Visual Data Exploration: How Expert Astronomers Use Flipbook-Style Visual Approaches to Understand New Data

Fernanda Monteiro Eliott[a] (fernanda.m.eliott@vanderbilt.edu)
Keivan Stassun[b] (keivan.stassun@vanderbilt.edu)
Maithilee Kunda[a] (mkunda@vanderbilt.edu)
[a]Department of Electrical Engineering and Computer Science, [b]Department of Physics and Astronomy
Vanderbilt University, PMB 351679, 2301 Vanderbilt Place, Nashville, TN 37235-1679, USA

Abstract

What are the cognitive processes in play when someone uses a visualization tool to interactively explore a new dataset? Here, we focus on one particular type of visualization—the scatter plot—which, despite (or perhaps because of) its simplicity, is still one of the most frequently used plot types in many data-intensive disciplines. We conducted a pilot study to investigate how expert astronomers interact with an unfamiliar dataset using a visualization tool called Filtergraph, which supports rapid and easy visualization of large datasets. We present both qualitative and quantitative results, including observations about the temporal dynamics of visual data exploration as well as interesting behavioral patterns that we saw in our participants, such as users taking "circular walks" through the data at various levels of abstraction.

Keywords: Data exploration; graph understanding; information visualization; scatter plots; visualization software.

Introduction

When astronomer Henry Norris Russell first introduced the now-called Hertzsprung-Russell (H-R) diagram, he wrote, "The appearance of [the figure] suggests the hypothesis that, if we could put on it some thousands of stars, instead of the 300 now available, ...we would find the points representing them clustered principally close to two lines, one descending sharply along the diagonal...the other ...running almost horizontally. ...These two classes of stars were first noticed by Hertzsprung, who has applied to them the excellent names of *giant* and *dwarf* stars" (Russell, 1914, p. 287).

In addition to Russell's obvious desire for more data, his wonderfully vivid description conveys the fundamentally *visual* nature of this discovery. Indeed, the H-R diagram has been called "perhaps the most spectacularly successful example of a simple scatterplot" in all of science (Spence & Garrison, 1993, p. 1). Today, like many disciplines, astronomy enjoys volumes of data that Russell could only have imagined. However, an astronomer's expertise to *make sense* out of data—to recognize which patterns represent actual scientific discovery—remains as vital today as it was in 1914.

Most human sense-making with data involves a visually-mediated interaction between the data and the perceptual/cognitive processes of the user—*data visualization*. Data visualization can be as short and simple as glancing at a printout of a plot on paper, or as lengthy and complex as spending months analyzing and modeling a large dataset. There is an increasing need for interactive data visualization tools that not only leverage the latest in pattern recognition and data mining algorithms, but also place the cognitive needs of the user front and center—to assist and augment *human* capabilities in the discovery process (Honavar, Hill, & Yelick, 2016).

One vital role for data visualization is the open-ended, open-minded exploration of data that leads to unexpected insight, often manifested at first as a "striking" or "interesting" multivariate plot, such as with the H-R diagram. To take a more recent example, an astronomy research team at Vanderbilt University developed a visualization tool called Filtergraph (see Figure 1) designed to allow people to rapidly and easily explore large datasets of up to a few million points (Burger et al., 2013). Using the Filtergraph software to visualize stellar variability data gathered by the *Kepler* spacecraft resulted in an unexpected and "visually interesting" scatterplot, which in turn led to the discovery of stellar granulation "flicker" and its utility for stellar and exoplanets research, a significant finding that was published in *Nature* (Bastien, Stassun, Basri, & Pepper, 2013).

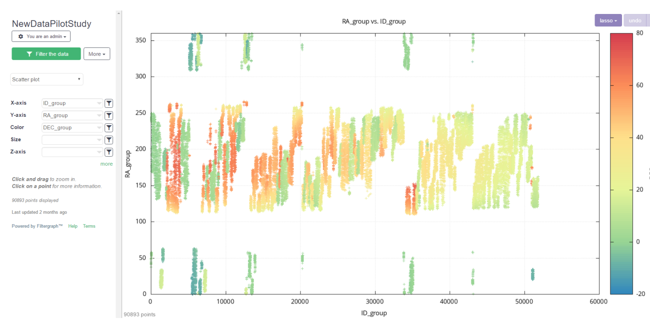

Figure 1: A screenshot of the Filtergraph data visualization interface (Burger et al., 2013), and also the initial view shown to participants in our pilot study.

Here, we report our preliminary results on a pilot study to investigate how expert astronomers interact with an unfamiliar dataset using scatterplots generated by Filtergraph. We use a novel analysis approach that is different from, but complementary to, existing methods that focus on measuring actions or tasks conducted by the data analyst. Instead, our approach measures interactions in terms of dataset attributes: which variables from a large dataset does an analyst look at, when, and why? We present both qualitative and quantitative results from this study, including observations about the temporal dynamics of visual data exploration as well as interesting behavioral patterns that we saw in our participants, such as users taking "circular walks" through the data at various levels of abstraction.

Related Work

There is a rich foundation of work in HCI (Human Computer Interaction), visual analytics, and infovis that aims to understand the processes by which people interact with and understand data. Sanderson and Fisher (1994) present a widely-used, general framework for thinking about user interactions with an interactive tool in terms of sequences of actions that represent different user functions, such as connecting ideas or introducing comments.

Specifically in relation to data visualization, Yi and colleagues (2007) identify seven modes of interaction with visualization tools, such as select, explore, and filter, that they believe are important for understanding the visual sense-making process. Brehmer and Munzner (2013) present a task typology that bridges low-level interactions with high level tasks and goals during visualization activities. Pirolli and Card (2005) present a detailed cognitive task analysis of visual sensemaking in the domain of intelligence analysis. ElTayeby and Dou (2016) present methods for studying exploratory data visualization that leverage the automated analysis of rich, quantitative interaction log data to identify and understand underlying patterns of interaction.

Saraiva and colleagues (2005) conducted a very interesting pilot study specifically focused on open-ended, exploratory data analysis in the domain of bioinformatics. They focus on how a visualization tool can complement a dataset in order to facilitate insights that may lead to a discovery. One important issue they addressed was how to define and measure *insight*. They defined insight based on the context of their work as well as based on characteristics such as time, hypotheses, correctness, and category. Their results show the influence of the visualization tool itself over the processes of human interpretation and insight.

Mayr and colleagues (2016) looked at how mental models parallel a user's use of external visualization tools. They identify key characteristics pertaining to mental models, such as content, structure, coherence, perspectivity, generalizability, and utility, and they review existing empirical methods for conducting user studies to get at these characteristics.

Finally, there is important work being done to understand data visualization and sensemaking in terms of core cognitive capabilities that people bring to bear on such tasks. Healey and Enns (2012) provide a research survey on cognitive theories of attention and perceptual processing as applicable to data visualization. They provide examples of factors that drive visual attention (such as visual feature hierarchies, memory, and prediction) and also factors that impair attention, such as what happens during change blindness, with observations about how improperly designed visualizations can significantly impact a user's mental models of the data.

Tversky (2003) reflects about humans, actions, and space, emphasizing the importance of differences in how people represent space across different spatial reference frames (e.g., in a navigation task versus a graphical understanding task). She includes discussion of the spatial references frames people use while processing external visualizations of information.

Methods

We conducted a pilot study to investigate how expert astronomers interact with an unfamiliar dataset using the Filtergraph visualization tool shown in Figure 1. We chose the domain of astronomy as being representative of today's data-intensive disciplines. In designing our study, we wanted to choose a data exploration task that was open-ended enough to provide a realistic challenge to our participants, but also that had at least some constraints so that we could analyze meaningful differences across participants. We decided to invite astronomers to our lab to participate in one-hour sessions, during which they would be instructed to "explore" an astronomy dataset that they had not previously seen. All necessary IRB approvals were obtained prior to the study.

Dataset. We chose a dataset that we believed would not require complex mathematical operations to make sense of, and also that would not be from too specialized a subfield within astronomy. The dataset we chose is described in Berlind et al. (2006) and contains data describing $90,893$ galaxies, which are individually discriminated through 10 attributes (including the galaxy ID), or group separated through 9 attributes (including the group ID), as shown in Table 1.

Table 1: The 19 attributes in the galaxy dataset used for our pilot study (Berlind et al., 2006), along with letter codes used throughout this paper. The last two shaded items represent arithmetic combination of attributes that we observed participants construct on the fly during the study sessions.

Attribute	Code	Attribute	Code
ID_group	A	DEC_gal	L
RA_group	B	Velocity_gal	M
DEC_group	C	AbsMag_g_gal	N
Velocity_group	D	AbsMag_r_gal	O
Ngal_group	E	sersic_gal	P
Velocity_dispersion_group	F	fiber_col_gal	Q
AbsMag_g_group	G	completeness_gal	R
AbsMag_r_group	H	distance_edge_gal	S
cen_sat_flag_gal	I	AbsMag_g_gal-AbsMag_r_gal	T
ID_gal	J	AbsMag_g_group-AbsMag_r_group	U
RA_gal	K		

Study protocol. We recruited 7 graduate students in astronomy from the Vanderbilt community, ranging in age from 23 to 28 years old, with 2 identifying their gender as female, 4 as male, and 1 as two-spirited. We ran participants in five sessions: two sessions (S1 and S4) each involved two participants conducting data exploration collaboratively, and the other three sessions (S2, S3, and S5) each involved a single participant. Note that as part of our pilot study design, we decided to include both individual and collaborative sessions to better inform our approaches for future studies.

Participants received gift cards as compensation for their time. Each session proceeded as follows. First, the participant(s) filled out a short demographic questionnaire. Then, we asked participants to sit down at a computer workstation and use Filtergraph to visualize and explore the galaxy data set. We provided a printout listing all dataset attributes and their semantic descriptions. A member of our study team sat with participants during the session, asking open-ended questions to better understand how the interaction was unfolding. At the end of the session, participants were asked to write down their own impressions about the study, and ideas about what software, tools, catalogs, or other data-related affordances would make their life as an astronomer easier.

Table 2: Session details from our pilot study. *M* gives the number of major observations in each session, and *N* gives the number of minor observations.

Session	# of participants	Duration (seconds)	M	N
S1	2	3378	12	67
S2	1	3160	7	47
S3	1	3289	20	73
S4	2	4394	19	107
S5	1	2921	-	-

Filtergraph settings. To constrain the visualizations used by participants, we asked participants not to change the Filtergraph setting that selects scatterplots as the visualization type. Within the scatterplot setting, Filtergraph offers many interactive options for changing how the dataset is visualized. The attributes assigned to X and Y axes can be changed, and a Z axis attribute can optionally be added. Attributes can also be assigned to the color dimension or used to select or filter out portions of the data. Anywhere individual attributes are used, arbitrary mathematical transformations or combinations of attributes can also be assigned. Additionally, there are options for changing the background plot color as well as the size and shape of data points. When the Z axis is in use, there are options to rotate the plot or change its scale.

All the sessions started from the same home screen, shown in Figure 1, with the X-axis set to attribute A, the Y-axis set to attribute B, and the color set to attribute C (see Table 1).

One concern we had was that participants might get bored or fatigued during the session and generate plots only to fill the time, and not through genuine interest and curiosity in exploring the dataset. Thus, the member of our team who sat with participants tried to be friendly and engaging, to help create a positive session environment. (Note that this member of our team comes from a computer science background, not astronomy, and so we do not believe this "social" aspect of the study sessions introduced significant biases in which parts of the dataset the participants would choose to focus on.)

In addition, after 45 minutes, the participants were informed that they could finish their current activity and end the session, if they wanted to, or they could continue to work for as long as they chose. As it turns out, all of our participants chose to stay past the 45 minute mark, with a minimum session duration of about 48 minutes and a maximum duration of about 73 minutes; see Table 2.

Analysis approach. We gathered data using a combination of note-taking by study personnel, paper forms, and interaction data collected on the computer workstation through screen recordings of each session. To analyze results, we defined the concepts of *major observations* and *minor observations* of the dataset, as illustrated in Figures 2 and 3.

Definition: A *major observation* is a grouping of contiguously viewed scatterplots within which the attributes assigned to X and Y axes remain constant. When a user changes the attribute assigned to either the X or Y axis (or both), a new major observation begins.

Definition: A *minor observation* is a grouping of contiguously viewed scatterplots that occurs during a major observation, within which one or more individual, highly related scatterplots are viewed. For example, looking at a 3D scatterplot but rotating the plot to see many different views would constitute a single minor observation of the data.

We used screen recordings to identify major and minor observations within each session. We did not count plots that were generated in the course of defining a single set of plotting parameters (i.e., because Filtergraph redraws plots nearly instantly, while the user might still be typing). Also, sometimes participants assigned the exact same parameters to the current plot, not actually changing the plot at all. We did not count these immediately repeated plots either. Finally, for continuously changing plots, which we saw especially when participants were smoothly rotating or moving 3D plots, we only counted the first and last views as separate plots within the same minor observation.

Exclusions. During session S2, an attribute was plotted by mistake; the participant had intended to plot a different attribute instead and only discovered their mistake after 12 minutes. We do not include these "mistake" plots in our current analysis, though certainly these kinds of mistakes will be considered in future work. In addition, session *S*5 was qualitatively very different from Sessions S1 through S4. In session *S*5, the participant appeared to pay very little attention to the semantics of the attributes that were being plotted (even when prompted to consider attribute meanings by our study team member). This participant declared to be selecting attributes randomly, and engaging in a primarily perceptual exploration of the dataset. While we find this pattern of interaction in session S5 extremely interesting, we felt that a different approach would be needed to analyze these data, and so we have left the analysis of session S5 for future work.

Results and Discussion

Here, we discuss preliminary results from our pilot study. While there is certainly work to be done in analyzing the specific, astronomy-related meanings of participants' data visualization choices, including a detailed criterial analysis on their mental models (Mayr et al., 2016), that type of analysis falls outside the scope of the current paper and will be part of future work. For now, we focus on describing the

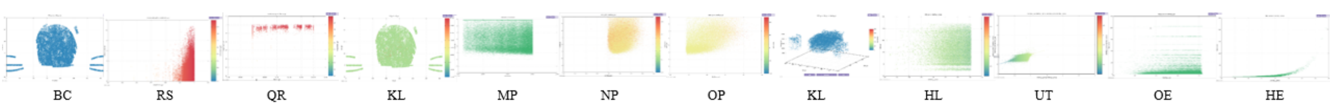

Figure 2: For session S1, plots showing the first minor observation for each of the 12 major observations.

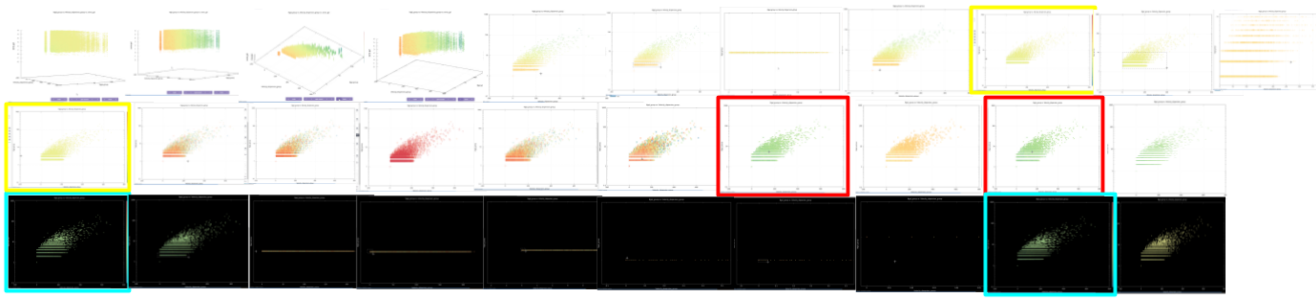

Figure 3: For session S2, in chronological order, from left to right: plots showing the 28 minor observations within the fourth major observation. Note the "circular walk": the yellow, red, and blue boundaries indicate 3 different plots that were each observed twice. The six plots with a black background are unclear due to the small number of plotted points.

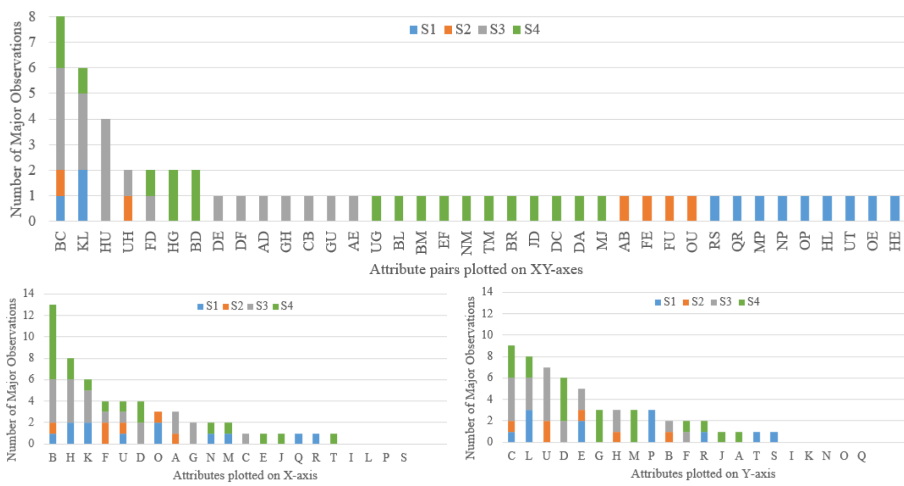

Figure 4: Number of major observations generated during each session for different combinations of attributes.

general behavioral patterns of data visualization that we observed, including the temporal dynamics of participant interactions with the Filtergraph tool. Table 2 gives details of the five sessions that we ran for this study.

Distribution of major observations by attribute. Consider the 17 data attributes contained in the galaxy dataset (see Table 1). Note that two of the original 19 attributes are ID numbers, for galaxies and groups of galaxies, and so do not capture "meaningful" in the same sense as the other 17. The number of possible major observations that could be made from this data, i.e., possible assignments of attributes to X and Y axes, is (17 choose 2) = 272 possibilities. If we allow mathematical transformations or combinations of attributes, then the number of possible major observations is infinite.

Across all sessions, participants viewed a total of only 38 distinct major observations, as shown in Figure 4. This figure also shows which attributes participants assigned to either the X or Y axes. Interestingly, only 4 major observations were shared by two or more sessions; this shows the high variability in data exploration paths taken by different individuals.

Attribute I was the only one never assigned to either the X or Y axes. It turns out that this attribute is Boolean; it indicates whether the galaxy is the brightest in its group, and so it makes sense to not be chosen for one of the primary axes. Attributes P (a measure of galaxy morphology) and Q (was there a problem while measuring the galaxy) are the other two categorical variables in the dataset; interestingly, these were assigned to axes at various times during one of the sessions.

The BC combination was the only major observation shared across all four sessions. B gives the longitude of the galaxy group center, and C gives the latitude. So, plotting the BC combination produces what is essentially a spatial "map" of the galaxies represented in the dataset.

Temporal dynamics of data exploration. We are partic-

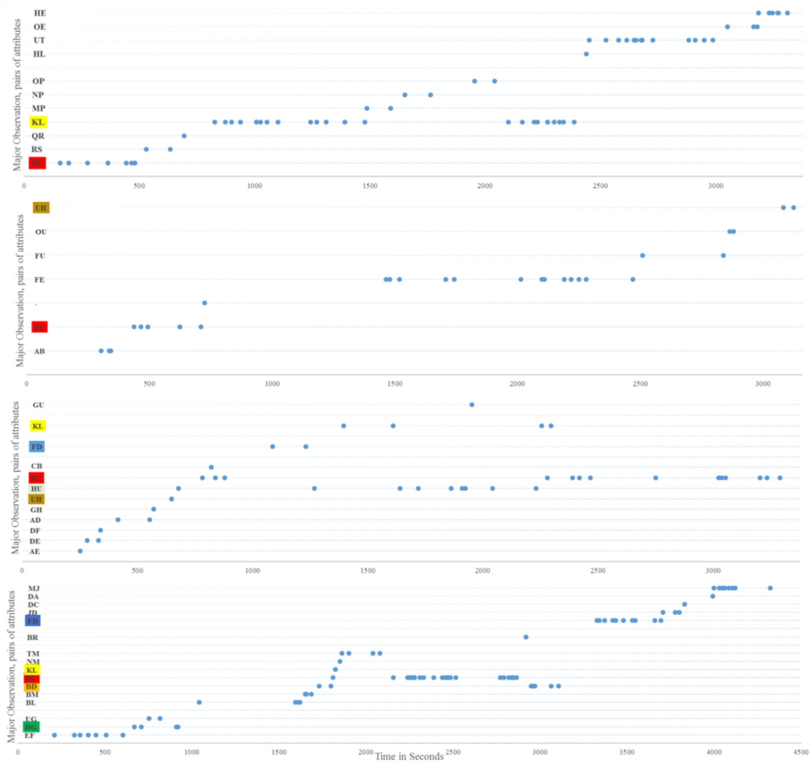

Figure 5: Distribution of major and minor observations over time, for sessions S1 (top) through S4 (bottom). Each point indicates a group of highly related individual plots (or, for example, a single continuous rotation of the same 3D plot). Major observations repeated within or across sessions are marked by colored labels.

ularly interested in understanding the *temporal dynamics* of open-ended data exploration. How frequently do our participants switch from one major or minor observation to the next?

Figure 5 illustrates the temporal distributions of major and minor observations for all four sessions. By looking at the blue dots in this figure, we can see the most active periods, the ones in which many plots were quickly generated. We can also see periods in which a single minor observation was studied at some length. (Note also that the large gap early in session S2 is due to data we omitted, as described earlier.) The first point of interest is that participants generally chose to "flip through" the dataset at a fairly brisk pace. Part of this could, of course, be due to the participants knowing they had only one hour to complete the study, but we expect the same to hold in more naturalistic settings as well.

The average duration of minor observations was about 61 seconds, though this distribution has a long tail that falls off fairly consistently and extends to the longest minor observations at around 554 seconds. For major observations, the average duration was about 229 seconds, with a maximum of about 1046 seconds. The durations of major observations seem to show a bimodal tendency, with many major observations falling under the 150 second mark, but another large grouping lying in the 150-400 second range.

Interestingly, participants often returned to the same major observation within the same session. This pattern occurs not at all in session S2, occasionally in session S1, and quite a lot in sessions S3 and S4. In sessions S2 and S4, we saw similar "circular walk" patterns *within* some of the major observations; the participant often returned to the same minor observation that they had started with, before moving on to the next minor or major observation. Figure 3 depicts an example of a "circular walk" within session S2.

Within each major observation, it seems as though participants are directing a type of "movement" from the first to the last plot (depicted via minor observations); a story is being told through the movement of data points across the plots. This notion of movement/story strongly brings to mind the idea of flipbooks. For instance, Figure 3 shows a story regarding the relationship between the attributes F and E.

Other qualitative observations. To conclude our presentation of preliminary results, we present a few high-level observations about patterns of data exploration in our study.

Starting point: Participants seemed to choose a starting point to anchor themselves in relation to the dataset. Sessions S1 and S2 began by plotting attributes that involve the spatial position of galaxies, perhaps to let participants establish a mental map of the spatial layout of the galaxy dataset. Sessions S3 and S4 began by looking at attributes like the number of galaxies in each group and the velocity with which each group is moving away from us, perhaps establishing an egocentric reference frame of "us versus the galaxies."

Mouse gestures: Participants frequently used the mouse as a communicative or attention-focusing tool, i.e., to gesture at the visualizations. While some of these were directed at our study team member who was observing the session, many of these mouse gestures also occurred when participants were interacting with Filtergraph and thinking about what to do next. Sessions S1 and S2 exhibited many more mouse gestures than did sessions S3 and S4.

Paper aids: Participants in sessions S1 and S2 relied heavily on the paper printout of attribute details throughout the sessions. Participants in S3 and S4, on the other hand, used the printout at the start, but, during the interaction itself, relied more on the list of attributes provided by Filtergraph.

Collaboration/leadership: In both of the sessions with two participants, we observed that one of the participants seemed to lead the exploratory line of thought. But note that *leading* does not necessarily means commanding the mouse and directly interacting with the workstation. In one session, the "thought" leader was also the one interacting with Filtergraph, but in the other session, the "thought" leader was not the primary tool interaction person.

Collaboration/corrections: We observed that, in the sessions with two participants, one helped the other to quickly correct mistakenly plotted attributes. In contrast, during session S2 (which involved a single participant), an attribute was plotted by mistake and not discovered for 12 minutes.

Conclusion and Future Directions

Our findings highlight a few interesting properties of how domain experts explore an unfamiliar dataset, particularly in terms of temporal and dimensional patterns. The next challenge is to understand how, as people follow exploratory paths through a dataset, they build meaningful cognitive representations of what they see, and how they are able to identify encounters with unexpected, significant data patterns.

We anticipate that temporal patterns are especially important from a cognitive perspective because they describe not just moment-to-moment attentional switches but also serve as a way of marking successive stages at which a person's mental model of a dataset is likely changing. Five minutes spent looking at a single plot, versus five minutes spent "flipping" through multiple plots, are both likely to be equally important modes of exploration. The key is figuring out the cognitive purpose served by each.

We are also intrigued by the frequency of "circular walks" in the exploratory paths taken by our participants. Viewing the same plot twice serves no obvious purpose from a purely statistical or data mining perspective. However, in humans, we predict that these "circular walks" actually serve key roles related to memory, attention, salience, etc.

Acknowledgments

We would like to thank Dr. Andreas Berlind who generously shared the galaxy dataset we used in our study. This research makes use of Filtergraph, an on-line data visualization tool developed at Vanderbilt University through the Vanderbilt Initiative in Data-intensive Astrophysics (VIDA).

References

Bastien, F., Stassun, K. G., Basri, G., & Pepper, J. (2013). An observational correlation between stellar brightness variations and surface gravity. *Nature*, *500*(7463), 427–430.

Berlind, A., Frieman, J., Weinberg, D., Blanton, M., Warren, M., Abazajian, K., ... others (2006). Percolation galaxy groups and clusters in the sdss redshift survey: identification, catalogs, and the multiplicity function. *ApJ Supplement Series*, *167*(1), 1.

Brehmer, M., & Munzner, T. (2013). A multi-level typology of abstract visualization tasks. *IEEE Transactions on Visualization and Computer Graphics*, *19*(12), 2376–2385.

Burger, D., Stassun, K., Pepper, J., Siverd, R., Paegert, M., De Lee, N., & Robinson, W. (2013). Filtergraph: An interactive web application for visualization of astronomy datasets. *Astronomy and Computing*, *2*, 40–45.

ElTayeby, O., & Dou, W. (2016). A survey on interaction log analysis for evaluating exploratory visualizations. In *Proc. of the beyond time and errors on novel eval methods for visualization* (pp. 62–69).

Healey, C., & Enns, J. (2012). Attention and visual memory in visualization and computer graphics. *IEEE trans. Visualization & Comp. Graphics*, *18*(7), 1170–1188.

Honavar, V., Hill, M., & Yelick, K. (2016). Accelerating science: A computing research agenda. *arXiv preprint arXiv:1604.02006*.

Mayr, E., Schreder, G., Smuc, M., & Windhager, F. (2016). Looking at the representations in our mind: Measuring mental models of information visualizations. In *Proc. of the beyond time and errors on novel eval methods for visualization* (pp. 96–103).

Pirolli, P., & Card, S. (2005). The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of international conference on intelligence analysis* (Vol. 5, pp. 2–4).

Russell, H. N. (1914, May). Relations Between the Spectra and Other Characteristics of the Stars. *Popular Astronomy*, *22*, 275-294.

Sanderson, P. M., & Fisher, C. (1994). Exploratory sequential data analysis: Foundations. *Human–Computer Interaction*, *9*(3-4), 251–317.

Saraiya, P., North, C., & Duca, K. (2005). An insight-based methodology for evaluating bioinformatics visualizations. *IEEE trans. visualization & comp. graphics*, *11*, 443–456.

Spence, I., & Garrison, R. (1993). A remarkable scatterplot. *The AM STAT*, *47*(1), 12–19.

Tversky, B. (2003). Structures of mental spaces how people think about space. *EAB*, *35*(1), 66–80.

Yi, J., ah Kang, Y., & Stasko, J. (2007). Toward a deeper understanding of the role of interaction in information visualization. *IEEE trans. visualization & comp. graphics*, *13*(6), 1224–1231.