

# The State of Reproducibility Stamps for Visualization Research Papers

Tobias Isenberg \*

Université Paris-Saclay, CNRS, Inria, LISN, France

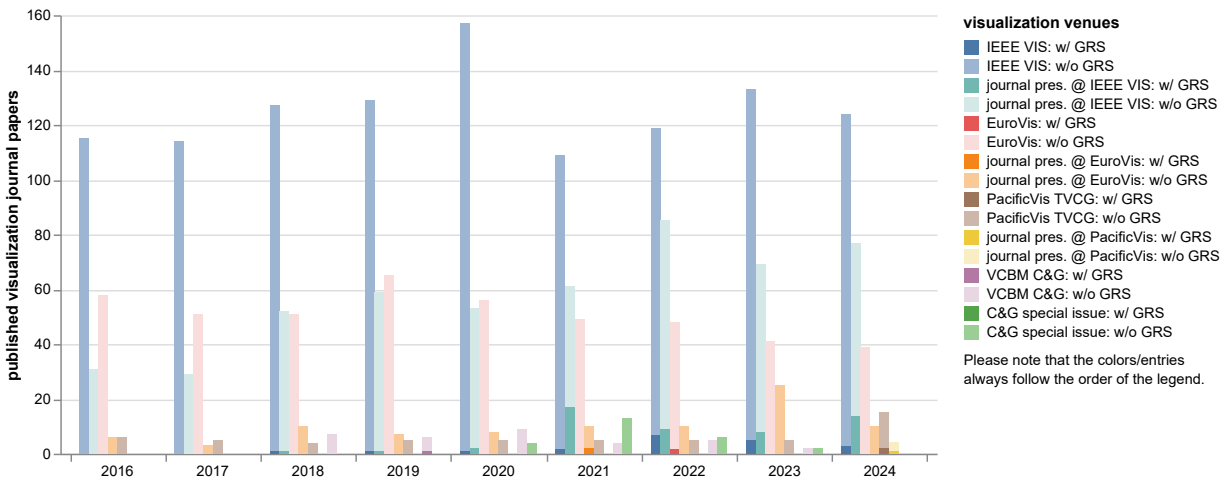


Figure 1: Evolution of publication counts in several visualization venues,<sup>14,15</sup> including their proportion of papers with GRS certification (for a normalized version with percentages see Figure 8). We can see that only few papers so far have received a GRS, in all venues.

## ABSTRACT

I analyze the evolution of papers certified by the Graphics Replicability Stamp Initiative (GRSI) to be reproducible, with a specific focus on the subset of publications that address visualization-related topics. With this analysis I show that, while the number of papers is increasing overall and within the visualization field, we still have to improve quite a bit to escape the replication crisis. I base my analysis on the data published by the GRSI as well as publication data for the different venues in visualization and lists of journal papers that have been presented at visualization-focused conferences. I also analyze the differences between the involved journals as well as the percentage of reproducible papers in the different presentation venues. Furthermore, I look at the authors of the publications and, in particular, their affiliation countries to see where most reproducible papers come from. Finally, I discuss potential reasons for the low reproducibility numbers and suggest possible ways to overcome these obstacles. This paper is reproducible itself, with source code and data available from [github.com/tobiasisenberg/Visualization-Reproducibility](https://github.com/tobiasisenberg/Visualization-Reproducibility) as well as a free paper copy and all supplemental materials at [osf.io/mvnbj](https://osf.io/mvnbj).

**Keywords:** Research replicability, research reproducibility, research repeatability, FAIR research data, open science, open research, open practices, Graphics Replicability Stamp Initiative (GRSI), visualization.

## 1 INTRODUCTION

“A scientific result is not truly established until it is independently confirmed” [2]—yet in the past this mantra has not really been the guiding principle for work within visualization. BELIV as a venue has looked at this question specifically in its 2018 edition [13, 20, 25, 27], and found that there is still much to be improved in our field.

\*e-mail: given\_name.family\_name@inria.fr

Two years later in 2020, Fekete and Freire [10] observed that “there are few visualization articles with a [graphics replicability] stamp (6 for TVCG, 4 for CGF), and most of these are about computer graphics.” They referred to the badges (which, in this paper, I call graphics replicability<sup>1</sup> stamps, GRS) awarded by the Graphics Replicability Stamp Initiative (GRSI; [replicabilitystamp.org](https://replicabilitystamp.org)) to published journal<sup>2</sup> papers within the larger computer graphics research field for ensuring that some of the presented work can indeed be replicated by independent others. As the GRSI started from within the computer graphics community and had only started to award papers in 2017, only few authors—in particular within the visualization community—up to 2020 really had the time to take advantage of this independent verification process, so the low numbers at the time are no surprise. So now, in 2024, with a lot of past and ongoing discussion within our field on how we can make work more reproducible, I think it is time to take another look and check again, to see if the status quo has improved since then.

Below I describe my work to collect the respective data, from the GRSI website, from the various digital libraries, from conference venues, and from other sources. Based on this data I then set out to analyze not only the GRSI-awarded work overall but also what part of it covers visualization work, how the various publication and presentation venues differ from each other, and how reproducibility is fostered in the different home countries of the researchers.

For full disclosure I note that I am independent<sup>3</sup> of the GRSI, but that I am personally interested in fostering reproducibility and replicability in the field. My collaborators and I did apply for and received several stamps from the GRSI for our work<sup>4</sup> (and also

<sup>1</sup>See the discussion of terminology in Section 2.

<sup>2</sup>The GRSI started with only certain journals as eligible publication platforms, but in 2024 also added SIGGRAPH and SIGGRAPH Asia conference-only papers as possible venues for papers to receive a GRS. In this paper I still often use “journal” to refer to the *publication* venues awarded by the GRSI, in particular since within visualization it is still true as of today. But notice that in the future other non-journal venues may be added.

<sup>3</sup>But my wife currently serves as a reviewer for the GRSI.

<sup>4</sup>[tobias.isenberg.cc/reproducibility](https://tobias.isenberg.cc/reproducibility)

encouraged others to do so)—and I thus also personally contributed to the numbers I report below, in the sense of trying to increase them.

## 2 TERMINOLOGY AND SCOPE

Unfortunately, the terminology with respect to being able to check the correctness of scientific results is far from clear [21], in particular in the GRSI case. Originally, *reproducibility* was defined to mean that one runs author-provided programs or tools to re-create the same results (e.g., images) using the author-provided input data [5], while *replicability* refers to “writing and then running new software based on the description of a computational model or method provided in the original publication, and obtaining results that are similar enough to be considered equivalent” [24]. The ACM, however, initially used the exact opposite definition of the terminology: *reproducibility* to mean “different team, different experimental setup” and *replicability* to mean “different team, same experimental setup” [2, 21],<sup>5</sup> adding the notion of *repeatability* for a re-creation of the same experimental setup by the original team (also see the discussion of the terminology in the experimental sciences by Plesser [21]). Luckily, as of August 24, 2020, the ACM aligned their definitions with the original ones, i. e., using *reproducibility* to mean “different team, same experimental setup” and *replicability* “different team, different experimental setup.”<sup>6</sup> As the Graphics *Replicability* Stamp Initiative was established in 2016, it likely followed the original definitions of the ACM and thus, despite using the term *replicability*, its focus is to verify independently (i. e., “different team”) that the author-provided code and data makes it possible to re-create the same results as the authors presented in the paper (i. e., “same experimental setup”). In this paper, however, to avoid further confusion I follow ACM’s updated notion as I discuss the independent verification of the author’s code by the GRSI with (usually) the authors’ data, using the *reproducibility* terminology—even in the title of this paper.

## 3 RELATED WORK

What motivates this discussion of reproducibility is the realization that way we traditionally conduct and report scientific research is subject to a *replication crisis* [6, 20]—work, once published, is rarely questioned or re-checked. Kosara [19] made this point for the field of visualization by describing it as “an empire built on sand,” giving many examples of where established canon should be questioned and re-examined. Yet in our field’s review process such replications often are considered to be of too low novelty to be publishable [22], such that it requires extensions of pure replications for making the cut as Quadri and Rosen [22] and even the VIS 2024 overall paper chairs<sup>7</sup> recommend. Kindlmann<sup>8</sup> at the VIS panel “Is There Science in Visualization?” already in 2006 argued the same point, and pointed to code availability for published papers as a pre-requisite.

In the past, this code availability that makes reproducibility and replicability possible has not been a core focus of authors. For the computer graphics field, Bonneel et al. [3] analyzed the reproducibility<sup>9</sup> of results published in ToG-level papers from SIGGRAPH 2014, 2016, and 2018, not only manually checking for available code but also rating its ease of reproduction.<sup>10</sup> While they found an overall increase of papers with code from  $\approx 30\%$  in 2014 to  $>50\%$  in 2018, at the time of publication only 5 papers had a GRSI-certified reproducibility, yet 4 of these with a top reproducibility rating. For the

<sup>5</sup>[acm.org/publications/policies/artifact-review-badging](https://acm.org/publications/policies/artifact-review-badging)

<sup>6</sup>[acm.org/publications/policies/artifact-review-and-badging-current](https://acm.org/publications/policies/artifact-review-and-badging-current)

<sup>7</sup>[ieevis.org/year/2024/blog/vis-2024-OPC-blog-replication](https://ieevis.org/year/2024/blog/vis-2024-OPC-blog-replication)

<sup>8</sup>[people.cs.uchicago.edu/~glk/talks/pdf/Kindlmann-ScienceInVisualization-VIS-2006-talk.pdf](https://people.cs.uchicago.edu/~glk/talks/pdf/Kindlmann-ScienceInVisualization-VIS-2006-talk.pdf)

<sup>9</sup>Also here, different from the paper, I use ACM’s updated terminology.

<sup>10</sup>The project’s database has since been extended to also include SIGGRAPH 2019–2021 as well as SIGGRAPH Asia 2015, at least partially.

visualization field, Haroz [13] analyzed IEEE VIS 2017 papers<sup>11</sup> and, among other things, checked for the availability of source code that was used for data collection, the availability of the raw collected data, and the availability of materials or source code needed to reproduce the data analysis, with overall low results. Some notable exceptions with good reproducibility were highlighted by Kosara and Haroz [20] in the same year. Looking at GRSI-certifications in 2020, Fekete and Freire [10] only found few visualization papers with a GRS and call for more reproducibility and replicability in the field, in particular within the GRSI. Overall, however, few if any reliable numbers on how much visualization paper authors emphasize on reproducibility exist at this point—which is what I add with my analysis of GRSI-awarded papers in this work.

But even with the few numbers on the support of reproducibility in our field that have been published so far we already know that we need to do better. Several groups of authors have discussed ways to improve the situation for various visualization subfields [4, 8, 25, 27] or how to include the subject into our taught curricula [26]. More generally, Reina [23] recommends and shows how to embed all information needed to reproduce a visual result (image) into the image itself. Another way that does not require source code is to provide executable demonstrators [3, 15], which avoid, in particular, the compilation issues of code [3]. Other than the replicability of specific approaches, Garkov et al. [12] argue that also the curation and long-term management of dedicated research datasets and artifacts can also be a means to facilitate research repeatability because, for instance, both ensure that new approaches can easily be compared to older results. Fekete and Freire [10] discuss possibilities to ensure reproducibility specifically tailored to the different contribution types we have in the visualization field, while Franke et al. [11] offer a typology of reproducibility aspects that allows paper authors to decide on reproducibility and replicability aspects based on their specific situation. There also used to be the EuroRV<sup>3</sup> workshop<sup>12</sup> as dedicated venue to discuss the topic, yet its last incarnation happened as TrustVis five years ago in 2019 and not again since then.

I want to emphasize that there are many aspects of the larger open science movement (e. g., [16]) that are important in this context. While work that focuses on empirical studies, for example, may require a certain type of shared material [20] (e. g., pre-registrations or registered reports), in this paper I am looking generally at the issue of to what degree reproducibility of some sort of source code aspects (be it code to produce visual results or code to reproduce the statistical analysis of experimental data from a user experiment) is actually done in our field using the GRSI as the only certification authority available to our community at the moment.

## 4 DATA ACQUISITION

To do so, I started by getting the data about reproducible papers from the listing on [replicabilitystamp.org](https://replicabilitystamp.org), which—as of Aug. 7, 2024—lists 346 total papers. Using Python code with the Beautiful-Soup library, I extracted each paper’s title, its authors, the journal, the DOI, and the active and archived repository locations. I then sanitized the downloaded data by, in particular, correcting author names as much as possible (needed due to typos, missing special characters, first name last name order, name changes from marriages, and other errors), resulting in the identification of 1244 unique authors. For each paper I then queried the digital libraries (i. e., IEEE Xplore, Elsevier, ACM, or Crossref) to obtain publication data (publication year, volume, number, pages, etc.) and I manually added the authors’ respective country of affiliation based on what authors reported on the published papers (often effective at the time of submission or publication; if multiple countries were named I added them all).

Next, I classified each paper as being a visualization paper or not. Because a manual or keyword-based classification is inherently bi-

<sup>11</sup>Later extended to IEEE VIS 2018–2019: [oaviv.org](https://oaviv.org).

<sup>12</sup>[diglib7.eg.org/handle/10.2312/980](https://diglib7.eg.org/handle/10.2312/980)

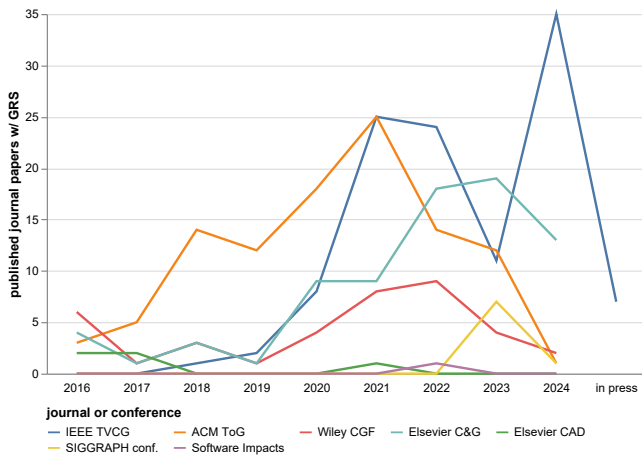


Figure 2: Overall development of papers with GRS, by publication venues (and their article *publication* years). For a bar chart version see Figure 14 in Appendix C.

ased, I attempted to use the paper authors’ own classification as much as possible. I thus used data about whether a paper was *presented* in a clearly visualization-focused venue, i. e., proper (journal-level) IEEE VIS/EuroVis/PacificVis papers,<sup>13</sup> journal papers presented at IEEE VIS/EuroVis/PacificVis, etc.,<sup>14</sup> or if the papers were published in a journal special issue focused on visualization topics (e. g., special-issue versions of VCBM, EuroVA, MolVA, EnvirVis in Elsevier’s C&G).<sup>15</sup> If this failed, I also checked the titles of the remaining papers for a number of visualization-related keywords (9×; see the list of keywords in Appendix A). Finally, I also manually classified a small number of papers based on an inspection (9×; see Appendix B for the rationales). As noted, these last two ways of classification are arguably biased, but it affected a relatively low number of papers (18 out of 98 papers classified as visualization papers in total, or 18.4%). This process resulted in 28.3% of all GRSI-awarded papers being classified as discussing visualization topics (98 out of 346), with 347 authors out of 1244 being classified as visualization authors (i. e., ≥50% of their papers being classified as visualization papers).

## 5 A VISUAL GRSI DATA ANALYSIS

I begin by looking at the temporal development of GRSI-awarded papers overall.<sup>16</sup> Figure 2 shows that the first stamps were awarded for papers published in 2016, with the major contributors being the *ACM Transactions on Graphics* (ToG) and the *IEEE Transactions on Visualization and Computer Graphics* (TVCG), followed by Elsevier’s *Computers & Graphics* (C&G) and Wiley’s *Computer*

<sup>13</sup>The conference data comes from the VisPubData dataset [14] (for IEEE VIS), from the Eurographics digital library (for EuroVis; including both regular papers and STAR papers), or was extracted from the conference webpage (for PacificVis).

<sup>14</sup>I extracted this data largely from the respective conference webpages or programs. The data of regular papers and presentations at IEEE VIS is updated up to the conference’s 2024 edition courtesy of Ross Maciejewski, Lane Harrison, and the VIS Papers Chairs, but the IEEE VIS 2024 data is preliminary at this point and there may be some changes as the final program for IEEE VIS 2024 is being finished (i. e., for the 2024 values in Figure 1, in Figure 8, and in Figure 13).

<sup>15</sup>I extracted this data from the special issue classifications in Elsevier’s digital library and from special issue overview articles in C&G. Please note that the 2023 and 2024 values of these types of publications are not yet final (i. e., the values for the bars in Figure 1, in Figure 8, and in Figure 13 are too low or do not exist yet), as they are still being assembled.

<sup>16</sup>Unlike the statistics on the GRSI website ([reproducibilitystamp.org/statistics.html](https://reproducibilitystamp.org/statistics.html)), I look at the temporal evolution of the *publication* data, not when a paper was awarded the GRS.

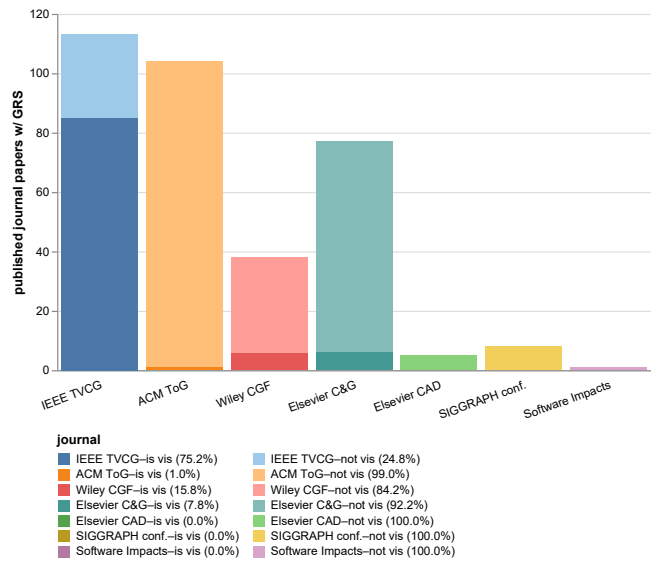


Figure 3: GRS papers by publication venue and their respective proportion classified as visualization papers. For a normalized version of the plot see Figure 15 in Appendix C.

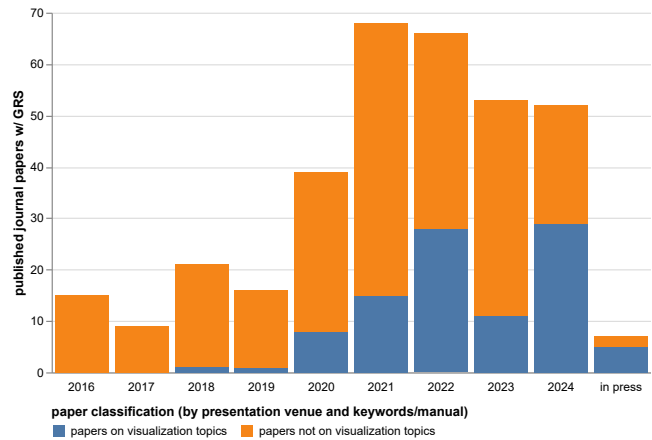


Figure 4: Overall proportion of GRS papers classified as visualization papers by article *publication* years. For a normalized version of the plot see Figure 17 in Appendix C.

*Graphics Forum* (CGF). Figure 3 shows that TVCG and ToG are at about the same level, with C&G and CGF having lower numbers, the remaining journals and the recently added SIGGRAPH conference papers do not (yet) play a big role. While it is normal that more recent years have lower numbers than those volumes a bit older (see also the total paper trend in Figure 4), what’s interesting about Figure 2 is the distinct drop in numbers for TVCG in 2023 and the following rise in 2024. The likely reason is that in 2023, for some reason, the promise by TVCG to assign a GRSI-awarded paper to the next available issue of the journal<sup>17</sup> somehow did not get upheld, and only after a friendly reminder to TVCG did several “in press” articles get fully published in early 2024. As such the 2023 and 2024 numbers for TVCG are likely to be outliers—a view that is supported by the more or less constant submission and acceptance numbers for TVCG articles as it is evident in the GRSI award time statistics that the GRSI publishes online.<sup>16</sup>

Looking next at the development of visualization contributions among all GRSI-awarded papers, in Figure 4 we can see that in the first four years only few if any papers were classified as covering

<sup>17</sup>This benefit is promised by TVCG in their paper acceptance messages.

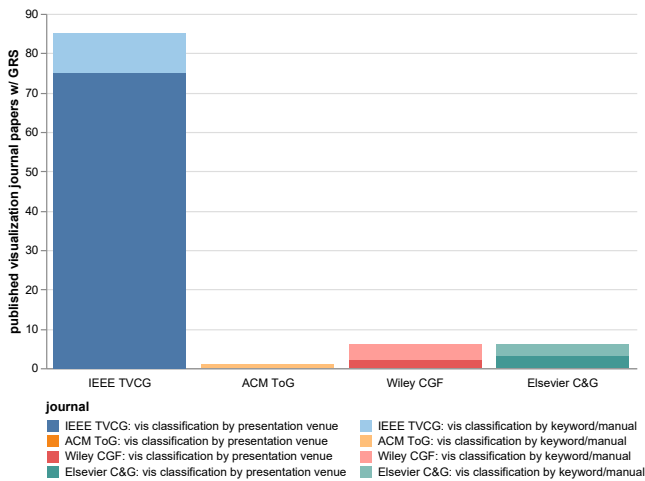


Figure 5: Sources of visualization papers overall, split between classification by visualization presentation venue and keyword/manual classification. A normalized version of this plot can be found in Figure 18 in Appendix C.

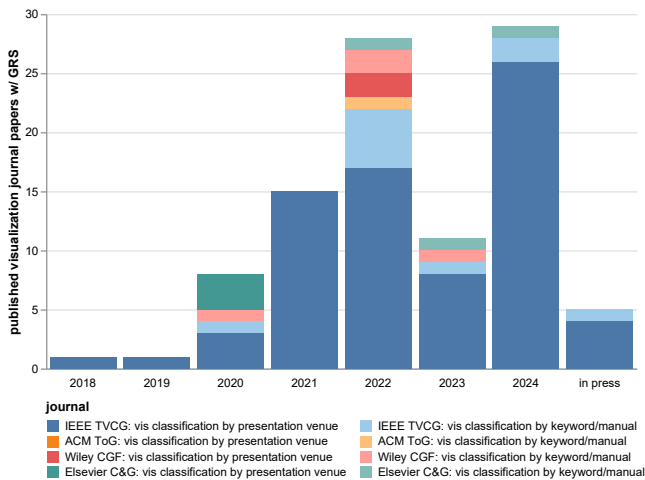


Figure 6: Same data as in Figure 5, but by article *publication* year. For a normalized version of the plot see Figure 19 in Appendix C.

visualization topics—likely an effect from the GRSI originating from the computer graphics field. TVCG as the prime journal of the visualization field only joined the GRSI in 2018, and in Figure 5 we see that TVCG is by far the dominating source of GRSI-awarded visualization papers. This figure also shows the split between author-classified (i. e., classified by presentation venue) and keyword- or manually classified papers separately for each journal, with 88.2% (75 out of 85) of GRSI-awarded TVCG papers on visualization topics actually being presented by their authors at visualization conferences. The percentages at the other venues are much lower, with 33.3% (2 out of 6) for CGF, 50.0% (3 out of 6) for C&G, and 0.0% (0 out of 1) for ToG (albeit at much lower total visualization papers in these three journals). For more detail, Figure 6 shows the same data as Figure 5, but with the development over the years. What is interesting to observe here is that almost all visualization TVCG papers are actually presented at conferences by their authors, except of those published in 2022—for which I really cannot envision a meaningful reason, other than it being a fluke due to the small-number statistics that this analysis admittedly is.

This small-number character is even more apparent if we drill deeper into the different venues where each of the visualization papers was presented. Figure 7 shows the evolution of GRSI-awarded

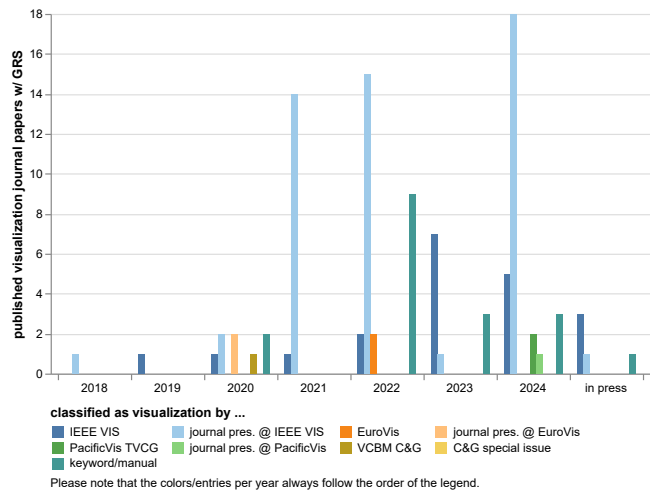


Figure 7: Different types of classification of papers as visualization work, by article *publication* year (i. e., not by paper *presentation* year). For a stacked bar chart version of the plot and its normalized version see Figure 20 and Figure 21 in Appendix C, respectively.

visualization papers by their *publication* years, split up into the different conferences at which they were accepted (20 at IEEE VIS, 2 at EuroVis, and 2 at PacificVis), the conferences where the papers were presented as journal papers (52 at IEEE VIS, 2 at EuroVis, and 1 at PacificVis), special-issue journal papers (1 in C&G special issue on VCBM, 0 in a C&G special issue on EuroVA, EnvirVis, or MolVA), or the 18 keyword- or manually classified papers. Figure 1 shows the same data, but focuses only on the conferences and the papers presented in their programs (regular and journal presentations) and also shows the papers by the year in which they were *presented*. Please note that in this figure (and in the normalized version in Figure 8) the IEEE VIS journal presentation counts include the IEEE Computer Graphics and Applications (CG&A) papers presented at the conference. This magazine, however, is not (yet?) included within the scope of venues for which a GRS is awarded, so that currently even in an ideal case the IEEE VIS journal papers could not achieve a 100% GRS rate (and also the presented rate is thus lower than it actually is for only presented TVCG papers).

Nonetheless, while we see in Figures 1 and 8 that the overall number of GRSI-awarded papers within visualization venues is still low, it is increasing. In particular for journal papers presented at IEEE VIS we see a substantial number of papers with a GRS—52 papers in total so far—, leading to percentages of  $\approx 10\%$  to up to  $\approx 20\%$ – $30\%$  in recent years. This is great news, and the fact that the percentage of GRSI-awarded papers among the TVCG journal papers is much higher than the percentage among pure IEEE VIS papers is also not surprising: pure IEEE VIS papers are guaranteed publication in (typically) the first TVCG issue of the year following the conference (i. e.,  $\approx 6$  months after paper acceptance and  $\approx 3$  months after presentation), while non-conference TVCG papers normally have to wait 1–2 years for full publication (i. e., being assigned a volume, an issue, and page numbers) if it were not for TVCG’s offer to fast-track GRSI-awarded papers to full publication.<sup>17</sup>

The data cleaning and visualization classification now also allows us to look at the contributions to the GRSI by author. Figure 9 shows the overall number of authors for a given number of GRSI awards—on the left for all authors who received a GRS, on the right only for people I classify as visualization authors by the fact that they, in my analysis, published  $\geq 50\%$  visualization papers. We see the clear signs of a power-law distribution, which is to be expected for datasets like this one. With our data, however, we can also ask to what degree the different communities between visualization and

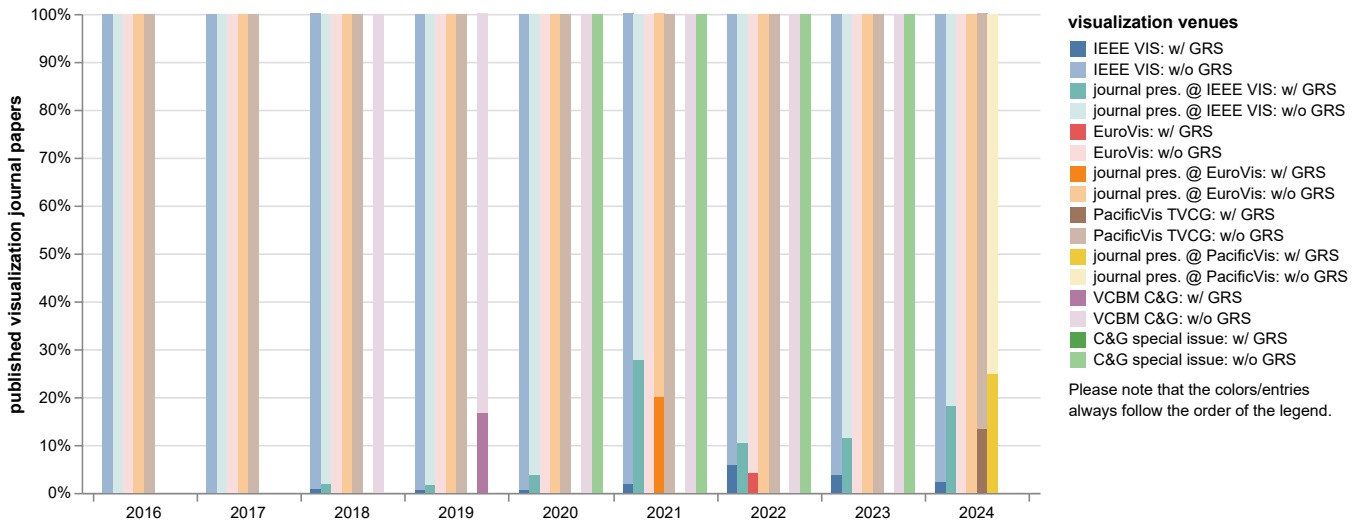


Figure 8: Papers that can clearly be classified as visualization work (based on being accepted to visualization conferences, being journal presentations at visualization conferences, <sup>14,15</sup> or having appeared in visualization-themed special issues in journals) by *presentation* year (i. e., based on the *conference* years, in which the papers were presented; same data as in Figure 1), with their GRS proportions, normalized.

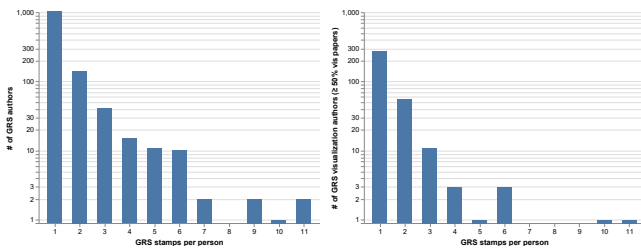


Figure 9: Number of authors with a given number of GRSI-awarded papers. Left: for all authors; right: only visualization authors.

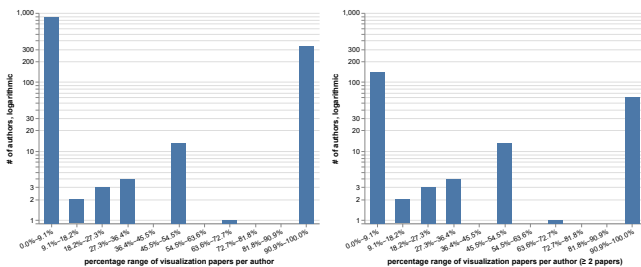


Figure 10: Histogram of the percentage of visualization papers per author (logarithmically plotted; left: all authors, right: only authors with  $\geq 2$  papers): a fairly bimodal distribution.

computer graphics intermix (I assume that papers not about visualization topics can be classified as being about computer graphics and related topics). Figure 10 thus shows a histogram of number of authors with a given percentage of visualization papers (based purely on the GRSI data). The left histogram in Figure 10 shows this analysis for all authors, but naturally this plot is biased by the many authors with a single paper only in the data (as we saw in Figure 9)—these can by definition only contribute to either group exclusively. On the right of Figure 10 I show, therefore, only those authors with  $\geq 2$  papers in the dataset. Still, we can see that we have a rather bimodal distribution between visualization authors and computer graphics authors, with only few authors publishing (GRSI-awarded) papers in both fields.

Finally, we can investigate the countries in which the GRSI-awarded authors are active. I base this analysis on the author affil-

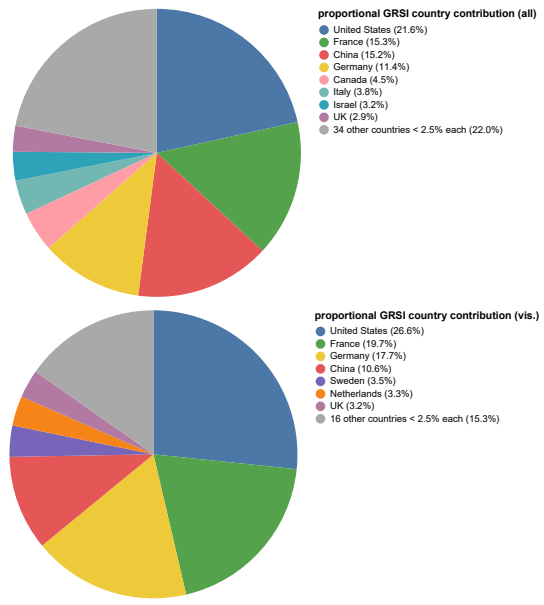


Figure 11: Proportional country contribution to GRSI awards (counting all paper authors). Top: overall; bottom: visualization papers. These plots are thresholded to 2.5% per-country contribution (for clarity), for full versions see Figure 23 and Figure 25 in Appendix C, respectively.

iation country or countries at publication time (as they reported it on the paper), as I described in Section 4, and for the purpose of the analysis I assign each paper the same weight of 1. I then distribute this weight evenly to each author of the paper, so an author of a paper with  $N$  co-authors gets weight  $1/N$ , which is then assigned to the respective country of affiliation of the author. If an author has  $M$  affiliations in different countries, then I only award each of these countries a weight of  $1/N \cdot M$  from the author. This way each paper counts equally, regardless of number of authors and their affiliations, and we get the proportional representation I show in Figure 11. The top pie chart in the figure shows the countries for all GRSI-awarded authors, the bottom one shows the version for only visualization-themed papers. We can see that four countries play the major role in both cases. For all papers with GRS, the United States has the largest

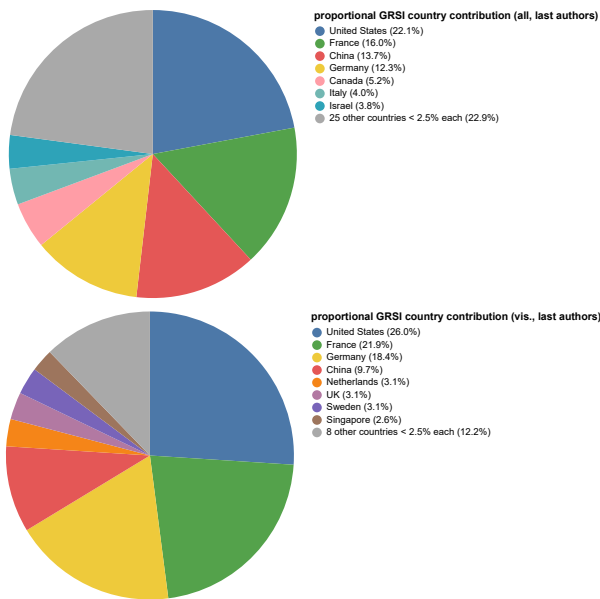


Figure 12: Country contribution of paper last authors to GRS awards. Top: overall; bottom: visualization papers. These plots are thresholded to 2.5% per-country contribution (for clarity), for full versions see Figure 24 and Figure 26 in Appendix C, respectively.

proportion with 21.6%, followed by France with 15.3%, China with 15.2%, and Germany with 11.4%. For visualization-themed papers with GRS, the United States also has the largest proportion with 26.6%, followed by France with 19.7%, Germany with 17.7%, and China with 10.6%. I also asked myself if this distribution would change drastically if we would only consider the last and thus likely the senior author of each paper, for which I show the analysis in Figure 12.<sup>18</sup> As we can see, the situation is similar to what we saw before. For all papers with GRS, the United States again has the largest proportion with 22.1%, followed by France with 16.0%, China with 13.7%, and Germany with 12.3%. For visualization-themed papers with GRS, the United States also still has the largest proportion with 26.0%, followed by France with 21.9%, Germany with 18.4%, and China with 9.7%.

In visualization we thus see that the mentioned four countries are responsible for  $\approx 75\%$  of all the reproducible research. Of course, we would need to compare this number to the overall contributions of the different countries to the field, such as looking at all papers presented at IEEE VIS as the primary publication venue, to know if these numbers differ from the general situation. Unfortunately I do not have the data to check it, and augmenting the VisPubData dataset [14] with countries of affiliations would be a substantial amount of work—beyond the scope of this paper. My expectation, however, is that these numbers would differ. For example, among the top five contributors to the GRSI from the visualization community, three (including the top two) are from France at the moment, at least partially explaining the high ranking of the country in the analysis above. So at the moment and due to the low number of reproducible papers overall (when compared to the overall number of scientific publications in both visualization and computer graphics) some countries may be over- or underrepresented, simply due to the initiative or lack of initiative of specific productive (usually senior) paper authors in the respective fields.

<sup>18</sup>Behind this analysis is the assumption that the last/senior author is more likely to be the instigator of a GRSI application.

## 6 DISCUSSION

So how are we doing, then? Visualization papers currently only represent 28.3% of all papers with a GRS, just a bit more than a quarter. After all, there is a difference between, on the one hand, some code being available as open source for a paper (which we do see often in today’s papers) that, with little or (usually much) more effort [3] can be made to run and, on the other hand, the actual certification of a sufficiently documented and runnable code—the latter being ensured by a GRSI certification. Research in psychology [7], for instance, found that “just” awarding badges to papers for the fact that data and/or code is available is not sufficient to guarantee an exact reproducibility of the reported results—which, in contrast, the certification by initiatives such as the GRSI does guarantee (at least for those visuals of the paper that are being checked). Also within visualization we thus need more certification of reproducibility because only this certification forces authors to check that their shared code actually runs out of the box and has a sufficient level of documentation. The 28.3% are too low—we can and we should do better. Several calls for more reproducibility exist in our field as I reviewed in Section 3 (e. g., [10, 13, 19])—yet why are the numbers of papers with a GRS still low then? Speculating about possible reasons for this situation, a few reasons could explain the low numbers.

First, there may be an **issue with (too) few people knowing about the GRSI**. In this case we would need more publicity for this service. Yes, the EICs of the affected journals repeatedly mention the initiative in their messages to the reader (e. g., [9, 17]), but maybe these messages are not as well read by the community? Also, while the GRSI is mentioned in the acceptance letter at least of some journals (such as TVCG) and conferences (such as IEEE VIS), maybe we could more clearly announce this initiative and its benefits for our community in the opening sessions of our conferences, so that more people hear about and are aware of it? On the bright side, however, as of today 3 of the 124 regular papers (2.4%) that are accepted to the (at the time of writing this paper still upcoming) 2024 VIS conference are already GRSI-certified. So at least some authors in our community are well aware of the GRSI and are eager to get their work certified—even before the conference.

Second, and despite this very positive observation just now, there may still be an **issue with few people willing to encourage the students to go the extra mile or the lack of tangible incentives**—after all, documentation of code still can require quite a bit of additional work, even if in many cases the GRSI application is rather lightweight. Yet being awarded a GRS currently often still boils down to “eternal fame and glory” [15]. One could argue (and some have [13]) that a way to success is to make reproducibility a requirement for publication, but I am not sure how realistic such a move would be (at least in the short run), and we would need to deal with cases where code cannot be shared or where other reasons arguably prevent us from checking reproducibility. For instance, exceptions are possible even in models that require shared materials [13] or the checking could be assigned to a trusted third party [1]. Yet a better approach may be to begin with more strongly incentivizing reproducibility such as through reproducibility awards at our conferences and in our journals—beyond a GRSI certification. Computers & Graphics already has a corresponding award, the “Computers and Graphics Best (Replicable) Paper Award” [18]—perhaps TVCG and CGF could introduce something similar as well? Maybe also other incentives could work, such as incentives at the level of the authors’ local institutions that should award GRS-like achievements or provide other forms of credit for such efforts.

Third, there may be an **issue with visualization being under-represented in the possible venues**. As we saw in Figure 3, the vast majority of GRSI-awarded visualization papers are published in TVCG. The other journals and conferences that are eligible for a GRS have little to no visualization-themed content. One way to address this issue would be to encourage the authors of visualization

papers in those other venues to apply for the stamp, another is to add more core visualization venues to the list of venues to the GRSI. For instance, journals like Sage’s *Information Visualization*,<sup>19</sup> Springer’s *Journal of Visualization*,<sup>20</sup> IEEE’s *Computer Graphics and Applications*,<sup>21</sup> and *The Journal of Visualization and Interaction*<sup>22</sup> could be added to the list. Also, similar to the GRSI recently having embraced SIGGRAPH’s and SIGGRAPH Asia’s conference-only papers, it could add at least some of the many significant smaller conferences and workshops of our field such as PacificVis (non-TVCG papers), ChinaVis, Lдав, VizSec, TopoInVis, VCBM (non-C&G papers), IVAPP, plus many others, and last but not least also BELIV. If such an expansion of topics into the visualization research field is something that is not desired by the GRSI because it wants to focus on computer graphics venues (as indicated by its name), then maybe the visualization community could consider to start an own initiative that could collaborate with the GRSI—such that papers in already covered venues such as TVCG, C&G, and CGF could apply to either group and, if certified, could be cross-listed in both initiatives.<sup>23,24</sup>

Finally, there is the potential **issue that some work is inherently difficult to reproduce**, such as research projects that require special dedicated hardware setups to run (e. g., VR or AR headsets, large wall setups, etc.) that the GRSI is not prepared to test for. I have encountered this very problem myself in the past, and there is no easy solution. Yet often it is possible to circumvent the issue by applying for a GRS for some other aspect of the work, such as the reproduction of the statistical analysis of a user experiment. But maybe the GRSI could consider asking its reviewers to state if they have access to some *standard* special hardware devices such as certain VR or AR headsets, and then to assign such applications to those reviewers who have access to the mentioned setups.

What is interesting to observe is that some authors who have explicitly advocated for reproducibility via the GRSI in the past so far at least have not received a GRS for their papers. I reached out to some of them to ask about the reasons for this lack of papers with GRS. In their responses some noted that they do not publish in the visualization field anymore but in other computer science fields with their own (and partially much longer-standing) forms of reproducibility initiatives, so a GRSI certification simply does not apply. Others said that, for some work, they were not the primary supervisor, which relates to the issue of encouragement and incentives I discussed above—a GRSI application usually does not happen without a primary supervisor strongly encouraging the responsible student to apply for the stamp. Yet other work primarily consisted of user experiments, which would call for an actual *replicability* verification which is currently beyond the capability of the GRSI—even if a full study protocol and the associated materials were provided. What this means is that more papers are actually reproducible or replicable than are certified by the GRSI—which would only be discovered by an initiative as the Code Replicability in Computer Graphics (CRCG)<sup>25</sup> project [3], which actively and without an application from the authors checks the code resources provided for published papers.

As for the *reproducibility* verification of the statistical analysis of existing study results, which the GRSI does check, the authors I contacted feel strongly that the stamp awarded by the GRSI would be misleading, as they claim to certify *replicability* but in fact they

certify *reproducibility*—as I had discussed in Section 2. They also feel that the verification suggests a validity that may not exist because authors could also manipulate their data. Here I personally disagree: To some degree we need to trust authors to report the true data. I also feel that a verification of the *reproducibility* of a statistical analysis of the results of an experiment not only enables others to check the numbers or conduct a similar analysis in the future, it also forces the authors of a paper to script, verify, and document the whole analysis code for their very own benefit: By starting with a reproducible analysis setup they become able to quickly sort out problems and re-run the whole analysis with little effort.

The authors also mentioned that **commercialization intents and proprietary code**<sup>26</sup> may prevent a certification, as do the problems of the already mentioned lack of incentives for paper authors. Finally, they pointed at the previously discussed publicity issue—observing that, as of 2024, neither the IEEE VIS call for papers nor the corresponding page on open practices<sup>27</sup> currently mention the GRSI, something that easily could be fixed.

So, ultimately, while we have already achieved quite a bit of progress compared to a few years ago, there is still work ahead of us. We need to work toward reducing the obstacles of making work reproducible and, if possible, even replicable, we need to reduce the confusion associated with the many terms being used, we need to increase the awareness of authors of initiatives such as the GRSI, we need to increase the incentives for people to provide the corresponding resources, and we need to discuss if we can extend the scope of venues that are covered by initiatives such as the GRSI.

Of course, any change that addresses the *replication crisis* in the form of more reproducibility or replicability verification also has **implications for the GRSI** as the initiative that checks it. More applications mean more work, and thus we would need more reviewers to support the initiative by volunteering their time. So we all need to be willing to accept this additional reviewing duty, on top of the existing *reviewing crisis* in (not only) our field. Finally I note that, maybe, the GRSI could consider renaming itself to *Graphics Reproducibility Stamp Initiative* to address the terminology confusion that I had mentioned before in Section 2.

## 7 REPRODUCIBLE PAPER WRITING

A paper about research reproducibility would not be authentic without being reproducible itself. I thus ensured that this is the case, not only for the data analysis I describe and the plots I produce, but also for the whole paper itself. I provide all materials, the script and the sources for the paper itself, as well as all external data that I used for my analysis. I normally pull the actually analyzed GRSI data dynamically from the web, and in this case minor manual updates (the country data) are needed for potentially newly added papers to the GRSI website. By default, however, the script runs based on the data status at publication time, but with a small reconfiguration of the script updated data can be used as well. The analysis itself is fully scripted (in a Python script), and produces all images shown in the paper (and more). Some of the numeric analysis results are also collected by the script and then written into L<sup>A</sup>T<sub>E</sub>X macros in dedicated files, which are then pulled into the L<sup>A</sup>T<sub>E</sub>X paper. So any future L<sup>A</sup>T<sub>E</sub>X compilation then uses the updated data, partially also with the pgf package which facilitates in-document calculations—even with the previously mentioned script-generated macros as input data. This process, b.t.w., not only facilitates the reproducibility of the work but also makes the paper writing itself much easier—any change in, correction of, or update of the analysis only needs to

<sup>26</sup>I can confirm this issue of proprietary code from my own experience, in which case only executable demos are a possibility [15]—which are not accepted by the GRSI as reproducibility artifacts.

<sup>27</sup>[ieevis.org/year/2024/info/open-practices/open-practices](https://ieevis.org/year/2024/info/open-practices/open-practices)

<sup>19</sup>[journals.sagepub.com/home/ivi](https://journals.sagepub.com/home/ivi)

<sup>20</sup>[link.springer.com/journal/12650](https://link.springer.com/journal/12650)

<sup>21</sup>[computer.org/csdl/magazine/cg](https://computer.org/csdl/magazine/cg)

<sup>22</sup>[journalovi.org](https://journalovi.org)

<sup>23</sup>Maybe a collaboration with initiatives in related fields as RepliCHI in HCI [28–30] would also be an option—even if RepliCHI (like the EuroRV<sup>3</sup> workshop) no longer seems to be running.

<sup>24</sup>An alternative to be considered could be the Papers with Code initiative in machine learning at [paperswithcode.com](https://paperswithcode.com).

<sup>25</sup>[replicability.graphics](https://replicability.graphics)

be done in or via the analysis script. The paper itself then automatically updates on the next compilation—seemingly by magic—, which means this approach is essentially an equivalent of a Jupyter Notebook for L<sup>A</sup>T<sub>E</sub>X document authoring.

To thus be able to reproduce my work, I encourage the reader to get the project from the linked repository (details below), and then to first run the script to do the data analysis and to produce the plots, before compiling the L<sup>A</sup>T<sub>E</sub>X sources to reproduce the paper. If the script was instructed to get new data from the web (in this case the country information for the newly added papers would need to be manually added as documented in the `readme.md`), the paper can be updated to the most recent data without any problems. Naturally, there is a limit to this reproduction process—after a while (likely months to 1–2 years) more authors will have been awarded a GRS and the situation, as I describe it, will have changed, hopefully for the better, and then the discussion in the paper itself will no longer be correct. Moreover, the structure of the website that I query or the used APIs of the digital libraries may also be different then, which could break the processing in the Python script I wrote.

## DATA SOURCES

The data used in this paper comes from the VisPubData dataset [14] as well as from downloads/extracts from the IEEE Xplore, ACM, Elsevier ScienceDirect, Wiley, Eurographics, and Crossref digital libraries; the GRS data comes from [replicabilitystamp.org](https://replicabilitystamp.org); the data is (and resulting plots are) current as of Aug. 7, 2024. I also note that the results of the analysis I described in this paper, specifically the fact that a given (IEEE VIS) paper has received a GRS, have already been added to the most recent 2023 edition of the VisPubData dataset [14] and will continue to be updated there.

## SUPPLEMENTAL MATERIAL POINTERS

The repository to reproduce the presented results as well as this paper (as discussed in Section 7) can be found at [github.com/tobiasisenberg/Visualization-Reproducibility](https://github.com/tobiasisenberg/Visualization-Reproducibility). A copy of the paper itself including its appendix as well as the figures from the paper can be found at [osf.io/mvnbj](https://osf.io/mvnbj).

## FIGURE CREDITS AND COPYRIGHT

I as the author of this paper state that all figures in this paper are my own as well as are and remain under my own personal copyright, with the permission to be used here. I also make them available under the [Creative Commons Attribution 4.0 International \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/) license and share them at [osf.io/mvnbj](https://osf.io/mvnbj).

## ACKNOWLEDGMENTS

Primarily I wish to thank the Graphics Replicability Stamp Initiative and, in particular, their reviewers for providing their invaluable free service to the computer graphics and visualization communities. In addition, I thank Stefanie Behnke for providing the EuroVis publication data extract from the CGF database, Ross Maciejewski for the list of planned IEEE VIS 2024 TVCG journal presentations, Lane Harrison for the list of planned IEEE VIS 2024 CG&A journal presentations, the 2024 VIS Papers Chairs for the list of accepted papers, as well as Petra Isenberg and Natkamon Tovanich for their continued work of keeping the VisPubData [14] database up to date and error-free. Finally, thanks to Lonni Besançon, Martin Skrodzki, and Petra Isenberg for feedback on this manuscript, to the authors whose replies I cite anonymously in Section 6 for their answers, as well as Cody Dunne for help with polishing the documentation of the [GitHub repository](https://github.com/tobiasisenberg/Visualization-Reproducibility) and with fixing some issues with the code for specific Python environments.

## REFERENCES

[1] L. Besançon, N. Peiffer-Smadja, C. Segalas, H. Jiang, P. Masuzzo, C. Smout, E. Billy, M. Deforet, and C. Leyrat. Open science saves

lives: Lessons from the COVID-19 pandemic. *BMC Med Res Methodol*, 21, art. no. 117, 18 pages, 2021. doi: [10/gkdzr6](https://doi.org/10/gkdzr6)

[2] R. F. Boisvert. Incentivizing reproducibility. *Commun ACM*, 59(10):5, 2016. doi: [10/gc5pts](https://doi.org/10/gc5pts)

[3] N. Bonneel, D. Coeurjolly, J. Digne, and N. Mellado. Code replicability in computer graphics. *ACM Trans Graph*, 39(4), art. no. 93, 8 pages, 2020. doi: [10/gg8xfh](https://doi.org/10/gg8xfh)

[4] S. Chopra, L. Labache, E. Dhamala, E. R. Orchard, and A. Holmes. A practical guide for generating reproducible and programmatic neuroimaging visualizations. *Aperture Neuro*, 3, art. no. 001c.85104, 20 pages, 2023. doi: [10/gtw4sn](https://doi.org/10/gtw4sn)

[5] J. F. Claerbout and M. Karrenbach. Electronic documents give reproducible research a new meaning. In *SEG Expanded Abstracts*, pp. 601–604. Society of Exploration Geophysicists, Houston, 1992. doi: [10/b6t7wj](https://doi.org/10/b6t7wj)

[6] A. Cockburn, P. Dragicevic, L. Besançon, and C. Gutwin. Threats of a replication crisis in empirical computer science. *Commun ACM*, 63(8):70–79, 2020. doi: [10/gjbnx4](https://doi.org/10/gjbnx4)

[7] S. Crüwell, D. Aphthorp, B. J. Baker, L. Colling, M. Elson, S. J. Geiger, S. Lobentanzer, J. Monéger, A. Patterson, D. S. Schwarzkopf, M. Zaneva, and N. J. L. Brown. What’s in a badge? A computational reproducibility investigation of the open data badge policy in one issue of psychological science. *Psychol Sci*, 34(4):512–522, 2023. doi: [10/grzd4n](https://doi.org/10/grzd4n)

[8] J. B. Cushing, D. Lach, C. Zanocco, and J. Halama. Scientific visualization and reproducibility for “open” environmental science. In *Proc. Big Data*, pp. 3211–3216. IEEE CS, Los Alamitos, 2018. doi: [10/gtw4zm](https://doi.org/10/gtw4zm)

[9] L. De Floriani. State of the journal. *IEEE Trans Vis Comput Graph*, 24(2):1036–1037, 2018. doi: [10/gtzdcx](https://doi.org/10/gtzdcx)

[10] J.-D. Fekete and J. Freire. Exploring reproducibility in visualization. *IEEE Comput Graph Appl*, 40(5):108–119, 2020. doi: [10/ghd59m](https://doi.org/10/ghd59m)

[11] M. Franke, G. Reina, and S. Koch. Toward reproducible visual analysis results. In *Proc. PacificVis*, pp. 102–106. IEEE CS, Los Alamitos, 2023. doi: [10/gtw42q](https://doi.org/10/gtw42q)

[12] D. Garkov, C. Müller, M. Braun, D. Weiskopf, and F. Schreiber. Research data curation in visualization: Position paper. In *Proc. BELIV*, pp. 56–65. IEEE CS, Los Alamitos, 2022. doi: [10/gttm5m](https://doi.org/10/gttm5m)

[13] S. Haroz. Open practices in visualization research: Opinion paper. In *Proc. BELIV*, pp. 46–52. IEEE CS, Los Alamitos, 2018. doi: [10/gtw4sp](https://doi.org/10/gtw4sp)

[14] P. Isenberg, F. Heimerl, S. Koch, T. Isenberg, P. Xu, C. D. Stolper, M. Sedlmair, J. Chen, T. Möller, and J. Stasko. [vispubdata.org](https://vispubdata.org): A metadata collection about IEEE visualization (VIS) publications. *IEEE Trans Vis Comput Graph*, 23(9):2199–2206, 2017. doi: [10/ggwwrv](https://doi.org/10/ggwwrv)

[15] T. Isenberg. Personal experiences of providing and using research prototypes. In *Proc. VisGap*, pp. 17–22. EG, Goslar, 2022. doi: [10/gr2d83](https://doi.org/10/gr2d83)

[16] Y. Jansen, J. B. Vornhagen, O. Iarygina, K. S. Niksirat, L. Besançon, P. Dragicevic, J. Gori, and C. Wacharamanotham. The many ways of being transparent in human-computer interaction research. OSF preprint, 2024. doi: [10/gt3f2v](https://doi.org/10/gt3f2v)

[17] J. Jorge. Editorial note. *Computers & Graphics*, 74:A1–A2, 2018. doi: [10/gtzdcz](https://doi.org/10/gtzdcz)

[18] J. Jorge. A note from the editor in chief. *Computers & Graphics*, 88:A1–A3, 2020. doi: [10/gtzdc3](https://doi.org/10/gtzdc3)

[19] R. Kosara. An empire built on sand: Reexamining what we think we know about visualization. In *Proc. BELIV*, pp. 162–168. ACM, New York, 2016. doi: [10/gfz5kr](https://doi.org/10/gfz5kr)

[20] R. Kosara and S. Haroz. Skipping the replication crisis in visualization: Threats to study validity and how to address them: Position paper. In *Proc. BELIV*, pp. 102–107. IEEE CS, Los Alamitos, 2018. doi: [10/gtw4sq](https://doi.org/10/gtw4sq)

[21] H. E. Plesser. Reproducibility vs. replicability: A brief history of a confused terminology. *Front Neuroinf*, 11, art. no. 76, 4 pages, 2018. doi: [10/gc5ptr](https://doi.org/10/gc5ptr)

[22] G. J. Quadri and P. Rosen. You can’t publish replication studies (and how to anyways). arXiv preprint 1908.08893, 2019. doi: [10/gtxgh9](https://doi.org/10/gtxgh9)

[23] G. Reina. Can image data facilitate reproducibility of graphics and visualizations? Toward a trusted scientific practice. *IEEE Comput Graph Appl*, 43(2):89–99, 2023. doi: [10/gtw4nt](https://doi.org/10/gtw4nt)

[24] N. P. Rougier, K. Hinsén, F. Alexandre, T. Arildsen, L. A. Barba,



- F. C. Y. Benureau, C. T. Brown, P. de Buyl, O. Caglayan, A. P. Davison, M.-A. Delsuc, G. Detorakis, A. K. Diem, D. Drix, P. Enel, B. Girard, O. Guest, M. G. Hall, R. N. Henriques, X. Hinaut, K. S. Jaron, M. Khamassi, A. Klein, T. Manninen, P. Marchesi, D. McGlenn, C. Metzner, O. Petchey, H. E. Plesser, T. Poisot, K. Ram, Y. Ram, E. Roesch, C. Rossant, V. Rostami, A. Shifman, J. Stachelek, M. Stimberg, F. Stollmeier, F. Vaggi, G. Viejo, J. Vitay, A. E. Vostinar, R. Yurchak, and T. Zito. Sustainable computational science: The ReScience initiative. *PeerJ Comput Sci*, 3, art. no. e142, 17 pages, 2017. doi: [10/gcx5kf](https://doi.org/10/gcx5kf)
- [25] P. T. Sukumar and R. Metoyer. Towards designing unbiased replication studies in information visualization. In *Proc. BELIV*, pp. 93–101. IEEE CS, Los Alamitos, 2018. doi: [10/gtw4sr](https://doi.org/10/gtw4sr)
- [26] U. H. Syeda, L. South, J. Raynor, L. Panavas, D. Saffo, T. Morriss, C. Dunne, and M. A. Borkin. Vis repligogy: Towards a culture of facilitating replication studies in visualization pedagogy and research. In *EuroVis Education Papers*, art. no. eved.20241054, 9 pages. EG, Goslar, 2024. doi: [10/gtxbsb](https://doi.org/10/gtxbsb)
- [27] A. C. Valdez, A. K. Schaar, J. R. Hildebrandt, and M. Ziefle. Requirements for reproducibility of research in situational and spatio-temporal visualization: Position paper. In *Proc. BELIV*, pp. 53–59. IEEE CS, Los Alamitos, 2018. doi: [10/gtw4ss](https://doi.org/10/gtw4ss)
- [28] M. L. Wilson, E. H. Chi, S. Reeves, and D. Coyle. RepliCHI: The workshop II. In *CHI Extended Abstracts*, pp. 33–36. ACM, New York, 2014. doi: [10/gt28zz](https://doi.org/10/gt28zz)
- [29] M. L. Wilson, W. Mackay, E. H. Chi, M. Bernstein, and J. Nichols. RepliCHI SIG: From a panel to a new submission venue for replication. In *CHI Extended Abstracts*, pp. 1185–1188. ACM, New York, 2012. doi: [10/m59w](https://doi.org/10/m59w)
- [30] M. L. Wilson, P. Resnick, D. Coyle, and E. H. Chi. RepliCHI: The workshop. In *CHI Extended Abstracts*, pp. 3159–3162. ACM, New York, 2013. doi: [10/gt28zx](https://doi.org/10/gt28zx)