Exploring How Personality Models Information Visualization Preferences

Conference Paper · October 2020

DOI: 10.1109/VIS47514.2020.00047

Tomás Alves
University of Lisbon

Bárbara Ramalho
Technical University of Lisbon

Joana Henriques-Calado
University of Lisbon

Daniel Gonçalves
Inesc-ID

Exploring How Personality Models Information Visualization Preferences

Tomás Alves,[*] Bárbara Ramalho,[†] Daniel Gonçalves,[‡] and Sandra Gama[§]
INESC-ID and Instituto Superior Técnico,
University of Lisbon, Lisbon, Portugal

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

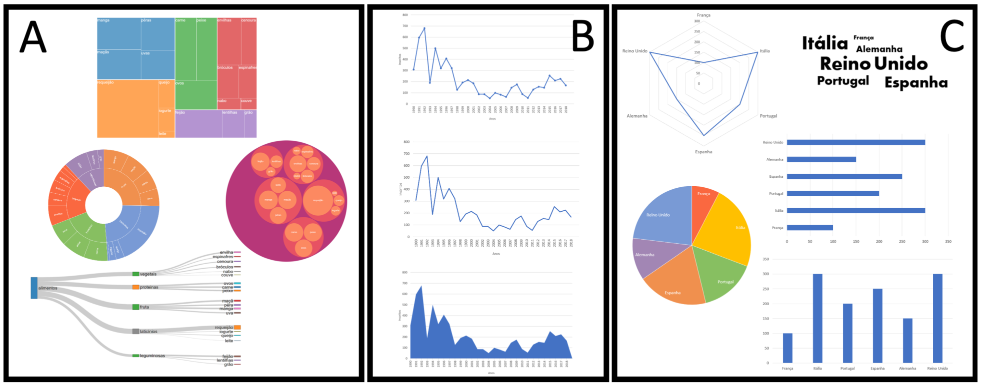

Figure 1: The three different information visualization contexts that we address in our study. **(A)** the hierarchy context which is studied with the treemap, circular packing, sunburst, and Sankey diagram idioms. **(B)** evolution over time context, where we apply line charts with and without points, and an area chart. Finally, **(C)** comparison context, which includes the radar chart, word cloud, horizontal and vertical bar charts, and pie chart idioms.

Abstract

Recent research on information visualization has shown how individual differences act as a mediator on how users interact with visualization systems. We focus our exploratory study on whether personality has an effect on user preferences regarding idioms used for hierarchy, evolution over time, and comparison contexts. Specifically, we leverage all personality variables from the Five-Factor Model and the three dimensions from Locus of Control (LoC) with correlation and clustering approaches. The correlation-based method suggested that Neuroticism, Openness to Experience, Agreeableness, several facets from each trait, and the External dimensions from LoC mediate how much individuals prefer certain idioms. In addition, our results from the cluster-based analysis showed that Neuroticism, Extraversion, Conscientiousness, and all dimensions from LoC have an effect on preferences for idioms in hierarchy and evolution contexts. Our results support the incorporation of in-depth personality synergies with InfoVis into the design pipeline of visualization systems.

Index Terms: Human-centered computing—Human computer interaction (HCI)—HCI design and evaluation methods—User studies; Human-centered computing—Visualization—Visualization design and evaluation methods

1 Introduction

Individual differences have shown promise as an adaptation metric of information visualization systems to tackle the limitations of one-size-fits-all approaches [5, 15, 24]. The inclusion of these factors empowers developers with guidelines on how individual characteristics impact human-computer interaction. Among several psychological constructs that differentiate individuals such as cognitive bias or abilities, personality stands as an established strong mediator given its stability throughout adulthood [21]. Compared to other well studied individual cognitive traits (e.g. spatial ability and visual working memory), promising results regarding the relationship between personality traits and information visualization are few [20]. In order to bridge this gap, we focus on two of the most extensively studied personality models in our research field: the Five-Factor Model (FFM) [6] and the Locus of Control (LoC) [16].

Although performance metrics such as speed or accuracy are important to address while users perform tasks in information visualization systems (e.g. [25, 26]), we believe that there is a lack of findings regarding how personality affects user preferences for information visualization techniques. Weighting how personality has an effect on user preferences, our study focuses on whether **personality promotes user preferences regarding idioms used for hierarchy, evolution over time, and comparison contexts**, as these contexts have been frequently applied in state-of-the-art research [4, 20, 31]. In particular, we address all traits and their facets from the FFM, and the three dimensions from LoC to provide an in-depth analysis. We take two distinct approaches to study this relationship: (i) correlation-based analysis – where we investigate whether a personality variable had correlations with the user preference regarding an idiom – and (ii) cluster-based analysis – where we aggregate individuals by common characteristics and extract preference patterns from each group. Our preliminary results suggest that personality has an effect on user preferences with both types of analysis.

2 Related Work

The Five-Factor Model consists of five general dimensions to describe personality and 30 subdimensions (facets) (Table 1): (i) Neuroticism distinguishes "the stability of emotions and even-temperedness from negative emotionality" [11]; (ii) Extraversion suggests "a lively approach toward the social and material world" [11]; (iii) Openness to Experience describes "the whole-

[*]e-mail: tomas.alves@tecnico.ulisboa.pt
[†]e-mail: barbara.ramalho@tecnico.ulisboa.pt
[‡]e-mail: daniel.goncalves@inesc-id.pt
[§]e-mail: sandra.gama@tecnico.ulisboa.pt
[¶]e-mail: joanahenriquescalado@gmail.com

Table 1: Traits and their facets of the Five-Factor Model.

Trait	Facets
Neuroticism (N)	Anxiety (N1), Anger (N2), Depression (N3), Self-consciousness (N4), Immoderation (N5), Vulnerability (N6)
Extraversion (E)	Friendliness (E1), Gregariousness (E2), Assertiveness (E3), Activity level (E4), Excitement-seeking (E5), Cheerfulness (E6)
Openness to Experience (O)	Imagination (O1), Artistic interests (O2), Emotionality (O3), Adventurousness (O4), Intellect (O5), Liberalism (O6)
Agreeableness (A)	Trust (A1), Morality (A2), Altruism (A3), Cooperation (A4), Modesty (A5), Sympathy (A6)
Conscientiousness (C)	Self-efficacy (C1), Orderliness (C2), Dutifulness (C3), Achievement-striving (C4), Self-discipline (C5), Cautiousness (C6)

ness and complexity of an individual's psychological and experiential life" [11]; (iv) Agreeableness distinguishes "pro-social and communal orientation toward others from antagonism" [11]; and (v) Conscientiousness suggests "self-use of socially prescribed restraints that facilitate goal completion, following norms and rules, and prioritizing tasks" [11]. Ziemkiewicz and Kosara [35] 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. Furthermore, Ziemkiewicz et al. [36] concluded that neurotic individuals attained high accuracy on hierarchical search tasks. Introverted participants were more accurate in answering the questions posed by the tasks. Other contributions [2, 3, 10, 24, 34] have addressed the traits of Neuroticism and Extraversion. Results have shown how these traits have an effect on task performance metrics such as the time to complete a task [10, 24] and accuracy [24]. Additionally, Neuroticism and Openness to Experience exhibited an effect on the attractiveness and dependability ratings from participants regarding driver state visualization systems [2].

The Locus of Control orientations are described as two different aspects, which are distinguished by different reinforcements. While internal LoC is related with internal reinforcement because the value of an individual is heightened by some event or environment, external LoC is linked to external reinforcement since it addresses how some event or environment yields benefit for the group or culture to which the individual belongs to [14]. Furthermore, external LoC can be differentiated in two types: *Powerful Others* – believe in an ordered world controlled by powerful others – and *Chance* – consider the world as unordered and chaotic [18]. Several studies have shown how LoC is related to search performance across hierarchical [10], time series [31], and item comparison [4] visualization designs, visualization use [34, 36], and behavioural patterns [26]. Although *Internals* are significantly faster than *Externals* when performing procedural tasks (search tasks to locate items) [10], *Externals* are faster and more accurate than *Internals* regarding inferential tasks such as comparing two items [34, 36]. In addition, *Internals* are usually faster than *Externals* in image-based search tasks [3].

Although performance metrics such as speed or accuracy are important to address, there are strong results regarding user preferences in information visualization. Ziemkiewicz et al. [36] focus on Neuroticism, Extraversion, and the LoC, while Lallé and Conati [15] address the latter. In contrast, Toker et al. [32] did not address personality. Nevertheless, research has not found effects for Agreeableness or Conscientiousness [20], and FFM traits' facets have been neglected. In our study, we propose an extension of the state-of-the-art research by including the remaining FFM traits and their facets, which may hinder relationships that are only represented at a finer granularity of personality variables.

3 Data Collection

In order to study ***how personality affects user preferences regarding information visualization techniques***, we started by choosing which contexts we wanted to address (Figure 1): (i) hierarchy, one of the most common in research (e.g. [36]); (ii) evolution over time, giving the importance of time series data analysis [31]; and (iii) comparison, as it is more appropriate to show differences or similarities between values at a fixed granularity [4]. We include a simple and familiar scenario with each context in order to stimulate users to reflect on the implications of using each idiom rather than the complexity of the data. We focused on minimizing the number of channels and marks of each graph and keeping them consistent across contexts, while keeping the same data within a context.

Regarding hierarchy, items are all related to each other by the principle of containment. We opted for a treemap, a circular packing diagram, a sunburst, and a Sankey diagram to display the distribution of food consumed by a household within a month. For evolution over time contexts, we chose line charts with and without points, and area charts. The scenario asked the participant to imagine that the data referred to the number of registrants and participants in a marathon held annually in the United States. Finally, we decided to use radar charts, word clouds, horizontal and vertical bar charts, and pie charts for the comparison context. In particular, the scenario represents the levels of the happiness index among six different countries (France, Italy, Portugal, Spain, Germany, and the United Kingdom).

Participants were recruited through standard convenience sampling procedures including direct contact and through word of mouth. Our final data set comprises 64 participants (30 males, 34 females) between 18 and 60 years old ($M = 24.27; SD = 7.10$). In addition, they were asked whether they were using glasses or contact lenses and the apparatus used while filling in the questionnaire. Neither factor had a significant effect on the experience.

Before the experiment, participants were informed about the experience and invited to agree with a compulsory consent form. They were also informed that they could quit the experiment at any time. We then collected the FFM five personality traits and its 30 facets, and the dimensions of LoC with the Portuguese versions of the Revised NEO Personality Inventory (NEO PI-R) [7, 19] and the IPC scale [17, 28], respectively. Afterwards, participants were presented an online questionnaire which contained a visual example of each idiom grouped by context. Participants were firstly prompt to read the scenario for the respective context and then assess their preference for an idiom by completing a seven-point Likert scale ranging from *Low Preference* (1) to *High Preference* (7). We allowed participants to freely change their ratings until they were satisfied with all ratings in order to avoid the anchoring bias.

4 Correlation-based Analysis

In order to find correlations between personality variables and user preferences, we used the Spearman's correlation method (Table 2), as it is preferable when variables feature heavy-tailed distributions or when outliers are present [8], and it has been shown as an appropriate statistical analysis with Likert scales [23]. Our hypothesis is that a personality dimension from the FFM and/or the LoC is correlated with how participants rated their preference for an idiom. Taking into account the large number of statistical models, we use a Bonferroni correction to counteract the problem of multiple comparisons. Therefore, significant p-values are reported at $\alpha = 0.0001$. Although we did not find any statistical significance, results suggest that, at a trait level, Neuroticism, Openness to Experience, and Agreeableness show weak negative effects with line charts with points ($r_s(64) = -.267, p = .033$), area charts ($r_s(64) = -.29, p = .02$) and sunburst ($r_s(64) = -.285, p = .022$), respectively. In addition, we found that 19 facets showed similarly weak effects. Among these facets, we can observe that facets from Agreeableness pointed towards more effects, followed by Neuroticism and Openness to

Table 2: Significant results from the Spearman's correlation tests.

Personality	Idiom	r_s	p-value
N	Line Chart with Points	-0.267	0.033
N1	Treemap	-0.364	0.003
N3	Line Chart with Points	-0.292	0.019
N4	Line Chart with Points	-0.276	0.027
N6	Line Chart with Points	-0.277	0.027
E1	Sunburst	-0.274	0.029
E4	Line Chart with Points	0.247	0.049
E6	Line Chart without Points	-0.283	0.024
O	Area Chart	-0.290	0.020
O1	Horizontal Bar Chart	-0.269	0.032
O2	Line Chart without Points	-0.251	0.046
O3	Area Chart	-0.320	0.010
O5	Area Chart	-0.340	0.006
O6	Sankey Diagram	-0.268	0.032
A	Sunburst	-0.285	0.022
A2	Sunburst	-0.317	0.011
A2	Treemap	-0.275	0.028
A3	Line Chart without Points	-0.249	0.047
A5	Radar Chart	-0.312	0.012
A5	Circular Packing	0.249	0.047
A6	Sunburst	-0.277	0.027
A6	Radar Chart	-0.263	0.036
C3	Line Chart with Points	0.255	0.042
C5	Line Chart without Points	-0.246	0.050
C6	Vertical Bar Chart	0.274	0.028
Powerful Others	Area Chart	0.320	0.010
Powerful Others	Line Chart without Points	0.313	0.012
Chance	Pie Chart	0.382	0.002

Experience. Although both Extraversion and Conscientiousness did not have an effect, three facets from each of these traits imply an effect. Regarding LoC, both External dimensions showed weak positive correlations. While Powerful Others may have modelled how participants rated both area ($r_s(64) = .32, p = .01$) and line chart without points ($r_s(64) = .313, p = .012$), Chance hinted an effect on ratings for the pie chart ($r_s(64) = .382, p = .002$). Taking into account the idioms, we can see that line charts suggest the largest number of effects related to personality-based user preferences, as most of Neuroticism, its facets, and facets from Conscientiousness suggested correlation effects. At a broader level, evolution over time context idioms indicated the largest number of effects (53.57%), followed by hierarchy (28.57%) and then comparison (17.86%).

5 Cluster-based Analysis

Following the work of Sarsam and Al-Samarraie [30], we applied hierarchical density based clustering [12, 22] to find that the most appropriate number of clusters to work with was three through silhouette and Davies–Bouldin index scores analysis [27] and Ward's cluster method. Then, we used the k-means clustering algorithm [33] to avoid the noise labels that hierarchical density based clustering yields. We started by normalizing our data and allow the algorithm to run 100 iterations with different centroid seeds using Euclidean distance. The final result contained the best output of 100 consecutive runs in terms of inertia. As a follow-up, we conducted an ANOVA to validate whether each personality trait from one cluster differs from the other instances in the other clusters. We found a significant difference ($p < .05$) in between the three clusters regarding Neuroticism, Extraversion, Concientiousness, all dimensions from Locus of Control, and 18 personality facets out of 30. These results show that all clusters have participants that differ among themselves in the aforementioned personality variables. Table 3 depicts the means and standard deviation values for all personality traits of the FFM and dimensions from the LoC. The first cluster ($N = 35$) notably has participants with the **highest levels of Conscientiousness and Internal dimension** across clusters. It also includes people with the **lowest values on Neuroticism and the External dimensions**. The remaining traits of Extraversion and Agreeableness show medium values, while Openness to Experience presents low levels. In contrast, the second cluster ($N = 11$) shows the **lowest values for Conscientiousness**. In addition, it features participants with

Table 3: Results of the K-means clustering algorithm for each personality trait and dimension.

Personality Variable	Cluster 1 M	Cluster 1 SD	Cluster 2 M	Cluster 2 SD	Cluster 3 M	Cluster 3 SD
Neuroticism	84.23	20.23	97.18	16.54	**124.22**	**18.10**
Extraversion	112.00	19.02	**122.00**	**13.82**	92.94	16.11
Openness to Experience	121.34	19.43	**136.09**	**23.70**	121.89	15.91
Agreeableness	125.40	15.70	129.09	21.42	123.67	16.84
Conscientiousness	**140.74**	**15.85**	96.73	20.03	105.67	18.24
Internal	**33.00**	**5.89**	32.82	4.31	29.33	6.00
Powerful Others	15.31	6.54	16.45	6.64	**19.39**	**5.25**
Chance	15.97	5.87	19.82	5.84	**20.17**	**5.84**

Table 4: Context and preferred idioms with their frequency on top rules for each cluster.

Context	Cluster 1	Cluster 2	Cluster 3
Evolution	Sunburst (50%)	Sunburst (76%)	Treemap (37%)
Hierarchy	Line Chart w/ Points (100%)	Line Chart w/out Points (70%)	Line Chart w/out Points (74%)
Comparison	Horizontal Bar Chart (71%)	Horizontal Bar Chart (71%)	Horizontal Bar Chart (71%)

the **highest values of Extraversion and Agreeableness**, while the remaining personality variables show medium values among the clusters. Finally, the third cluster ($N = 18$) includes participants with the **highest levels on Neuroticism and on both the External dimensions**. Nevertheless, it presents medium values for Conscientiousness and the remaining variables have each the lowest values of the set. As we mentioned, it is possible to observe that the trait of Agreeableness presents similar values across clusters, while Openness to Experience, in spite of not showing significant differences between clusters, has very dissimilar values on Cluster 2 compared to the others. In the real world, Cluster 1 contains people that are organized and believe in their efforts. In contrast, Cluster 3 includes moody people that believe the external world has a large influence over their life. Finally, Cluster 2 contains outgoing and open people.

In order to extract information visualization preferences for the different contexts among individuals of those three clusters, we opted for the Apriori algorithm [13], an association rules method to find common patterns. Data preprocessing included the creation of an array for each participant containing the idiom that they preferred the most for each context. In case of a tie between two or more idioms in their preference ratings, we included all idioms that tied together. Afterwards, we divided users by their cluster labels and used the Apriori algorithm in each cluster. Each run was performed with lower bound minimal values of 0.1 for support, 0.8 for confidence, and 3.1 for lift. An Apriori association rule is often represented as $itemA \rightarrow itemB$, which translates into $itemB$ being frequently present in a set of preferences that also contains $itemA$.

We continued our analysis by choosing which rules to focus on the information visualization techniques according to the frequency of each rule. We started by choosing the rule with the highest frequency and then choosing rules that had similar item sets until there was no rule with common or contradictory associations. Finally, if a context did not have a style associated to it, we chose the most frequent preferred idiom for that context among participants of the cluster, which was the case for the hierarchy context for Cluster 1. Based on the final set of rules for each cluster, we were able to derive which idioms were the most preferred according to the different contexts (Table 4). Notably, there are differences in the contexts of evolution over time and hierarchy. Regarding the evolution over time context, both Clusters 1 and 2 prefer a sunburst idiom and Cluster 3 participants rate treemaps higher. Compared to the other contexts, the chosen evolution over time idiom was less prominent in Clusters 1 (50%) and 3 (36.7%), while Cluster 2 was more consistent in their preference (76.2%). For hierarchy contexts, while Cluster 1 completely prefers line charts with points (100%), the remaining clusters would rather omit the use of those marks. Finally, all clusters state that an horizontal bar chart is the most preferred idiom to use for comparison data, with frequency values around 71%. This may hint that the appropriateness of an idiom for a specific problem context acts as a stronger regulator compared to personality.

6 Discussion

After analysing our results with both approaches we were able to have a better understanding of how personality has an effect on information visualization technique preferences. From the correlation-based analysis, results pointed towards effects from the Neuroticism, Openness to Experience, and Agreeableness traits in user preference regarding different idioms. Several more facets from all FFM traits and both External LoC dimensions also suggested a correlation. This lack of significance results is a consequence of the Bonferroni correction we applied in order to counteract the multiple comparisons problem. We believe the correlation-based approach is sound for a smaller number of Spearman correlations, which points our next step in this research towards the separate analysis of these personality variables to verify whether the results of this study have a high false discovery rate. Regarding Neuroticism, the trait itself and three facets showed a weak negative effect, suggesting that people with higher levels of Neuroticism dislike line charts with points. In fact, we were able to verify the same effect on Cluster 3, where users had the highest levels of Neuroticism and they preferred line charts without points. Additionally, only the cluster with the lowest levels of Neuroticism (Cluster 1) showed a preference towards line charts with points. This effect may be given to how people with high Neuroticism experience more stress when the idiom contains more marks, thus more information that may be harder to perceive. Extraversion only showed strong results in the cluster-based approach. While individuals with high and medium levels preferred sunburst as an idiom to represent hierarchy, people with low levels showed a preference towards treemaps. Interestingly, individuals with medium levels would rather use line charts with points, contrary to the remaining participants which showed an inclination towards excluding points on those charts. This effect may be explained by interaction effects with the remaining personality variables. In the correlation-based approach, three of its facets may have had an effect on preferences for sunburst and line chart idioms, yet all effects were weak in size.

The best clusters produced by the k-means algorithm did not divide individuals significantly based on Openness to Experience. Nevertheless, we found that independently of the remaining personality variables, results suggested it could foster the preference for area charts. Furthermore, five of its facets hinted negative weak effects with several idioms, mostly regarding evolution over time idioms. This is a rather interesting effect, considering how Openness to Experience has been shown to model how individuals process evolution [1, 9]. Agreeableness was also not significantly different among the different clusters, yet, similarly to Openness to Experience, it suggested some significant effects on the correlation-based approach along four of its facets. Most of these effects are referring to hierarchy, which may be related to how Agreeableness models how people evaluate hierarchical structures of collectivism [29]. Finally, Conscientiousness showed more effects while interacting with the other personality variables in the cluster-based approach then by an analysis with correlations. People with high levels tend to prefer line charts with points compared to the remaining population. We believe that this preference may be given to how these people prefer an organised approach to life, thus preferring to see idioms with more detail. We also found that people with high and low values on this trait prefer a sunburst in comparison to a treemap, similar to the Extraversion trait. Concerning the dimensions from LoC, both External dimensions suggested positive weak correlation effects. While Powerful Others hinted an effect on the evolution over time context, Chance did it for the comparison context. In addition, cluster-based analysis showed that people with higher values on these dimensions and the lowest Internal levels among the clusters have a preference for a treemap compared to the sunburst idiom. In contrast, the highest values for Internal and lowest for both the External dimensions showed a preference for line charts with points. This effect may be a result of *Internals* being faster than *Externals* when the former search for items [10] because they use additional marks such as points to guide their search.

In the light of this, our results suggest that personality is a differentiating factor when it comes to designing information visualization systems. Looking into our approaches, for example, while facets from Conscientiousness hinted in the correlation-based approach, participants from Cluster 1 preferred line charts with points. In addition, results from the Cluster 3 were indicated by the correlation results where higher values on Neuroticism or its facets led participants to choose a line chart without points. The same effect happened with Powerful Others. In contrast, we found dissimilar effects for Extraversion and Agreeableness. Regarding the former, we expected that Cluster 2 would have a preference on the evolution over time context for line charts with points and on the evolution over time context different from a sunburst idiom. Moreover, the latter was not significantly different between clusters. In the light of this, we hypothesize that this lack of significance led to an omission of interaction effects from Agreeableness. In this case, we must also address how participants rated each idiom independently of personality. As the Spearman's correlation effects were all small in size, we believe that the interaction effects were stronger, as one individual perceives information through interactions of all their personality constructs and not only one. Thus, we consider the cluster-based approach to be more appropriated. There are some important factors that may explain the lack of significance observed in some of our results. First, since we are tackling a lot of personality variables, a larger number of participants would allow conclusions with a stronger impact in both approaches. In particular, we could have a better sampling regarding Openness to Experience, Agreeableness, and the Internal dimension of LoC. Secondly, although there are more idioms in information visualization, there are more idioms from these contexts to explore. We also did not control for the familiarity cognitive bias, which may have had an effect on the results. Thirdly, the scenario and the complexity of the dataset used to illustrate the different contexts may have had an effect on how people perceived the idioms. Finally, not asking users to perform any task rather than rating their preference for the aesthetics of an idiom may not impact visual task analysis.

7 Conclusions and Future Work

This exploratory study focuses on personality with two different psychological constructs (FFM and LoC) models user preferences regarding information visualization techniques in three different contexts: hierarchy, evolution over time, and comparison. Besides identifying which idioms are modelled by personality-based user preferences, our results suggest important implications that may be used in the design pipeline to customize information visualization systems. Future work includes the implementation and testing of different information visualization systems developed based on our results to assess how they affect user preference, performance, experience, and satisfaction. In addition, task types, task complexity, and contexts should be further explored as they may lead to distinct interactions of users given their individual differences. Finally, we aim to recruit a larger number of participants so that we can explore more in-depth the personality variables that were not significantly different between clusters.

Acknowledgments

This work was supported by national funds through Fundação para a Ciência e a Tecnologia (FCT) with references UIDB/50021/2020 and SFRH/BD/144798/2019.

REFERENCES

[1] A. Acerbi, M. Enquist, and S. Ghirlanda. Cultural evolution and individual development of openness and conservatism. *Proceedings of the National Academy of Sciences*, 106(45):18931–18935, 2009.

[2] M. Braun, R. Chadowitz, and F. Alt. User experience of driver state visualizations: A look at demographics and personalities. In *IFIP Conference on Human-Computer Interaction*, pp. 158–176. Springer, 2019.

[3] 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.

[4] 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.

[5] C. Conati, G. Carenini, D. Toker, and S. Lallé. Towards user-adaptive information visualization. In *Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015.

[6] 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

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

[8] J. C. de Winter, S. D. Gosling, and J. Potter. Comparing the pearson and spearman correlation coefficients across distributions and sample sizes: A tutorial using simulations and empirical data. *Psychological methods*, 21(3):273, 2016.

[9] G. J. Feist and T. R. Brady. Openness to experience, non-conformity, and the preference for abstract art. *Empirical Studies of the Arts*, 22(1):77–89, 2004.

[10] 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.

[11] 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

[12] J. Han, J. Pei, and M. Kamber. *Data mining: concepts and techniques*. Elsevier, 2011.

[13] M. Ilayaraja and T. Meyyappan. Mining medical data to identify frequent diseases using apriori algorithm. In *2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering*, pp. 194–199. IEEE, 2013.

[14] H.-S. Kim and S. S. Sundar. Motivating contributions to online forums: can locus of control moderate the effects of interface cues? *Health communication*, 31:1–13, 09 2015. doi: 10.1080/10410236.2014.981663

[15] 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.

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

[17] H. Levenson. Multidimensional locus of control in psychiatric patients. *Journal of consulting and clinical psychology*, 41(3):397–404, 1973. doi: 10.1037/h0035357

[18] H. Levenson. Reliability and validity of the i, p, and c scales-a multidimensional view of locus of control. 1973.

[19] M. Lima and A. Simões. Neo-pi-r manual profissional. *Lisboa: CEGOC*, 2000.

[20] Z. Liu, R. J. Crouser, and A. Ottley. Survey on individual differences in visualization. *arXiv preprint arXiv:2002.07950*, 2020.

[21] R. R. McCrae and P. T. Costa. Self-concept and the stability of personality: Cross-sectional comparisons of self-reports and ratings. *Journal of Personality and Social Psychology*, 43(6):1282, 1982.

[22] L. McInnes, J. Healy, and S. Astels. hdbscan: Hierarchical density based clustering. *Journal of Open Source Software*, 2(11):205, 2017.

[23] G. Norman. Likert scales, levels of measurement and the "laws" of statistics. *Advances in health sciences education*, 15(5):625–632, 2010.

[24] N. Oscar, S. Mejía, R. Metoyer, and K. Hooker. Towards personalized visualization: Information granularity, situation, and personality. In *Proceedings of the 2017 Conference on Designing Interactive Systems*, pp. 811–819, 2017.

[25] A. Ottley, R. J. Crouser, C. Ziemkiewicz, and R. Chang. Manipulating and controlling for personality effects on visualization tasks. *Information Visualization*, 14(3):223–233, 2015.

[26] 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.

[27] S. Petrovic. A comparison between the silhouette index and the davies-bouldin index in labelling ids clusters. In *Proceedings of the 11th Nordic Workshop of Secure IT Systems*, pp. 53–64, 2006.

[28] P. J. P. Queirós. *Burnout en el trabajo y conyugal en enfermeros portugueses*. PhD thesis, Universidad de Extremadura, 2004.

[29] A. Realo, J. Allik, and M. Vadi. The hierarchical structure of collectivism. *Journal of Research in Personality*, 31(1):93–116, 1997.

[30] S. M. Sarsam and H. Al-Samarraie. A first look at the effectiveness of personality dimensions in promoting users' satisfaction with the system. *SAGE Open*, 8(2):2158244018769125, 2018. doi: 10.1177/2158244018769125

[31] 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.

[32] D. Toker, C. Conati, G. Carenini, and M. Haraty. Towards adaptive information visualization: on the influence of user characteristics. In *International conference on user modeling, adaptation, and personalization*, pp. 274–285. Springer, 2012.

[33] J. Wang, J. Wang, Q. Ke, G. Zeng, and S. Li. Fast approximate k-means via cluster closures. In *Multimedia data mining and analytics*, pp. 373–395. Springer, 2015.

[34] 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.

[35] 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.

[36] C. Ziemkiewicz, A. Ottley, R. J. Crouser, A. R. Yauilla, S. L. Su, W. Ribarsky, and R. Chang. How visualization layout relates to locus of control and other personality factors. *IEEE transactions on visualization and computer graphics*, 19(7):1109–1121, 2012.