

ToonNote: Improving Communication in Computational Notebooks Using Interactive Data Comics

Daye Kang
dayekang@kaist.ac.kr
Industrial Design, KAIST
Daejeon, Republic of Korea

Tony Ho
tho121.th@gmail.com
Independent Researcher
Vancouver, Canada

Nicolai Marquardt
n.marquardt@ucl.ac.uk
University College London
London, United Kingdom

Bilge Mutlu
bilge@cs.wisc.edu
Computer Sciences, University of
Wisconsin-Madison
Madison, USA

Andrea Bianchi
andrea@kaist.ac.kr
Industrial Design, KAIST
Daejeon, Republic of Korea

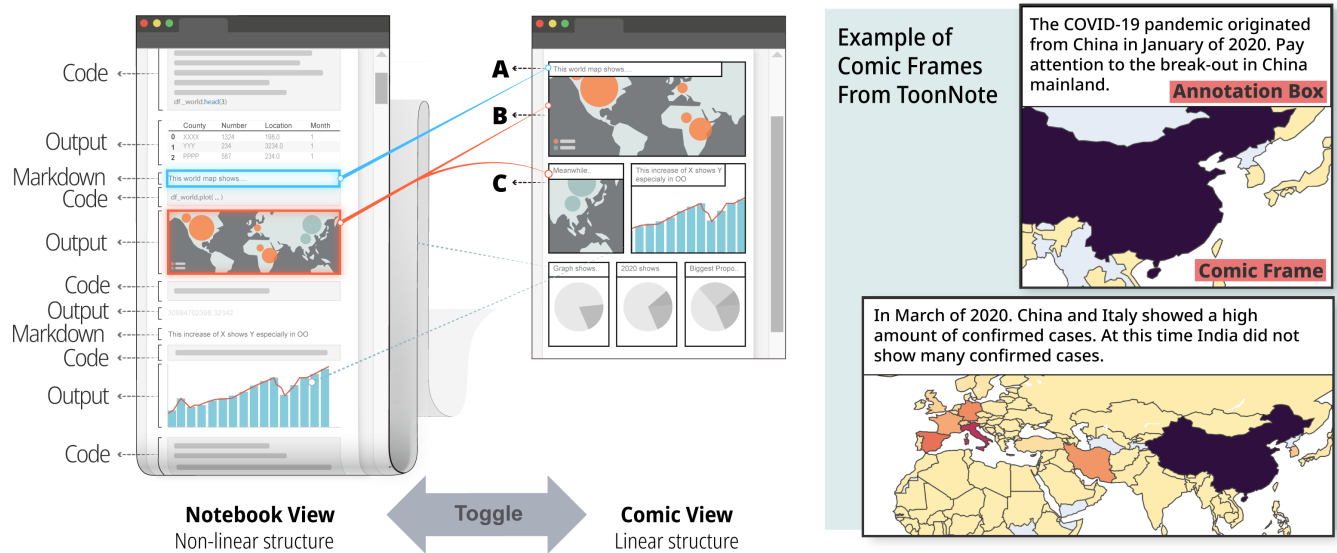


Figure 1: ToonNote is a novel technique for representing computational notebooks in the form of interactive data comics: *Notebook View* uses the traditional computational notebook’s format (using a combination of markdown/code/output), which affords detailed analysis, but makes it difficult to easily comprehend the high-level story behind the data. Therefore, ToonNote provides a high-level, curated narrative of the dataset in *Comic View*. The author can choose different markdown (A) and outputs (B) which can be combined in single or multiple (C) comic frames. When viewing the data in the Comic View, readers can focus on data storytelling, and not be hindered by code, unnecessary outputs, or markdown – and can switch back to the Notebook View when needed.

ABSTRACT

Computational notebooks help data analysts analyze and visualize datasets, and share analysis procedures and outputs. However,

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

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

© 2021 Association for Computing Machinery.

ACM ISBN 978-1-4503-8096-6/21/05...\$15.00

<https://doi.org/10.1145/3411764.3445434>

notebooks typically combine code (e.g., Python scripts), notes, and outputs (e.g., tables, graphs). The combination of disparate materials is known to hinder the comprehension of notebooks, making it difficult for analysts to collaborate with other analysts unfamiliar with the dataset. To mitigate this problem, we introduce *ToonNote*, a JupyterLab extension that enables the conversion of notebooks into “data comics.” ToonNote provides a simplified view of a Jupyter notebook, highlighting the most important results while supporting interactive and free exploration of the dataset. This paper presents the results of a formative study that motivated the system, its implementation, and an evaluation with 12 users, demonstrating the effectiveness of the produced comics. We discuss how our findings

inform the future design of interfaces for computational notebooks and features to support diverse collaborators.

CCS CONCEPTS

• **Human-centered computing** → **Human-computer interaction (HCI)**.

KEYWORDS

computational notebooks, data summarization, data comics, multi-level interface

ACM Reference Format:

Daye Kang, Tony Ho, Nicolai Marquardt, Bilge Mutlu, and Andrea Bianchi. 2021. ToonNote: Improving Communication in Computational Notebooks Using Interactive Data Comics. 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/10.1145/3411764.3445434>

1 INTRODUCTION

Computational notebooks are powerful tools for data analysts, because they allow to document and share analysis procedures and findings [16]. These tools have seen rapid adoption by researchers, data analysts, and journalists in diverse fields [30]. Data analysis is rarely done alone, and sharing of these notebooks is currently the norm in many fields. Despite their popularity, computational notebooks have notable limitations and usability issues [7, 17, 30]. For example, a typical notebook tends to be long and filled with interim notes and findings [14, 30] that may not contribute to the communication of the higher-level ideas and insights of the author. The use of disparate types of content in a free-flowing and unstructured way can make it difficult to navigate and understand the notebook [17]. Furthermore, readers seldom have an interest in reading code when trying to understand the document at a high-level [28].

There have been attempts to improve communication of the computational notebook by using a multi-level interface. An example is the use of “annotated cell folding” [28], which enables authors to curate their notebooks, improving the reading experience. Mathisen et al. [24] proposed “InsideInsight”, which uses dynamic and interactive features to adapt analysis reports to the readers’ desired level of detail. Although this body of work expands the design space of computational notebooks, prior systems have not been able to show an increase in terms of comprehension among computational notebook experts. These experts use their own, well-established strategies to understand unfamiliar notebooks and view the use of these systems as a trade-off [28]. We hypothesize that by further condensing the document, providing more structured guidance, and using a more visual-oriented format, we can improve user understanding of the contents of the document. We chose the “comic” format [4], because previous work has shown data comics to increase engagement [37]. In this paper, we follow McCloud’s definition of comics, which is described by the following: “juxtaposed pictorial and other images in deliberate sequence, intended to convey information and/or to produce an aesthetic response in the viewer” [27]. We chose the *icon*, *panels*, and *gutter* as the core elements for a comic. An icon is “to mean any image used to represent a person, place, thing or idea” [27]. An icon may have different levels of abstraction, from a realistic representation to

a simplified image. Text is also a form of an icon [27]. Frequent use of visualizations and annotations by data analysts make this a natural fit. Panels are the frames that contain icons, and they are used to deliver a deliberate sequence. The space between frames is called the ‘gutter’, and readers use their imagination to fill the gaps between frames to read the sequence of frames as a story [27]. Intending to combine comics with computational notebooks, we created a JupyterLab extension called *ToonNote* that allows data analysts to read notebooks as a data comic and switch back and forth between the original presentation of the notebook content and the data comic. ToonNote presents a comic format by putting visualizations and annotations within a sequence of frames.

This paper explores how we can leverage the data comic format in designing computational notebooks to encourage clearer communication and deeper engagement with complex data analyses. In this paper, we used the word “communication” as it was used by Scott McCloud in “Understanding Comics” [27] to describe the purpose of comics: a possible “form that communication can take” between the author and the reader. This interpretation of comics as an “intended” and “one-way” communication medium is extensively used by McCloud and was implicitly adopted in this paper. Our exploration involves two user studies. First, we conducted a formative study on how data analysts use and share computational notebooks. Particularly, we sought insight into the challenges they faced in reading notebooks. We also asked analysts to assess and comment on content presented in a notebook and a data comic. Based on their feedback, we created the ToonNote JupyterLab extension. The second study aimed to determine whether a ToonNote enabled notebook benefited the reader. Our results demonstrate that, when reading a computational notebook that utilizes a data comic representation, readers are more engaged, recognize the author’s narrative better, and experience a much lower cognitive load. Finally, we report on participants’ suggested scenarios of how to use ToonNote to benefit their respective fields.

2 RELATED WORK

A limiting factor in the sharing of computational notebooks is that they are workspaces containing intermediate work that, in most cases, is not intended for presentation [30]. It becomes difficult to understand the document, which in turn can discourage collaboration. Previous work has leveraged the multi-level interface to address these issues [28], but did not show a significant increase in comprehension. We believe data comics can make strides in this category due to the heavy use of visualizations, intrinsic narrative structure, and dense layout.

2.1 Collaboration In Data Science

Data analysis is rarely done by a single person. Multiple collaborators are usually involved in the same analysis [38], and intermediate results can be shared with different stakeholders. Therefore, an important aspect of collaborative data analysis is to enable communication of insights before the analysis is completed [12, 24, 38]. Data analysis is meaningless if the insights derived from them cannot be effectively communicated to collaborators or stakeholders [9]. Gaining a good understanding of the document is very important for ensuring effective coordination [31]. Even though all the code

and resources are contained within a single document, making it understandable to collaborators remains a challenge [29, 30].

In fact, since computational notebooks are adopted by a wide variety of users, their technical expertise will vary [24, 38]. Data science projects consist of complex tasks that require multiple skills [18, 25]. A team may be composed of expert programmers, some with knowledge of data science methods and others possessing domain knowledge. Sharing work and insights within such a team is challenging, especially when sharing partial progress [24].

When preparing a notebook for sharing, analysts often delete parts to curate the content [14, 30]. Computational notebooks tend to become long, poorly formatted, and difficult to understand [13, 34]. This hinders sharing between collaborators [17]. Not every detail is beneficial to every collaborator [24] and some parts include intermediate work. Yet, any omitted information should be available for review when required [24]. Often, it is necessary to understand how a piece of data is produced or processed. Insights into how the data was processed is an essential part of trusting the data [32]. Previous work mainly addresses this by creating tools that allow the author to curate their work to be more presentable [28] while allowing the reader to explore these hidden components for further detail. This is called the *overview + detail style* of the multi-level interface [8]. Our work follows the same principle but offers a different format, the data comic, as an overview for the curated document.

2.2 Multi-level Interfaces

Multi-level interfaces [8], in computational notebooks, aim to aid navigation by presenting an overview of the content and allowing the reader to explore relevant sections in detail. These interfaces allow authors to hide irrelevant sections, making the notebook more comprehensible. Compressing information in computational notebooks have shown to be beneficial for the reader. Navigating a long notebook where the reader has to scroll frequently is distracting and makes reading the notebook more tedious. Previous works [24, 28] allow the author to group cells and/or visualizations together and controlling their visibility using a hierarchy. The result is a curated, presentable document that is easily shareable and less strenuous to read. However, these versions didn't show improvements in comprehension. Previous work has also found that even with the closest collaborators, they were usually not interested in the code [28]. This implies that data visualization plays an important role in conveying insights. Our approach aims to address this by giving users a visually oriented format, the data comic, shown to improve comprehension.

Approaches such as cell folding compress the notebook length by hiding cells [28]. Other approaches rely on potentially deep hierarchies that require non-trivial navigating [24]. Our approach instead makes use of the horizontal space, allowing for further compression of the notebook as well as high visual density, which has shown to increase memorability and comprehension [35]. Our data comic format supports up to three frames in a single row, potentially resulting in using 3 times less space. We compress this even further by showing annotations in the comic frame itself, what would otherwise be markdown in a typical notebook view.

The result is an overview capable of communicating more insights that are related to each other, all in the same view.

Previous implementations of the multi-level interface noted they would have benefited from structured guidance [24]. Although it is important to let users explore the data to the degree they desire, the author's intended narrative might become less clear when not following the clear directions from the author, making it more difficult for authors to be sure they are conveying their ideas and insights effectively. The benefit of using a comic layout is that there is an intrinsic linear narrative due to the structure of the layout, which guides the reader. Clear sequencing increases the readers' ability to focus and navigate spatial-temporal information [37], is easier to understand, and helps with memorability [11].

2.3 Data Comics

Communicating complex data in an accessible graphical form has been addressed in earlier work [19, 21], and the past few years have witnessed research regarding visual genres in storytelling [6, 22, 33]. Particularly, the unique characteristics of comics [27] have been studied for data storytelling [3, 20].

Data comics [4] is a genre using the well-established visual language of comics to deliver narratives about data. Data Comics have been shown to be very effective at communicating information [5]. The format improves understanding and engagement and is a preferable form of consuming information when compared to slideshows or infographics [37, 39]. When compared to text only [2] or illustrated text [26], comics have been shown to demonstrate increased memorability. When compared to a graph or a long paragraph of continuous text, comics are less strenuous to read [32]. The linear narrative structure serves as a guide by the author that shows increased understanding [11]. Also, the comic frames or panels help divide the information into easily memorable chunks [37]. The data comic format has a high integration of text and visuals, which minimizes spatial distance—and has been shown to increase understanding, decrease cognitive load, and contribute to reader preference [37].

However, previous studies of the use of comics did not target data scientists, and did not yet integrate comics with computational notebooks. We studied whether these benefits are transferred when the comic is integrated in the computational notebook and whether the format helps expert data analysts to comprehend the notebook better. Furthermore, previous explorations did not support interactive data manipulations, as the produced comics were static [3, 5, 20, 37, 39]. In contrast, ToonNote is a novel tool for data scientists, consisting of integration directly in computational notebooks. It provides an alternative output (e.g., comic frames) to visualize the computation results, yet allowing data scientists to both inspect/modify the original code and to interact with the raw data by direct graphical manipulations within a comic frame.

3 STUDY 1: FORMATIVE STUDY

We conducted a formative study to better understand the drawbacks of the standard computational notebook format and the possible benefits of data comics for data analysts. We performed an online

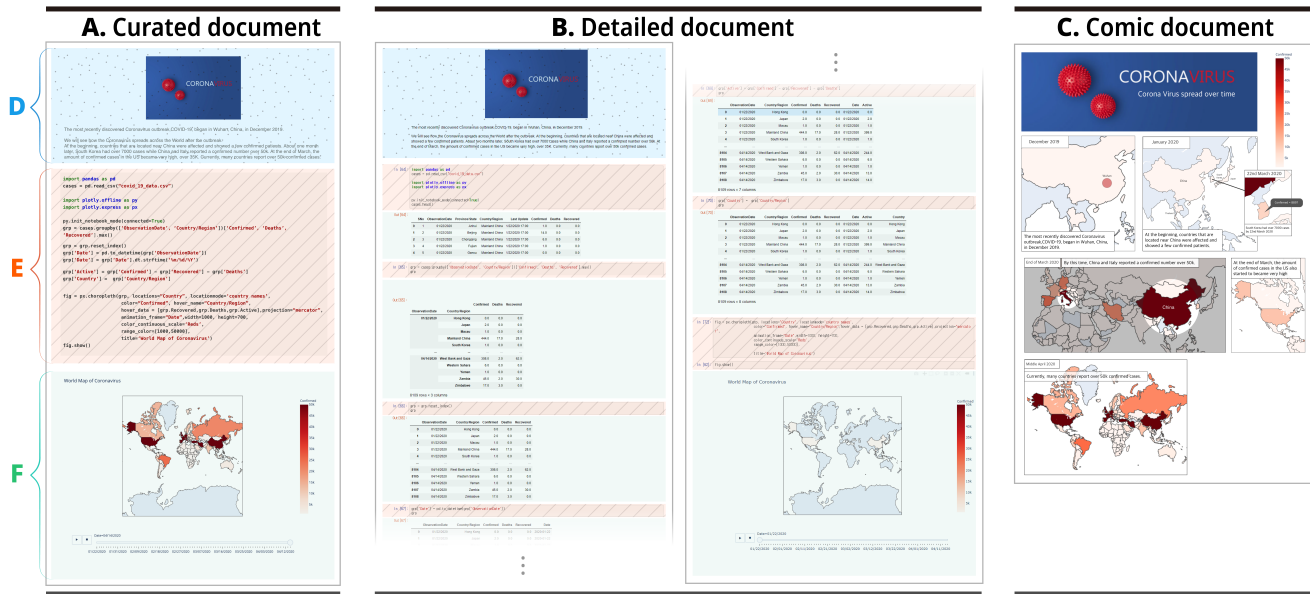


Figure 2: Study Material: In this figure, we show the study materials we used for the formative study. The different types of cells are highlighted as follows, (D) is a markdown cell, (E) is a code cell, and (F) is output that is generated from the code cell. The (A) Curated document and (B) Detailed document are both documents that were made using the typical computational notebook format. The Curated document consisted of one final output (a world map graph) and one code cell without intermediate steps. The Detailed document includes the same markdown and code as the Curated document but also shows the intermediate steps. The (C) Comic document is a comic made stories and images from the other documents.

survey evaluating each format and conducted semi-structured interviews with the participants afterward. The findings from this study directly inform our design for developing a new interface.

3.1 Design Exploration and Survey

When data analysts have to provide high-level explanations, they often try to use less technical jargon when communicating [15]. Even though the reader may be technically savvy, it is not always easy to understand code and analysis without context [30]. Such an example would be the data analyst who works on different data sets from different fields. Another example would be a manager or decision-maker who needs to understand only the insights. We hypothesized that different narrative genres can help deliver a high-level understanding of the document. From the previous visual narrative research, we decided to use the comic medium for the reason that comics are primarily visual and feature some annotations. This matches the nature of the high-level information presented in computational notebooks, which are mainly comprised of text and figures [37, 39].

We designed three documents and asked participants about their understanding and sharing practices. Figure 2 shows the three types of document (Curated document, Detailed document, and Comic document) we prepared for the study. All three notebooks used the same data set about COVID-19, with information about Coronavirus, where it originated from, and how it spread over the world over time. While the Comic document used only images and text to deliver data storytelling, the other two documents follow the

typical computational notebook format (Markdown/Code/Output) using Jupyter Notebook. The code is about importing data sets, data wrangling, and uses the interactive visualization library 'plotly' to draw interactive graphs. The graph shows how the virus spread all over the world from January to August 2020.

3.2 Participants

We recruited twelve participants (5 female, 7 male, mean age: 24) with different backgrounds related to data science: members of faculty (1), professional data scientists (2), college students in both undergraduate (3) and graduate (4) programs with a special interest in data science. The average experience in data analysis was 1.80 years (standard deviation: 1.34 years). Also, eleven out of twelve participants had collaboration experience. All participants were compensated with \$5 USD gift cards.

3.3 Questionnaire

The online survey evaluated the three types of documents. First, we asked participants to read both the Curated document and Detailed document in random order and then solve one high-level question about the dataset to facilitate the reading. Participants rated survey questions about navigation, understanding, and focus on a 7-point Likert scale. For example,

- Q. How easy is it to navigate the document? (very hard to very easy)
- Q. How hard did you have to focus on reading and understand it? (very little to very much)

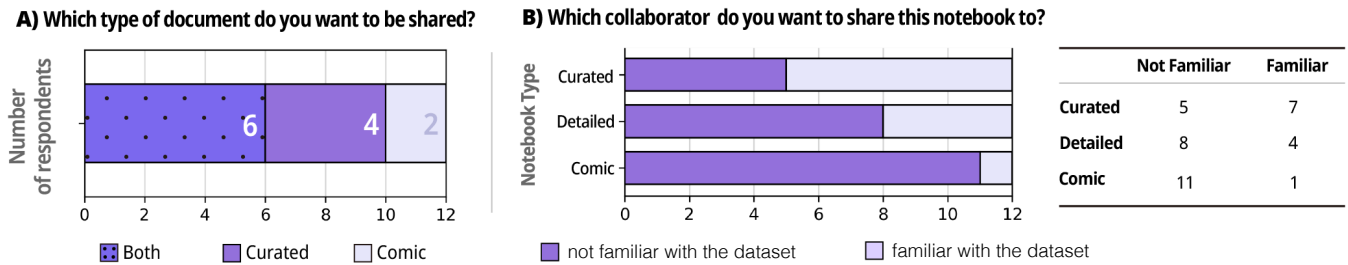


Figure 3: Online Survey results: A) Half of the respondents wanted to get both Curated and Comic documents if they were shared a document with their collaborators. B) Participants responded that comics are better for people who are not familiar with the dataset.

Next, we asked participants to comment on the benefits and drawbacks of each document. Finally, we asked about sharing the document.

- Q. If you share this document with your collaborator, which type of collaborator do you think is appropriate for this format? Choose between ‘a collaborator familiar with the dataset’ and ‘a collaborator not familiar with the dataset’.

We continued to ask questions regarding the Comic document. The questionnaire included:

- Q. Which version is easier to understand? (Among Curated document and Comic document)
- Q. When your collaborator delivers a document, which version do you want to get?

3.4 Procedure

The survey was made using Google survey forms. We first shared the link to the survey with the participants. Participants began the survey by filling in their demographic information and data analysis experience. Next, they navigated through three pages (one per representation). On each page, participants received an explanation of the representation with an HTML link to open the document. After they completed the tasks in each section, participants entered their email address for the gift card. After the survey was over, we conducted a semi-structured interview about the study and their data analysis activities using Zoom.

3.5 Results

3.5.1 Quantitative Results.

Document sharing: Figure 3 shows that half of the participants would like to receive a document capable of showing both a Curated document as well as a data comic document. Figure 3 also shows the ideal audience for each type of documents, according to the participants. Over half of the participants thought the Curated document was more suitable for collaborators who are familiar with the dataset due to its condensed nature. For the Detailed document, 8 out of 12 people reported that they want to share the document with the collaborators who are not familiar with the dataset. Binomial test of Comics (11: not familiar with the dataset, 1: familiar with the dataset) showed statistical significance ($p < 0.001$). Binomial test between Curated and Detailed documents did not show statistical difference (Curated: $p = 0.45$, Detailed: $p = 0.23$). However, Fisher’s

exact test showed a statistical frequency difference from the three documents ($p < 0.05$).

Navigation, Understanding, Focus: Regarding the ease of navigation of the document, participants rated the Curated document ($MD = 6$, $M = 5.58$, $SD = 0.90$) higher than the Detailed document ($MD = 6$, $M = 5.33$, $SD = 1.15$). Regarding the ease of understanding of the document, participants rated the Detailed document ($MD = 6$, $M = 5.33$, $SD = 1.37$) higher than the Curated document ($MD = 5$, $M = 4.92$, $SD = 1.50$). Regarding focus required for understanding, participants rated the Detailed document ($MD = 3.5$, $M = 3.83$, $SD = 1.53$) higher than the Curated document ($MD = 4$, $M = 3.92$, $SD = 1.31$). Wilcoxon test showed that there is no statistical difference between the Curated document and Detailed document in navigation, understanding, and focus (navigation: $p = 0.46$; understanding: $p = 0.57$; focus: $p = 0.92$). Even though there was no statistical difference among Curated and Detailed documents, regarding the question, “Which version is easier to understand,” 10 out of 12 participants answered “comic document”.

3.5.2 Qualitative Feedback.

For the interviews, we transcribed all interview recordings into textual form. Then the first author did a line-by-line Open Coding on the script data. Then the second author used the code on the transcript. Disagreements were discussed then we chose only the codes with consensus between both coders.

Benefits and Drawbacks: Participants mentioned the benefits and drawbacks of each document. For the curated document, participants mentioned the interactive plot as a benefit. However, they also reported that it is difficult to understand the analysis process because of its condensed and limited formatting.

“World map is very useful [...] I can zoom in and use the slider to change the date.” - P8

“It didn’t show many details. For example I’m curious about how the data looks like.” - P1

“Hard to hover over different places on the map especially places with relative less land cover.” - P11

For the Detailed document, participants considered it easy to understand the analysis process, attributed to the detailed explanations.

However, the long document length made it more difficult to understand the intent of the analysis.

“Each step is visualized well and is more understandable.” - P6

“I didn’t know what the author wanted to do with the data at the beginning of the notebook.” - P2

For the Comic document, the participants reported that it was more welcoming and easy to understand, and needed less effort to understand the output of the analysis.

“More welcoming, less daunting, and seemed easier to look at.” - P3

“Comic tells a story and it helps people unfamiliar with the topic better understand it.” - P7

“[...] highlights important data with very little work from the viewer.” - P9

Individual Preferences: Feedback from participants who wanted both the Comic document and the Curated document suggested that each document has advantages that are suited for different situations (P9: “I think that both are needed depending on what it is you are trying to know”). The ones who preferred Curated documents over the Comic document said it was because the Curated document includes more detail and that the Comic document was static and didn’t have interactive features (P12: “It seems like the comic version is static, so for me, that’s a big drawback. It doesn’t seem to allow exploration of the data”). The ones who preferred the Comic document said it was because the Comic document is engaging and easy to understand the contents (P1: “More graphs which will make me more interested in it”).

Limitations of the computational notebook: During the post-survey interview, we asked our participants about their typical data analysis practices and pain points when they use computational notebooks. They mentioned the limited layout as a pain point of the current notebook. Due to the vertical scroll base format, they considered lengthy documents as not practical.

“I think scrolling is something that is a little dizzying.” - P12

“[...] short and simple is always better than long and unfocused.” - P11

Another pain point was having to make versions of their documents because of various collaborators having different backgrounds and technical expertise.

“If I’m presenting to someone who is, who is still like a technical person but not working in the same area, I may still use some of the technical words, not all of them and maybe you know, try to explain a bit of the work I’m doing.” - P11.

3.6 Design Insights

From the formative study, it is shown that the Comic document and the notebook documents (Curated document, Detailed document) all provide value, depending on the scenario. By allowing authors to integrate both formats, we hypothesize that readers will be able to gain an understanding of the document more effectively while still being able to view the details as needed.

A limitation of the Comic document was the loss of interactive features. At times, participants mentioned they preferred the notebook instead due to this aspect. In the formative study, we learned that users require at least the same capabilities as typical computational notebooks, including the ability to interact with generated graphs. Therefore, instead of entirely replacing the current computational notebook, it can be beneficial to integrate an additional layer that provides an alternative graphical representation of the computed output through a comic format. To preserve the capabilities of the computational notebook, we decided to support view switching where readers can use the computation notebook features by switching to the notebook view at any stage. However, to minimize context switching, the comic view should support data exploration by showing the relevant code in the comic view when desired. To further support data exploration, users should be able to explore the graphical data exploration, while at the same time the authors’ intended view of the generated graphs should be preserved. However, computational notebooks are unable to easily preserve the visual states of the graphs and figures when interacted with. The author may interact with the figure by using the panning or zooming function to highlight a specific aspect of the figure. Unfortunately, further interaction by readers will change the visual framing of the figure, and without being able to explicitly revert these changes, the author’s original intention is lost. Therefore, when creating a data comic that enables interaction, it is essential to be able to revert to the author’s intended visual state. Through this feature, the reader would be free to explore the data in the generated graphs and then reset the graph to the author’s intended view and preserve the story of the comic.

4 TOONNOTE SYSTEM

We developed a JupyterLab extension – called *ToonNote* – that allows the development of data comics right in the notebook environment. The extension was built using Javascript, HTML, CSS, and the extension framework provided by JupyterLab. Authors of comics can tag outputs and markdown cells as a ‘comic’, and only those tagged cells will be included as a comic frame in the Comic View, while all other untagged cells are hidden. The author can, at any stage, create or edit the comics output and interactive graphs. In addition to comic tags, users can use additional sizing tags to determine frame size (i.e., frame’s width and height). If the comic frame contains an interactive element, such as a graph, the author can record a sequence of actions delimited by the “start” and “end” recording buttons. The sequence of mouse movements and actions within the comic frame are saved as metadata that can be re-instantiated to recreate the interactive graph. Alternatively, to edit the graph, the author can “reset” the stored metadata. Once the data comics are created, the reader can easily toggle between the two views: Comic View and Notebook View. When a person receives a notebook that was developed using ToonNote, they are able to view a data comic that shows visualizations and graphics contained within the comic frames. These comic frames often have annotations that contribute to a linear narrative, usually describing the contents shown in the frame.

Readers can interact with the Comic View provided by ToonNote in three different ways. First, at any stage, readers can switch

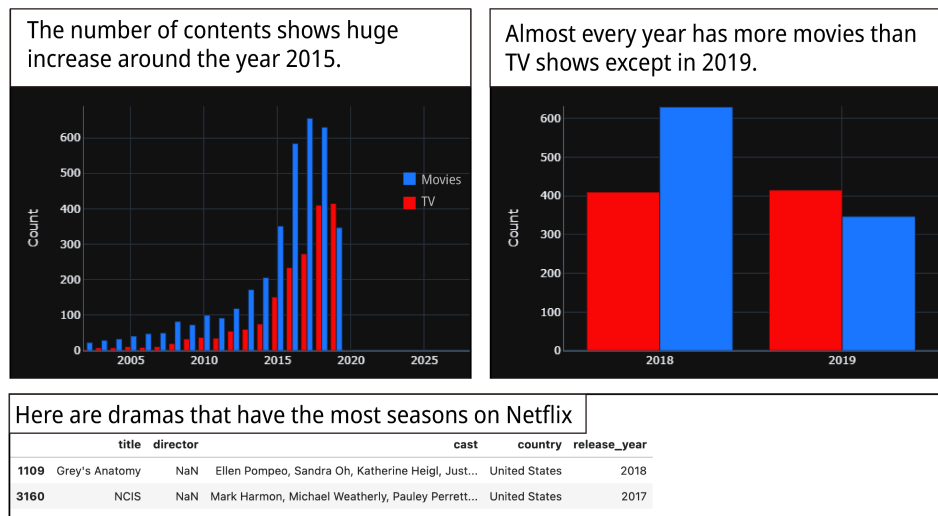


Figure 4: ToonNote Comic View: This figure shows a sample of the Comic View created with the Netflix dataset. The top-left frame shows a typical bar graph, while the top-right frame shows the same graph but zoomed in to highlight a specific portion of the x-axis. The bottom frame shows that the Comic View can display any output, a data table in this case, not just graphs.

back to code using the “toggle” feature and view or edit the code. Figure 1 describes the toggling between Comic view and Notebook View. Second, if the reader wants to inspect the code only for a specific comic frame, a “Peek Code” button in the toolbar allows them to toggle the respective code cell while in Comic View. The “Peek Code” feature enables users to do low-level data exploration through code changes without requiring context switching. Also, by quickly viewing the code, readers can understand the data analysis process that was used to make the graph, which promotes transparent content consumption. Finally, as the extension allows authors to record their manipulations of the generated visualizations (e.g. graphs, figures), readers can execute the respective code cells using the “undo” feature to regenerate the visualizations. The ToonNote system replays the recorded manipulations, recreating the intended framing by the author. The “undo” feature is motivated by the users’ need for a way to graphically manipulate output results as mentioned in the “Design Insights” section. The “undo” feature supports high-level data exploration in interactive graphs without affecting the curated comics output. With the “undo” feature, in practice, any manual change to the comics can be reset. The ToonNote extension is open source and the code is accessible in the Github repository¹

5 STUDY 2: USER STUDY OF TOONNOTE

The purpose of this user study is to find out the usefulness of the comic medium in a computational notebook and whether data analysts see benefits in using such a medium for sharing their work. The questions we sought to answer focused on whether Comic View presented by the ToonNote-enabled notebook can effectively improve comprehension of the document and understanding of the author’s insights. We pose these questions in the form of the following research questions.

¹<https://github.com/tho121/comicconverter>

- **RQ1: Does the ToonNote enabled notebook help readers efficiently understand the dataset at a high-level?** We assume that the format of Comic View, with the focus on visual explanations and placing annotations in the comic frame, helps readers understand the related outcomes more efficiently than reading a standard notebook. This leads to *H1: Data Analysts will grasp the high-level information about the dataset more efficiently when using a ToonNote enabled notebook than a standard notebook.*
- **RQ2: Does the ToonNote enabled notebook help readers to get a general idea about the data with less effort?** We assume that the sequential narrative, visual explanations, and condensed layout of Comic View is easier to understand than reading an article mixed with code. This leads to *H2: Data Analysts will grasp the high-level information about the dataset with less effort when using ToonNote enabled notebooks than using a standard notebook.*
- **RQ3: Does the ToonNote enabled notebook help readers to understand the author’s narrative of the analysis?** We assume that ToonNote enabled notebooks help readers understand the narrative of the analysis due to the author-driven storytelling characteristics of the Comic View. This leads to *H3: Data Analysts understand the intention of the author better using the ToonNote extended notebook than the standard notebook.*
- **RQ4: Does the ToonNote enabled notebook help readers to comprehend the entire document?** We assume that toggling between Comic View and Notebook View can deliver high-level and low-level information in a way that helps the reader’s comprehension of the whole document. This leads to *H4: Data Analysts will comprehend the contents of the ToonNote enabled notebooks more effectively than standard notebooks.*

- **RQ5: Does the ToonNote enabled notebook provide a more enjoyable and engaging reading experience?** We assume that visual explanations in Comic View help readers enjoy studying the analysis. This leads to *H5: Data Analysts will enjoy and be more engaged when reading ToonNote enabled notebooks more than standard notebooks.*

5.1 Study Design and Preparation

To prevent any bias from us in the study material and task, we recruited an external data scientist with 4 years of data analysis experience to make a data comic from an existing notebook using ToonNote. We prepared two sets of notebooks that cover data exploration of different data sets. The notebooks were modified from existing notebooks written in Python from Kaggle [1]. We chose the notebooks with various visual outputs and that use plotly or matplotlib for visualization. To balance the scope of the task within the available time of the study, we focused on exploratory data analysis, which is the basic stage for all data scientists. We provided a basic analysis for participants to begin with. The data scientist we recruited to create the comic with ToonNote was compensated with a \$80 USD gift card.

We asked this participant (the external data scientist recruited to author the data comic) to create two comic documents within four hours. This included 30 minutes of a demonstration on how to use the ToonNote extension, and 10 minutes of practice time to become familiar with the system. Then we asked the participant to read the original document and come up with meaningful insights from the notebook. The data scientist was then asked to make a comic that's able to deliver relevant insights from the original notebook, and to create seven questions regarding these insights in such a way that would promote user exploration. Figure 5 shows the final outputs that were created by the external data scientist using ToonNote and then used for our study.

5.2 Participants

We recruited 12 data scientists (different to the ones from Study 1, 3 Female, 9 Male, mean age: 28) from a public university and a data science special online interest group. The participants met the following requirements: pursuing a degree that majors in data science or working in data analysis, and proficiency in literate programming (about two years). We compensated each participant with a \$20 gift card. They self-reported an average of 3.2 years of data science experience ($SD = 1.1$). They self-reported an average of 2.7 years of using Jupyter notebooks ($SD = 0.9$). They self-rated their coding skills with mean of 4.5 ($SD = 0.9$), using a seven-point Likert scale from Novice (1) to Expert (7).

5.3 Procedure

We asked the participants to join a video conferencing call via Zoom where we then streamed our screen to them while they also shared their screen with us. For the online study, a remote computer was set up to host the JupyterLab sessions with the relevant extensions, libraries, data files, and notebook files. A JupyterLab session hosted by this computer could be accessed from any browser by inputting the IP address in the address bar. This allowed participants to use our controlled instance of JupyterLab.

Tutorial Session: Before the task began, we first introduced the ToonNote extension and the interactive features they can use while reading the notebook. We introduced:

- (1) Toggling: Changing view mode (Notebook View/Comic View)
- (2) Peek Code: Inspecting code in Comic View
- (3) Undo: Resetting the comic frame after interaction with the content.

Participants were then shown a demo of how to use these features. These explanations took about 15 minutes in total.

Tasks: After the tutorial session, the researcher informed the participants of how to access the desired Jupyter Notebook file from their browser. In this within-subjects study, we asked participants to imagine they received the notebook from another data analyst and their task was to understand the notebook and solve seven questions designed to test comprehension. The participants streamed their screen while they read the notebook and answered the questions. There was no time limit. We asked participants not to use the find function in their browser, to encourage reading and understanding of the notebook.

The order of presentation of the four representations (COVID-19 Standard notebook / COVID-19 ToonNote enabled notebook / Netflix Standard notebook / Netflix ToonNote enabled notebook) was randomized across participants. Additionally, for the ToonNote enabled notebook, we asked participants to rate the usefulness of each interactive feature provided by ToonNote and the overall experience of using it while completing the tasks.

The participants then filled out a survey regarding their experience using the extension that asked questions such as "Was it easy to understand the intent of the author," and "If you can make this to whom do you want to share?" The survey also included the NASA Task Load Index questions. After the survey, we conducted semi-structured interviews.

6 RESULTS

6.1 Quantitative Results

Efficiency of high-level information comprehension: We measured the total time it took the participants to solve the seven questions after reading the notebook. The ToonNote enabled notebook ($M = 7.83$, $SD = 3.08$) took less than the standard notebook ($M = 16.31$, $SD = 7.10$). Fig 6A shows the mean time difference between the standard notebook and ToonNote enabled notebook. The Wilcoxon test revealed that there is a significant difference between the two types of notebook ($p < 0.001$). Based on the accuracy of the answers given by participants, we measured accuracy of the seven comprehension questions. Accuracy was computed as the number of correct answers divided by the total number of questions. The ToonNote enabled notebook's accuracy ($M = 0.95$, $SD = 0.093$) is slightly higher than the standard notebook ($M = 0.89$, $SD = 0.19$). Fig 6B shows a mean accuracy difference between the standard notebook and ToonNote enabled notebook. The Wilcoxon test revealed that there is no significant difference between the two types of notebook in terms of accuracy ($p = 0.34$). These results provide support for our hypothesis H1 that data analysts understand high-level information more efficiently in ToonNote enabled notebooks since accuracy was similar but the ToonNote enabled notebook took much less time.

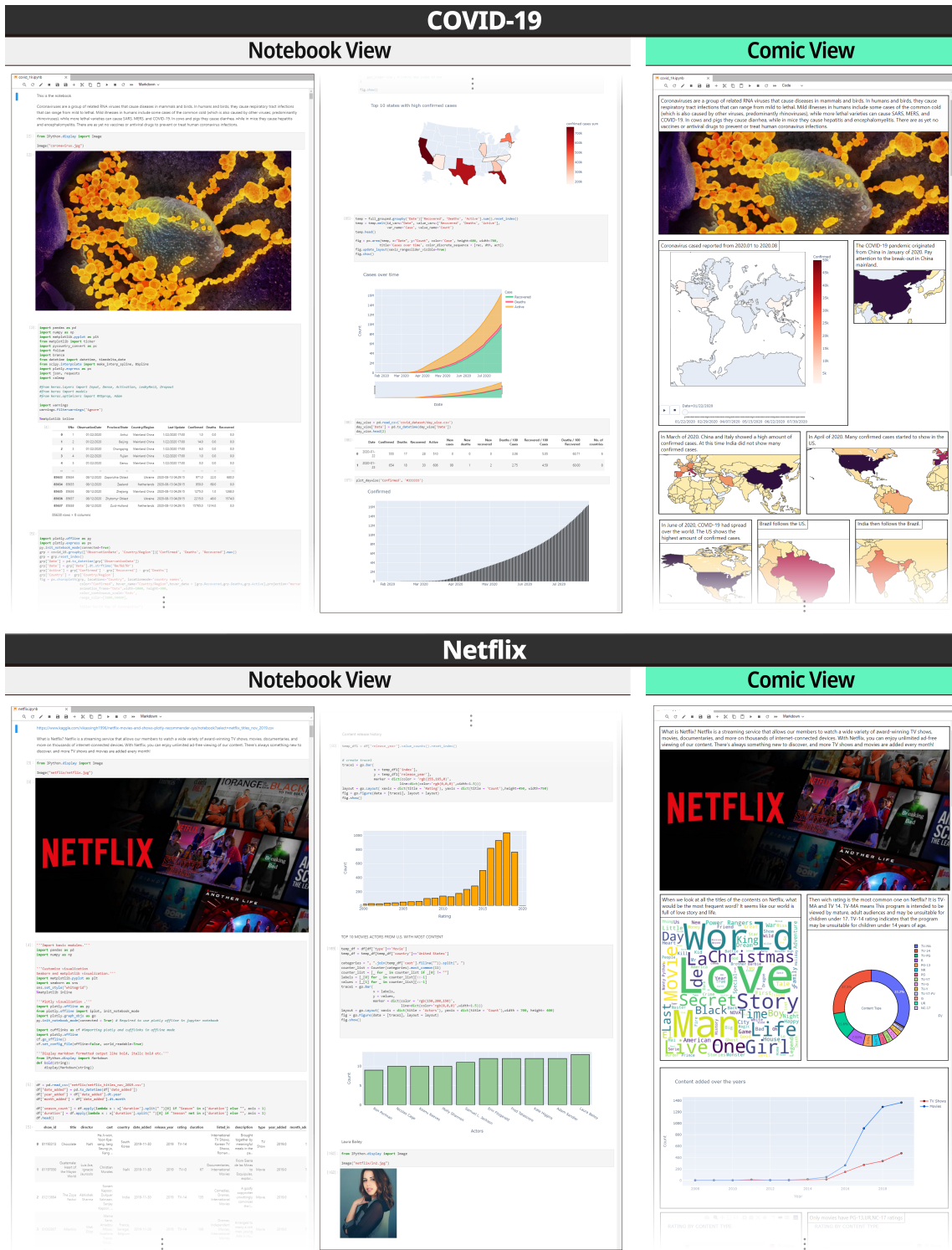


Figure 5: Study material: The first notebook (top) was an exploratory data analysis of a COVID-19 related dataset, containing just over 400 lines of code and 28 data visualizations. The COVID-19 related dataset contained data about the number of COVID-19 cases around the world from January 2020 to August 2020. The second notebook (bottom) was an exploratory data analysis of a Netflix related dataset, containing around 300 lines of code and 24 data visualizations. The Netflix related dataset contained data about the catalog of Netflix, including data such as year added, rating, and content type.

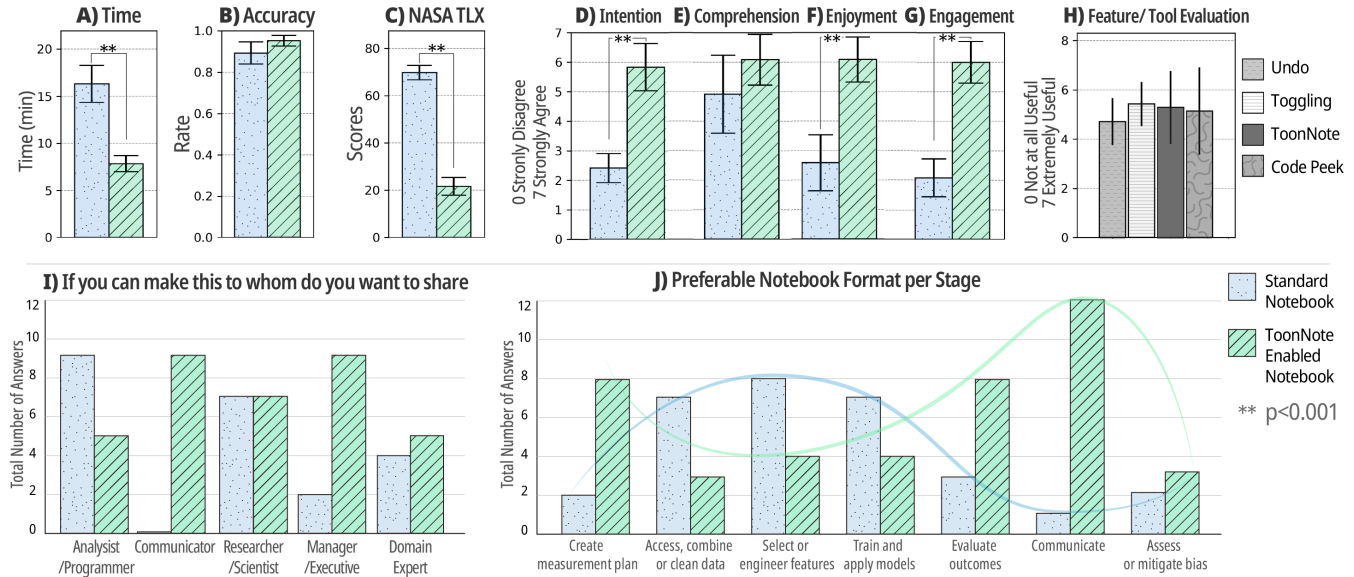


Figure 6: Results from Study 2 (error bars represent standard deviations).

Task load: The ToonNote enabled notebook showed lower task load score ($M = 21.67$, $SD = 13.51$) than the standard notebook ($M = 69.86$, $SD = 11.01$). Fig 6C shows a mean difference of NASA TLX between the standard notebook and ToonNote enabled notebook. The Wilcoxon test revealed that there is a significant difference between the two types of notebooks in terms of task load ($p < 0.001$). These results provide support for our hypothesis H2 that data analysts experience less cognitive load when understanding high-level information in ToonNote enabled notebooks.

Author's intention: Participants reported that the ToonNote enabled notebook ($MD = 6$, $M = 5.83$, $SD = 0.83$) was easier to understand the intention of the author than the standard notebook ($MD = 2$, $M = 2.41$, $SD = 0.51$). Fig 6D shows a mean difference in self-reported ease of understanding the author's intention. The Wilcoxon test revealed that there is a significant difference between the two types of notebook in the understanding of the intention of the notebook author ($p < 0.001$). These results provide support for our hypothesis H3 that data analysts understand the intention of the author more easily in ToonNote enabled notebooks.

Overall comprehension: According to the Participants's answers, the ToonNote enabled notebook ($MD = 6$, $M = 6.08$, $SD = 0.90$) did not show significant improvement in comprehension over the standard notebook ($MD = 5.5$, $M = 4.92$, $SD = 1.38$). Fig 6E shows a mean difference of self-reported comprehension of the document. The Wilcoxon test revealed that there is no significant difference between the two types of notebook in comprehension ($p = 0.06$). These results are inconclusive for our hypothesis H4 that data analysts comprehend the document more effectively in ToonNote enabled notebooks.

Enjoyment and engagement: Participants reported the ToonNote enabled notebook ($MD = 6$, $M = 6.08$, $SD = 0.79$) was more enjoyable to read than the standard notebook ($MD = 2$, $M = 2.58$, $SD = 0.10$). Fig 6F shows a mean difference in self-reported rating

of enjoyment. Participants reported the ToonNote enabled notebook ($MD = 6$, $M = 6.08$, $SD = 0.79$) was more engaging than the standard notebook ($MD = 2$, $M = 2.08$, $SD = 0.67$). Fig 6G shows a mean difference in self-reported rating of engagement. The Wilcoxon test revealed that there is a significant difference between two types of the notebook in terms of enjoyment and engagement (enjoyment: $p < 0.001$, engagement: $p < 0.001$). These results provide support for our hypothesis H5 that ToonNote enabled notebooks increase enjoyment and engagement in reading.

Usefulness of interactive features and the extension: Fig 6H shows self-reported usefulness of ToonNote's features and the extension itself in helping comprehension. The participants found all the features of the system beneficial (view toggling: $MD = 6$, $M = 5.58$, $SD = 0.90$; peek code: $MD = 6$, $M = 5.41$, $SD = 1.78$, undo: $MD = 5$, $M = 4.7$, $SD = 0.95$; average of three features: $M = 5.33$, $SD = 1.26$). For the general usefulness of ToonNote in helping readers understand the insights of the author and the data analysis process, participants found it beneficial overall ($MD = 5.5$, $M = 5.3$, $SD = 1.48$).

Collaborators and sharing: Fig 6I shows sharing by collaborator type. Participants reported if they were to make any comic notebook by themselves, they would want to share it with communicators (9 out of 12 participants), managers/executives (9 out of 12 participants), and researchers/scientists (7 out of 12 participants). For the standard notebook, analysts/programmers (9 out of 12 participants) and researchers/scientists (7 out of 12 participants) got the most votes. Fig 6J shows the preferable notebook format per stage. Participants reported the ToonNote enabled notebook is beneficial when used at the beginning (Create measurement plan: 8 out of 12 participants) and the end stages (Evaluate outcomes: 8 out of 12 participants; Communicate with clients and/or stakeholders and/or others: 12 out of 12 participants) of the workflow process. On the other hand, the standard notebook was preferred to be used at

the engineering oriented stages (Access, combine and curate data: 7 out of 12 participants; select or engineer features: 8 out of 12 participants; train and apply models: 7 out of 12 participants).

6.2 Qualitative Results

We again transcribed all interview audio recordings into textual form. The first author did a line-by-line Open Coding on the script data, before the second author used the code on the transcript. Any disagreements were discussed then we chose ones that got consensus from both coders. Participants mentioned benefits and drawbacks of the extension and suggested various scenarios where the comic format can be used, outside of sharing between expert analysts.

Multi-level interface and interaction: The ToonNote uses the established overview + detail style of multi-level interface. The Comic View provides an overview, while Notebook View and the 'Peek Code' action allows readers to investigate the details. Readers can interact with the figures in Comic View as well to explore the data. Readers made positive remarks about these features as they proved to be convenient for navigation and understanding of the content. In general, participants were satisfied with the three features that support data exploration. Furthermore, the interviews suggest that we reached our goal of providing a curated story with support for data exploration. However, for some participants, those interactions were considered as possible distractions.

"It is easy to understand the contents as a whole." - P3
 "[...]quickly toggling the document and navigate to the desired part using the shortcut was useful" - P1
 "[...]static figure can emphasize key messages[...]The advantage of interactive contents is that you can directly modify the code and explore. By having both, it can fulfill both needs."- P4
 "Interactive features can hinder you from focusing on the core message of the content." - P1

Comic View: The participants commented that the Comic View condensed the document well, resulting in less scrolling. They also mentioned that annotations within the frames provided clear explanations of the graph, and together with the sequential layout, provided a linear narrative that was engaging for the reader.

"Aesthetically pleasing, story is more engaging than the previous code folding features."- P10
 "Length of text in annotation box of comic frame is more concise than the length of text normally used in PowerPoint. I think it has the benefit of faster understanding."- P11

Collaboration: While some participants believed the features of ToonNote are universally beneficial (P1: "It seems beneficial for both practitioner and non-practitioner because its ease of understanding is good for everyone"), other participants offered more specific scenarios where the features of ToonNote would be beneficial. Since Comic View can be toggled from the Jupyter notebook, there is no need to leave the JupyterLab environment. This allows for faster iterations, as well as being able to verify the data used to make the data comic. This scenario highlights the benefits that can be gained when working with another data analyst.

"I don't read all things written in notebooks anyway. It's better to have a summary in comic format. Especially analysts check out the graph first before they jump into the code and investigate. Usually, they check code when they think there's something weird in the graph." - P2

Another scenario where participants suggested ToonNote would be useful is the exploration of data with a supervisor or manager. Commonly, a supervisor or manager might have less specific technical knowledge but instead has effective communication skills. In these situations, Comic View can be used to organize and convey the authors' insights more effectively. Most participants said ToonNote could be used to facilitate discussion with supervisors, especially in the early stages of the work process to decide the analysis strategy and also at the end to discuss the output.

"It can be used with decision-makers because they like clear communications but also sometimes they want to see the other side of the output." - P8
 "[...] in case of managers who are adjacent to customers and do sales with data, they can use this type of note-book to communicate with customers." - P3

6.3 Highlights of the Results

In this subsection, we highlight the main findings. The main findings of our study are as follows.

- Participants finished the task faster with less cognitive loads in the ToonNote enabled notebook.
- There was no significant difference in accuracy in both formats.
- Participants reported that the ToonNote enabled notebook is more enjoyable and engaging.
- Participants found features that support data exploration generally useful.
- From the qualitative interview, participants mentioned potential benefits for non-technical collaborators.

7 DISCUSSION

From the results, we can draw several insights. Although both formats produced answers to the sets of questions with similar accuracy, overall the questions when using the data comic format were completed faster. The cognitive load from analyzing the data comic was significantly lower when compared to the Notebook View, while engagement was much higher. Our qualitative results suggest that the comic format in a computational notebook can be used for more than just collaborating with other data analysts who are not familiar with the dataset. Several participants suggested it could be well suited for diverse types of collaborators from various backgrounds.

Our attempt to significantly increase comprehension amongst expert data analysts showed little evidence that there was any improvement, if any. Accuracy comparisons between the ToonNote enabled notebook and the standard notebook showed little difference. This suggests that participants were able to understand the contents of the document very well from both formats. Participants self-reported their comprehension using both formats and our statistical analysis of the result was inconclusive. When comparing the

mean difference in self-reported comprehension, the mean difference of 1.2 slightly favors the ToonNote enabled notebook, but the p-value of 0.057 suggests this is not significant. This result seems to support the findings of previous work [28] that expert participants are very familiar and effective at reading standard notebooks. Despite leveraging the comic format to create a more condensed, visually oriented, author guided document compared to previous works, the combination of these benefits did not overcome the current methods employed by expert data analyst for comprehension. Comprehension would not be a core reason to use a comic format in this case.

Previous studies focused on comprehension of the whole document, particularly the code, but since we are leveraging a narrative genre, it is important to measure the degree of success in understanding the author's narrative. Participants reported to understand the author's intention at a higher degree when using the ToonNote enabled notebooks. This shows readers were more confident in their understanding of the author's narrative. This confidence is justified since the accuracy is very high when reading the ToonNote enabled notebook. If the goal was to communicate the author's ideas and insights, then the data comic format has shown to be very successful. This shows that authors can leverage this format to present their notebooks to collaborators and most likely a more diverse audience.

The different scenarios suggested by the participants of the user study show how versatile the integration of a comic layout could be for sharing. Participants believed the format would help communicate their analysis to their collaborators and other stakeholders, who often benefit from a format that caters to a non-technical background. This can include people with non-technical business background, non-technical domain experts, and even the general public. Even though we aimed to understand one-way communication between authors and readers, communication can be in various forms (e.g., one to many, bidirectional). Furthermore, we used COVID and Netflix data sets in our study and the notebook was about an exploratory data analysis. However, more complex data sets or different data analyses (e.g., predictive modeling) might have resulted in more distinctive differences in comprehension between the formats.

8 LIMITATIONS

A limitation of our study is the quality of the notebooks developed for our user study. Although we aimed to minimize bias by using notebooks created by data analysts who were not researchers on this paper, we believe they could have produced higher quality notebooks if they were more familiar with using ToonNote. Also, the authors only had a few hours to understand and produce insights about the datasets they received. The questions they developed could have been more challenging and complex, resulting in more obvious differences between the formats. Furthermore, the quality of the notebooks created for the user study could have been further increased if the ToonNote extension would include additional features. For instance, a highlighting feature could be used to draw attention to specific points within the comic frame. This could be used to add more detail in each frame, potentially reducing the number of comic frames required to convey an idea and guiding

reader attention even more. Although we've concluded that gains in comprehension for the expert data analyst group would be minimal, these improvements would benefit future work studying non-expert collaborators. Furthermore, the sharing aspect was only evaluated from the "data analyst" perspective. Our participants mentioned the benefits of sharing data comics to audiences such as executives or domain experts. However, this is different from receiving feedback from those audience members themselves. Finally, we used self reported methods to evaluate usefulness of the features, yet standardized testing could have been used instead.

9 FUTURE WORK

Since we have shown the benefits of having both Comic View and Notebook View available to readers, the next step would be to develop ToonNote further to enable authors to easily and effectively convert their existing notebooks. Although this study focused on the readers and the usability of the format, we recognize the need to minimize the time and effort required to create such a format. The more convenient the authoring process, the more likely data comics could be used as a presentation format [39] as well as see higher rates of adoption by collaborative teams [23]. There are a few inherent benefits that ToonNote provides which are not found or lacking in other formats, such as direct integration with the notebook, or a linear, visually oriented narrative. For future work, it would be interesting to investigate how to automate the authoring process. Currently, the author manually tags each cell to mark which cells to convert into a comic frame. However, an automated process can at least provide a better starting point to lessen the burden on the author in creating an effective data comic.

One of the most frequent comments we received was the desire to share the converted notebooks as a data comic with the non-technical audience. Some of the participants specifically said they wanted to export the data comic as a PDF file. We agree that the ability to export would expand the potential audience and the data comic format would be a good fit for sharing it in this particular format. This would allow authors to share the ToonNote notebook with a wider audience of people who would not require the Notebook View or interactive features, and where the data comic by itself – if constructed carefully – would suffice. Other methods for increasing notebook accessibility could include using JupyterHub to share the notebook on the web with anyone with a link. This method would enable readers to access and interact with the data comics without any technical requirements other than a web browser. Providing various export options can facilitate collaboration with novice and non-technical collaborators, but a future user study should be conducted to determine if this would actually be beneficial to their workflow.

Another study that would be meaningful to conduct would be to compare *interactive data comics* to non-interactive data comics, slides and infographics. Although ToonNote allows interactivity, enhancing the interactive experience was not part of this study. Previous work suggests that readers often do not realize which parts are interactive, preventing readers from making full use of the features of the medium. Since data comics have shown better engagement and comprehension than currently popular formats, it

would be interesting to study if meaningfully adding interaction would continue this trend.

Finally, there are several interesting ideas for future work. Diversifying levels of the multi-level interface into a curated notebook, a detailed notebook and a comic notebook to address the problem of *messy notebooks* would be an interesting topic for future work. This paper, as also previous studies that aimed to improve computational notebooks [10, 28, 36], conducted studies on data scientists as participants. Our participants mentioned that data comics integrated in the Jupyter Notebook can benefit non-technical collaborators, suggesting that meaningful future work should include conducting a user study or developing systems for non-technical collaborators (e.g., executives, domain experts, general public). Furthermore, in this paper we target one-way communication between the author and the reader, using the icon, panels and gutter as the core elements of a comic. However, depending on the communication type (one to many, or two way) the interface can adopt different elements (e.g., characters, speech bubbles). If there are multiple readers and they all require different versions of the comic, the comic can have author-specific speech bubbles or characters.

10 CONCLUSION

In this paper we explored how computational notebooks might be designed to encourage clearer communication by integrating a comic format. We developed ToonNote, an extension for Jupyter-Lab notebooks that enables readers to toggle between the typical notebook view and a comic layout. We conducted a formative study and a user study, where a total of 24 data analysts participated, and we found that data comics in computational notebooks supports not only the original analysis, but communicates it more easily and in a more engaging way. Comprehension does not show a significant increase when using the data comic format between expert collaborators when compared to the standard notebook. However, readers are able to identify the author's ideas and insights at a much higher level of confidence, than when reading the computational notebook alone. The data analysts who participated in our user study believe the integrated comic layout could be beneficial for sharing between collaborators and stakeholders, especially if they are a non-technical person.

ACKNOWLEDGMENTS

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2020R1A2C1012233).

REFERENCES

- [1] 2020. Kaggle. <https://www.kaggle.com/>.
- [2] Paul A Aleixo and Krystina Sumner. 2017. Memory for biopsychology material presented in comic book format. *Journal of Graphic Novels and Comics* 8, 1 (2017), 79–88.
- [3] Benjamin Bach, Natalie Kerracher, Kyle Wm. Hall, Sheelagh Carpendale, Jessie Kennedy, and Nathalie Henry Riche. 2016. Telling Stories about Dynamic Networks with Graph Comics. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) (CHI '16). Association for Computing Machinery, New York, NY, USA, 3670–3682. <https://doi.org/10.1145/2858036.2858387>
- [4] Benjamin Bach, Nathalie Henry Riche, Sheelagh Carpendale, and Hanspeter Pfister. 2017. The Emerging Genre of Data Comics. *IEEE Computer Graphics and Applications* 37, 3 (May 2017), 6–13. <https://doi.org/10.1109/MCG.2017.33> Conference Name: IEEE Computer Graphics and Applications.
- [5] Benjamin Bach, Zezhong Wang, Matteo Farinella, Dave Murray-Rust, and Nathalie Henry Riche. 2018. *Design Patterns for Data Comics*. Association for Computing Machinery, New York, NY, USA, 1–12. <https://doi.org/10.1145/3173574.3173612>
- [6] M. Brehmer, B. Lee, B. Bach, N. H. Riche, and T. Munzner. 2017. Timelines Revisited: A Design Space and Considerations for Expressive Storytelling. *IEEE Transactions on Visualization and Computer Graphics* 23, 9 (2017), 2151–2164.
- [7] Souti Chattopadhyay, Ishita Prasad, Austin Z. Henley, Anita Sarma, and Titus Barik. 2020. What's Wrong with Computational Notebooks? Pain Points, Needs, and Design Opportunities. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '20). Association for Computing Machinery, New York, NY, USA, 1–12. <https://doi.org/10.1145/3313831.3376729>
- [8] Andy Cockburn, Amy Karlson, and Benjamin B. Bederson. 2009. A Review of Overview+detail, Zooming, and Focus+context Interfaces. *ACM Comput. Surv.* 41, 1, Article 2 (Jan. 2009), 31 pages. <https://doi.org/10.1145/1456650.1456652>
- [9] Kristin A Cook and James J Thomas. 2005. Illuminating the Path: The Research and Development Agenda for Visual Analytics. (5 2005).
- [10] Ian Drosos, Titus Barik, Philip J. Guo, Robert DeLine, and Sumit Gulwani. 2020. Wrex: A Unified Programming-by-Example Interaction for Synthesizing Readable Code for Data Scientists. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '20). Association for Computing Machinery, New York, NY, USA, 1–12. <https://doi.org/10.1145/3313831.3376442>
- [11] Ana Figueiras. 2014. Narrative visualization: A case study of how to incorporate narrative elements in existing visualizations. In *2014 18th International Conference on Information Visualisation*. IEEE, 46–52.
- [12] Nahum Gershon and Ward Page. 2001. What Storytelling Can Do for Information Visualization. *Commun. ACM* 44, 8 (Aug. 2001), 31–37. <https://doi.org/10.1145/381641.381653>
- [13] Philip Jia Guo. 2012. *Software tools to facilitate research programming*. Ph.D. Dissertation. Stanford University Stanford, CA.
- [14] Andrew Head, Fred Hohman, Titus Barik, Steven M. Drucker, and Robert DeLine. 2019. Managing Messes in Computational Notebooks. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland UK) (CHI '19). Association for Computing Machinery, New York, NY, USA, 1–12. <https://doi.org/10.1145/3290605.3300500>
- [15] John T. Van Stan II Jan Friesen and Skander Elleuche. 2018. Communicating Science through Comics: A Method. *Publications* 6 (2018), 2.3.
- [16] Jeffery M. Perkel. [n.d.]. Why Jupyter is data scientists' computational notebook of choice. <https://www.nature.com/articles/d41586-018-07196-1>. Accessed: 2020-09-03.
- [17] Mary Beth Kery, Marissa Radensky, Mahima Arya, Bonnie E. John, and Brad A. Myers. 2018. The Story in the Notebook: Exploratory Data Science Using a Literate Programming Tool. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) (CHI '18). Association for Computing Machinery, New York, NY, USA, 1–11. <https://doi.org/10.1145/3173574.3173748>
- [18] Miryung Kim, Thomas Zimmermann, Robert DeLine, and Andrew Begel. 2016. The emerging role of data scientists on software development teams. In *2016 IEEE/ACM 38th International Conference on Software Engineering (ICSE)*. IEEE, 96–107.
- [19] N. W. Kim, B. Bach, H. Im, S. Schriber, M. Gross, and H. Pfister. 2018. Visualizing Nonlinear Narratives with Story Curves. *IEEE Transactions on Visualization and Computer Graphics* 24, 1 (2018), 595–604.
- [20] Nam Wook Kim, Nathalie Henry Riche, Benjamin Bach, Guanpeng Xu, Matthew Brehmer, Ken Hinckley, Michel Pahud, Haijun Xia, Michael J. McGuffin, and Hanspeter Pfister. 2019. DataToon: Drawing Dynamic Network Comics With Pen + Touch Interaction. In *CHI*. ACM, 105.
- [21] N. W. Kim, E. Schweickart, Z. Liu, M. Dontcheva, W. Li, J. Popovic, and H. Pfister. 2017. Data-Driven Guides: Supporting Expressive Design for Information Graphics. *IEEE Transactions on Visualization and Computer Graphics* 23, 1 (2017), 491–500.
- [22] B. Lee, N. H. Riche, P. Isenberg, and S. Carpendale. 2015. More Than Telling a Story: Transforming Data into Visually Shared Stories. *IEEE Computer Graphics and Applications* 35, 5 (2015), 84–90.
- [23] Yaoli Mao, Dakuo Wang, Michael Muller, Kush R Varshney, Ioana Baldini, Casey Dugan, and Aleksandra Mojsilović. 2019. How Data Scientists Work Together With Domain Experts in Scientific Collaborations: To Find The Right Answer Or To Ask The Right Question? *Proceedings of the ACM on Human-Computer Interaction* 3, GROUP (2019), 1–23.
- [24] Andreas Mathisen, Tom Horak, Clemens Nylandstedt Klokmose, Kaj Grønbaek, and Niklas Elmquist. 2019. InsideInsights: Integrating Data-Driven Reporting in Collaborative Visual Analytics. In *Computer Graphics Forum*, Vol. 38. Wiley Online Library, 649–661.
- [25] Kate Matsudaira. 2015. The science of managing data science. *Queue* 13, 4 (2015), 30–34.

- [26] Richard E Mayer and Joan K Gallini. 1990. When is an illustration worth ten thousand words? *Journal of educational psychology* 82, 4 (1990), 715.
- [27] Scott McCloud. 1993. *Understanding comics: The invisible art*. Northampton.
- [28] Adam Rule, Ian Drosos, Aurélien Tabard, and James D Hollan. 2018. Aiding collaborative reuse of computational notebooks with annotated cell folding. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–12.
- [29] Adam Rule, Amanda Birmingham Id, Cristal Zuniga Id, Ilkay Altintas Id, Cheng Huang Id, Rob Knight, Niema Moshiri Id, and Mai H Nguyen. 2019. Ten simple rules for writing and sharing computational analyses in Jupyter Notebooks. (2019), 1–8.
- [30] 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* (Montreal QC, Canada) (*CHI '18*). Association for Computing Machinery, New York, NY, USA, 1–12. <https://doi.org/10.1145/3173574.3173606>
- [31] Ali Sarvghad and Melanie Tory. 2015. Exploiting analysis history to support collaborative data analysis. In *Proceedings of the 41st Graphics Interface Conference*. 123–130.
- [32] Andreas Schreiber and Regina Struminski. 2017. Visualizing Provenance using Comics. In *9th USENIX Workshop on the Theory and Practice of Provenance (TaPP 2017)*. USENIX Association, Seattle, WA. <https://www.usenix.org/conference/tapp17/workshop-program/presentation/schreiber>
- [33] E. Segel and J. Heer. 2010. Narrative Visualization: Telling Stories with Data. *IEEE Transactions on Visualization and Computer Graphics* 16, 6 (2010), 1139–1148.
- [34] Aurélien Tabard, Wendy E. Mackay, and Evelyn Eastmond. 2008. From Individual to Collaborative: The Evolution of Prism, a Hybrid Laboratory Notebook. In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work* (San Diego, CA, USA) (*CSCW '08*). Association for Computing Machinery, New York, NY, USA, 569–578. <https://doi.org/10.1145/1460563.1460653>
- [35] Chao Tong, Richard Roberts, Rita Borgo, Sean Walton, Robert S Laramée, Kodzo Wegba, Aidong Lu, Yun Wang, Huamin Qu, Qiong Luo, et al. 2018. Storytelling and visualization: An extended survey. *Information* 9, 3 (2018), 65.
- [36] April Yi Wang, Zihan Wu, Christopher Brooks, and Steve Oney. 2020. Callisto: Capturing the “Why” by Connecting Conversations with Computational Narratives. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (*CHI '20*). Association for Computing Machinery, New York, NY, USA, 1–13. <https://doi.org/10.1145/3313831.3376740>
- [37] Zehong Wang, Shunming Wang, Matteo Farinella, Dave Murray-Rust, Nathalie Henry Riche, and Benjamin Bach. 2019. Comparing Effectiveness and Engagement of Data Comics and Infographics. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland UK) (*CHI '19*). Association for Computing Machinery, New York, NY, USA, 1–12. <https://doi.org/10.1145/3290605.3300483>
- [38] Amy X. Zhang, Michael Muller, and Dakuo Wang. 2020. How do Data Science Workers Collaborate? Roles, Workflows, and Tools. 1, 1 (2020), 1–23. arXiv:2001.06684 <http://arxiv.org/abs/2001.06684>
- [39] Zhenpeng Zhao, Rachael Marr, and Niklas Elmqvist. 2015. Data comics: Sequential art for data-driven storytelling. *tech. report* (2015).