README.md: A Tutorial on Reproducible Visualization Research

Velitchko Filipov[1] and Tobias Isenberg[2] and Alexander Lex[3,4]

[1]TU Wien, Austria [2] Université Paris-Saclay, CNRS, Inria, LISN, France [3] TU Graz, Austria [4] University of Utah, USA

Abstract

Reproducibility is fundamental to scientific progress, yet visualization research remains difficult or impossible to reproduce due to missing code, inaccessible URLs, undocumented dependencies, and incomplete documentation. This tutorial provides practical guidance for reproducible visualization research across development and empirical studies. We address common barriers including resource constraints, disclosure concerns, and publication anxieties, and present workflows for version control, documentation, dependency management, and deployment. Participants will learn to structure projects with reproducibility in mind, engage with the Graphics Replicability Stamp Initiative, and learn about frameworks such as reVISit 2 for replicable empirical studies. Participants will be supported by interactive activities and an open repository of reusable resources.

1. Introduction

Reproducibility is fundamental to scientific progress, yet most published visualization research, including recent works from 2024 or 2025, remains difficult or impossible to reproduce. These reproducibility challenges come up in visualization research in multiple contexts, i.e., software artifacts (tools, systems, and implementations), empirical studies (user experiments and evaluations), and the research process itself (design decisions, iterations, and rationale). There are commonalities between these contexts, such as missing code, broken URLs, and undocumented dependencies, however, they also present distinct challenges that require different combinations of technical and documentation-related methods and practices. In this tutorial, we would like to stress practical, low-overhead, and accessible approaches to setting up reproducible environments, papers, and results for academics. We hope to highlight the benefits of incorporating this into your research workflows and how quickly these can pay off. With that our **goal** is to present why reproducibility matters, how you can approach it through practical approaches and resources and demonstrate user study replicability. We realize that there are real reason why most people do not develop and publish research with reproducibility in mind and have organized these into three categories that we intend on addressing:

Resource Constraints: Excessive time commitment and resources can be managed with incremental steps and minimal documentation. Maintenance burdens can be reduced and eliminated with practical versioning and dependency packaging approaches.
Disclosure Concerns: Anonymity requirements can be addressed via technological and organizational means, including anonymous repositories and embargoes. Proprietary data can be addressed by providing synthetically generated data or partial reproducibility.
Publication Issues: Concerns about messy code are framed with realistic expectations, any code is better than none. Anxiety about competition are balanced by citation benefits [Har18]. To make replication studies impactful, Quadri and Rosen [QR19] propose re-evaluation, expansion, and specialization strategies.
We believe reproducibility can be categorized into three distinct levels: (i) None - which encompasses a complete lack of any methods to reproduce the contributions cited by Fekete and Freire as the most common [FF20]; (ii) Minimal reproducible publication - which describes a minimal set of reproducible components achievable by everyone with low-overhead; and (iii) Complete - which is the highest effort reproducibility level, where each aspect of the publication is completely reproducible. Within this 100-minute tutorial we provide participants with information, resources, tools, and workflows to make their research reproducible as a minimal reproducible publication with low-overhead. We present a range of workflows, introduce the Graphics Replicability Stamp Initiative (GRSI), and present case studies using reVISit 2. Participants are provided resources and templates participants can apply immediately. The tutorial targets graduate students, early-career researchers, and experienced researchers seeking better practices.

2. Motivation

The visualization community faces a reproducibility crisis [FF20]. The community has responded to this challenge. Since 2018, there has been a focus on reproducibility in visualization research. Researchers have examined threats to study validity [KH18], developed guidelines for designing unbiased replication studies [SM18], proposed methods for embedding replications into novel work [QR19], analyzed practices for open data and material sharing [Har18], proposed typologies of reproducibility aspects for visualization design [FRK23], and introduced lightweight methods for embedding pipeline state in images [Rei23]. Recently, reVISit 2 emerged as a framework supporting the full experiment lifecycle from design

DOI: 10.2312/evt.20261000

Table 1: *Preliminary Tutorial Schedule*

Time	Duration	Content
9:00–9:05	~5 min	Welcome & Introduction
9:05–9:25	~20 min	**Part I: Why?**
9:25–9:45	~20 min	**Part II: How?**
9:45–10:05	~20 min	**Part III: Replicate!**
10:05–10:20	~30 min	Discussion (Q&A)
10:20–10:25	~5 min	Comments & Closing

through dissemination [DWS*23, CWS*26]. The system facilitates reproducible study workflows through domain-specific languages, provenance tracking, and participant replay and received a best paper award at IEEE VIS 2025 [CWS*26]. Isenberg's [Ise24] analysis of GRSI-certified papers quantifies how few published works meet reproducibility standards. The amount of recent work on the topic shows the topic is relevant and highlights the urgency and the community's commitment to addressing it. Meyer and Dykes' [MD20] criteria for methodological rigor in design studies established transparency as fundamental, positioning reproducibility as part of a broader movement toward rigorous research practice across visualization's diverse methodologies. Fekete and Freire [FF20] outline that most researchers cannot recreate their own results from two years ago without proper practices. Scripted workflows enable reproducible results and comparison with previous version. Well-documented code provides great benefits to the academic community that would like to use, build upon and extend existing techniques. Maintaining reproducible research produces comprehensive appendices that enhance paper value, lead to more citations and wider adoption [Ise24]. As AI is making its way through and into visualization research, it is further compounding the reproducibility crisis. A noticable trend over the last few years is the increasing amount of research done with LLMs, presenting results that are no longer obtainable due to the rapid development and expansions of the available models' capabilities. Researchers need reliable and reproducible baselines to advance the state of the art.

3. Organization

Velitchko Filipov is a postdoctoral researcher in the Visual Analytics research unit at TU Wien. His research interests include information visualization and visual analytics of dynamic graphs and networks, focusing on novel representations, interactions, and engaging methods for interactive visual analysis.
Tobias Isenberg is a senior research scientist at Inria, also affiliated with the Université Paris-Saclay. His research focuses on interactive scientific visualization, illustrative rendering, and human-computer interaction. He is pursuing the reproducibility of his work via open code, OSF repositories, and GSRI stamp applications [Ise22, Ise24].
Alexander Lex is a professor of Human Computer Interaction at TU Graz, Austria and affiliated with the University of Utah. He and his lab design, build, and study interactive visual data analysis to understand and communicate complex data sets. He also investigates how to make visualization reproducible and how to integrate visualizations with computational workflows. [DWS*23, CWS*26].

4. Content

The tutorial is structured in three 30-minute parts (Table 1). We will use interactive polling to gauge audience experience and identify common issues they face. We provide a GitHub repository containing all resources from the workshop including: templates, checklists, and example projects. References to key papers, a comparison between platforms for disseminating code and results. For GRSI applicants, we present the process and practical tips to apply.

Part I: Why Reproducibility Matters?

We start with a poll: "*Can you reproduce your work from two years ago?*". Here, we establish personal and community stakes, quantifying the "costs" in wasted effort. It is estimated that researchers spend 20-30% of their time attempting (and often failing) to reproduce others' work [FRK23], representing massive wasted effort and slowed scientific progress. We establish clear terminology distinguishing reproducibility (`same data` + `code` → `same results`) [KH18] from replicability (`different data` + `same method` → `consistent findings`) [SM18], drawing on established definitions from the visualization community [FF20, Ise24]. The core of this session demonstrates why reproducibility is important to consider from the start. We present scenarios and show-case the life-cycle of a typical publication and the stakeholders that are affected by the lack of reproducible methods, results, and approaches. We end this session with an interactive walk-through of how a few minutes can be used to structure a low-overhead reproducible environment. We present some practical workflows that should quickly pay off rather than spending countless hours retrofitting your work after publication. We cover the following aspects:

Version Control: Managing research artifacts throughout the project lifecycle is fundamental to reproducibility. Without a form of version control, there is no way to recover the exact code states that produced published results and they can also be used to document design choices, justifications, and trade-offs. These can help resolve issues like "it worked on my machine last month."
Documentation: Writing effective `README.md` files that guide strangers (including future you) through installation, execution, and deployment, plus `CONTRIBUTING.md` for extending the codebase. We cover meaningful comments that explain decisions and trade-offs rather than obvious functions and code-blocks, helping others understand why the code does what it does.
Dependency Management: Capturing exact environments through package manifests (`requirements.txt, package.json, environment.yml`) and `lock-files` for version pinning. We introduce containers (i.e., Docker, Singularity) for complete environment setup that helps reviewers (and others) to reproduce results without dependency conflicts. *(Note: These are not a catch-all solution, however, they bring many benefits when it comes to being able to quickly setup and reproduce visualization research.)*
LLM & AI Reproducibility: Documenting AI-assisted research is an emerging challenge. We cover prompt versioning frameworks, conversation archival (shared links and exports), and best practices for reporting model details (i.e., models, timestamps, and parameters). As LLMs assist with code generation and analysis in visualization research, proper documentation is essential for reproducibility.

Deployment Platforms: Deploy interactive visualizations through static hosting (GitHub Pages, Netlify), full-stack platforms (Vercel, Render), notebook hosting (Binder, Observable), and temporary sharing tools (ngrok) for review and demonstration.

Part II: How to Approach Reproducible Research

We focus on navigating the GRSI process and leveraging available infrastructure to make reproducibility achievable and beneficial for you. We outline the benefits of pursuing GRSI recognition, such as improved citation rates, enhanced credibility during review, development of better research habits, and increased visibility. Building reproducibility into workflows (e.g., by following the steps from **Part I**) from the start makes the GRSI process straightforward.

We introduce key platforms that can be used for packaging all research artifacts together (i.e., the manuscript, repository/code, supplemental materials, data from the evaluation/analysis), such as Zenodo, OSF, Software Heritage for long-term archival, AsPredicted and OSF for study pre-registration, and arXiv, Figshare, Dryad for pre-print dissemination and establishing precedence. The core of this session will be to provide a practical approach (and guidance) to navigate the GRSI process.

Strategic Planning: The challenge and expensive part about retrofitting reproducibility after publication is that you are attempting to reconstruct decisions made months (even years) ago, discovering that your packages are outdated, and the intermediate version of a dataset you used no longer exists. To address this issue, we recommend to start with reproducibility during research rather than after publication to improve organization, collaboration, revision handling, and quality throughout the project. This can be immensely effective and save valuable time and resources. Investing a few minutes to setup the structure and initialize the project, and a few minutes each time you need a new result or a new asset (e.g., figure) to ensure that is reproducibly quickly pays off.
Paper Documentation: Provide brief statements in main text of the paper and a detailed appendix with structure and availability of the repository or other supplemental materials. Scripting can also be used to manage the reported numbers in publications.
Reviewers' Expectations: Reviewers need to be able to reproduce the key results with reasonable effort. We will illustrate a process for how to achieve an executable system/approach.
Common Issues: More often than not, there are typical issues that make most approaches challenging to reproduce, such as missing dependencies, undocumented pre-processing, data (& schema) availability issues, missing installation instructions and hardware specification, and version drift between code and publication.
Making it Easy: Provide quickstart scripts, map figures to specific commands (e.g., bash scripts), offer multiple complexity levels (demo/partial/full), pre-test on fresh systems and/or containers.

Part III: Replicating Studies in Visualization Research

This part focuses on replicability challenges specific to empirical visualization research, particularly perceptual studies. While sharing code and data supports reproducibility, replicating findings requires access to experimental designs, interaction logic, and participant behavior that are difficult to fully specify in papers. The core of this session demonstrates how replicability can be achieved in practice using dedicated infrastructure.

Replicability in Perceptual Studies: We clarify what replication means in visualization experiments, including re-running studies with new participants, modifying conditions, or extending prior work while preserving the original experimental logic.

Experimental Structure: Perceptual studies depend on tasks, stimuli, timing, counterbalancing, and interaction constraints. We discuss why incomplete specification of these elements is a major barrier to replication and extension.

Provenance and Replay: We highlight the importance of recording interactions and study execution to allow detailed inspection of individual sessions and verification of reported results.

reVISit 2: We present reVISit 2 [DWS*23, CWS*26], a framework supporting experiment design, automatic provenance tracking, and participant replay. The system enables researchers to explore and replicate actual study executions rather than relying solely on textual descriptions. We plan a live walkthrough of an existing perceptual study replicated using reVISit 2 demonstrating how the framework lowers barriers to replication, supports extensions of prior work, and enables transparent verification of experimental results.

5. Intended Outcomes

The tutorial concludes with an open discussion in which participants share experiences, challenges, and open questions related to reproducibility and replicability. All materials, including resources, checklists, guides, and practical suggestions, will be collected in an openly accessible repository on GitHub. The repository is designed as a living resource that can be extended by community members with additional tools, practices, and examples they commonly use. As visualization research spans diverse methods and contexts (e.g., information visualization, scientific visualization, virtual reality), it does not present an exhaustive set of resources and we aim to extended it through continued contributions. Furthermore, due to the nature of the various libraries, languages, platforms, and frameworks commonly employed it is challenging to provide a single script or method to deliver for all of these. We intend to present reproducible practices, processes, and guides on how this can be addressed rather than an exhaustive list for each possible technology stack.

6. Intended Audience & Equipment

The intended audience comprises researchers who are concerned with the replicability and transparency of their work. While the session is particularly relevant for PhD students and early-career researchers, it is equally aimed at established researchers who supervise students and shape methodological practices, offering concrete ideas they can integrate into their own research and mentoring. A standard seminar room equipped with a projector is sufficient. No special equipment or software is required. The expected audience size is approximately 20 participants.

Acknowledgements: This work was funded by the FWF WEAVE project SANE [10.55776/I6635] and ArtVis [10.55776/P35767]. Open access funding provided by Technische Universität Wien/KEMÖ.

References

[CWS*26] CUTLER Z., WILBURN J., SHRESTHA H., DING Y., BOLLEN B., NADIB K. A., HE T., MCNUTT A., HARRISON L., LEX A.: ReVISit 2: A full experiment life cycle user study framework. *IEEE Trans Vis Comput Graph* (2026). To appear. doi:10/hbkxwp.

[DWS*23] DING Y., WILBURN J., SHRESTHA H., NDLOVU A., GADHAVE K., NOBRE C., LEX A., HARRISON L.: reVISit: Supporting scalable evaluation of interactive visualizations. In *Proc. Visualization* (2023), IEEE Computer Society, Los Alamitos, pp. 31–35. doi:10/hbdjdh.

[FF20] FEKETE J.-D., FREIRE J.: Exploring reproducibility in visualization. *IEEE Comput Graph Appl 40*, 5 (2020), 108–119. doi:10/ghd59m.

[FRK23] FRANKE M., REINA G., KOCH S.: Toward reproducible visual analysis results. In *Proc. PacificVis* (2023), IEEE Computer Society, Los Alamitos, pp. 102–106. doi:10/gtw42q.

[Har18] HAROZ S.: Open practices in visualization research: Opinion paper. In *Proc. BELIV* (2018), IEEE Computer Society, Los Alamitos, pp. 46–52. doi:10/gtw4sp.

[Ise22] ISENBERG T.: Personal experiences of providing and using research prototypes. In *Proc. VisGap* (2022), Eurographics Association, Goslar, Germany, pp. 17–22. doi:10/gr2d83.

[Ise24] ISENBERG T.: The state of reproducibility stamps for visualization research papers. In *Proc. BELIV* (2024), IEEE Computer Society, Los Alamitos, pp. 97–105. doi:10/nt3t.

[KH18] KOSARA R., HAROZ S.: Skipping the replication crisis in visualization: Threats to study validity and how to address them : Position paper. In *Proc. BELIV* (2018), IEEE Computer Society, Los Alamitos, pp. 102–107. doi:10/gtw4sq.

[MD20] MEYER M., DYKES J.: Criteria for rigor in visualization design study. *IEEE Trans Vis Comput Graph 26*, 1 (2020), 87–97. doi:10/gf6724.

[QR19] QUADRI G. J., ROSEN P.: You can't publish replication studies (and how to anyways). In *Proc. IEEE Workshop Vis X Vision* (2019). Available as arXiv preprint 1908.08893. doi:10/gtxgh9.

[Rei23] REINA G.: Can image data facilitate reproducibility of graphics and visualizations? Toward a trusted scientific practice. *IEEE Comput Graph Appl 43*, 2 (2023), 89–100. doi:10/gtw4nt.

[SM18] SUKUMAR P. T., METOYER R.: Towards designing unbiased replication studies in information visualization. In *Proc. BELIV* (2018), IEEE Computer Society, Los Alamitos, pp. 93–101. doi:10/gtw4sr.