Human-Notebook Interactions: The CHI of Computational Notebooks

Jesse Harden
Virginia Tech
Blacksburg, Virginia, USA
jessemh@vt.edu

April Wang
ETH Zürich
Zürich, Switzerland
april.wang@inf.ethz.ch

Rebecca Faust
Virginia Tech
Blacksburg, Virginia, USA
rfaust@vt.edu

Katherine E. Issacs
University of Utah
Salt Lake City, Utah, USA
kisaacs@sci.utah.edu

Nurit Kirshenbaum
University of Hawaii at Manoa
Honolulu, Hawaii, USA
nuritk@hawaii.edu

John Wenskovitch
Pacific Northwest National Laboratory
Richland, Washington, USA
john.wenskovitch@pnnl.gov

Jian Zhao
University of Waterloo
Waterloo, Ontario, Canada
jianzhao@uwaterloo.ca

Chris North
Virginia Tech
Blacksburg, Virginia, USA
north@vt.edu

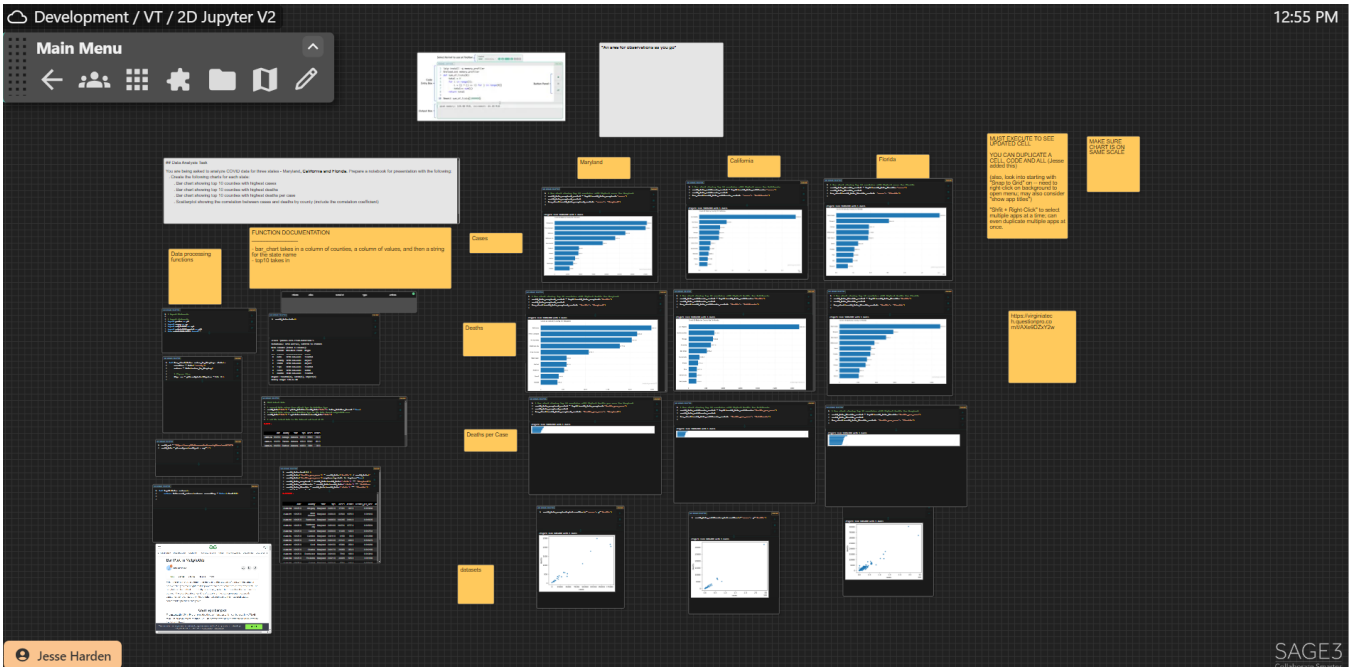

Figure 1: Example 2D Computational Notebook in SAGE3

CHI EA '24, May 11–16, 2024, Honolulu, HI, USA

ACM ISBN 979-8-4007-0331-7/24/05
https://doi.org/10.1145/3613905.3636318

ABSTRACT

The overall goal of this workshop is to bring together researchers from across the CHI community to share their knowledge and build collaborations at the intersection of computational notebook and HCI research, focusing on both the effective design and effective use of interfaces and interactions within computational notebook environments. This includes innovating upon the computational notebook metaphor, designing new tools, interfaces, and interactions for use with computational notebooks, and more. We aim

to pull expertise from across all fields of CHI to deliver novel research and generate open discussion about the current state of computational notebooks, how it can be improved from an HCI standpoint, and how these potential improvements can direct future research. To achieve this goal, we propose a full-day, hybrid workshop with discussions of challenges and opportunities, paper and demo presentations, lightning talks, and a keynote. Participants in this workshop will exchange ideas and help define a roadmap for future research at the intersection of HCI and computational notebook design.

CCS CONCEPTS

• **Human-centered computing** → **Collaborative and social computing**; **Collaborative and social computing systems and tools**; *Open source software*; **Interaction design**; *Visualization*; **Interactive systems and tools**.

KEYWORDS

Computational Notebooks, Human-Computer Interaction, Interface Design, Interaction Design, Computational Narratives, Data Science, Data Analysis

ACM Reference Format:
Jesse Harden, April Wang, Rebecca Faust, Katherine E. Issacs, Nurit Kirshenbaum, John Wenskovitch, Jian Zhao, and Chris North. 2024. Human-Notebook Interactions: The CHI of Computational Notebooks. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '24), May 11–16, 2024, Honolulu, HI, USA.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3613905.3636318

1 MOTIVATION

Computational Notebooks are becoming the standard media for much data science work [13, 15]. However, notebook systems are in their infancy in terms of user interface and interaction design. Research has described notebook system limitations [1], such as more complex non-linear data science and analysis workflows [3] and narratives [14], making the most of larger, widescreen displays [3]. While notebooks have become popular, their current limitations present significant difficulties for data scientists. In this workshop, we are intere7sted in exploring, cataloging, and innovating with respect to issues of user interface and interaction design of computational notebooks, such as the following:

- How can we make computational notebooks more usable for constructing, refining, and presenting computational narratives?
- How can we refine and/or innovate on the current state of computational notebooks?

At their best, computational notebooks actualize 7Knuth's concept of literate programming [8], where authors weave "human language with live code and the results of the code" to produce a computational narrative [12]; computational notebooks support incremental and iterative analysis, explanation of an analyst's thoughts and processes, and sharing of code, text, and visuals in one document [14]. In addition, recent work has shown how computational notebooks can support interactive visual analysis [11].

However, computational notebooks current have various pain points [1] for users, such as with exploring and analyzing data and the resulting issues with managing code; some of these issues, such as messiness [4] lead to the goals of data analysis and presentation with computational notebooks to be at odds [6]. Messiness in computational notebooks is a common problem [4], exacerbated by the lack of version control and dependency management systems. Tools such as Verdant [5] for version control and Albireo [17], as seen in Figure 2, for visually graphing dependencies between cells have been developed in response to these limitations of current computational notebooks. Debugging, another pain point of computational notebooks [1], was recently addressed with a visual tool, Anteater [2].

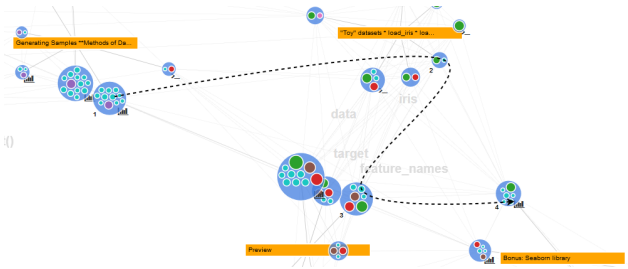

Figure 2: Demonstrated in their Albireo tool, Wenskovitch et al. proposed a visual approach for analyzing how variables change dynamically across notebook cells.

Some developed tools help expand the computational notebook metaphor and computational notebook capabilities. Work by Yifan Wu on B2 [18] expands computational notebook capabilities by capturing user interactions with interactive visualizations in computational notebooks, thus allowing their use in code cells. Kery et al., in their work on mage [7], provide data workers the ability to fluidly switch between working with code and working with a graphical user interface, expanding the modes in which workers can interact with computational notebooks. Work by Harden et al., as seen in Figure 1 explored the use of 2D space for organizing computational notebook cells [3], while Fork-It by Weinman et al. [16] utilized bounded 2D space to improve upon the standard computational notebook structure; both of these works expanded the computational notebook metaphor through their thoughtful use of available space.

Additional visual tools that rely on analyzing large repositories of computational notebooks to assist computational notebook users have been developed, as well. Examples include NBSearch [9] for semantic code searching of what has been done in other notebooks and EDAssistant [10] for help with exploratory data analysis; both of these tools visualize results from a repository of computational notebooks rather than visualizing properties of just one notebook.

Given the concerns above and the plentiful opportunities for potential research, this workshop offers a golden opportunity to coalesce, present, and discuss a research agenda centered on how to effectively improve computational notebooks with respect to their capabilities, interactions, and user interfaces. To do this, we will pull expertise from across all fields of human-computer interaction (HCI or CHI) to promote open discussion, establish the current state-of-the-art, and to produce an agenda for future research in this area.

1.1 Contrast with Similar Workshops

The topic of the workshop will focus on issues and opportunities related to the design, development, and use of computational notebooks. We believe this topic needs more focused attention from CHI community. Other related workshops have different foci:

- The Visual Data Science VDS symposium, which can be found at http://www.visualdatascience.org/2022/index.html, focuses exclusively on visualization in data science.
- The Workshop on Human-Centered Study of Data Science Work Practices (https://husdatworkshop.github.io/) does not focus directly on computational notebooks or literate programming tools and their interfaces and interactions.
- The Machine Learning from User Interactions (MLUI) workshop (https://learningfromusersworkshop.github.io/) focused on interaction provenance and system responses to inferred user intent, with no direct connection to computational notebook usage or research.
- The VisXAI Workshop (http://visxai.io) uses computational notebooks as its submission media to demonstrate eXplainable-AI concepts.
- The Interrogating Human-centered Data Science workshop (https://sites.google.com/view/hcds-workshop-chi2022/home) aims to reflect on HCI work towards humanizing data science.

In contrast, this proposed workshop centers on computational notebooks and their design, interfaces, and interactions as it relates to accomplishing the goals of literate programming in data science.

1.2 Workshop Themes

By focusing on the intersection of computational notebook research and CHI, we seek to support endeavors that better enable computational notebook users to perform data science tasks, craft rich and compelling computational narratives, and more effectively interact with computational notebooks. This workshop invites contributions that study, analyze, or address questions related to the following theme questions:

- How can we improve the accessibility of relevant properties of computational notebooks, such as kernel state, provenance, and more, through novel interface and interaction design?
- How can we improve upon the Computational Notebook Metaphor for literate programming, such as with alternative structures that support non-linearity in analysis and narrative?
- How can computational notebooks exploit recent technological advances, such as those in virtual and augmented reality, mobile devices, and artificial intelligence tools like ChatGPT?
- How can computational notebooks more effectively assist collaborative data science through methods such as best practices, novel interfaces and interactions?
- What challenges and opportunities exist for computational notebook use in the wild?
- How can we bridge the gap between computational notebook researchers, developers, and end users?
- How can we better evaluate both the actual and the potential value of computational notebooks and innovations to and/or for them?

2 ORGANIZERS

Jesse Harden is a Ph.D. student in Computer Science & Applications under Chris North at Virginia Tech. His current research is on the intersection of Data Science and Human-Computer Interaction, with a focus on computational notebooks and the use of 2D space. He conducts this research as part of the Smart Amplified Group Environments (SAGE) multi-university research group. He received his Master's in Data and Information Management from Radford University in 2019.

April Wang is an Assistant Professor at the Department of Computer Science at ETH Zurich, where she directs the Programming, Education, and Computer-Human Interaction lab. She received her Ph.D. from the University of Michigan. Her work aims to understand different collaboration needs and challenges around data science, and design better programming tools to support collaborative data science.

Rebecca Faust is a postdoctoral fellow at Virginia Tech. She received her Ph.D. from the University of Arizona. Her research focuses on enabling visualization and human interaction in data analytics to support analyst understanding. Her work includes a visual debugging method for computational notebooks.

Nurit Kirshenbaum is an Assistant Professor at the University of Hawaii-Mānoa (UHM). She specializes in Human-Computer Interaction and Tangible User Interfaces (TUI) such as ProjecTable - a physicalization system using a tangible model augmented with digital data. She is the User Experience designer of SAGE3- NSF's flagship middleware for working in data-rich immersive environments.

John Wenskovitch is a Visual Analytics Research Scientist at Pacific Northwest National Laboratory and an Adjunct Professor of Computer Science at Virginia Tech. He also received his Ph.D. from Virginia Tech in 2019. His current research focuses on Human-AI co-learning and teaming, trust in machine assistance, human-in-the-loop data science, analytical provenance, and the application of visualization to both the sciences and to cybersecurity. He teaches courses on information visualization, data analytics, computer graphics, and human-computer interaction.

Katherine E. Isaacs is an Associate Professor with the Scientific Computing and Imaging (SCI) Institute and the Kahlert School of Computing at the University of Utah. Her research focuses on visualization challenges in complex exploratory analysis scenarios, such as those of active data scientists and researchers across domains such as high performance computing, program analysis, and environmental justice. In meeting the needs of these data workers, she has integrated multiple interactive visualizations into computational notebook workflows.

Jian Zhao is an Assistant Professor in the Cheriton School of Computer Science, University of Waterloo, where he directs the WatVis (Waterloo Visualization) group. He received his Ph.D. from the University of Toronto. His research interests include information visualization, human-computer interaction, and data science. His work contributes to the development of advanced interactive

visualizations that promote the interplay of human, machine, and data.

Chris North is a Professor of Computer Science at Virginia Tech and Associate Director of the Sanghani Center for AI and Data Analytics. His research and education agenda seeks to enable human-AI interaction in visual analytics, immersive analytics, and data science, including novel interactive visual interfaces for computational notebooks. +

3 PLANS TO PUBLISH WORKSHOP PROCEEDINGS

We plan to publish accepted papers on our workshop website, and provide links to the online portals of the accepted demos. We are also considering publishing these materials through https://ceur-ws.org/ per recommendations from the workshop proposal website.

4 A HYBRID WORKSHOP EXPERIENCE

We plan to offer this workshop as a hybrid experience to enable more participants to join and learn from each other. All in-person talks will be broadcast and recorded over Zoom; this will enable participants to engage with the materials after the workshop, as well as enable any remote participants with internet difficulties to engage asynchronously. Furthermore, we plan to make use of SAGE3, an infinite canvas computational notebook environment with many applications integrated into it, including Python programming cells, sticky notes, web browsers, multi-page PDF viewers, and more. The use of SAGE3 infinite canvas boards, along with Zoom, will enable remote participants to interact with materials alongside in-person participants. These boards will be kept active for participants to engage with after the workshop.

For questions, participants can submit them either via Zoom on their laptop or mobile device, or via adding a sticky note on the SAGE3 board for the workshop. We will have at least one of the organizers watching for questions from remote participants for each part of the workshop.

4.1 What is SAGE3?

SAGE3 provides an environment where local and remote users can collaborate with one another, similar to a shared workspace in the cloud. SAGE3 enables users to create public or private SAGE3 boards, which are collected into SAGE3 rooms. Within a board, users can: upload/download various file types (e.g., images, videos, PDFs, code) to and from the boards; launch applications, such as, Sticky Notes, Maps, SAGECells, Notepad, Screenshare, and Webview; resize and move windows; collaboratively edit applications; draw on a board; view other users' cursors and viewports; and, execute Python code. Additionally, SAGE3 offers built-in AI functionality.

A SAGECell (See Figure 3) is comprised of two main GUI components: (1) a code area where users input and modify code, and (2) an output area that can display diverse data types, including images, HTML, and video. The user selects a kernel from the toolbar. SAGECells lets users switch between Python kernels, which enables task segregation, so users can independently experiment with or prototype analyses in separate kernels.

Figure 3: The SAGE3 SAGECell app allows local and remote participants program collaboratively.

5 ASYNCHRONOUS ENGAGEMENT

As stated in the prior section, we will keep SAGE3 boards with the workshop materials available, as well as Zoom recordings for later viewing. In addition, we will provide the recordings and instructions on how to access the SAGE boards on our workshop website.

6 WORKSHOP ACTIVITIES

We divide our discussion of workshop activities into a brief discussion of how we will encourage and select participants, an example workshop schedule, and workshop planning dates.

6.1 Workshop Participants

We will issue a Call for Papers, Demos, and Position Abstracts from the broader CHI community, as well as related communities such as the VIS community, the intelligent user interfaces community, and data science communities. We envision three main types of attendees at this workshop: paper presenters, demo presenters, and position abstract submitters.

Paper presenters will be selected based on paper submissions. Our full schedule presented in the Workshop Planning Dates section gives us time to review and select paper submissions, including a best paper or two. We have set aside time for 5 to 9 papers, but may adjust the schedule if we accept fewer.

Demo presenters will be selected based on demo submissions, which will be required to have an online portal for remote participants. Our full timeline presented in the Workshop Planning Dates section gives us time to review and select demo submissions, including potentially a best demo.

We seek position abstracts to ensure that attendees not presenting a paper or demo have a vested interest in the intersection of computational notebook research and CHI. Position abstract submitters will have an opportunity to briefly share their thoughts during time allotted for lightning talks and demos.

We anticipate 30 to 60 attendees may attend the workshop.

6.2 Workshop Schedule

To achieve our goals and make the desired impact on the CHI community, we propose a full-day workshop. Based on our uncertainty on the number of paper and demo submissions, we propose the following possible schedule:

- 9:00 AM - 9:15 AM: Introduction and Welcome
- 9:15 AM - 9:45 AM: Keynote Presentation
- 9:45 AM - 10:30 AM: Best Paper Presentations
- 10:30 AM - 11:10 AM: Morning Coffee Break
- 11:10 AM - 12:35 PM: Paper Presentations
- 12:35 PM - 2:30 PM: Lunch Break

- 2:30 PM - 3:15 PM: Lightning Talks & Demos Session
- 3:15 PM - 3:55 PM: Small Group Discussion
- 3:55 PM - 4:35 PM: Afternoon Coffee Break & Extended Demo Session
- 4:35 PM - 5:15 PM: Full Group Discussion
- 5:15 PM - 5:30 PM: Closing Comments & Wrap-Up

The workshop begins with a welcome from the organizers, explanation of the workshop format, and introduction of the Keynote speaker. The keynote speaker will give their presentation on a relevant topic related to computational notebooks, data science, and CHI. After the keynote, we transition to the paper presentations. We currently plan to give each best paper presenter 15 minutes to present and 7 minutes for questions, and other paper presenters 7 minutes to present plus 5 minutes for questions; this schedule assumes 2 best papers and 7 other papers, as well as a morning coffee break after the best paper presentations. We plan to follow up the paper presentations with a lunch break so that workshop participants can network and take a breather before jumping into the next set of sessions.

After the break, we will start a queue for participants to present short, 3-minute maximum talks on research they are doing, demos they are presenting, and/or their position abstracts on topics relevant to the workshop, with each talk followed up by a couple questions from a panel of the organizers and the audience. This session will naturally feed into the small group discussion session, where we will discuss current issues with computational notebooks and directions for further research in small groups. Once the small group discussion is closed, we will open the floor for demos and poster presentations, giving participants another chance to network as well as view ongoing research and quality additions to the computational notebook space before reconvening for the full group discussion.

For the small group discussion period, we will have a set of potential topics based on our guiding questions from section 1, the position abstracts we receive, and proposed topics by attendees. Participants will be able to self-select into a small group of their choice based on the topic they would most like to discuss; virtual participants will be able to join breakout rooms on the Zoom session and utilize SAGE3 boards to discuss, and some in-person organizers will join the virtual breakout rooms to ensure that the virtual participants are included and their findings presented in the full group discussion even in the case of tech issues, such as internet connectivity difficulties. For the full group discussion period, we will give each small group a chance to present their main findings and solicit feedback from the rest of the attendees, which should lead to a fuller understanding of the challenges and opportunities facing computational notebook research for all involved.

Finally, after the full group discussion the organizers will make closing comments and wrap up the workshop.

6.3 Workshop Planning Dates

We have tentatively made the following plan to ensure this workshop is a success:

- Call for Papers released: December 14, 2023
- Paper Submission Deadline: February 22, 2024
- Paper Author Notification: March 15, 2024
- Paper Camera-Ready Deadline: April 1, 2024
- Speaker Schedule Available: April 8, 2024
- Deadline for Position Abstracts & Demos: April 15, 2024

This schedule can be adjusted as needed.

7 POST-WORKSHOP PLANS

In addition to publishing the accepted materials as workshop proceedings, we are considering post-workshop plans. Based on participants' interests and discussion outcomes, we may organize additional activities, such as:

- Crafting a paper on challenges and opportunities based on the discussions from this workshop. Participants can volunteer to help author this paper.
- Organizing a special issue in a journal to invite expanded versions of papers in a venue such as ACM TIIS, ACM TIST, ACM TOCHI, ACM Data Science, or JOVI.
- Taking what we learn from this workshop to conduct outreach workshops and surveys at appropriate venues that bring together computational notebook researchers, developers and end users.

8 CALL FOR PARTICIPATION (250 WORDS)

Calling all computational notebook and literate programming researchers! We invite you to participate in a hybrid workshop at CHI aimed at bringing together researchers at the intersection of computational notebooks and HCI research, especially as it relates to interface and interaction design. We have multiple avenues for participation including papers, demonstrations, lightning talks, and attendance. To attend, please at a minimum fill out the form below with a position abstract related to computational notebooks and human-computer interaction. Position abstracts should be limited to 500 words, excluding references, and allow, but do not require, a presentation. Paper submissions should be of a length reflecting the contribution and in the ACM Conference Proceedings Primary Article Template format. Demo submissions should include an online portal(s) (e.g. website, YouTube videos, etc.) showcasing what the submitter wishes to demonstrate. All submissions should be related to computational notebook research and computer-human interaction. Paper Submissions will be peer reviewed by a minimum of one organizer and two knowledgeable reviewers recruited by the organizers. Accepted papers and demos will be published on the workshop website. At least one author of each accepted submission must attend the workshop, either in person or online. All participants must register for both the workshop and for at least one day of the conference.

More detailed information is located at our workshop website: link

Form link: link2

REFERENCES

[1] Souti Chattopadhyay, Ishita Prasad, Austin Z Henley, Anita Sarma, and Titus Barik. 2020. What's Wrong with Computational Notebooks? Pain Points, Needs, and Design Opportunities. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–12.

[2] Rebecca Faust, Carlos Scheidegger, Katherine Isaacs, William Z Bernstein, Michael Sharp, and Chris North. 2022. Interactive Visualization for Data Science Scripts. In *2022 IEEE Visualization in Data Science (VDS)*. IEEE, 37–45.

[3] Jesse Harden, Elizabeth Christman, Nurit Kirshenbaum, John Wenskovitch, Jason Leigh, and Chris North. 2022. Exploring Organization of Computational Notebook Cells in 2D Space. In *2022 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*. IEEE, 1–6.

[4] Andrew Head, Fred Hohman, Titus Barik, Steven M Drucker, and Robert DeLine. 2019. Managing messes in computational notebooks. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12.

[5] Mary Beth Kery and Brad A Myers. 2018. Interactions for untangling messy history in a computational notebook. In *2018 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*. IEEE, 147–155.

[6] Mary Beth Kery, Marissa Radensky, Mahima Arya, Bonnie E John, and Brad A Myers. 2018. The story in the notebook: Exploratory data science using a literate programming tool. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–11.

[7] Mary Beth Kery, Donghao Ren, Fred Hohman, Dominik Moritz, Kanit Wongsuphasawat, and Kayur Patel. 2020. mage: Fluid moves between code and graphical work in computational notebooks. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*. 140–151.

[8] Donald Ervin Knuth. 1984. Literate programming. *The computer journal* 27, 2 (1984), 97–111.

[9] Xingjun Li, Yuanxin Wang, Hong Wang, Yang Wang, and Jian Zhao. 2021. Nbsearch: Semantic search and visual exploration of computational notebooks. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–14.

[10] Xingjun Li, Yizhi Zhang, Justin Leung, Chengnian Sun, and Jian Zhao. 2022. EDAssistant: Supporting Exploratory Data Analysis in Computational Notebooks with In-Situ Code Search and Recommendation. *ACM Transactions on Interactive Intelligent Systems (TiiS)* (2022).

[11] Jorge Piazentin Ono, Juliana Freire, and Claudio T Silva. 2021. Interactive data visualization in jupyter notebooks. *Computing in Science & Engineering* 23, 2 (2021), 99–106.

[12] Fernando Perez and Brian E Granger. 2015. Project Jupyter: Computational narratives as the engine of collaborative data science. *Retrieved September* 11, 207 (2015), 108.

[13] Jeffrey M Perkel. 2018. Why Jupyter is data scientists' computational notebook of choice. *Nature* 563, 7732 (2018), 145–147.

[14] Adam Rule, Aurélien Tabard, and James D Hollan. 2018. Exploration and explanation in computational notebooks. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–12.

[15] Jeremy Singer. 2020. Notes on notebooks: Is Jupyter the bringer of jollity?. In *Proceedings of the 2020 ACM SIGPLAN International Symposium on New Ideas, New Paradigms, and Reflections on Programming and Software*. 180–186.

[16] Nathaniel Weinman, Steven M Drucker, Titus Barik, and Robert DeLine. 2021. Fork It: Supporting Stateful Alternatives in Computational Notebooks. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–12.

[17] John Wenskovitch, Jian Zhao, Scott Carter, Matthew Cooper, and Chris North. 2019. Albireo: An interactive tool for visually summarizing computational notebook structure. In *2019 IEEE Visualization in Data Science (VDS)*. IEEE, 1–10.

[18] Yifan Wu, Joseph M Hellerstein, and Arvind Satyanarayan. 2020. B2: Bridging code and interactive visualization in computational notebooks. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*. 152–165.