P J I M PARSONS JOURNAL FOR INFORMATION MAPPING

THE PARSONS INSTITUTE FOR INFORMATION MAPPING
68 Fifth Avenue
New York, NY 10011
212 229 6825
piim.newschool.edu

Suitability of Form for Data: An Analysis of Data Visualization Projects

DEEPIKA SAHAI, LALIT CHAUDHARI, SINDHU MS, & SINDHU S. RAO

KEYWORDS *Instructional diagrams, perspectives, visual communication, visualization, wordless diagrams*

ABSTRACT The intent of information visualization is to exploit the internal structure of data and its associated relationships to find quick patterns, trends, or events positioned within large data sets, and to help the reader comprehend unseen and often complex relationships. Determining the appropriate visual properties for each element is a constant effort. This paper attempts to depict effective methods to visually communicate information through a set of four well-selected visualizations. One form of visualization is through wordless diagrams, which investigate the appropriateness of a chosen rendering technique, geometry, orientation, grouping, and sequence, and their impact on the recognized meaning and its emotional impact on the readers. A static visualization of data pertaining to the 2013 Delhi assembly elections is attempted in the second visualization. Adding multiple perspectives to depict noise levels in a locality by means of tangible data visualization is explored in the third visualization. An interactive visualization tool is created to facilitate exploration of a time use survey data set having multiple dimensions to quickly draw attention to the key insights from the data. The paper concludes with a discussion on the key takeaways from the iterative design process and set of recommendations for visualization of similar data.

INTRODUCTION

Raw data is ubiquitous. Comprehension and understandability is limited due to multidimensionality of raw data; it needs organization, structuring, and analysis. From these initial investigations comes the requirement to present data in a manner that would make it informative, sensible, and impactful. Every data set (numeric or non-numeric) has myriad possibilities for conveying the differing aspects of source information. The final presentation will affect how the users will interpret the data. In order to comprehend the data, in-depth analysis and multilayered visualization is crucial. It is the prime responsibility of the visualizer to be able to convey the intended information accurately without causing miscommunication. This paper is an exploration of data ranging from static to interactive visualization. Following are four different sets of visualized data:

1 WORDLESS DIAGRAM: HOW TO IRON A SHIRT

An everyday activity of ironing a shirt is explored through visuals. When visualized on paper, a simple task can be difficult to represent. Pictorial representation cross social and linguistic boundaries with ease, however, creating a diagram without any words to support them can be more challenging. Nigel Holmes, author of *Wordless Diagrams* (2005) noted "while I have always preached that the best diagrams are a seamless and careful combination of words and pictures, I have always secretly wanted to create a graphic language that speaks for itself: pictures but no words."[1] On similar lines, the objective is to explain the task of ironing the shirt by enabling the reader to sequentially follow it step-wise through the visuals without any words.

ORGANIZING THE INFORMATION

The inspiration is taken from comic techniques and sequential art. The first step is to organize the information, determining if all the information is actually required, the items/objects to be included or discarded based on their importance, and the emphasis for narrating the story. The key is to keep the scenario as generic as possible taking the reader's point of view into consideration. To keep the readers on the same page, the ironing board was not considered. Organization of the information with right amount of detail and emphasis was to be done effectively to make sure the reader is able to comprehend and use the instructions easily.

DECIDING THE FLOW

After organization, the flow of the instruction is decided. The steps are arranged in a sequence having disclosure structure: having a beginning, middle, and end of the narration.[2] The sequence approximates the real experience. It starts with a wrinkled shirt being displayed with water sprayed on it, and then the ironing of the collar to sleeves, finally an ironed shirt is shown as the completed task. The choice of keeping wrinkles and outlines is to give the viewers an element to differentiate between an un-ironed shirt and an ironed one.

ESTABLISHING THE CONTEXT

Establishing the context was next wherein the view (top, side, front) and perspective was decided. As per the experiment done by Perrett and Herries (1988), Harries, Perrett, and Lavender (1991), Perrett.et.al (1992) on obtaining the preferred view for a potato, tetrahedral, clay head, and a machine tool part, it was found that the preferred views are aligned with an object's major axis relative to off-axis views.[3] The chosen view for the shirt is top view which is aligned with the preferred view. Also the location rule says that it is important to show orientations of the object in a manner that is physically realizable.[4] This is further inspired by the real life scenario on how the reader would experience the task while actually ironing the shirt. The diagram is kept simple for easy interpretation with use of only three colors. It is observed that colors have the tendency to capture better attention level, and thus, better memory. The choice of colors are kept contrasting to differentiate between the inner and outer areas of the shirt thereby making it easy to comprehend.[5] The red dot on the iron, along with its conical form, helped in taking the attention on the direction of the iron on the shirt. [FIGURES 1, 2, 3]

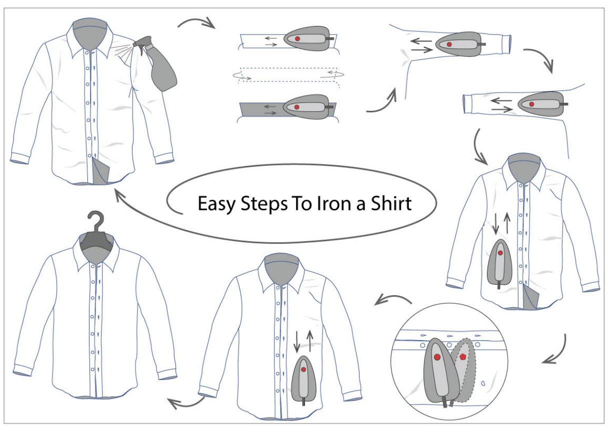

FIGURE 1: *First iteration had arrows to show the flow of steps.*

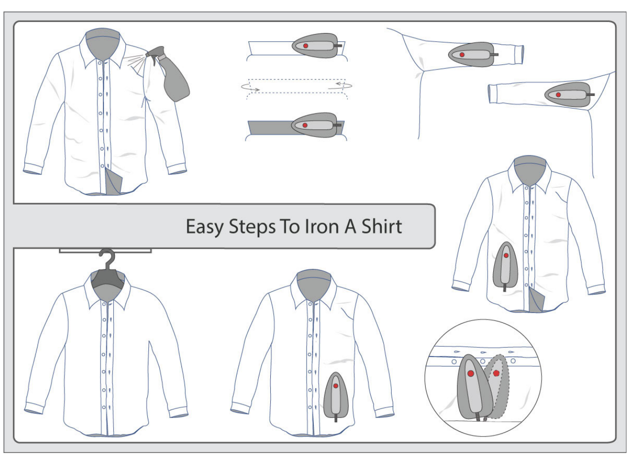

FIGURE 2: *In second iteration, the arrows were removed allowing the middle separator to guide the flow.*

REMOVING REDUNDANCY

To avoid the overburdening with redundant steps, the total number of steps is carefully kept to a minimal number while providing adequate information. One of the nine strategies of infographic design says that redundant information helps readers easily digest the information while too much of it can bore the readers. Maintaining optimal balance between predictability and uncertainty is needed. There is some information that needs to be emphasized more and this should be highlighted. Object of focus rule says that the user should be able to clearly identify which part(s) is being acted on in each step.[6] In our case ironing between the buttons need to be highlighted to avoid damaging them. This information has been shown separately by zooming in. This makes sure that readers take note of it (therefore, being subsequently more cautious while ironing between the buttons). Finally, displaying the outcome of the task as an ironed shirt was kept as the last step to give readers a sense of pride and satisfaction on what they would achieve in the end. It helped in catering to their emotions and rewarding them in the end.

PUTTING IT TOGETHER

By following the above-mentioned methods, basic elements took shape and the steps to be included were clear; it is the placement and overall layout and overall

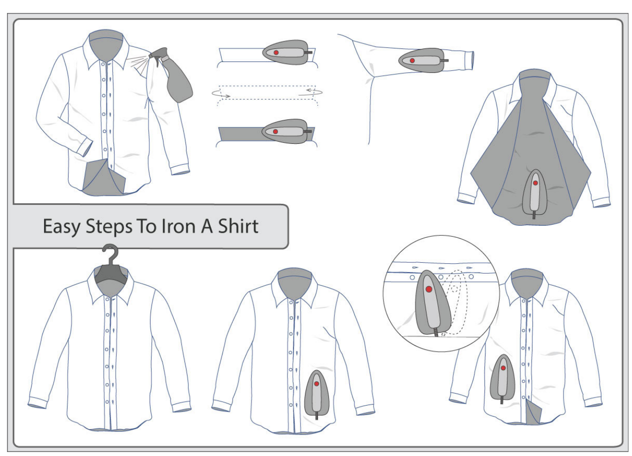

FIGURE 3: *Final iteration had better organized elements with highlighted steps.*

organization that needed to be looked at. After finalizing the basic elements, various iterations were tried on the overall layout and the final diagram created. The steps, graphics, orientation, and elements were kept simple to keep the communication easily visible, immediate, and cohesive. Words were minimal—or not needed at all.

2 DELHI ASSEMBLY ELECTIONS VISUALIZATION

VISUALIZATIONS IN NEWSPAPER

Storytelling through data visualization has gained acceptance. Newspapers are one of the platforms that have witnessed this change profoundly, from a complete text-supported information forum, to graphical and pictorial representation of the stories. They portray complex data in a more comprehensive and engaging tone through the use of visualization. Annotation, in addition to the graphics, is used to highlight the story provide additional context.[7] Thus, the newspaper draws the reader's attention, instigates curiosity in the featured story, and generates an impact on the minds of the reader.

The usage of data visualization in media is has been seen in the US since the 1930s. *Fortune* magazine's appealing visualization garnered attention. "The colored weather forecast graphic" designed by George Rorick in 1982 for the newspaper *USA Today* is accepted as the beginning of a new period for infographics in newspapers. He is considered as one of the pioneers of news graphics and honored with Lifetime Achievement Award in 2005.[8]

VISUALIZATION ASPECTS

The style of visualization is dependent on the medium. Nevertheless, prior to sketching visualization, a large amount of time is necessary to handle the data. First, the acquired raw data is processed, simplified, and normalized in order to be able to draw patterns or story from it. Classification of the data and ordering it based on its importance is an integral part of data analysis. Choosing the right kind of visualization to narrate the story is the next step. The right usage of graphical elements, combined with colors and annotations is of utmost necessity in order to bring out the story.

DEPICTING THE PERCENTAGE OF CRIMINAL CANDIDATES

In this paper, visualization of the 2013 Delhi Assembly elections for print media is undertaken. The number and percentage of elected candidates who were accused of crimes is the focus the presentation. Unfortunately, the connection between politicians and criminals has assumed alarming association in the country. Three major

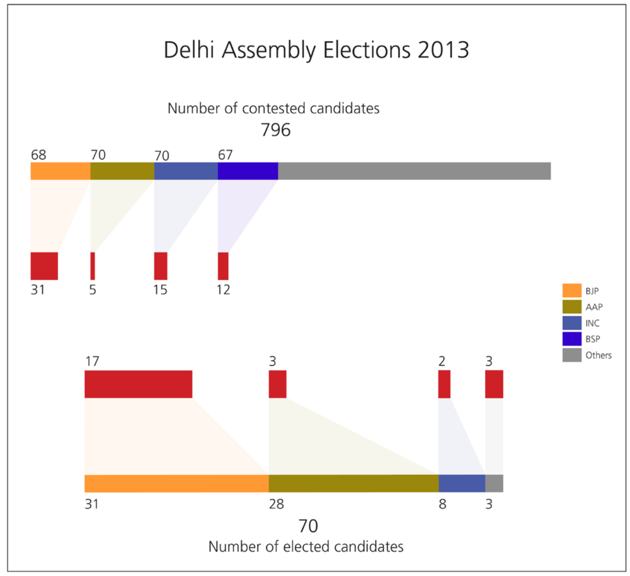

FIGURE 4: *The first iteration of the election visualization.*

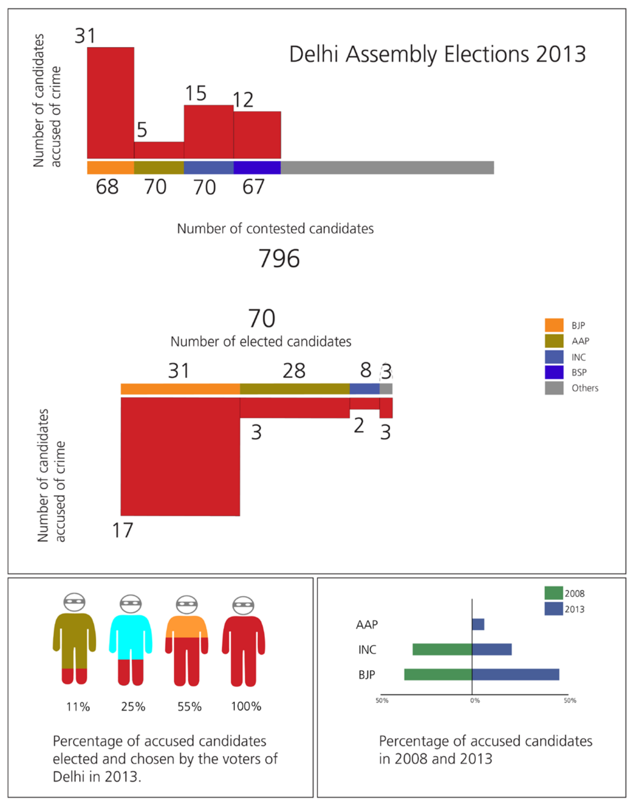

FIGURE 5: *The second iteration of the election visualization.*

parties: Aam Aadmi Party, Bhartiya Janata Party, and Indian National Congress and the percentage of accused candidates in both the 2008 and 2013 elections indicates the rate of this nexus. Twelve parties (including independent candidates) contested for the Delhi Assembly elections for a total of 796, out of which 129 (16%) were accused of crime. Out of the 129, 93 (12%) have been involved in serious criminal cases.[9] Serious criminal cases are those which involve: offences for which maximum punishment is of five years or more and is non-bailable, electoral offence (171E or bribery), corruption, murder, kidnap, rape related, and crimes against women.

The upper half of the visualization pertains to number of contested candidates with respect to accused candidates from their respective parties, and the lower half depicts the number of elected candidates with criminal accusation. Party color is used as an identification factor for each party. The linear representation of the parties is to give an overview of the distribution of seats by each. In this visualization, the use of red is deliberate as it emphasizes the number of criminals in the ministry.

The first iteration shows the number of criminal candidates, both contesting and elected. Varied width of the block shows the total number of criminal seats. In the second iteration, the number of candidates contesting from a particular party is represented horizontally, and the number of accused candidates is indicated through vertical lines. The use of red vertical blocks symbolizes the number of politicians with criminal background.

Pictorial representation of the criminal offense is also included. The major parties that won the seats in Delhi Assembly Election are illustrated with the human icon, providing an overall picture. On the bottom right, the comparison is made between Delhi assembly elections held in 2008 and 2013 in order to show the rise or fall in the total number of criminals in Aam Aadmi Party, Bhartiya Janata Party, and Indian National Congress. An interesting aspect to be noted is the decrease in the number of people in one party and increase in the other.

Many members of the national parliament or state assemblies have been indicted with serious charges, including murder. Between the criminal and non-criminal groups, a statistical report stated the chances of the accused candidate winning an election are 23% and that of the latter is 12%, as candidates possess electoral advantage. This has attracted official attention with the appointment of an independent commission to analyze the phenomenon and suggest remedial measures.

3 PHYSICAL DATA VISUALIZATION

One of the other approaches for data visualizations is creating physical, tangible data visualizations. Readily available data can be visualized in various forms and techniques. However, the key differentiating factor in selecting the form of the visualization is its ability to be grasped by the viewer with a minimum number of annotations added. Conventionally, data visualizations use a 2D graphical approach. The reason could be the ostensible ease of creating these visualizations on-screen using software tools. Physical spatial visualizations tend to grab the viewer's attention enough to invoke curiosity about the data. This curiosity serves as the base for the viewer to grasp the data from the visualization. Secondly, the choice of the media/artifacts for the visualization further complements in communicating the domain of the visualized data. This results in effective communication of the data to the viewer.

The earliest data visualizations were in fact, likely physical, built by arranging stones or pebbles, and later, clay tokens. Pierre Dragicevic and Yvonne Jansen have compiled a list of physical visualizations chronologically dating back to 5500 BC.[10]

VISUALIZING NOISE LEVELS ON THE CAMPUS

With the data for the visualization not being readily available, it had to be manually accumulated. For simplicity, only the ground floor of the building was taken into consideration as most of the active areas viz., workshop, prototyping lab, conference hall, meeting rooms etc., are on the ground floor.

The intended visualization would be a spatial visualization of noise levels at a given time of the day. Hence, to ensure there was little or no time variance factor in the captured noise levels, the data-capturing time window was kept as short as possible for the entire process. A 90-minute window starting from six PM was fixed considering the activities in the areas of interest. A single noise measurement instrument was used for the data capturing in the form of an android application called Sound Meter on a smartphone.[11]

For mapping the noise-level readings, an outline map of the ground floor along with the surrounding areas was created. It was overlaid by 8x7 grid, which would divide the space into 56 sectors.

The noise was measured twice in each sector for a span of 30 seconds each, and the mean value was taken as the final measurement for that sector. The process extended to a little over an hour. The generated data map was then digitized using Adobe Illustrator software. Each sector was

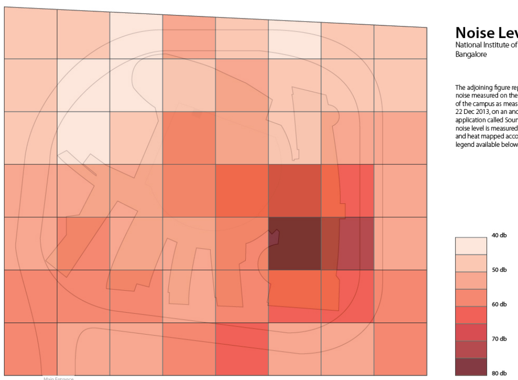

FIGURE 6: *The use of a simple choropleth system over a gridded floor plan provides an interim digital visualization*

mapped with the corresponding noise level using a color legend. FIGURE 6 shows the interim digital visualization of the captured data.

For the actual physical visualization, only five data points were chosen for the representation that corresponded to the most significant areas in and around the campus.

They were:

1) Street outside the campus

2) Prototyping Lab

3) Meeting Room

4) Workshop

5) Design Studio [Design for Retail Experience]

The physical visualization is essentially a tangible installation which is able to communicate the represented data in multiple forms viz., visual and audible. The extra sensory element invokes the curiosity and eagerness for the viewer to grasp the data. Hence the recall capacity of the viewer is increased.

Aluminum pipes are used for the construction since they are readily available at local hardware stores and also produce more noise when striked, compared to steel or plastic counterparts. The length of each pipe corresponds to the noise level captured in dB at that point on the map. The mapped data is as seen in FIGURES 7–9.

The base for the visualization was made out of five-millimeter thick foam board. The map of the institute was laid onto it in such a way that each data point appears

LOCATION	MEASURED NOISE (DB)	PIPE HEIGHT (mm)
Street outside the campus	*53*	*106*
Prototyping Lab	*78*	*156*
Meeting Room	*41*	*82*
Workshop	*60*	*120*
Design Studio	*48*	*96*

FIGURE 7: *In table form the information is accurate and specific, but unlike the diagram, does not yield a rapid contextual understanding.*

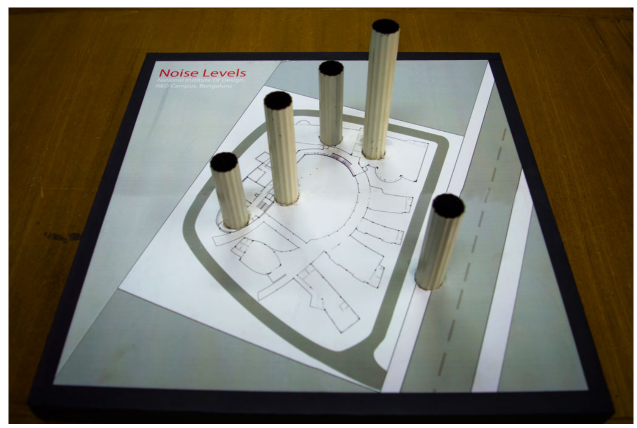

FIGURE 8: *The noise data was converted into a physical model through the use of aluminum tubular sections; each can produce a sound value if struck.*

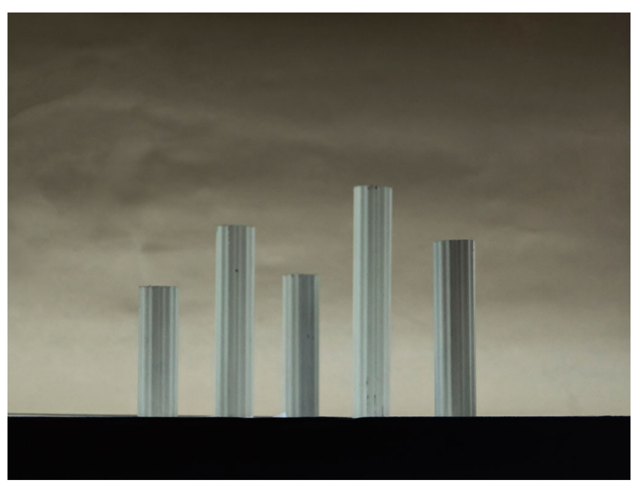

FIGURE 9: *An elevation view of the model represents a bar graph.*

to be equidistant from the adjacent data point when viewed from the front side of the installation. The pipes were then fixed at the respective locations on the base.

When viewed from the front side of the installation, the viewer can see a bar graph formed, which provides an accurate representation of the noise levels. A small hammer was used to gently strike any of the pipe to give an audible measure of the noise level at that point. This way the data was communicated in multiple forms, as well as maintaining the fun element of the visualization.

4 UNDERSTANDING TIME USE SURVEY STATISTICS

A time-use survey is a household data survey which aims to record and report data on average time spent by people on various activities. It provides an opportunity to measure the choices and the routine of daily life. It offers a unique tool for exploring a wide range of social concerns such as the division of labor or allocation of time for the household work, among other tasks. Every day, all activities that occur within the one-hour interval are recorded as is total time spent on each activity. The deluge of data sources with various categories necessitates the use of software-based tools to fill the gap and enhance the understanding and analysis. The multi-dimensional data is made comprehensible using interactive data visualisation.

The time-use survey in India was conducted from July 1998 to June 1999 by the Directorate of Economics and Statistics of the six participating states: Madhya Pradesh, Gujarat, Haryana, Orissa, Tamil Nadu, and Meghalaya for 18,591 households.[12] The field survey was spread over a year to account for seasonal variation in the activity pattern of the individual. The data set was available in multiple files with over 1,000,000 records in the data source. The data available in multiple data sets was combined, normalized, and filtered using MS Excel and MS Access. The visualization is generated using Tableau as platform.

WHO IS HAPPIER?

One of the main objectives of the survey was to collect data to study the gender discrimination in the household activities. Women's freedom, (especially married women) within both the family and market sphere, has expanded over the past 35 years. However, the measure of women's subjective well-being has fallen both absolutely and relatively to that of men based on the research paper "Paradox of Declining Female Happiness" by Betsey A. Stevenson & Justin Wolfers, University of Pennsylvania. Ambiguity in division of household responsibilities between couples often results in ongoing negotiations, resentment, and tension. This raises the concern of declining women's freedom and independence and thus the happiness after marriage. Happiness and contentment are exceedingly hard to quantify, but with the Time-use survey data we can derive insights that come close to reality.

Analysis of time-use survey data in six representative states of India helped to prove the hypothesis "Unmarried women have freedom to spend more time on activities of their choice compared to married women." Though the statement posed involves two categories of women (married and unmarried), the third category, men's activities, helped in analyzing them. The visualization was created for average time spent (in hours) per day by women and men (married/unmarried; well educated/less educated/non literate) on activities related to education/learning, entertainment, working outside, personal care, and community service. More time spent on these activities translates to freedom and independence which in turn translates to happiness; family care and household activities when shared among women and men leads to happiness, ignoring the exceptions.

INTERACTIVE VISUALIZATION

This visualization was created using a combination of standard representation. Tableau software allowed connecting data, visualizing and creating sharable, and interactive dashboards. It provided scalability and in-memory analytics by loading the data into computer's memory and allowing the user to slice, dice, and analyze data in real time. The data demanded easy interactivity and to exploit the familiarity of the readers considering the limitations of the tool. The challenge of the visualization was to show the underlying data without distortion.

The bubble chart in FIGURE 11 shows the high level of information of member participation in decision-making. Married women participate the most and are made to feel part of decision making in the family. Unmarried women have the least say in the family.

The tree chart in FIGURE 11 shows a compact visualization of the directory determining which member consumed the largest share of tasks.

The visual structure of the bar graph matches much of the data and makes it simple to understand. Other relationships of the structure of data can be indicated by grouped and stacked column charts. It helps in comparing elements across categories and in comparisons between each element in the categories.

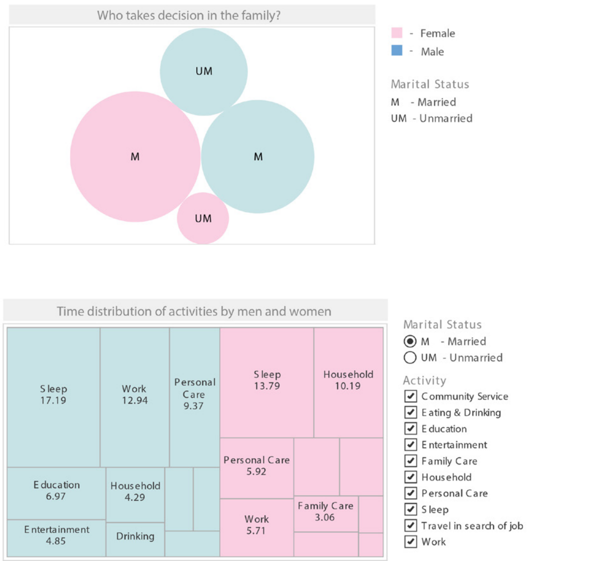

FIGURE 10: *Bubble chart and tree map models representing the scope of decision-making as dependent on married/ unmarried status male/female within families.*

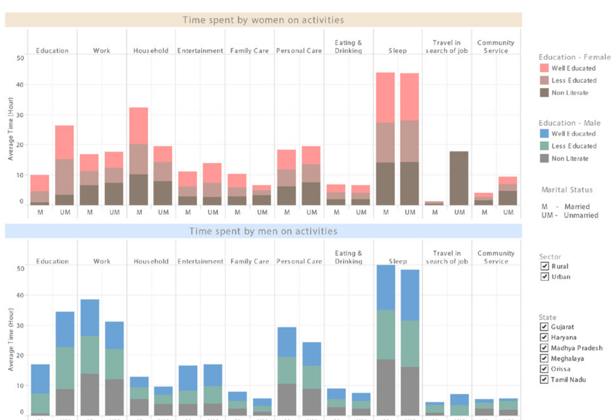

FIGURE 11: *Bar charts comparing activity levels under varying segments for married/unmarried, and male/female cohorts.*

The bar chart on the next page [FIGURE 10] has two parts: women and men (unmarried/married; well educated/less educated/non literate) spending time (in hours) on various sets of activities. One common, although unsurprising trend can be seen with married men; they spend greater time on entertainment, personal care, and work outside home compared to women. Unmarried women spend more time in learning, working outside house, entertainment, personal care, and community service compared to married women. Married women spend significant time in household activities compared to unmarried women. The anomaly here is that illiterate unmarried women spend a surprising amount of time travelling in search of job within the urban sector. Interactive visualizations as seen in the above statistical survey help deduce insights from multidimensional data set.

The interactive visualization can be found here: https://public.tableausoftware.com/profile/sindhu.rao#!/vizhome/HappinessQuotientSingleMarriedWomen/NewsPaper

CONCLUSION

The above illustrative examples depict a few examples of the many ways data can be visualized. A visualization can be considered successful when its interpretation meets its core intention (author/creator intent). According to Cleveland, "When a graph is made, quantitative and categorical information is encoded by a display method. Then the information is visually decoded. This visual perception is a vital link. No matter how clever the choice of the information, and no matter how technologically impressive the encoding, a visualization fails if the decoding fails."[13] The four examples follow the iterative design process of data acquisition, data processing, design explorations, and final visualization. Explorations help to match the intentions behind encoding and decoding. It helps the visualizers to make decisions on the functional and aesthetic design issues and finally choose appropriate means of visualization which effectively represent the data. The different types of explorations including wordless diagram, static, physical, and interactive visualization gives the visualizer the freedom to devise novel methods of representing the data.

ACKNOWLEDGEMENT

This paper is written under the guidance of Professor Venkatesh Rajamanickam.

BIOGRAPHY

The authors are currently pursuing Information & Interface Design at National Institute of Design, R&D Campus, Bangalore. They have a background in Engineering and significant experience in software industry. As part of their course module, they have worked on various projects in the area of Information Visualization. Their fields of interests are: information visualization, information architecture, and design research.

NOTES

1 *Design: The Journal of the Society of Newspaper Designers*. Issues 94-97 (1 Sep 2009):52

2 Barbara Tversky, Julie Heiser, Paul Lee, and Marie-Paule Daniel. "*Explanations in Gesture, Diagram, and Word*," Spatial Language and Dialogue (Oxford: Oxford University Press, 2009), 119-131.

3 Volker Blanz, Michael J. Tarr, Heinrich H. Bulthoff, Thomas Vetter. "What Object Attributes Determine Canonical Views?" *Perception 28* (1999): 575-599.

4 Raymond D. Frost, Sarah Pels. "Nearly Wordless Techniques for Software Instruction: Increasing Comprehension and Higher Level Learning." *Issues in Information Systems* 11, no 2 (2010): 191-198.

5 Pan Y. "Attentional capture by working memory contents." *Canadian Journal of Experimental Psychology* 64, no. 2 (2010): 124-128.

6 "Infographic Design: A Framework," schroeppel-design.de

7 Edward Segel and Jeffrey Heer. "Narrative Visualization: Telling Stories with Data" *IEEE Transactions on Visualization & Computer Graphics* 16, no. 6 (2010): 1139-1148.

8 Banu Inanc Uyan Dur. "Analysis of Data Visualizations In Daily Newspapers In Terms Of Graphic Design" *Procedia-Social and Behavioral Sciences* 51 (2012), 278-283

9 "Analysis of Criminal Background, Financial, Education, Gender and other details of Candidates contesting in the Delhi 2013 Assembly Elections," Nov. 2013, ADR Press release

10 "List of Physical Visualizations and Related Artifacts" http://dataphys.org/list/

11 "Sound Meter Pro" Google Play https://play.google.com/store/apps/details?id=kr.aboy.sound

12 "India Time Use Survey: 1998", Ministry of Statistics & Programme Implementation. http://mail.mospi.gov.in/index.php/catalog/130

13 W. S. Cleveland. *The Elements of Graphing Data, Revised Edition*. (New Jersey: Hobart Press, 1994).