

THE CREATION OF DATA VISUALIZATIONS WITHOUT PROGRAMMING

by

Drew Skaau

A dissertation proposal submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Computing and Information Systems

Charlotte

2013

Approved by:

Dr. Robert Kosara

Dr. Zachary Wartell

Dr. Celine Latulipe

Christopher Beorkrem

Dr. Mary Lou Maher

©2013
Drew Skaau
ALL RIGHTS RESERVED

ABSTRACT

DREW SKAU. The Creation of Data Visualizations Without Programming. (Under the direction of DR. ROBERT KOSARA)

Visualizations are typically created using a pipeline approach that relies on pre-defined algorithms to map data to visuals. This prevents creativity from entering the process of building data visualizations. My work explores a new model for creating visualizations that involves designing data visualizations with an object- oriented approach. This allows the creation of novel visualizations and visual structures. I plan to explore this space further using prototype tools developed using the model. Evaluation of the tools will focus around creativity support.

TABLE OF CONTENTS

LIST OF FIGURES	vii
LIST OF TABLES	ix
CHAPTER 1: INTRODUCTION	1
CHAPTER 2: RELATED WORK	3
2.1 Theory	3
2.2 Tools	4
2.3 Creativity Support	6
CHAPTER 3: COGNITIVE MOTIVATION	8
3.1 The Information Visualization Model	8
3.2 Cognitive Models	9
CHAPTER 4: VISUALIZATION PRIMITIVES	11
4.1 Prototypes	11
4.2 Visual Properties	12
4.2.1 Internal Geometric Relationships	13
4.3 Iteration	14
4.4 Data	15
4.4.1 Mappings	15
4.4.2 Connections	16
4.5 Axes and Layout	17
CHAPTER 5: EXAMPLES	19
5.1 Bar Chart	20

5.2	Scatterplot	22
5.3	Line Chart	23
5.4	Heat Maps	24
5.5	Waterfall Chart	24
5.6	Glyphs	25
CHAPTER 6: EXPERIMENT		28
6.1	Materials	29
6.2	Procedure	32
6.3	Results	33
6.3.1	Experience and Creativity Support	34
6.3.2	Experience and Productivity	36
6.3.3	General CSI Results	37
6.3.4	Example Visualizations	37
CHAPTER 7: DISCUSSION		38
7.1	Lessons Learned	39
7.2	Interface Improvements	40
7.2.1	Default Values	40
7.2.2	Primitives Graph	41
7.2.3	Connections Between Primitives	42
7.2.4	Set of Primitives	43
7.2.5	Creativity Support	44
CHAPTER 8: FUTURE WORK		45

CHAPTER 9: CONCLUSION	48
REFERENCES	50

LIST OF FIGURES

<p>FIGURE 1: A primitive rectangle in our implementation.</p> <p>FIGURE 2: VIE-VISU is a glyph based visualization for displaying patient visits [13].</p> <p>FIGURE 3: Layout of a bar chart is defined mostly by administrative data.</p> <p>FIGURE 4: The interface for the current implementation of Visualization Primitives.</p> <p>FIGURE 5: A stacked bar chart and grouped bar chart in our implementation. The data in both is from Anscombe's Quartet [3]. Stacked bars (a) and grouped bars (b) differ only in the mapping of their sides. The numbers are added to indicate mappings, in the interface mouseover provides this information. The red highlights indicate the differences in the mappings between the two charts.</p> <p>FIGURE 6: A scatterplot in our implementation using the UCI cars dataset. Miles per gallon is attached to vertical position, weight is attached to horizontal position. Number of cylinders is connected to color. The numbers are added to indicate mappings, in the interface mouseover provides this information.</p> <p>FIGURE 7: A line chart in our implementation is created by turning on a connection between points. The numbers are added to indicate mappings, in the interface mouseover provides this information.</p> <p>FIGURE 8: A heat map of temperature using TMY3 data [33]. The numbers are added to indicate mappings, in the interface mouseover provides this information.</p>	<p style="margin: 0;">2</p> <p style="margin: 0;">17</p> <p style="margin: 0;">18</p> <p style="margin: 0;">20</p> <p style="margin: 0;">21</p> <p style="margin: 0;">22</p> <p style="margin: 0;">24</p> <p style="margin: 0;">25</p>
---	--

FIGURE 9: A waterfall chart showing the lifespans and terms of the presidents of the United States. The numbers are added to indicate mappings, in the interface mouseover provides this information.

26

FIGURE 10: A glyph technique in our implementation using the cars dataset. The technique is similar to the one used in VIE-VISU [13]. The numbers are added to indicate mappings, in the interface mouseover provides this information.

27

FIGURE 11: The bubble chart users were left with after completing the tutorial.

32

FIGURE 12: Histograms of the rankings for each factor (built using the visualization primitives prototype).

34

FIGURE 13: A selection of example visualizations submitted by study participants.

35

LIST OF TABLES

TABLE 1: Factor ranking results.

36

CHAPTER 1: INTRODUCTION

Creating new types of visualization invariably requires programming. Tools like Tableau [26] create visualizations based on data and user input, but are limited to a relatively narrow selection of visualization types. Many designers are trying their hands on visualization frameworks like D3 [7], but are not familiar enough with the programming concepts involved to be very effective. At the same time, the theory of visual representation has not advanced much since the seminal work by Bertin [4] and Mackinlay [17]. I believe that both problems can be solved by a fresh look at the nature of visual representation in visualization.

I propose a new model for the representation of data in information visualization: visualization primitives. Like graphical primitives, visualization primitives are simple geometrical objects that can be combined into more complex ones. The model provides easy access to the visual variables that each primitive object has (Figure 1). The visual variables are similar to Bertin’s retinal variables [4], and are customized per primitive to match the affordances of the primitive. For example, a rectangle visualization primitive would provide access to height, width, left side, right side, top, bottom, horizontal midpoint, vertical midpoint, color, linewidth, etc. In addition to just the graphical component, however, visualization primitives also connect to data and, in turn, produce output data. In the rectangle example, all inputs have corresponding outputs, and outputs are calculated automatically

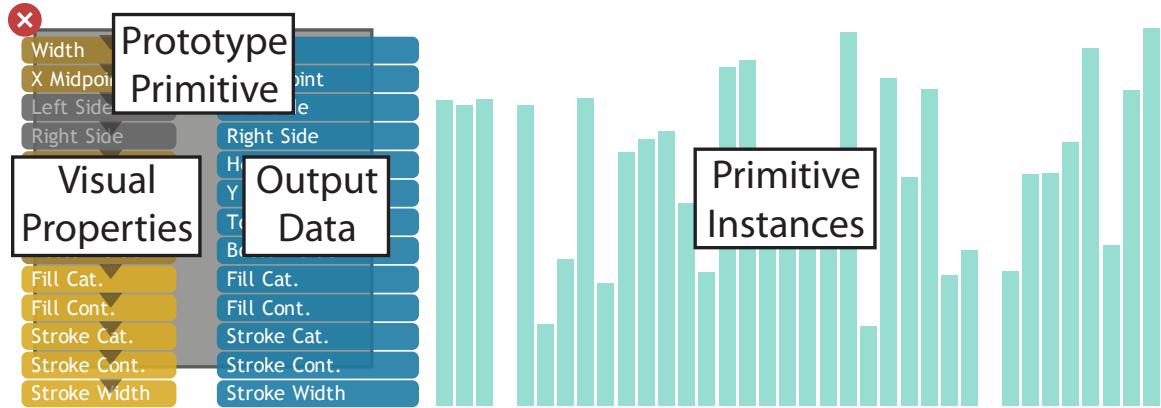


Figure 1: A primitive rectangle in our implementation.

given any input. Automatic iteration through the data gives the user an experience of rapid visualization prototyping with immediate visual feedback. By designing simple prototypes and applying data to them, users can quickly create many different visualizations in a very short time.

The goals of this work are as follows. On the theory side, I want to develop a new model of visual data representation. I believe that current models that are based on pipelines are too limited and do not adequately define the connection between visual appearance and data. On the practical side, I believe that this model will translate into tools that better support creativity in the creation of data visualizations. By producing interactions similar to design tools that work with graphical objects, visualization primitives will allow users to rapidly iterate with immediate feedback. Immediate feedback and direct manipulation are key components of much creativity support software.

CHAPTER 2: RELATED WORK

2.1 Theory

Information visualization is mostly understood as the visualization of discrete data points [18].

By contrast, Scientific visualization typically uses discretely sampled data that represents continuous data. This distinction means that the visual representations can be treated as individuals for each data item. In InfoVis, each data point is represented by a visual *mark*, according to Bertin's highly influential work from the 1960s [4]. Bertin defined marks as graphical objects with visual (or *retinal*) properties such as color, size, or position.

Mackinlay's APT system [17] was directly based on Bertin's work, and the first one to create tailored visualizations for a dataset. APT's goal was the automatic generation of the visualization, however, not specifically supporting the user's creativity. SAGE [23] provided similar functionality, but also primarily oriented at data rather than visual design. Tableau [26], which is partly based on the ideas behind APT, goes in a similar direction, though the user can construct visualizations from scratch. Tableau's expressive power is limited to a relatively small number of plot types, however. We want the user to be able to create entirely new types of visualizations quickly and easily. In order to do this, a system that effectively supports creativity in the creation process may be necessary.

Card and Mackinlay developed a notation for analyzing the visual mappings used in visualizations [8]. Their goal was to connect point designs into a coherent design space,

which then could be explored more effectively. However, their system is mostly analytical, not constructive: changes in the analysis do not translate back into novel visualization designs.

The Grammar of Graphics [32, 34] describes a way of defining almost any visualization imaginable, but from a mathematical and computational perspective that does not necessarily match up with practical visualization creation or best practice. For example, the idea that a pie chart is nothing but a stacked bar chart that has been transformed into polar coordinates may be a correct mathematical explanation, but we argue that few non-technical users would even understand this; much less construct a pie chart this way.

Visualization primitives are an extension of glyphs [2], which are graphical objects whose properties (lengths of parts, angles, colors) represent data. Glyphs are typically used to show small numbers of high-dimensional data points, but we argue that the idea can be applied in a much more general way. Tools for constructing glyphs [22] have been proposed, but tend to be driven by form-based interfaces and limited in the scope of available properties and ways of combining them into more complex (and “traditional”) visualizations. Glyphs are a technique ripe for creativity because of their propensity for semantically meaningful forms, but the established approaches limit creativity by constraining the possible results to a pre-determined set.

2.2 Tools

Visualization toolkits like prefuse [12], Protovis [6], and D3.js [7] put many of Bertin’s and Wilkinson’s ideas into practice, and make it possible to create a wide variety of visualizations, including entirely new ones. While they elegantly abstract away many common

tasks, they do require considerable programming knowledge, however. This is a major hurdle for many potential users of such tools, like journalists and designers.

Other tools that do not require programming, like Quadrigram [5, 20] and DEFOG [16], are still strongly influenced by computing constructs. Building a visualization using Quadrigram’s data flow paradigm requires many steps that are effectively function calls, and are based on typical operations available in programming languages. This approach is conceptually different because it focuses the user on algorithms to reshape the data. These algorithms are modifiable and adaptable, but still result in conceptualizing the visualization as a pipeline from original data to what gets drawn on screen. DEFOG is also driven by data rather than the visual representation; that is a useful property for a data analysis tool, but it does not help designers construct new and interesting visualizations.

Several systems infer visualizations from hand drawn sketches, or abstractly created objects. NapkinVis [10] infers a common visualization type and uses Protopis to draw the resulting charts. CavePainting Visualizations by Keefe et al. [14] shows the value of having non-technical users involved in the visualization creation process, and explores how creativity can improve resulting visualizations. Brett Victor has been developing a system for Drawing Dynamic Visualizations [29] that appears to infer data connections from the structure of what has been drawn. These systems rely on expertise of technical users to create the actual visualizations by embedding it into the system, and by relying on known techniques.

On the other side of the spectrum, there are tools that allow users without programming experience to create visualizations. Many Eyes [30], Excel, Adobe Illustrator, Tableau, all allow users to create data visualizations through simple input of data. Unfortunately, the

visualizations that these tools produce have all been predefined by the tool creator. They each have a set number of visualizations types that they can create. This constrains users from creating new visualizations and possibly embedding domain knowledge or exposing new relationships within the visualization. Adobe Illustrator does technically allow creative freedom, however the process is entirely manual, and certainly does not support any sort of flow state for data visualization creation. These creativity constraints are inherent to the approach each of these tools takes, so a new approach is necessary to provide support for creativity.

2.3 Creativity Support

There have been so many improvements in the computing world Since Alan Kay's early work [?, ?] suggesting computers could teach kids and support their creativity. Computer software we use every day supports creativity in many different disciplines, ranging from music, to literature, to the visual arts. Unfortunately, there has still not been a wide array of work exploring creativity in the creation of information visualizations, however some work in visualization does still involve creativity. In the visual analytics space, it has been noted that creativity plays a role in the visual analytics workflow [24], however this creativity lies more on the analysis and data side than on the visualization creation and development side.

Lee et al. [15] have done some work investigating interaction with visualizations beyond keyboard and mouse. Their design considerations very closely parallel design guidelines for supporting creativity. Other significant work by Walny et al. [31] explores visualization creation on whiteboards. The study never explicitly investigates creativity, but the whiteboard format is relatively open ended, free-form, and collaborative, and the visual-

izations generated during the study are clearly creative. CavePainting [14] intentionally incorporates creative users as a means to explore creative techniques applied to scientific visualization creation. None of this work directly evaluates or explores creativity.

As creativity is a difficult concept to define, quantitatively evaluating creativity is also a difficult task. The best option for evaluating creativity may be the Consensual Assessment Technique (CAT) [1]. The technique requires a panel of experts within the domain to evaluate the work, and rate its creativity. The ratings revolve around how novel and useful the produced work is.

Evaluating a tool's support of creativity is somewhat easier, as it does not depend on the abilities or mental state of the user, only on the capabilities of the tool. The Creativity Support Index (CSI) [9] provides a system for measuring a tool's ability to support creativity. It uses a set of six factors, ranked with Likert scales, followed by a set of fifteen ranked pairs to prioritize the factors. The results of the index are a value out of 100, with 100 supporting creativity perfectly. (For more on the CSI, refer to section 6.1.)

CHAPTER 3: COGNITIVE MOTIVATION

Part of the motivation for this work, both for the theoretical side and the practical uses for visualization designers, is based on the idea of designing visual objects as actual objects (rather than a pipeline). We believe that this approach more closely matches an effective mental model that can benefit both education in information visualization and creativity in creating visualizations.

3.1 The Information Visualization Model

The theory of information visualization is built on discrete data items. Bertin's marks are individual objects, just as the data points rendered by APT [17] and the data items fed to a piece of Protopvis [6] code. One of the two dimensions Tory and Möller [18] use to distinguish between scientific and information visualization is also the discrete domain of the data (the other dimension being whether the spatial layout is given or chosen).

This distinction is fundamental and important, because it leads to entirely different types of solutions than work being done in scientific visualization. Ray casting, splatting, and other techniques that assume and require a continuous data domain (where values can be meaningfully interpolated between data points) are of no use in information visualization; though some InfoVis techniques are applicable in scientific visualization, because any real dataset necessarily consists of a limited number of data values.

Distinct data values translate into distinct, well-defined visual objects. While a volume

dataset can be rendered at many different resolutions, and can even be improved by better reconstruction and interpolation techniques, the number of items to be drawn for a given dataset in information visualization is not up for debate: one visual shape for each data point (except when there is filtering or aggregation).

This concept of maintaining the same discretization throughout the path from data to visual representation is important. It helps the simplicity of the model, and makes it possible to connect individuals in the visualization to individuals in the data source. It also means that an object based approach to creating data visualizations is a good fit. In particular, preface [12] and Protovis [6] provide classes that closely resemble primitives, as well as the means to iterate over such definitions to render data onto the screen.

3.2 Cognitive Models

Our perception is based on the notion of objects. Very early on, infants develop the ability to differentiate and track objects [25]. We do not see our world as a collection of pixels or surfaces, but distinct, physical objects.

This also applies to our perception of two-dimensional shapes. Pinker [21] describes a model of the mental processes behind reading and understanding a chart as an exercise in separating it into objects and examining the relationships between them. The user translates the properties of all these objects on the chart into cognitive representations that answer conceptual questions. Pinker shows that the cognitive representations of these objects are symbols with visual descriptions.

Recent work on the perceived relationships between elements of visualizations [35] has reinforced this idea: items in a bubble chart seemed to attract each other when they were

similar or close to one another. Such an effect is only possible if these elements are, in fact, seen as actual objects, with mass and other physical properties.

Grammel et al. [11] calls for tools supporting iterative refinement, as well as tools to help teach novices about visualization. The model we propose can support tools that have iterative refinement. We also believe the experimentation process that this model allows can help users to learn about information visualization.

We believe that having a model that closely aligns with our perceptual and cognitive processes can assist creativity by allowing users to think in more natural ways. As a result, it may help to support creativity when using tools based on the model.

CHAPTER 4: VISUALIZATION PRIMITIVES

We propose that a complete set of visualization primitives along with an appropriate set of visual properties will cover the possible visualization design space for tabular data. Visualization primitives allow for the creation of prototype objects that are instantiated for data points.

Visualization primitives uses a familiar concept from object-oriented programming; prototypes and instances [27]. Prototypes contain the essential information about themselves. They are the blueprint that allow copies to be created. Prototypes in the visualization primitives model know what data they get mapped to, as well as some internal geometric relationships. A designer creates prototype primitives, and assigns values to their properties, to generate a visualization. Instances are the objects in a visualization; the specifications of an instance come from the prototype and work together with the data to create the form of the visualization.

4.1 Prototypes

The concept of visualization primitives is built on a system of primitive shapes and their visual properties. The shapes correspond to commonly used components of a visualization, and their properties correspond to commonly used visual properties.

There is no optimal set of primitives because the goals of different implementations may be different. The set of primitives, and the properties they each have determines what

visualizations the model is capable of easily producing. A set optimized for one goal may leave gaps in the implementation's capabilities to achieve another goal. For example, a line primitive with angle, length, and midpoint properties can easily achieve texture effects for things like flow visualization, however it would be difficult to create a line chart or parallel coordinates with it. On the other hand, a line primitive defined with endpoints makes it trivial to create line charts and parallel coordinates, but textured effects are nearly impossible.

It is possible to have an optimal set of primitives for a given set of goals. One danger in implementing the model is having too much or too little complexity in a primitive. Adding angle, length, midpoint, and endpoints to one line primitive greatly increases the complexity of the primitive, and may confuse the user as to which combinations of properties can be assigned simultaneously. The key is to balance the complexity of the primitives with the capabilities it provides.

As a first pass, we covered five different primitives to explore the model. It should be noted that our set is not necessarily complete. The line primitive in our implementation does not include assignments for endpoints. Instead, we included line chart functionality in the point primitive, however this approach may not be intuitive for the user, and it does not allow for the creation of parallel coordinates. Adding endpoints to the line primitive in the future may solve these issues.

4.2 Visual Properties

Data is represented through visual (or Bertin's *retinal*) properties of the object. The visual properties of a primitive are what defines the capabilities it has. Providing properties

that closely align with the affordances of the visual object creates the most useful set of primitives. For example, a circle's size could be defined by a width, however it would be more useful to provide properties for area, radius, and circumference. These properties can encourage best practice in producing visualizations, simply by exposing the appropriate visual variables. Assigning quantity to radius creates a quadratic increase in area (a poor representation for human perception). But assigning that same quantity to area creates a linear increase, so the resulting visualization is perceived correctly. Radius is still provided as a property, as artificially limiting the user is not the goal of the system, but just making area available encourages good practice.

In order to provide appropriate properties, each primitive must have internal geometric relationships. The area, radius, and circumference properties are all interconnected. From any one of these properties, a circle's size can be calculated for drawing to screen. In addition, assigning one of the properties enables the calculation of the others. This supplies the user with an output data source that describes the existing objects. That data source can be used as input data for other objects.

4.2.1 Internal Geometric Relationships

Many visualization primitives have these internal geometric relationships between their properties. Embedding relationships into the software's representation of them lets the user take advantage of them when building the visual structure of their visualization. For example, this system makes it trivial to connect the left side of one rectangle to the right side of another.

These geometric relationships also mean that every primitive has a different set of prop-

erties. The differing properties help to determine what each primitive will be good at in a visualization. The implementations of the model should closely align the visual properties with the affordances of the visual objects. For example, a point is more intuitively positioned from its center, rather than a corner of its bounding box. The user of the system decides what data is mapped to what property, but the creator of the system decides what properties are available.

4.3 Iteration

We want the user to design a prototype object, not an algorithm. The visualization designer defines the loop once, and this creates each primitive seen in the visualization. When a user designs a loop, they are thinking about a data path or algorithm that gets the data to the final destination, but we want the user to be able to design an object. In visualization primitives we use implicit loops to allow the user to design a prototype object instead of a pipeline.

The core concepts of the information visualization model (one data point translates to one visual object) means that implicit iteration is something that can be done in every case. This iteration is built into the primitives and repeats for every item of data. The scaling and internal geometric relationships are calculated during the iteration, and the resulting data is stored in the primitive. Implicit iteration along with mappings are the key to allowing a prototype design strategy. Together they turn a prototype primitive into the repeated instances of primitives in the visualization.

4.4 Data

Since iteration handles the assignment of the individual data values to each instance, in the interface for the user, our data can be reduced to fields. But not all data used in a visualization comes from the original data source.

Algorithmic approaches to generating visualizations often have hardcoded values, or sequences of values generated during the drawing process. We call these values administrative data because they set a size for line weights, or provide position information for bars in a bar chart. Visualization primitives do not use an algorithmic approach, but this administrative data is still necessary to produce a visualization. Our implementation calculates administrative data during the time of data import and provides it as a data source just like the actual data dimensions.

We provide two fields of administrative data. Intervals provide values that count up by the specified scale value for each data item (Scale:1 Data: [0, 1, 2, 3...]). Constants provide the same value specified by the scale value for each data item (Scale: 1 Data: [1, 1, 1, 1...]).

4.4.1 Mappings

Since instances of the primitive in the visualization are not designed individually, it is necessary for there to be a way to abstractly assign data to a prototype primitive's properties. Mappings provide this abstract connection. Each property knows its data source, and a scaling factor. The primitive uses this information in the generation of its own data.

4.4.2 Connections

Since primitives store their own sets of data, they can also be used as sources of data. Having output data is important for being able to build relationships between multiple primitives. Creating any sort of glyph-based visualization requires that there be relationships between the multiple objects.

The data primitives store has often been processed, either through scaling, or through the primitive's internal geometric relationships. Using data from a primitive provides a simple way to make meaningful spatial relationships between two different primitives. An example use case of this would be VIE-VISU (Figure 2), a glyph visualization using small multiples to represent data over time [13]. VIE-VISU shows up to twelve dimensions simultaneously in one graphic, with a thirteenth dimension of time represented by each multiple of the graphic. The primitives within the multiples have the same spatial relationships in each multiple. An example of this technique implemented in our system can be seen in Figure 10 (page 27). Output data from each primitive provides a way to create the spatial relationships. Connecting the data of one primitive to another sets up a parent-child relationship between primitives. This allows the child to refer to the parent's data in order to generate its own data.

In some interface implementations, it is conceivable that this connection could be accomplished based on the spatial relationships between prototypes. For example, placing the left side of one prototype next to the right side of another prototype could trigger a connection between the respective data and property of the primitives. This visual method of defining connections only works for some spatial relationships. Relationships between

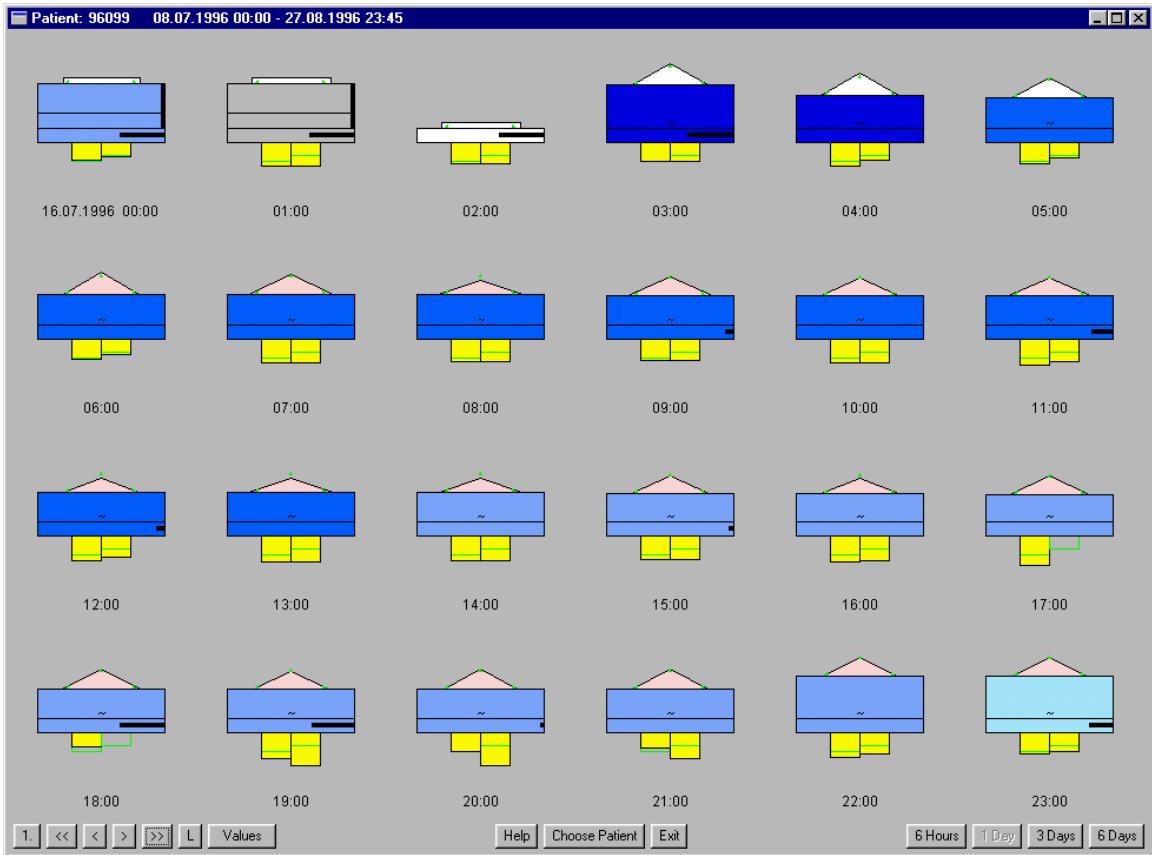


Figure 2: VIE-VISU is a glyph based visualization for displaying patient vitals [13].

color, line weight, or other more abstract spatial properties (area for example) would have to be defined using another interface.

4.5 Axes and Layout

Spatial layout is an important aspect of visualization design. Administrative data is one of the most common sources of data that influences layout. For example, in a bar chart, the only actual data on the chart is the vertical height of the bars. Their horizontal positions come from a sequence, and their vertical position comes from a constant. The majority of the layout control comes from the step size in the sequence, and the constant mapped to bar width (Figure 3).

The system is limited to layouts that are non-iterative, and do not require spatial infor-

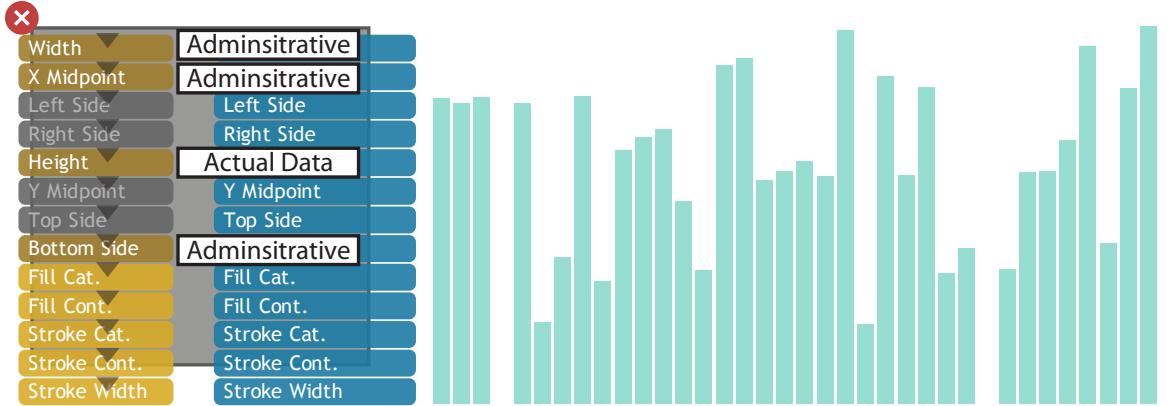


Figure 3: Layout of a bar chart is defined mostly by administrative data.

mation from the layout. This means that visualizations like tree-maps are not possible, because they subdivide a total space. This subdivision requires knowing the extents of the total, and iteratively dividing it based on the data. Word clouds also require iterative layout calculations using information about what is already drawn on the screen. This feedback loop is not present in the visualization primitives model, as it would require a pipeline approach to achieve.

Axes are not in our current implementation (axes weren't critical for testing creativity), but an axis primitive could provide the labeling necessary. Its properties would include the label data source, direction, size, etc. The type of axis could also be driven by a property, switching between a spatial axes or color scale. The actual labeling could be automatically handled using the extension of Wilkinson's Algorithm by Talbot et al. [28]

Axes are less of an issue for glyphs; the size and position of the objects is entirely relative, there is no axis to compare the spatial dimensions to. In the case of multiple views (small multiples, trellis displays) that have axes in common, scales exchange information so they can use the same transforms. This exchange is possible using the output data from each primitive to create connections between the axes present in a visualization.

CHAPTER 5: EXAMPLES

In order to help explain the property mapping process, we present an introduction to the interface, followed by a few examples.

The interface (Figure 4) has four main components. The left side contains a list of all of the data dimensions loaded into the tool. The green figures at the top of the screen are buttons for creating new primitives. In Figure 4, the circle on the right side with the lists of items is a prototype circle primitive. The golden list on the left side of the circle is the input properties, and the blue list on the right side is the output data from the primitive. The triangles above each property are scale handles.

Mousing over each property shows the current scale value, and if the property is mapped, highlights the data field that the property is connected to (in this case, area is connected to EmploymentRate). The grayed out properties are not available because area is mapped, and through the internal geometric relationships that defines the radius and circumference of the circle. Property mappings are accomplished by clicking on any blue data buttons, and then clicking on any gold property buttons. Scaling of a mapped property is accomplished by clicking and dragging the triangular handle left or right.

The blue circles are the instances of the primitive that are being specified by the prototype circle. Together they show the visualization that is produced by the mappings of the prototype. To prevent occlusion of the visualization, the prototype can be moved around

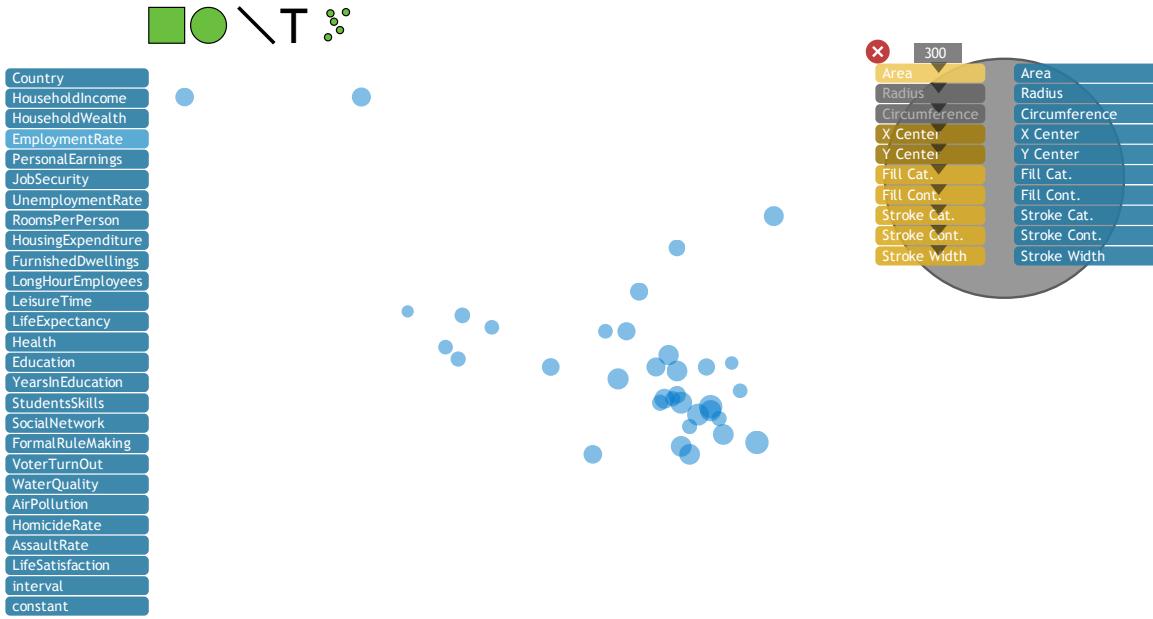


Figure 4: The interface for the current implementation of Visualization Primitives.

the canvas area by clicking and dragging. The visualization itself can also be moved around the canvas to help keep all instances on screen (in the case of extremely high or low values). Dragging the visualization will move all instances of all primitives as a group, in order to keep the data defined relationships intact. (To try the interface yourself, you may visit <http://visualizationprimitives.net>.)

The following examples show how to construct existing familiar visualization types. Some of them show interesting concepts that are highlighted when producing visualizations from a primitives based approach. Along with the description of how to construct them, we present versions of each visualization created with our current implementation of visualization primitives.

5.1 Bar Chart

Bar charts are created with a rectangle primitive. In a bar chart, the rectangle only has one data driven visual property, height. The rectangle's other properties are all derived

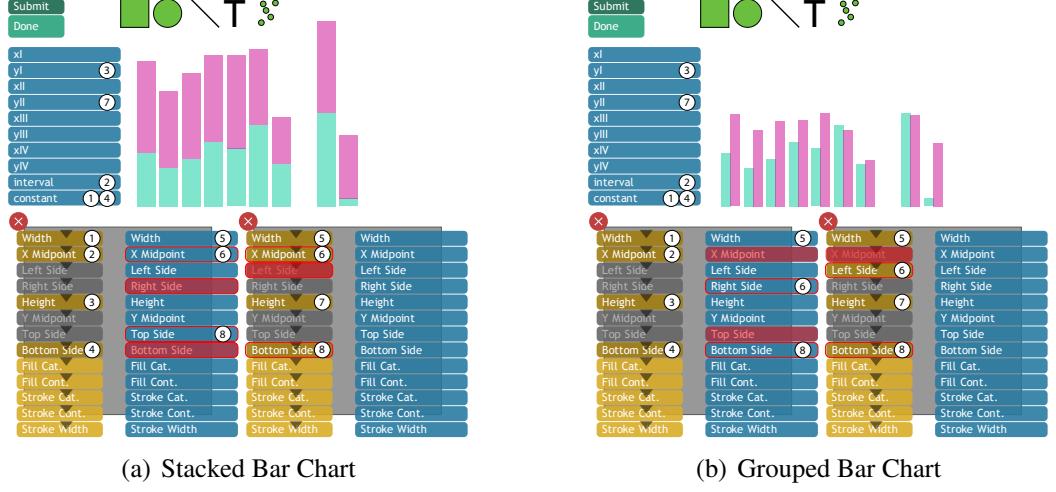


Figure 5: A stacked bar chart and grouped bar chart in our implementation. The data in both is from Anscombe’s Quartet [3]. Stacked bars (a) and grouped bars (b) differ only in the mapping of their sides. The numbers are added to indicate mappings, in the interface mouseover provides this information. The red highlights indicate the differences in the mappings between the two charts.

from administrative data. Horizontal position is tied to a sequence, while other properties are all tied to a constant.

To create a stacked bar chart, a new rectangle primitive can be connected to the existing rectangle primitive (Figure 5(a)). To make the connection, the top side of the original rectangle is used as the input for the bottom side of the new rectangle. They share the same horizontal positions and widths, and the height comes from a separate data source. Users design not only the prototype primitive, but also the relationships between the prototype primitives.

To create a grouped bar chart, the data assignments are similar (Figure 5(b)). The two rectangles share the same bottom side position, and the same width, and the height of the new rectangle still comes from external data. But now the left side of the new rectangle is connected to the right side of the original rectangle.

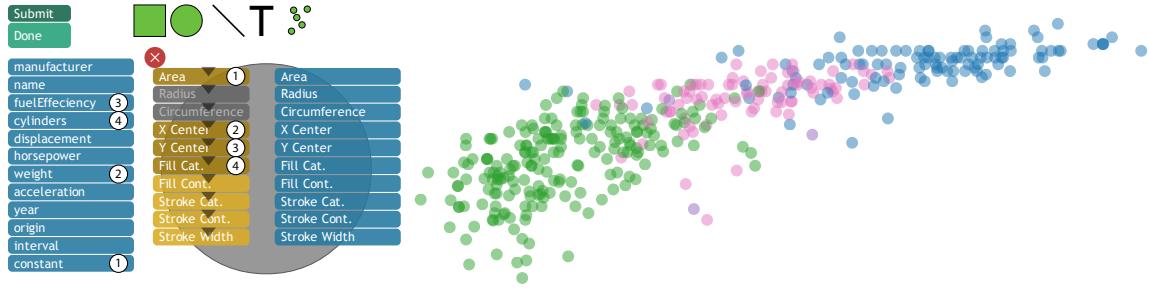


Figure 6: A scatterplot in our implementation using the UCI cars dataset. Miles per gallon is attached to vertical position, weight is attached to horizontal position. Number of cylinders is connected to color. The numbers are added to indicate mappings, in the interface mouseover provides this information.

5.2 Scatterplot

A standard scatterplot showing only one category is easily created with a circle primitive. Horizontal and vertical position of the shape are the two data driven properties. The other properties are all administrative. For some datasets, a scatterplot can represent additional values with the color or size properties.

Scatterplots are very simple to create with our implementation (Figure 6). In this example a circle has been used, but any primitive in our implementation has the horizontal and vertical position properties that are necessary. Connecting a categorical value (in this example, number of cylinders) to the fill color property, maps the value to a member of a categorical color scale.

Other primitives designed for scatterplots could provide a categorical visual variable of shape. This variable would provide access to a set of shapes for representing categorical data.

5.3 Line Chart

Line charts are a unique case. They break the information visualization model slightly by indicating continuity within a data dimension. This special case requires special treatment in the visualization primitives model.

Line charts need to have the iteration of some of their data to be offset by one. This makes the two endpoints of the line instances refer to different items in the data, and creates the visual continuity between each data item. Without this connection, a line chart would simply be a series of points.

There are a number of ways this can be accomplished in the visualization primitives model. Since a line chart without connections decomposes into points, we opted to make the points primitive allow connections between each point instance. This allows the line primitive to keep an angular property for creating texture effects.

In a line chart, just like a bar chart, the primitive has one essential visual property. The remaining properties are all administrative, and only serve to create the layout.

For point primitives in our implementation, the connectedness of the points is also a property. This allows connections to be driven by data, although in a line chart, connections are mapped to a constant value to turn them on for all instances (Figure 7). When the point primitive draws its instances in the visualization view, it grabs the position for the connection's other end from its previous sibling in the data. In our current implementation, this uses the existing order of data, however future implementations will include ways to sort data by any of the data dimensions.

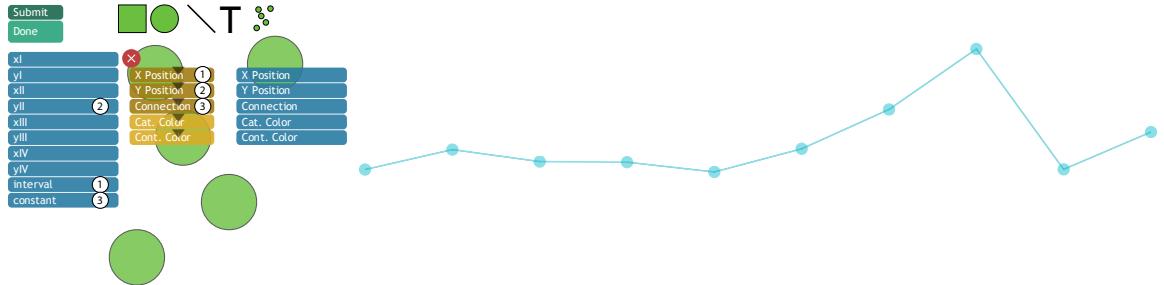


Figure 7: A line chart in our implementation is created by turning on a connection between points. The numbers are added to indicate mappings, in the interface mouseover provides this information.

5.4 Heat Maps

Periodic data is often represented in heat map grids where color is tied to data, while time data build the grid (Figure 8). With the right breakdown of time in the data, this is also a simple visualization to create with visualization primitives. Our example uses Typical Meteorological Year 3 data for Charlotte, NC [33]. One data field indicating day of the year is mapped to the vertical position, while horizontal position comes from the hour of the day. The width and height of the primitive instances comes from a constant value. The color scale is tied to the dry bulb temperature.

5.5 Waterfall Chart

Related to the gantt chart, a less well known chart is the Waterfall Chart (Figure 9). This chart type is useful for seeing trends in timeline data. In our implementation, the chart is created using three primitives showing data on United States Presidents. One rectangle primitive shows the lifespan of the presidents. The left side is mapped to birthday, the right side is mapped to the day of their death (or the current date). The height is assigned a constant value, and the vertical position is a sequence. Another rectangle primitive displays the presidents' time in office. The left side is mapped to their inauguration, while the

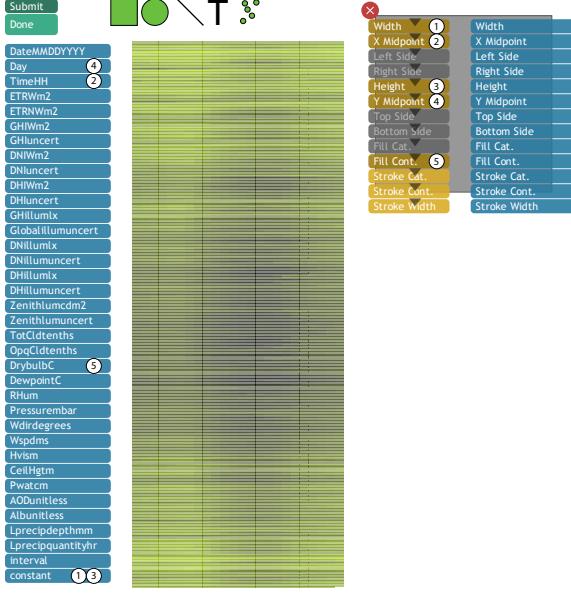


Figure 8: A heat map of temperature using TMY3 data [33]. The numbers are added to indicate mappings, in the interface mouseover provides this information.

right is mapped to the end of their term. The height and vertical position come from the respective outputs on the other rectangle primitive. A text primitive labels the names of each president. The text string is mapped to the names field. The left side comes from the right side of the lifespan rectangle, while the vertical position and height are mapped to the respective properties of the lifespan rectangle.

5.6 Glyphs

Glyphs [13] are often used to create small multiples for comparison. An example of this technique implemented in our system can be seen in Figure 10. The visualization is built using the cars dataset. This visualization is not an existing chart type, but is an example of a novel visualization created using visualization primitives. Four primitives make up each glyph, with each primitive representing a dimension of the data. Horsepower is the width of the blue rectangles, fuel efficiency is the width of the green rectangles. The yellow-green

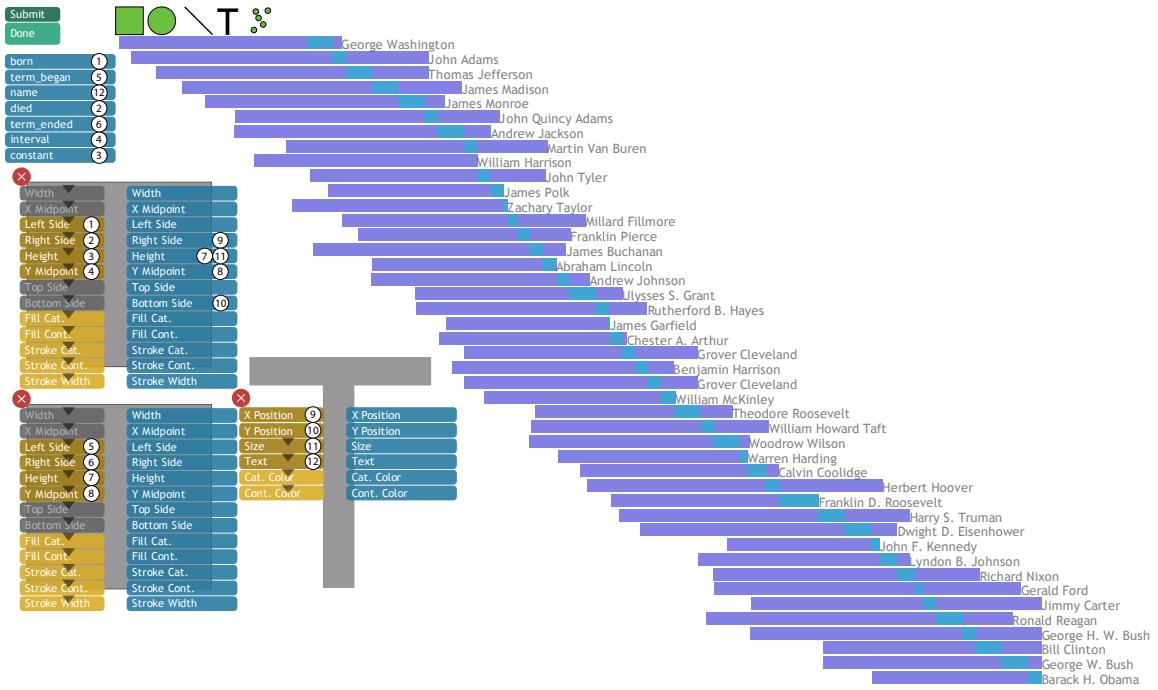


Figure 9: A waterfall chart showing the lifespans and terms of the presidents of the United States. The numbers are added to indicate mappings, in the interface mouseover provides this information.

“wheels” show weight, while the blue “wheels” show acceleration.

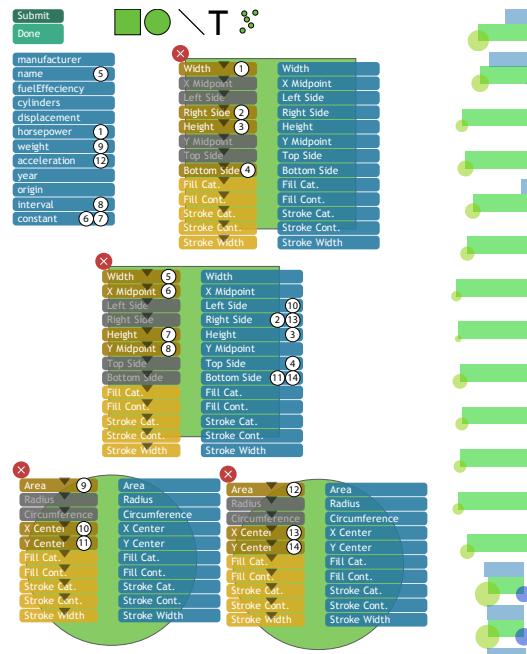


Figure 10: A glyph technique in our implementation using the cars dataset. The technique is similar to the one used in VIE-VISU [13]. The numbers are added to indicate mappings, in the interface mouseover provides this information.

CHAPTER 6: EXPERIMENT

One aim of the visualization primitives model is to enable creativity during the creation of data visualizations. Given this goal, tests to measure the implementation need to evaluate the ability of the tool to support creativity. This means that a user study will not evaluate visualization primitives against other tools using tests of task efficiency, or other similar metrics. Instead, we employ the Creativity Support Index (CSI) [9] to evaluate the tool’s ability to support creativity.

CSI scores are an index out of 100. The value is intended to provide a summation of a tool’s ability to support creativity. Along with the indexed score, there are ranked pairs that are also useful for ranking individual components of a tool’s creativity support. For example, support of exploration, effort vs. reward tradeoff, engagement, etc. The CSI is based closely on a tools ability to support a flow state for the user.

The study is intended to measure the tool’s ability to support creativity, not to quantify the quality of visualizations that are produced. These concepts are related, however they are not identical. Creative results are not guaranteed even with tools that support creativity perfectly. Creativity depends on several components, including the mind of the creator, and the capabilities of the tool at hand. We have included examples of “creative” visualizations generated with visualization primitives, however, these are anecdotal. Evaluating how creative they are is a secondary issue to how well visualization primitives support creativity.

Tasks that ask a user to find a certain thing in the data could provide inspiration, but they could also constrain thought processes. Time limits impose pressure that could help or hinder the creative thought process. We have opted to remove external pressures in the study, allowing participants to build visualizations at their own pace, setting their own tasks along the way.

To test the ability of Visualization Primitives to support creativity in visualization design, we conducted an online experiment. We chose an open-ended session design in which participants used Visualization Primitives freely, but with a dataset chosen by the experimenters and with questions preceding and following the session.

Specifically, we formed the following hypotheses regarding Visualization Primitives and creativity:

- More experience in visualization would lead to a higher score in the Creativity Support Index.
- More experience in visualization would lead to more visualizations generated using the tool.
- Visualization Primitives can significantly impact creativity in visualization design for users with sufficient experience with visualizations.

6.1 Materials

The materials used in this experiment consist of three main elements: the dataset, visualization-specific demographics, and the Creativity Support Index.

We chose to use the Better Life Index [19] as a dataset for several reasons. Primarily, the dataset has been used previously for a creative flower based visualization [19]. The

dataset is also socially relevant since it contains information on several different countries. This means a wider group of people may find the data relevant to themselves, encouraging participation. Finally, the dataset size is large enough to provide interesting possibilities for visualization and analysis, yet small enough to run well in a browser.

Besides typical demographics such as age, education, and gender, we also included questions to measure participants' experience level with visualization tools. Specifically, participants were asked to rank their experience with other data visualization tools (using a 20-point scale), and to describe their experiences in a text area.

The CSI is an analogue to the NASA Task-load Index (TLX), but for measuring how well a tool supports creativity [9]. In the CSI, participants are first given six 20-point Likert-scale questions that measure the tool's ability to support exploration, collaboration, engagement, expressiveness, perceived effort, and the tool's ability to become transparent to the design process. Next, participants answer fifteen ranked-pair questions, choosing which of the aforementioned aspects was more important to them during the activity (e.g. exploration versus collaboration). The CSI questions are shown below.

These questions are answered using a 20-point Likert-scale for each.

- It was easy for me to explore many different options, ideas, designs, or outcomes without a lot of tedious, repetitive interaction.
- I was able to work together with others easily while doing this activity.
- I was very absorbed/engaged in this activity - I enjoyed it and would do it again.
- What I was able to produce was worth the effort required to produce it.

- While I was doing the activity, the tool/interface/system "disappeared," and I was able to concentrate on the activity.
- I was able to be very expressive and creative while doing the activity.

For each pair below, please select which factor is most important to you when you were doing this activity.

Exploration	Collaboration
Exploration	Engagement
Exploration	Effort/Reward Tradeoff
Exploration	Tool Transparency
Exploration	Expressiveness
Collaboration	Engagement
Collaboration	Effort/Reward Tradeoff
Collaboration	Tool Transparency
Collaboration	Expressiveness
Expressiveness	Effort/Reward Tradeoff
Expressiveness	Tool Transparency
Expressiveness	Engagement
Engagement	Tool Transparency
Engagement	Effort/Reward Tradeoff
Effort/Reward Tradeoff	Tool Transparency

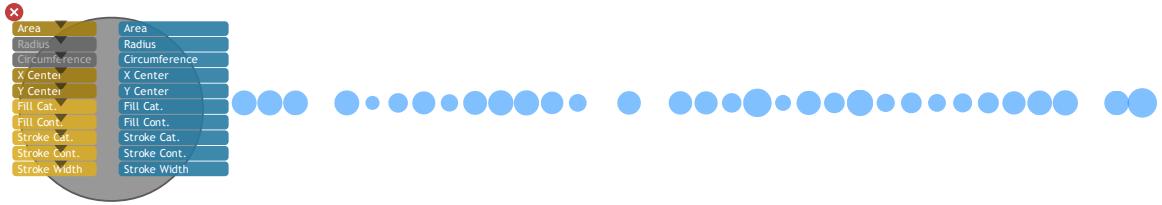


Figure 11: The bubble chart users were left with after completing the tutorial.

6.2 Procedure

Participants were recruited via advertising on Twitter to the URL <http://visualizationprimitives.net>. No payments or incentives were given. It took approximately six days to gather all responses.

After following the link, participants were shown a briefing which described the experiment as an open-ended visualization creation tool. When participants indicated they were ready, they were asked for basic demographics (age, gender, education level).

After completing the demographics, a multiple step tutorial was started. In this tutorial, participants were required to perform the actions necessary to make a bubble chart (Figure 11). Participants were also instructed to submit “results they were happy with” at any time via a submit button. That sent their current screen configuration to our server as a SVG (scalable vector graphics) file.

Following the tutorial, participants were allowed to use Visualization Primitives with the OECD’s Better Life Index [19] dataset for any amount of time, submitting notable visualizations as desired. In addition to participant-submitted visualizations, the system also auto-submitted when any new mapping was assigned.

When participants indicated they were finished with the open-ended session, they clicked a done button and were presented with the final questionnaire, which contained three parts:

visualization-experience, the Creativity Support Index [9], and a space for any additional comments.

6.3 Results

As with many online studies, participation drastically fell off with each step. 875 participants made it past the introduction page. 674 answered the demographic information. 95 participants completed the tutorial. The tutorial was considered a necessary component because without it, the users were not fully introduced to the tool's features. Of those who completed the tutorial, 44 completed the CSI, and only 31 of those built anything beyond the tutorial before answering the CSI. Those 31 participants made an average of 14 (median of 8) mappings. Participants who did not answer the CSI, but completed the tutorial, made an average of 9 (median of 6) mappings. For the participants who answered the CSI, their average time with the primitives was 21 minutes, with a median of 13, and standard deviation of 23. Not including submissions made during the tutorial, a total of 1060 SVG files were generated.

Since each participant worked with the same dataset and answered the same questions, our experiment adhered to a between-subjects design for the CSI evaluation. We divide our analysis into two cases to examine the relationship between experience, perceived creativity support, and productivity (i.e., the number of visualizations produced).

We excluded 4 participants who did not complete the CSI due to technical issues, and 3 participants who provided meaningless answers for the final questionnaire resulting in $n = 23$ for the final analysis.

Participants were then divided according to their reported experience with visualization

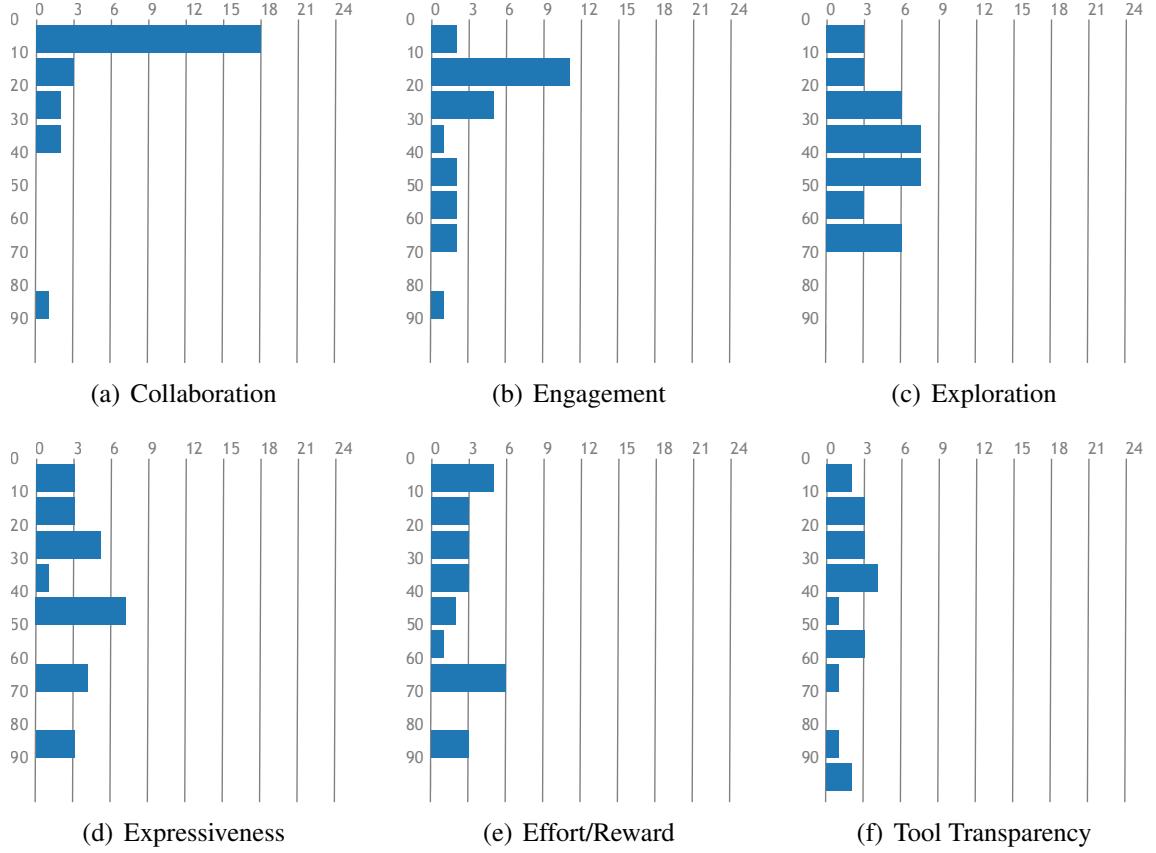


Figure 12: Histograms of the rankings for each factor (built using the visualization primitives prototype).

tools. Since a 20-point scale was used for the visualization experience question, we divided participants evenly into a high ($experience \geq 17, n = 12$) experience group and a low ($experience < 17, n = 11$) group.

6.3.1 Experience and Creativity Support

CSI scores are an index out of 100. They are calculated by combining ranked pairs with Likert scale results. For each factor, the Likert scale results are multiplied by the total factor count from the ranked pair portion. This value is then added to all the other factor results, and divided by 3 to arrive at an index out of 100. Table 1 and Figure 12 show the CSI results for our experiment.

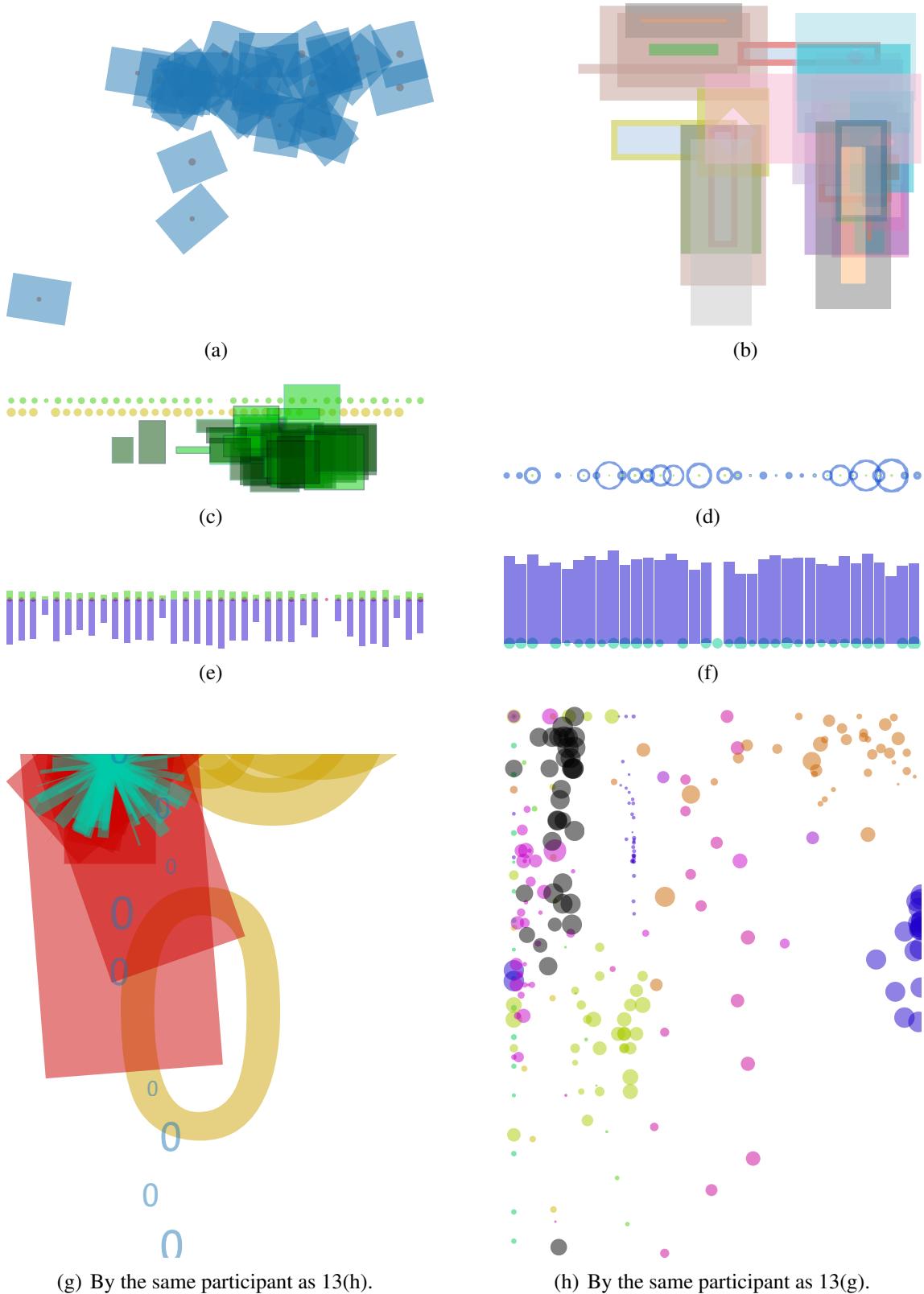


Figure 13: A selection of example visualizations submitted by study participants.

Factor	Mean	Median	Standard Deviation
exploration	41.62	39.5	24.46
collaboration	9	0	18.22
engagement	26.31	18.5	19.87
tradeoff	37.73	37.5	27.3
transparency	53	45	35.48
expressiveness	38.23	40	23.52

Table 1: Factor ranking results.

We compared CSI scores for the low-experience and high-experience groups using a t-test. This yields a significant effect for the CSI-score $t(21) = 2.0843; p = 0.0495$, with CSI-score in the high-experience group being higher than that of the low-experience group. The lower mean CSI-score appeared in the low-experience group ($M = 54.27, SD = 22.50$), meaning participants in the high-experience group felt that Visualization Primitives supported their creativity in visualization design more ($M = 73.75, SD = 22.28$). (The paper presenting the CSI offers a sample ranking of 77.3 [9]) These results are consistent with our hypothesis that Visualization Primitives can significantly impact creativity in visualization design for participants who have sufficient experience with visualizations.

6.3.2 Experience and Productivity

We also compared experience with the total number of SVG graphics submitted. This did not yield a significant effect for the the total number of SVGs submitted, although participants in the high-experience group submitted more SVGs on average ($M = 21, SD = 19$) than those in the low-experience group ($M = 17, SD = 10$). These results do not support our hypothesis that users with more experience in visualization would submit a higher number of visualizations.

6.3.3 General CSI Results

In addition to discussing the total CSI scores, it is also useful to look at which categories the participants ranked most important to them while using Visualization Primitives. These results are shown in Figure 12. Tool transparency and exploration ranked most important, while collaboration and engagement ranked at the bottom of users' priorities.

6.3.4 Example Visualizations

The vast majority of users generated simple visualizations with a single primitive. Since the tutorial left users with a circle primitive, scatterplots became one of the most common visualizations that was created. Many users created visuals with only one or two primitives, and often there was no data driven spatial relationship between those primitives. Connections between primitives were rare, possibly because that ability was not explicitly shown to users during the tutorial, only described in text. The connections that were created were often built off of the bubble chart created during the tutorial (Figures 13(e), 13(f)). Often the creations fell somewhere between art and data visualization (Figures 13(a), 13(c), 13(d)), and some resulted in purely artistic work (Figures 13(b), 13(g), 13(h)).

CHAPTER 7: DISCUSSION

The results of the experiment indicate that visualization primitives support creativity, especially with users who have had adequate experience with visualization tools.

User interface issues were a limiting factor in users' exploration of the design space. Specifically, having to mouse over properties to see what they were bound to made it difficult to build and maintain a mental model of what was being shown. While our goal was to reduce clutter from connecting lines, the alternative may have created more issues than it solved.

The unfamiliarity of the interface definitely contributed to many users frustrations. Many comments expressed users' frustration with not knowing what the parts of the interface did. This could be addressed by further iterations of the tutorial and improvements to the UI.

The internal geometric representations of the primitives also may have contributed to confusion. Drawing elements to the screen is often not possible until multiple properties have been specified. This could be alleviated with default values, however the solution is far from elegant, and can produce unexpected behavior when assigning properties.

Not all the results that participants created and submitted were strictly visualizations in the sense of being useful. We see some playful examples that are more artistic than useful, which is in line with our goals. We did not ask people to create anything in particular so as not to constrain their exploration.

The considerable time spent on the site shows that they were engaged and interested in exploration, despite the user interface flaws. This suggests that tapping into users' creativity is a promising way of getting them interested in visualization.

7.1 Lessons Learned

This experiment was very obviously flawed in many ways, however it revealed to me that there were many components of visualization primitives that need work. The interface issues became a major hinderance to testing the model, and the experiment design itself needs a lot of improvement. The best course of action for proceeding is to take what I learned about the interface and spend a significant amount of time improving it. After the interface has been revisited, the experiments need to be planned to evaluate both the creativity of the product the participants create, and the creativity support of the application.

In addition, the experiments need to provide more direction to the participants. Rather than a period of free exploration, the participants need to be given a task to achieve. This constraint should help to focus participants, and will make it possible to evaluate the creativity involved in what they produce. Some possible tasks would be: "Create three visualizations that answer x." or "Show a visualization that answers both x and y." Still including a period of open exploration before the tasks may help participants to become more familiar with the tool, and make it easier for them to achieve the given tasks later.

Without knowing the purpose the participants had when creating a visualization, it is impossible to rate the usefulness of the visualization, and therefore impossible to evaluate how creative it was. Providing a task structure to participants means that the usefulness can be evaluated, and even ranked against other participant's produced visualizations.

Evaluating creativity of the participant's product may be best accomplished through the use of the Consensual Assessment Technique (CAT) [1]. The technique involves a panel of expert judges who spend time examining the "creative" works, and evaluating their creativity. A panel of three to five judges will be able to rate the creativity of what is created. This information can then be used in conjunction with the Creativity Support Index to better determine the creativity support of the tool, and validate the participant's answers to the CSI.

Another change that should be made to the study is the participant selection. The shotgun approach of attracting random participants on the internet is likely partially responsible for the extremely low completion rate of the study. It also means that many participants were not as familiar with the concept of visualization, so their produced results were more like data art than visualization. Manually recruiting participants can help to constrain the group to individuals who are aware of things like Bertin's hierarchy of retinal variables [4]. This will increase the likelihood that their visualizations will be useful for a task.

7.2 Interface Improvements

The visualization primitives model has proven to be effective at some things, however there are still limitations due to the implementation. The areas that could result in the most striking improvements in creativity support deal with the interface, and are needed mostly to help users achieve a flow state.

7.2.1 Default Values

In the current implementation, there are no default values assigned to any of the properties. This results in there being no feedback to the user when they map properties that

require multiple other properties to also be mapped. Once the necessary subset of properties has been mapped, the primitive can draw itself, but in the intermediate time, the primitive does nothing. For example, a rectangle primitive needs two horizontal properties mapped out of four, and two vertical properties mapped out of four. Without default values, the primitive cannot draw itself until four out of those eight properties have been mapped.

To complicate things further, knowing which properties to assign defaults to depends on the internal geometric relationships of the primitive. The best way to solve this is probably a decision tree that chooses which defaults to assign based on what else has been assigned already. The important part of this decision tree structure is how it fits into the internal geometric relationship architecture, and how it gets initiated in the mapping assignment.

Fixing this so that there is always immediate feedback should help the user experience

7.2.2 Primitives Graph

The visualization primitives model allows for dependency relationships to occur between the properties of primitives. These dependencies are represented through mappings.

Mappings in the current implementation record connections between data output, and a property. The system for storing them is simplistic, as each only belongs to the parent primitive. There are currently no methods for traversing the structures of mappings that can be created. This presents problems with updating child primitives when a parent primitive is modified.

A proper system for storing mappings would allow traversal of the mappings, and prevent circular traversals when a loop is present in the structure. The current implementation also means that scaling of values is clumsy when those values come from a parent primitive

and not the original data. A better structure would allow the properties to retrieve the scale value from the parent and modify it rather than just multiplying their own scale value on top.

This structure will allow for a better interface and more consistent experience with scaling data inputs to properties. This should help the user to achieve a flow state by making the interface control behavior more intuitively.

7.2.3 Connections Between Primitives

Currently, the interface provides buttons for each visual property that a primitive has, and output buttons that pass the primitive's data through. This interface works for testing the model, but it is hardly user-friendly. Ideally, the interface should feel much more natural, and have the same feeling of direct connections that come from the implementation's immediate feedback.

As mentioned, one possible way to accomplish that is through using the location of the prototypes relative to each other to help determine the spatial relationships between primitives. This type of connection feels physical, and would be relatively intuitive to users. However spatial relationships cannot define mappings for properties like color, rotation, or line weight. These mappings still need some sort of input and output interface, and may require things like icon representations or other interface elements to allow their mapping. The interactions of connecting abstract visual properties of different objects are new, and there are not many metaphors to the physical world. Inspiration for these elements could, however, be drawn from layout or drawing tools.

For example, rotation of an object in many vector drawing programs is accomplished

using a handle on the object. Perhaps using a similar handle for controlling rotation would improve the intuitiveness of the connection process. Unfortunately, interface elements borrowed from drawing tools may not be intuitive for users that are not familiar with them. Testing of several different approaches may be necessary to determine what interactions are simplest to grasp and use for most users.

7.2.4 Set of Primitives

The set of primitives provided by the system has the most profound impact on what can be created with the system. The current set is definitely not ideal. Future tweaking of the set should help to optimize the balance between simplicity and flexibility.

One potential way to proceed with this is to build a list of visualizations that fall within the theoretical constraints of the system (tabular data, non-iterative layouts, etc.) Establishing which visualizations in the list cannot be created with the current set of primitives will help to identify possible primitives that should be created. It also may be possible to tweak the properties and internal geometric representations of existing primitives to allow the creation of other visualizations.

Identifying which primitives to add, and which to remove (when there is excessive overlap in their capabilities) is the difficult part. The best way to achieve this may be to select the primitives that cover the greatest range of visualization techniques that are not possible with the current set. By aiming for the best coverage first, corner cases can be identified, and helpful modifications to existing primitives should be easier to find.

Once the list of theoretically possible visualizations can all be created with the implementation, the implementation will be complete. This still will not mean that the set of

primitives is optimal for a given task, but it does guarantee maximum flexibility of the system.

7.2.5 Creativity Support

The initial evaluation of creativity support had several issues. Aside from using an earlier version of the CSI, the tutorial teaching people how to use the system probably needed to be enforced better. Several people didn't finish the tutorial entirely before exploring on their own and continuing the study. This is a minor issue, and easily fixable in the next study.

It would also be good to provide a task for participants to perform. The participants of the last study seemed to just play around with the visuals, and not think much about A goal will help to focus people and give them constraints to work within.

The difficulty in this task is finding a goal that strikes a good balance between open ended solutions, and providing focus. If the goal is too open ended, it will not serve as a motivator. But if the goal is too constrained, there will be a small handful of solutions and no creative solutions will be generated. One possible way to alleviate this is to pick a more focused goal, and follow it up with a period of free exploration for users to express their creativity.

CHAPTER 8: FUTURE WORK

Putting the lessons learned into practice is the next step of the process. The publication opportunities that are most appropriate for this work set the timeline for improvements to the interface and conducting studies. The interface needs to be given a lot of attention. I would like to achieve all of the interface improvements in Section 7.2. This puts the implementation on a good footing to have several studies run against it with only minor changes between each study.

The study design needs to be adjusted heavily to account for the addition of the CAT, and to encourage participants to complete the tasks more thoroughly. Providing a task will help with completion. One possible task would be showing three visualizations that show airlines that have fewer delays, using an Airline On-Time Performance dataset. Another possible task is creating three visualizations that show data dimensions that correlate with miles per gallon in a dataset on cars. A concrete task will provide participants with boundaries and restrictions, giving them restrictions to be creative within. The task will also make it possible to judge how creative their solutions are.

It may also be important to run the study with individuals in a lab setting rather than over the internet. This also makes it simpler to monetarily reward participants for their participation, and provide further motivation (funding provided by Visually). I can ensure that the participant fully understands the interface, and has an opportunity to ask questions

before they are left to the task. It also means browser compatibility will not be an issue.

Using contacts in the visualization domain, I have access to a handful of students studying visualization, as well as a large collection of information designers working in the visualization space. The student members of this group may be pulled from data visualization classes at UC Berkeley, and the designers may be pulled from a pool of contacts I have through my work at Visually. This group provides a set of participants that are likely familiar with data visualization concepts, and able to perform creative tasks using the tool.

The study will test two different interfaces. One interface will be visualization primitives, while the other will be Tableau Public. Tableau is the closest available software to the visualization primitives model and allows the most freeform creation of visualizations. Both interfaces will have the same data set loaded, and the tasks will be the same in both cases. Participants will experience both interfaces, in random order.

The visualization primitives portion will begin with an information session on the tool and the components of its interface. Participants will be trained on how to make a visualization, and how to connect primitives together. After training, they will be presented with the tasks, and asked to solve them. Following that, they will be given a time period of free exploration where they can explore the interface at their will, and ask questions freely. During this stage, they will be allowed to designate any of the things they produce as "creative". Once they have completed the assignment, they will be given the CSI to answer for the visualization primitives interface.

The Tableau portion will follow mostly the same model. It will begin with an information session on how to use Tableau Public, and the components of the interface. Participants will be trained on how to make and manipulate a visualization, before being presented with the

tasks and asked to solve them. This will also be followed by a period of free exploration with an opportunity to mark produced things as creative. After completion, they will be given the CSI again to answer for the Tableau interface. Once participants have used both interfaces, their participation in the study is over.

The expert judges will also be drawn from my contacts through Visually. They will be chosen based on their expertise in visualization, and will come from both the academic realm (professors of data visualization), and from practicing leaders in the field. They will be monetarily rewarded for their participation in the study (funding provided by Visually). An interface will be provided to allow the judges to rate the graphics over the internet. The judges will be trained, either in person, or over a video call, on the rating system, and on how to use the interface rate each participant produced visualization. The judges will be instructed to do these activities alone, without consultation with others, to ensure they do not influence each other.

I expect to find that the visualization primitives interface allows for more creative results than Tableau. I expect it to have a higher CSI score, and I expect the judges to rate the products as more creative. This would mean that the approach provides better freedom, and better support of creativity in the creation of data visualizations.

Studies beyond this one will be based somewhat on the results of the study, but the main goals will be to establish the creativity support of the visualization primitives model. Another possible angle to examine is whether the model supports creativity better because of how closely it matches the mental model of visualization creators. If this turns out to be true, then the work will make a significant contribution to our understanding of information visualization, and can be used in the development of visualization theory.

CHAPTER 9: CONCLUSION

Our model presents a novel way to approach visualization design, and to support creativity. Visualization primitives in conjunction with visual properties create an environment optimized for rapidly prototyping new visualization techniques. The environment supports instant visual feedback and helps to develop an efficient flow for visualization design.

The theoretical contributions of this model are based on a cognitive argument that is closely aligned with visualization theory. Our study shows considerable interest in and potential for the creation of new visualizations using visualization primitives.

The most compelling argument for this approach is the control and flexibility given to non-technical users who want to create novel visualizations. We hope that some of the principles it presents can be integrated into existing and future visualization and visual analytics tools.

The planned future work will help to solidify the visualization primitives model as a useful way of conceptualizing data visualization. The lessons learned from testing the limitations of the model will help to contextualize existing visualizations. This can help to add to our theoretical understanding of visualization, especially with regards to how a visualization's visual characteristics impact what it is capable of displaying. Improving the creativity support of the implementation will help to prove the the model's capacity as a tool for successfully building data visualizations. It will also help to understand the

role of creativity in the visualization creation process, and perhaps help to determine what characteristics other visualization tools need to have to help support creativity.

REFERENCES

- [1] T. Amabile. *Creativity in Context: Update to the Social Psychology of Creativity*. Westview Press, 1996.
- [2] E. Anderson. A Semigraphical Method for the Analysis of Complex Problems. *Proceedings of the National Academy of Sciences of the United States of America*, 43(10):923–927, 1957.
- [3] F. Anscombe. Graphs in Statistical Analysis. *The American Statistician*, 27(1):17–21, 1973.
- [4] J. Bertin. *Semiology of graphics*. University of Wisconsin Press, 1983.
- [5] Bestiario. Quadrigram. <http://www.quadrigram.com/>.
- [6] M. Bostock and J. Heer. Protovis: A Graphical Toolkit for Visualization. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):1121–1128, 2009.
- [7] M. Bostock, V. Ogievetsky, and J. Heer. D³ data-driven documents. *Visualization and Computer Graphics, IEEE Transactions on*, 17(12):2301–2309, 2011.
- [8] S. Card and J. Mackinlay. The Structure of the Information Visualization Design Space. *Proceedings Information Visualization*, page 92, 1997.
- [9] E. A. Carroll and C. Latulipe. The creativity support index. In *CHI'09 Extended Abstracts on Human Factors in Computing Systems*, pages 4009–4014. ACM, 2009.
- [10] W. Chao and T. Munzner. Poster: Rapid Pen-Centric Authoring of Improvisational Visualizations with NapkinVis. *IEEE Transactions on Visualization and Computer Graphics*, 16(6), 2010.
- [11] L. Grammel, M. Tory, and M. Storey. How information visualization novices construct visualizations. *Visualization and Computer Graphics, IEEE Transactions on*, 16(6):943–952, 2010.
- [12] J. Heer, S. K. Card, and J. A. Landay. Prefuse: A Toolkit for Interactive Information Visualization. In *CHI '05: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 421–430, New York, NY, USA, 2005. ACM.
- [13] W. Horn, C. Popow, and L. Unterasinger. Metaphor graphics to visualize ICU data over time. *Intelligent Data Analysis in Medicine*, 1998.
- [14] D. Keefe, D. Acevedo, J. Miles, F. Drury, S. Swartz, and D. H. Laidlaw. Scientific sketching for collaborative VR visualization design. *IEEE Transactions on Visualization and Computer Graphics*, 14(4):835–847, July-August 2008.

- [15] B. Lee, P. Isenberg, N. H. Riche, and S. Carpendale. Beyond mouse and keyboard: Expanding design considerations for information visualization interactions. *Visualization and Computer Graphics, IEEE Transactions on*, 18(12):2689–2698, 2012.
- [16] L. Lins, D. Koop, J. Freire, and C. T. Silva. Defog: A system for data-backed visual composition. Technical Report UUSCI-2011-003, SCI Institute, University of Utah, 2011.
- [17] J. Mackinlay. Automating the design of graphical presentations of relational information. *ACM Transactions on Graphics (TOG)*, 5(2):110–141, 1986.
- [18] T. Möller. Rethinking Visualization: A High-Level Taxonomy. In *IEEE Symposium on Information Visualization*, pages 151–158, 2004.
- [19] OECD. Better Life Index. <http://www.oecdbetterlifeindex.org/>.
- [20] S. Ortiz. Use Cases of Impure, an Information Interface. In *VisWeek 2010 Discovery Exhibition*, pages 2–4, 2010.
- [21] S. Pinker. A Theory of Graph Comprehension. In R. O. Freedle, editor, *Artificial Intelligence and the Future of Testing*, pages 73–126. Lawrence Erlbaum Associates, 1990.
- [22] W. Ribarsky, E. Ayers, J. Eble, and S. Mukherjea. Glyphmaker: creating customized visualizations of complex data. *Computer*, 27(7):57–64, 1994.
- [23] S. F. Roth, J. Kolojejchick, J. Mattis, and J. Goldstein. Interactive Graphic Design Using Automatic Presentation Knowledge. In *Proceedings CHI*, pages 112–117. ACM Press, 1994.
- [24] J. Scholtz. Beyond usability: evaluation aspects of visual analytic environments. In *2006 IEEE Symposium On Visual Analytics Science And Technology*, pages 145–150, 2006.
- [25] E. S. Spelke. Principles of Object Perception. *Cognitive Science*, 14(1):29–56, 1990.
- [26] C. Stolte, D. Tang, and P. Hanrahan. Polaris: A System for Query, Analysis, and Visualization of Multidimensional Relational Databases. *IEEE Transactions on Visualization and Computer Graphics*, 8(1):52–65, 2002.
- [27] A. Taivalsaari. Classes vs. Prototypes: Some Philosophical and Historical Observations. *Journal of Object-Oriented Programming (JOOP)*, 10(7):44–50, 1997.
- [28] J. Talbot, S. Lin, and P. Hanrahan. An Extension of Wilkinsons Algorithm for Positioning Tick Labels on Axes. *Computer*, 16(6):1036–1043, 2010.
- [29] B. Victor. Drawing Dynamic Visualizations. <http://worrydream.com/#!/Showreel2012>.

- [30] F. Viegas, M. Wattenberg, F. Van Ham, J. Kriss, and M. McKeon. Manyeyes: a site for visualization at internet scale. *IEEE Transactions on Visualization and Computer Graphics*, pages 1121–1128, 2007.
- [31] J. Walny, S. Carpendale, N. Henry Riche, G. Venolia, and P. Fawcett. Visual thinking in action: Visualizations as used on whiteboards. *Visualization and Computer Graphics, IEEE Transactions on*, 17(12):2508–2517, 2011.
- [32] H. Wickham. A Layered Grammar of Graphics. *Journal of Computational and Graphical Statistics*, 19(1):3–28, Mar. 2010.
- [33] S. Wilcox and W. Marion. *Users manual for TMY3 data sets*. National Renewable Energy Laboratory Golden, CO, 2008.
- [34] L. Wilkinson and G. Wills. *The Grammar of Graphics*. Springer, Chicago, IL, 2nd edition, 2005.
- [35] C. Ziemkiewicz and R. Kosara. Laws of Attraction: From Perceived Forces to Conceptual Similarity. *Transactions on Visualization and Computer Graphics (Proceedings InfoVis)*, 16(6):1009–1016, 2010.