Generating data comics about personal activities

Rui Wu

Master of Science
School of Informatics
University of Edinburgh
2022

Abstract

Data comics provide benefits of engagement and strong guidance through reading compared to current widely distributed data visualization tools. However, the current data comics are mainly manual work which take time, effort and skills drawn from many disciplines. As a result, there is no data comic template that enabling automatically generating about personal activities data. This research respond to the gap by an application case of CleverCogs and its data. It was found a potential family communication scenario with updated personal activities data. Hence, a generative data comic can play a role in this. This research intends to design a template that enabling generating data comics about personal activities. The design process involves exploratory data analysis and many other design considerations drawn from data comics, techniques from storytelling and interactivity. The result reflects on the design considerations of data comics, as well as limitations regarding to customizability and generalizability of data comics template.

Acknowledgements

Contents

Chapter 1

Introduction

1.1 Background

Massive personal activities data is generated everyday, as a result of the proliferation of digital electronics and trend of ubiquitous computing. These data enables tracking of personal life in various life aspects. Corporations are commonly using online behavior data to analyze user's preference and interest for enhancing business value. App usage data are frequently used as reference to improve user experience. Regarding to personal informatics, existing visualization tools include monitoring health, sleep, home automation, physical activities and online activities. Moreover, delivering feedback through visualizing personal activities has been commonly researched and applied in behavior change technologies(BCT) (Schneider, Schauer, Stachl, Butz, 2017) (Hermsen, Frost, Renes, Kerkhof, 2016). These personal informatics tools are argued for values in self-monitoring, self-reflection and self-optimization (Ruckenstein, 2014)(Murnane, E. L., Walker, Tench, Voida, Snyder, 2018).

If we overview the current visualization tools for supporting different roles in engaging with personal activities data, there are three main streams identified. First, statistical data visualization libraries such as matplotlib (Matplotlib, 2012), seaborn (seaborn, 2012) for professionals, data analysts and researchers to draw insights from research domain data. Second, dashboards that widely adopted in business through supporting of data visualization tools such as Tableau (Tableau, 2019), Google Analytics (Using google charts, n.d.), which allow managers to monitor the usage and users activities. Third, personal informatics tools that adopts various visualization pattern from simple counts and graphs, to incorporation of metaphoric, narrative, physicalization techniques (Schneider et al., 2017), supporting improvement of self-knowledge

and self-reflection (Pousman, Stasko, Mateas, 2007).

However, the above visualization tools in three main streams have their limitations in target audience and application scenarios. For data visualization tools, transforming dataset into graphs, charts, maps usually require data skills, catering for professionals. Casual audience can barely benefit from those tools without training. For digital dashboards, they enable key messages and broad information to be covered in a holistic view, and suitable to deliver snapshots in a timely manner. Dashboard templates are caters for monitoring changing responding to updated data stream; nevertheless, a lack of engaging and compelling attributes may exist. For personal informatics tools, on the one hand, simple graphs are commonly seen in practical (e.g. sleep and health tracking with apple watch). On the other hand, the incorporation of artistic, stylistic display or physicalization though provides freedom and flexibility, but usually require mapping knowledge between visual elements and implicit data, hence may not straight forward in conveying messages. Lastly, a bit of notes sum up here: data visualization libraries may not friendly for casual audience. The dashboards may not be engaging and compelling enough. Personal informatics tools commonly seen and founded either risk of being restriction in numbers and graphs, or being not straight forward to encapsulate a range of messages.

One novel genre of visualization which provides benefits compared to above three main streams, is the data comic. It incorporates features of comic, sequential arts and narrative techniques into the visualization of data(Bach, Riche, Carpendale, Pfister, 2017). It is accessible for wider audience even including the younger children and elderly compared to data visualization tools. It engages audience by the characteristic of storytelling compared to traditional dashboards. It breaks down complexities through panels and make piece of messages explicit stated. (Bach et al., 2017). However, the current data comic cases are mainly manual work, partially because the outcome is highly based on communication context, the narrative dynamic and creating styles of authors. Besides, the manual work usually requires plenty of time and effort, data skills and art skills, result in comics which are not capable of updating. As a result, there is no data comic template that enabling automatically data comic generating about personal activities data.

This research identifies a gap where no such templates supporting generating data comics about personal activities. Hence, to motivate a space of creating data comic through generating from templates, based on certain type of data, this research explores, experiments with design, and informs the possibility with a use case at hand.

Using CleverCogs personal activities as a use case, I try to design a template which serves as basis for generating data comics about personal life.

1.2   Design Motivation

Motivation data source

The use case in my research is motivated by online activities data from a digital entertainment system, CleverCogs, developed by Blackwood(Scott, n.d.). It is an digital entertainment system delivered through dedicated tablet, aiming to promote digital participation and independence for elderly and people with disabilities. The online activities data here is defined as a series records of activities types in terms of entertainment, health services, social connection, home automation, wellbeing, etc. Each record is associated with a user ID and timestamp, indicating what link visited by a particular user at a specific time. For example, I would know which radio channel the user is listening, what shopping website the user is visit, what keyword the user is searching through internet and so on. It is neither the log data, nor the data about interaction behavior (e.g. clicks on UI elements). It only shows what kind of service the user is visited, basically suggesting what kind of activity the user is doing.

There are two motivations provided by the online activities data from CleverCogs. First, it provides opportunities to visualize personal activities for life reporting and summary. Second, regarding to visualizing personal activities in this case, data comics have potential benefits in communicating with family members, relatives of users of CleverCogs. I will clarify benefits and potential advantages in use case section below.

Use case

Generating data comic upon online activities data from CleverCogs can depict life of users. The users of CleverCogs are mainly elderly and people with disabilities who needs care and support. One thing in common for them is they usually live alone and separate from their family and thus in need some of independent support. Due to the isolation and separation, their life habits, interests may not be known to their family, could result in loss of a sense of connection. Even though video call nowadays allows mutual conversation, not all life aspects can be covered through speaking. Their daily activities data can give more insights into personal life, for people who care about

them. Hence, the visualization of personal life could serve as a communication tool which enhances a connection bond for their families, and potentially remote relatives.

Under this context, generating data comic about personal activities over time is motivated by the need to visualize personal activities of users towards their family. Considering of target audience is casual audience in this case, data comics can take advantage of engaging audience to users life story. Rationales for applying data comic in this case over other genres are supported by literature review in chapter 2. Moreover, since everyone's data is different, it is not possible to create data comic one by one for each person. To adapt data comic to this use case, I intend to design a generic data comic template that can potentially apply to any individual to generate their own story.

1.3 Design Question

To report personal activities through the data comic template, possibilities would be report on different time span: daily, weekly, monthly or yearly basis. I select weekly basis for this use case considering the moderate amount of information and the situation where family visits on weekends every week. Hence, the design question is framed to be: how can (potentially remote) family members keep track of weekly life with their relative? In this case, I respond to this question with a designed template, through which family members can generate data comics about their relative's weekly life by themselves.

1.4 Design Methods

To design the data comic template about the personal activities, based on activities data from CleverCogs, I first conduct data cleaning and exploratory data analysis to deliver insights. Based on the messages I obtained in data exploration stage, I choose one topic as an example, around which I blend messages with storytelling techniques, and experiment with comic layout design. Referring to data comic design patterns (Bach et al., 2018), and data comic creation model (Wang, 2022), I experiment and iterate with various elements and design considerations of data comic through two sketches and three versions of design. Finally, the data comic demo, which I called "weekly journal" is presented as an example to illustrate one week life about a user Lora (pseudonym).

There are multiple tools and techniques involved in the design. To give a broad overview, the data is processed with Python. Besides, I use pandas and associated visualization packages seaborn(seaborn, 2012), matplotlib (matplotlib, 2012) during the exploratory data analysis. Regarding to the comic design, I adopt existing characters library "Olá" (and "The Little Things" (Sales, n.d.) and adapt them to my design by modification in PS. The graphic and interactive elements of comic are arranged in an interface design tool Figma.

1.5 Contribution, Novelty

While dashboards, infographics and traditional charts are widely distributed in industry and consumption products, few examples of data comics found to be reusable across fields and scenarios. Especially I find there is no data comic template presented, which prevents data comics to be deployed with updated data in a long-term monitoring or tracking. This research explores to generate data comic through template as a pioneer.

Rather than manually creating through scratch, generating data comic through the template can be beneficial in efficiency, especially regarding to data updated periodically. The use case in this research can be a good reference for future researchers, practitioners and designers. In addition, it motivates a larger picture to create such templates upon other domains of data in the future.

1.6 Outline

Chapter 2: It starts by building a theoretical background drawn from fields of data visualization and visual storytelling. In each field, features, strengths, weakness of relatives state-of-art visualization genres and forms are given. Then, the features, current methods and tools in supporting creating data comics, challenges in creation and gaps are given about data comics.

Chapter 3: It describes the design process of creating data comics through defined templates. It starts from data exploratory data analysis on Clevercogs data, followed by stages to establish the context and categorize messages. Then, the design features two sketches and three versions of templates through iterations to deliver final demo as a 'weekly journal" about Lora. The feedback from audiences and results are presented to inform the reflection in chapter 4.

Chapter 4: It reflects on successful aspects of design considerations in the design practice and questioning in a larger space of applying comics and storytelling in visualization of data. Then, it discuss challenges posed in practice of design template to generate comics, as well as limitations in terms of generalizability and customizability.

Chapter 5: It gives a brief summary regarding to response to the research questions, the contribution of this research and suggested future work informed by the design practice.

Chapter 2

Related Work

Data comic, as a novel genre, has its root in field of data visualization, and draws features from comic(Bach el al., 2018). Visualization is necessary to make data available to wide audience. In order to generate life reporting from data based on application case of Blackwood, the field of visualization and more specific genre: data comics are referred. This chapter cover three subjects of study: data visualization, visual storytelling, data comics to argued for three question: 1) Why visualization; 2) Why storytelling; 3) Why data comics. Together they inform the options of data comics for this design and implies what advantages could be taken. In section 2.1, I set up the fundamental theoretical background of data visualization, and step into features of relative mainstream visualization tools. Section 2.2 introduce the notion of storytelling to be applied in scenarios, and a brief discussion on relative genres in visual storytelling. The terms "visual storytelling" and "narrative visualization" are used interchangeably in the paragraph. Section 2.3 focus on a discussion of data comics, the genre upon which the research works. Specifically, it critically compare features of data comics with the alternatives, and draw the design considerations involved in relative methodologies for creation. Finally, some challenges and considerations in data comics creations and the gap are summarized, which informs me to take care of in this research.

2.1 Data Visualization

Visualization helps communicating with data to audience. Robert points that visualization is originally for present and explain rather than analyze (Kosara Mackinlay, 2013). From the viewer's perspective, the benefit of visualization is obvious. It saves

memory and cognition resources by exploiting value of visual system compared to simply viewing tabular raw data (Kosara Mackinlay, 2013).

Kirk(2016) defines data visualization as "the visual representation and presentation of data to facilitate understanding". This perspective implies two essential activities involved in data visualization: (1) Visual representation by encoding data into visual elements; (2) presentation by adding design decisions of interactivity, annotations, compositions,etc (Kirk, 2016). Visual representation is built from "marks" and "channels", where "marks" refer to dots, lines, areas and "channels" are properties such as color, size, position (Kirk, 2016). The basic building blocks together form the common chart types used in data visualization: line chart, bar chart, pie chart, scatter plot, etc. Consuming data through charts provides benefits in understanding and recognition (Gershon Page, 2001)(Kosara Mackinlay, 2013). As opposed to raw tabular view, the chart affords visual presentation of patterns, trends and outliers to interpret (Kirk, 2016).

2.1.1 Dashboard

Data visualization encompasses a variety of graphic forms. It has its root in early maps and diagrams(Friendly, 2008). William Playfair is credited for the first use of graphic forms that are widely seen today: line charts, bar charts, pie charts and circle graphs(Friendly, 2008). The statistics and scientific reporting has contributed to modern visualization forms of data. With the power of computational graphics in modern days, there are many libraries to generate statistic figures from data through codes (e.g., matplotlib, seaborn, etc.). These data visualization libraries has been used extensively in scientific exploration and reporting. However, these libraries are usually implemented in a programming environment which require data analysis and programming skills. To make the process of data visualization more accessible to wider public, there has been data visualization tools that feature a GUI to interact with data such as Tableau(Tableau, 2019), Microsoft Power BI. These tools are widely employed in business under the data-driven world. These popular data visualization tools mainly deliver solutions of "dashboard": an overview panel view of data through multiple chars and graphs. It has been extensively used in organizations for decision making (Rahman, Adamu, Harun, 2017).

However, the dashboard encompass a range of charts and graphs, which require related data literacy and visual literacy for readers to interpret. The key messages

extracted from observation and interpretation on patterns, trends and anomaly, sometimes are not explicitly stated. Another challenge in consuming messages through dashboard is the lack of engagement and compelling experience it provides(Sarikaya, Correll, Bartram, Tory, Fisher, 2019)..

2.1.2 Personal Informatics

Regarding to visualizing personal activities data, there are personal informatics tools that track sleep (Hoefer, Schumacher, Voida, 2022), health (Shin Biocca, 2017), heart rate (Burma, Lapointe, Soroush, Oni, Smirl, Dunn, 2022), emotions (Fairclough Dobbins, 2020), etc., in support of self-reflection or behavior change. The studies of personal informatics usually involve: (1) data-driven insights elicited from personal data; (2) map the data-driven insights to visual elements to create the visualization (Choe, Lee, Schraefel, 2015) (Pousman et al. 2007). Regarding to data-driven insights, there has been eight insights types (including trend, comparison, self-reflection, correlation, etc.) concluded from previous analysis (Choe et al., 2015), which are similar to the notion of "messages" used as materials to build data comics. In terms of visual representation of insights, theoretically visual forms can be varied in a range of style, but bar chart, line chart, map are mostly found in practice (Choe et al., 2015) (Fairclough Dobbins, 2020) (Janzen McGrenere, 2022). Similarly, a analysis has found line chart, bar chart and scatter plot are three visualization types with top usage adopted by people to communicate personal data insights (Fairclough Dobbins, 2020).

2.1.3 Implications

The review of current data visualization main streams suggest an extensive adoption of data charts(line charts, bar charts, scatter plots, maps, etc.) in a dashboard. It is partially attributed to intuitive visual encoding (e.g. amount reflected by the height in bar charts), and established protocol to generate such charts in an efficient way. However, a single visualization may not engage audiences with data. The next section 2.2 introduces the notion of storytelling, also referred as narrative, to be incorporated into the visualization.

2.2 Visual Storytelling

While visualization promotes data communication through the use of visual-perceptual system, visual storytelling (also known as narrative visualization) leverages visualization by taking advantage of human mind. The human mind is born to understand story (Gershon Page, 2001).

The storytelling languages involves plot, point-of-view, characters, themes, dialogues, setting, etc. The plot, sometimes referred as narrative, represents a series of events. The typical storytelling structure employed today has been evolved from Freytag's Pyramid consisting of exposition, rising action, climax, falling action, and resolution(Freytag, 1984).

The difference of visualization and visual storytelling lies in the use of narrative. Story is commonly seen in media of TV, video, animation, film in modern day, while the earliest introduction of storytelling in visualization was in 2003, where Nahum used story-like visual presentation to describe a hostage situation (Gershon Page, 2001). The story is able to accommodate more complex and massive information than static image (Gershon Page, 2001). A sequence of image is suitable for describing time-dependent events (Gershon Page, 2001). "Engaging", "appealing" are terms that frequently used in literature to appreciate using storytelling as techniques to convey message (Gershon Page, 2001). Storytelling also provides "uncertainty" which leaves a space for audience to interpret in a range of possibilities (Gershon Page, 2001).

Storytelling has long been used in journalism, newspaper stories, presentation and scientific research (Kosara Mackinlay, 2013). Most of previous literature worked on seven genres of narrative visualization established by Segal and Heer(2010): magazine style, annotated chart, partitioned poster, flow chart, comic strip, slide show, film/video/animation. More novel genres have drawn from or beyond the existing seven genre, such as data video (Amini, Henry Riche, Lee, Hurter, Irani, 2015) (Amini, Riche, Lee, Monroy-Hernandez, Irani, 2017), data comics (Zhao, Marr Elmqvist, 2015), data-gif (Shu, Wu, Tang, Bach, Wu, Qu, 2021), performance (Latulipe, Wilson, Huskey, Gonzalez, Word, 2011), data storytelling in AR (Zhao Elmqvist, 2022). The choice of genre depends on audience, medium, story, and data (Zhao Elmqvist, 2022). Genres are associated with context scenario: slideshow for presentation, dashboard for BI monitoring, video for commercial ads (Zhao Elmqvist, 2022).

Segel and Heer 2010 discussed three narrative visualization structure: (1) martini glass structure by initiating with questions, observations through a narrow mouth,

then open the space for exploration after finishing intended narrative; (2) interactive slideshow that allow mid-narrative interaction in the slideshow process; (3) drill-down story by presenting the theme while allowing further investigation of details on selected instances. These three structures are good reference for implementing a narrative visualization. However, there is no studies of comparison on different narrative structures employed regarding to the effectiveness, engagement, and other user experience factors.

The techniques largely used in visual storytelling include: (1) labels that indicate the narrative structure; (2) arrows that help readers to follow some orders such as chronological events; (3) on-chart annotations that direct audience's attention; (4) tool-tips for interactive explanation[ah]. These techniques were further classified as narrative structure tactics, including: (1) orders that imply paths for audience to follow; (2) messages that refer to text annotations, labeling, captions; (3) interactivity such as tool-tips, filtering, searching, and time-sliders (Zhao Elmqvist, 2022).

The above three narrative structure tactics suggest three aspects of design strategies in narrative visualization design. (1) Order: it is usually superimposed by author. While an open space allows greater flexibility to explore, an explicit ordering is an appropriate reading aid for passive readers. The ordering can better guide the audience through the reading process. (2) Message: it provides clarity through explanation and annotation, but it also induces clutter (Zhao Elmqvist, 2022). (3) Interactivity: it creates entertaining effect and invites an involvement. However, it can also be distracting from the intended messages (Zhao Elmqvist, 2022).

2.3 Data Comics

Data comics are adapted from comic-strip visualization for telling stories about data (Zhao Elmqvist, 2015). It draws features from comics, and incorporates storytelling techniques in delivering data insights. Bach et al. (2017) proposed four essential components of data comics: (1) Visualization: Visual illustration of messages that range from realistic to abstract style.

(2) Flow: It is also known as sequencing that imply an undirected or directed reading order. It is closely related to transition between panels.

(3) Narration: Narrative built upon context that transforms factual information to story.

(4)Words and pictures: The integration degree of verbal and visual elements.

2.3.1 Comparison: Data Comics, Data Videos, Infographics

In terms of delivery data insights through an integration of text and pictures, there are two alternative addressed: infographics and data videos. They both afford more aesthetic and engagement attributes by illustrations compared to single data visualization charts. The narrative are basically involved in data comics and data videos, whereas infographics are not necessarily the same.

Study conducted in previous research demonstrated benefit of data comics by promoting engagement and memorability, facilitating understanding, providing clear reading order and enabling complexity breakdown over infographics (Wang, Wang, Farinella, Murray-Rust, Henry Riche, Bach, 2019). However, there is a trade-off between benefits of complexity breakdown and risks of visual repetition, redundancy, given by panels and visual graphics (Wang et al., 2019). The significant difference between data comics and infographics lies on how strong the guidance given for audience (Wang et al., 2019). Data comics superimpose explicit reading order by sequence of panels (Wang et al., 2019)(Bach el al., 2018). Infographics, however, allow greater flexibility in exploring and comparison by presenting levels of information through one hierarchical panel (Wang et al., 2019). However, there is no study of differences of data and visual literacy entailed by data comics and infographics. I suspect that infographics are more suitable for active audience while data comic can be consumed by passive reading where audiences simply follow the sequences of panels.

Unlike the infographics, both of data comics and data videos share the explicit reading ordering: panel-by-panel for data comic, and frame-by-frame for data video. Data comic involves spatial arrangement, whereas data videos features on temporal characteristic. The benefit of involving the spatial layout is an overview of all contents provided in one page, whereas data videos can only be navigated through playback. It suggests an message can be easily located in data comics, but not for data videos.

2.3.2 Data Comics Creation Support

To assist data comic creating practice, previous research proposed relative design patterns (Bach el al., 2018), and data creation model (Wang, 2022). Both of them set the fundamental framework for the data comic creation support.

Design patterns for data comics

The design patterns proposed by Bach. et al. in 2018 informed a design space for data comics (Bach el al., 2018). The design patterns describe a solution space for various narrative purpose through two dimensions: spatial-panel layout that implies a spectrum from linearity to non-linearity and content relation between panels that implies transition (Bach el al., 2018).

The design patterns demonstrate usefulness in facilitating the data comic creation by assessment in a workshop (Bach el al., 2018). The series of design patterns are referred as useful guidance to inspire possibilities of story structure, transition and panel layouts during data comic creation process.

However, there is no discussion on the appropriate use cases of and differences among pattern solutions. Theoretically, more than one pattern available to apply upon one case. Sometimes, it is hard for authors to select the suitable pattern solution for specific story (Bach el al., 2018) (Wang, Dingwall, Bach, 2019). Basically, the design choices depend on subjective understanding and interpretation on dataset, messages need to be communicated, the communication purpose, and story.

The factors of pattern solutions are not evaluated upon user-experience goals such as engagement, understandability, effectiveness, recall. Thus, if the ultimate design goal involves user-centered objectives, for example, prioritize engagement and fun over recall, multiple design alternatives with respective evaluation are desired in this case. Nonetheless, the investigation into standalone factor is challenged. This is because the perceived final presentation is a contribution of layout, transition, visual graphic elements, even the literacy of audiences. Then, how can we argue that "time-ordering" is superior to "overview+detail" approach for particular data story case?

Data comic creation model

Wang introduced data comic creation model that describes components of data comic through five layers built from the bottom (Wang, 2022):

(1) Communication objective: the audience, communication context and scenarios that inform the subsequent design.

(2) Message: the information fragments derived from data that need to be communicated. The messages selection often subject to communication objective layer at the bottom.

(3) Narrative structure and explanatory visualization: storytelling structures (re-

viewed in section 2.2) employed to connect fragmented messages into narrative with data-based evidence supported.

(4) Visual-narrative integration: a range of design considerations to be considered when transform narratives into visual graphic layout. It draws from work of design patterns by Bach[], and complement it with other components. The design considerations include pace, transition, layout and narrative guide.

(5) Graphic appearance: Visual embellishment, visual arrangement and appearance added to the comic. Currently, it is mostly depend on author's personal preference and communication objective, and yet to be investigated.

The data creation model provide a high-level framework to guide the creating of data comics. The data analysis stage to prepare messages is not covered in the model. Beyond that, it covers basically all essential elements that related to data comic design, and suggests what factors should be concerned during the process. While it not going to the specification for individual component, I draw ideas from other fields to supplement the component with concrete techniques and design strategies.

It should be noted that this model is a recently proposed work by Wang (2022). Hence, it requires further evaluation on its usability and effectiveness in practice. Wang also pointed out the dynamics in creation process, so it is not a solution to generalize data comic creation process (Wang, 2022). I refer to this data comic creation model for some key elements as well as design considerations mentioned in the visual-narrative integration layer, and adapt it my workflow of design a template.

I adapted this model to my design of data comic template. I referred each layer as a design stage in Chapter 3, but complement it with stage of exploratory data analysis for narrative. Each design stage may share the same name as each layer, but the meaning may slightly different in my context, in which the relative terms are clarified.

2.3.3   State of Art, Challenges & Gaps

While given the benefits of engagement, understandability the data comics provide. To enable it apply for the updated data, a template is desired. However, there is currently no research working on how to generate data comics through template. This research would response this by an application case of CleverCogs.

Regarding to the challenges demonstrated in the data comics creation process, with design patterns available, it is found that people were struggle to select a pattern (Wang, Dingwall, Bach, 2019). It is also found that people feeling challenged to blend nar-

ratives with messages in comic creating (Wang, Dingwall, Bach, 2019). This design considerations are discussed in my design, and I tend to use a template to scaffolding the generation in practice.

Another gap found in the data comics study is the lack of examples about reporting personal activities data as well as no practice of generate data comics through templates. The state-of-art in data comics involves extending the application case, e.g. applying data comics for reporting user studies (Wang, Ritchie, Zhou, Chevalier, Bach, 2021), use data comics in co-design process(Hasan, Wolff, Knutas, Pässilä, Kantola, 2022), incorporate data comics in computational notebooks (Kang, Ho, Marquardt, Mutlu, Bianchi, 2021). Other researches touch on fields on customizing the interactivity data comics, e.g. incorporate interactivities in navigation of network-based data, use pen and touch to draw data comics about dynamic network (Kim, Henry Riche, Bach, Xu, Brehmer, Hinckley, Pahud, Xia, McGuffin, Pfister, 2019). Although some data comics involve interactivity, but they are limited to the space of changing different view or filtering the data. More innovative skills is to relate audience and invites them to put personal data that can take part into the data exploration (Wang, Romat, Chevalier, Riche, Murray-Rust, Bach, 2021).

To draw the benefits of data comics, as well as to extend the use of data comics, this research refer to the design considerations of data comics discussed above and explore to use template in producing the data comics in design process.

Chapter 3

Design

3.1 Design Objectives

The design section aims to take advantage of data comic to address the design question: "how can (potentially remote) family members keep track of weekly life with their relative?" To keep track of weekly life entails a template to support generating in a weekly basis. Hence, based on CleverCogs data as a use case, I first need to specify the structure of template, and then present a data comic demo, named a weekly journal, through this template. In addition, I shall explain on the demo to specify which parts are subject to dynamic generation from data in the future.

Aiming to generate data comic about weekly personal activities, I identify five sub-objectives to achieve:

1. Define the communication context in terms of data, target audience and the scenario.

2. Based on communication context, extract intended messages from raw data.

3. Sort through and weave messages into storytelling patterns.

4. Design a template which implies how storytelling patterns are integrated into the visual layout.

5. Present a data comic demo, named weekly journal, by applying the template.

The final design outcome is an example of generating data comic through pre-defined template. The example is based on selected data excerpt about a week of one anonymous user. However, the underlying template is intended to fit data given about any weeks of any users as much as possible. Hence, it suggests variations of weekly journals can be generated in the future.

3.2   Design Process

The design process involves two main parts with subdivided stages: (1) from data to messages (from stage 1 to stage 3); (2) data comic template design (from stage 4 to stage 5).

Stage 1: Communication Objective. I set up the communication context for this use case by specifying the data, target audience, and the communication scenario. Personas and storyboards are used during the process. This stage informs the perspective to be taken in the subsequent data analysis.

Stage 2: Data preparation. The aim is to derive facts, messages and insights, which serve as materials for visual storytelling. I perform the data cleaning and conduct exploratory data analysis prompted by a series of inquires. Python skills, pandas for data analysis and seaborn, matplotlib for data visualization are used in this stage.

Stage 3: Messages. Messages obtained in stage 2 are sorted and classified into different levels. Based on obtained messages, I determine the topic to assemble messages for storytelling purpose.

Stage 4: Sketching. I sketch to explore concepts and basic composition of comic template. Though sketches, I play with layout, transition, narrative guide and interactive elements of data comics. I first create an initial sketch, followed by some discussion with supervisor and classmates. Then I modify and create a second sketch to set up basic ideas, concepts and direction for building up the templates.

Stage 5: Data comic template. There are three versions of data comic templates through iterations. Finally, an example of "weekly journal" is presented through applying the template.

3.3   Stage 1 : Communication Objective

It is necessary to set up communication context in the beginning as it helps to understand the needs of target audience, as well as the potential application scenario of this design. Since I am not able to access to relatives of CleverCogs users in real life, the user research is not possible in this case. However, based on the data at hand, I create persona to explain the value, goal and pain point of target audience, family of Clevercogs users. Besides, the storyboard is created to illustrate the obstacles faced by target audience and how the design solution might help.

The communication objective influence how comic template would be designed,

including what information should be included based on audience interest, in what manner it presents. WangZ's Data Comic Creation Model mention how audience group and setting varies in different conversation scenarios. Informed by data comic workshop by WangZ, in this stage, I defined three components to scope the communication objective: Dataset to be communicated, target audience, communication scenario. To better support clarification, I used design methods of personas[ref] to specify the target audience, as well as storyboard to frame the communication scenarios.

Target Audience

A typical target audience can be characterized by Gloria (left character) shown in persona. Lora (right character), who is Gloria's mother, represents an user of CleverCogs Digital System, contributes the data source.
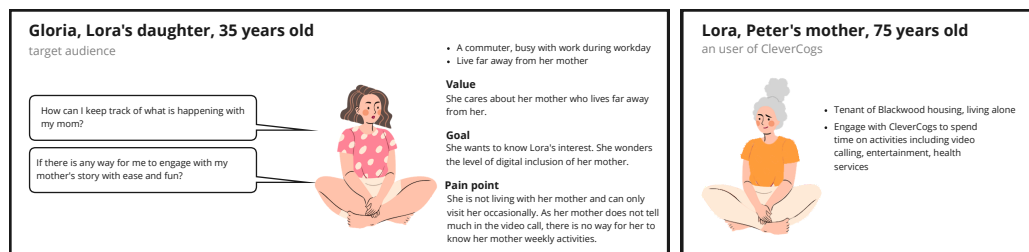

Figure 3.1: The persona specifies the target audience (Glora on the left side) and the user of CleverCogs (Lora on the right side).

Lora as shown in right side, can represent the tenant of Blackwood housing, who has been used CleverCogs to access digital services. The data hence reflect the personal activities about Lora's life. She lives alone and may not be seeing her family everyday, such as her daughter Gloria.

Gloria refers to family or relatives of users of CleverCogs. They are probably informal care providers of elderly or disabled who living alone in this case. The goal of them is to gain more knowledge into life of their relative, who in this persona is Gloria's mother Lora. However, due to the restriction of distance, Gloria can only go visiting her mother occasionally. Therefore, Gloria wonders how to keep track of what is happening with her mother who does not live by her side. It prompts her to seek approaches to engage with her mother's story.

Communication Scenario

One of solution to engage Gloria with her mother's life story is through data comics based on Lora's personal activities data.
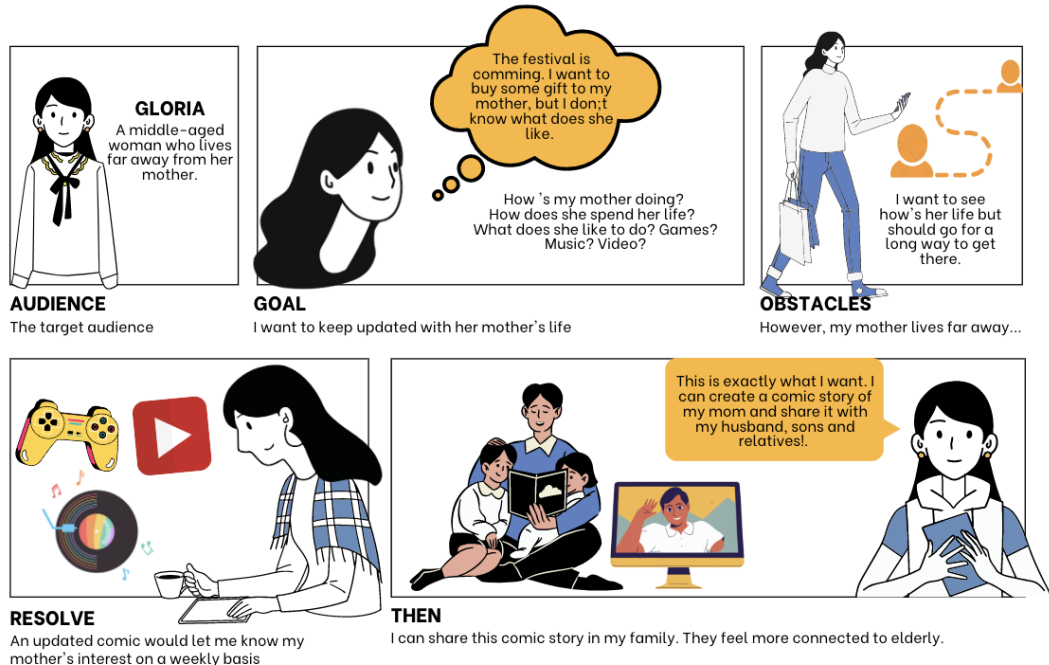

Figure 3.2: The storyboard about Gloria attempting to use updated data comics to learn her mother's interest.

The storyboard above illustrates a possible scenario. Gloria wants to know: "how her mother doing recently"; "how her mother spend her life"; "what her mother like to do across music, video, games". Especially when festival is coming, she wonder what can be prepared. However, Gloria is not visiting her mother every day or weeks. To address this challenge, an updated comic would enable her to have a close-up on her mother's interest. Moreover, the comic story can be shared within the family, potentially strengthen the bonds among family members.

Data to be Communicated

The data to be communicated imply personal activities that convey different aspects of user's life: entertainment, video call, health, etc. In above case, Lora's data about her personal activities are intended to be communicated over time, in a weekly basis. In this way, Gloria can better understand on her mother's interest, preference and habits.

In this research, I select one month of records of an individual anonymous user as data to be communicated for the design. The selected data can only reflect the personal activities about that particular user, but how I work with example data excerpt can also be applied to other users as well.

3.4   Stage 2 : Data Preparation

The aim of the data preparation is to explore what messages and data insights to convey through further presentation. The provided dataset is presented in two spreadsheets, and is examined to have some null values and inconsistent format. Hence, it is necessary to clean the data for the sake of efficient exploratory data analysis. I start by brief introduction of the dataset to work with, followed by description of data cleaning and exploratory data analysis.

3.4.1   Dataset

The dataset is from CleverCogsTM Digital System, developed by Blackwood Homes and Care. The original dataset contains three separate spreadsheets describing users demographic, medical condition and link activities respectively. Especially, the personal activities are stored in form of tabular records with corresponding timestamp when the user visits a "cog". The "cog" represents an user's online behavior such as listening to particular music channel or watching football game. Because this research mainly focus on activities data, sensitive information including health condition and location was not used in the research. The raw data spanned from 2018 to 2021, but only part of them are selected for design.

Raw data is given as two excel spreadsheets:

- CCDate2018-2021.xlsx : Each record represents a link visited by a unique user ID. There are three major fields for each record: (1) "Link type": it represents the visit categories such as "Category" or "Internet"; (2) "Link title": it contains semantic information about what the activity is about; (3) "Time": it suggests the time when there is a visit to a activity link.

- Users2021.xlsx : Each record represents a user with associated information of "Birth date", "Gender" and "Location". "External ID" uniquely identifies an individual user, and can be used as the key to join with other tables.

3.4.2   Data cleaning

To improve the quality of data for further analysis, I perform data cleaning in interactive jupyter notebook environment provided by Noteable[]. Python library: pandas[] is mainly used in data cleaning process, in which four aspects are examined (1) null values; (2) typo errors; (3) data conversion; (4) join tables into one table for analysis. Specifically, I perform the data cleaning through following five main steps:

Step 1: Read spreadsheet by Pandas and convert it into a dataframe, a pandas data structure of tabular data. Check some basic property of dataframe. Drop some unnecessary columns which are not used for analysis.

Step 2: Drop rows for users with "External ID" equal to null. The field of "ExternalID" is used to identify specific user. There are 26.8% missing values found in "ExternalID", which can not be linked to associated activities records. Hence, I dropped rows if "ExternalID" equal to null.

Step 3: Make sure the "ExternalID" is unique. Check whether a column value is unique by Pandas built-in function.

Step 4: Adding additional column of "Age" by calculating based on "Birth date". I define a function that calculate the Age from birth date [picture]. I assume the valid age for users were between 10 to 120. Therefore, I drop the user whose age less than 10 and substitute value of age with NaN for whose age larger than 120.

Step 5: Select subset of activities data sheet if "Role" equal to "User". Merge users data table and activities data table into one table by matched "ExternalID".

Finally, two original data tables were merged into a single table, exported into the "dataset_cleaned.xlsx". For the cleaned data, there is no missing values in field of "ExternalID" which uniquely identifies a user. Besides, it contains important information of activities name and respective timestamp.

3.4.3   Exploratory data analysis

Exploratory data analysis (EDA), as opposed to statistical test, is argued for its ability of informing the discovery, models and hypothesis (Morgenthaler, 2009). Hence, my focus is to motivate take-away messages in the interest of audience, rather than examining the various statistical measurement.

The EDA process is primarily about transformation of data into visualization by plotting[d2]. The graphical summary of data can reveal the "shape" of data, and suggest relationship between data[d1]. To visualize "shape" of data, I conduct exploratory

data analysis based on cleaned data sheet. More specifically, I randomly select one month of data, which is the October of 2021, as an example to explore for further design.

Unfortunately, there is no guidelines and related works found about using EDA for the purpose of narrative. It is partially because an open mind is required during the process. In this case, I take the perspective of target audience by asking what might be interest to them. Then, I devised a series of exploratory inquires about data from the overview into the details.

The EDA process involves three parts. I conduct this practice in jupyter notebook, an interactive python environment. I use libraries of pandas, seaborn, matplotlib and ipywidgets for data analysis, visualization and interactive exploration. The workflow goes through from part 1 to part 3 as described below.

Part 1: data scope

The part 1 is to narrow down the data scope by selecting a user's one-month activities for weekly analysis. Specifically, I select October of 2021 as the targeted exploring case. Then, I investigate and rank according to the number of visits in October for each user, and discover a long tail effect in the distribution of visit times per user in October. Among all users who have shown an active visit in October, only 37.6% of users whose number of visits per month are more than 100. I can hardly find any patterns if the number of visits is quite few. Thus, for a consideration of informative richness, I select an user with highest number of visits per month, who I give a pseudonym Lora. The data comic story then is built about Lora's personal activities.

Part 2: week analysis

The part 2 is to conduct a week analysis based on Lora's data obtained in Part 1. The data contains the following fields to investigate:

- ExternalID : It implies an individual user.

- Time : It specifies the time in a timestamp format when an user visit a link.

- Link Title : It suggests what activity the user has done.

The fields suggest an opportunity to investigate the relationship between the time and various activities. What did Lora do during that week? What can be inferred

from what Lora did in that week? Does it suggest the Lora's interest? Moreover, I expect to see more details regarding to what was routine for particular activity. For example, when did Lora usually do that? Besides, I speculate if there is any patterns underlying showing a significant increase or drop in terms of what kind of activities. These patterns can attract audience attention, and inform either a notice or an alert.

The main feature of the dataset is the sequence of data points over time. It is referred as time series data. However, each data point is categorical value at respective time, thus not for showing the quantitative measurement change over time. In order to take advantages of time series data, I aggregate data points over a period of time and count observations according to "weeks", "weekday", "day", and "hour" time unit.

Step 1 : Creating "month", "week_of_year", "weekday", "day", "hour" labels from time series

```
In [5]:  # add time attributes columns from timestamp
         df['week_of_year'] = df["Time"].dt.isocalendar().week
         df['weekday'] = df["Time"].dt.day_name()
         df['month'] = df["Time"].dt.month
         df['hour'] = df["Time"].dt.hour
         df['day'] = df["Time"].dt.day_of_year
         df
```

Figure 3.3: Code snapshot: Extracting time unit features to prepare for later "group by" on observations.

The challenges lies in the massive types of activities in various life aspects. There are 826 different activities demonstrated in dataset of 2021. Hence, it is not efficient to examine activities one by one. To deal with large quantities, I classify multiple activities into a broad category. For example, the activities with title of "play games", "tablet games", "Jigsaw Puzzle Game" can be placed under the "game". Similarly, "Shopping" covers a list of title such as "tesco", "waitrose", "sainsbury", "lidl", "aldi", etc. "Health" is defined to include "my care", "care notes", "care team", "my care survey". Because the classification is the manual encoding from my personal understanding, it is unlikely to exhaust all the activity types. In the future, I hope additional information on activities categories can be provided.

During the data analysis, I use following seven questions to guide through process. Q1 to Q3 give an overview in last week, while Q4 to Q6 step into details of game playing of Lora. Q7 is to explore trends regarding to the activity visits shift over time. HereI clarify a term "last week" used in the questions below. "Last week" refers to time from 25/10/2021 to 31/10/2021. It tends to simulate a scenario of reporting last week personal activities.

Q1: How many visits in total during last week?

266 visits totally found by calculating the length of dataframe about last week.

Q2: How does total visits last week compare to other weeks in October?

I call the countplot() function from seaborn library to counts the observation over weeks in October by "week_of_year". I visualize it through bar chart and change x-axis tick values to corresponding time range.
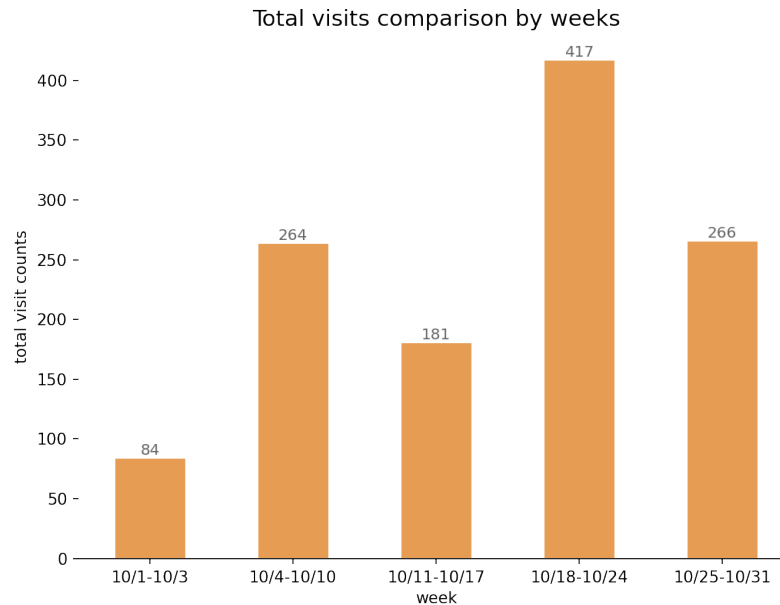

Figure 3.4: Total visits change across weeks

We can see from chart, the total visits of 10/25/2021 to 10/31/2021 dropped significantly from 417 to 266 compared to one week ago. However, if we compare it to earlier weeks in October, we can find the number is in an average level. It implied Lora was much more engaged in terms of one week before the last week.

Q3 : What did Lora do during last week?

I count the observation regarding to all activities of last week and rank them in a descending order. The chart shows the only 6 dominating activities contributing to total visits, while others are only observed 1 or 2 times in a week. There are 20 activities observed for last week, ranging from game, music, family connection, and other entertainment.

Q4 : Which activities made up the majority, suggested by the top 5 for last week?

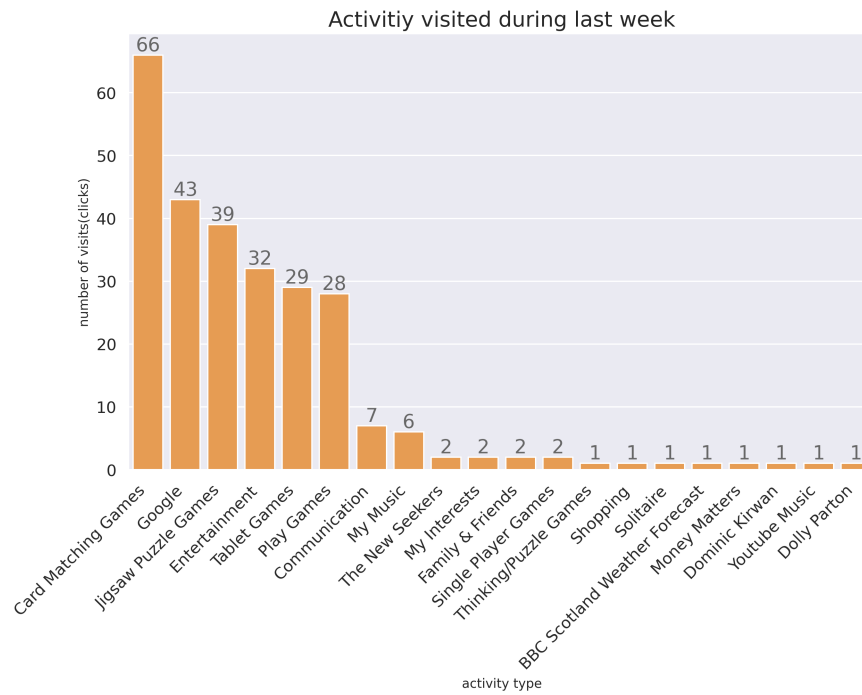If there are too many activities shown in a bar chart which demonstrates a long tail

Figure 3.5: All activities visited by Lora are ranked in a descending order.

effect, we can focus on the main contribution. I assume the interests of Lora can inferred from the several top activities during last week. Hence, I looked up for top 5 activities, among which 3 of them are games. This result informs me to investigate further into the game categories of Lora.
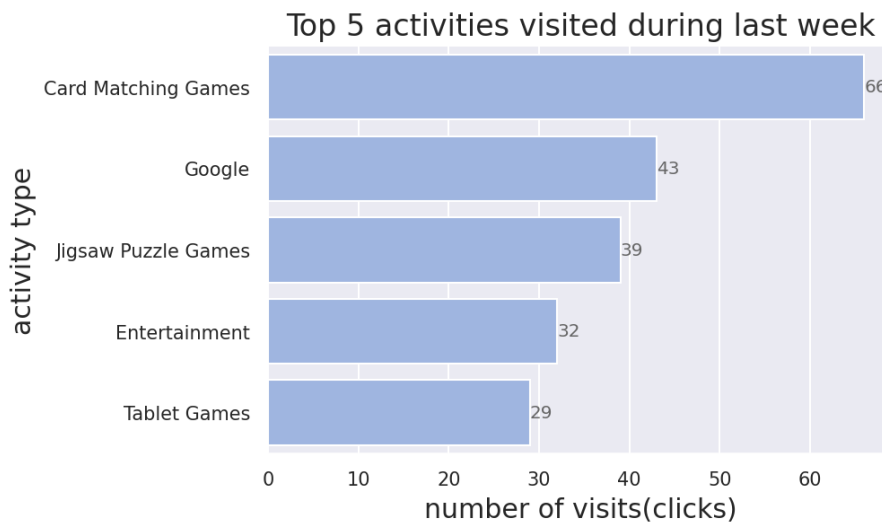

Figure 3.6: Among top 5 activities, three of them are games.

Q5 : What day of the week did Lora usually play games?

Learning that game might be the primary activities of Lora, I step further into the weekday routines of it. First, I select all observations of games by search for "game" words contained in the string "link title". This method however imposes the problem of missing out games whose title does not include string "game". Hence, it requires some domain knowledge to annotate each activity into categories in the future. Then, I count the observations based on weekday. The chart shows that the number of visits on Sunday is significant higher than other days. It suggests that Lora usually plat games on Sunday.
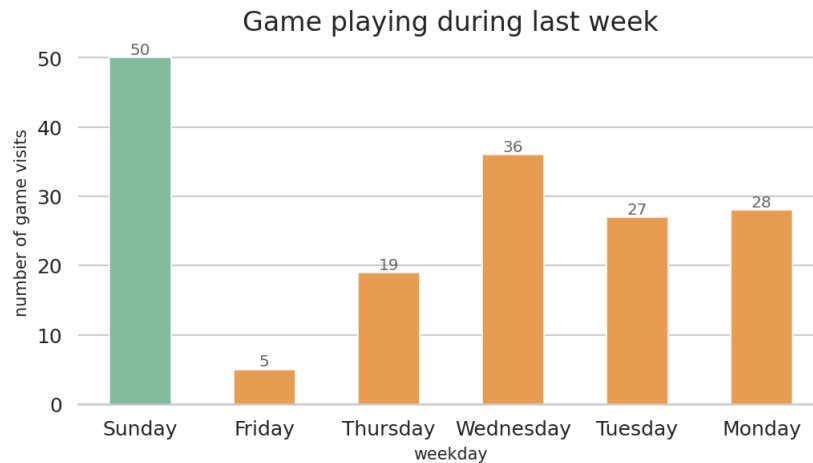

Figure 3.7: Gloria mostly played games on Sunday

Q6 : When did Lora usually play games last week?

In addition to examine the weekday routines for Lora playing game, I also look at when Lora usually played games during that week. It is evident that Lora usually played game at around 8 o' clock in the evening.
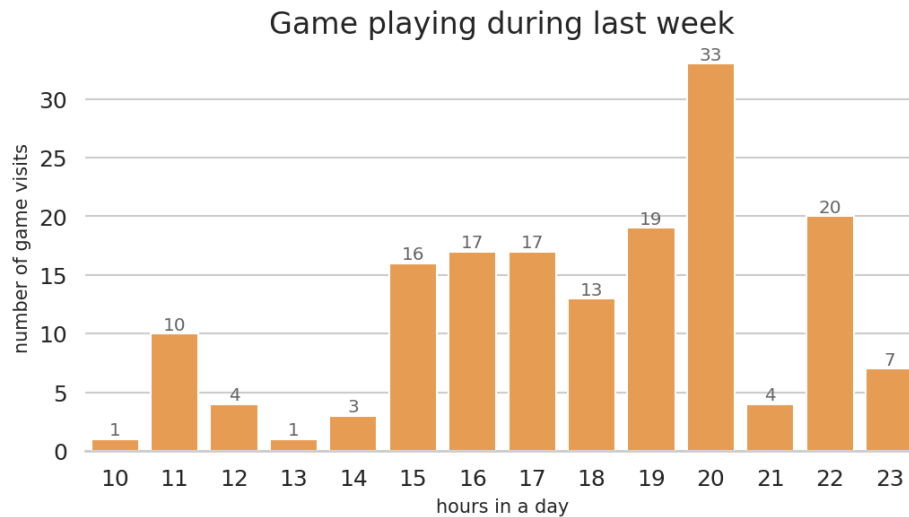
Game playing during last week

Figure 3.8: Gloria mostly played games on Sunday

Q7 : What has Lora been up to lately? To visualize trend of various activities visited by Lora across weeks in October, I create a table where columns referring to various activities, index referring to the week ordinal of the year. In this way, I aggregate the number of visits over weeks for each activity and visualize them through line chart. In the line chart, I look for trend exhibiting a sharply increase or decrease.

A sharply trend is located in below chart from 41 to 42 (week ordinal of the year 2021) on x-axis. There is a sharply increase of number of visits for "Google" while only slightly increase for other activities from week 41 to week 42. It implies that the dramatic increase of total visits in week 42 is largely because the contribution of "Google". Lora might be engage into searching something during week 42.
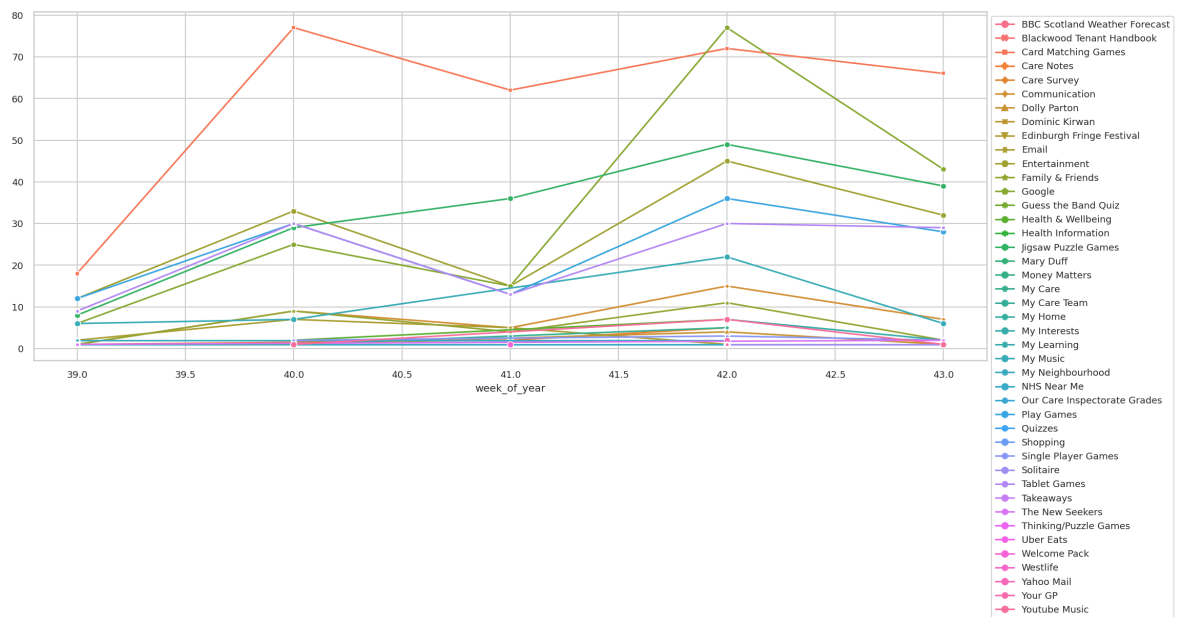
Figure 3.9: The number of visits by activity types encoded in color for each week

However, this discovery is user-specific. While it could be identified through interpretation from visualization on a particular userit raise difficulties in applying to other users. Not all users share the similar trends in their data. To report such trends for every user, more advanced data mining techniques are desired to automatically identify the pattern, which can be further developed in the future.

Furthermore, based on Q5 and Q6, I apply similar approach to investigate the weekday routine and daily routine of activities other than games for Lora. I include an interactive drop-down menu in the notebook, where all activities visited by Lora can be manually selected. The corresponding visualization in terms of visits over weekday and visits over hours in a day are shown below with respond to selected activity.

3.5 Stage 3 : Messages

The messages are selected "data excerpts" from exploratory data analysis[e1], the building blocks for visual data stories. The messages are determined in consideration of two aspects: (1) what would family and relatives of CleverCogs's users like to know; (2) what are valuable information obtained from exploratory data analysis? There are some messages such as interests, total visits which can be applied to broad users, while others such as evident trends over time are not identical with respect to each user. For example, I find Lora once visited an Edinburgh Fringe Festival that

has never visited before. It implied the festival is coming and may be interested to Lora. Such discovery is not generic to other users and can not be applied to everyone. Hence, I try to summarize common message patterns that are generic to report every individual's personal activities.

Referred to data creating workshop conducted by WangZ[], the messages are classified into three groups: high-level messages, facts and data-literacy knowledge[] for building the narrative. The similar approach is adopted in my design for three reasons: (1) In terms of data-literacy knowledge, I do not use complex charts and visualization to communicate the facts to casual audience. I intend to inform messages through simple and evident descriptive data measurement such as sum, maximum number and frequency distribution; (2) it is useful in design stage to determine what facts and high-level messages to involve in comic, as well as to evaluate the understandability of final design in terms of facts acquisition and high-level messages interpretation.

High-level message	Facts	Data-literacy knowledge
① How 's his/her level of engagement in this week?	Number of activities records in this week	Sum
② What is his/her interest?	Top five activities ranked by visit times.	Maximum. Top five number..
③ When did he/she usually do that?	Frequency distribution on weekdays and hours.	Frequency distribution

Figure 3.10: Classifying messages into high-level messages, facts and data literacy knowledge.

I sort through the results found in exploratory data analysis from Q1 to Q7, and summarize them into three high-level messages with corresponding facts and data-literacy knowledge, which could apply to any specific user in the dataset. The messages indicate what I want audience to get out from the data comic. (1) Audience can get overall idea of user's level of engagement in this week by number of visits in total. (2) Audience can know the favorites and interests of users by the rank of visit number with respect to each activity. (3) Audience can know the weekday routine and daily routine for one activity.

3.6   Stage 4 : Sketching

Moving from messages towards visual storytelling, I start by some sketches to explore possible structure and layout of data comic template.

Sketch version 1

The template is used to generate data comic automatically, implying that users do not need to draw by themselves. Hence, I intend to enable audience to build comic by playing with some parameters and eventually create the data comic through one click.
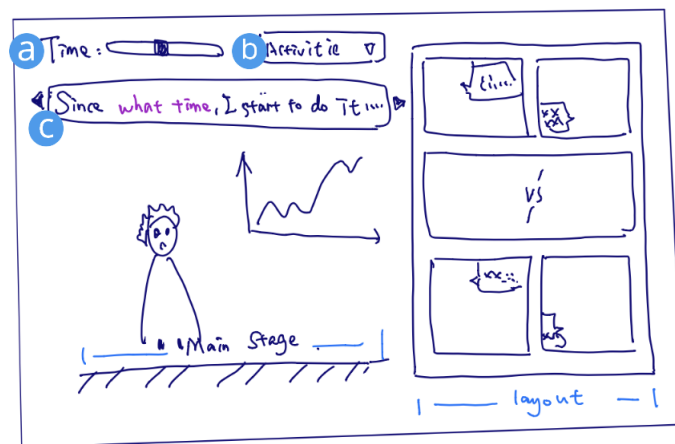

Figure 3.11: The first version of sketches features a sentence pattern generator(c).

The first sketch shown above features a main stage where allow the user to build the scene. It involves an character, the environment and an illustrated graph. The character on the stage represents the user of CleverCogs. The right side is the layout of generated comic through panels. In each panel, there are some dialogue boxes placed at the corner. After defining some basic visual elements of data comic, I start to work on text visual integration considering how to integrate messages with corresponding pictures. The intuitive way is to use text to explain finding directly. I specify some sentences patterns to describe intended messages. The sentence pattern here is the generic expression to state the facts. Variables are used in the sentence to generate variations.

Table 3.1: Orange for "something", represents activity variable. Green for "X" represents number variable. Purple for "during what time" represents time variable

The totally visits for me last week was X
My favorite activity is doing something during last week
I like to do something during what time

For example, in above three generic expressions: "X" is the number of total visits; "something" is one of activities of last week; "during what time" is the time variable that can be weekday, hour, in the morning, at night, etc. These three variables can be returned by processing on the data, or can be parameters as input for audience (a)(b)(Figure 3.11).

After the user identify one day of week through slide bar(a), and select an activity from drop-down menu(b), the related expression would then automatically shown in the speech box(c), such as "I like to listening to music around 9:00pm". The drop-down menu(b) provide option of "all activities" so that corresponding expression will be "The total visits for me last week was X". After the expression has been generated(c), the corresponding illustration of the character listening to music then display on the stage. The relative chart revealing the routine for listening to music would appear on the right-side corner. The character, chart, speech altogether composite a scene that added to the panel by click the "Create" button.

The first sketch is presented to supervisor and people in the wild. The feedback report positive on ideas of introduce parameters and character to engage audience. The generic expression enable speech can be generated through template. However, it fails to imply a sequence through the interactive experience. The comic layout on the right side fails to inform the audience the transition and sequence to comply.

Sketch version 2

To involve the flow and sequence into the design consideration, I refer to data comic design patters about panels layout and content relation (Bach, 2018). Among layout range from non-linear to linear, I find some layouts imply case for a specific purpose, for example, large panel is usually used for granular patterns and rarely seen for temporal patterns. I suppose the grid layout, branch layout and linear layout are beneficial for the generative purpose as an individual panel can be displayed in an successive

order through generation process. It is also better to employ a linear sequence in the template as it is able to accommodate more patterns under dimension of content relation. Since the design is intended for audience to read through digital device such as web in tablets, the parallel layout is more adaptive for this reading environment. An individual panel can be generated one by one from top to bottom. Therefore, I employ a linear layout from top to the bottom by utilizing the features provided by screen space for the second sketch.
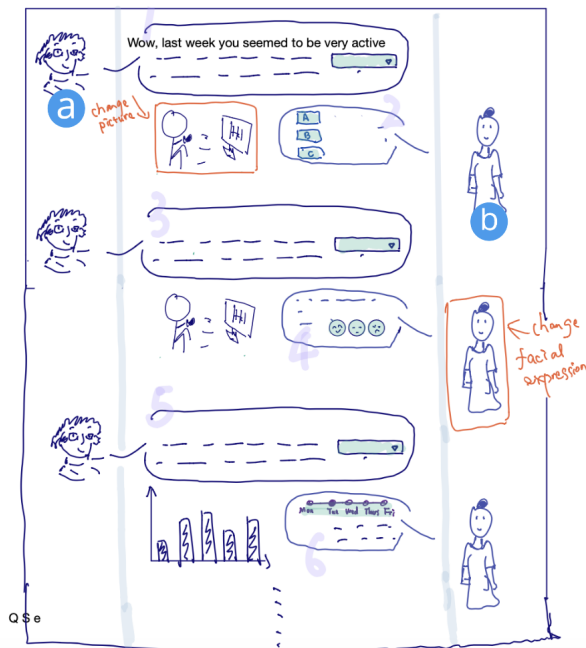

Figure 3.12: The second version draft

The second sketch retains the use of characters, illustration about activity and interactive elements such as drop-down menu. Besides, it involves a clear reading sequence by parallel layout from top to bottom. To propel the engagement, I adopt the narrative pattern of question answer. There are two characters introduced: narrator and main character. The narrator(a) is intended to attract audience attention by raising questions and make inferences. The main character(b) is intended to answer questions from the first-person perspective. The interactive elements such as slider, buttons for options are added into the dialogue text box. In this way, audience read the comic guided by dialogue sequence from top to bottom. They can easily toggle the options (interactive elements highlighted with green) through the process, leading to different generative contents.

3.7   Stage 5 : Data Comic Template

In this stage, I continue to use "Lora" to refer to the main character to whom the story happened. Based on data visualization charts obtained in exploratory data analysis in Stage 2 (Data preparation), and the messages I sort through in Stage 3 (Messages), I further develop the design considerations brainstormed in Stage 4 (Sketching). I apply a linear layout in a visual vertically sequence combined with narrative patterns, faceting patterns and granular patterns in creating the data comic. The use of narrative guide is inspired from sketch version 2 to incorporate with question&answer pattern to arise audiences' interest and draw attention to essential messages.

3.7.1   An iterative process

The template is demonstrated based on case of "Lora". The entire workflow is an iterative process with modification and refinement based on previous version. I elaborate in detailed about layout, structure, elements involved and design considerations in data comic template version 1. The basic idea and overall structure remain for version 2 and version 3. In version 2, I further refine the template in terms of context, text and pace of story. Then, an interactive prototype is tested with students before version 3 in an informal pilot testing to simulate an audience-driven data comic generation process. Based on the feedback from test, I address two issues with modified design in version 3 in terms of structure, sequence and transition.

3.7.2   Visual resources

In terms of graphic appearance, I use an illustration collections "Olá" (Kauffman, 2021). and "The Little Things"(Salas, n.d.) from a resource website blush (Blush, n.d.) The benefit are the graphic variations of characters in different poses and scenes. I can also optimize the characters in terms of their face expression, hair style, clothing and poses. This feature provides opportunities to dynamic generation in data comic.

Figure 3.13: The customization on the facial expression to suggest an emotion change

However, I identify a biggest issue through the creation process along with feedback from supervisor. The clothing provided with characters are not in uniform style with different poses. If I use variations of poses for one character in comic, the different clothing confuse audience as it implies the transition from one day to another. Hence, I modify the original resources (Figure 3.14) into uniform clothing (Figure 3.15) in Photoshop. The challenge encountered in my case also implies the demand to create such libraries to assist creating and generating data comic in the future.

Figure 3.14: Jonanthan: The original characters set from "The Little Things"

Figure 3.15: Jonanthan: Change clothing color into consistent

3.7.3 Data comic template version 1

The objective of template version 1 is to depict overall structure in more details by implementing a linear layout. The linear layout basically implies a linear reading structure and requires the continuity in logical connection and content relation. The comic applies the notion of linear layout from sketch. Additionally, I add one column of control panels on the left to separate user interactive elements from the generative comic (the right column). The control panels walk audience through tour of generating the comic step by step. It comprises titles to imply the structure and steps, explanatory texts to guide the audience, interactive buttons and drop-down menu that allow user's input. The right column comic panels is the space to deliver the generative content. The layout, interactive options, characters are predefined as the template, but the text and charts highlighted in gray are generative content subjective to the data and users' input. The comic consists of four parts I – IV. For each part, I will describe the design patterns, objectives in terms of messages delivery, applied design elements and techniques respectively.
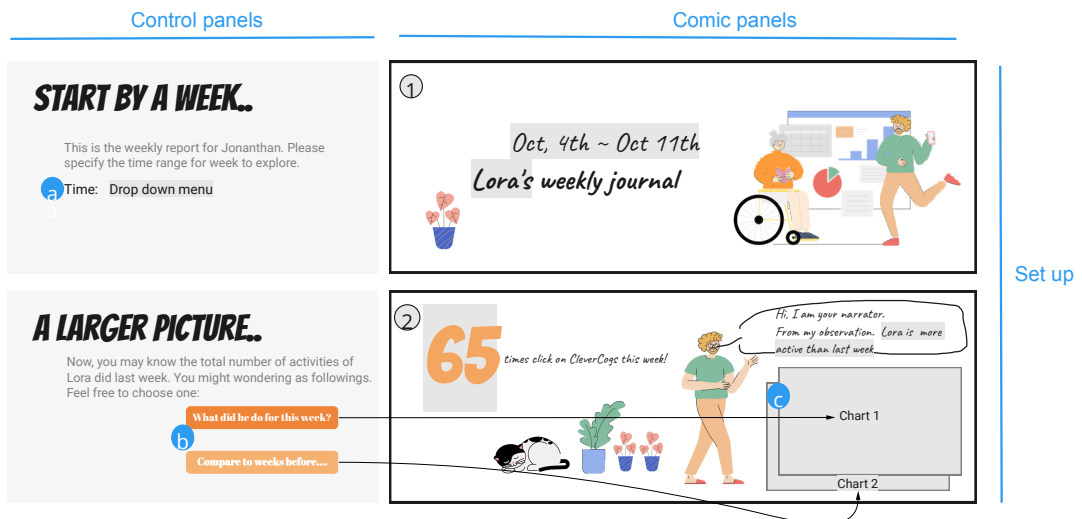

Figure 3.16: The set up scene comprises of a cover panel and a larger picture.

I. Set up: The set up scene (Figure 3.16) informs the time range of weekly panel on the cover panel ①, followed by a larger-picture showing the overview of total visits during last week ②. The set up is intended to enable audience to grasp essential summary of facts by highlighting the week and number of visitsfollowed by details to account for activities contribution and active state. A granular pattern, overview+detail is used in panel ② by providing overview of total visits and supplement with two charts

that relate to number of visits: chart 1 account for how the total visits distributed across ranges of activities; chart 2 relate it in dimension of time by comparing with previous weeks. The details illustrated by two charts are not shown in a sequence, but alternatives that depends on audience's choice ⓑ. The generative charts ⓒ subjective to audience's option to display, providing multiple facets about number of visits last week. Except from the button ⓑ as options for audience, another interactive element is the drop-down menu ⓐ to allow audience specify the week upon which the story is built at the beginning.

II. Branch: The branch (Figure 3.17) provides a scene with options for audience to select a topic to start a story. It acts as a transition that improves coherence between the set up I and the narrative III. The options ⓓ in control panel branch narration into "The favorite" or "One day". The story is then began with a question: "what were the Lora's favorites last week?" or "how did Lora spend her day last week?" It drives the focus of audience and suggests what kind of messages are going to reveal in next stage.
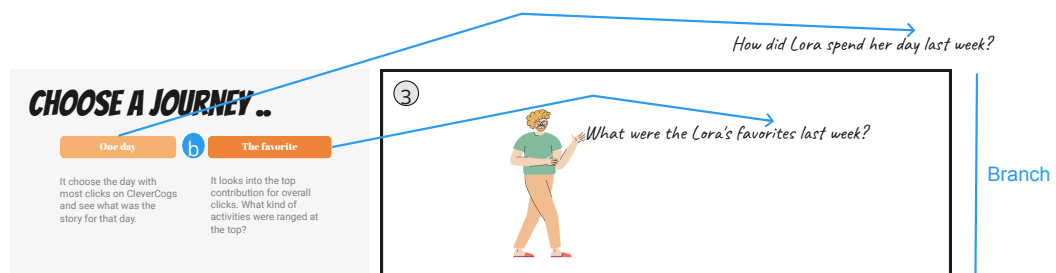

Figure 3.17: The branch provides options for audience to branch into different plot of story.

III. Narrative (Figure 3.17): The showcase is based on the "the favorite" as an example. With questions of "what were favorites of Lora, the narrative part guides audience through the experience of the interpreting, inferring, scenario-based narrative, and facets exploring. First, I use the notion of user segmentation to separate users into groups based on their interests. I assume the interests are reflected by the number of visits per activity in a week. Links that are frequently visited are indicators for users' favorites. The panel ④ highlights the top activities by a generative chart ⓔ based on filters from drop-down menu ⓕ. To engage audience with facts given in charts, the take-away messages are summarized with text and prompt for an interpretation by asking "do you think she is a game lover?ⓖ" In this way, I tend to involve audience into

the interpretation process with facts and their inferences make up part of generative content. After having the idea of the preferable type of activities, which is game in this case, I apply the temporal pattern and question&answer narrative pattern to convey messages of "what was Lora's favorite game", "when did Lora usually play the game", and "what weekday did Lora usually play the game" through panels ⑤ – ⑨. Rather than simply presenting the facts of "when" and "what", I twist it with a temporal sequence by setting up a scene, where Lora is thinking it is time for her to play some game (panel ⑤). Then, an illustration is followed showing that Lora is playing the game (panel ⑥). Next, an interruption is created through a second-person point of view, commenting: "It seems like you are playing the game" (panel ⑦). This is an use of audience surrogate to speak in behalf the audience. The narrator then starts to chat with Lora by asking "what game are you playing?" The answer from Lora reveals what she is playing is exactly her favorite. It achieves the "climax" point in the conversation (panels ⑤ – ⑨). Then, the pace slow down by a further inquiry about "is it a daily routine for you", so audience can know that Lora did this mostly on Sunday (panel ⑨). Through the sequence from panel ⑤ to ⑨, this template uncovers the facts step by step by attracting the attention, emphasizing certain facets, prompting questions and inducing curiosity.
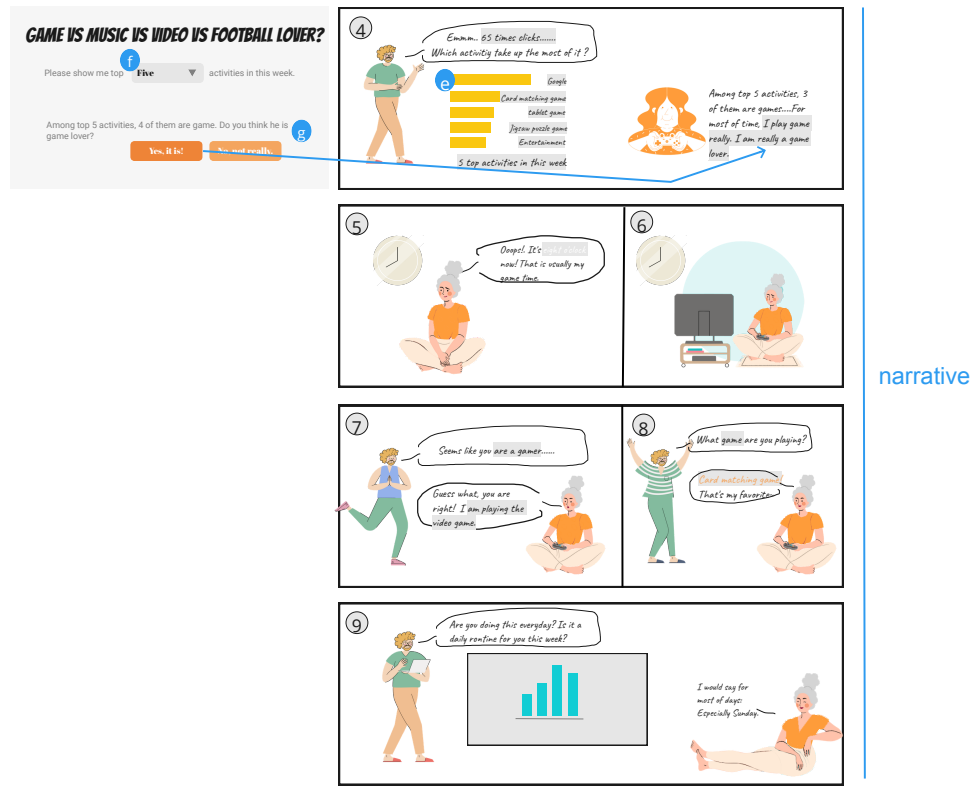
Figure 3.18: A set of panels in narrative is to develop a plot on a selected topic

IV. Ending: The comic is ended with options to involve in exploring routines for other activities. The buttons ⓗ allow users to either close the reading or explore activities other than favorites visited during last week. For example, if audience want to see activities related to health services, the drop-down menu allow selection from all activities during last week, and generative charts ⓘ ⓙ in terms of weekday and daily routines will be shown on the right.
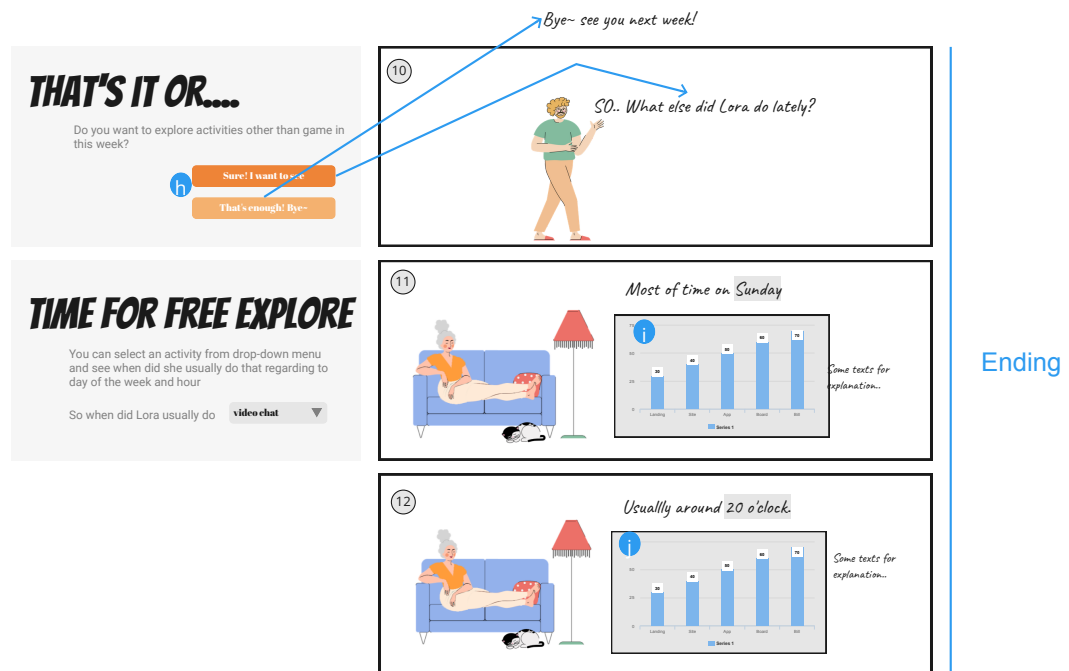
Figure 3.19: The ending opens up a space of free exploration for audiences

Data comic template version 2

The data comic template version 1 is presented to the supervisor, relatives and classmates in the wild. Overall, the features of an overview-and-detail structure and scenario-based narrative are widely accepted for a clear structure and abilities to draw attention to key messages. The feedback suggests two directions: (1) add more context relative information to explain the scene. (2) work on granularity and rhythm affected by panels.
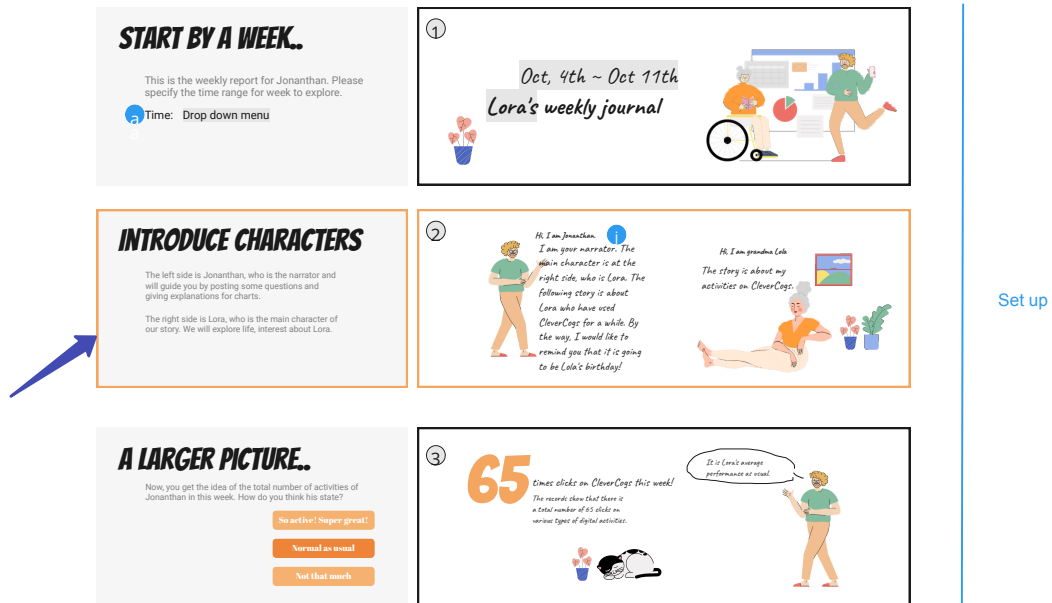
Figure 3.20: Adding context information to introduce characters in template version 2

Regarding to the context information, I include another panel ② in set up (I) to introduce the two characters involved in comic (Figure 3.20) . This gives audience hints on the roles that two characters play before jumping into the story. Moreover, some additional text ⓙ is added to the speech of narrator, such as "it is going to be the Lora's birthday" based on "birth date" from dataset, or other news about what happened to Lora's neighbourhood. Such information creates a context for audience to relate personal activities of their relatives in the real scenario, for example, birthday gift selection inspired from their favorites. Besides, supplementary text are added to explain the scene and actions in panels ⑦ − ⑧ (ⓚ,ⓛ), especially for non-speech environment.
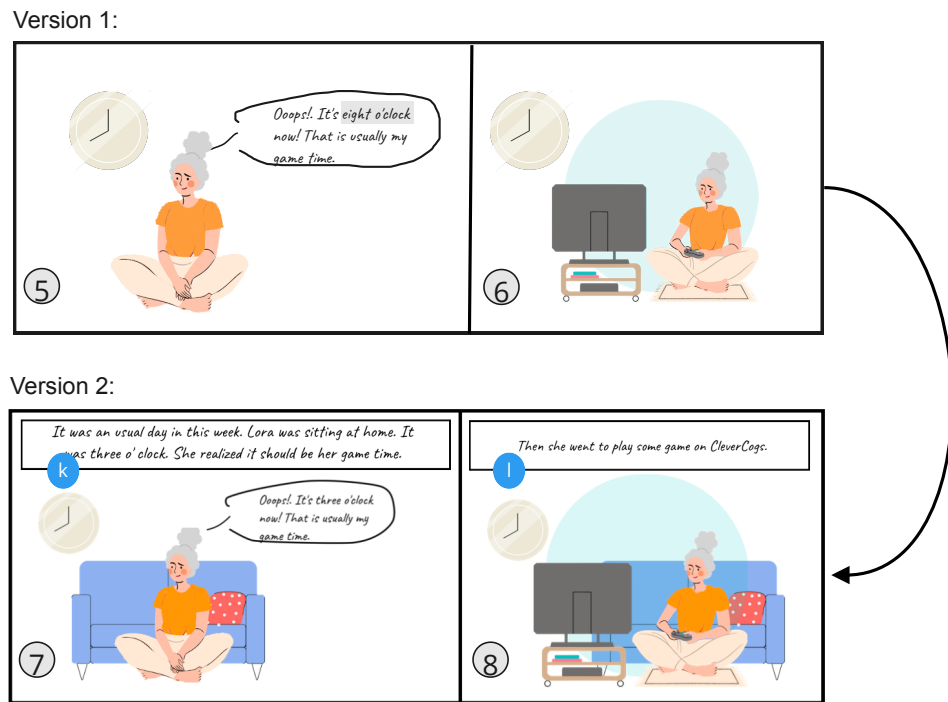
Figure 3.21: The refinement from the panels version 1 to version 2, by adding more supplementary text to explain the non-speech environment.

In terms of Granularity and rhythm, the number of panels affect the pace and flow of the story(Wang, 2022). Originally, there are three messages covered in the larger picture (Figure 3.16) with interactive options and charts. Too much information can sometimes risk of being cluttered. To ensure the comic is easy-to-understand to broader audience group, for example, the children or elderly group in the family, I break down one larger picture panel into two panels to slow down the pace (Figure 3.22). Moreover, I prompt for a reflection and attention of audience on the number of visits per week by asking "how do you think of Lora's state?" It provides an opportunity for audience to make an assumption or guess, then I uncover more insights through frequency distribution charts and comparison charts in next panel ④.
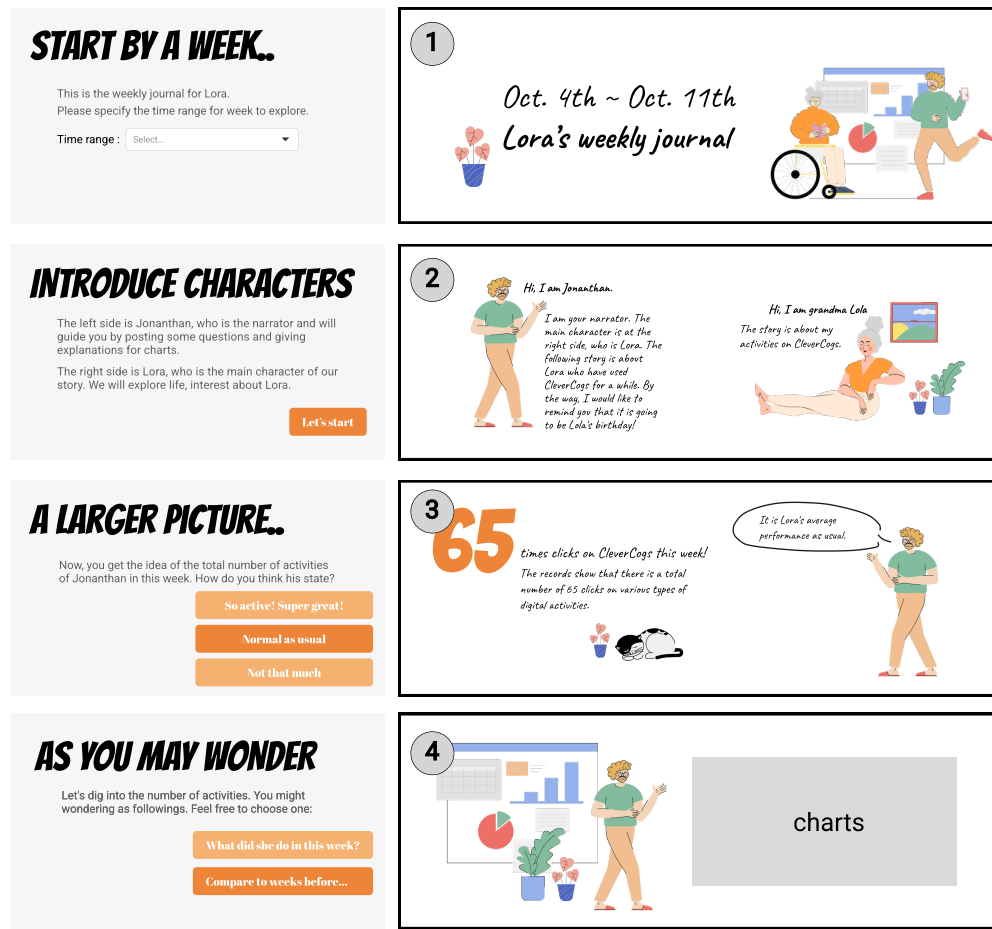
Figure 3.22: Modifications for second version: break down one larger picture into panel 3 and 4.

Data comic template version 3

The second version of data comic template is added with interactive functionalities and transition animation in Figma. Then, the interactive prototype is tested with students from design background. This pilot testing is to examine if there is any unreasonable design demonstrated in terms of structure, transition and overall flow. There are two issues identified which I address in the third version of data comic template.

The first issue is regarded to the topic selection panel placed between the set up (part I) and narrative (part III). The panel ⑤ provides two topics of journey to start the story. However, the topic selection is usually placed at the beginning by convention. Having a topic selection in the middle of the story might confuse audienc (Figure

3.23). It is probably because the the introduction and set up stages are too long as it takes up to four panels (panel ①– ④).

Figure 3.23: Template version 2: The place where to put the topic selection panel is reflected to not conform to the convention.

The moderate amount of information contained in the set up (part I) motivates me to turn the original structure of "overview & detail" into branches of "check the overview(ⓜ)" and "start the story(ⓝ)"(Figure **??**). In this way, the overview is branched out of original linear sequence and serves as an alternative if audience just want to quickly grasp some key information in a short time. The improved audience reading experience is started with an option of "overview" or specific "story", if the "story" is selected, then the audience moves to step of picking a topic(panel ③) to start the story.

Before: template version 2
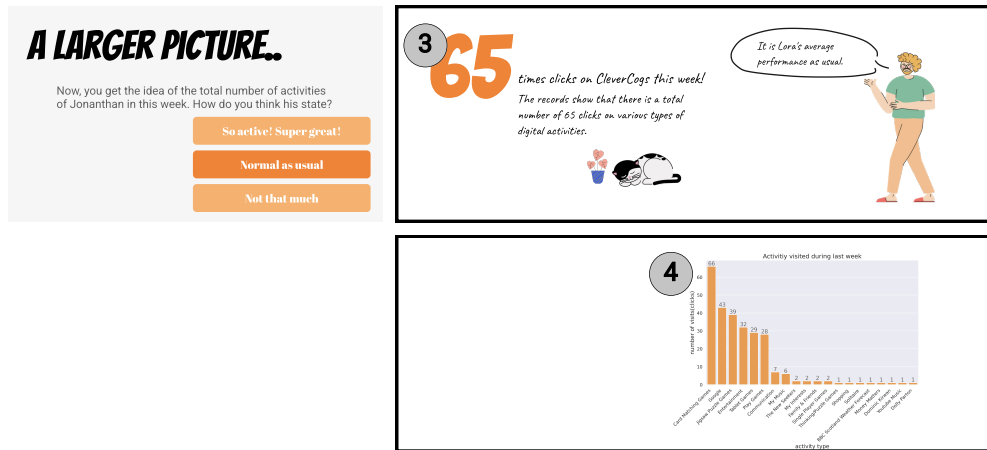
After: template version 3

Figure 3.24: Template version 3: Rearrange the panels so it provides a clear reading strategy.

The second issue is regarded to the interactive commenting in original panel ③. Someone points it to be unreasonable if only the total number of visits are given, yet not seeing the supported charts in panel ④. The relative feedback reports "how can I comment on my mom's state if I only see the number of visits?", or "the comment usually takes place at last by convention, rather than put it at the beginning". Another problem is "how can my point of view become part of speech from Jonathan, who is the male narrator in the data comic", and "is Jonathan represent my perspective in comic?"

Informed by those feedback, I realize that there are three roles involved in the en-

gagement of data comic: the main character Lora, the narrator Jonathan, the audiences themselves. The audience are viewers to witness the conversation between Lora and Jonathan. Thus Jonathan can perform to attract audience attention, but is not able to represent the view of audience. Hence, I turn the generative text into the annotated text on the right side (ⓞ). Besides, I shift the comment step to the end after presenting all charts (ⓟ).

3.8 Final demo: weekly journal

In the end, after been through three iterations, I apply data, charts and messages obtained in Stage 2 and Stage 3, regarding to personal activities of "Lora"(a pseudonym given to a user selected from source dataset), onto the data comic template and finalize it in terms of composition, layout, text, graphics and interactions. The interactive demo of this use case is called weekly journal, reporting personal activities of Lora from 25/10/2021 to 31/10/2021 (Figure 3.25).
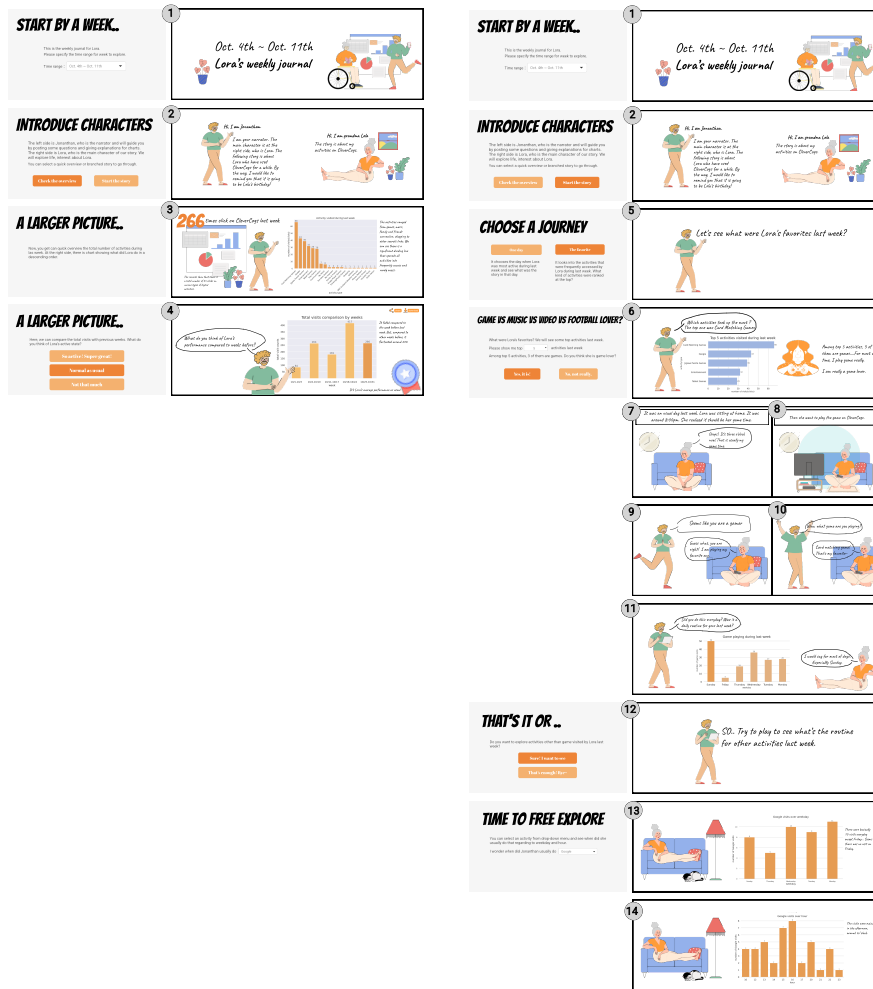
Figure 3.25: The final demo: a weekly journal based on data of Lora and showcases her favorites for narrative as an example

I.Set up: Audiences start by specifying the time range in Panel ①. Then. in panel ②, Lora and Jonanthan are introduced to audience, followed by options of "check the overview" or "start the story".

II.Overview: The "check the overview" brings audience into panel ③ – ④. Three messages regarding to number of visits, what did Lora do, and comparison with previous weeks are shown in a sequence manner. First, the number of visits of last week is highlighted, accompanied with frequency distribution of all activities done by Lora last week. Then, a close-up comparison among weeks are provided, inviting audience to draw their own conclusion through buttons and corresponding generative text will be displayed at lower right corner. Finally, the generative comic overview allow audiences

to download or share it within family.

III.Narrative: The "start the story" in panel ② bring users into another sequence of panels ⑤ – ⑪. It starts by selecting a topic. Each topic defines a structure with series of panels to develop the plot. The demo showcase one topic example of "The favorite", which encompasses a series of panels ⑥ – ⑪ to support narrative about favorite activities. The narrative structure begins with panel ⑥ to investigate answer of "what was the favorite for last week". After the favorite type is drawn from the ranked bar charts of "top activities", the panel ⑦ – ⑩ craft the presence by placing the conversation into the context and environment of doing the favorite thing. The narrative employs in a reverse manner that presenting what game the Lora was playing at the end (panel ⑩)and introducing what time Lora usually play at the beginning. At last, a follow-up question to investigate what day of the week does Lora play is presented in panel ⑪.

IV. Ending: The ending is presented as a close of story that allow audience to download or share the story, or open up a space to explore "what day of the week" and "when" regarding to audience selected activity from the drop-down list.

3.9 Feedback & Result

Since reaching to target audience group of this data comic template is not permitted by Blackwood, I present the demo to some students in University of Edinburgh, and invite them to play with weekly journal demo through generating the data comic page of Lora. This is a rather informal test to communicate the final presentable for some feedback and suggestions. I first introduce the CleverCogs. Then, I invite them to imagine if their grandparents are using such product, and this weekly journal is provided as a medium to communicate the personal activities of their grandparents. Finally, I ask them some freely open questions(Q1 – Q5) to gain some subjective feedback after they have experienced the journey of generating data comic. The results further inform my discussion in next chapter.

Q1: What messages do you get from the weekly journal? Q1: All audiences can easily identify the key messages in the data comic, including key facts such as the total number of visits in one week, comparison of number of visits among previous week. They also understand that the Lora was a game lover as well as the favorite game for her. The game playing routines in terms of "what day of the week Lora usually play game" is quite obvious from the chart and speech of narrators. In contrast, someone

reports that the message of "Lora mostly play game around 8:00pm" is implicit and require further interpretation on what Lora has said.

Q2: How can these messages help you to understand your relatives or inform the actions? Most of audiences report that they feel Lora is pretty cool if that is the grandparent and is surprised by the "fashion" game Lora is playing. Someone feel she is lovely to play such game:"in that case, I can play with my grandma when I visit her. I can even introduce more game to her." Someone notices the significant visit times of game on Sunday, followed by a sharp decrease on Monday: "Lora might feel boring or nobody visited her on Sunday. It will be interested to look at if there is some correlation with the visits of family and friends video call".

Q3: What else messages are you expected? The common feedback is that they care about how they can support their relative. "If that is my grandma, I would love to see what problem she might have encountered recently", commented by one audience. There are other voices: "I care about her physical and mental health, and let me know if she is not feeling well recently". "I would like to see if my grandma is in need of something, or want to learn something, so I can help her".

Q4: How do you feel about the way to deliver messages through data comic? Most of audiences reflect that it is their first time to read the data comic. They generally feel novel about this form and prefer it compared to the single chart: "sometimes I do not know what to look at or what to look for if simply presenting me the charts, but I can easily take away messages from what the character is saying. It is quite straightforward, as they point out directly on what to look at", reflected by one audience. Some audiences are fond of how they can play a role in the comic creation by providing their thoughts through interaction. The techniques used to create the narrative and presence is favoured by some audiences: "I love the way it uncover the truth in narrating the time in the scene and the favorite game". However, there are others reflecting that: "the messages are simple. It wastes too much space and involves so much elements to convey messages which can be done through several short sentences".

Q5: Would you like to receive such comic page once a week to know your relative's life? Most audiences reflect that they would like to receive the updated comic but expect it could be different in story. "Like the comic series", someone describes. "It is a fresh new experience to generate such comic because at the beginning I do not know what would be generated. But after having done this, I know how this plot develop. The similar plots and characters might bored me in a long run", someone reflects.

In summary, the feedback from audiences suggest some benefits (R1–R3) and lim-

itations (R4,R5) of the design outcome summarized into the following points:

R1: The weekly journal take advantages of data comic by imposing a strong reading sequence, enabling take-away messages to be broken down through successive panels. The messages can easily be captured by the speech of characters. Compared to simply visualization charts, the questions posed by narrators address the attention to the intended messages.

R2: The engagement is enhanced through crafting the presence. It is achieved by the use of "audience surrogate" or "proxy" in storytelling, and conversation to address the scene, environment and atmosphere. Besides, the interactive buttons to support audience involvement is similar to the effect of "breaking the forth wall" in plays, which creates a sense of intimacy.

R3: Use the existing visual library of characters save time and efforts while providing variations in graphic appearance. The comic-style appearance arise the audience's interest and motivate them to read. However, to promote the variations on the template, visual assets library specially for generative data comics are demanded.

R4: The amount of deliverable messages are limited. It is partially due to: (1) The features provided by source data, in this case, is largely about online activities, which can not reflect physical actions and behavior in life; (2) The selected dimension of time in data analysis, in this case, is weekly basis. I suppose more seasonality patterns can be found on a monthly-based or yearly-based analysis; (3) The template entails generic messages to be examined across all users. It implies some user-specific pattern can hardly be reported in a generative data comic.

R5: Generating data comic through template expose some issues in terms of generalization and customization, which is elaborated in chapter 5.

Chapter 4

Discussion & Conclusion

In this research, to response to research question "how can (potentially remote) family members keep track of weekly life with their relative", I design a template and present an example of application case of template: weekly journal, to demonstrate a generative data comic about weekly life around the topic of "the favorite". The evaluation results in Section 3.8 report the benefits as a result of design considerations in data comics, and limitations in generative data comics through template respectively. This chapter is informed by the design practice and evaluation results, divided into two sections. In section 4.1, I further reflect on the successful aspects of design considerations in this practice, and supplement them with critical rethinking. In section 4.2, I discuss challenges encountered in crafting a template to generate data comics, limitations in terms of generalizability and customizability, and future prospects.

4.1 Data comics design considerations

The design considerations in this design practice involve all elements mentioned in layer of "visual narrative integration" in data comic creation model[2]: layout, pace, granularity, transition and narration guide. Among them, the layout and transition are mostly drawn from techniques introduced in design patterns of data comics[1]. Breaking down messages through panel and reveal the facts gradually are main ideas involved in my design. However, there is no research of influence of narration guide used in data comics. In the weekly journal, the use of "audience surrogate"(Jonanthan) is suggested to create a presence and engagement.

The results from the audiences feedback reconfirm the benefits of data comic to facilitate understanding. There are three design considerations that contribute to this:

(1) Deliver one piece of message in one panel, so each panel has its focus; (2) Provide a high integration of text and image, so that they work closely to convey intended messages; (3) Raise questions regarding to the data visualization charts to encourage an initiative thinking for audiences. Among these three features, the characteristic of text-picture integration has been appreciated in previous research[h1], where data comic was reported to have higher accuracy in understanding compared to infographics and illustrated text through a lab study.

Another significant benefit provided by data comics is the engagement. This feature has been demonstrated in previous research for high rating of engagement, fun, staying attentive, enjoyability compared to powerpoint, illustrated text and infographics[4][3]. I find the engagement is highly appreciated in feedback from audiences, which is in line with the previous research findings. The design considerations that contribute to the engagement are: (1) Comic elements; (2) Storytelling. In terms of comic elements, qualitative study previously reported the comic figure and speech bubble help focus and improve a sense of involvement. The similar results have suggested by the feedback in this research. In terms of storytelling, I apply a structure that similar to martini glass structure [h3], where it starts by an observation or question (e.g. it is around 8:00pm now so what should I do), and then moves into an open space that allow audience to investigate other activities. I do not strictly follow the structure of beginning-middle-end, and use elements of climax, conflict and resolve, which are typically used in traditional storytelling. This is due to the implications of messages. Instead, I aim to promote engagement by creating the presence(e.g. use of "audience surrogate", text like "it's eight'o clock now" to craft an atmosphere of being at site). Besides, the "break the forth wall" technique is used when narrator in the scene speaking directly to audiences by asking about how the audiences think. It addresses the audiences and crafts a sense of intimacy.

Aside the benefits provided by the features of data comics, there are two design considerations identified to influence and introduce some limitations on the outcome: (1) Messages; (2) Graphic appearance.

(1) Messages underlie the core idea and the content of data comics. Feedback from audiences suggest that messages are strongly related to the effectiveness and usability when such generative data comic play an role in certain context. Audiences expect more messages that can possibly inform the actions, which poses challenges to the data analysis stage. On the one hand, different messages can be elicited from data analysis on a monthly basis or on a yearly basis. I find the number of visits for card matching

game demonstrating a drastic fluctuation across the days in a month through a post-hoc analysis. On the other hand, different aspects can be examined with supplementary of domain knowledge on data features (e.g. what can "link title" tell us).

(2) Graphic appearance: The comic style and lovely figures create an appealing effect when audiences read the weekly journal. However, hand-drawn graphics are unable to meet the demands of updated data comics. To call third-party graphic assets bring another challenge in variability of graphics in libraries.

At last, given the benefits and rationale to employ the data comic, I would like to motivate a larger space of reflection: is data comic a good medium to tell? While I appreciate the benefits provided by data comic, it is not a generic solution for all data. In terms of personal activities data such in my case, the data comic will not make sense if the data amount is extremely scare for an inactive user. The narrative features can hardly been built regarding to a lack of messages. In other words, the storytelling techniques leverage the speak of data, but do not speak on behalf of data. Another reflection is about the pros and cons to deliver messages by incorporating narrative. While the benefits of storytelling is largely reviewed in literature, there are some voices reflecting a redundancy of elements by this research feedback. The intended messages in this design practice are brief. Integrating narrative in delivering messages involves non-associated elements such as narrators, environment graphics into the outcome, which might risk of tedious. This is similar to a result has been reported in the lab study about downside of comic in terms of visual repetition[h1]. In summary, there is a trade-off between engagement and redundancy for visual narrative integration in data comics, which should be taken care in future design.

4.2 Generating data comics through template

There is no precedent of generating data comic through templates in previous research and design practices. Hence, this research take the first step by presenting the template with a use case of Lora based on data from CleverCogs. The template in this context is defined as a series of panels with pre-defined sequence and layout. Regarding to each panel, what messages are intended to deliver and in what ways to deliver are specified.

However, the template implies a group of pre-defined message types. For example, in this design of weekly journal, the message types encompasses: (1) high-level messages of engagement level, interest, routine for selected activity; (2) associated facts of number of visits, top activities, frequency distribution for selected activity. They are

both with respect to a week. The message types are generic applied to every user's data, but not necessarily cater for every audience. The feedback from audiences shows that they expect different types of messages in context. It prompts me think about: "can the template be designed in a user-center approach?" While this research is basically data-driven, in which the template design is based on my point-of-view investigation upon data. In the future, is it possible to conduct it from the user research? I presume the design of data comic template should involve the user-centered design considerations, but due to the privacy and protection consideration, I am not able to access the target audience in my case. Hence, the design of template only suggest a possibility, but is not validated through the lab study or an actual use.

The approach of applying the template, in generating the data comics that fits to respective user's data, implies the limitations in (1)generalizability and (2)customizability.

(1) Generalizability: The demo, weekly journal, presented as design outcome is only based on a particular user's data. Can it be applied to other users of CleverCogs? The answer is yes given that the data features and format are the same. It is because the selected messages reported through this template are generic sets that do not include user's specific patterns. However, the template presented in this research can only report the personal activities in terms of "the favorite" topic. It motivate such more topics to be defined in the future. Hence, this research has not provided a generic solution to report all personal activities, but it offers some experience and reference for researchers and practitioners to design more variations that cover other life aspects. Furthermore, this research intends to motivate a larger space of generative data comics in other scenarios that involve updated dataset: e.g. periodically report on health data, financial data, etc.

(2) Customizability: Informed by the feedback Q5 reported in section 3.8, involving audiences in the generation process is a appealing experience for the first time, but not necessarily in a long-term use. The key is related to what level of flexiability the data comic template can provide. In this research, I do not provide options for audiences in terms of design considerations, but only interactive elements that address audience attention. In the future, to leverage the audience involvement from generating data comics into creating a variety of data comics with fun, the template can allow more flexibility by enabling audiences to customize the characters, select the messages from list to build the story, opt for a range of layouts with design patterns[](e.g. time-sequence,larger-picture, questionanswer) and free to add context-related graphics. A

powerful system is desired, even integrate techniques drawn from artificial intelligence to enable automatically composition based on users' specified elements.

Chapter 5

Conclusion

To response to "how can (potentially remote) family members keep track of weekly life with their relative", this research works on the use case of Clevercogs, and contributes to a practice of generating data comics through templates. This practice reflects the benefits of using comics and storytelling in communicating with data, and successful in delivering generic data insights applicable to every user, but still in need of further practice to extend the messages and explore context that the generative data comics can play. This research works on particular data but can motivate a larger space of designing such templates in other fields that involves updated communication with data. The further practice can take steps in supplement relative visual assets that enable generative contents, experimenting templates with other narratives and structures, as well as explore templates in terms of customizability that involves audience into the selection of factors that has on effect on the generative outcome.

Bibliography

Amini, F., Henry Riche, N., Lee, B., Hurter, C., & Irani, P. (2015). Understanding data
videos. *Proceedings of the 33rd Annual ACM Conference on Human Factors in
Computing Systems*. https://doi.org/10.1145/2702123.2702431

Amini, F., Riche, N. H., Lee, B., Monroy-Hernandez, A., & Irani, P. (2017). Authoring data-
driven videos with dataclips. *IEEE Transactions on Visualization and Computer
Graphics*, *23*(1), 501–510. https://doi.org/10.1109/tvcg.2016.2598647

Bach, B., Riche, N. H., Carpendale, S., & Pfister, H. (2017). The emerging genre of data
comics. *IEEE Computer Graphics and Applications*, *37*(3), 6–13. https://doi.org/
10.1109/mcg.2017.33

Bach, B., Wang, Z., Farinella, M., Murray-Rust, D., & Riche, N. H. (2018, April 21). *Design
patterns for data comics*. Www.research.ed.ac.uk; ACM. https://doi.org/
10.1145/3173574.3173612

Blush: Illustrations for everyone. (n.d.). Blush.design. Retrieved August 26, 2022, from
https://blush.design/en

Burma, J. S., Lapointe, A. P., Soroush, A., Oni, I. K., Smirl, J. D., & Dunn, J. F. (2022).
Letter to the editor: Sampling rate and heart rate variability: On metrics and health
outcomes. *Journal of Biomedical Informatics*, *131*, 104115. https://doi.org/10.1016/
j.jbi.2022.104115

Choe, E. K., Lee, B., & schraefel. (2015). Characterizing visualization insights from
quantified selfers' personal data presentations. *IEEE Computer Graphics and
Applications*, *35*(4), 28–37. https://doi.org/10.1109/mcg.2015.51

Fairclough, S. H., & Dobbins, C. (2020). Personal informatics and negative emotions during
commuter driving: Effects of data visualization on cardiovascular reactivity & mood.

International Journal of Human-Computer Studies, *144*, 102499. https://doi.org/
    10.1016/j.ijhcs.2020.102499

Freytag, G. (1904). *Technique of the drama; an exposition of dramatic composition and art*.
    Scott Foresman and Co.

Friendly, M. (2008). A brief history of data visualization. *Springer Handbooks
    Comp.Statistics*, 15–56. https://doi.org/10.1007/978-3-540-33037-0_2

Gershon, N., & Page, W. (2001). What storytelling can do for information visualization.
    Communications of the ACM, *44*(8), 31–37. https://doi.org/10.1145/381641.381653

Hasan, M. T., Wolff, A., Knutas, A., Pässilä, A., & Kantola, L. (2022). Playing games through
    interactive data comics to explore water quality in a lake. *CHI Conference on Human
    Factors in Computing Systems Extended Abstracts*. https://doi.org/
    10.1145/3491101.3503561

Hermsen, S., Frost, J., Renes, R. J., & Kerkhof, P. (2016). Using feedback through digital
    technology to disrupt and change habitual behavior: A critical review of current
    literature. *Computers in Human Behavior*, *57*, 61–74. https://doi.org/10.1016/
    j.chb.2015.12.023

Hoefer, M. J. D., Schumacher, B. E., & Voida, S. (2022). Personal dream informatics: A self-
    information systems model of dream engagement. *CHI Conference on Human
    Factors in Computing Systems*. https://doi.org/10.1145/3491102.3517669

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science &
    Engineering*, *9*(3), 90–95. https://doi.org/10.1109/mcse.2007.55

Janzen, I. F., & McGrenere, J. (2022). Reflective spring cleaning: Using personal informatics
    to support infrequent notification personalization. *CHI Conference on Human Factors
    in Computing Systems*. https://doi.org/10.1145/3491102.3517493

Kang, D., Ho, T., Marquardt, N., Mutlu, B., & Bianchi, A. (2021). ToonNote: Improving

communication in computational notebooks using interactive data comics. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3411764.3445434

Kauffman, L. (2021). *Leni Kauffman*. Lenikauffman.com. https://www.lenikauffman.com/

Kim, N. W., Henry Riche, N., Bach, B., Xu, G., Brehmer, M., Hinckley, K., Pahud, M., Xia, H., McGuffin, M. J., & Pfister, H. (2019). DataToon. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3290605.3300335

Kirk, A. (2016). *Data visualisation : a handbook for data driven design*. Sage Publications Ltd.

Kosara, R., & Mackinlay, J. (2013). Storytelling: The Next Step for Visualization. *Computer*, *46*(5), 44–50. https://doi.org/10.1109/mc.2013.36

Mathisen, A., Horak, T., Klokmose, C. N., Grønbæk, K., & Elmqvist, N. (2019). InsideInsights: Integrating data-driven reporting in collaborative visual analytics. *Computer Graphics Forum*, *38*(3), 649–661. https://doi.org/10.1111/cgf.13717

Matplotlib. (2012). *Matplotlib: Python plotting — matplotlib 3.1.1 documentation*. Matplotlib.org. https://matplotlib.org/

Matthew, Shadare, Adebowale E, Musa, S. M., & Akujuobi, Cajetan M. (2016). DATA VISUALIZATION. *International Journal of Engineering Research and Advanced Technology (Ijerat) (E-ISSN 2454-6135) DOI: 10.31695/IJERAT*, *2*, 11–16. https://ijerat.com/index.php/ijerat/article/view/191

McKinney, W. (2010). Data structures for statistical computing in python. *Proceedings of the 9th Python in Science Conference*. https://doi.org/10.25080/majora-92bf1922-00a

Morgenthaler, S. (2009). Exploratory data analysis. Wiley Interdisciplinary Reviews: Computational Statistics, 1(1), 33-44

Murnane, E. L., Walker, T. G., Tench, B., Voida, S., & Snyder, J. (2018). Personal informatics
in interpersonal contexts. *Proceedings of the ACM on Human-Computer Interaction*,
2(CSCW), 1–27. https://doi.org/10.1145/3274396

Özdal, Ö., & Çatak, G. (2021). Breaking the fourth wall in video games: A new terminology
and methodology. *International Series on Computer Entertainment and Media
Technology*, 45–59. https://doi.org/10.1007/978-3-030-81538-7_3

Pousman, Z., Stasko, J., & Mateas, M. (2007). Casual information visualization: Depictions
of data in everyday life. *IEEE Transactions on Visualization and Computer Graphics*,
13(6), 1145–1152. https://doi.org/10.1109/tvcg.2007.70541

Rahman, A. A., Adamu, Y. B., & Harun, P. (2017). Review on dashboard application from
managerial perspective. *2017 International Conference on Research and Innovation
in Information Systems (ICRIIS)*. https://doi.org/10.1109/icriis.2017.8002461

Ruckenstein, M. (2014). Visualized and Interacted Life: Personal Analytics and Engagements
with Data Doubles. *Societies*, *4*(1), 68–84. https://doi.org/10.3390/soc4010068

Salas, S. (n.d.). *The Little Things by Susana Salas - Illustrations on Blush*. Blush.design.
Retrieved August 26, 2022, from https://blush.design/collections/
ZSZYg3gPA3kfEIibGw1I/little-things

Sarikaya, A., Correll, M., Bartram, L., Tory, M., & Fisher, D. (2019). What do we talk about
when we talk about dashboards? *IEEE Transactions on Visualization and Computer
Graphics*, *25*(1), 682–692. https://doi.org/10.1109/tvcg.2018.2864903

Schneider, H., Schauer, K., Stachl, C., & Butz, A. (2017). Your data, your vis: Personalizing
personal data visualizations. *Human-Computer Interaction – INTERACT 2017*, 374–
392. https://doi.org/10.1007/978-3-319-67687-6_25

Schreiber, A., & Struminski, R. (2017). Tracing personal data using comics. *Universal Access
in Human–Computer Interaction. Design and Development Approaches and Methods*,

444–455. https://doi.org/10.1007/978-3-319-58706-6_36

Scott, C. (n.d.). *CleverCogs - blackwood*. Www.blackwoodgroup.org.uk. Retrieved August

26, 2022, from https://www.blackwoodgroup.org.uk/clevercogs

seaborn. (2012). *Seaborn: Statistical data visualization — seaborn 0.9.0 documentation*.

Pydata.org. https://seaborn.pydata.org/

Segel, E., & Heer, J. (2010). Narrative visualization: Telling stories with data. *IEEE*

Transactions on Visualization and Computer Graphics, *16*(6), 1139–1148. https://

doi.org/10.1109/tvcg.2010.179

Shin, D.-H., & Biocca, F. (2017). Health experience model of personal informatics: The case

of a quantified self. *Computers in Human Behavior*, *69*, 62–74. https://doi.org/

10.1016/j.chb.2016.12.019

Shu, X., Wu, A., Tang, J., Bach, B., Wu, Y., & Qu, H. (2021). What Makes a Data-GIF

Understandable? *IEEE Transactions on Visualization and Computer Graphics*, *27*(2),

1492–1502. https://doi.org/10.1109/tvcg.2020.3030396

Tableau. (2019). *Tableau: Business intelligence and analytics software*. Tableau Software.

https://www.tableau.com/en-gb

Using google charts. (n.d.). Google Developers. https://developers.google.com/chart/

interactive/docs

Wang, Z. (2022). Creating data comics for data-driven storytelling. *Era.ed.ac.uk*. https://

doi.org/10.7488/era/2047

Wang, Z., Dingwall, H., & Bach, B. (2019). Teaching data visualization and storytelling with

data comic workshops. *Extended Abstracts of the 2019 CHI Conference on Human*

Factors in Computing Systems. https://doi.org/10.1145/3290607.3299043

Wang, Z., Ritchie, J., Zhou, J., Chevalier, F., & Bach, B. (2021). Data Comics for Reporting

Controlled User Studies in Human-Computer Interaction. IEEE transactions on

visualization and computer graphics, 27(2), 967–977. https://doi.org/10.1109/
TVCG.2020.3030433

Wang, Z., Romat, H., Chevalier, F., Riche, N. H., Murray-Rust, D., & Bach, B. (2021).
Interactive data comics. *IEEE Transactions on Visualization and Computer Graphics*,
1–1. https://doi.org/10.1109/tvcg.2021.3114849

Wang, Z., Wang, S., Farinella, M., Murray-Rust, D., Henry Riche, N., & Bach, B. (2019).
Comparing effectiveness and engagement of data comics and infographics.
Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems.
https://doi.org/10.1145/3290605.3300483

Zhao, Z., & Elmqvist, N. (2022). The stories we tell about data: Media types for data-driven
storytelling. *ArXiv:2202.00047 [Cs]*. https://arxiv.org/abs/2202.00047

Zhao, Z., Marr, R., & Elmqvist, N. (2015). *Data comics : Sequential art for data-driven
storytelling*. Www.semanticscholar.org. https://www.semanticscholar.org/paper/Data-
Comics-%3A-Sequential-Art-for-Data-Driven-Zhao-Marr/
43f6a7a70a9cc3dfdaec99f0c240a04830191827