



Comparing Bar Chart Authoring with Microsoft Excel and Tangible Tiles

Tiffany Wun, Jennifer Payne, Samuel Huron, Sheelagh Carpendale

► To cite this version:

Tiffany Wun, Jennifer Payne, Samuel Huron, Sheelagh Carpendale. Comparing Bar Chart Authoring with Microsoft Excel and Tangible Tiles. Computer Graphics Forum, 2016, Computer Graphics Forum, 35 (3), pp.111 - 120. 10.1111/cgf.12887 . hal-01400906

HAL Id: hal-01400906

<https://imt.hal.science/hal-01400906>

Submitted on 10 Oct 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Comparing Bar Chart Authoring with Microsoft Excel and Tangible Tiles

Tiffany Wun¹, Jennifer Payne¹, Samuel Huron^{1,2}, and Sheelagh Carpendale¹

¹University of Calgary, Canada

²I3-SES, CNRS, Télécom ParisTech, Université Paris-Saclay, 75013, Paris, France

Abstract

Providing tools that make visualization authoring accessible to visualization non-experts is a major research challenge. Currently the most common approach to generating a visualization is to use software that quickly and automatically produces visualizations based on templates. However, it has recently been suggested that constructing a visualization with tangible tiles may be a more accessible method, especially for people without visualization expertise. There is still much to be learned about the differences between these two visualization authoring practices. To better understand how people author visualizations in these two conditions, we ran a qualitative study comparing the use of software to the use of tangible tiles, for the creation of bar charts. Close observation of authoring activities showed how each of the following varied according to the tool used: 1) sequences of action; 2) distribution of time spent on different aspects of the InfoVis pipeline; 3) pipeline task separation; and 4) freedom to manipulate visual variables. From these observations, we discuss the implications of the variations in activity sequences, noting tool design considerations and pointing to future research questions.

Categories and Subject Descriptors (according to ACM CCS): H.5.2 [Computer Graphics]: Information Interfaces and Presentation—User Interfaces

1. Introduction

The use of information visualization (InfoVis) is becoming more common in everyday life. Today, we see visualizations such as bar charts and network diagrams on television [HVF13, Dan08, DJ], in newspapers [SH10], and on websites [PSM07]. Despite the prevalence of information visualization, research suggests it is difficult for novices to author their own visualizations [GTS10]. Creating accessible visualization authoring tools has been recognized as a major challenge [LIRC12, HCL05, MJM*06] in visualization research. To inform the design of such tools, we need to better understand how people create a visual mapping for a data set. “Applying a visual mapping” is defined as assigning visual variables [Ber77, CM97] to parameters of a dataset. Different tools are used in this task. Visual mappings can be done manually – for example, by sketching on a napkin [CMvdP10], or with scissors and paper [DHM95]. They can be built with tangible representations [HJC14], coded in a programming language [BOH11], or made automatically, using a software that provides ready-made templates [Mic, Goo, Taba]. Each tool implies a different design paradigm and process. Currently, the most popular tool for generating visualizations from templates is likely Microsoft Excel [Mic]. However, not only does Excel limit the possible visualizations to those preprogrammed as templates, but previous work

has indicated that using Excel templates can be misleading for novices [Su08]. Other research has shown that people do not seem to encounter such limitations when authoring visualizations with tangible tiles [HJC14]—using this method, people who had not studied visualization could author their own representations in a short period of time, without additional help.

Both Excel and tangible tiles can be used to author visualizations, but these tools are different in several ways – Excel is digital and semi-automatic, while tiles are tangible and entirely manual. With this research, we use qualitative methods to uncover differences in the authoring procedures between these tools. We are interested in how people author visualizations using Microsoft Excel, how authoring might be different when using tangible tiles, whether a tool implies a specific process and what these processes might be, and how the tool might impact the activity of visual mapping.

If we want to address the challenge of creating accessible visualization authoring tools, we need to better understand similarities and differences between commonly-used tools, as well as the benefits and limitations of each. We present a qualitative study in which we examine these differences. Our main goals are to investigate: *what* differences exist between these tools during data manipulation and visual mapping tasks, *when* people perform each sub-task during visualization authoring, *how* each tool impacts data manip-

ulation and visual mapping procedures, and *which* freedoms are provided by each of these tools.

To analyze the processes within visualization authoring, we used the InfoVis pipeline as presented by Jansen et al. [JD13]. We reveal differences in both process and freedom of visual variable attribution. Our analysis makes the following contributions:

- an identification of different sequences of actions in visualization authoring using each of the tools;
- a between-tool comparison of the distribution of time spent on each of the tasks of the InfoVis pipeline;
- a schema of the interrelation between stages of the InfoVis pipeline according to tool used;
- a model of the correspondences between the authoring tasks, and the freedom to manipulate visual variables; and
- suggestions, recommendations and implications for future research and design.

2. Motivation & Background

Information visualization authoring, which focuses on mapping between data and a visual encoding [Ber77], is a complex design task that can be performed with different types of tools [GBTS13]. The community has produced important variety of digital tools [VWVH*07, Tabb, Taba] and toolkits [Fek04, BOH11] for authoring visualizations, as well as several variations of model of the process [CMS99, Car99, JD13]. Despite this work, there remains the challenge of providing accessible visualization authoring tools to a wide audience [LIRC12, HCL05, MJM*06]. To provide context for this research, we discuss visualization literacy, template editors, and tangible authoring and construction.

2.1. Visualization Literacy

Boy et al. [BRBF14] designed tests to assess visualization literacy, which they defined as “the ability to use well-established data visualizations (e.g. line graphs) to handle information in an effective, efficient, and confident manner.” The tests are based on item response theory, with three versions – one based on line graphs, one based on bar charts, and a third based on scatter plots. These tests use six tasks: minimum and maximum tasks involve finding the lowest and highest values; variation tasks involve “detecting a trend, similarities or discrepancies”; intersection tasks involve finding a specific data point; average tasks involved estimating the mean, and comparison tasks involved comparing either trends or particular values. We used these tests as they provide efficient means of assessing participant visualization comprehension.

2.2. Template Editor Software

Gammel et al. [GBTS13] reviewed visualization authoring software published within visualization and human-computer interaction venues. They noted six different visualization authoring approaches: visual builders, visualization spreadsheets, textual programming, visual dataflow programming, template editors, and shelf configuration (similar to template editors, but with a greater

number of configurable options). Gammel et al. define the category of *template editors* as software in which a person “select[s] the data to visualize and then pick[s] a visual structure to represent it”. These tools include web services, such as: ManyEyes [VWVH*07], WolframAlpha [Wol], Google Chart Editor [Goo], and Tableau Public [Tabb]; and desktop applications like Open Office [Apa], Tableau [Taba], Spotfire [Spo], and Microsoft Excel [Mic]. This type of software is often composed of two views: a view containing the spreadsheet with the data set, and a view containing the visualization. People enter data into the spreadsheet, select it, and generate a visualization by selecting a predefined visual mapping template from a group of choices.

Template editors pose several challenges. By analyzing forum posts, Chambers et al.’s [CS10] revealed a series of challenges in using Microsoft Excel for authoring visualizations: a high frequency of problems in creating novel charts, i.e. charts not supported directly; difficulties in mapping goals to multiple features; and difficulties with finding features, i.e. finding specific functionality through the Microsoft Excel interface. Other work examining use of Excel suggests that certain features of the default chart types limit peoples’ ability to comprehend data [Su08].

Gammel et al. [GTS10] also studied how information visualization novices author visualizations. Their participants did not interact directly with template editor software, but rather with an expert operator using Tableau; participants had a task sheet containing the data attributes, possible operations and a set of visualization samples. In authoring a visualization, participants encountered three major barriers: (i) the selection of data attributes to explore; (ii) the design of visual mappings; and (iii) the interpretation of the visualizations. Later, Huron et al. [HCT*14] studied novice visualization authoring, using tangible tiles, and did not observe these same barriers. Though these two studies are not directly comparable, they point to potential differences between the two authoring approaches, motivating us to explore such differences in the present study.

2.3. Tangible Authoring and Construction

Though most visualizations are produced with software tools and distributed via digital media, there are other practices used to generate representations – e.g. sketching on a napkin [CMvdP10], on paper [WHC15], or on a whiteboard [WCR*11]. In addition, there are many diverse tools for creating representations using physical materials, as illustrated by the extensive list of tangible visualizations collected by Dragicevic and Jansen [DJ]. These include: scissors and paper [DHM95], gumballs [DJ], yarn [DJ], LEGO bricks [Hun08], and stones [Rei12].

Recently, Huron et al. [HCT*14] described a common approach to visualization creation, based on the process of construction. Using this method, information visualization novices were able to author a visualization of a simple data set in a short period of time with relative ease [HJC14]. Thus different visualization authoring tools have been created and explored, with varying levels of accessibility. Differences between visualization authoring with template-based tools, like Microsoft Excel, and tangible approaches, like construction using tiles, are not yet clear.

3. Study Design

In this study, we were interested in: *i*) understanding and exploring differences in peoples' actions and sequences of actions using Excel and tiles, *ii*) understanding how people manipulate visual and presentation variables [Ber77,Mac86] when creating bar charts with each tool, and *iii*) exploring the impact of the tool on authoring procedures. We first describe our setup, conditions, datasets and tests used, then outline our procedure.



Figure 1: Tiles in different colors. Two boxes of these tiles were provided to participants.

3.1. Setup

The room held two working spaces, the **tangible table** and the **digital table** (see Fig. 2). The **tangible table** contained a workspace inspired by Huron et al.'s tile-based study [HJC14]. It contained A2 canvas paper, boxes of tiles (3), and a monitor (1) to display questions for the reading test. We provided square tiles with the same colours as the cells within the printed datasets: purple, white, orange, and green (see Figure 1). At the **tangible table**, one camera captured the workspace from above, and the other captured participants' gestures. The **digital table** contained a computer with a monitor and a mouse, screen capturing software, and Microsoft Excel. A camera was set up to capture participants' gestures. Both tables contained a sheet of A4 paper with a printed dataset.

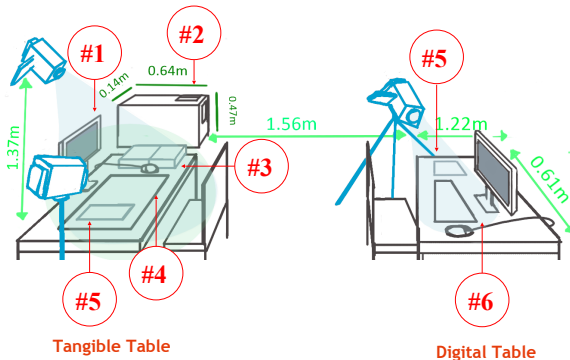


Figure 2: Room setup: 1) computer screen; 2) box to cover canvas during tangible reading test; 3) boxes of tiles; 4) canvas for tiles authoring; 5) dataset; and 6) computer used for Excel.

3.2. Conditions and Tasks

The three tasks performed by participants were authoring, editing and reading. Each task was done in each of two conditions – using Microsoft Excel, and using tangible tiles.

Condition 1 – Microsoft Excel. In the Excel condition, participants sat at the digital table (Fig. 2) and performed the following tasks:

- **Authoring task:** In the authoring task, participants were given a dataset on an A4 sheet (Fig. 2, #5), and asked to “make a bar chart” in Excel.
- **Editing task:** When the participant declared the task complete, they were asked to verify their chart. If something was missing after verification, the experimenter recommended adding the missing element(s).
- **Reading task:** The experimenter removed the dataset sheet from the table, took a screenshot of the chart, and closed the Excel window. The participant's chart was then inserted into the reading test. The participant then completed the reading test.

Condition 2 – Tangible Tiles. In the tiles condition participants were invited to sit in front of the tangible table (Fig. 2) and performed the following tasks:

- **Authoring task:** Participants were given a dataset in tabular format on A4 paper (Fig. 2), and asked to “make a bar chart”, using tiles on the A2 canvas (Fig. 2, #3).
- **Editing task:** After the construction phase, participants were given a pen, and asked to verify and annotate their chart. As in the Excel condition, if something was missing after verification, the experimenter recommended adding the missing element(s).
- **Reading task:** The experimenter removed the dataset sheet from the tangible table. Mimicking the procedure with Excel, the experimenter used a box to hide the visualization as needed during the reading test.

Excel Dataset			Tiles Dataset		
	Planet	Values		Planet	Values
1	Nacre	110	1	Cordina	120
2	Zahir	120	2	Pandora	130
3	Osiris XI	130	3	Hedeby	70
4	La Maetelle	140	4	Nepenthe	90
5	Jorassia	130	5	Banoi	80
6	Gor	130	6	Devon	110
7	Corneria	120	7	Abari	100
8	Tiamat	140	8	Isthmus	90
9	Strogos	150	9	Sodor	90
10	Altair IV	100	10	Gielinor	100

Figure 3: The two datasets for the authoring tasks.

3.3. Tests and Dataset

During the experiment, we used reading tests to assess: *i*) if participants knew what a bar chart was, and if they knew how to read one; *ii*) if they understood the dataset provided; and *iii*) if they could read the visualization they authored.

Primer and Reading tests. Based on Boy et al.'s visualization literacy test [BRBF14], we generated a primer test and a reading test.

The goal of the primer test was to determine if participants were able to detect features in a dataset similar to the one used to author the visualization. The goal of the reading test was to see if participants were able to detect the same data features from their authored charts. These tests contained one question for each data feature. As in the Boy et al. test, participants were given eleven seconds to answer each question.

Datasets. Our two datasets (Fig. 3) were based on those in the Visualization Literacy test [BRBF14], as these datasets were simple to understand, and not associated with potentially familiar real-world data. To simplify data processing for participants, we rounded values to the nearest ten.

3.4. Participants

Seventeen people (12 female) participated in the study. All but one were university students. Twelve participants were studying science, technology, engineering, and mathematics fields, and the remaining five participants were medical and humanities students. They ranged in age from 18 to 33, with a median age of 20. Participants were recruited via word-of-mouth, social media, and posters. They were required to have normal colour vision, and not be visualization experts. 8/17 participants declared that they were at least “somewhat familiar with creating visualizations in Excel”, while none had used tangible tiles before.

3.5. Procedure

Participants filled out a consent form, demographic questionnaire, then completed the bar chart visualization literacy test and primer test. Participants then authored a bar chart in two conditions – using Microsoft Excel or tangible tiles. The order of conditions was counter-balanced; half of participants first created a bar chart using tiles; the other half first created a bar chart using Excel. Next, we asked participants to verify and annotate their chart. Following this, participants completed the reading test and another visualization literacy test. Next, participants completed this sequence of activities in the other condition. After all conditions were completed, participants were invited to share their thoughts in a semi-structured interview. Participants received \$20 for participation.

3.6. Data Collection

To analyze differences between the software and tiles conditions, we gathered several types of data, as detailed below.

- **Video:** We videoed the entire process in both conditions, with the described viewing angles. Three video cameras were used to record participants during the software authoring task, along with screen-capturing software.
- **Questionnaires:** We gathered demographic information, and had questionnaires about impressions after each condition.
- **Visual Literacy Tests:** We collected responses to visualization literacy tests to discover to what degree participants were visually literate, and to see if this changed over the course of the study.

Actions	Definition
Loading data	Transferring data from the provided format to the visualization authoring tool.
Data transformation	Changing the data organization or structure, involving such actions as ordering, sorting, aggregating, etc.
Visual mapping	“This transformation gives an initial visual form to the processed data by mapping data entities to visual marks, and data dimensions to visual variables” [JD13]
Presentation mapping	“This transformation turns the abstract visual form into a fully-specified visual presentation” [JD13]
Tool exploration	Manipulating the tool without resulting in a direct transformation to any part of the visual representation.
Requesting help	Asking the experimenter for help.
Coloring the bars	Coloring the bars of the bar chart in Excel.
Correcting errors	When a participant performs an action and then reverses it a few seconds after.
Verifying	Visual indication that the participant is verifying the visualization or the dataset.

Table 1: Actions identified and coded in the video data. The first part of the table is from the InfoVis pipeline, while the second part was identified through open coding.

- **Interview:** At the end of the study, we conducted a semi-structured interview during which we asked participants to elaborate on the use of each tool. We asked them if they perceived any difference in reading and understanding the visualization with each tool, and if they perceived benefits in using one tool or the other, which tool they preferred, and if they learned something during the process.

3.7. Data Analysis Method

We collected 265 minutes of video from the three cameras and screen capture. To analyze the video, we used a qualitative approach based on open coding and axial coding, as described by Creswell et al. [Cre12]. The coding of the video was performed in three passes, using an iterative process and three coders. The first pass involved open coding, allowing us to identify several categories of interest (see Table 1). For the second pass, we decided on codes to compare activities, using terms from the InfoVis pipeline by Jansen et al. [JD13] to relate our coding to previous models of InfoVis authoring.

4. Results

Since we are primarily interested in comparing the process of authoring visualizations in the two conditions – Excel and tiles – our study results focus on process details, which are primarily available from the collected video data.

4.1. Observations on Conditions and Tasks

All participants were able to author and read a bar chart in both conditions. Participants were faster using Excel than using tiles

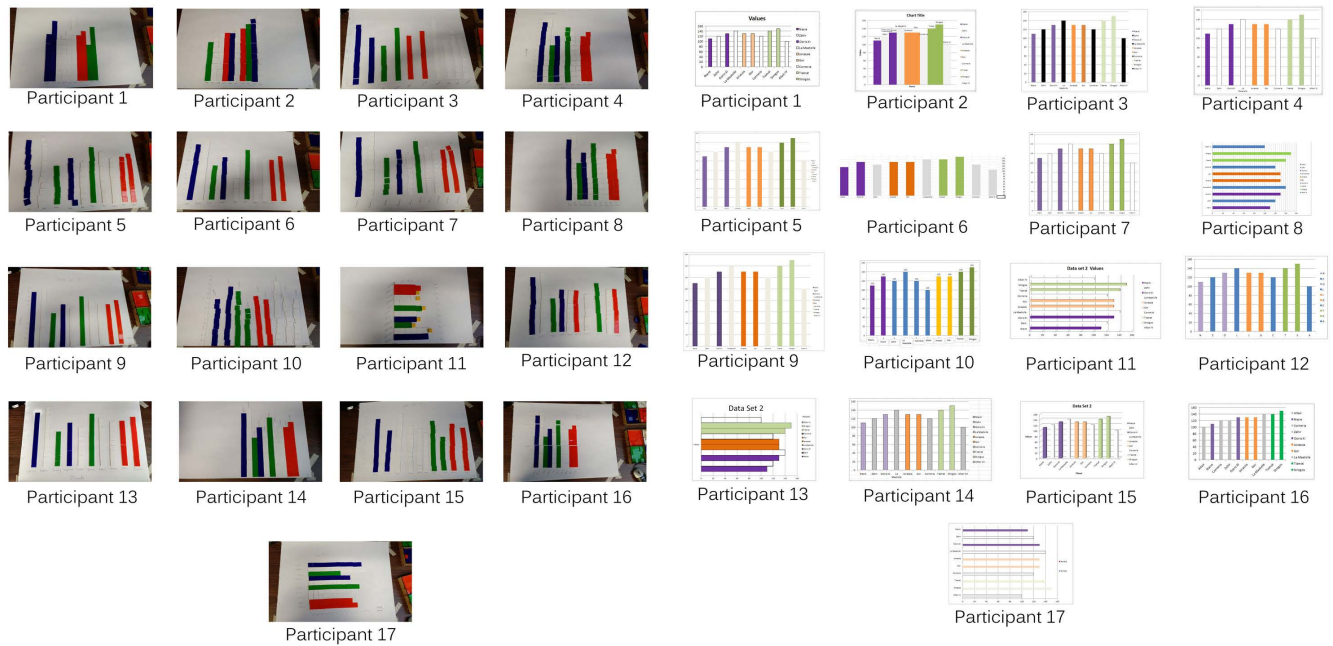


Figure 4: Bar charts authored by participants with the tangible tiles and Microsoft Excel. A larger version is available on our supplementary website at <http://innovis.cpsc.ucalgary.ca/supplemental/Comparing-excel-tiles/>.

(Figure 6). It took approximately eight and half minutes for participants to author and edit a visualization using tiles, and approximately seven minutes with Excel. Figure 5 illustrates the coded visualization authoring processes. Each line represents the sequence of actions for a participant in a particular tool. The top half shows visualization authoring using Excel, and the bottom half using tangible tiles. The highlighted colours show particular actions.

Figure 6 illustrates the time spent by action. While using tiles, participants spent more time loading and transforming the data, performing visual and presentation mappings and verifying the visualization, in comparison to the amount of time they spent on these actions when using Excel. In Excel, participants spent more time exploring the tool and colouring the bars of the bar chart.

All participants scored very well on the initial visualization literacy test (9/17 had an initial perfect score); subsequent score fluctuations were negligible.

4.2. Comparing Common Actions

The actions of loading data, transforming data, creating or changing a visual mapping or presentation mapping occurred both when using Excel and when using tiles; the distribution of time spent on each of these actions is shown in Figure 6. However, these actions were performed differently depending on the tool used. We discuss how each of these actions was performed with each tool below.

Loading Data. Loading data in Excel involved copying the number from the provided paper dataset into the spreadsheet, using the keyboard and mouse. Using tiles, this action corresponded to selecting tiles from the boxes, counting them and transporting them to a position to the canvas. On average, participants spent 97 seconds performing this task with Excel, and 149 seconds performing

it with tiles. With Excel, all participants waited until the dataset was fully loaded before performing any other actions. In the tiles condition, all participants started doing other actions, such as visual mapping and presentation mapping, while they were loading the data.

Transforming Data. In both conditions, the only transformation of data we observed was reordering. These tasks were not suggested to participants, but were done spontaneously. Two participants reordered with Excel, while six reordered with tiles. To reorder data in Excel, one can use spreadsheet features that automatically sort the data according to the selected specification. However, none of our participants used such features. One processed the reordering manually on the spreadsheet, using copy and paste, and the other did this task mentally prior to first typing the data into the spreadsheet. These two participants sorted their charts by colour, and by value, respectively. With tiles, all the participants used spatial rearrangement of tiles to reorder the data. Two ordered the bars by colour alone, one by value alone, two by both colour and value, and one (P12) participant kept half in their original order, while sorting the second half by value.

Visual Mapping. In Excel, visual mapping was performed by all participants using the same sequence of actions: selecting the cells containing the data in the spreadsheet, opening a menu containing icons of the offered templates, then clicking the icon representing the desired chart. Sometimes a participant spent some time choosing which chart they wanted. Sometimes they generated several charts before deciding on one. On average, participants spent 64 seconds doing these actions in Excel. Using tangible tiles, the process was different. Participants transported tiles to the canvas to load the data, and arranged them on the canvas to complete the loading. Through these actions, they defined the mapping between

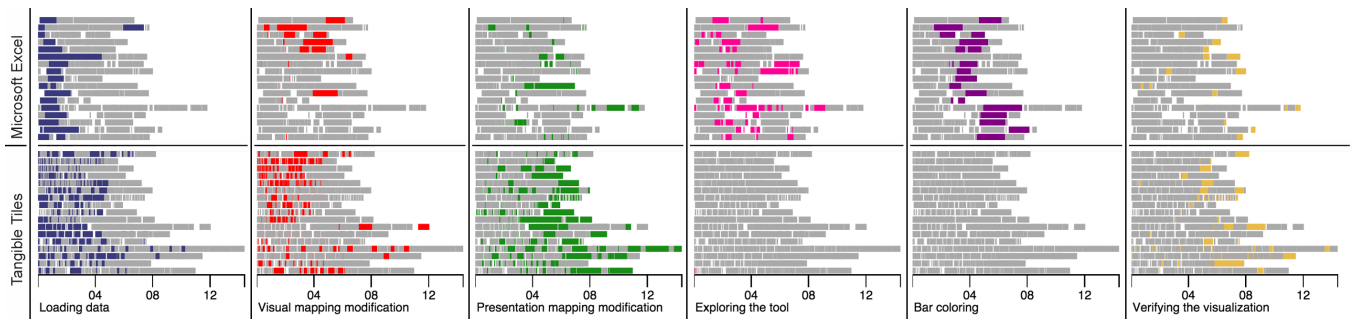


Figure 5: Actions over time for each participant, using Excel and using tiles, for authoring and editing tasks: loading data (■), visual mapping (■), presentation mapping (■), exploring the tool (■), coloring the bar (■) and verifying the visualization (■).

the data dimension and the position of the tiles. For all participants, this was done as they loaded the data, even if some participants redefined the positions later. Participants spent an average of 160 seconds on this process with tiles.

Presentation Mapping. Presentation mapping refers to many different actions, such as labeling, providing a title, positioning the chart, changing the aspect ratio, annotating, and specifying the scale. With Excel, the chart is produced with a set of default presentation elements, such as axes ticks and a legend. Most participants modified these. Some participants used the automatic features to do tasks such as changing the label position for all the bars, while others did this manually. In Excel, participants spent an average of 43 seconds on these types of actions. With the tiles, some of these actions – such as changing the position of the chart or the aspect ratio – were cumbersome or difficult. However, participants did create axes, axes ticks, legends, annotations and labels. With tiles, participants spent an average of 209 seconds on these types of actions.

4.3. Condition 1 – Microsoft Excel.

Throughout this condition, participants followed a general sequence of actions, illustrated in Figure 5. This sequence was generally load data, visual mapping, presentation mapping, colour the bars, sometimes followed by more presentation mapping. At various times during this sequence, we observed participants *exploring the tool*. Only one participant (P6) did not adopt this same sequence of actions, instead using a completely different process. This participant coloured spreadsheet cells to represent the data values, as if they were constructing bars with tiles, but instead using cells as tiles. Seven participants requested help with the tool. Two actions, *exploring the tool*, and *colouring the bars*, were specific to Excel condition and are described below.

Exploring the tool. ‘Exploring the tool’ encompassed activities during which a participant was acting on the Excel interface without the action leading to any result – either on the spreadsheet, or on the visualization. A sample exploration action is to open a menu, go through the sub-menus, and not click anywhere. On average, participants spent 92 seconds on this type of action, and the average number of these actions per participant was four. Such actions were spread out over the authoring time sequence, visible in Figure 5. Only four participants explored the tool at the beginning of the task, before loading the data; other participants explored at various other times during the authoring period.

Colouring the bars. The colour was a necessary and valuable piece of information that provided the grouping of planets. Participants added this information in Excel by colouring the bars of the bar chart. To do so, they selected the bar in the chart by clicking on it, and then opened a menu to choose the correct colour and validate the change. Participants repeated this action for each bar of the chart. The total duration of this action was 92 seconds, on average.

4.4. Condition 2 – Tangible Tiles.

In this condition, all participants constructed a visualization without any requests for assistance specific to the tool. The only questions asked consisted of clarifications of the overall task.

The order of actions while using tiles was more repetitive and cyclical than the authoring process in Excel. Often, actions were of short duration compared to Excel, visible in Figure 5. Participants still followed particular sequences. They loaded a portion of the data, visually mapped it to bars, and repeated this until all data was loaded and mapped. Sometimes, participants did presentation mapping in between actions, while other times, they waited until all bars were generated. Lastly, they verified the visualization.

4.5. Interviews

Participants were asked a series of questions to elicit their thoughts and opinions during the study. While participants enjoyed both two conditions, they generally preferred one or the other: tiles for its simplicity and control over the tool, and Excel for the automation of menial, repetitive tasks. 10/17 participants thought that building their chart in Excel was easy, and 13/17 stated using the tangible tiles was easy. 9/17 participants answered that authoring their chart in Excel helped them to understand their chart better, while 7/17 participants thought that authoring with the tiles helped them to understand their chart better. 7/17 participants said that reordering data based on values or colors made it easier for them to read their chart. Similar to Huron et al.’s earlier study with tangible tiles [HJC14], participants were asked “*What did you manipulate during the process of making your visualization?*” in both questionnaires. In Excel, “*the coloring of the bars*” was the most popular answer, while with the tangible tiles, “*tiles/blocks*” was the most popular answer.

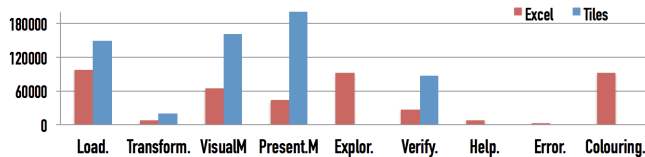


Figure 6: Distribution of time spent in average by all participants by conditions. Vertical axis represents the time in milliseconds, horizontal axis the different participant's actions. Times were computed from analysis of the video codings.

5. Discussion

We discuss findings in terms of the distribution of time across the InfoVis pipeline operations, the interrelation between pipeline operations, and a schema for freedom of visual variable manipulation.

5.1. Distribution of Time Spent on Pipeline Operations

Our results illustrate clear differences in the order and sequence of tasks (see Figure 5). Using Excel, participants tended to follow a linear pattern, while using tiles, participants tended to follow a cyclical pattern that repeated for each bar within the bar chart. Using tiles, the majority of participants spent more time on each of the pipeline operations than they did on these operations in Excel: loading data (+52 seconds), transforming data (+12 seconds), visual mapping (+96 seconds), and presentation mapping operations (+166 seconds). Using Excel, participants spent more time exploring how the tool works (+91 seconds), and on colouring bars of the bar chart (+92 seconds). Some actions, such as colouring the bars, are difficult to position in the InfoVis reference model, and are not an option in the tiles condition. Selecting tiles of a particular colour could be considered part of the 'loading data' action, as it adds new data to the representation. When coded this way, participants were forty seconds quicker at loading coloured data using tiles (149 seconds), than loading the numerical data (97 seconds) and colouring the bars (92 seconds) in Excel. Alternatively, selecting tiles could instead be considered part of visual mapping, as the participant assigns a visual variable to data by choosing tiles of a particular colour, and picking a certain number of tiles. In this case, participants spend approximately the same amount of time performing visual mapping in tiles (160 seconds) as the combined time for visual mapping (64 seconds) and colouring bars in Excel (92 seconds).

The distribution of time across activities in the two conditions suggests that Excel's automation features are effective in reducing the amount of time spent on visual mapping, and default presentation mapping tasks (96 and 166 seconds less than with tiles, respectively). However, doing a simple modification – such as adding a data point and mapping it to the appropriate colour of bar – is very time-consuming. The bar colouring action was particularly time consuming in comparison to other actions in Excel (92 seconds, on average), taking almost as much time as loading data (97 seconds). Bar colouring also took place in a menu which often obscured the view of the chart, removing the action from the representation of the data. Thus, initial generation of the chart with Excel is well-supported, but later editing – including both alterations of the visual mapping and addition of new elements – is cumbersome.

Similarly, the time spent exploring the tool in Excel suggests that people have difficulties finding the features they need, echoing previous work [CS10]. This is evident when participants want to change default visual mapping options, such as the colour of the bars. In Excel, there are fewer visual mapping operations than in tiles, as the chart is created in one action. Tiles do not have automated features, making it necessary for participants to spend more time on visual and presentation mapping tasks. These observations raise questions about what a visual mapping tool should be – should it automate the visual mapping, or should it be a tool that enforces spending time on the visual mapping? This depends on an individual's goal – if they want to generate a bar chart based on a pre-defined template, template editor software solves the problem. If they want to work on transforming the visual representation of their data, it may be useful to use other types of tools.

5.2. Operation Interrelations

Excel provides two different representations of the data: the spreadsheet, which is a symbolic representation (a matrix filled with numbers, letters, and sometimes formulas) and the generated bar chart, which is a visual representation. The two representations allow a clear separation between *processed data* and *visual representation*, two elements of the InfoVis reference model. For instance, loading and transforming data occur within the spreadsheet representation, while modifying aspects of the visual and presentation mapping occurs on the visual representation (i.e. the bar chart). However, some of Excel's features link these two views – for instance, if the value of a cell is changed in the spreadsheet, the chart is updated automatically.

Multiple (linked) representations do not exist in tangible tiles. As a result, when using tiles, some operations in the pipeline are not separable from others. A good example of *interrelated* operations is the combination of actions involved in loading the data. *Loading data* involves transporting a number of tokens (representing a certain value) to the canvas. When doing this loading, the work is being done directly on the visual representation. For this reason, participants must define some of the parameters of the *visual mapping* at this point – for instance, how the tokens are positioned relative to each other. They also have to define *presentation mapping* parameters – for instance, where the token is to be placed on the canvas. These three transformations are separated in the InfoVis reference model. However, in using tiles, they occur simultaneously as part of the task of loading data – it is impossible to perform a data loading action without also defining some aspects of visual and presentation mapping.

These observations allow us to define the *interrelation* principle as follows: *operations are interrelated if one operation in the pipeline specifies – either implicitly or explicitly – some parameters of one or more other operations.* The *interrelation* between operations is a characteristic of the authoring tool used. These inseparable operations are likely to influence the cognitive processes of the author. For instance, while one loads data using tiles, the tool enforces thinking about the spatial organization of visual elements and *visual mapping* – this is not the case with Excel. Varying interrelation of tasks between authoring tools is one reason for the observed differences in sequences of action.

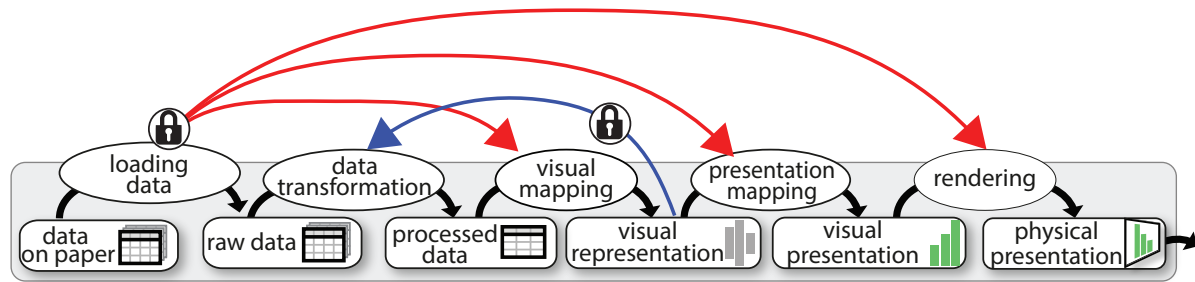


Figure 7: Schema representing interrelations between operations along the first part of the InfoVis pipeline [JD13]. The lock icon represents operations that cannot be separated; red lines represent linked operations with tiles, and blue lines linked operations within Excel.

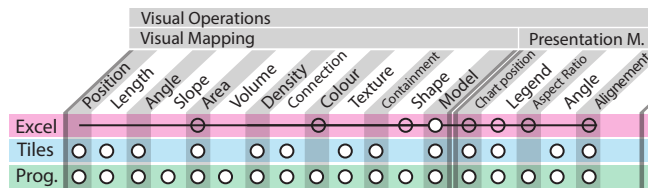


Figure 8: Freedom of the visual variables [Mac86] by tool (Excel, Tiles, Programming). Filled circles represent a variable that can be modified by the tool operator, transparent circles are setup by default according to the template chosen in Excel; the line illustrates the dependencies between variables.

These interrelated operations are beneficial in certain cases. For instance, when a person changes the position of column of tiles on the canvas, by acting on the spatial organization, the person is also reorganizing the underlining data. Having multiple representations also has advantages in allowing one to focus on one aspect of visualization creation at a time, such as changing the colours of individual bars (visual mapping/cell coloring) without replacing the entire bar (as one would have to do in tiles in order to change the colour of a bar).

5.3. Visual variable freedom

Visual mapping is defined by Card et al. as the process of mapping data to visual variables [CMS99]. Visual variables were initially defined by Bertin [Ber77], extended by Mackinlay [Mac86] and later augmented to include physical variables for physical visualization [JDI*15]. Since only one tool in the current study has tangible aspects, we will focus our analysis on visual variables.

Our observations of participant behaviours reveals that visual variable manipulation freedom changes according to the tool used. In Excel, after loading the data, the participant uses menus to select one visualization template from several options. Selecting this template automatically applies a predefined visual and presentation mapping, specifying all the visual variables necessary to display the visualization. After this transformation, the participant can manipulate only a restricted set of visual and presentation variables. The form of the chart is limited by the template list provided by the software, and each template predefines a set of visual variable relations. The participant is only free to modify variables such as the colour of the bars, the stroke, the aspect ratio and other presentation variables. Thus, this tool restricts the participants' choice. These options are shown in the first row of figure 8.

In contrast, with tangible tiles, visual variables are restricted by the types of tiles in use. In the present study, we choose to limit the availability of tiles to one shape, one volume and eight colours. This could be changed by simply adding more tokens types, as requested by one participant. Despite this choice, the tool is still less restricted than Excel in that it allows participants to assign several visual variables. In Figure 8 row 2, we can see that position and length, amongst others, are free variables. The tangible aspect of the tiles does limit some presentation variables – for instance, it would be difficult to change the aspect ratio of the representation.

6. Future Considerations

In this section, we discuss open research questions and implications for design that have arisen from our study.

From our observations, we have distilled several implications for design. It seems important to consider the way people find features in template editor software and to look for ways to make this process more efficient. In addition, since edits to Excel template visual mapping take as much time as generating a chart, it could be useful to consider designing interactions that simplify template visual mapping modifications. Visualization creation software may benefit from exploration and utilization of interrelated actions. For instance, when a participant colours a bar in Excel, they are in effect inputting data, and the software does not replicate this action in spreadsheet view; this leads to a loss of information if the chart view is closed. Creating more congruent relations between views in template editor software could be useful.

Direct manipulation of the visual representation is something we can learn from working with tiles. People generally performed data transformations spontaneously using tiles – they directly manipulated visual elements that represent columns to reorder the dataset. Analogous actions are not possible in Excel – one cannot drag and drop columns in the bar chart to reorder them. Actions like this – in which participants used the visual variables to manipulate and transform the data – could be added to visualization creation software. It might also be possible to go beyond redesigning interactions, to designing a new class of digital tools that allow people to create their own visual representation, while still offering some automation.

This research has also led to several interesting future research questions. Firstly, what parts of the InfoVis pipeline would be useful to automatize, which are best left as manual actions, and in what

circumstances is this different? The interrelation principle enforces different cognitive operations during visualization authoring – how these influence the perception of the data and the visualization is another open research question. Aside from Microsoft Excel and tangible tiles, what are the interrelated operations with others tools, for instance with drawing, or paper and scissors? How can we classify tools according to interrelated actions? Is it beneficial to design tools which allow for certain actions to be either interrelated or separable (or even to switch between modes)? With different tools, different visual and physical variables are free. Can we classify tools according to their types of freedoms? Does the automaticity of visual variable assignment influence people's perception of the resultant visualization?

7. Limitations

In this section, we discuss the limitations in our comparison of Microsoft Excel and tangible tiles as visualization authoring tools. If we aimed to explore the parameters of automaticity, constructive vs. non-constructive, digital vs. tangible individually and separate, more specific studies might have been more appropriate. However, the aim of this investigation was to explore the whole tools, as they are currently used.

Tangible vs. Digital, Constructive vs. Non-Constructive. Each of the tools implies a different authoring paradigm – constructing, for tangible tiles, and using, for Excel. This study cannot be considered a direct comparison of constructive and non-constructive paradigms, as one tool is both constructive and tangible, and the other is non-constructive and purely digital. For similar reasons, this study cannot be considered a comparison of physical and digital visualization design tools. Future work could compare constructive digital tools with other non-constructive digital tools, constructive tangible tools with non-constructive tangible tools, and constructive digital with constructive physical tools, to explore differences in the authoring process that might be attributed to method of creation (constructive vs. non-constructive) and physicality (digital vs. tangible).

Manual vs. Automatic. This study is not a comparison of automatic and manual tools. Excel is a significantly more automated tool than tangible tiles. For this reason, comparison between the two tools in terms of total time taken, as well as time spent on particular actions might be deemed an unfair comparison. Automation and manual actions have different benefits; future work could explore what portions of the visualization design process are best to automate, and when it is best to provide both manual and automatic options.

Type of Chart & Dataset. This study is a non-exhaustive examination of Microsoft Excel and tangible tiles in that the tools are only used to generate one type of visualization, based on only one type of dataset. Though the visualization and the dataset type were chosen because they are common, our results are not necessarily generalizable to the creation of all types of visualizations using these tools, nor to the representation of all types of data. More complex datasets and different types of visualizations might reveal nuances of each tool that were not evident in the present investigation.

Readability. Major differences exist in reading setup between

the tools: the tiles and Excel visualizations have different i) aspect ratios, (ii) sizes, and (iii) orientations. Thus, reading test performance differences could be a result of some or all of these, as opposed to fundamental differences between the tools.

8. Conclusion

In this paper, we present a study examining the differences between two visualization authoring tools – tangible tiles and Microsoft Excel. We describe our observations of people using these tools to author, edit and read bar charts. These observations allow us to qualify *what* differences exist between tools, *when* people performed different actions using each tool, and *how* each tool influences the authoring process. One might expect Excel to impose a certain workflow in the creation of a representation, and tangible tiles to give the visualization creator more freedom in how they go about the design and construction process. Our results offer the first empirical evidence this. We specifically observed how the distribution of time spent, and sequence of actions in different operations from the InfoVis pipeline is different according to the tool used. When using Microsoft Excel to generate visualizations, participants spent more time exploring the tool. When using tangible tiles, participants spent more time on actions directly related to the data, such as visual mapping operations. In addition, our observations indicated that some operations are linked, and that the linked operations are different between Excel and tiles; this operation linkage may influence the cognitive processes of visualization authors.

9. Additional Materials

Additional materials are available online at <http://innovis.cpsc.ucalgary.ca/supplemental/Comparing-excel-tiles/>.

10. Acknowledgements

This research was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC), SMART Technologies, and the Alberta Innovates Technology Futures (AITF).

References

- [Apa] APACHE: Open Office, 2011. URL: <https://www.openoffice.org/>. 2
- [Ber77] BERTIN J.: *La graphique et le traitement graphique de l'information*. Flammarion Paris, 1977. 1, 2, 3, 8
- [BOH11] BOSTOCK M., OGIEVETSKY V., HEER J.: D³ data-driven documents. *Visualization and Computer Graphics, IEEE Transactions on* 17, 12 (2011), 2301–2309. 1, 2
- [BRBF14] BOY J., RENSINK R., BERTINI E., FEKETE J.-D.: A principled way of assessing visualization literacy. *Visualization and Computer Graphics, IEEE Transactions on* 20, 12 (2014), 1963–1972. 2, 3, 4
- [Car99] CARPENDALE M. S. T.: *A framework for elastic presentation space*. PhD thesis, Simon Fraser University, 1999. 2
- [CM97] CARD S. K., MACKINLAY J.: The structure of the information visualization design space. In *Information Visualization, 1997. Proceedings., IEEE Symposium on* (1997), IEEE, pp. 92–99. 1
- [CMS99] CARD S. K., MACKINLAY J. D., SHNEIDERMAN B.: *Readings in information visualization: using vision to think*. Morgan Kaufmann, 1999. 2, 8

- [CMvdP10] CHAO W. O., MUNZNER T., VAN DE PANNE M.: Poster: Rapid pen-centric authoring of improvisational visualizations with napkinvis. *Posters Compendium InfoVis 2* (2010). 1, 2
- [Cre12] CRESWELL J. W.: *Qualitative inquiry and research design: Choosing among five approaches*. Sage, 2012. 4
- [CS10] CHAMBERS C., SCAFFIDI C.: Struggling to excel: A field study of challenges faced by spreadsheet users. In *Visual Languages and Human-Centric Computing (VL/HCC), 2010 IEEE Symposium on* (2010), IEEE, pp. 187–194. 2, 7
- [Dan08] DANZIGER M.: *Information visualization for the people*. PhD thesis, Massachusetts Institute of Technology, 2008. 1
- [DHM95] DOUGLAS S., HUNDHAUSEN C., MCKEOWN D.: Toward empirically-based software visualization languages. In *Visual Languages, Proceedings., 11th IEEE International Symposium on* (1995), IEEE, pp. 342–349. 1, 2
- [DJ] DRAGICEVIC P., JANSEN Y.: List of Physical Visualizations, 2012. URL: <http://dataphys.org/list>. 1, 2
- [Fek04] FEKETE J.-D.: The infovis toolkit. In *Information Visualization, 2004. INFOVIS 2004. IEEE Symposium on* (2004), IEEE, pp. 167–174. 2
- [GBTS13] GRAMMEL L., BENNETT C., TORY M., STOREY M. A.: A Survey of Visualization Construction User Interfaces. In *EuroVis-Short Papers* (2013), The Eurographics Association, pp. 19–23. 2
- [Goo] GOOGLE: Google Chart Editor, 2013. <https://support.google.com/drive/answer/63824>. 1, 2
- [GTS10] GRAMMEL L., TORY M., STOREY M. A.: How information visualization novices construct visualizations. *Visualization and Computer Graphics, IEEE Transactions on* 16, 6 (2010), 943–952. 1, 2
- [HCL05] HEER J., CARD S. K., LANDAY J. A.: Prefuse: a toolkit for interactive information visualization. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (2005), ACM, pp. 421–430. 1, 2
- [HCT*14] HURON S., CARPENDALE S., THUDT A., TANG A., MAUERER M.: Constructive visualization. In *Proceedings of the 2014 conference on Designing interactive systems* (2014), ACM, pp. 433–442. 2
- [HJC14] HURON S., JANSEN Y., CARPENDALE S.: Constructing visual representations: Investigating the use of tangible tokens. *Visualization and Computer Graphics, IEEE Transactions on* 20, 12 (2014), 2102–2111. 1, 2, 3, 6
- [Hun08] HUNGER, MICHAEL: On LEGO Powered Time-Tracking: My Daily Column, 2008. URL: <http://jexp.de/blog/2008/08/on-lego-powered-time-tracking-my-daily-column/>. 2
- [HVF13] HURON S., VUILLEMOT R., FEKETE J.-D.: Bubble-TV: Live visual feedback for social tv broadcast. In *ACM CHI 2013 Workshop: Exploring and enhancing the user experience for television* (2013). 1
- [JD13] JANSEN Y., DRAGICEVIC P.: An interaction model for visualizations beyond the desktop. *Visualization and Computer Graphics, IEEE Transactions on* 19, 12 (2013), 2396–2405. 2, 4, 8
- [JDI*15] JANSEN Y., DRAGICEVIC P., ISENBERG P., ALEXANDER J., KARNIK A., KILDAL J., SUBRAMANIAN S., HORNBÆK K.: Opportunities and challenges for data physicalization. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (2015), ACM, pp. 3227–3236. 8
- [LIRC12] LEE B., ISENBERG P., RICHEL N. H., CARPENDALE S.: Beyond mouse and keyboard: Expanding design considerations for information visualization interactions. *Visualization and Computer Graphics, IEEE Transactions on* 18, 12 (2012), 2689–2698. 1, 2
- [Mac86] MACKINLAY J.: Automating the design of graphical presentations of relational information. *ACM Transactions On Graphics (Tog)* 5, 2 (1986), 110–141. 3, 8
- [Mic] MICROSOFT: Microsoft Excel, 2015. URL: <https://products.office.com/en-CA/excel>. 1, 2
- [MJM*06] MOORHEAD R., JOHNSON C., MUNZNER T., PFISTER H., RHEINGANS P., YOO T. S.: Visualization research challenges: A report summary. *Computing in Science and Engineering* 8, 4 (2006), 66–73. 1, 2
- [PSM07] POUSMAN Z., STASKO J. T., MATEAS M.: Casual information visualization: Depictions of data in everyday life. *Visualization and Computer Graphics, IEEE Transactions on* 13, 6 (2007), 1145–1152. 1
- [Rei12] REININGER H.: Hans Rosling’s shortest TED talk, 2012. URL: <https://www.youtube.com/watch?v=UNs-ziziPyo>. 2
- [SH10] SEGEL E., HEER J.: Narrative visualization: Telling stories with data. *Visualization and Computer Graphics, IEEE Transactions on* 16, 6 (2010), 1139–1148. 1
- [Spo] SPOTFIRE: SpotFire, 2013. URL: <http://www.spotfire.com/>. 2
- [Su08] SU Y.-S.: It’s easy to produce chartjunk using Microsoft Excel 2007 but hard to make good graphs. *Computational Statistics & Data Analysis* 52, 10 (2008), 4594–4601. 1, 2
- [Taba] TABLEAU: Tableau, 2013. URL: <http://www.tableausoftware.com/>. 1, 2
- [Tabb] TABLEAU: Tableau Public, 2014. URL: <http://www.tableausoftware.com/public>. 2
- [VWVH*07] VIEGAS F. B., WATTENBERG M., VAN HAM F., KRISS J., MCKEON M.: Many Eyes: a site for visualization at internet scale. *Visualization and Computer Graphics, IEEE Transactions on* 13, 6 (2007), 1121–1128. 2
- [WCR*11] WALNY J., CARPENDALE S., RICHEL N., VENOLIA G., FAWCETT P.: Visual thinking in action: Visualizations as used on whiteboards. *Visualization and Computer Graphics, IEEE Transactions on* 17, 12 (2011), 2508–2517. 2
- [WHC15] WALNY J., HURON S., CARPENDALE S.: An Exploratory Study of Data Sketching for Visual Representation. *Computer Graphics Forum* (2015), 231–240. 2
- [Wol] WOLFRAM RESEARCH: Wolfram Alpha, 2015. URL: <http://www.wolframalpha.com/>. 2