Datamations: Animated Explanations of Data Analysis Pipelines

Xiaoying Pu
xpu@umich.edu
University of Michigan, Ann Arbor
Ann Arbor, Michigan

Sean Kross
seankross@ucsd.edu
University of California San Diego
La Jolla, California

Jake M. Hofman
jmh@microsoft.com
Microsoft Research
New York, New York

Daniel G. Goldstein
dgg@microsoft.com
Microsoft Research
New York, New York

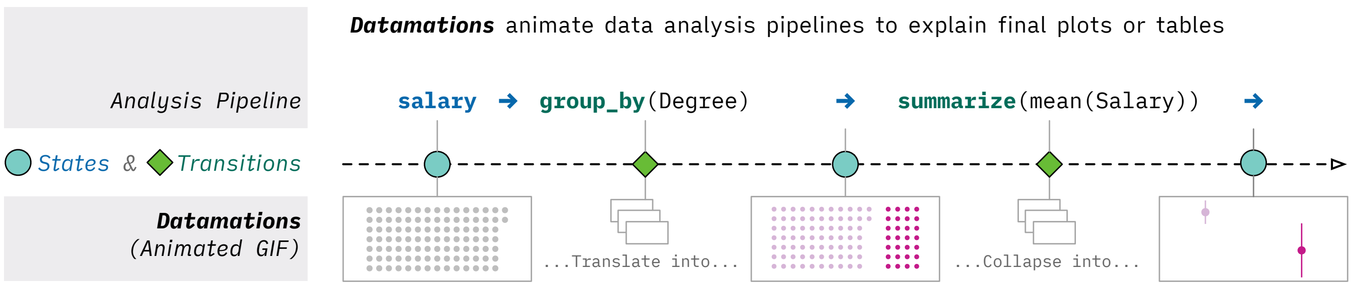


Figure 1: An overview of the idea behind datamations.

ABSTRACT

Plots and tables are commonplace in today's data-driven world, and much research has been done on how to make these figures easy to read and understand. Often times, however, the information they contain conveys only the end result of a complex and subtle data analysis pipeline. This can leave the reader struggling to understand what steps were taken to arrive at a figure, and what implications this has for the underlying results. In this paper, we introduce *datamations*, which are animations designed to explain the steps that led to a given plot or table. We present the motivation and concept behind datamations, discuss how to programmatically generate them, and provide the results of two large-scale randomized experiments investigating how datamations affect people's abilities to understand potentially puzzling results compared to seeing only final plots and tables containing those results.

KEYWORDS

data, analysis, visualization, animations

ACM Reference Format:
Xiaoying Pu, Sean Kross, Jake M. Hofman, and Daniel G. Goldstein. 2021. Datamations: Animated Explanations of Data Analysis Pipelines. In *CHI Conference on Human Factors in Computing Systems (CHI '21), May 8–13, 2021, Yokohama, Japan.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3411764.3445063

CHI '21, May 8–13, 2021, Yokohama, Japan

ACM ISBN 978-1-4503-8096-6/21/05…$15.00
https://doi.org/10.1145/3411764.3445063

1 INTRODUCTION

As the world becomes more data-driven, we are increasingly presented with plots and tables that convey the results of complex analyses involving intricate datasets. A great deal of work has been done on how to make these figures easy for readers to comprehend, for instance in helping people decode data values [10, 24, 38], make high-level inferences [51], and perceive related uncertainties [16, 28]. All of this work has led to vast improvements in data visualization and data communication, and yet it has focused on conveying only a small part of what is involved in creating these plots and tables.

By the time a reader is presented with a figure, the underlying data have most likely been extensively processed (*e.g.*, filtered, grouped, aggregated, augmented, and reshaped), but conventional plots and tables show only the end results of the analyses that led to them. As such, while the reader is often able to decode the values in a plot or table, they may be left wondering how those values were arrived at and what they actually mean. In short, it is easy to see a figure and be unsure of exactly what went into creating it and how that affects what one should (and shouldn't) take away from it.

One solution to this problem is giving the reader more context around the steps that led to any given plot or table. Often times this is done through text and captions written to accompany a figure. Such write-ups can be time-consuming to produce and are often only an approximation to the actual steps that were taken in any data analysis pipeline. As a result, these descriptions can introduce ambiguities, gloss over important steps in an analysis, or fail to convey analysis steps in an easy-to-comprehend manner. Another potential solution is to simply share the code that generated a figure, but this can have the opposite problem: while code is precise and provides exact information on the process that led to a figure, it is often difficult for novices and experts alike to read and understand someone else's code.

This is where we see an opportunity to improve how data analyses are communicated to readers. In this paper, we introduce *datamations*, automatically generated animations that explain the data analysis pipeline that produced a given plot or table. The idea behind datamations is relatively simple: as illustrated in Figure 1, each step in a data analysis pipeline can be programmatically mapped onto a visual transformation of the underlying dataset, and these transformations can be chained together to produce an animated explainer for a plot for table. Datamations can, for instance, show data being filtered, split into different groups, and aggregated into summary statistics.

To clarify our contribution, we are aware of existing examples of animations in data journalism [4] and education [37] that have been created to explain data analyses. These serve as inspirations for our work, but many of them are custom animations that require a great deal of time and manual effort to create, and do not readily generalize to other scenarios. Furthermore, custom animations for data analysis pipelines have not to our knowledge been tested in controlled experiments to assess the degree to which they help readers understand plots and tables they are shown. Thus, our contribution is threefold: first, to formalize the idea of datamations for explaining entire data analysis pipelines; second, to provide a framework by which such animations can be automatically generated to explain both plots and tables; and third to run large-scale experiments to better understand the potential benefits of (and issues with) these animations.

In the remainder of the paper, we first review previous work related to the development of datamations and then present a framework for formalizing and automatically generating datamations. Then we apply this framework to generate and test example datamations that explain an instance of Simpson's paradox [40], where a seemingly paradoxical reversal in final results can occur based on a small but important change to the underlying data analysis pipeline. We chose to study Simpson's paradox because it is a case where understanding the data analysis pipeline is critical for understanding the corresponding results. The problem that we presented participants involved a dataset of salaries where, looking across industry and academia, people with master's degrees make more money on average than people with PhD degrees overall, yet within both academia and industry people with master's degrees make less money on average than people with PhD degrees. We created plots and tables describing these results along with accompanying datamations that explain the process leading to these figures. We then used these datamations as stimuli in two pre-registered experiments that involve over 1,200 participants to test whether datamations are able to improve comprehension of Simpson's paradox. We investigated whether seeing datamations helps readers correctly identify that such a reversal is possible and/or whether it helps them choose the correct explanation for the reversal. We also used this experiment as a chance to collect qualitative feedback from participants on the benefits of datamations and ways in which they can be improved. We conclude with a discussion of participants' feedback, and we present thoughts for future research to be done in this area.

2 RELATED WORK

2.1 Understanding data and analysis behind visualizations

When readers are presented with a plot or table it can often be difficult to understand what led to the results encoded in that figure. Information about the raw data and the analysis pipeline are often placed elsewhere, in the form of code and written paragraphs. The disconnect between data representations and analysis is evident in Rule *et al.* [45], where users (data analysts) reportedly share their data analysis results as emails and slides, excluding the computational notebooks that generated the findings [45]. When users lack easy access to the context of data, they can misunderstand the data patterns they see. A concrete data scenario is Simpson's paradox, where data trends in one grouping contradicts trends in another grouping [40]. Without explicit efforts such as a detection algorithm [21], users may miss out on such aspects of the data or struggle to understand these apparent contradictions [41]. We introduce datamations to enhance conventional plots and tables with details from the data analyses that generated them, and test their effectiveness using an instance of Simpsons' paradox.

2.2 Using animations to communicate data

By using animations to explain plots and tables we hope to take advantage of their visual appeal and potential explanatory benefits. As one example, the GapMinder[1] animation communicates how life expectancy changed throughout the world in an engaging and informative manner[17]. Fisher suggests that such animations can be broadly useful when they follow the appropriate design principles [17]. By adding animated transitions between statistical graphics, Heer & Robertson find improvements in graphical perception in two ways—tracking objects and estimating changes [26]. Tracking objects—in our case data points in a plot or table—can be essential for understanding data analyses, as the user may want to know which data points are related to which results. In addition, users have been found to prefer animations over their static counterparts [2, 34]. Hypothetical outcome plots (or HOPs) are one recent and promising example of using animations to add to the information shown in static plots and boost reader comprehension [28, 31]. Specifically, HOPs augment static visualizations such as error bars with animated frames of random draws from the underlying sampling distribution. With HOPs, animation conveys randomness and uncertainty in final results, but unlike datamations, the goal of HOPs is not to communicate the underlying data analysis pipeline that led to these results.

With datamations we aim to surface visually the *entire* analysis pipeline behind final plots and tables. The high-level idea behind datamations is similar to that of animations that teach complex systems and algorithms, as reviewed in Tversky *et al.* and Hundhausen *et al.* [29, 48], but with a more specific scope and different goal. As we discuss in Section 3, datamations communicate a particular data

[1] gapminder.org

analysis process with concrete steps and visual analogies. Instead of helping readers learn abstract algorithms in an educational setting, datamations are meant to help people understand specific analysis results in everyday settings.

Since creating animated transitions can be low-level and time consuming, previous work has developed ways to automate the creation of these transitions. Kim & Heer [36] propose a visualization grammar and a recommender system on the visual mark level (*cf.* the Grammar of Graphics [53]) to augment specification. In addition, Drucker & Fernandez design a framework for transitioning between unit visualizations [14], and Canis is a declarative language for creating chart animations in SVG [19]. These approaches for automating animated transitions tend to work with a single data state and do not explicitly consider the semantics of animation such as aggregation or grouping. Our efforts on datamations take inspiration from and extend this work in two ways. First, we outline a general purpose framework for communicating entire analysis pipelines, such that there is a direct mapping from code that executes an analysis to an animation that helps to explain it. Second, we offer a large-scale empirical evaluation of the degree to which these animations aid readers' understanding of potentially puzzling results through our experiments involving Simpson's paradox.

2.3 Probing the data analysis pipeline

Datamations animate sequential, separable data operations, each of which can be thought of as a step in a "pipeline", where verbs operate on the data at each step and pipes chain together the results of each operation. Here we present an implementation of datamations centered around the programming language R, which has recently seen broad adoption of the pipeline paradigm through packages such as `magrittr` (which creates an explicit pipe operator within R, denoted by `%>%`) [3] and verbs to operate on data within a pipeline through the `tidyverse/dplyr` package [52]. That said, the pipeline approach is a general purpose one inspired by relational algebra [50] and has been adopted in systems ranging from Unix to Javascript [23]. As a result our work on datamations is applicable to this entire range of data analysis frameworks.

Ordinarily, data analysis pipelines do not directly expose the intermediate data states involved in an analysis. To this end, there are existing solutions for probing pipelines, such as text-based tools [13, 42, 44] and static illustrative comics [49] to surface intermediate results. Furthermore, a grammar for "data tweening" has been proposed to generate easy-to-follow, static data table visualizations to step through and explain database queries [33]. However, these solutions are often targeted at developers or data scientists and have not been evaluated in randomized experiments. Our work on datamations builds on these efforts by presenting a framework to automatically generate visually compelling animations that achieve similar goals, but for everyday consumers of data visualizations.

2.4 Provenance

Our efforts on visualizing data analysis pipelines are also related to work on visualizing *data provenance* [43]. Provenance is a broad research topic on the history of changes in the process of analysis and the creation of visualizations [43], and Chevalier *et al.* identify the use of animation as "the most under-explored" for replaying and summarizing history (provenance) [9].

The research community has designed many systems that capture and communicate data provenance during visual exploratory analysis and present it to users [6, 11, 15, 30, 46, 47]. These provenance-capable systems focus on preserving and presenting all the *alternatives histories* of user action. As a result, such systems often visualize provenance as abstract node-link diagrams [25]: the nodes represent data states, and forking links represent alternative actions or transitions between states. Partly due to the spatial layout of node-link diagrams, the exact data actions are often presented abstractly, in text or as glyphs [25]. In comparison, datamations only communicate *one version* of an analysis pipeline, but emphasize more of the semantics of analysis operations ("verbs") in the corresponding animations. In addition, most visual provenance tools operate within visual analytics interfaces and lack the flexibility of datamations, which translate potentially more complex code to animations.

3 DESIGN OF DATAMATIONS

Datamations are animations that explain data analysis pipelines. Our design of datamations is informed by an abstraction of states and transitions, summarized in Table 1. Given a data analysis pipeline, we define states as all intermediate and final data values, and transitions are operations on these data. From there, we build mappings from data values to plots and tables, as well as mappings from data operations to different types of animations. We provide a prototype implementation of datamations as an R package.[2]

Table 1: The datamations framework. Data values are mapped to states (shown as plots or tables) and operations on those values are mapped to transitions (shown as animations between plots or tables).

	Code	*map onto*	Visual
State	Data Values	→	Plot or Table
Transition	Data Operation	→	Animation

3.1 Data value-plot/table mapping

Datamations present intermediate and final data values as plots or tables. For tables, we assume that the data is in a "tidy" format [50], where each row corresponds to an observation and each column a variable, and there is one value per cell. The table visuals for datamations directly reflect the rows and columns in the data, supplemented with labels of variable names and values to help comprehension. Figures 5 and 6 are examples of table visuals, with the latter being a sample of the frames in the table-based datamations used in our experiments.

We have more flexibility when presenting data values as plots, and a large body of literature compares visualization types [5, 10, 28, 35] and even looks at automating the process of selecting the best type of visualization for a given dataset [39]. To facilitate the design

[2]https://github.com/seankross/datamations

Table 2: Data operations in datamations and their corresponding animations.

Operation		Animation
`group_by`	→	Translate observations to depict splitting into groups
`arrange`	→	Reorder observations to depict sorting by a variable
`mutate`	→	Highlight observations to depict creating or modifying variables
`filter`	→	Observations disappear to depict removal from dataset
`summarize`	→	Observations collapse into summary values for each group

of plot-based datamations, we make two design choices: first, we use natural-frequency encodings as often as possible, and second, we constrain the datamation so that the final plot (keyframe) is identical to the plot it is meant to explain. For frequency encodings, we use icon arrays and jittered scatter plots. The cells in tables and icon arrays, and the points in scatter plots, convey to the reader the sizes of different subgroups in the data, so that readers can intuitively judge the relative proportions of these groups. We use these frequency encodings because they have been shown to lead to improved comprehension [1, 20] and decision making in various settings [18, 22, 32]. Figures 2 and 3 are examples of plot-based visuals, with Figure 3 showing frames in the plot-based datamations used in our experiments.

3.2 Data operation-animation mapping

Datamations support a set of operations in data analysis pipelines as described in Table 2. Given the plots and tables we generate from data states, the process of animating often amounts to translating visual markers representing individual data points from one coordinate to the next. As discussed above, here we show operations corresponding to R's `tidyverse/dplyr` verbs [52] informed by data science tutorials[3] and common transition and interaction types in the visualization literature [17, 27], but the framework we outline is general purpose and not language specific. For instance, regardless of the language an analysis is carried out in, the idea that points might be grouped together and summarized by an average remains broadly applicable.

This abstraction also incorporates *modularity* in the composition of datamations, which facilitates their automatic generation. Data analysis pipelines are very flexible and small changes—placing an aggregation operation before or after a group by operation—can have large consequences, some of which introduce logical (but not syntactical) errors. Since datamation transitions are directly determined by data operations (*e.g.*, `summarize()`) and intermediate data states are already computed in the code, datamations automatically reflect any such changes in the underlying pipeline, which can have the effect of alerting users to these important differences.

3.3 Following animation design principles

We follow design principles and empirical findings in the visualization literature when creating datamations. Dragicevic *et al.* find that when animating between views, it is best to use "slow-in/slow-out" transitions [12]. Accordingly, we use an exponential function[4] for easing the transitions between datamation states. When there is grouping in the data, all groups move at once instead of one group after another. This design choice is based on there being little advantage to "staggering" group animations [8]. In addition, data points within a group move in the same way. This is supported by *Gestalt* psychology: people perceive things similar in motion as belonging to the same group [7].

If we view the analysis pipeline as a whole, datamations are staged in that they animate one data operation at a time. *Staging* makes the animation easier to follow [54]. Datamations also generate only *necessary motion* and *meaningful motion* [17], as each animated transition is directly linked to a step in the analysis.

4 OVERVIEW OF EXPERIMENTS

We present two experiments where we gathered data from over 1,200 participants to gain a deeper understanding of the potential benefits of (and issues with) datamations for explaining data analysis pipelines. Both experiments introduced the same scenario to each participant: they were presented with the results of a hypothetical survey about employment analyzed in two different ways, with a seemingly contradictory reversal of results in average earnings due to an instance of Simpson's paradox. As the main intervention, we showed participants these results as either a set of static images only or with a set of datamations. Participants were then asked whether it was possible that these two sets of results could have come from the same underlying dataset. After answering this question, we revealed to participants that both results were in fact from the same dataset, and asked participants to select an explanation (from a set of eight possible choices) for how this potentially puzzling outcome is possible. We concluded by asking participants for their preferences about datamations compared to static images and for free text feedback on datamations.

For our main analyses we tested two pre-registered[5] hypotheses with this design:

- **H1: Acknowledging that a reversal is possible.** Participants who see datamations of data analysis pipelines will be able to correctly identify that a reversal is possible in the dataset more often than participants who see only static images of final results.
- **H2: Identifying the correct explanation.** Participants who see datamations of data analysis pipelines will choose the correct explanation for the existence of the apparent paradox in the dataset more often than participants who see only static images of final results.

[3] https://rstudio.com/wp-content/uploads/2015/02/data-wrangling-cheatsheet.pdf
[4] https://gganimate.com/reference/ease_aes.html
[5] Pre-registration at: https://aspredicted.org/72qc9.pdf

We tested each of these hypotheses for two different settings in order to evaluate the two types of datamations we have developed: one where people are shown plots and another where they are shown tables. Plot-based datamations resemble animated scatterplots and point-range plots, and in this experiment they were tested against a static image of point-range plots. In Table-based datamations, cells in a data table are animated as part of a grid, and they were tested against a static image of a data table. In total we conducted two between-subjects experiments (one for plot-based figures and one for table-based figures), each with two conditions (comparing static images to datamations).

For both experiments, we report all sample sizes, conditions, data exclusions, and measures for the main analyses mentioned in the pre-registration document, with all data and analyses available as supplementary material.[6] We determined the total sample size of 1,300 participants for our experiments based on estimates from a pilot study to enable detection of a difference of at least 10 percentage points in the probability of correct answers between the static final figures condition and the datamation condition with 80% power in a one-sided[7] comparison of proportions using the `prop.test()` function in R. All plots and datamations described below visualize synthetic data that we created for the purpose of being used in these experiments. Our experiments were approved by the IRB committee at Microsoft Research. A live version of the experiments can be found online[8], and the plots, tables, and animated GIFs containing datamations used as stimuli in the experiments are available in the online supplement. We ran the two experiments using one common task on Amazon's Mechanical Turk for ease of administration, but describe and analyze them separately (as declared in our pre-registration) as they apply to different use cases (plots versus tables). As per our pre-registration, we excluded the 27 participants who took part in previous pilots of the experiment before conducting any analyses.

5 EXPERIMENT 1: PLOT-BASED DATAMATIONS

The goal of our experiments is to understand whether datamations help people better comprehend the plots and figures they are shown by exposing more information about the underlying data analysis pipeline that led to these figures. To evaluate this we designed experiments to see if datamations could help people understand an occurrence of Simpson's paradox within a dataset compared to a static figures containing the same results. We chose to study an instance of Simpson's paradox because it is a case where awareness of the underlying data analysis pipeline is crucial for resolving seemingly contradictory results.

We created a synthetic dataset which contains the results of a survey about employment, where each respondent provided information about whether they have a master's degree or a PhD, whether they work in industry or in academia, and their annual salary. This dataset was constructed to show that on average respondents with master's degrees made more money than respondents with PhDs, however when the data are grouped according to whether respondents work in academia or industry, then PhDs make more money on average within each group. In this experiment we created static plots that showed this reversal along with datamations that did the same, and randomly exposed participants to one of these two stimuli.

5.1 Stimuli

The first stimulus for this experiment (the entirety of Figure 2) is a static image of two point range plots that are the result of a modest data analysis pipeline. Both plots show average salaries and standard errors between groups of different degree holders, and the second plot further distinguishes between degree and work setting (academia versus industry).

The second stimulus for this experiment (illustrated in Figure 3) is a looping animated GIF file meant to show transitions between different parts of the a data analysis pipeline, which ultimately ends in a plot similar to what is displayed in Figure 2. While the static version shows two charts, this version shows two datamations: one that only accounts for degree when calculating the average salary, and a second datamation that shows the contrasts for both degrees in industry and in academia separately. These plot-based datamations start as a grid of points that are then colored and arranged according to what type of degree an individual has (in the first case) or their degree and their work setting combined (in the second). These points then shift into a bee swarm plot that show salary on the Y axis and category (according to degree and work setting) on the X axis. Finally these points contract onto one point representing the group average salary, ending the transformation as a point range plot.

5.2 Procedure

First we provided participants with an overview of the study and presented them with a consent form. Specifically we told participants that they would be asked questions about a dataset and that they should not consult external resources to formulate their answers. After this introduction participants saw the following:

> Imagine that you are an analyst working for a think tank. You conducted a salary survey with 100 respondents in June 2018. Each respondent worked in either industry (companies) or academia (colleges and universities) at the time of the survey. Also, each respondent had either a master's or a PhD degree. Each of the 100 respondents reported:
>
> - **Work setting**: whether they worked in academia or industry at the time of the survey.
> - **Degree**: their highest education level obtained (master's or phd degree).
> - Their current annual **salary**.

After clicking a button acknowledging that they had read this information, participants were shown one of two conditions: either a static image, or a series of datamations. The static image that participants saw is in Figure 2, and an illustration of the series of datamations that participants saw can be found in Figure 3. For participants who saw datamations, they were presented with six datamations in sequence, with each datamation corresponding to

[6] See https://osf.io/85njc/ for all supplemental material.
[7] We conducted one-sided tests because we are only interested in whether datamations improve upon the status quo of static figures.
[8] https://jhofman.github.io/datamations-chi2021-paper/simpsons_multiple_choice_all/ randomly redirects to one of the experimental conditions.

Figure 2: In our first experiment we tested whether plot-based datamations help participants to understand data analysis pipelines. Participants were randomly assigned to see either two static plots (shown in this figure) or a animation (illustrated in Figure 3).

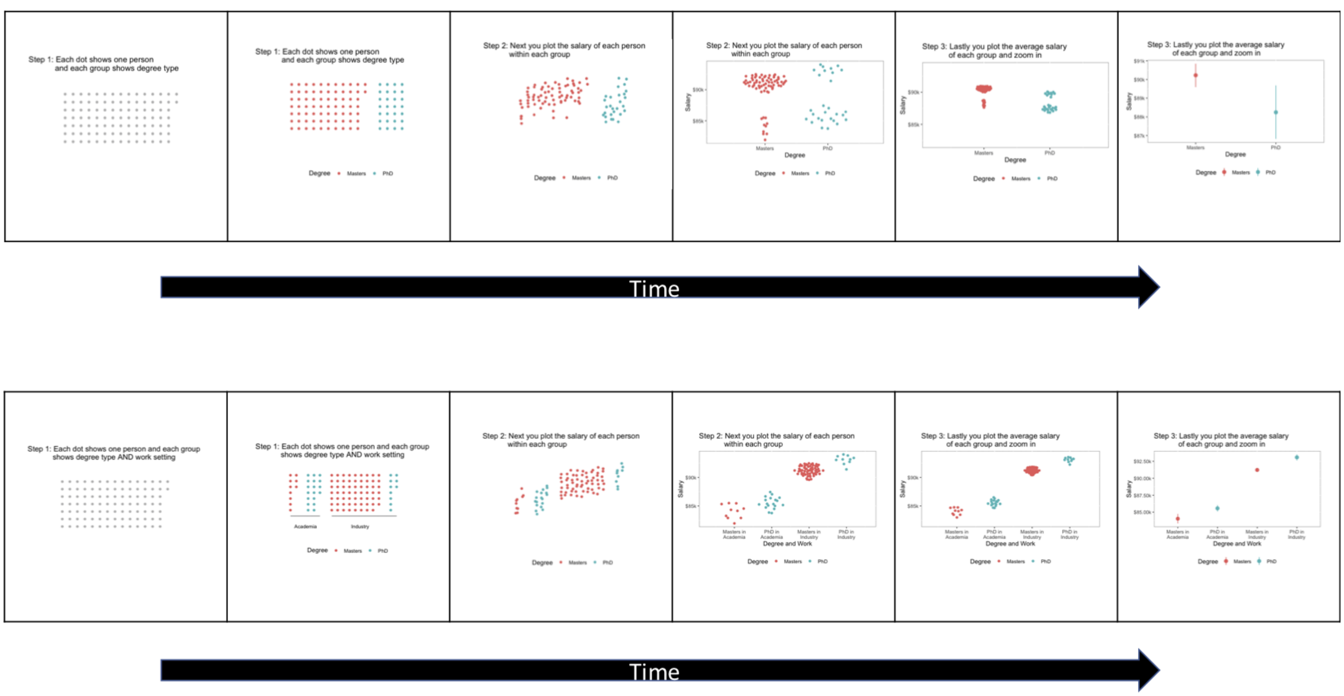


Figure 3: The timeline above illustrates two series of animations that were shown to participants in sequence. In the first datamation, individuals in the survey are displayed as grey circles, then they are colored in according to whether they have a master's degree or a PhD. The points then move to show a bee swarm plot according to each individual's salary by their degree. Finally, the points in the bee swarm converge to show line plots featuring means and 95% confidence intervals. The average salary for an individual with a master's degree appears to be higher than the average salary for someone with a PhD. The second datamation is similar to the first, except each point is separated further according to whether that individual works in industry or academia. In the last frame we can see that the trend is reversed: it appears PhD holders make more money on average than people with master's degrees.

a step in the data analysis pipeline. Each of these datamations played on a loop so that the animation restarted after it reached the last frame. After seeing each abbreviated datamation in sequence, participants in this condition saw the entire datamation pieced together, showing an animation of the entire data analysis pipeline. Finally participants in the datamation condition were shown both the complete datamations and the final static images in Figure 2.

After seeing either only a static image in one condition, or a series of datamations and a static image, participants were asked:

> Note that compared to people with PhDs, people with master's degrees make more in the left chart, but make less in the right chart. Does it seem impossible to you that the results would come out this way?

Participants could wither select *Yes, it seems impossible* or *No, it seems possible*, the latter being the correct answer.

After submitting their answer participants were provided with the correct answer (that the results are in fact from the same dataset, without any data exclusions or manipulations) and then asked to select one of eight multiple choice explanations that they thought best accounted for the seemingly paradoxical reversal. The text informing them of the correct answer is below:

> **It turns out that these two charts are made up of the exact same data, just grouped differently.** That is, 100 people's salaries are represented in the left graph and the same 100 people's salaries are represented in the right graph. There is no mistake in the charts, but it may seem like a **paradox** when both are true:
>
> - The left chart shows that people with master's degrees make **more** money than people with PhD degrees on average.
> - The right chart shows that people with master's degrees make **less** money than people with PhD degrees on average, both inside industry and inside academia.
>
> Which of the following could explain how both statements are true? **This is the main point of this experiment. Please take it seriously.**

Participants could then select from the following answer choices:

- Most people with a master's degree work in industry, which pays more and drives up the average master's salary in the left chart.
- People with neither master's nor PhDs are factored into the right chart, which biases the averages in opposing directions.
- In the right chart, there are many kinds of industry jobs, but fewer kinds of academic jobs.
- The differences in the left chart are not statistically significant and therefore could be due to chance.
- Due to outliers, the master's point on the left chart can be higher than both master's points on the right chart.
- The chart on the left includes salaries from people who work in neither industry nor academia and who are unrepresentative of the general trend.
- The left chart shows data from more respondents than the right chart, so it is not appropriate to compare the two charts.
- None of the above explain the difference.

The correct answer is the first answer we show here, however the order of all of the answer choices (except for "None of the above") were randomized for each participant. We derived these answer choices from a pilot study we conducted using Amazon Mechanical Turk workers where we solicited free response explanations for what could be causing the apparent paradox and coded them into these eight categories.

After submitting their answers participants were asked to indicate whether they preferred static charts or datamations, and they were encouraged to explain their preference in free response text.

5.3 Participants

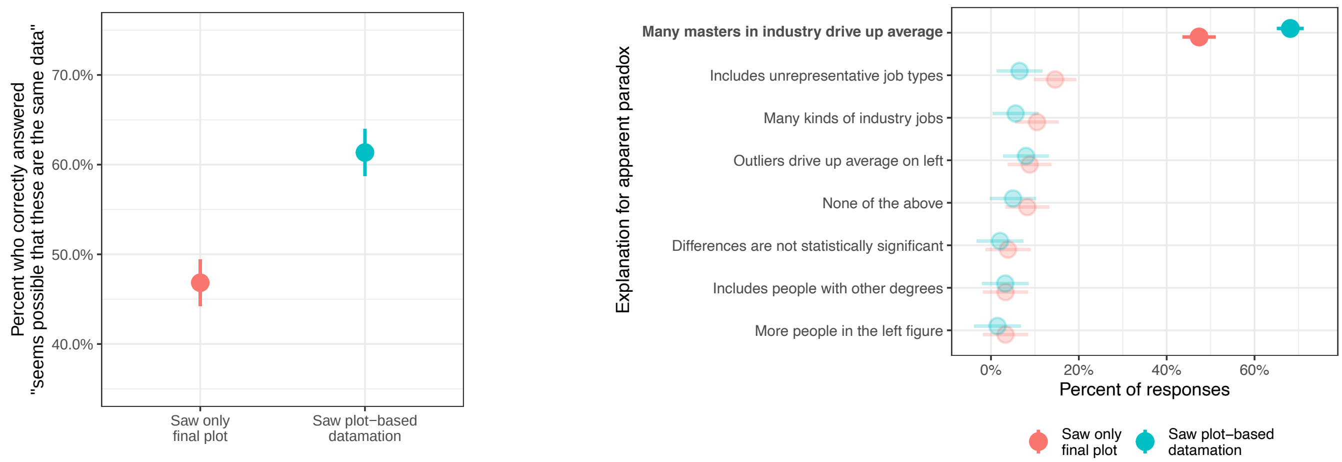
We recruited participants from Amazon Mechanical Turk and, after excluding those who had taken part in any pilots of this study, randomly assigned 368 participants into the condition that only saw static plots and 340 participants into the condition that saw datamations. Each task was available to Turk workers with an approval rating greater than or equal to 99%, and to be eligible to participate the worker had have previously completed at least 100 tasks on AMT. Workers were paid a one-time fee of $1.50.

5.4 Results

After collecting responses from all participants, we conducted the analyses specified in our preregistration plan. Accordingly, we removed participants who finished the experiment too quickly (five participants who finished in under 45 seconds for the static condition, and one who finished in under 90 seconds for the datamations condition). This left us with 363 participants who saw only the final static plots and 339 participants who saw datamations. Median completion time for the experiment was 6.9 minutes.

Acknowledging that a reversal is possible. We first looked at participants' ability to correctly identify that a reversal in average earnings between workers with masters and PhD is possible based on whether one conditions on the work setting or not. For each condition we computed the fraction of participants who correctly stated that "it seems possible" that results could come out this way. As shown in Figure 4a, we find that that while only 47% of participants who saw only the final plots answered this question correctly, 61% of participants who saw plot-based datamations did so, a statistically significant ($\chi^2(1, N = 702) = 14.30, p < .001$, one-tailed) and sizeable difference of 14 percentage points (Cohen's h = 0.29).

These results support our hypothesis that plot-based datamations can improve people's ability to understand data analyses over static data visualizations alone. Why might this be? While our experiments were not designed to uncover a precise mechanism to explain our results, we can offer a guess. In this case, participants saw two sets of datamations: one for salaries split only by degree type and one split on degree type and work setting. While the final plots showed different directional trends that might be puzzling on their own, *the initial frames of each of these datamations are identical*, and the animations showed smooth transitions from these identical initial frames to the final ones. Although this does not prove to the reader that both datasets are identical or that the reversal is possible, it might nonetheless increase the chances that



(a) **(b)**

Figure 4: Main results of Experiment 1, comparing static plots to plot-based datamations. The left panel shows the fraction of participants who were correctly able to resolve Simpson's paradox in each condition, stating that it was not impossible that the two different plots they were shown could have come from the same dataset. The right panel shows the explanation participants chose to resolve the paradox from a multiple choice list. The top option (in bold) is the correct one. Error bars in both plots show one standard error on the estimated mean.

participants find it plausible that this is the case, leading to a boost in correct answers to this question.

This potential explanation relies on a simple strength of datamations over static plots: they offer strictly more information about the underlying analyses than plots alone. This of course could be done by making a series of plots for different stages of an analysis with text explaining the relationship between these plots, but this can quickly become cumbersome for authors to produce and for readers to consume. Datamations, in contrast, offer an easy and compelling way for this information to be shared with readers.

Identifying the correct explanation. After answering the identification question all participants were told that both figures they saw were in fact from the same underlying dataset and asked to choose one of eight multiple choice explanations for how this could be the case. Figure 4b shows the the distribution of explanations chosen within each of the two conditions, ranked from most to least commonly chosen. The answer on top—that "Most people with a master's degree work in industry, which pays more and drives up the average master's salary in the left plot"—is the correct one. Comparing the top-most blue and red points, we again see a large (21 percentage point, Cohen's $h = 0.42$) and statistically significant ($\chi^2(1, N = 702) = 30.05, p < .001$, one-tailed) improvement between those who saw datamations and those who did not. Even after being told that these were in fact the same dataset, under half (47%) of people who saw only the final plots answered this question correctly, whereas more than two thirds (68%) of people who saw datamations did so.

The distribution over the remaining (incorrect) answers provides some additional insights as to why datamations might help readers understand data analyses. The two most frequent incorrect responses to this question for participants who saw only the final plots were explanations that involved heterogeneity in or representativeness of the data. The next involved the presence of outliers. As discussed with the previous question, one explanation consistent with these results is that participants who saw datamations were shown smooth transitions between identical initial frames and final results, reducing the chances that they thought differences in the datasets were responsible for the reversal.

6 EXPERIMENT 2: TABLE-BASED DATAMATIONS

This experiment mirrored our first experiment, but was designed for settings where people are presented with tables instead of plots. As many of the details are the same, here we simply discuss the changes that were made from the plot- to table-based setting.

6.1 Stimuli

The first stimulus for this experiment, seen in Figure 5, shows two tables that could be computed at the end of a relatively small data analysis pipeline. The first table shows the average salary for workers according to what kind of degree they have, while the second table shows the average salary for combinations of degree and work setting. These tables and the following datamations were created using the same synthetic salary survey data from the first experiment.

The second stimulus for this experiment, which is illustrated in Figure 6, is another looping animated GIF that highlights the transitions and transformations required to go from the raw table of survey data to a summarized table that contains average salaries for each group. Both datamations start with the same raw table of survey data, but then they differentiate depending on how many

Figure 5: For our second experiment we tested whether tabular datamations help participants to understand data analysis pipelines. Participants were randomly assigned to see either two images of tables (shown in this figure) or a animation (illustrated in Figure 6).

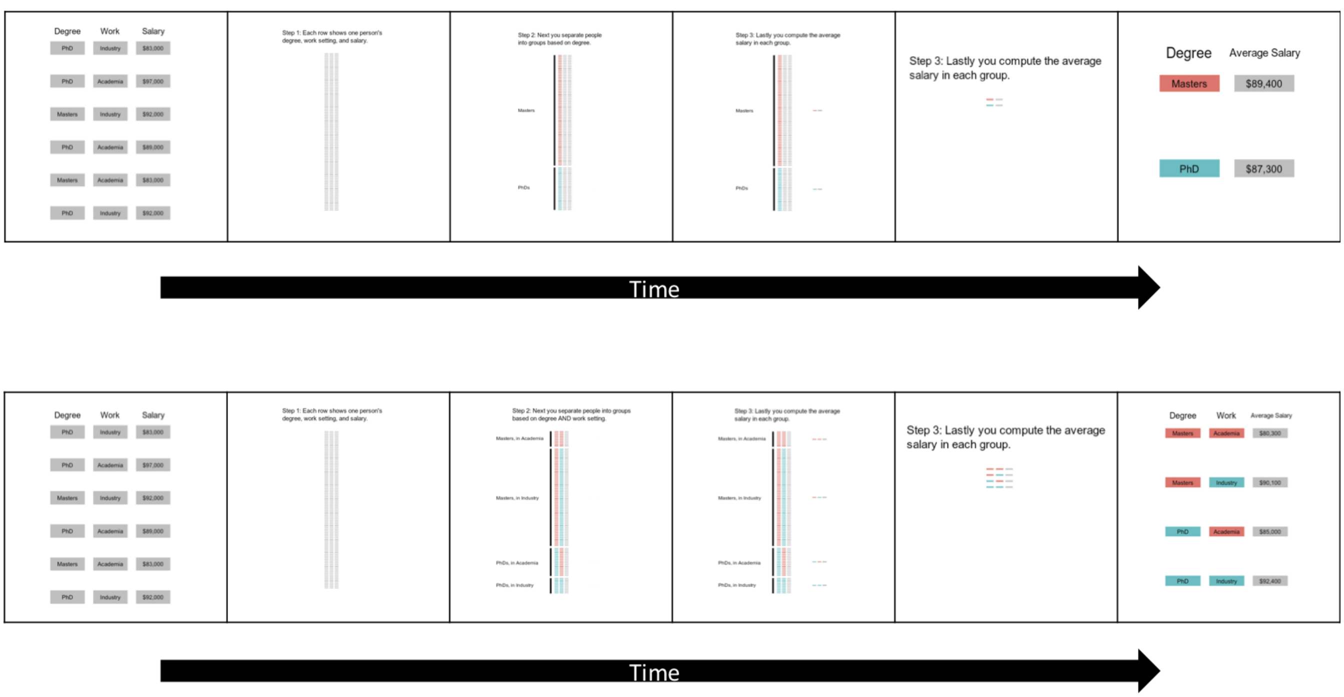


Figure 6: This timeline illustrates the tabular datamations in our second experiment with two series of animations that were shown to participants in sequence. In the top datamation every survey respondent is represented by a row, where the first column in that row shows their degree, the second column shows whether they work in academia or industry, and the third column shows their salary. The datamation then zooms out to show the entire table, and the table cells in the first column are colored in depending on which degree an individual has. New, summarized values representing the average salary of each group then appear to the right of the table. The original table then disappears and the table of summarized average values is centered. Finally, the datamation zooms back in to show the summarized table with the average value for each type of degree. The mechanics of the bottom datamation are similar, however both the columns representing degree and work environment are colored in, and groups of cells are distinguished by the four possible combinations of degrees and work settings. Finally the final table shows average salaries for each pair of degree type and work setting. In the last frame we can see that the trend from the first datamation is reversed: PhD holders make more money on average than people with master's degrees in both academia and industry.

subgroups are required by the analysis: there are only two subgroups when calculating the average salary across two different types of degrees, however four subgroups are highlighted when calculating average salary across degree types and work settings. Once the average salary values have been calculated the raw values fade away and the summarized values take focus. The two static tables are the end frames for the two datamations: both tables are transformed differently depending on how they are grouped.

6.2 Procedure

The experiment for tabular datamations proceeded similarly to the experiment for plot based datamations. Participants received the same information about our study up front and were presented with the same consent form. In the first section of the study they were presented with the same prompt (*Imagine that you are an analyst working for a think tank...*) and after acknowledging they had read that information they were in either shown static images (featured in Figure 5) or a series of tabular datamations, illustrated in Figure 6.

As in the first experiment, participants who saw datamations were presented with six datamations in sequence, with each datamation corresponding to a step in the data analysis pipeline. Each of these datamations played on a loop. Participants in this condition then saw the datamations pieced together. Finally both the complete datamations and the final static images (as in Figure 2) were shown to the participants in the datamation condition, whereas participants in the other condition saw only the final images.

Participants were asked the same question they were asked in the first experiment (*Does it seem impossible to you that the results would come out this way?*) except the language was changed slightly to discuss tables instead of plots. After submitting their response participants were provided with the correct answer and then asked to select one of eight multiple choice explanations from the first experiment except with different language talking about tables instead of plots where appropriate. The order of answer choices was randomized for each participant. Finally participants were asked to indicate whether they preferred static figures or datamations, and they could provide a written explanation.

6.3 Participants

As in the first experiment we recruited participants from Amazon Mechanical Turk and randomly assigned 298 participants into the condition that only saw static tables and 267 participants into the condition that saw tabular datamations. Workers were held to the same standard as they were in the first experiment: they had to have 99% or better approval ratings, experience completing 100 or more tasks, and they could not have participated in our previous pilots. Workers were paid a one-time fee of $1.50.

6.4 Results

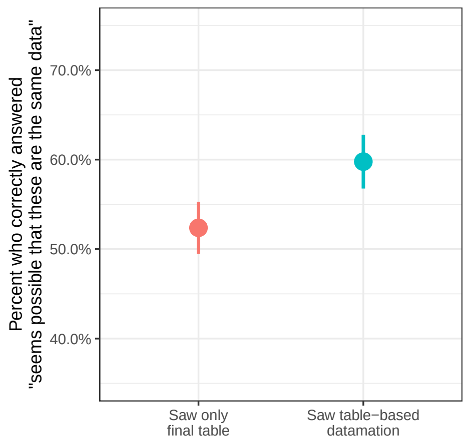
After collecting responses from all participants, we again removed participants who finished the experiment too quickly (four participants who finished in under 45 seconds for the static condition, and one who finished in under 90 seconds for the datamations condition). This left us with 294 participants who saw only the final static tables and 266 participants who saw datamations. Median completion time for the experiment was 6.9 minutes.

Acknowledging that a reversal is possible. As with the previous experiment, we first looked at people's ability to correctly identify that a reversal is possible. As shown in Figure 7a, we see a similar boost in ability to answer this question correctly for those who saw datamations compared to those who saw only final tables, supporting our first hypothesis. While only 52% of participants who saw static tables recognized that the reversal was possible, 60% of participants who saw datamations did, a statistically significant ($\chi^2(1, N = 560) = 2.80, p = .05$, one-tailed) 8 percentage point difference (Cohen's $h = 0.15$).

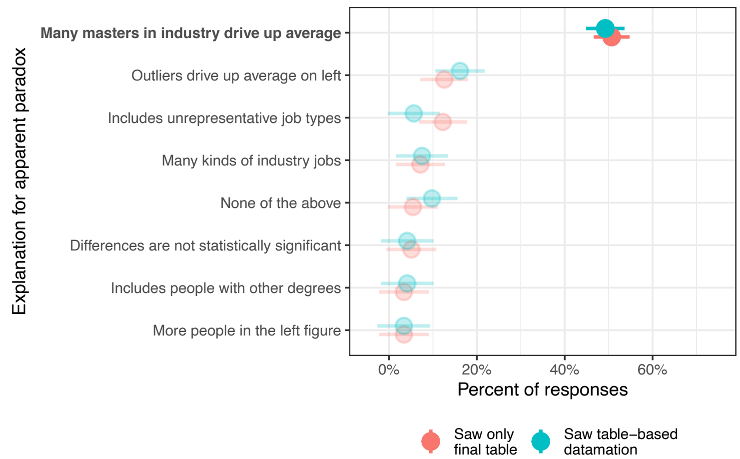
One explanation consistent with these results is that improvements are due to the increased transparency offered by datamations: participants saw identical starting frames of the animations, which could have increased the chances that they would recognize these results as possibly coming from the same dataset. That said, in these table-based datamations, only a subset of the actual salary values are visible, as it quickly becomes prohibitive to display hundreds of values in plain text such that readers can actually see and process them. As a result, this leaves reasonable ambiguity as to whether the datasets in the initial frames of the datamations are indeed identical. Nonetheless, we see that table-based datamations can help readers recognize that Simpson's reversal is indeed possible.

Identifying the correct explanation. After all participants were told that the datasets were in fact identical and the reversal was possible, they again chose one of eight explanations for why the reversal occurs. Figure 7b shows the distribution of responses within each condition, ranked from most to least popular, with the correct answer in bold on the top. Comparing the the top-most blue and red points, we see a lack of support for our second hypothesis, as we find no evidence of a statistically significant difference ($\chi^2(1, N = 560) = 0.06, p = .60$, one-tailed) between the proportion of participants who chose the right answer (51% for those who saw only final tables versus 49% for those who saw datamations).

It is difficult to explain why we see that datamations boost ability to identify that the reversal is possible, but see no evidence for an improvement in ability to explain why the reversal occurs. That said, we can offer two potential explanations. The first is simply that laypeople may be less comfortable or facile with tabular representations of data. The second is based on the same ambiguity mentioned above: while table-based datamations show the relative sizes of different subgroups in the data, they may be unable to clearly convey all of the underlying individual salary information. The cells can simply become too small to insert the salaries as text when several hundred rows are shown, and even if they were visible they would likely be difficult for participants to process holistically. If this is correct, it may be the case that table-based datamations are good for some users (e.g., data scientists) and not others, or may be more helpful for certain types of data analysis operations than others (e.g., joins or data reshaping) not explored here.



(a)



(b)

Figure 7: Main results of Experiment 2, comparing static tables to table-based datamations. The left panel shows the fraction of participants who were correctly able to resolve Simpson's paradox in each condition, stating that it was not impossible that the two different figures they were shown could have come from the same dataset. The right panel shows the explanation participants chose to resolve the paradox from a multiple choice list. The top option (in bold) is the correct one. Error bars in both plots show one standard error on the estimated mean.

7 FUTURE PREFERENCES AND FEEDBACK ON DATAMATIONS

As mentioned above, the final page in both experiments asked participants for their future preferences about datamations. Regardless of which condition participants were assigned to, everyone was shown both static figures (either plots or tables, depending on the experiment) and the corresponding datamations, and asked whether they would like to see future figures presented as either a) only static or b) static "accompanied by animations like these". As shown in Figure 8, close to two thirds of participants who saw plot-based datamations expressed that they would like to see these animations in the future, while almost half of participants who saw table-based did.

After making this choice, participants were asked to provide free text feedback about whether these animations were helpful or not. While a good deal of the feedback was generally supportive of the goal behind datamations and in line with insights from our main analyses ("It makes me think more about the data rather than the static charts", "It gives a good visual of how the change takes place when averaging on the chart", "Since it shows exactly how the charts are made it gives a little bit more information and understanding to the reader"), there were also informative critiques that offered insights for how datamations could be improved.

One fairly common piece of feedback was that refined timing of the animations and the ability to control playback would be very helpful ("It is easier for me to read the tables rather than animations like these because I can read at my own pace and not have to feel rushed to obtain the information", "I would have preferred to be able to see each static slide as well, or to have some control over the animation (pause, slow down, rewind, etc.)", "I think they're helpful, but I would like to be able to pause them so I could read and compare them").

In designing these experiments, we initially considered including controls for playback with the datamations, but ultimately decided against them to keep the experiment as tightly controlled as possible. This provided all participants with the same experience and avoided concerns about endogeneity, but it would be interesting to conduct further studies on how datamations with playback controls perform in terms of both the main hypotheses we studied and people's preferences going forward. Another issue with timing may be that participants were Mechanical Turk workers who are often trying to complete as many microtasks as possible in a given time frame to earn the highest hourly wage possible. As such, it is likely the case that our participants had less tolerance for watching animations play out than, say, a reader who opts into viewing an article containing a datamation because they are inherently interested in the topic it concerns.

Finally, there were several participants who pointed out issues in reading details of the table-based datamations ("I like the animation, but it is very small and hard to read so would prefer the static tables", "I liked the tables and animation, but they were a little small"). While we believe that some of these concerns could be addressed with different design choices in our stimuli, we also recognize that as datasets grow in size, it becomes increasingly difficult to create datamations that contain all of the underlying data, especially in the table-based animations. In these settings we might consider aggregation or sampling to reduce the amount of information shown to readers, and leave this as future work.

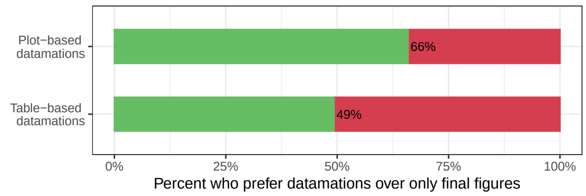


Figure 8: People's preferences for seeing datamations in the future for both the plot-based and table-based experiments.

8 DISCUSSION

In this paper we have described an opportunity to improve how data analyses are communicated to readers through the use of datamations, which are animations that reveal details of the data analysis pipeline that led to a given plot or table. We presented a modular framework that allows analysts to programmatically translate data analysis code into datamations that attempt to explain that code. We then used this framework to generate datamations and ran experiments comparing them to static plots and tables.

The results of our experiments show that datamations, in most cases, improved people's performance on test items about a subtle data analysis pipeline. Specifically, through two, large-scaled preregistered experiments we found evidence that both the plot-based and table-based datamations we showed people offered sizable increases on comprehension questions about an instance of Simpson's paradox. Furthermore, many participants reported gaining important insights from these datamations as well as a preference for seeing animations like these in the future. Participants also provided helpful critiques of the specific implementations of datamations we showed them and how they could be improved. Key insights are that the timing of the animations and ability to control their playback are important for an enjoyable and informative user experience. Our experiments also surfaced the insight that table-based datamations can become difficult to read with medium to large datasets due to constraints in displaying many small table cells at once.

There are, of course, many limitations to this work, as we have explored only a small subset of possible research directions related to the task of communicating and explaining data analysis pipelines. Specifically, we looked at how datamations help one population (laypeople being paid to participate in a lab experiment) to perform one task (resolve Simpson's paradox) with a particular visual implementation (self-playing GIFs with the visual transitions we tested). These experiments have helped us answer some questions, but raise many others.

First, even in this particular setting there are several additional questions one might ask. For instance, did participants truly "understand" the datamations they saw? Unfortunately our experiments cannot answer this question, as comprehension is abstract construct that cannot be directly measured or assessed. One could, however, imagine more detailed studies designed to assess what participants do and don't take away from datamations using techniques such as think-aloud protocols and in-depth interviews. Likewise, what is the precise mechanism responsible for results we saw, and are there alternative, possibly simpler, interventions for communicating data analysis pipelines that could be just as effective as datamations? Again, our experiments cannot answer these questions, as they were designed to compare the status quo (static plots and tables) to what we thought was a viable, practical alternative (animated GIFs) for the purpose of detecting if an effect exists at all. Now that we know this is the case, one could conduct more tightly controlled experiments to isolate effects and identify underlying mechanisms. For instance, is it crucial that datamations are animated, or even visual? Or could a series of static panels depicting a data analysis pipeline or a simple paragraph of descriptive text be just as effective? And what future design choices could make datamations even more effective than those tested here? Some of these questions are notoriously difficult to answer [48], but might be appealing subjects for future research.

Second, beyond the setting we studied here there are several ways this work could be extended by testing datamations on different audiences, for different purposes, and with different implementation details. For instance, it would be interesting to see how datamations could be used to teach students learning data analysis about different concepts, or to see how seasoned data scientists make use of datamations for understanding and debugging code they are writing. One could also evaluate datamations for a host of other types of data analysis pipelines—for instance involving more complex operations like data joins, reshaping, modeling, statistical estimation, etc.—and investigate for which settings they do (and don't) provide value over the status quo. Likewise, there are many opportunities to add to, experiment with, and optimize the visual elements and transitions used in datamations, akin to the research done in the HCI community to improve the details of static visualizations. Finally, there is the opportunity to develop software packages that make it easier for developers to explain their data analysis pipelines to their audiences. We have created one such implementation based on the framework described here, but there are many ways it can be extended in the future. It is currently designed to work for R's tidyverse, but we hope to see it extended to other programming languages and packages. We view this as the first of many steps towards developing more tools to explain data analysis pipelines and their results to students, analysts, and their audiences.

REFERENCES

[1] Richard B. Anderson, Laura Marie Leventhal, Don C. Zhang, Jr Daniel Fasko, Zachariah Basehore, Christopher Gamsby, Jared Branch, and Timothy Patrick. 2019. Belief bias and representation in assessing the Bayesian rationality of others. *Judgment and Decision Making* 14, 1 (2019), 1–10. https://econpapers.repec.org/article/jdmjournl/v_3a14_3ay_3a2019_3ai_3a1_3ap_3a1-10.htm
[2] Benjamin Bach, Emmanuel Pietriga, and Jean-Daniel Fekete. 2014. GraphDiaries: Animated Transitions andTemporal Navigation for Dynamic Networks. *IEEE Transactions on Visualization and Computer Graphics* 20, 5 (May 2014), 740–754. https://doi.org/10.1109/TVCG.2013.254
[3] Stefan Milton Bache and Hadley Wickham. 2014. *magrittr: A forward-pipe operator for r.* manual. RStudio. https://CRAN.R-project.org/package=magrittr
[4] Emily Badger, Claire Ann Miller, Adam Pearce, and Kevin Quealy. 2018. Income Mobility Charts for Girls, Asian-Americans and Other Groups. Or Make Your Own. *New York Times* -, - (March 2018), –. https://www.nytimes.com/interactive/2018/03/27/upshot/make-your-own-mobility-animation.html
[5] Enrico Bertini, Michael Correll, and Steven Franconeri. 2020. Why Shouldn't All Charts Be Scatter Plots? Beyond Precision-Driven Visualizations. *arXiv:2008.11310 [cs]* -, - (Aug. 2020), –. http://arxiv.org/abs/2008.11310 arXiv: 2008.11310.
[6] Christian Bors, Theresia Gschwandtner, and Silvia Miksch. 2019. Capturing and Visualizing Provenance From Data Wrangling. *IEEE Computer Graphics and Applications* 39, 6 (2019), 61–75.

[7] Amira Chalbi, Jacob Ritchie, Deokgun Park, Jungu Choi, Nicolas Roussel, Niklas Elmqvist, and Fanny Chevalier. 2019. Common Fate for Animated Transitions in Visualization. *IEEE Transactions on Visualization and Computer Graphics* 26, 1 (2019), 1–1. https://doi.org/10.1109/TVCG.2019.2934288

[8] Fanny Chevalier, Pierre Dragicevic, and Steven Franconeri. 2014. The Not-so-Staggering Effect of Staggered Animated Transitions on Visual Tracking. *IEEE Transactions on Visualization and Computer Graphics* 20, 12 (Dec. 2014), 2241–2250. https://doi.org/10.1109/TVCG.2014.2346424

[9] Fanny Chevalier, Nathalie Henry Riche, Catherine Plaisant, Amira Chalbi, and Christophe Hurter. 2016. Animations 25 Years Later: New Roles and Opportunities. In *Proceedings of the International Working Conference on Advanced Visual Interfaces - AVI '16*. ACM Press, Bari, Italy, 280–287. https://doi.org/10.1145/2909132.2909255

[10] William S. Cleveland and Robert McGill. 1984. Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods. *J. Amer. Statist. Assoc.* 79, 387 (Sept. 1984), 531–554. https://doi.org/10.1080/01621459.1984.10478080

[11] Zachary T Cutler, Kiran Gadhave, and Alexander Lex. 2020. Trrack: A Library for Provenance Tracking in Web-Based Visualizations. In *OSF Preprints*. Open Science Framework, -, 5. https://doi.org/10.31219/osf.io/wnctb

[12] Pierre Dragicevic, Anastasia Bezerianos, Waqas Javed, Niklas Elmqvist, and Jean-Daniel Fekete. 2011. Temporal distortion for animated transitions. In *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11*. ACM Press, Vancouver, BC, Canada, 2009. https://doi.org/10.1145/1978942.1979233

[13] Matt Dray. 2019. Fix leaky pipes in R. https://www.rostrum.blog/2019/04/07/fix-leaky-pipes/

[14] Steven M. Drucker and Roland Fernandez. 2015. *A Unifying Framework for Animated and Interactive Unit Visualizations*. Technical Report. Microsoft.

[15] Tommy Ellkvist, David Koop, Erik W. Anderson, Juliana Freire, and Cláudio Silva. 2008. Using provenance to support real-time collaborative design of workflows. In *International Provenance and Annotation Workshop*. Springer, Berlin, Heidelberg, 266–279.

[16] Michael Fernandes, Logan Walls, Sean Munson, Jessica Hullman, and Matthew Kay. 2018. Uncertainty Displays Using Quantile Dotplots or CDFs Improve Transit Decision-Making. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, 144:1–144:12. https://doi.org/10.1145/3173574.3173718 event-place: Montreal QC, Canada.

[17] Danyel Fisher. 2010. Animation for visualization: Opportunities and drawbacks. In *Beautiful Visualization*. O'Reilly Media, E-book, 24.

[18] Mirta Galesic, Rocio Garcia-Retamero, and Gerd Gigerenzer. 2009. Using Icon Arrays to Communicate Medical Risks: Overcoming Low Numeracy. *Health Psychology* 28, 2 (2009), 210–216. https://doi.org/10.1037/a0014474 ISBN: 0278-6133.

[19] T. Ge, Y. Zhao, B. Lee, D. Ren, B. Chen, and Y. Wang. 2020. Canis: A High-Level Language for Data-Driven Chart Animations. *Computer Graphics Forum* 39, 3 (June 2020), 607–617. https://doi.org/10.1111/cgf.14005

[20] Gerd Gigerenzer and Ulrich Hoffrage. 1995. How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review* 102, 4 (1995), 684–704. https://doi.org/10.1037/0033-295X.102.4.684 ISBN: 0033-295X\r1939-1471.

[21] Yue Guo, Carsten Binnig, and Tim Kraska. 2017. What you see is not what you get! Detecting Simpson's Paradoxes during Data Exploration. In *Proceedings of the 2nd Workshop on Human-In-the-Loop Data Analytics (HILDA'17)*. Association for Computing Machinery, New York, NY, USA, 1–5. https://doi.org/10.1145/3077257.3077266

[22] Sarah T. Hawley, Brian Zikmund-Fisher, Peter Ubel, Aleksandra Jancovic, Todd Lucas, and Angela Fagerlin. 2008. The impact of the format of graphical presentation on health-related knowledge and treatment choices. *Patient Education and Counseling* 73, 3 (2008), 448–455. https://doi.org/10.1016/j.pec.2008.07.023 ISBN: 0738-3991 (Print)\r0738-3991 (Linking).

[23] Jeffrey Heer. 2020. An Illustrated Guide to Arquero Verbs. https://observablehq.com/@uwdata/an-illustrated-guide-to-arquero-verbs

[24] Jeffrey Heer and Michael Bostock. 2010. Crowdsourcing graphical perception: using mechanical turk to assess visualization design. In *Proceedings of the 28th international conference on Human factors in computing systems - CHI '10*. ACM Press, Atlanta, Georgia, USA, 203. https://doi.org/10.1145/1753326.1753357

[25] Jeffrey Heer, Jock Mackinlay, Chris Stolte, and Maneesh Agrawala. 2008. Graphical histories for visualization: Supporting analysis, communication, and evaluation. *IEEE transactions on visualization and computer graphics* 14, 6 (2008), 1189–1196. Publisher: IEEE.

[26] J. Heer and G. Robertson. 2007. Animated Transitions in Statistical Data Graphics. *IEEE Transactions on Visualization and Computer Graphics* 13, 6 (Nov. 2007), 1240–1247. https://doi.org/10.1109/TVCG.2007.70539

[27] J. Hullman, S. Drucker, N. Henry Riche, B. Lee, D. Fisher, and E. Adar. 2013. A Deeper Understanding of Sequence in Narrative Visualization. *IEEE Transactions on Visualization and Computer Graphics* 19, 12 (Dec. 2013), 2406–2415. https://doi.org/10.1109/TVCG.2013.119

[28] Jessica Hullman, Paul Resnick, and Eytan Adar. 2015. Hypothetical Outcome Plots Outperform Error Bars and Violin Plots for Inferences about Reliability of Variable Ordering. *PLOS ONE* 10, 11 (Nov. 2015), e0142444. https://doi.org/10.1371/journal.pone.0142444

[29] Christopher D. Hundhausen, Sarah A. Douglas, and John T. Stasko. 2002. A Meta-Study of Algorithm Visualization Effectiveness. *Journal of Visual Languages & Computing* 13, 3 (June 2002), 259–290. https://doi.org/10.1006/jvlc.2002.0237

[30] Nazanin Kadivar, Victor Chen, Dustin Dunsmuir, Eric Lee, Cheryl Qian, John Dill, Christopher Shaw, and Robert Woodbury. 2009. Capturing and supporting the analysis process. In *2009 IEEE Symposium on Visual Analytics Science and Technology*. IEEE, Atlantic City, New Jersey, 131–138.

[31] A. Kale, F. Nguyen, M. Kay, and J. Hullman. 2019. Hypothetical Outcome Plots Help Untrained Observers Judge Trends in Ambiguous Data. *IEEE Transactions on Visualization and Computer Graphics* 25, 1 (Jan. 2019), 892–902. https://doi.org/10.1109/TVCG.2018.2864909

[32] Matthew Kay, Tara Kola, Jessica R Hullman, and Sean A Munson. 2016. When (ish) is My Bus?: User-centered Visualizations of Uncertainty in Everyday, Mobile Predictive Systems. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*. Association for Computing Machinery, New York, NY, USA, 5092–5103. https://doi.org/10.1145/2858036.2858558

[33] Meraj Khan, Larry Xu, Arnab Nandi, and Joseph M. Hellerstein. 2017. Data tweening: incremental visualization of data transforms. *Proceedings of the VLDB Endowment* 10, 6 (Feb. 2017), 661–672. https://doi.org/10.14778/3055330.3055333

[34] Younghoon Kim, Michael Correll, and Jeffrey Heer. 2019. Designing Animated Transitions to Convey Aggregate Operations. *Computer Graphics Forum* 38, 3 (June 2019), 541–551. https://doi.org/10.1111/cgf.13709

[35] Younghoon Kim and Jeffrey Heer. 2018. Assessing Effects of Task and Data Distribution on the Effectiveness of Visual Encodings. *Computer Graphics Forum* 37, 3 (2018), 157–167. https://doi.org/10.1111/cgf.13409

[36] Y. Kim and J. Heer. 2020. Gemini: A Grammar and Recommender System for Animated Transitions in Statistical Graphics. *IEEE Transactions on Visualization and Computer Graphics* -, - (2020), 1–1. https://doi.org/10.1109/TVCG.2020.3030360 Conference Name: IEEE Transactions on Visualization and Computer Graphics.

[37] Daniel Kunin. 2020. Seeing Theory. https://seeing-theory.brown.edu/index.html

[38] Jock Mackinlay. 1986. Automating the design of graphical presentations of relational information. *ACM Transactions on Graphics (TOG)* 5, 2 (April 1986), 110–141. https://doi.org/10.1145/22949.22950

[39] D. Moritz, C. Wang, G. L. Nelson, H. Lin, A. M. Smith, B. Howe, and J. Heer. 2019. Formalizing Visualization Design Knowledge as Constraints: Actionable and Extensible Models in Draco. *IEEE Transactions on Visualization and Computer Graphics* 25, 1 (Jan. 2019), 438–448. https://doi.org/10.1109/TVCG.2018.2865240

[40] Judea Pearl. 2011. *Simpson's Paradox: An Anatomy*. Technical Report. UCLA, Los Angeles, CA. 11 pages.

[41] Judea Pearl. 2014. Understanding Simpson's Paradox. *The American Statistician* 68, 1 (Jan. 2014), 8–13. https://doi.org/10.1080/00031305.2014.876829

[42] pig. 2017. ILLUSTRATE. https://pig.apache.org/docs/r0.17.0/test.html#illustrate

[43] E. D. Ragan, A. Endert, J. Sanyal, and J. Chen. 2016. Characterizing Provenance in Visualization and Data Analysis: An Organizational Framework of Provenance Types and Purposes. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (Jan. 2016), 31–40. https://doi.org/10.1109/TVCG.2015.2467551 Conference Name: IEEE Transactions on Visualization and Computer Graphics.

[44] Evangeline "Gina" Reynolds. 2020. flipbookr: Parses code, creates partial code builds, delivers code movie. Note: R package version 0.1.0.

[45] 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 - CHI '18*. ACM Press, Montreal QC, Canada, 1–12. https://doi.org/10.1145/3173574.3173606

[46] Yedendra Babu Shrinivasan and Jarke J. van Wijk. 2008. Supporting the analytical reasoning process in information visualization. In *Proceedings of the SIGCHI conference on human factors in computing systems*. Association for Computing Machinery, New York, NY, 1237–1246. https://doi.org/10.1145/1357054.1357247

[47] Claudio T. Silva, Juliana Freire, and Steven P. Callahan. 2007. Provenance for visualizations: Reproducibility and beyond. *Computing in Science & Engineering* 9, 5 (2007), 82–89. https://doi.org/10.1109/MCSE.2007.106 Publisher: IEEE Computer Society.

[48] BARBARA Tversky, JULIE BAUER Morrison, and MIREILLE Betrancourt. 2002. Animation: can it facilitate? *International Journal of Human-Computer Studies* 57, 4 (Oct. 2002), 247–262. https://doi.org/10.1006/ijhc.2002.1017

[49] Zezhong Wang, Jacob Ritchie, Jingtao Zhou, Fanny Chevalier, and Benjamin Bach. 2020. Data Comics for Reporting Controlled User Studies in Human-Computer Interaction. *IEEE Transactions on Visualization and Computer Graphics* -, - (2020), 1–1. https://doi.org/10.1109/TVCG.2020.3030433

[50] Hadley Wickham. 2014. Tidy Data. *Journal of Statistical Software* 59, 1 (Sept. 2014), 1–23. https://doi.org/10.18637/jss.v059.i10

[51] H. Wickham, D. Cook, H. Hofmann, and A. Buja. 2010. Graphical inference for infovis. *IEEE Transactions on Visualization and Computer Graphics* 16, 6 (Nov. 2010), 973–979. https://doi.org/10.1109/TVCG.2010.161

[52] Hadley Wickham, Romain François, Lionel Henry, and Kirill Müller. 2020. dplyr: A grammar of data manipulation. https://CRAN.R-project.org/package=dplyr
[53] Leland Wilkinson. 2005. *The Grammar of Graphics*. Springer-Verlag, New York. http://link.springer.com/10.1007/0-387-28695-0
[54] Douglas Zongker and David H. Salesin. 2003. *Creating animation for presentations*. PhD Thesis. Citeseer.