highlighted by increasing their opacity. Users can also click on any tag to toggle its assignment to the selected variable. Then, the tag view displays these tags for each factor through a horizontal stacked bar chart (Figure 1Db). Each rectangle within this chart symbolizes a tag, with its width proportional to the count of associated variables exceeding the threshold. The design ensures only pertinent tags are shown, with each rectangle's size proportional to the tag's relevance to a factor, aiding in factor interpretation. When a rectangle's width is insufficient to display a tag's text, hovering over it fully activates a tooltip that reveals the tag's name and value (the rectangle's width). When the "Normalized" checkbox on the top bar of the tag view is checked, instead of the raw count, the width of a rectangle will become proportional to the total counts (Figure 2b), which is more helpful when users only want to interpret the meaning of a factor, instead of exploring potential values of the threshold.
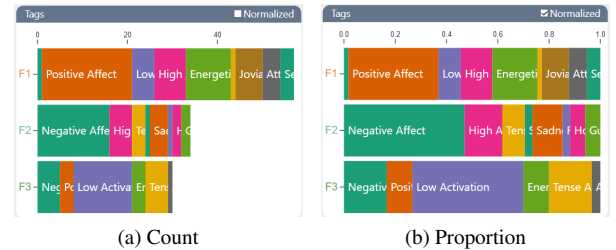

(a) Count   (b) Proportion

Figure 2: Comparison of tag views: (a) showing counts and (b) showing proportions of variables with associated tags. (a) is beneficial for comparing the effectiveness of different factors in measuring theorized psychological constructs, while (b) aids in interpreting the meaning of individual factors in terms of these constructs.

Word cloud view (Figure 1E) is the most suitable for interpreting the meaning of a factor without relying on tags (**G1**). For a selected factor, each text element corresponds to the name of a variable, with its size and color determined by the factor loading. The color reflects the factor loading value, following the same color scale as the matrix view. Similarly, the font size is scaled linearly according to the absolute value of the factor loading. Therefore, users can interpret the meaning of a factor just by concentrating on a set of highlighted tags, which is particularly helpful when dealing with numerous variables, and the matrix view cannot display every variable simultaneously.

Threshold view (Figure 1F) plots the cumulative distribution of the factor loadings and provides a slider to adjust the threshold used in the matrix and network views (**G3**). The threshold determines information loss to improve the interpretability of the model, as factor loadings below the threshold will not be shown in the matrix or network views. Information loss is depicted as the horizontal line in the plot, which corresponds to the portion of the factor loadings that will be discarded in the matrix and network views. The threshold value is shown as the vertical line. As users update the threshold value using the slider, the matrix and network views will also be dynamically updated. Users are expected to change the threshold value, which minimizes the number of cross-loadings, preserving most of the factor loadings (**G3**). A good model is expected to have few cross-loadings even with a lower threshold value, as factors are well separated in such a case (**G2**).

Factor correlation view (Figure 1G) shows the correlations among factors using a graph with the force-directed layout (**G2**). Each node represents a factor, and each edge represents the correlation between two factors. This view will not display any edge if no factors are correlated, which can happen when we use PCA or orthogonal rotations.

Additional interactions and workflow. In the top bar (Figure 1H), there is a line input for searching variables by their names and filtering out these variables in the matrix, network, and parallel-coordinates views (**G4**). We can also limit the maximum number of variables or factors displayed in these views. In FAVis, many interactions among views are implemented (**G4**). For example, suppose users click on a variable/factor on the matrix view. In that case, the factor loadings will be sorted by descending order, helping them identify the most relevant factors/variables, which will also be reflected in the parallel-coordinates views. Additionally, if users click a rectangle within the matrix view, the corresponding row and column are highlighted in red. This action also triggers highlights in the network and parallel-coordinates views, where irrelevant visual elements become less opaque, focusing attention on the selected variable and factor. Similarly, clicking a node in the network view highlights the selected variable consistently across other views. On hovering a rectangle in the matrix view or a node in the network view, a tooltip will appear, which contains relevant information about its corresponding variable and/or factor, such as its value of the factor loading, codebook information (e.g., question text), and associated tags to facilitate understanding of the results.

5 EVALUATION

We conducted a user study to evaluate the usability of FAVis by following the procedures used in past studies [3, 20]. Five domain experts (one professor, three quantitative psychology graduate students, and one statistics graduate student) familiar with EFA and/or PCA were selected to evaluate FAVis. The participants received no compensation for taking part in the study.

Data (CFMs). Participants (denoted as $E1, E2, \ldots, E5$) were given two CFMs from two datasets during the study. The first CFM and dataset were a 12-factor CFM derived from a questionnaire for suicidal thought and behavior [18], which were used for a tutorial on FAVis. The second CFM and dataset were a three-factor CFM derived from the Motivational State Questionnaire dataset [22], which consists of items taken from other scales [29, 27, 15].

Procedure. The study was conducted in person or via Zoom session (one participant). The participants were asked to (1) answer pre-study questions, (2) run FAVis in a browser and receive a tutorial using the first model, (3) reload the page to evaluate using the second model, and (4) answer post-study questions, including usability measures rated one to five. The participants were asked to conduct specific tasks related to the design goals. Since they did not have a theory for the dataset, they were first asked to evaluate the CFM without tags, and then we showed them pre-labeled results to re-evaluate the same CFM using tags. Each study session was recorded and lasted about 30 to 60 minutes.

Results. Overall, all the participants rated the basic usability measures highly: an average of 4.1 for *Easy to Understand*, 4.4 for *Easy to Use*, 5.0 for *Enjoyable to Use*, and 5.0 for *Would use in the future*. The effectiveness of FAVis is also demonstrated by the task-specific measures: an average of 4.8 for *Helps you understand the meaning of each factor*, 5.0 for *Helps you understand factor relationships*, 4.4 for *Helps you identify the optimal threshold*, and 5.0 for *Helps you relate a theory to a model*. These results suggest that although FAVis is not necessarily the easiest to learn, as its average rating is the lowest, the overall satisfaction is very high. It is very useful to conduct these tasks.

We mainly conducted three tasks: interpreting the meanings of factors with/without tags, explaining relationships among factors (e.g., cross-loadings and correlations), and identifying the best

items for each factor. As demonstrated by the high average ratings of task-specific measures, all participants completed these tasks, accurately describing the meanings of factors and their relationships. However, there are differences among participants in approaching some of the tasks.

When identifying cross-loadings, *E3*, *E4*, and *E5* kept adjusting the threshold until they found cross-loadings in the network view. In contrast, *E1* used filters on the parallel-coordinates view to emphasize cross-loadings in other views. Both approaches are generally considered efficient for this task. Nevertheless, *E2* took a different route, setting the threshold to zero to reveal all factor loadings on the matrix view, which was unexpected. This could be because some researchers are used to reading matrix values to the point where they find it challenging to unlearn their habitual methods, highlighting the possible difficulty in learning to use FAVis effectively. Regarding the parallel-coordinates views, *E1*, who used the views for this task, said, "It has a lot of potential but it is also the one that's got the highest ceiling for use." *E1* also used the views for identifying the best items for each factor, saying, "I think this tool is very powerful." On the other hand, *E3* did not use them because, as he said, "the views looked like spaghetti." This suggests that while the parallel-coordinates views have the range filtering feature, a more straightforward interface for range filtering might be preferable to some people. When interpreting the meaning of each factor, *E5* especially preferred using the word cloud view (she mentioned that the word cloud view is instrumental when the names of variables are informative; otherwise, it is less useful), while others tended to use the matrix view.

The usability measures show that all the participants enjoyed using the tool and liked different aspects, including the network view, thresholding, matrix view and sorting feature, and word cloud view. For example, *E1* and *E3* particularly liked the network view: *E1* mentioned that he enjoyed watching the changes and movements of graphics when adjusting the threshold, and *E3* said he enjoyed watching clusters and cross-loadings appearing in the network view as he changes the threshold.

6 CONCLUSION, LIMITATIONS, AND FUTURE WORK

This study introduces FAVis, a visual analytics tool designed to help psychological researchers explore and understand CFMs. This research was motivated to fill a gap in the visualization literature, as few focus on visualizing behavioral data to identify psychological constructs. Having multiple views, FAVis can help users perform various tasks such as interpreting factors, understanding between-factor information, and relating a theory to a model. Our user study also revealed the tool's effectiveness in completing the different tasks, which led to high satisfaction. The multiple views allowed users with varying levels of familiarity with visual analytics and factor analysis to accomplish tasks using their preferred methods.

This study has several limitations. First, some users noted that the factor correlation view could benefit from supporting different visualization modalities. Currently, the view uses a force-directed graph, effective when factor correlations have more significant variability, leading to distinct graph shapes. However, the force-directed graph becomes less informative if the correlations exhibit less variability. In such a case, a heatmap representation would be more beneficial, as it provides a better interface to find a value associated with a specific pair of factors. Second, the tool's initial layout and ordering of the rows and columns in a matrix are fixed. FAVis should include a method to determine the optimal ordering for the rows and columns in a factor loadings matrix. Third, the tool cannot compare between different CFMs for the same dataset. Finally, the results from our user study primarily provide usability data and do not offer strong evidence of efficacy. Comparing this tool against a baseline method could be useful to demonstrate its efficacy. In the future, we aim to address these limitations to enhance FAVis.

REFERENCES

[1] D3 - the JavaScript library for bespoke data visualization. https://d3js.org. Accessed: 2024-04-30. 2

[2] Vue.js - the progressive JavaScript framework. https://vuejs.org. Accessed: 2024-04-30. 2

[3] R. G. Bayrak, N. Hoang, C. B. Hansen, C. Chang, and M. Berger. PRAGMA: Interactively constructing functional brain parcellations. In *Proceedings of IEEE VIS Short Papers*, pp. 46–50, 2020. doi: 10.1109/VIS47514.2020.00016 2, 4

[4] A. Costello and J. Osborne. Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research, and Evaluation*, 10(1), 2019. doi: 10.7275/jyj1-4868 2

[5] L. M. Crocker and J. Algina. *Introduction to Classical and Modern Test Theory*. Cengage Learning, 2008. 1

[6] M. de Ridder, Y. Jung, R. Huang, J. Kim, and D. D. Feng. Exploration of virtual and augmented reality for visual analytics and 3D volume rendering of functional magnetic resonance imaging (fMRI) data. In *Proceedings of Internation Symposium on Big Data Visual Analytics*, pp. 1–8, 2015. doi: 10.1109/BDVA.2015.7314293 2

[7] M. de Ridder, K. Klein, and J. Kim. TemporalTracks: Visual analytics for exploration of 4D fMRI time-series coactivation. In *Proceedings of Computer Graphics International*, pp. 1–6, 2017. doi: 10.1145/3095140.3095153 2

[8] M. de Ridder, K. Klein, and J. Kim. A review and outlook on visual analytics for uncertainties in functional magnetic resonance imaging. *Brain Informatics*, 5(2):5, 2018. doi: 10.1186/s40708-018-0083-0 2

[9] A. Endert, C. Han, D. Maiti, L. House, S. Leman, and C. North. Observation-level interaction with statistical models for visual analytics. In *Proceedings of IEEE Conference on Visual Analytics Science and Technology*, pp. 121–130, 2011. doi: 10.1109/VAST.2011.6102449 2

[10] S. Epskamp, A. O. J. Cramer, L. J. Waldorp, V. D. Schmittmann, and D. Borsboom. qgraph: Network visualizations of relationships in psychometric data. *Journal of Statistical Software*, 48(4):1–18, 2012. 2

[11] J. F. Hair, B. W. C., B. J. Babin, and R. E. Anderson. *Multivariate Data Analysis*. Pearson Prentice Hall, 2021. 2

[12] L. House, S. Leman, and C. Han. Bayesian visual analytics: BaVA. *Statistical Analysis and Data Mining*, 8(1):1–13, 2015. doi: 10.1002/sam.11253 2

[13] D. H. Jeong, C. Ziemkiewicz, B. Fisher, W. Ribarsky, and R. Chang. iPCA: An interactive system for PCA-based visual analytics. *Computer Graphics Forum*, 28(3):767–774, 2009. doi: 10.1111/j.1467-8659.2009.01475.x 2

[14] C. Ji, J. J. van de Gronde, N. M. Maurits, and J. B. T. M. Roerdink. Visual exploration of dynamic multichannel EEG coherence networks. *Computer Graphics Forum*, 38(1):507–520, 2019. doi: 10.1111/cgf.13588 2

[15] R. J. Larsen and E. Diener. Promises and problems with the circumplex model of emotion. In M. S. Clark, ed., *Emotion*, pp. 25–59. Sage Publications, Inc., 1992. 1, 4

[16] S. Liu, D. Maljovec, B. Wang, P.-T. Bremer, and V. Pascucci. Visualizing high-dimensional data: Advances in the past decade. *IEEE Transactions on Visualization and Computer Graphics*, 23(3):1249–1268, 2017. doi: 10.1109/TVCG.2016.2640960 2

[17] R. C. MacCallum, K. F. Widaman, S. Zhang, and S. Hong. Sample size in factor analysis: Psychological methods. *Psychological Methods*, 4(1):84–99, 1999. doi: 10.1037/1082-989X.4.1.84 2

[18] K. McClure, R. Jacobucci, and B. A. Ammerman. Initial calibration of a multidimensional assessment of suicide risk in time-intensive longitudinal designs: Item-bank development and calibration. In *ABCT Annual Meeting Posters*, 2023. 4

[19] S. A. Mulaik. *Foundations of Factor Analysis*. Chapman and Hall/CRC, 2009. 1, 2

[20] D. Munechika, Z. J. Wang, J. Reidy, J. Rubin, K. Gade, K. Kenthapadi, and D. H. Chau. Visual Auditor: Interactive visualization for detection and summarization of model biases. In *Proceedings of IEEE VIS Short Papers*, pp. 45–49, 2022. doi: 10.1109/VIS54862.2022.00018 2, 4

[21] F. Rasheed, D. Jonsson, E. Nilsson, T. B. Masood, and I. Hotz. Subject-specific brain activity analysis in fMRI data using merge trees. In *Proceedings of IEEE Workshop on Topological Data Analysis and Visualization*, pp. 113–123, 2022. doi: 10.1109/TopoInVis57755.2022.00018 2

[22] W. Revelle and K. J. Anderson. Personality, motivation and cognitive performance: Final report to the Army Research Institute on Contract MDA 903–93-K-0008. Technical report, Northwestern University, 1998. 4

[23] D. Sacha, L. Zhang, M. Sedlmair, J. A. Lee, J. Peltonen, D. Weiskopf, S. C. North, and D. A. Keim. Visual interaction with dimensionality reduction: A structured literature analysis. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):241–250, 2017. doi: 10.1109/TVCG.2016.2598495 2

[24] D. Slayback, S. Abdali, J. Brooks, W. D. Hairston, and P. Groves. Novel methods for EEG visualization and virtualization. In *Proceedings of IEEE International Symposium on Circuits and Systems*, pp. 1–5, 2018. doi: 10.1109/ISCAS.2018.8351726 2

[25] C. J. Soto and O. P. John. The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of Personality and Social Psychology*, 113(1):117–143, 2017. doi: 10.1037/pspp0000096 1

[26] B. G. Tabachnick and L. S. Fidell. *Using Multivariate Statistics*. Pearson, 2007. 2

[27] R. E. Thayer. *The Biopsychology of Mood and Arousal*. Oxford University Press, 1989. 1, 4

[28] M. E. Tipping and C. M. Bishop. Probabilistic principal component analysis. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 61(3):611–622, 1999. doi: 10.1111/1467-9868.00196 1, 2

[29] D. Watson, L. A. Clark, and A. Tellegen. Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, 54(6):1063–1070, 1988. doi: 10.1037/0022-3514.54.6.1063 1, 4

[30] J. O. Willis, R. Dumont, and A. S. Kaufman. Factor-analytic models of intelligence. In R. J. Sternberg and S. B. Kaufman, eds., *The Cambridge Handbook of Intelligence*, pp. 39–57. Cambridge University Press, 2011. doi: 10.1017/CBO9780511977244.004 1