

# GalStamps: Analyzing Real and Simulated Galaxy Observations

Nina McCurdy\*  
University of Utah

Miriah Meyer†  
University of Utah

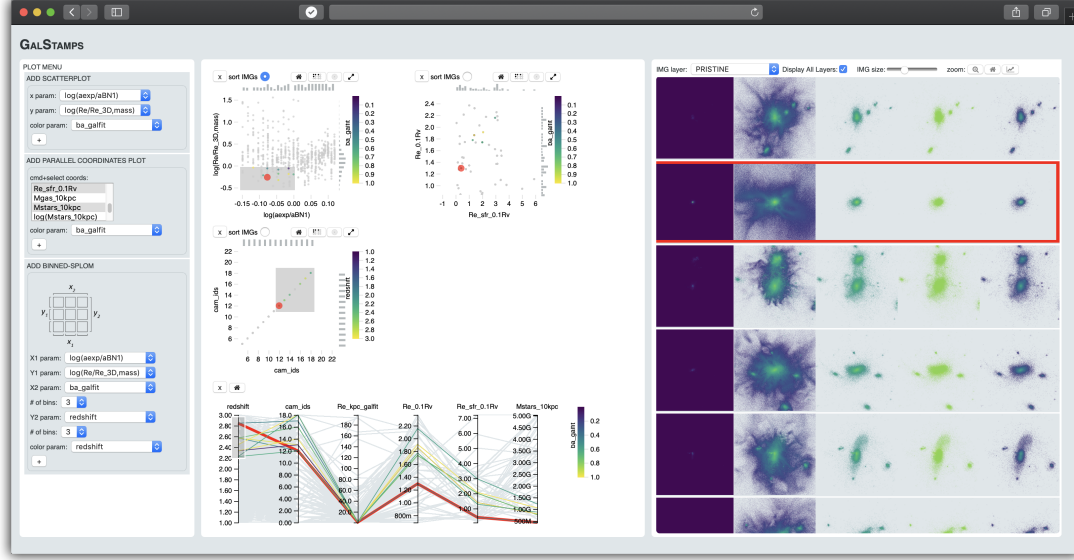


Figure 1: the GalStamps interface supports the linked statistical and visual analysis of galaxy observation data. Statistical plots are generated via a plot menu (left pane) and displayed in the plot view (middle pane). Selections within plots reveal the associated 2D scalar field data (rendered as images) in the image view (right pane). Hovering over images in the image view highlights the associated statistical data in the plot view (shown in red).

## ABSTRACT

One way astronomers and astrophysicists study galaxy formation and evolution is by analyzing and comparing *real* galaxy observations, captured by telescopes, and *simulated* galaxy observations, generated from theoretical models. They approach this through a combination of statistical and visual analysis, conducted either independently or sequentially. During the first year of an ongoing design study with astronomers and astrophysicists, we explored approaches to integrating statistical and visual analysis to enhance understanding of these data. Contributions from this stage of the study include a data and task abstraction for statistically and visually analyzing real and simulated galaxy observations, as well as an initial design, implemented in a prototype called GalStamps, and evaluated through two case studies with domain experts.

**Index Terms:** Visualization—Design study—Astronomy

## 1 INTRODUCTION

When we look out in space, we are looking back in time. Astronomical observations of galaxies, like those taken by the Hubble Space Telescope, provide snapshots at different stages of galaxy formation and evolution — as if billions of flip books were torn apart and spread across the sky. Theoretical astrophysicists develop models that effectively reassemble these flip books. Through a sequence of simulated snapshots, these models predict how galaxies form,

interact, merge, and evolve over time. Matching *real* observations from telescopes with *simulated* observations from snapshots of theoretical models can help astronomers predict the histories and futures of individual galaxies. Comparing collections of real and simulated observations can help validate models and clarify key processes in galaxy formation and evolution.

Astronomers and astrophysicists studying galaxy formation and evolution analyze and compare these data through a combination of statistical analysis of plots and visual analysis of images and movies. These two forms of analysis are most often completed independently or sequentially. Integrated analysis of these data is not supported in existing astronomical data analysis tools. The goal of this work is to begin to fill this gap by exploring integrated approaches to statistical and visual analysis to enhance the analysis of these data.

This work evolved out of a long-term, collaborative relationship with astronomers and astrophysicists at the University of California Santa Cruz. We report here on the initial results of an ongoing design study. Current contributions of the study include: a data and task abstraction for statistically and visually analyzing real and simulated galaxy observations; and an initial design, implemented in a prototype called GalStamps and evaluated through two case studies with domain experts. Our results lay the groundwork and introduce interesting visualization research opportunities for further design study research. As supplemental material, we include a rich process description in order to support transparency and transferability of our research.

## 2 PROBLEM DOMAIN BACKGROUND

One of the major questions in astronomy is how galaxies were formed and how they have evolved over time. Observational astronomers study galaxy formation and evolution by analyzing observations of galaxies captured by space-based and ground-based

\*e-mail: nina@cs.utah.edu

†e-mail: miriah@cs.utah.edu

telescopes. These telescopes survey regions of the sky, imaging galaxies with instruments that measure different wavelengths of light. Astronomers then derive additional statistical measurements from these images to characterize and analyze the observations.

Theoretical astrophysicists, on the other hand, study galaxy formation and evolution by developing computational models that simulate how galaxies form, merge, and evolve over time. Comparing the results of these models to real observations helps researchers validate and improve models and clarify what is being exhibited in observations. To support this comparison, theoretical astrophysicists generate collections of *simulated* observations by taking snapshots of their simulations at different angles and different stages of evolution, rendering images at different wavelengths of light, and degrading the images to resemble those taken by telescopes. Similar to real observations, comparable statistical measurements are then derived from these simulated images. Astronomers have termed these real and simulated images *galaxy postage stamps*.

In both cases, scientists analyze galaxy postage stamps and derived statistical measurements in order to identify trends, correlations, and outliers that help to answer open questions about galaxy formation and evolution. Analyzing statistical measurements reveals features that are either not captured or less evident in the images, and vice versa. Visually inspecting the images associated with features in the statistical measurements allows scientists to verify their findings and to discover correlations between statistical and visual attributes. Finally, comparing features across real and simulated collections of observations is critical to these scientists’ analyses: agreement in features allows scientists to validate existing models, and thus the underlying theory, whereas disagreement in features helps scientists refine theory and even uncover new, key phenomena. Neither of these forms of analysis — visually verifying statistical features, and comparing features across real and simulated collections of observations — are supported in an integrated way by existing astronomical pipelines and technologies. As a result, these powerful techniques are used only on occasion. We speculate that this has resulted in missed opportunities for verification and discovery.

### 3 RELATED WORK

Within the astronomy and astrophysics community, a host of services have been developed to support the dissemination and analysis of real and simulated observation data. These platforms support the navigation and exploration of astronomical surveys and catalogues [1, 2, 24, 37] and offer a range of tools for mapping and cross-matching observations across datasets [5, 6], statistically analyzing associated tabular data [38], and visually classifying real and simulated galaxy observations [9]. Numerous visualization packages have also been developed to support statistical and visual analysis of astronomical data more generally [17, 25, 39]. These packages target a broad range of data types — from catalogue data to time-evolving numerical simulation data — and support an equally broad range of visualization techniques — from volume rendering to 2D and 3D scatterplots. While some of these tools support both statistical and visual analysis, these tasks are typically performed separately.

Within the visualization community, research has mainly focused on the *visual* representation of astronomical data. This work spans efforts to visualize astronomical and cosmological simulation data [3, 21, 35] to reconstructing 3D models from 2D observations [22, 28, 41], to developing 3D Universe software, which visualizes and maps observation data from a range of sources for use in planetariums and other public outreach venues [7, 33]. Other instances of visualization research focus on representing uncertainty in astronomical data [23] and supporting the mapping and cross-matching of multisurvey observational datasets [27]. These instances focus purely on visual analysis and not on statistical analysis.

A handful of successful approaches have combined visual and statistical analysis of astronomical data using linked views [15, 16, 31].

This work, however, falls primarily in the context of visualizing and analyzing cosmology simulation data. A more general push for linked view analysis of high-dimensional astronomical data has come from the astronomy visualization community [13]. Glue [4], a widely used Python library designed to support multidimensional linked-data exploration, is a direct response to this call. With Glue, users can manually generate scatterplots, histograms, and images from multiple related datasets. The platform is built around brushing and linking, such that selections in any one graph or image are propagated to all others. Glue, however, focuses its analysis on individual images, allowing users to query and visually inspect image pixels corresponding to statistical properties. Our work extends the Glue framework by supporting this analysis across collections of images.

### 4 DESIGN PROCESS

As part of an ongoing investigation into recording and reporting on design study research, we include a rich process description — detailing our methods and resulting artifacts, and reflecting on the learning that occurred along the way — in supplemental materials<sup>1</sup>. What follows is a high-level summary of this process description.

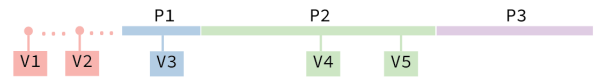


Figure 2: the design study was conducted in 3 core phases (P1-P3) and structured around 5 visits (V1-V5) to the domain experts’ home institution.

As shown in Figure 2, the design study was carried out in 3 core phases (P1-P3) structured around 5 visits (V1-V5) to the University of California Santa Cruz (UCSC), and concluding with a period of remote, iterative design and development. Our collaborators consisted of faculty, graduate and undergraduate students, and a post-undergraduate researcher, all conducting research in either observational astronomy or theoretical astrophysics and with a focus on galaxy formation and evolution. Preconditioning and formulation of the domain problem [36] occurred over the course of two visits to UCSC (V1-V2) in the two years prior to the start of study. Through reviewing the literature, attending talks and group meetings, and conducting unstructured and semistructured interviews the visualization researchers developed an understanding of the problem space and honed in on an initial visualization research opportunity [36].

The first core phase of the study (P1) was devoted to developing an initial data and task abstraction. This phase spanned 2 months and was structured around a single visit (V3), in which visualization researchers conducted interviews, think-alouds, and group brainstorms involving existing, related tools. The final data and task abstraction, developed and refined over the remainder of the study, is presented in Section 5. Building on the initial data and task abstraction, the second phase (P2) was devoted to eliciting design requirements and developing an initial design. Over the course of 6 months and 2 visits (V4-V5), the team iteratively developed technology probes to brainstorm and experiment with features, functionalities, and interactions, and to explore applications to a range of datasets.

The final phase (P3), spanning the last 4 months of the study, was devoted to translating the study findings and the results of the technology probes [20] into a deployable prototype. During this time, visualization researchers met remotely with one collaborator, on a weekly or biweekly basis, to iteratively design and develop the prototype using a case study that was central to the collaborator’s research. Results of this phase are presented in Section 6.

### 5 DATA AND TASK ABSTRACTION

Within our collaborators’ analysis, data items correspond to individual **observations** and datasets correspond to **collections of observa-**

<sup>1</sup> <http://bit.ly/GalStamps>

**tions.** Whether real or simulated, a single observation corresponds to a region in the sky where a number of measurements are taken. These measurements are recorded as images and represented as a set of 2D scalar fields, where each field stores a different measurement, or *visual attribute*, across a 2D grid. Measurements derived from the set of 2D scalar fields are recorded as *statistical attributes* of the observation, and stored as a row of tabular data along with the statistical attribute values of all other observations within the given collection. Thus, each observation in a collection is represented both as a set of 2D scalar fields and as a row of tabular data. The two representations are linked via an observation ID.

The goal of our collaborators’ analysis is twofold: to verify that their theoretical simulations, and thus the underlying physics, are accurately capturing what is being observed in nature; and to use these simulations to help characterize what is being captured by observations. In order to do this, our collaborators compare real and simulated galaxy observations. The associated visualization tasks involve *identifying* important features within each collection of observations, and *comparing* these features across collections [8].

Identifying features within a single collection of observations involves analyzing statistical attributes within the tabular data, visual attributes within the 2D scalar data, and the relationships between these attributes. Visualization tasks include *locating*, *exploring*, and *browsing* [8] patterns, trends, correlations, and outliers within and across statistical and/or visual attributes. Comparing features across collections of observations involves analyzing the relationships [12] (similarities, differences, correlations, etc.) between features across datasets. Designing for this comparative level of analysis is a focus of future work in our ongoing design study.

## 6 VISUALIZATION DESIGN

Our second contribution is a visualization design for statistically and visually analyzing real and simulated galaxy observations. We implemented our design in a prototype called GalStamps, named after the *galaxy postage stamps* that our collaborators analyze. The GalStamps interface, shown in Figure 1, comprises a plot menu and two linked views: a plot view and an image view. Our design was grounded in the visualization literature and the results of our design study, and was heavily informed by prior work surrounding combined statistical and visual analysis using linked views [18, 30]. GalStamps was implemented in D3.js. A video demonstration of the system is included in supplemental materials.

### 6.1 Plot Menu

The plot menu, shown as the leftmost pane in Figure 1, allows users to generate custom plots from the statistical attributes available for a given dataset. We found manual selection of plot attributes to be an efficient approach, as each domain expert was interested in exploring unique and specific sets of attributes. Users can generate three types of plots: scatterplots, parallel coordinate plots, and binned-SPLoms. We detail these plots in the proceeding sections.

### 6.2 Plot View

The plot view, comprising the middle pane of the prototype shown in Figure 1, serves two primary function. First, it supports the analysis of statistical observation attributes. As described in Section 5, this involves identifying patterns, trends, correlations, and outliers within and across statistical attributes. We support this through the analysis and comparison of plots. Second, the plot view provides a selection mechanism for querying 2D scalar field data based on selections of statistical attribute data. Selections are made via brushing. The plots are linked, such that selections in one plot are propagated to all other plots. For selections in multiple plots, the intersection of the selections is utilized.

Examples of scatterplots and parallel coordinate plots are shown in Figure 1. In the **scatterplots**, observations are encoded as points

of uniform radius, and histograms encode the distribution of points along each axis. Users can toggle to display isodensity contours. Point selections are made via a single 2D rectangular brush; isodensity contours are selected via mouse-click. Scatterplots support zooming and panning, and may be expanded to the width of the plot view. In the **parallel coordinate plots**, observations are encoded as joined line-segments, and selections are made via 1D rectangular brushes along one or multiple axes. Users can zoom and pan along each axis and reorder axes via dragging. Our initial inclusion of parallel coordinate plots was exploratory, as our collaborators had never seen them before. Interestingly, the domain experts predominately used the parallel coordinates plots as interactive legends for filtering high-dimensional data, which echoes the findings of others [19]. As described in Section 7, they also proved useful for summarizing the statistical attributes for a given observation.

Examples of the third type of plot, which we have termed **binned-SPLoms**, are shown in the plot view panes in Figure 3. In a binned-SPLom, the bottom ( $x$ ) and left ( $y$ ) axes are preserved across all cells. Data in each cell is filtered into 2D bins, defined by secondary, top ( $x'$ ), and right ( $y'$ ) axes. The static form of these plots is gaining popularity in the astronomy and astrophysics research community as an effective way to identify trends along pairs of secondary attributes [11, 42]. In this work, we experimented with adding interactivity to these plots. Zooming and panning allows users to adjust the ranges of the primary  $x$  and  $y$  axes. Histogram-scented filters allow users to adjust bin sizes and ranges along the secondary  $x'$  and  $y'$  axes. Observations are encoded as points, users can toggle to display isodensity contours, and selections are made via a 2D rectangular brush, or via mouse-clicks for isodensity contours.

### 6.3 Image View

The image view, shown in the rightmost pane in Figure 1, supports the analysis of visual attributes for a collection of observations. This includes tasks for analyzing individual observations, represented as sets of 2D scalar fields, and for comparing multiple observations.

As a preprocessing step, 2D scalar fields are assigned a colormap and rendered as images. For a given observation, users can switch between different 2D scalar fields, via a dropdown menu, or can view the set of 2D scalar fields side by side. Image size can be adjusted via a slider, and users can toggle in and out of *zoom mode*, which supports zooming on images via mouse-scroll. Zooming on one image translates to all other images. For a given selection in the plot view, the associated images are displayed in rows and are ordered to reflect the  $xy$  distribution of the selection (the *selection space*) in the plot view. We liken this to an abstracted version of image scatterplots [40]. For intersected selections, this defaults to the latest selection. Given multiple plots in the plot view, users can toggle between selection spaces via radio buttons. Images can also be reordered manually via dragging — we found this to be useful both for comparing images side by side and ordering groups of images based on visual patterns.

## 7 CASE STUDIES

In this section, we present two case studies used to evaluate our initial design and its implementation in GalStamps. These case studies validated our integrated approach to enhancing and accelerating our collaborators’ analysis, and verified that our initial design was effective in supporting this approach. Additionally, the first case study revealed the need to support comparison of multiple collections of observations; the second case study illustrates the use of parallel coordinates for summarizing distributions along each attribute.

### 7.1 Comparing Real & Simulated Galaxy Observations

The first collaborator used the tool to examine findings from a recent paper [42]. The paper employed data derived from both real and simulated galaxy observations to model the evolution of galaxy

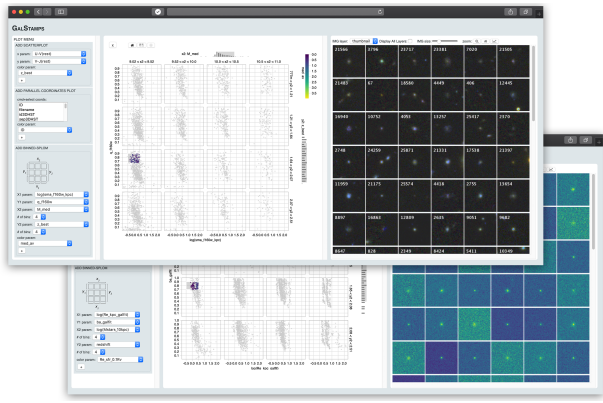


Figure 3: our first collaborator used the prototype to compare real (*foreground*) and simulated (*background*) galaxy observations and to confirm prior claims based on deep learning analysis.

shapes within the CANDELS galaxy survey [14]. This collaborator began by examining a *real* collection of galaxy observations. Using the plot menu, he recreated a central figure of the paper, a binned-SPLOM, shown in Fig. 3 (*foreground*). In the plot view, he adjusted the histogram-scented brushes to reflect the original binning in the paper, and began brushing to select various regions of interest within the different cells. With each selection, he studied the resulting images, hovering over individual images to reveal and assess the corresponding plot point. He was reassured that the images confirmed his expectations, and that he could reason about any disagreements.

In a separate browser, he then recreated and examined the same binned-SPLOM, this time generated using a *simulated* collection of galaxy observations, shown in Figure 3 (*background*). Selecting the same regions of interest in the plot view, he confirmed that the associated simulated galaxy observations appeared as he expected. He commented that being able to easily pull up and inspect the associated images was useful. As a final step, he examined the two browsers side by side, selecting the same regions in each of the binned-SPLOMs, and visually comparing the resulting sets of images. In each case, the simulated images looked very similar to the observed images, thereby validating the paper findings [32]: “*we made this claim [...] based on deep learning analysis, but this is much better confirmation! [...] looking at the images and seeing how similar they are is really quite nice!*” In discussing plans to support the comparative analysis of multiple datasets (e.g., real and simulated), he responded “*why make things more complicated? This is fine!*” Finally, this collaborator remarked on the broader utility of the system. Speaking about a related project, he commented: “*tools like this one would be enormously helpful in analyzing these data.*”

## 7.2 Deconstructing 2D Scalar Fields

The second collaborator intended to use the system to investigate a puzzling statistical feature in a simulated observation dataset. Questions about the image data, however, caused her to redirect her analysis, using the system instead to identify the visual attributes stored within each of the 2D scalar fields. By plotting related statistical attributes, selecting specific regions, and cross-verifying the associated images, she was able to test various hypotheses and eventually draw a tentative conclusion. A snapshot of this process is depicted in Figure 1. Building on her conclusion, she was then able pursue her initial investigation, generating preliminary hypotheses.

In reflecting on this process, she commented that the system enabled her to ask and answer her questions quickly and efficiently, in what would otherwise be a long, iterative process involving python scripts. She remarked that seeing the tabular and image data together allowed her to explore the data in new ways. She also commented on the usefulness of the parallel coordinate plot for summarizing

the statistical attribute values for a given observation, and for cross-verifying them with the images: “*instead of jumping around from plot to plot, I can just look at how each of [the attributes] are related.*”

## 8 DISCUSSION

The case studies presented in Section 7 provide validation for our data and task abstraction and our initial design, and illustrate two ways in which the design can be used to enhance our collaborators’ analysis. Our collaborators found the GalStamps interface to be intuitive, and that the workflow — iteratively generating plots, selecting regions of interest, and cross-examining the associated observations — allowed them to effectively and efficiently pursue existing questions of their data, while also prompting new kinds of questions. These results give us confidence to continue with our initial design and reveal interesting design considerations moving forward. First, while our ultimate goal is to support the comparative analysis of real and simulated galaxy observations, and while our findings throughout the study indicate the importance of designing for this level of comparison, we wonder whether utilizing two side-by-side versions of the system is in fact sufficient, as evidenced by our first case study, or whether there is significant added benefit integrating this analysis. Second, feedback gathered with the technology probe and in the second case study suggests a different use of parallel coordinate plots beyond exploring correlations between attributes: as an interactive legend for filtering and summarizing high-dimensional data.

Our findings throughout the first stage of this design study point to two interesting mismatches between visualization guidelines and domain practice. First, as we describe in the rich process description included in supplemental materials, interactive 3D scatterplots play a valuable role in our collaborators’ current analysis workflow. As one collaborator put it: “*I’ve built my career analyzing 3D scatterplots.*” 3D scatterplots, however, are argued against within the visualization community [10, 26, 29]. Building on existing work in this space [34], we see this collaboration as opportunity to further investigate the potentially underappreciated value of interactive 3D scatterplots for analyzing and detecting features in 3+ dimensional data. Second, an underlying question for us throughout this study was: *why hadn’t this system already been developed?* The GalStamps prototype uses a standard multiple linked view approach with rich interactions, but these design standards had not yet made their way into the astrophysical community. We see this collaboration as a valuable opportunity to investigate the barriers that domains face in adopting established visualization techniques.

## 9 CONCLUSIONS AND FUTURE WORK

We present the initial results of an ongoing design study exploring the integration of statistical and visualization techniques to enhance the analysis and comparison of real and simulated galaxy observation data. Contributions from this stage include a data and task abstraction for statistically and visually analyzing galaxy observation data, and an initial design, implemented in a prototype called GalStamps. Evaluation through two case studies validates the design and informs the next stages of our study.

Moving forward, our collaborators plan to use the prototype in several upcoming and ongoing projects. We plan to use these opportunities to continue to test and develop our design, including integrating support for comparative analysis of real and simulated galaxy observations. In addition to developing the standalone system, we have begun working with lead researchers for yt and Glue to develop our system as extensions of these visualization and analysis packages. We hope this will help disseminate our results to the broader astronomy and astrophysics community and beyond.

## ACKNOWLEDGMENTS

We thank our collaborators at the University of California Santa Cruz. This work was supported in part by NSF grant GRF-1747505.



## REFERENCES

- [1] Mikulski archive for space telescopes. [archive.stsci.edu](http://archive.stsci.edu).
- [2] The virtual astronomical observatory. [virtualobservatory.org](http://virtualobservatory.org).
- [3] K. R. Almyre and A. G. Forbes. Halos in a dark sky: Interactively exploring the structure of dark matter halo merger trees. In *2015 IEEE Scientific Visualization Conference (SciVis)*, pp. 73–77. IEEE, 2015.
- [4] C. Beaumont, T. Robitaille, A. Goodman, and M. Borkin. Multidimensional data exploration with glue. In *Proceedings of the 12th Python in Science Conference*, p. 8, 2013.
- [5] G. Berriman, A. Laity, J. Good, J. Jacob, D. Katz, E. Deelman, G. Singh, M. Su, and T. Prince. Montage: The architecture and scientific applications of a national virtual observatory service for computing astronomical image mosaics. In *Proceedings of Earth Sciences Technology Conference*, 2006.
- [6] T. Boch, A. Oberto, P. Fernique, and F. Bonnarel. Aladin: An open source all-sky browser. In *Astronomical Data Analysis Software and Systems XX*, vol. 442, p. 683, 2011.
- [7] A. Bock, E. Axelsson, C. Emmart, M. Kuznetsova, C. Hansen, and A. Ynnerman. Openspace: Changing the narrative of public dissemination in astronomical visualization from what to how. *IEEE computer graphics and applications*, 38(3):44–57, 2018.
- [8] M. Brehmer and T. Munzner. A multi-level typology of abstract visualization tasks. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2376–2385, 2013.
- [9] H. Dickinson, L. Fortson, C. Lintott, C. Scarlata, K. Willett, S. Bamford, M. Beck, C. Cardamone, M. Galloway, B. Simmons, et al. Galaxy zoo: morphological classification of galaxy images from the illustris simulation. *The Astrophysical Journal*, 853(2):194, 2018.
- [10] N. Elmqvist, P. Dragicevic, and J.-D. Fekete. Rolling the dice: Multidimensional visual exploration using scatterplot matrix navigation. *IEEE transactions on Visualization and Computer Graphics*, 14(6):1539–1148, 2008.
- [11] J. J. Fang, S. Faber, D. C. Koo, A. Rodríguez-Puebla, Y. Guo, G. Barro, P. Behroozi, G. Brammer, Z. Chen, A. Dekel, et al. Demographics of star-forming galaxies since  $z = 2.5$ . i. the uvj diagram in candel. *The Astrophysical Journal*, 858(2):100, 2018.
- [12] M. Gleicher. Considerations for visualizing comparison. *IEEE transactions on visualization and computer graphics*, 24(1):413–423, 2017.
- [13] A. A. Goodman. Linking visualization and understanding in astronomy. In *American Astronomical Society Meeting Abstracts*, vol. 223, 2014.
- [14] N. A. Grogan, D. D. Kocevski, S. Faber, H. C. Ferguson, A. M. Koekemoer, A. G. Riess, V. Acquaviva, D. M. Alexander, O. Almaini, M. L. Ashby, et al. Candel: the cosmic assembly near-infrared deep extragalactic legacy survey. *The Astrophysical Journal Supplement Series*, 197(2):35, 2011.
- [15] P. Hanula, K. Piekutowski, C. Uribe, K. Almyre, A. Nishimoto, J. Aguilera, and G. E. Marai. Cavern halos: Exploring spatial and nonspatial cosmological data in an immersive virtual environment. In *2015 IEEE Scientific Visualization Conference*, pp. 87–99, 2015.
- [16] S. Haroz, K.-L. Ma, and K. Heitmann. Multiple uncertainties in time-variant cosmological particle data. In *2008 IEEE Pacific Visualization Symposium*, pp. 207–214, 2008.
- [17] A. Hassan and C. J. Fluke. Scientific visualization in astronomy: Towards the petascale astronomy era. *Publications of the Astronomical Society of Australia*, 28(2):150–170, 2011.
- [18] C. Henze. Feature detection in linked derived spaces. In *Proceedings Visualization '98 (Cat. No. 98CB36276)*, pp. 87–94. IEEE, 1998.
- [19] I. Hotz. Visualizing line-like features – cleaning up the mess. Lecture. The University of Utah. Salt Lake City, Utah, January 2019.
- [20] H. Hutchinson, W. Mackay, B. Westerlund, B. B. Bederson, A. Druin, C. Plaisant, M. Beaudouin-Lafon, S. Conversy, H. Evans, H. Hansen, et al. Technology probes: inspiring design for and with families. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 17–24. ACM, 2003.
- [21] R. Kaehler, O. Hahn, and T. Abel. A novel approach to visualizing dark matter simulations. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2078–2087, 2012.
- [22] H. Li, C.-W. Fu, and A. Hanson. Visualizing multiwavelength astrophysical data. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1555–1562, 2008.
- [23] H. Li, C.-W. Fu, Y. Li, and A. Hanson. Visualizing large-scale uncertainty in astrophysical data. *IEEE transactions on visualization and computer graphics*, 13(6):1640–1647, 2007.
- [24] C. J. Lintott, K. Schawinski, A. Slosar, K. Land, S. Bamford, D. Thomas, M. J. Raddick, R. C. Nichol, A. Szalay, D. Andreescu, et al. Galaxy zoo: morphologies derived from visual inspection of galaxies from the sloan digital sky survey. *Monthly Notices of the Royal Astronomical Society*, 389(3):1179–1189, 2008.
- [25] D. R. Lipsa, R. S. Laramée, S. J. Cox, J. C. Roberts, R. Walker, M. A. Borkin, and H. Pfister. Visualization for the physical sciences. 31(8):2317–2347, 2012.
- [26] S. Liu, B. Wang, J. J. Thiagarajan, P.-T. Bremer, and V. Pascucci. Visual exploration of high-dimensional data through subspace analysis and dynamic projections. In *Computer Graphics Forum*, vol. 34, pp. 271–280. Wiley Online Library, 2015.
- [27] T. B. Luciani, B. Cherinka, D. Oliphant, S. Myers, W. M. Wood-Vasey, A. Labrinidis, and G. E. Marai. Large-scale overlays and trends: Visually mining, panning and zooming the observable universe. *IEEE transactions on visualization and computer graphics*, 20(7):1048–1061, 2014.
- [28] M. Magnor, G. Kindlmann, C. Hansen, and N. Duric. Reconstruction and visualization of planetary nebulae. *IEEE Transactions on Visualization and Computer Graphics*, 11(5):485–496, 2005.
- [29] T. Munzner. *Visualization analysis and design*. AK Peters/CRC Press, 2014.
- [30] P. M. Nina McCurdy. Interactive exploration of solar magnetic field lines. Presented at SC17, Denver CO, November 2017.
- [31] A. Preston, R. Ghods, J. Xie, F. Sauer, N. Leaf, K.-L. Ma, E. Rangel, E. Kovacs, K. Heitmann, and S. Habib. An integrated visualization system for interactive analysis of large, heterogeneous cosmology data. In *2016 IEEE Pacific Visualization Symposium*, pp. 48–55. IEEE, 2016.
- [32] J. Primack, A. Dekel, D. Koo, S. Lapiner, D. Ceverino, R. Simons, G. Snyder, M. Bernardi, Z. Chen, H. Domínguez-Sánchez, et al. Deep learning identifies high- $z$  galaxies in a central blue nugget phase in a characteristic mass range. *The Astrophysical Journal*, 858(2):114, 2018.
- [33] A. Sagrista, S. Jordan, T. Müller, and F. Sadlo. Gaia sky: Navigating the gaia catalog. *IEEE transactions on visualization and computer graphics*, 25(1):1070–1079, 2019.
- [34] H. Sanftmann and D. Weiskopf. 3d scatterplot navigation. *IEEE Transactions on Visualization and Computer Graphics*, 18(11):1969–1978, 2012.
- [35] K. Schatz, C. Müller, M. Krone, J. Schneider, G. Reina, and T. Ertl. Interactive visual exploration of a trillion particles. In *2016 IEEE 6th Symposium on Large Data Analysis and Visualization (LDAV)*, pp. 56–64. IEEE, 2016.
- [36] M. Sedlmair, M. Meyer, and T. Munzner. Design study methodology: Reflections from the trenches and the stacks. *Visualization and Computer Graphics, IEEE Transactions on*, 18(12):2431–2440, 2012.
- [37] A. S. Szalay. The national virtual observatory. In *Astronomical Data Analysis Software and Systems X*, vol. 238, p. 3, 2001.
- [38] M. Taylor. Topcat: Working with data and working with users. *arXiv preprint arXiv:1711.01885*, 2017.
- [39] M. J. Turk, B. D. Smith, J. S. Oishi, S. Skory, S. W. Skillman, T. Abel, and M. L. Norman. yt: A multi-code analysis toolkit for astrophysical simulation data. *The Astrophysical Journal Supplement Series*, 192(1):9, 2010.
- [40] C. Vertan, M. Ciuc, C. Fernandez-Maloigne, and V. Buzuloiu. Browsing image databases by 2d image similarity scatter plots: updates to the iris system. In *Proceedings of International Conference Communications*, pp. 397–402, 2002.
- [41] S. Wenger, M. Ament, S. Guthe, D. Lorenz, A. Tillmann, D. Weiskopf, and M. Magnor. Visualization of astronomical nebulae via distributed multi-gpu compressed sensing tomography. *IEEE transactions on visualization and computer graphics*, 18(12):2188–2197, 2012.
- [42] H. Zhang, J. R. Primack, S. Faber, D. C. Koo, A. Dekel, Z. Chen, D. Ceverino, Y.-Y. Chang, J. J. Fang, Y. Guo, et al. The evolution of galaxy shapes in candel: from prolate to discy. *Monthly Notices of the Royal Astronomical Society*, 484(4):5170–5191, 2019.