Data Visualization In Review

Part of the Strategic Evaluation on Communicating Research for Influence

Evaluation managed by
Jacqueline Strecker, Research Award Recipient IDRC;
& Tricia Wind, Evaluation Officer IDRC.

Report author: Jacqueline Strecker
Feature review by: Amanda Cox

Winter 2012

Introduction

By Jacqueline Strecker

There has been a recent surge of interest in data visualizations and their potential to communicate effectively. This rise has been influenced by the increasing availability of tools for creating visualization, and the spike in data visualization use. News outlets around the globe have been at the forefront of this charge, experimenting with unique and appealing ways to present information to the public. Their investments and dramatic outputs have provoked questions amongst other industries about what benefits data visualizations could contribute to communicating information more effectively.

While in the past data visualization was viewed as an important analytical tool for researchers, it is quickly being recognized as an essential aspect of effective research communication. Although data visualization is fairly new for development researchers, it affords opportunities to both transform and display data (Lindquist, 2011). Visualization proponents also highlight that these capabilities are extremely useful within complex and changing environments, which are akin to the contexts surrounding IDRC-supported projects. As Evert Lindquist argues,

> *visualization techniques loom as potentially important sense-making, analytic and communications tools for capturing and addressing complexity. The promise is that, if properly chosen and calibrated, they can show the breadth and evolutions of problems and interventions, permit more detailed explorations of facets and strands, as well as how these facets and strands link to the whole (Lindquist, 2011: 3).*

The importance of data visualization is further heightened by the increasing digitization of the world, which has created information-overloads in a time-deprived policy and development sector. One of the strengths of IDRC-supported projects has been the under-researched regions and fields that are explored. However, while the collection of this information is often ground-breaking and innovative, the expounded findings still have to be heard within saturated information markets. The utility of this research is therefore dependent on how it is communicated and the level of interest and investment from stakeholders and policymakers.

This study assesses the potential of data visualization to assist in effectively communicating research for influence. Section 1 of this review, provides an overview of the data visualization field; highlighting the rich history, and scientific rational supporting visualization use. This section also provides a review of current trends, and good practice techniques, as articulated by the leading scholars and practitioners. Section 2 situates IDRC within the greater data visualization landscape by assessing how IDRC-supported research has utilized data visualizations, and to what effect. The final section of this report, Section 3, provides further information on resources, good practice guidelines, and general recommendations from IDRC staff, supported partners, and the leading data visualization proponents on how to ensure data visualizations are used effectively.

Table of Contents

Introduction 2

Section 1: 4

Defining Data visualization 4

Data Visualization as a Field 4

Calibrating Visualizations for Success 9

Current Trends and Interesting Studies Shaping the Field of Data Visualization 17

The Risks of Data Visualization 18

Section 2: 21

Stage 3: External Data Visualization Review 27

Models for Success 40

Conclusion 49

References 50

Section 1:

Assessing the Field

Defining Data visualization

The definition of data visualization is far from simple, since the term and its corresponding synonyms lack clear distinction and agreed-upon definitions. Terms like data visualization, information visualization, and infographics have also been used interchangeably, despite arguments for clear distinctions. During a recent discussion on *Quora,* designers, practitioners, and scholars debated this topic with several proponents maintaining that the terms are in fact synonyms. In contrast, others sought to distinguish data visualization by its clear linked to raw data. Others emphasized that data visualizations were general, context-free renderings; as opposed to information visualizations which were argued to presented editorial representation of data. While the debate still rages, this study explicitly adopted the term data visualization, because of the reference and emphasis on word *data*; which lies at the heart of IDRC-support research. The use of data visualization within this report is therefore not limited to the display of raw data sets, but rather all static and interactive visual representations of research data, which may include infographics.

Data Visualization as a Field

Although data visualization has only recently been recognized as a distinct discipline; it has deep roots, dating back to the 2nd century cartographers and surveyors. The early origins of data visualization can be traced to the ancient Egyptians surveyors who organized celestial bodies into tables to assist with the laying out of towns and the creation of navigational maps to aid exploration (Friendly, 2006). It was only during the 17th century, when French philosopher and mathematician, Rene Descartes developed a two-dimensional coordinate system for displaying values along horizontal and vertical axis, that graphing began to take shape (Few 2012). During the late 18th century, Scottish social scientist William Playfair changed the field of visualization, by pioneered many of today's widely used visualizations – including the line graph and bar chart (Playfair, 1786), and then later the pie chart and circle graph (Playfair, 1801).

During the 19th century there was a radical increase in the use of statistical graphics and thematic mapping, which according to Michael Friendly, occurred at a rate which has not been matched until modern times (Friendly, 2006, 9). It was during this period that all modern forms of statistical graphs were invented including: pie charts, histograms, time-series plots, contour plots, scatterplots, and many more. Scholars were also experimenting with thematic cartography, in order to display an array of economic, social, medical, and physical data (Friendly, 2006). Part of the increasing push for data visualizations, was caused by the establishment of state offices throughout Europe, which were utilizing numbers in social planning, commerce and transportation. The popularity and support for visualizations during the 19th century was regarded as an Age of Enthusiasm, but was quickly followed by what Friendly claims to be the Golden Age, with "unparalleled beauty and many innovations in graphics and thematic cartography" (Friendly, 14).

Today, the world is experiencing another surge in data visualization popularity. This interest can be partially linked to the increased availability of new technologies and software products which enable every user to dabble in the world of visualization. However, these resources have not come about on their own, but have been the by-products of years of research and development from an international community of scholars and practitioners. In a recent publication Evert Lindquist examines visualization by parsing the field into three unique disciplinary streams: information visualization, graphics and information display, and visual facilitation for thinking and strategy (Lindquist, 2). While each of these streams is distinctive in both their approach and focus; there are larger overlaps which undercut any hard fast boundaries. That said, exploring these three streams provides a stronger understanding of the rich and diverse scholarship which has contributed to the field of data visualization[1].

Graphics and information display is the first stream of the visualization, which focuses on the aesthetics of displaying information graphically, rather than enabling the data to determine the form. Lindquist's summary of this area highlights that there is an astounding diversity of approaches; covering everything from designing algorithms to enable visualization production, to understanding cognitive interpretations of different graphical forms, to exploring the applications and theoretical constructs of data visualization (Lindquist, 2011). Overall, what unites the divergent approaches of this stream is a concentration on the design of visualizations and how form can strengthen utility for purposes of communication, marketing and illumination.

Information visualization ('InfoVis') is arguably the newest of the three streams, emerging towards the end of the 1990s. It was motivated by the desire to represent increasingly large amounts of data and was influenced by computing, graph-making, and informed by the findings of the other streams. Proponents of InfoVis do not make a firm division between information and scientific findings, but rather focus on visualizing all forms of data (Lindquist, 5). The main thrust behind InfoVis is to aid human cognition by transforming abstract data into visual-spatial forms to amplify human intelligence (Shneiderman, 2004). Research has focused on a variety of areas including different ways to distil statistical or graphical data efficiently; approaches for automating the transformation of data into graphical representations; and ways for facilitating the exploration of data within different data streams; and large datasets (Lindquist, 2011). Although this area is still fairly young, there is a wide collection of publications, journals, conferences, and university courses, which have sought to distinguish and build information visualization as a unique field.

---

[1] For a more detailed analysis of these three streams please see: Evert Lindquist...

[2] Each document was initially coded for the following: document type, presence of visualizations, presence of tables, visualization category, and number of visualizations.

Facilitation & strategic thinking is the last stream of visualization and focuses largely on the engagement of users, rather than on the data or display. There has been a growing practitioner community which has become connected through the International Forum of Visual Practitioners (IVFP). The focus of this area is to use visualizations as a tool for facilitation, assisting groups to interact and understand one another, and/or challenges from a different viewpoint (Lindquist, 3). Within this field, there has also been a recent and growing interest in the role of visualizations in approaching issues of complexity or systems thinking, which is discussed in the policy section of this report.

Overall, identifying these three streams as unique areas provides a set of useful entry points for understanding the various aspects of the data visualization. However, it is also important to remember that while Lindquist distinguishes between these streams, he is also quick to highlight how they overlap.

> *All of the domains seem equally inspired by [Edward] Tufte's work and earlier efforts at mapping and drawing; all see visualization as having great promise as a superior way to render information for illumination and decision-making; and all try to balance and improve the aesthetic and practical qualities of visualization, albeit in varying ways (Lindquist, 21).*

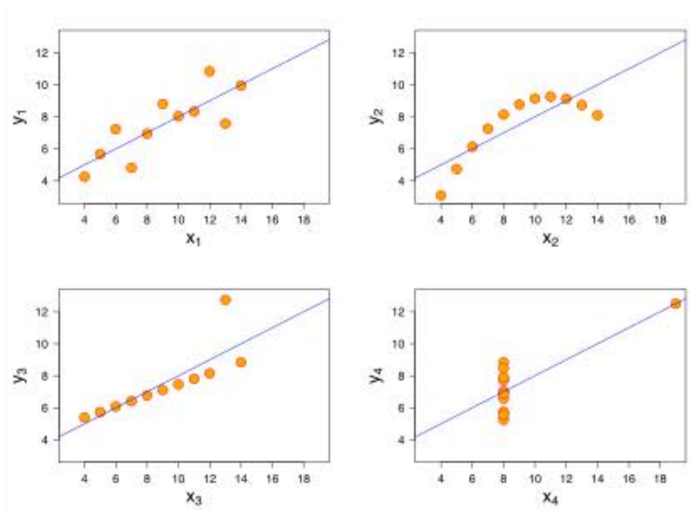
Other scholars have adopted alternative method for distinguishing approaches to data visualization based on those which are focused on exploring data or explaining data (Steele & Iliinsky 2011), a distinction which cuts across all of Lindquist's three streams.

Exploring data through visualization was popularized in 1973 by statistician Francis Anscombe, who designed *Anscombe's quartet*, a series of four datasets with identical means, modes, and averages (figure 1). Anscombe used these datasets to demonstrate the radical difference of these datasets when graphed, as demonstrated in figure 2. This example provided an important rationale for the importance of graphing data before analysing it.

Figure 1: Anscombe's quartet.

Anscombe's quartet

I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89



In 1977 Princeton statistics professor John Tukey, further demonstrated the exploratory function of visualization, by introducing *exploratory data analysis,* an innovative methodology which focused on visualizations to structure analysis (Few, 2007). In the last thirty-five years significant scientific and technological advancements have enabled visual data analysis to take on new forms, and enter into different fields. Recent projects and studies have employed visualizations to examine everything from migration patterns of birds (Ferreira et al. 2011), biases in the thematics of grant application (Dou et al., 2011), and even perform risk assessments on potential coastguard station closures (Malik et al., 2011).

While some IDRC-supported projects have started to experiment with visualization analysis, the majority have utilize visualizations to explain data. This means that data visualizations are employed as tools for communicating findings to a targeted audience. Visual science has demonstrated that data visualizations are particularly effective in communicating or explaining data to an identified audience, if the visualizations are calibrated correctly to draw on the brains ability to detect certain properties. If visualizations are properly designed they cannot only increase the speed at which data is comprehended but can also increase the retention of data. It is for this reason that proponents of data visualization draw heavily on visual perception scientists who maintain that visualizing data is typically more effective for communicating information than using text-based renderings (Lindqust, 3). This is because data visualizations shift the balance, between seeing and thinking, towards greater use of visual perception, taking fuller advantage of the brains abilities (Few, 2012).

Visual perception utilizes the eyes, a channel which has one of the largest bandwidths to the brain (Kosara et al., 2002). The eyes transmit information from 100 million receptors through a million fibres in the optic nerve (Ware, 2004). Visual details are registered at greater detail from the very center of our visual field, as appose to the periphery. In the center, the eye can resolve about 100 points at the edge of a pin (held at an arm's length away); whereas at the edge of our visual field objects need to be the size of a head to be registered (Ware, 2004). Ware highlights that:

> *The non-uniformity of the visual processing power is such that half our visual brain power is directed to processing less than 5 percent of the visual world... non-uniformity is also one of the key pieces of evidence showing that we do not comprehend the world all at once (Ware, 6).*

Instead of perceiving the entire visual field in a single glace, the eyes are thus forced to move and scan throughout this field, refocusing and registering different details. This information is distilled in the visual cortex, which is extremely fast and efficient, as compared to the cerebral cortex, which is slower and is largely used for other cognitive tasks (Few, 2012).

In his 2004 publication, *Visual Think: for Design,* Colin Ware highlights that visual perception (the process of seeing and interpreting) involves two types of processes:

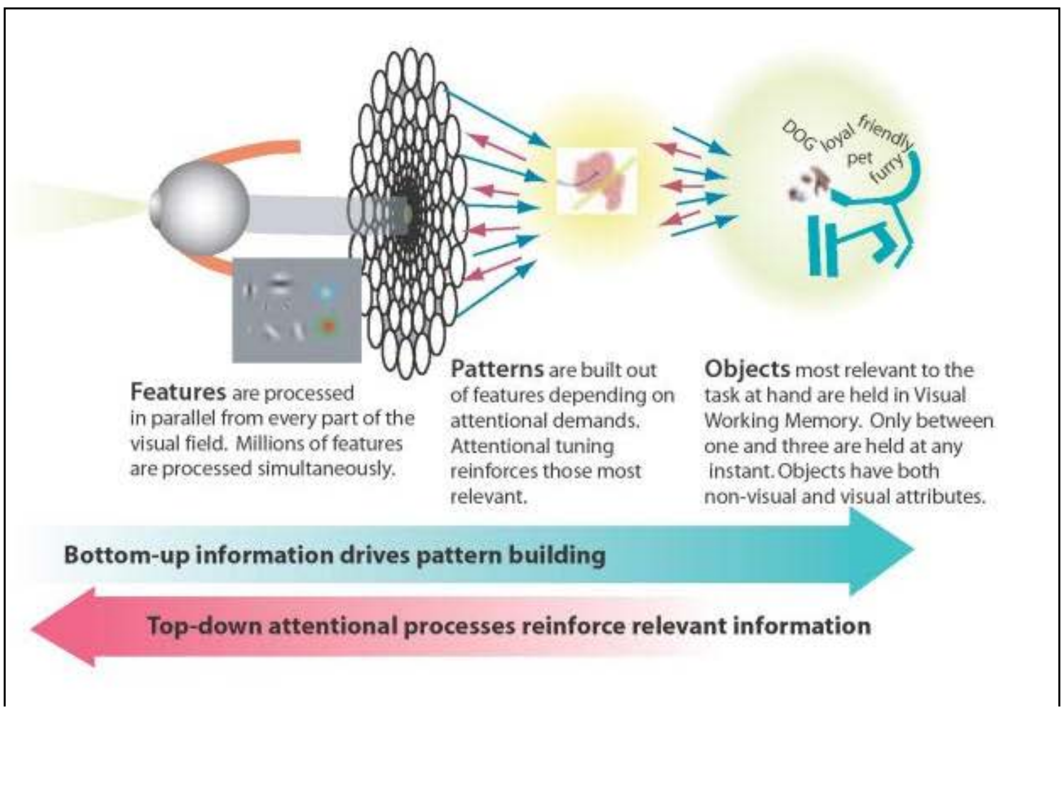
- **Bottom-up processing** - driven by the visual information in the pattern of light falling on the retina
- **Top-down processing** - driven by the demands of attention, which in turn are determined by the needs of the task
  (Ware, 8).

Bottom-up processing occurs in three main sequential stages: parallel processing; pattern perception and sequential goal-directed processing. The first stage, parallel processing involves the extraction of orientation, colour, texture and movement from our field of vision. This occurs rapidly and without conscious thought. This is why Ware notes that “if we want people to understand information quickly, we should present it in such a way that it could easily be detected by these large, fast computational systems in the brain” (Ware, 21). The efficiency of this process is related to the large amount of neurons, up to five billion, which are simultaneously processing different features.

In the second stage of processing, both the working and long-term memory are engaged; partitioning the detected features from stage one into regions and simple groupings or patterns. This stage is influenced by both the information acquired from stage one, as well as from top-down attention driven inquires (Few, 2007). The third and final stage is where a small number of visual objects are distilled through the previous pattern-processing stage. At this level, objects are temporarily stored within the short-term memory for quick recall and processing, however only a small amount of data can be held in attention at one time (Ware, 2004). This stage is also influenced by goal-oriented processing, based on directed or stimulated questions. “We see something that catches our interest and provokes a question, which we pursue by searching through the patterns in our visual field (a visual query) to satisfy our interests and answers the question” (Few, 3).

Within each of these three stages top-down processing is also influencing and directing our attention in pursuit of a specific goal. This influences our perception in two main forms. First top-down processing creates bias during low-level pattern analysis. If one is trying to detect red lines, then these detectors transmit a louder signal to the brain (Ware, 13). Ware highlights that “this biasing in favor of what we are seeking or anticipating occurs at every stage of processing. What we end up actually perceiving is the result of information about the world strongly biased according to what we are attempting to accomplish” (Ware, 13). Top-down processing also directs our eye movements based on a ‘just-in-time strategy’ where observers acquire the specific information they need at the point it is required to accomplish the task (Hayhoe & Ballard, 2005). Since we perceive the world through just-in-time visual queries, effective data visualizations should anticipate and provide answers to the cognitive

Figure 2: Colin Ware Illustration of Visual Processing (Ware, 2004).



tasks the graphic is intended to support (Ware. 14).

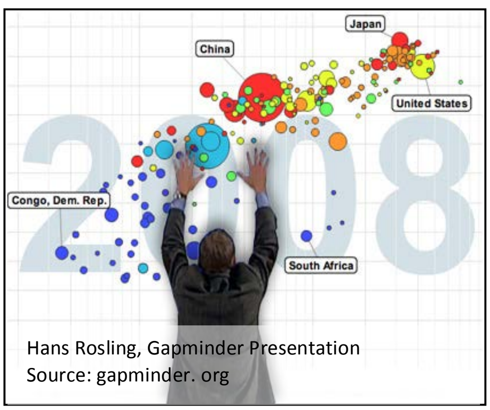
Understanding these three stages of visual processing is vital to ensuring that visualizations are designed to be readily and rapidly decoded by the human brain. If designed well, visual processing allows viewers to process abstract information with a fullness and speed not afforded from textual-renderings. This is why the prominent data visualization author, Stephen Few highlights that visual perception is a gift, which should be nourished and not ignored (Few, 2007). Part of being able to nourish these abilities, relies on our understanding how to calibrate visualizations so that they can be effectively decoded by the brain.

Calibrating Visualizations for Success

Although the science of visual processing suggests a clear potential of data visualization to enhance communication, some proponents caution that creating effective visualizations involves care and precision. In other words, just like verbal language, visual communication depends on semantics and syntax, and it is therefore important to understand the rules in order to communicate effectively (Few, 2007). Edward Tufte's, *The Visual Display of Quantitative Information* was the first publication of its kind to highlight the difference between graphing data, and effectively visually it. In this 1983 publication, Tufte highlighted that the majority of individuals using visualization were doing so poorly (Tufte, 1983). Today, scholars like Stephen Few have echoed this criticism further emphasizing that the availability and rapid expansion of visualization use, has been accompanied by a general lack of understanding. Few argues that this lack has often undermined the potential benefits of data visualization, and many of the current visualization trends "are actually producing the opposite of the intended effect, confusion rather than understanding" (Few, 2007:2). This is why practitioners emphasize the importance of becoming literate in visualization techniques (Lindquist, 11).

While there is still debate over certain principles to data visualizations, some rules are general accepted as good practice. Noah Iliinsky, co-author of *Designing Data Visualizations* states three general guidelines for strong visualizations. These include understanding your data; understanding what you want to show; and understand the format of your visualization and its strengths and limitations (Iliinsky, 2011).

Understanding Your Data



Hans Rosling, Gapminder Presentation
Source: gapminder. org

Knowing your data is vital for effectively communication, be it through text, presentations, or data visualizations. The quality of a data visualization is contingent on the strength of the data and the analysis underlying it. Hans Rosling presentation of Gapminder is a good example of how important it is to know your data. Although the enthusiasms which underpins Rosling's presentations is contagious, and akin to the pace of a sportscaster, his presentations are captivating because of the degree to which he knows his data.

One of the most important dimensions of understanding data is acknowledging the relationships or patterns in a datasets. At a broad level, data can be classified as either discrete or continuous. Discrete (or nominal) data represents separate items which have no intrinsic order in relation to one another (e.g. apples and oranges); as oppose to continuous data which specifies a particular ordered pattern (e.g. temperatures, days of the week, income brackets) (Whitney, 2011). Visualization conventions infer that these types of data are displayed differently to ensure that their relationship is easy to identify. For example, if you are representing continuous data which is connected chronologically, forms such as timelines, line graphs or family trees will help viewers quickly acknowledge this relationship. Discrete data on the other hand could be graphed using nominal scales or ordinal scales; for example a pie chart displaying the percentage of people who prefer apples to oranges.

Further distinctions of data types have been made in recent years, the culmination of which was a 1996 publication by Ben Shneiderman, which outlined seven different kinds of data: one-dimensional, two-dimensional, three-dimensional, temporal data, multi-dimensional data, tree data, and network data (Shneiderman, 1996). While some of these titles provide clues for graphing options, it is also important to identify any patterns within a dataset. Nathan Yau highlights that patterns can be found in aggregates that help you compare groups, people, or things. Or they can also be derived from observing changes over time; or over geographical regions (Yau, 20122). Understanding the patterns and relationships of the data will also assist in identifying what important data you want to highlight for your viewer.

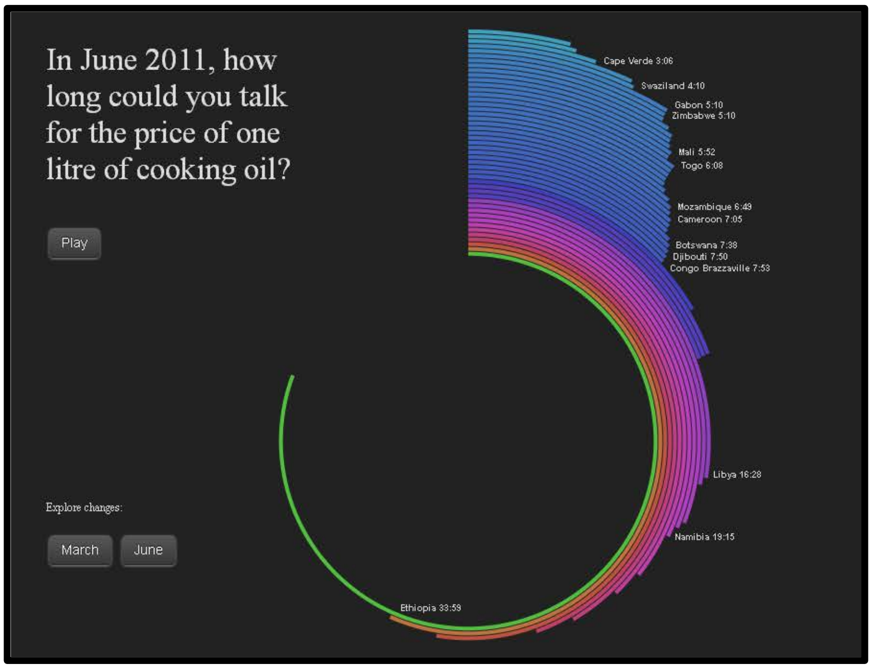
Understanding What You Want To Show

It is important to know who you are communicating to, and what you want to communicate. The audience and purpose of the visualization should always be top of mind when considering what you want to visualize. Although, some researchers have a habit of producing “visual data-dumps” to show their work; presenting unfiltered entire renderings of datasets, is often overwhelming for viewers. As one IDRC-supported partner commented, “the strength of data visualizations come from their appeal and their usability for readers. You often only have 5-to-10 minutes to capture the attention of a policymaker; therefore the visualization needs to have a strong focus” (Subrat, 2011). Iliinsky and Steele stress the point that it is important to consider the context of your viewer including their motivation, level of interest, and the time available (Iliinsky and Steele, 2011). If designing for a broad audience, it is best to identify the most important viewers within this group and design for them. Knowing what you want to show, as well as knowing your audience and their context, will help you decide what organizational structure to apply to your data visualization.

All data visualizations fall somewhere along the *author-driven* to *reader-driven* spectrum. An author-driven approach displays data in a specific order, includes no interactivity and includes a structured message or narrative. It is essentially like a traditional storytelling structure, where the author controls the speed, order and information provided. In contrast, a strict read-driven approach provides information without a specific narrative. It has no prescribed ordering, includes a high level of interactivity, and has little-to-no messaging (Segel & Heer, 8). While the author-driven approach is better at providing a specific message to the reader, the latter approach can create a sense of ownership for very interested, engaged, and knowledgeable audience. This is because read-driven approaches allow viewers to interact or interpret the data without guidance, arriving at their own conclusions. The danger of this approach is that viewers can misinterpret the information, which can result in misguided conclusions. This is why the majority of effective and compelling data visualizations fall somewhere in the middle of the author-driven/user-driven dichotomy. They utilize visual narrative tactics of visual structuring, highlighting, and visual transition guidance readers are guided through the visualization, while still providing a sense of control (Segel & Heer, 2010).

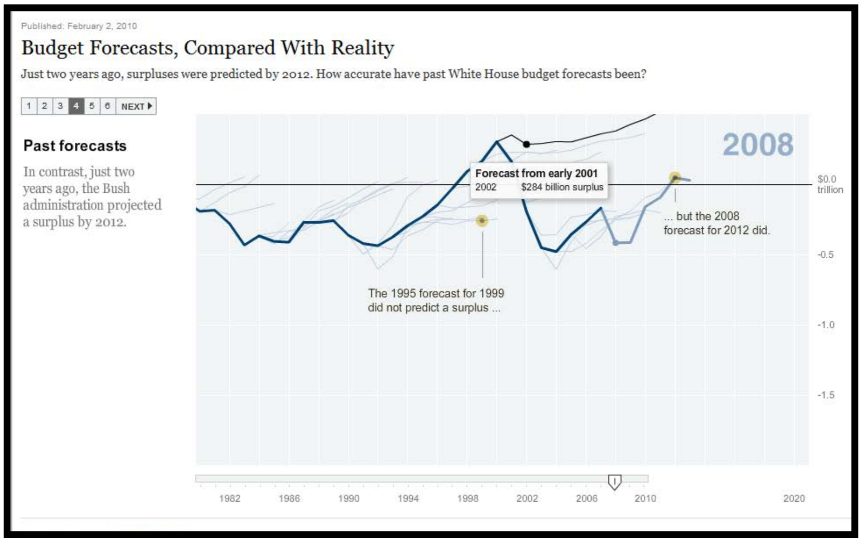
As a general principle, it is advisable to keep displays simple and allow a specific narrative to organize the information. Not all data should be of equal importance within a visualization (Segel & Heer, 2010). This is why it is important to establish hierarchies which can help communicate what information is most important and which is included to provide context. Titles can be used to also help draw attention to the main message of the visualization, and colour, size, and orientation can then be utilized to highlight data which supports this story. Using colour, size and orientation is particularly effective because these features are innately processed during the first stage of visual perception, making them readily identifiable for the viewer.

Figure 3: Fair Mobile (Cox, 2012).



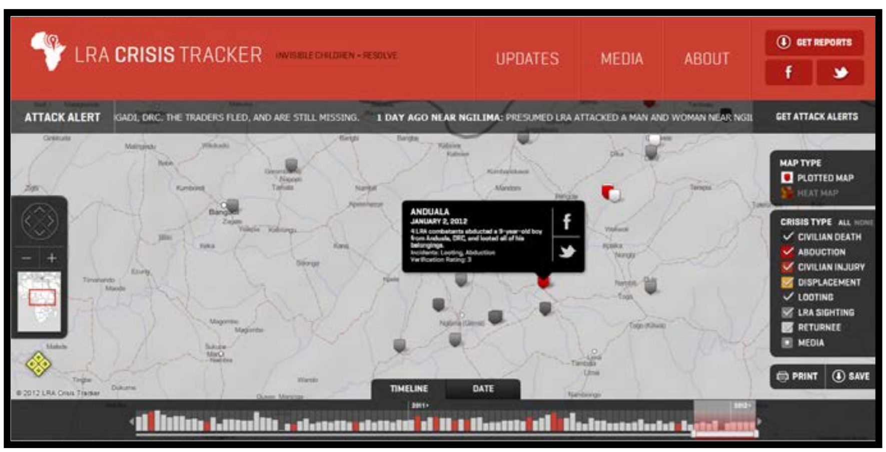
While these techniques are important for every type of visualization, knowing what you want to show, and establishing data hierarchies is of particular importance when designing interactive visualizations. Segel and Heer's 2010 empirical study on visual narrative techniques revealed three common schemas which fall within the spectrum of narrative approaches. The first of these designs is the *Martini Glass Structure*, which was found to be the most commonly used, in the Segel and Heer study. This approach provides a stricter author-driven display initially, and then once the author's narrative has completed, the functionality opens-up enabling viewers to interact with the data. This structure is said to resemble a martini glass with "the stem representing the single-path author-driven narrative and the widening mouth of the glass representing the available paths made possible through reader-driven interactivity" (Segel & Heer, 9). This type of display enables the author's initial narration to provide a frame or entry point for viewer exploration. Figure 3 presents an example of this type of interactive display, which compares the cost of mobile phone use to cooking oil. This display begins by showing an animated video of the information displaying the lowest costs of mobile phones use per country. After the video has finished readers are then able to play with the different lines to reveal further information and select different filters to display the data at different periods of time.

Figure 4: Budget Forecasts (www.nytimes.com, 2010)



The *Interactive Slideshow* is the second structure type, which provides further opportunity for reader-driven inquires. This approach utilizes a regular slideshow format to display data in truncated pieces and enables the viewer to explore particular points of interest on each slide before proceeding to the next segment. This increased interactivity makes this type of display particularly effective for displaying complex datasets, since the author can provide step-by-step guidance for the viewer, while also designating discrete boundaries between different narrative segments (Segel & Heer, 9). An example of this type of structure is New York Time's 2010 visualization of Budget Forecasts, Compared with Reality (Figure 4). This visualization provides six different slides, which animate the timeline slider, revealing new segments of the line graph and textual references. Between each slide, the animation pauses and the viewer is invited to select different line segments to reveal the forecast for that particular period; before clicking next to continue to the next slide.

Figure 5: LRA Crisis Tracker (www.lracrisistracker.com, 2012)



The last structure is the *Drill-Down Story*, which provides even more reader-driven exploration. This structure presents a general theme then allows the user to select particular data points to extract further information. While the viewer is the one controlling which stories are investigated, the structure still relies on the author to select what stories to include and what details to include for each drill-down item. An example of this type of structure is the LRA Crisis Tracker (Figure 5). This visualization indicates the location and date of attacks from the Lord's Resistance Army in the Democratic Republic of Congo. Viewers can interact with this information by specifying a period of time to look at, or by selecting a particular location to access more specific details.

Understanding these types of structures, and the different options for enabling reader-driven inquires is important when considering what you want to communicate and the overall purpose of your visualization. Striking the appropriate balance between author-driven and read-driven approaches still seems to be context and content specific. However, this study does suggest that "data stories appear to be most effective when they have constrained interaction at various checkpoints within a narrative, allowing the user to explore the data without veering too far from the intended narrative" (Segel &

Heer, 9). It is thus important to strategize about what you want to show, and the narrative you plan to tell, before deciding on what form of data visualization is best suited for conveying this story.

Understanding Form

In the past couple of years, there has been a dramatic rise in publications highlighting good data visualization practice, and providing step-by-step instructions for creating effective charts and graphs. While there are always exceptions to these rules; what is most important if for researchers and designers to understand the strengths and limitations of different formats. Below is a brief summary of the commonly used data visualizations.

Bar charts are one of the most commonly used chart types, used to depict nominal data and often used to illustrate comparisons (Yau, 2010). Bar charts should always start with the value axis at zero and use a consistent scale (either linear or logarithmic). If your axis does not start at zero, it is likely that your display will depict incorrect relationships, which can bias your data visualization (Vlamis 2010, 3). It is also important that axes are labelled and scales are clear. Multiple colours can be used to provide emphasis between different bars, but should be used to highlight patterns or relationships between barplots. Text captions can also be assigned to different bars to provided added context or information. Lastly, bars should always be depicted as two-dimensional objects. Three-dimensional bar charts add distortion since it is often not clear where the bars end.

Pie Charts are used to depict pieces of a whole. Each wedge of the pie represents a category or value; with the sum of all wedges equalling 100 percent. Many data visualization proponents, such as Stephen Few, recommend avoiding the use of pie charts all together. This is because it is difficult for humans to visually perceive areas and angles (Vlamis 2010). However, despite the stigma around pie chart use, it is accepted that "you can use the pie chart without any problems as long as you know its limitations" (Yau, 137). Part of these limitations include limiting the number of wedges which appear in the chart to under 5 slices. Pie slices should also never be arranged clockwise from smallest to largest, because the brain intuitively reads from top to bottom. This is why Don Wong from the Wall Street Journal, recommends that "it is most effective to place the largest segment at 12 o'clock on the right to emphasize its importance...the second biggest slice at 12 o'clock on the left; the rest would follow counter clockwise. The smallest slice will fall near the bottom of the chart, in the elast significant position" (Wong, 2010). Slices can also be emphasized by shading them in a different colour then the other segments to draw the eye. Pie Charts should also never be depicted as three dimensional objects since the relative size of pieces of a pie are distorted to achieve the illusion of perspective. *Donut charts* are often preferred to pie charts, because the eye is better able to distinguish the length of a segment than the overall area of a wedge. In a donut chart, the value of a segment is proportionate to the arc length, and relative to the donut's circumference (Yau, 142).

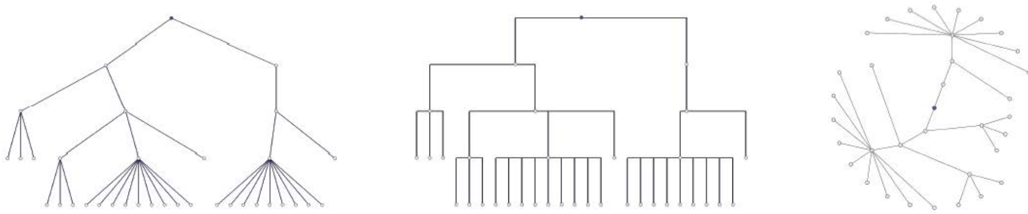
Line Charts are very effective at depicting patterns over a continuous range. Unlike a bar chart which value should always start at zero, line charts afford greater granularity by depicting any value range, without distorting the data. However it is important that data ranges are clearly marked and the overall chart retains a rectangular shape. Vlamis highlights that the "if the chart is excessively tall and narrow, the data will show an excessive amount of change. If the chart is short and wide, the change will be minimized" (Vlamis, 2010). In order to prevent data distortions it is recommended that line charts adopt an approximate height-width ratio of 5:8. One of the drawbacks of a standard line chart is that they can imply a steady change from one data point to another (Yau, 2010). For occasions where data remains stagnate and then suddenly changes, step charts can used to visualize this difference. "Instead of connecting point A to point B directly, the line stays at the same value until there is a change, at which point it jumps up (or down) to the next value" (Yau, 123).

Scatter Plots can be used to map relationships between two variables (one which is depicted on the x axis, and the other along the y axis. These graphs depict single points which coordinate to a particular location along the x and y axis. One of the strengths of a scatter plot is that it can be used to represent large datasets, to reveal overall patterns, or correlations between variables. Yau notes that one of the dangers of a scatter plot is the temptation to assume a 'cause-and-effect relationship' between variables within a scatter plot. For example Yau notes, "just because the price of a gallon of gas and world population have both increased over the years doesn't mean the price of gas should decrease to slow population growth" (Yau, 183). To prevent misguided interpretations it can be useful to provide guiding text (in the form of titles or textboxs). Trend lines can also be added to scatter plots to help highlight the overall patterns of the dots.

Another form of a scatter plot is the *Bubble Chart* which is used to introduce a third variable. While x and y axis present two variables, like a standard scatter plot, the size of the data point (or the bubble) is used to represent a third variable. In order to do this effectively the number of individual data points depicted should be significantly lower than those depicted in traditional scatter plots (Vlamis 2010). As well Yau highlights that if you are representing the data points as circles, the size of the circle should be determined by the area, rather than the diameter, radius, or circumference (Yau, 193).

Heatmaps can be extremely helpful in displaying multiple variables. A heatmap is essentially a table but instead of using number values, colours are used. The resulting image is a grid, which is often about the same size as the initial table, but the colour differentiation makes it easier to readily spot high or low values (Yau, 229). Within heatmaps it is very important to select the colour palette wisely because it will highly impact the charts tones. Yau recommends selecting neutral colours, or muted tones for somber topics, and selecting more vibrant colours for more uplifting or casual topics (Yau, 234). Rollovers (or tooltip features) can also be added to heatmaps to provide additional details to individual data points.

Node-link diagrams provide a way to represent the hierarchical ordering or structure of data. These displays are often call tree diagrams because they resemble a tree with the leaves and branches at the top, and roots at the bottom of the structure. There are three main types of node-link diagrams. Traditional diagrams (a) connect parent data points with sub-points through a series of connecting lines. Orthogonal (b) diagrams differ from traditional, through the use of a chain of horizontal and vertical segments connected by through ninety degree angled edges. Radial diagrams (c) benefit from being space efficient, growing a circle center radically. The same hierarchy is still retained but the empty space between nodes is reduced. Despite these features, most viewers are said to favour diagrams which have the parent node starting at the top of the graph, compare to on the side or middle (Burch et al. 2011). As well traditional and orthogonal diagrams provide additional ease for viewers to decode the relationship between nodes. In contract radial diagrams, tend to encourage cross-checking (Burch et al. 2011).[i]

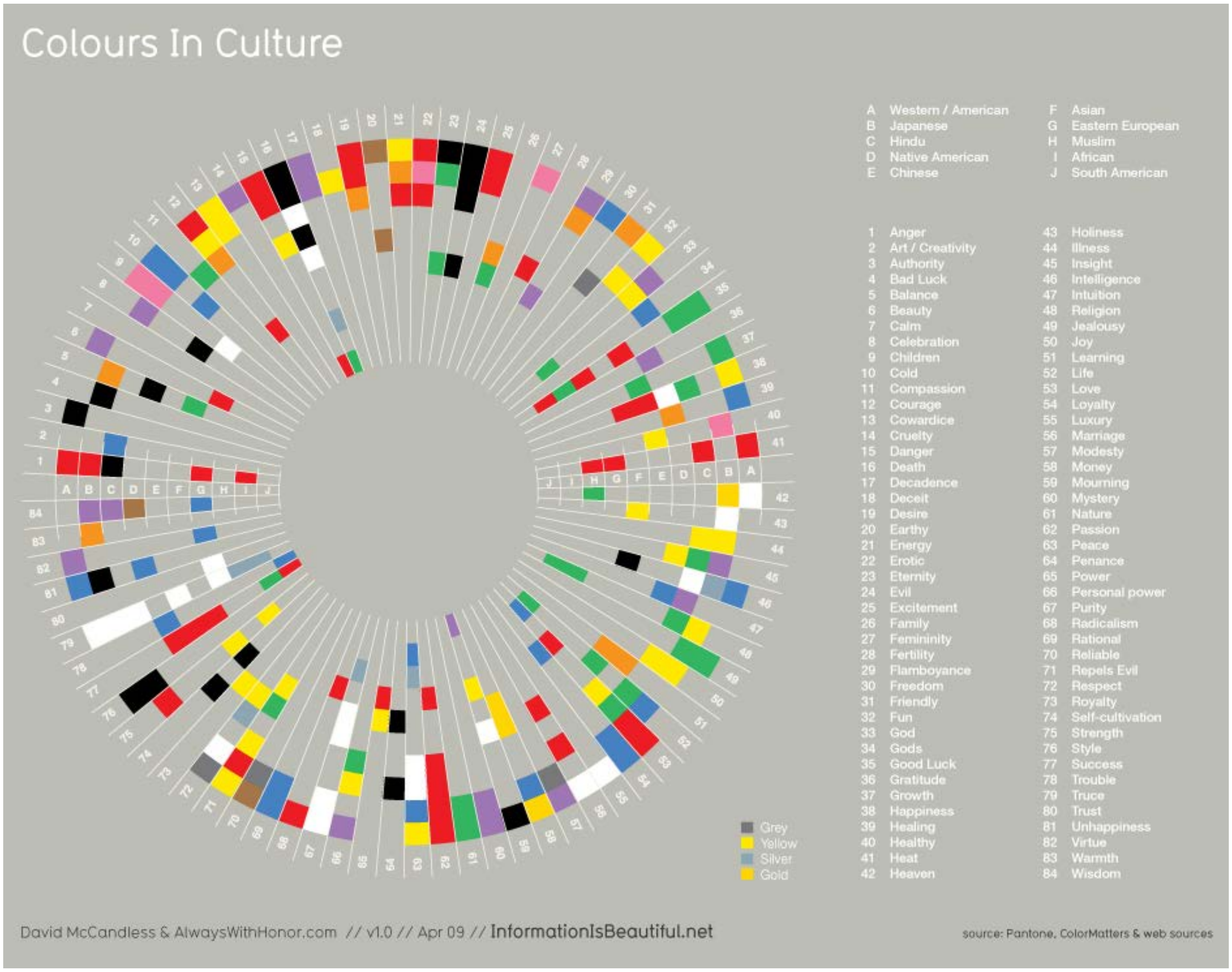


Spider charts, also referred to as star or radar charts, can be effective in displaying multivariate data. These charts select an initial stating point, and then map multiple (often three or more) quantitative variables outwards along different representative axes, or spokes. The length of the spoke corresponds to the value, and connecting lines are often drawn to other spokes to show the relationship between variables. While spider charts are effective at showing outliers and similarities between sets, it is difficult for viewers to compare the lengths of the different spokes, because the eye is able to effectively discern radial distances.

The aforementioned graphical forms do not represent an exhaustive list of data visualization forms, but rather present a small sampling of the strengths and weaknesses of commonly applied forms. Every type of representation involves trade-offs, so it is important to understand a form before applying it to your data (Lindquist, 2011). Each of these forms can be significantly weakened by the failure to title, label, or colour appropriately. Section 2 of this report provides additional guidance on how to effectively apply these features to a variety of graphs; however it is worth discussing the implications of colour briefly.

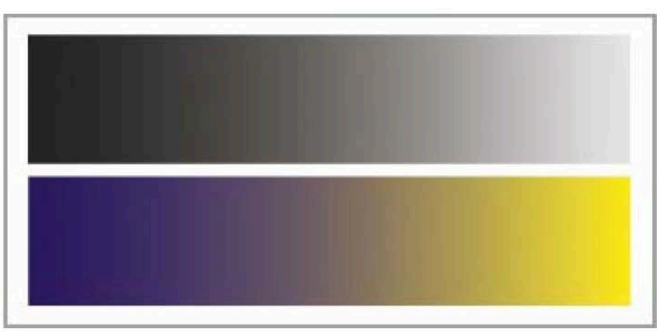
Colour can be a very important tool for setting the tone of the data visualization, or for highlighting certain data points, such as outliers. When considering colour use, there are three important elements: the hue (the actual pigment of the colour), the chroma (the bright or muted colouring of the pigment), and the value (the lightness or darkness of that pigment) (Stone, 2006). Hue charts are represented in many different formats, but all reserve the same ordering of colours placing analogous hues together, and contrasting hues on the opposite side. Different hues tend to carry significant connotations, which can vary across cultural backgrounds or contexts. As, IIiinsky and Steele, adamantly affirm that it is vital to consider the different knowledge, assumptions which your viewer brings to your design; this includes their cultural conventions pertaining to colour (IIiinsky and Steele, 2011). The colour red for example has been known to represent danger in the United States but symbolizes success within China. David McCandles iconic visualization of cultural connotations of common colours is a prolific example of how dramatically and frequently colours convey different meaning. One IDRC-supported partner highlighted that during the creation of their data visualization they were careful not to select colours associated with any political party. This example is useful because it reveals that colours hold different meaning across different cultures and contexts.

Figure 6: Colours In Culture (McCandles, 2009)



The chroma of a colour can also carry connotations, but is more often used as a tool for adding emphasis to a particular data point. Stephen Few purposes that more muted-tones should be used to convey the overall data points, whereas vibrant colours should be reserve for adding a few additional highlights (Few, 2006). Stone however argues that the value dimension is visually most important for legibility. It is very easy to detect value changes in shades of grey, or in shades of a single colour, however legibility concerns arise when trying to discern value amongst different hues. Figure 7 illustrates colour gradations between two shades that have the same value scale (Stone, 2006).

Figure 7: Colour gradations illustrate (approximately) the same value scale (Stone, 2006).



Value considerations are particularly important for documents that might be printed in colour and in black-and-white, as well as for individuals who might be colour blind. Stephen Hanks suggests a good test for ensuring that there is adequate value differentiation between colours is to change your image to a grey scale or try to fax your visualization to see whether the differences are legible.

Maureen Stone states that “Color used well can enhance and clarify a presentation. Color used poorly will obscure, muddle and confuse.” (Stone, 1). It is therefore extremely important that colour is not interjected into displays, but is carefully considered. Colour can be effectively added to all of the aforementioned graphical examples, presented in this section. These graph types represent some of the oldest types of data visualizations. While these ‘oldies’ are still extremely effective in communicating data, there have been several new trends, which have been changing the forms and field of data visualization.

Current Trends and Interesting Studies Shaping the Field of Data Visualization

Geo-spatial visualization

One of the current trends which is building in popularity, and use, is geo-spatial mapping. This type of data visualization utilizes maps to order data by their connection to specific locations. Geo-spatial mapping is effective in part because it provides viewers with a familiar entry point (a geographic location) from which they can understand the additional data presented. From a business perspective, Stephen Few has found that “much of the information that businesses must monitor and understand is tied to geographical locations” (Few, 2007, 5). Within the field of international development and research, geo-spatial visualization programs such as Ushahidi, have provided a means for crowdsource data to ascribe to an interactive map. The Ushahidi platform was developed in 2007 during the aftermath of Kenya’s disputed presidential elections. Since this time the open-source program has been used to collect testimonies from individuals around the globe during times of crisis. The World Bank and AidData have also begun utilizing geo-spatial mapping to increase aid transparency by providing a geographical reference map of the locations of projects receiving funding (open.aiddata.org). IDRC is also currently engaging in an exercise which will help map the location of supported-research projects around the globe (Heloise, 2011).

Understanding Complex Systems

Another area which has received recent attention for its increasing reliance on visualization has been within the field of systems thinking. Although modern system analysis rarely perceive themselves as ‘visualists’, Lindquist affirms that many authors and facilitators utilize visualizations as a way of addressing complex situations (Lindquist, 2011). For example as often participants are encouraged to illustrate their perspectives, and emotions through diagrams, which are then discussed and debated in groups. The rationale for utilizing pictures within soft systems methodology espouses from the fact that,

> *the complexity of human affairs is always a complexity of multiple interacting relationships; and pictures are a better medium than linear prose for expressing relationships. Pictures can be taken in as a whole and help to encourage holistic rather than reductionist thinking about a situation (Checkland, 1999, 12).*

The use of visualizations within systems thinking is not just tied to the use of illustrations, but a range of other visual approaches which blend exploratory and communicative functions to help stimulate sense-making, and strategic dialogue. Some of these approaches include the use of data visualizations to present simulations of different trajectories or situations; to develop ‘shared-mental-maps’ which can be used in scenario-building amongst groups, or even the use of logic models, or outcome tables to map performance thinking (Lindquist 20). Although the use of data visualization is just starting to be acknowledged within systems thinking, this area does reveal rich potential for utilizing a growing range of visualizations to communicate elements of complexity (Lindquist, 21).

A Cause for Chart-Junk

A recent study conducted argued that adding visual difficulties, or what some refer to as chart junk, can actually aid cognition - depending on the objective of the visualization. For example, using hard-to-read fonts, providing static-staged diagrams (vs. animated visualizations), and adding 3D to charts are all examples of chart junk which have been shown to improve people's retention of data. In the past design principles have been premised on making data intuitive and clear. What this study reveals is that strategically breaking these principles can force people to work harder to understand what is portrayed, resulting in improved cognitive efficiency and data retention. While this study has caused many to pause and reflect, include chart junk is still likely to be a deterrent for most viewers, unless they have a particular motivation or interest in the data presented.

Utilizing Data Visualizations with Policy Contexts

While much research has yet to concentrate on the effect of data visualization within policy contexts, Lindquist proposes that within these highly information-saturated environments, provide an exciting opportunity of data visualization. "Ministers, citizens, stakeholders, and officials alike function in environments with information overload and time compression, and often paradoxically have too little and too much information for addressing specific issues" (Lindquist 3). It is for this reason that Lindquist contends that policy contexts are becoming a ripe environment for using data visualizations. While the complex nature of policy contexts provides one rational for using visualizations, the other comes from the recognition there is a multitude of ways to communicate and consume information. Even IDRC-partners have admitted that they often do not have time to read every document which crosses their desk. Instead, today's readers, be it a policymaker, researcher, or citizen, tend to adopt a selective approach to reading documents; perusing a publication or brief to see if any interesting information jumps out at them. While research is just starting to explore this potential area, policymakers and governments are also examining if and how they should invest in utilizing data visualizations to communicate with their citizens.

The Risks of Data Visualization

While there are many potential benefits of data visualizations, there are some known risks or disadvantages, which are primarily associated with the limitation of a form or misuse. Tufte highlights that all of the undesirable effects of visualizations are either caused by the designer or by the user (or their interpretation) (Tufte, 1986). Knowing the limitations of visualization's form, as well as appropriately using design conventions, is one important countermeasure for these risks. However, in 2008 Bresciani and Eppler categorized different types of visualization 'disadvantages' which were attributed to each actor. To do this, Bresciani adopted Roos' classification of cognitive, emotional, and social effect distinctions (Roos, Bart et al. 2004). Figure 8 outlines some of the effects which were identified within each of these categories.

Figure 8: Visual Disadvantages Categorized (Bresciani et al. 2008)

	Designer induced	User induced
Cognitive	- Ambiguity *Visual notation may contain unlabeled symbols that may be ambiguous and thus difficult to interpret.*	- Change blindness *Important changes in pictures may go unnoticed by the viewers.* - Channel thinking

	- Breaking conventions *A visualization may employ different visual rules or symbols than normally expected.* - De-focused *Visualization may distract use to represent data may not be universally understandable and confuse some audiences.* - Implicit meaning *Many visualizations contain allusions that are not fully described or explained and may go unnoticed or may be misinterpreted.*	*The visualization can direct thinking in an inappropriate direction -caused by a metaphor or familiarity level.* - Depending on perceptual skills *People see differently, depending on physical (e.g. colour blindness) or cultural factors (attention to background or foreground).* - Wrong salience *The reader concentrates on the wrong issue, for example on the tool or on the visual appearance instead of on the task.*
Emotional	- Disturbing *Some images may cause emotional harm to the reader because of their stocking or repellent content.* - Boring *Some graphic representations are perceived as un-interesting.* - Wrong us of colour *The inadequate use of colours or their combinations make images confusing.*	- Visual stress *Some kind of patterns (stripes or flickering) may cause illness in the reader.* - Personal likes and dislikes *Some visualizations may get more attention than others, not because of their importance, but because they fit the cognitive preferences of a particular viewer.*
Social	- Affordance conflict *A visualization may signal the wrong kind of required (inter-)activity to its viewer.* - Inhibit conversation *Having one's contributions visualized (for example in a group context) may lead to participants being less outspoken about certain issues.* - Rhythms of freezing and freezing *A visualization may make a certain view point or idea too rigorous and fixed too soon, thus not leaving enough room to invest alternative views or options.*	- Cultural and cross-cultural differences *The meaning of symbols and colours are not universal and hence some graphic representations may be misinterpreted in other cultural contexts.* - Framing effect *The meaning of a visualization is not interpreted in a vacuum but as part of a broader context, that depends on what the user has been previously exposed to.* - Different perspectives *Different people look at issues from different points of view.*

Bresciani, et al. 2008

This report also discussed different countermeasures, which could be adopted to offset some of these effects. For example, defocused visualizations are often the product of a designer who has not identified the visualization's main message, and as a result has distracted the viewer from what is really important. "Sources of distraction can be: unnecessary ornaments, visual background noise, flashy animated graphics, or including unrelated elements in a diagram" (Bresciani, et al. 23). Defocused effects can be countered through utilizing position (emphasizing key data at the top of the visualization) or through emphasizing tools like size, colour, or accentuating symbols. Another recommendation, suggested by Bresciani, is to avoid the use of unrelated elements or decorative features that may distract or detract from the visualizations message.

This list and suggested countermeasures further reinforce the importance of good design principles; since the majority of data visualization disadvantages can be overcome by effective and strategic designs.

What this literature review presents is a condensed overview of some of the history, rationale, and theories about data visualization, as well as the current debates and discussions appearing in the field today. While this is merely a concise summation, it is intended to provide an entry point to begin assessing where IDRC is situated within the spectrum of data visualization users. Section 2 begins to unpack this issue further by highlighting that although IDRC is no stranger to data visualization use, it is also not a front-runner in leading innovative designs.

Section 2:

IDRC & Data Visualization

Data visualization is not a new concept for IDRC, or its partners. IDRC-supported research has dabbled in visualizations use for years. Although the majority of these visualizations involved simple graphs and charts, the concept of complementing presentations with illustrative representations of data is not new. That said the term *data visualization* can be intimidating for those less familiar with the field. This is partly because of the recent explosion in dynamic and interactive data visualizations which have flooded the internet and media publications. While these innovative displays often create quite a splash, the key principles for producing effective visualizations remains the same regardless of whether your visualization is static or dynamic.

In order to gain a better understanding of how IDRC-supported research has used data visualizations, and to what effect; a three stage analysis was conducted using IDRC-supported research outputs. The first two stages were conducted internally and provide context on the frequency of data visualizations used. Stage 1 provides a snapshot of how often visualizations are used and within what type of documentation; while stage 2 examines what kinds of visualizations are used most often. Stage 3 provides an external assessment of the Centre's data visualization use; exploring whether these data visualization design follow good practice, and ways of improving data visualization use to ensure effective communication of research findings. Overall these three stages provide a useful assessment and entry point to begin a conversation on data visualization use.

Methodology

The first stage of this process, questioned the degree to which IDRC-partners were currently using data visualizations to communicate their findings. I collected a randomized sample from all documents filed in IDRC's Digital Library (IDL) since 2009. A total of 330 documents were examined and coded for their document type and the occurrence of visualizations[2]. This sample size provided a confidence level of 95% and a confidence interval of 5. The documents were selected from across the IDL collections, ensuring representation from all programs within the Centre. Since the documents were selected at random there was no control over the type of documents included. Overall, the review was composed of academic publications (80), professional publications (71), media documents (40), event documentation (48), project reports (71), and evaluations (20). The following definitions were used to identify the document type:

> **Academic Publication:** All publications produced for and published for an academic audience (eg. journal articles, books, book chapters, literary compositions or dissertations, and scoping or exploratory studies).
>
> **Professional Publication:** Documents published for the development, policy or general community (eg. policy briefs, project briefs, manuals, curriculum, and training materials).
>
> **Event:** Associated event documentation (eg. text of conferences, proceedings, speeches, slide presentations, workshop reports).
>
> **Evaluation:** All internal or consultant report which evaluates project(s).
>
> **Media:** Final versions of communication material produced to inform the community about the project (eg. website, social media posts, newsletters, bulletins, pamphlets, newspaper articles, pictures, and videos).

One of the limitations to the selection of documents was that the query of IDL communities could not account for the multiple entries from translated documents or the existence of dead links, which resulted from documents being withdrawn from the system. To address these issues, all duplicated or translated versions were deleted from the initial list from which the 330 documents were selected. As well, in the cases were documents had been withdrawn from the system, an alternative was selected from the corresponding community to replace the file. Another selection limitation is the inability for IDRC systems (namely the IDL) to capture and record online or interactive outputs; which might include more dynamic visualizations. As a result, stages 1 and 2 were only able to include static documents; however to account for this limitation, additional efforts were made to source dynamic data visualizations for inclusion in stage 3.

After initially coding the sample documents from stage 1, all documents which contained data visualizations were separated out for inclusion in stage 2. During stage 2, a second sample of documents were coded to examine what types of visualizations were being used, and further specifications on colours, axis labels, source information, and data clarity, were also noted. In total 36 documents were selected for inclusion in stage 2. Documents were selected at random, but the number of each document type was calculated from the percentage of visualization occurrence per document type.

For stage 3, data visualizations were selected through a number of processes. First an open call was made to IDRC staff to nominate projects which contained strong or innovative data visualizations[3]. Invitations and notifications were also sent to a variety of Centre staff including regional communication officers, program management officers, and communication staff in Ottawa. The

[2] Each document was initially coded for the following: document type, presence of visualizations, presence of tables, visualization category, and number of visualizations.
[3] This call was published on the Centre's internal website

nominated examples make up the majority of 21 visualizations included in stage 3. Four of the included visualizations were pulled from documentation found during stage 1's randomized review.

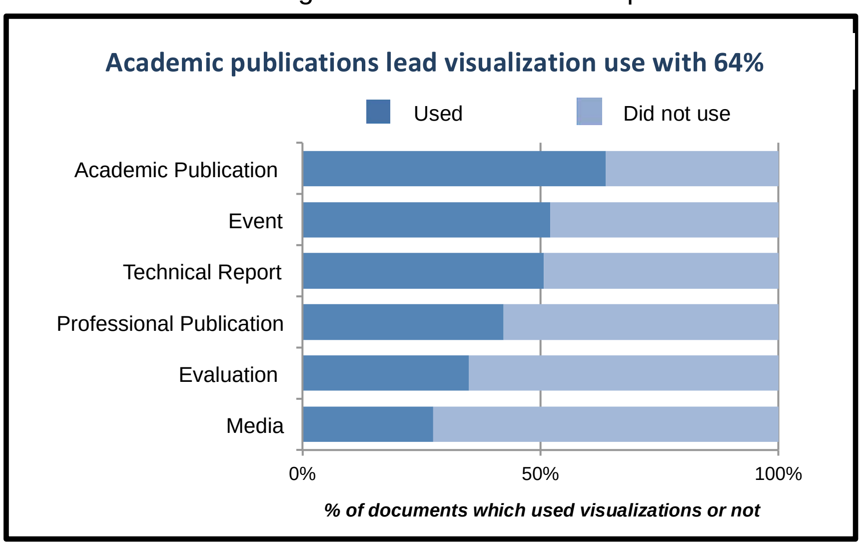
The framework for stage 3 was predicated on the understanding that the most effective data visualizations are clear, focused and compelling. While these characteristics can be subjective, and audience dependent, they provide a strong starting point for assessing data visualizations which are intended to communicate research. The following questions were used to assess each of examples included in stage 3:

- **Clarity:** Is the charting form appropriate? Are titles appropriate? Are the units of the data familiar to the intended audience? Does the visualization anticipate the questions it raises?
- **Focus:** Does the language used in the visualization support at least one specific idea? Do design choices such as colors, typography or highlighted areas support at least one specific idea? In more complicated visualizations, is it clear that some parts of the information are more important than other parts?
- **Compelling:** Will your audience want to talk about or act upon this data? Does the richness of the data justify a visualization? Would incorporating photography or annotations make the data more relatable?[4]

To complement the findings of this review process, five IDRC staff members (from programs branch, communications, and information management) participated in structured interviews, which last forty-five minutes. As well over fifteen other staff members reviewed a selection of IDRC-supported research visualizations during an interactive lunch-time presentation. Lastly three partners participated in one-on-one phone interviews which provided insights on their experience and reflections on data visualizations.

Stage 1 & 2: Review Findings

Overall stages 1 and 2 reveal that IDRC-supported research has, and continues to; invest time and resources in creating data visualizations to present research findings. 48 percent of sampled IDL documents included some form of data visualization. However, the majority of this visualization-use was focused on the insertion of charts and graphs into publications or presentation slides. There were very few examples of next generation visualizations, and none of the documents demonstrated truly innovative or ground-breaking design use.[5] Instead, the documents tended to use fairly standard visualizations as a way of



[4] The framework for stage 3 was developed by Amanda Cox, based on feedback and input from the Evaluation Unit.
[5] While this could in part be linked to the inability of the IDL to capture the more interactive and online displays, there very few online examples brought forward during stage 3.

complementing the textual explanations of the research findings, or to provide a visual representation of models or systems.

In total, 72 percent of documents with visualizations included standard chart types, such as bar and pie charts, while the remaining 28 percent utilized other forms which varied from tree maps to venn-diagrams.

The review also found that not all document-types utilized data visualizations to the same degree. Academic publications had the highest rate of visualization use; and contained on average more visualizations per document than any other type.[6] In contrast, media documents included the fewest instances of data visualizations. The majority of media documents reviewed were under ten pages and used maps. For academic publication line charts were the most frequently used form; compared to bar charts which were most common overall, and received the highest rate of use in all other document types.

While it is important to acknowledge which types of visualizations are being most used in research outputs, it is also important to assess whether the forms are being used well. As discussed in Section 1, the effectiveness of data visualizations is predicated on proper design and integration. Although stage 3, will discuss good design practice in great detail, stages 1 and 2 did revealed several broad concerns impacting the integration of data visualizations overall. These concerns were primarily regarding issues of inconsistency, reliance on 3D graphics; and the ineffectual use of text.

Issues of inconsistency were found around the use of colour schemes and charting graphics. Nearly 40 percent of documents reviewed in stage 2 included the unjustified use of multiple colour schemes which shifted from one-visualization-to-another. The most extreme instance of this was a document which used over eight different colour palettes in its different visualizations. Establishing a consistent colour palette is important for providing coherence and unity to a document. Colours also carry different connotations based on different cultures, or contexts, which can make the addition of multiple colour schemes even more confusing. If a document readily changes its colour palette, the effect can be jarring and distracting for the reader, since their mind is automatically decoding colour differentiation and searching for implied meaning. In contrast to these examples of disruptive use of colour, several other documents utilizing colours extremely effectively, drawing attention to highlight sections of the data begin discussed.

The use of different graphics was another area of inconsistency. There were several documents which changed the graphic-bars used within the bar chart to cylinders half way through the document, without justification. Alternatively, other documents would introduce 3D-bars into one chart and 2D-lines in the next. Much like alterations in colour, the mind's visual processing registers these changes and can misinterpret them as patterns which hold deeper meaning. Good practice recommends that there should be a consistent graphic used for all visualizations of the same form, unless there is a particular meaning for deviating from this form.

It is important that charting graphics are not conflated with different data visualization forms. One of the fundamental principles of data visualization is that the form should be determined by what is being communicated and the nature of the data. Selecting an appropriate chart form was one of the

[6] It should also be noted that, academic publications tended to have a higher number of pages per document then other documents which provides the potential for additional room for the inclusion of data visualizations.

strengths of the documents reviewed, and three-quarters of the sample decided to integrate more than one type of data visualization.

Unfortunately, despite the demonstrated competence in determining the appropriate chart form, there were many poor design decisions including the use of 3D-chart-graphics. 38 percent of documents which utilized bar or pie charts, also elected to add 3D graphics. The 3D perspective distorts the overall length of the bars, and misrepresent the surface area of slices. Although 3D-chart-graphics were present within a small proportion of the overall documents, there existence is still cause for concerns since recommendations against their use are well-documented and publicized.

The final concern which was revealed in stage 2 was the ineffective use of text. The majority of data visualizations were situated within the body of a publication or report, without a focused title to help the reader initially interpret the information being displayed. Stephen Hanks, IDRC's resident graphic designer, affirms that "headlines and decks can serve as important qualifiers for data visualizations. It can also help viewers who are less likely (or comfortable) interpreting the data to have a textual reference to explain what the graphic is showing" (Hanks, 2011). Titles can also play an important role in creating appeal. Hanks uses ICT4D's internet cable map which included the title "OUT OF AFRICA", as an example of a quick, witty, and dynamic headline which pulls viewers into the data. The benefits of creating focused and compelling titles are numerous, which is why many data visualization proponents advice working with writers to draft supporting texts, and headlines.

Overall, these three areas of concern do not suggest the need for extensive reform to data visualization use; but rather refinements, and a need for greater education and capacity building around good design practice. The large prevalence of visualizations within the documents sampled, indicate that researchers have acknowledged a use, and/or need, for including visual representations of their data. However, as Stephen Few highlights, it is important for this recognition to be coupled with an appreciation and knowledge for appropriate design.

One way of increasing knowledge is through the use of identifying models of good practice. The concept of pulling from existing models can be extremely helpful from a design sense, but can also assist in identifying innovative designs, which could be applied to a similar dataset. Identifying models is also useful for acknowledging which types of visualizations researchers naturally gravitate to. Many of the staff interviewed at IDRC, supported the use of data visualization, but felt that it was still an untapped resource; commenting that most of IDRC-supported projects had barely scratched the surface of its potential. One program leader suggested that it should be the programs leading the charge and pushing research partners to attempt new and creative ways of utilizing data visualizations to communicate their research.

Perspectives of IDRC Staff and Partners

When asked to identify reasons why research partners weren't experimenting with data visualization use, a program officer identified three main rationales from the projects he worked with. First, he felt that there was a general lack of knowledge around how to produce data visualizations. Second, some of the researchers did not see their role as being a communicator, but rather felt their responsibility was to share their work with fellow academics. Lastly, he noted that researchers were becoming increasing aware of the pedigree and status of international research, and were therefore focused on building accreditation through citation counts, instead of producing professional or communication documents. Even for partners who were trying to communicate with audiences outside of academia, there was little attention to the benefits on data visualizations. One IDRC-partner commented that it was only recently that they have 'woken-up' to the idea of integrating data visualizations to assist with translating their data into information into a format which had greater appeal for policymakers;

> *"We have woken-up to this idea of data visualization very recently. Over the last few years we have recognized that when you are presenting data, the data does not necessarily become information… our readers are more interested in getting the information as quickly as possible. They do not want to go through a lot of data to get the information that we want to convey. We have to treat our data so a person with a short attention span can get it... Politicians and bureaucrats have short attention spans. They will devote a couple of minutes to look at a document, and in that time if there is something that catches their attention, something they can retrain, then they are likely to devote more time."*

Partners in various regions have also noted that newspapers have been investing and experimenting with data visualizations, creating greater interest and demand for this type of presentation. According to Lindquist, this trend is part of the reason why policy contexts are ripe for data visualizations use (Lindquist, 2011).

This does not mean that IDRC-supported partners should immediately start investing in larger dynamic data visualizations; but rather highlights that it is important to monitor these trends and begin examining relevant aspects of this field. Seeing as how visualizations have been a part of human communication since the 2nd century, it is unlikely that their use and importance will entirely dissipate in the coming years. What is more likely is that, as in the past, some trends in data visualization will fade, while others will be adopted into common practice. That said, IDRC, and it supported partners should not wait to investigate the use of data visualization. They should be investing resources and time in learning the principles for effective data visualizations, so they can ensure that both the designs they are currently using, and those which may be considered tomorrow, are well informed and based on good practice. As with all skills, before we consider sprinting to the front of the data-visualization pack, it is essential to ensure that we are able to *walk before we run*. Knowing the basic principles of data visualization design is a fundamental step to utilizing visualizations effectively to communicate research for influence.

In the final stage of this review process, the Evaluation Unit sought the expertise of an external consultant to evaluate the degree to which data visualization's created by IDRC-supported partners are in fact adhering to good design practice. For this stage, Amanda Cox, graphics editor at the New York Times, was brought on board to discuss in richer detail, how individual visualizations could be more effective in communicating for influence. Amanda Cox reviewed a total of 21 data visualizations which were gathered from IDRC staff nominations of strong data visualizations created by IDRC-supported partners. The report that follows is designed to assist learning; using each example to illustrate larger design principles which could be applied to data visualizations across the Centre's work.

Stage 3: External Data Visualization Review

Written by Amanda Cox

Summary

A review of 21 data visualizations produced by the International Development Research Centre's projects found that the Centre's data visualization work is generally clear. About three-quarters of the projects used the best possible charting form, such as a map or a bar chart, for the data shown. (See Appendix 1 and the discussion of each visualization for detailed assessments.) Units for the data were typically included and appropriate. Nearly all of the projects incorporated a title that described the data.

These titles, however, were overwhelmingly generic descriptions, which would have been appropriate regardless of the research results. They simply described the topic of data, instead of what was learned from analyzing it. Presumably, the purpose of most visualizations is not to simply convey that data exists, but to help reach some sort of a conclusion. For many readers, titles will be the first piece of information they read. Using generic titles forces readers to draw their own (possibly misguided) conclusions about the patterns shown in the visualization.

In fact, very few of the visualizations used any language in support of a specific, focused idea. Firm conclusions could be found in the reports accompanying the visualizations: “Cuba and South Africa are the most active in South-South collaborations”, “Five out of the 19 projects completely lack a gender component, while nine consider the issue only superficially” or “The global average cost [of transferring money] has not come down.” But the words used within the visualizations were often very timid.

Design choices also tended to be generic. Color, typography or highlighted areas were rarely used to draw attention to points of interest. Using color in a way that supports a message, as well as sorting tables by a value of interest, are among the most frequent criticisms in this review.

Among the more complicated visualizations, about half established a clear hierarchy, in which some of the information was clearly more important than other parts. Hierarchies were established with position, color and size.

Nearly all of the visualizations included a sufficient amount of variation to justify a graphic, but few of the visualizations described trends or anomalous points, anticipated questions that the visualization raised or indicated areas that experts found interesting.

Most of the interactive work in this review allowed users to look up data of interest to them. In general, these visualizations functioned well and navigation was clear, though none of the examples used technology that allows smooth transitions between views.

Key issues and trends

Within the last five years, presenting large amounts of data – especially in an interactive way – has become substantially easier, and the volume of this type of work has grown rapidly.

Much current attention is being devoted toward making interactive work that functions on mobile and tablet devices. Increasing amounts of attention is also being paid to real-time, streaming visualization, and collecting data from non-traditional sources, such as crowd-sourcing.

With non-traditional sources, transparency and proper sourcing is a larger issue than it is with data gathered as part of traditional research projects or by governmental organizations. Regardless of the size, provenance or complexity of final visualizations, providing access to full data tends to generate goodwill and greater faith in results, assuming confidentiality can be maintained.

As interactive work matures, more analysis is being incorporated into visualizations. Links to interesting findings can be part of the visualization itself or part of a blog-type post that sits on top of the visualization. Other mature work involves combining different types of media. For example, photography of research projects linked to a map might make data feel more relevant than simple circles on the same map. Audio of experts explaining their results alongside charts might also help clarify difficult ideas.

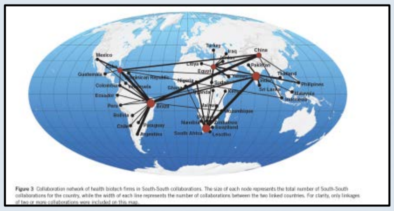
Many of the examples in this review display relatively small amounts of data in a static way. Even when a visualization is intended to be viewed online, this may often be the most effective way to communicate research results. Why? Static visualizations tend to give the creator more control over the message. In the same way that editing is an important part of writing, distilling information to what is important is crucial for effective visualization. In contrast, interactive displays of larger amounts of information may be more engaging for topics that are very familiar or personally relevant for an intended audience.

The following section considers 21 examples, chosen from IDRC-supported research. These examples have been grouped into five broad subject areas: color, sorting tables, choosing a chart type, clarity and interaction.

All of the examples have positive elements, but the review mainly focuses on opportunities for improvement, in the hopes that relatively simple changes could result in more effective or more powerful communication.

Color

Example 1: Designing for emphasis



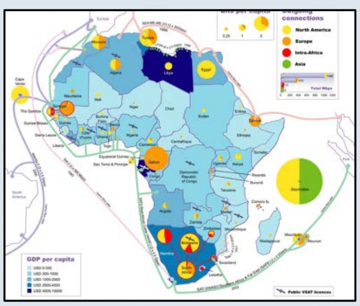
Design choices should help a reader determine what is important. In this example, some choices appear to have been made without considering the data.

For example, terrain shown in the background is unlikely to be very relevant in a map of South-South collaborations between biotech firms. Instead, simple country outlines might have been used to convey relevant information. Countries like Mexico and Nigeria, which may be underrepresented because they were not surveyed, could be colored in a slightly lighter shade.

At first glance, the data is forced to compete for attention with a deep blue ocean and bold typography. Bold type — a great tool to emphasize salient points or to help readers skim through a graphic — should seldom be used for every label. And, whenever possible, type should not be obscured by data. Using great circle arcs (e.g. http://paulbutler.org/archives/visualizing-facebook-friends/) may further reduce clutter.

Finally, a key should be part of almost every graphic. Does a thick line represent 40 current collaborations? Or three within the last five years? Without reading the accompanying text, it is impossible to know. Conclusions from the accompanying text can also be drawn into the graphic. Consider which of the following is a more compelling introduction: “The size of each node represents ... ” or “Biotech firms in South Africa have many collaborations with India, but none with China.”

Example 2: Making some data secondary



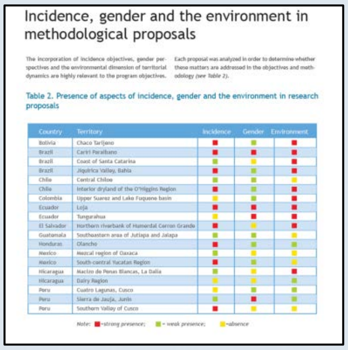
With four separate keys, it's clear that this map has a lot going on. As a look-up table, it is reasonably successful. But if someone comes to the map without knowing what they are looking for, where should they start? What is most important?

The title — “The Internet: Out of Africa” — is one clue. But the colors chosen to represent each country's wealth make the fact that Egypt is wealthy jump out.

Using a very light gray palette to encode wealth would visually suggest that the wealth data is secondary, in the same way it is clear that Spain is secondary.

Another option would be to remove the income data from the background altogether. Trying to layer too many pieces of information into one view is unwise. It is better to make one strong point in three simple graphics that are displayed next to each other. A small map of income could still accompany the main map. Size, like color, is very useful in immediately conveying relative importance.

A secondary point: charts that label every data point do not need scales.



Incidence, gender and the environment in methodological proposals

The incorporation of incidence objectives, gender perspectives and the environmental dimension of territorial dynamics are highly relevant to the program objectives. Each proposal was analyzed in order to determine whether these matters are addressed in the objectives and methodology *(see Table 2)*.

Table 2. Presence of aspects of incidence, gender and the environment in research proposals

Country	Territory	Incidence	Gender	Environment
Bolivia	Chaco Tarijeno	■	■	■
Brazil	Cariri Paraibano	■	■	■
Brazil	Coast of Santa Catarina	■	■	■
Brazil	Jiquirica Valley, Bahia	■	■	■
Chile	Central Chiloe	■	■	■
Chile	Interior dryland of the O'Higgins Region	■	■	■
Colombia	Upper Suarez and Lake Fuquene basin	■	■	■
Ecuador	Loja	■	■	■
Ecuador	Tungurahua	■	■	■
El Salvador	Northern riverbank of Humedal Cerron Grande	■	■	■
Guatemala	Southeastern area of Jutiapa and Jalapa	■	■	■
Honduras	Olancho	■	■	■
Mexico	Mezcal region of Oaxaca	■	■	■
Mexico	South central Yucatan Region	■	■	■
Nicaragua	Macizo de Penas Blancas, La Dalia	■	■	■
Nicaragua	Dairy Region	■	■	■
Peru	Cuatro Lagunas, Cusco	■	■	■
Peru	Sierra de Jauja, Junin	■	■	■
Peru	Southern Valley of Cusco	■	■	■

Note: ■=strong presence; ■ = weak presence; ■=absence

Example 3: Matching expectations

In North America and many other countries, a red-yellow-green color scheme invokes the idea of traffic lights, and their stop-slow-go messages.

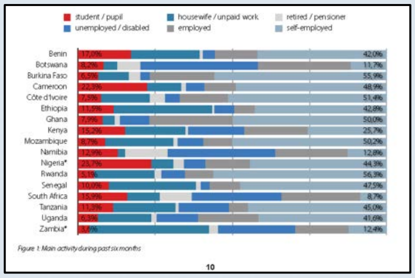
Even without traffic lights, the color spectrum suggests that yellow comes between red and green.

But, here, red (stop) is used to identify a strong presence, green (go) to identify a weak presence and yellow (slow) to identify absence.

Perhaps this choice was made out of respect for color-blind readers, who would be able to distinguish between some kind of presence (red and green) and absence (yellow). Even if this is the case, a more natural solution could be found. For example, the squares representing absence could be removed altogether, and a dark and light shade could be used to suggest the strength of presence.

Basic words describing the table (“Incidence, gender and the environment in methodological proposals”) are repeated in four places. Some of this space could be used more effectively by describing what the data actually shows. (e.g. “Gender perspectives are strongly present in four proposals, while the environment is strongly present in 10 proposals.”)

Sorting



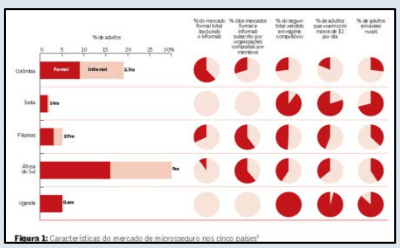
Example 4: Sorting and color

Alphabetical sorting is usually not the best choice in a chart with numerical data. In this example, color suggests the primary focus of the chart is the percentage of people who were students during the last six months, so the data could be sorted by that value. Sorting by one value allows easier comparisons for the other values. Does the pattern for the self-employed (the final bar) follow the same pattern as that for students? With sorting, it would be easy to tell.

Again, color should be used thoughtfully. Here, certain categories are more similar to others. For example, the unemployed, employed and self-employed are all part of the labor force. Thoughtful color choices could make this clear.

Choosing to label only key values in a chart with many numbers is a good idea. To reduce clutter, units like percentage signs are only necessary on the first value.

Example 5: Consistency

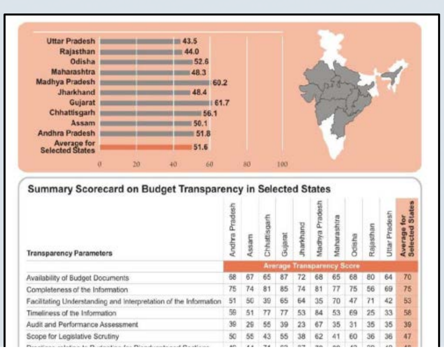


This example further emphasizes why alphabetical sorting is rarely the best way to present graphical data. (The translated Portuguese version of this chart underscores this point, with “Africa do Sol” making the order appear to be random, because the order of the rows retains the original alphabetical ordering from the English design.)

Consistency is very important in small-multiple charts. In the first four columns of pie charts, the red highlighted portion of the pie moves counterclockwise. In the last column, the red highlighted portion moves clockwise. This inconsistency forces readers to guess which portion represents the quantity described by the column's label.

In a data set of this size, including numbers is a good idea. Turning the pies into so-called “doughnut” charts, with a hole in the middle, will leave room for this number and make the chart easier to scan. (As a side note: Doughnut charts also encourage readers to focus on arc-length, instead of angles, which can help with accurate perception in pie charts with more than two categories.)

Example 6: Using the same information twice



Summary Scorecard on Budget Transparency in Selected States

Transparency Parameters	Andhra Pradesh	Assam	Chhattisgarh	Gujarat	Jharkhand	Madhya Pradesh	Maharashtra	Odisha	Rajasthan	Uttar Pradesh	Average for Selected States
	Average Transparency Score										
Availability of Budget Documents	68	67	65	87	72	68	65	68	80	64	70
Completeness of the Information	75	74	81	85	74	81	77	75	56	69	75
Facilitating Understanding and Interpretation of the Information	51	50	39	65	64	35	70	47	71	42	53
Timeliness of the Information	59	51	77	77	53	84	53	69	25	33	58
Audit and Performance Assessment	39	29	55	39	23	67	35	31	35	35	39
Scope for Legislative Scrutiny	50	55	43	55	38	62	41	60	36	36	47

Labels should not be separated from data. In this example, identifying that the top chart shows overall budget transparency scores is much harder than it should be, because that label can be found only at the very bottom of the page.

It is usually best to sort a table or a chart by a meaningful metric instead of alphabetically (or, as in the case of the top chart here, reverse alphabetically). To make variation more immediate, bar charts can be incorporated within a table. This would prevent the information from being repeated at the top and the bottom. Alternatively, the highest and lowest values in a table can be identified with shading to show variation and patterns at a glance.

The size of different elements is one good way to convey relative importance. Here, the size of the map is too large to merely identify the selected states. Depending on the intended audience, labels should be provided, or the map should be much smaller.

A side note: while this example is a print graphic, some recent interactive league tables have successfully allowed users to place different weights on individual metrics to develop their own averages. (See e.g. http://nymag.com/realestate/neighborhoods/2010/65355/)

Example 7: Repetitive information

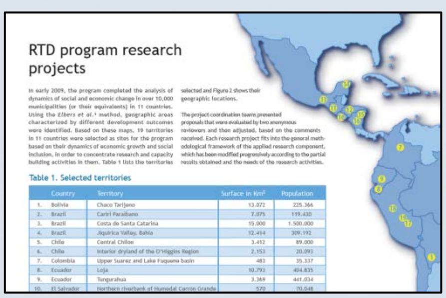
Good charts anticipate questions that readers are likely to have. Here, anomalies like why the data for Argentina is so out of date are explained. Sorting the table by the value of interest, instead of alphabetically by country, makes it easy to identify the highest and lowest values.

Choosing the breaks for the groups in "friendly" round numbers like 10% is another nice touch, but the "Grupo" column is too dominant in the table. Grid lines could separate the groups, making the label necessary only once per group. This would also connect the table to the map in a stronger way.

In tables where it doesn't make sense to have grid lines separate groups, one rule-of-thumb is to use a line after every third row. This helps with reading because it makes each row very easy to identify: the row either has one grid line above it, one below it, or neither.

While a continuous color scheme is the right choice here, it could be more aggressive. With the current palette, it is not trivial to distinguish between the middle two colors on the map.

Example 8: Precision



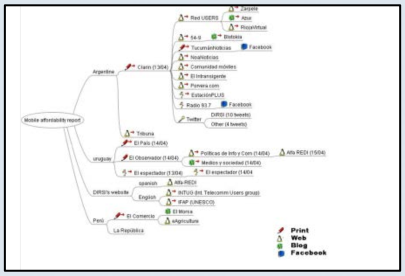
Including actual footprints of the research projects on this map would make the visualization more sophisticated, especially since one of the purposes of this table seems to be to show the size of the projects. This might not be meaningful for the smaller territories, but it would certainly be possible for the larger ones.

Precision seems to vary across the table. To facilitate comparisons and to make the table easier to read, population figures could be rounded to the nearest thousand or hundred.

Sorting the table by something more meaningful than country name would make any patterns within the data easier to recognize. One option would be latitude, so the table pairs better with the map. (Readers who are hoping to look up an individual country are likely to start with the map anyway.) This would make it clear that, without the facing page in print, Brazil is missing. Another option would be one of the columns in the table.

Charting forms

Example 9: When a list is just a list

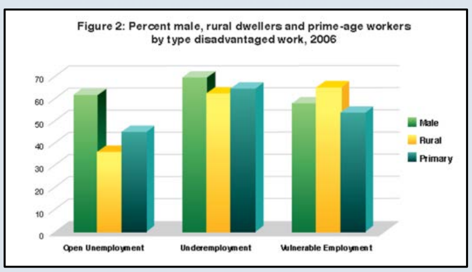


The structure of your data should help determine the kind of chart that is used. In this case, an indented list or an outline would be simpler and more effective than a network diagram, because the structure of the data is a hierarchy, and not a set of connections, which is where networks excel.

Titles should be specific to what is actually shown. Neither “tarriffs” nor an “affordability gap” seem to appear in the diagram, though it is difficult to tell.

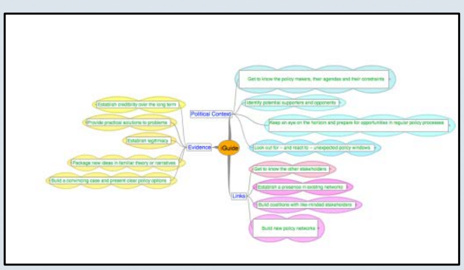
Drawing coherent icon sets is quite difficult. But when well-known icons are available – for example, in the case of Twitter and Facebook – it often makes sense to use them.

Example 10: Avoiding 3D



Almost all visualization experts recommend avoiding 3D. The reason is simple: it makes charts more difficult to read accurately. For example, here, the percentage of the open unemployed who are male appears to touch the 60% axis. But the actual value is likely to be around 58%. The perspective 3D implies makes it difficult to know for sure.

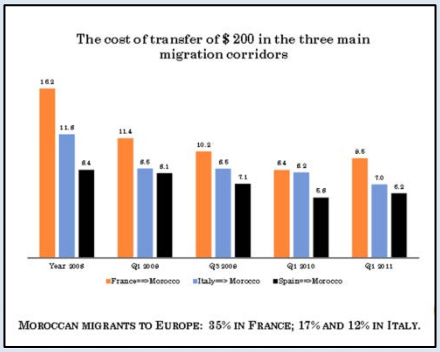
In charts with only nine numbers, the numbers should usually be included on the chart. Why? If your reader finds something surprising, you want it to be easy for them to write or talk about your data.



Example 11: Avoiding scavenger hunts

Flash is a poor way to display large amounts of text. If you are interested in the entire guide shown here, it is difficult to remember which sections you have already visited. If you are looking for a specific piece of information, it's not easy to copy it to your own notes, to share a particular section with a coworker or to search the text for key words. New guides should be in HTML and CSS, perhaps with a small amount of JavaScript to show and hide different levels of information. Without strong links between the different sections of the guide, a well-designed list is much easier to browse and skim.

Example 12: Displaying changes over time



Research on how people interpret charts suggests that line charts are best at conveying movement across time, particularly when the quantity being measured does not start over at zero with each new time period. A line chart would make the patterns shown in this chart more immediately obvious. Lines can also easily convey that the data is not spaced equally over time, so slopes are not misinterpreted.

For certain audiences, the language used to describe this chart might be friendlier. For example, a headline might read “How much does it cost to transfer $200 to Morocco?” The lines would be directly labeled “From France,” “From Italy” and “From Spain.”

Example 13: What's unique about your data?

In this example, the cost of mobile voice service is compared to the cost of cooking oil. Using units that are likely to be familiar to the intended audience is one of the most important steps in making data meaningful.

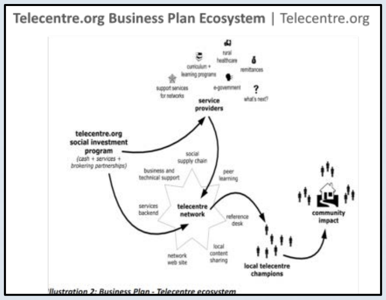
Recognizing that a map is not always the best form for geographic data is admirable and using a picture of cooking oil makes the video memorable. (Attempting to “crowd-source” Coca Cola prices – which may be a better base unit than cooking oil, but were not readily available – is also admirable.)

But the video becomes a bit repetitive, in part because it is difficult to store more than a handful of numbers in working memory. The video for one time period – here, June – is unlikely to feel any different from the video for any other period, even if the data changes dramatically.

One of the unique aspects of this data is that it is about time. Even better: all of the times are less than one hour. A clock metaphor would allow more positions to be stored in viewers' memory.

Clarity

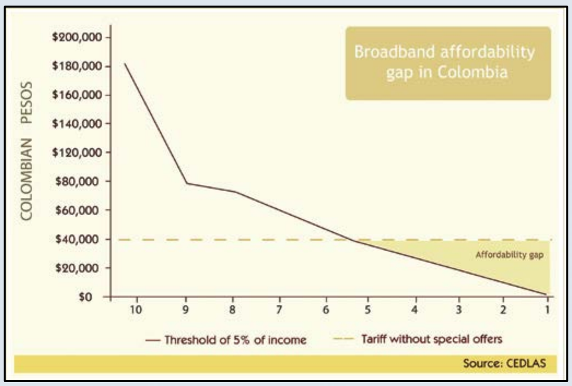
Example 14: Where to start?



In this diagram, the flow of the arrows suggests that a good starting point would be the "Teleconferencing social investment program" node. But, in English, people read from left-to-right and from top-to-bottom, so the "service providers node" is also competing for the starting position. Placing the "investment program" node on top (or the title on the left) would resolve this conflict.

Presumably, the arrows do not all represent the same action. Clarity could be improved by placing text on each connection, describing what the arrow actually means ("Provides funding," say).

Example 15: Clear labels



This example emphasizes the importance of clearly labeling a chart. It is not clear what the x-axis on this chart represents. Income deciles seem likely, though if the headline read: "Half of Columbia cannot afford broadband," readers would not be forced to guess, even without a label.

A good rule-of-thumb in designing both simple and complicated charts is to minimize eye movement. Minimizing eye movement turns reading a chart into less of a decoding exercise. Here, that would mean placing labels directly on the lines.

Notice how the "affordability gap" label is more successful because it is placed directly on the gap, instead of being moved into the legend at the bottom of the chart.

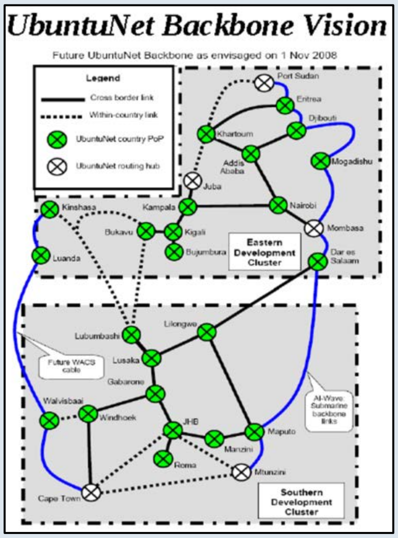
Example 16: But what does it mean?

Edward Tufte uses the term "small multiples" to describe a group of similar charts that display different slices of a data set. Because small multiples allow readers to quickly and easily make comparisons, it is often a very effective technique, and one that works well here.

But the visualization could be made stronger by describing what experts see in each map directly next to it (or in text on top of it in the case of a blog article). For example: why are so many Swahilli Wikipedia articles written in Turkey? "The answer is simply a few dedicated editors creating stub articles about relatively structured topics." This explanation feels disappointing. Is every interesting pattern as easily explained? Could the data be filtered to remove stubs?

The maps are visually attractive, though. Compare the country outlines and ocean here to Example 1. Because of the design choices, the data is prominent here, not the background information.

Example 17: Emphasizing what's important



Data visualization is about abstraction. So it is fine – and perhaps even helpful – to move away from literal geography in some cases, such as this example, even though the underlying data has a strong connection to a map. But once you move into abstraction, choices should be clear. Is there a reason the future WACS cable moves outside the Southern cluster? Do the horizontal positions of the cluster boundaries mean anything?

Small changes would make this sketch clearer. For example, there is no need to outline the development clusters with a thick dashed border, especially when a dashed line holds some meaning within the diagram. A blue line should appear in the legend, even when the blue lines are labeled individually. If the blue lines are the focus of the graphic, their labels should be bold, while the labels for the development clusters should be placed in a consistent way.

Interaction

Example 18: Details-on-demand

Ben Shneiderman, a computer scientist who developed some of the early ideas on interaction design, has a few words he often repeats. He says: “overview, zoom & filter, *details-on-demand.” This graphic provides an overview and zoom capabilities, but it does not allow filtering, or, more importantly, substantial details-on-demand. Clicking the countries should update the table below the map with details on individual measures. (Consider which of the following is more compelling: “Changed the rules on importing aquatic animals” or “40.”) Filtering by date would allow returning users to track what is new.*

Critically, the circles on the protectionist and liberalizing maps should be scaled in the same way to allow easy comparisons between the maps.

User interaction might also be improved. With the type of rollovers used here, the mouse must directly touch a circle before its information box is displayed. This type of interaction is known as hit detection. Instead of hit detection, many modern visualization toolkits find the nearest element as the mouse is moved, which would prevent the information box from flickering on and off. Compare the experience of traveling over the map with the smoothness of an example like this: http://mbostock.github.com/d3/ex/voronoi.html

Example 19: Meaningful interaction

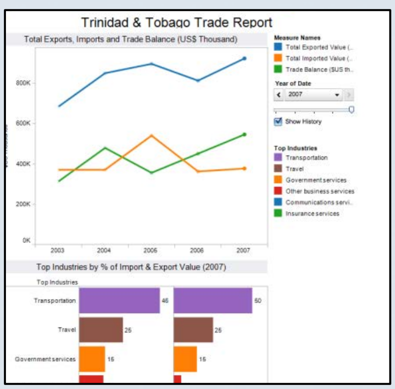
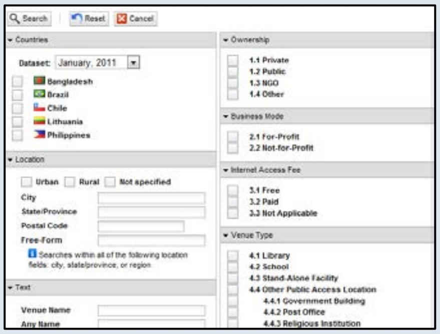


Tableau is a useful tool for exploratory analysis, and it has a low learning curve for creating certain types of interactivity. But some of Tableau's drawbacks for presentation reveal themselves in this interactive example. Keys get cut off. The legend for the bar chart is oddly disconnected from the chart. (In fact, it's not clear why this legend is necessary at all, since the labels are repeated on the actual graph.)

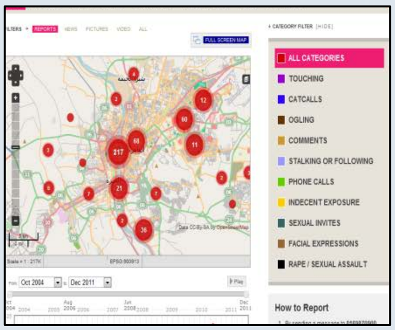
This visualization also features a time slider which reveals each segment of the line chart. However, without annotation describing why certain points are interesting, the slider is distracting. Interactivity that merely hides data shown in a simple static line chart is not useful. Analysts may find Tableau – and the drilling into data it sometimes encourages – revelatory, but presentation for wider audiences may require additional finesse or other tools.

Example 20: Determining intent



Data visualization should be judged according to how well it does what it intends to do. This example, a filterable database of different venues, is clearly intended for professional users. (A barrier that forces users to sign in makes that clear.) For casual users, a blank default screen is intimidating. But that may be exactly what professional users appreciate. The experience of the map and charts could be improved by not requiring a full refresh when query parameters are changed, but that may require more work than is justified.

Example 21: Defining success



In some cases, the mere existence of data may be what is powerful for outsiders. And very local, real-time data may be the sort that most affects people's lives.

This example does both: it demonstrates that data exists, and allows people to look up incidents in their own neighborhoods.

The refresh on the map is too slow to encourage much interaction, though, assuming users do not already know what they are looking for. With a fast internet connection, using the filters takes as long as a second. This is a short amount of time, but it is at least ten times longer than the time frame that feels immediately responsive. After even a second, it can be difficult to remember the pattern that was previously shown on the map to compare or contrast with the new view. .Finally, the scale on the chart at the bottom suggests the project has run much longer than its developer anticipated.

That said, those quibbles – or concerns about how representative the data is likely to be – are unlikely to matter here. The individual data points are compelling for both outside observers and the local population.

Conclusion

Two simple steps would improve the power of the Center's visualization work.

First, every static visualization should include a headline or other language that describes the findings of the visualization in a meaningful way. What is its takeaway message? In many cases, the projects have made strong and thoughtful conclusions about what the data means and why it is important in the text accompanying the visualization. These conclusions should be repeated, in a concise way, within the visualization. A quick check: does the headline or other prominent text include a verb?

Second, the conclusions of the visualization should shape its design. Designers should think about how the choices they make with color or type help guide readers to interesting findings. Would a line or two of text pointing directly to the most interesting parts allow readers to see patterns or relationships they might otherwise miss?

For interactive work, the first step is to decide on a goal. Work that primarily allows people to look up information about themselves or their communities will likely be quite different from work that intends to show broader patterns or trends. So far, a lot of successful interactive work – within the Centre and the larger data visualization community – falls into the former category. But interactive work that incorporates explanations or annotations is becoming more common, and it may mean that interactivity plays a more prominent role in communicating research in the near future. Already, this trend is clear in text books.

Finally, the Centre should critically examine results from projects like the UN's Global Pulse (http://www.unglobalpulse). This lab has been a leader in data visualization within the development space, and its work may help the Center consider whether experimenting with larger data sets or new forms of data collection would be useful for its own mission.

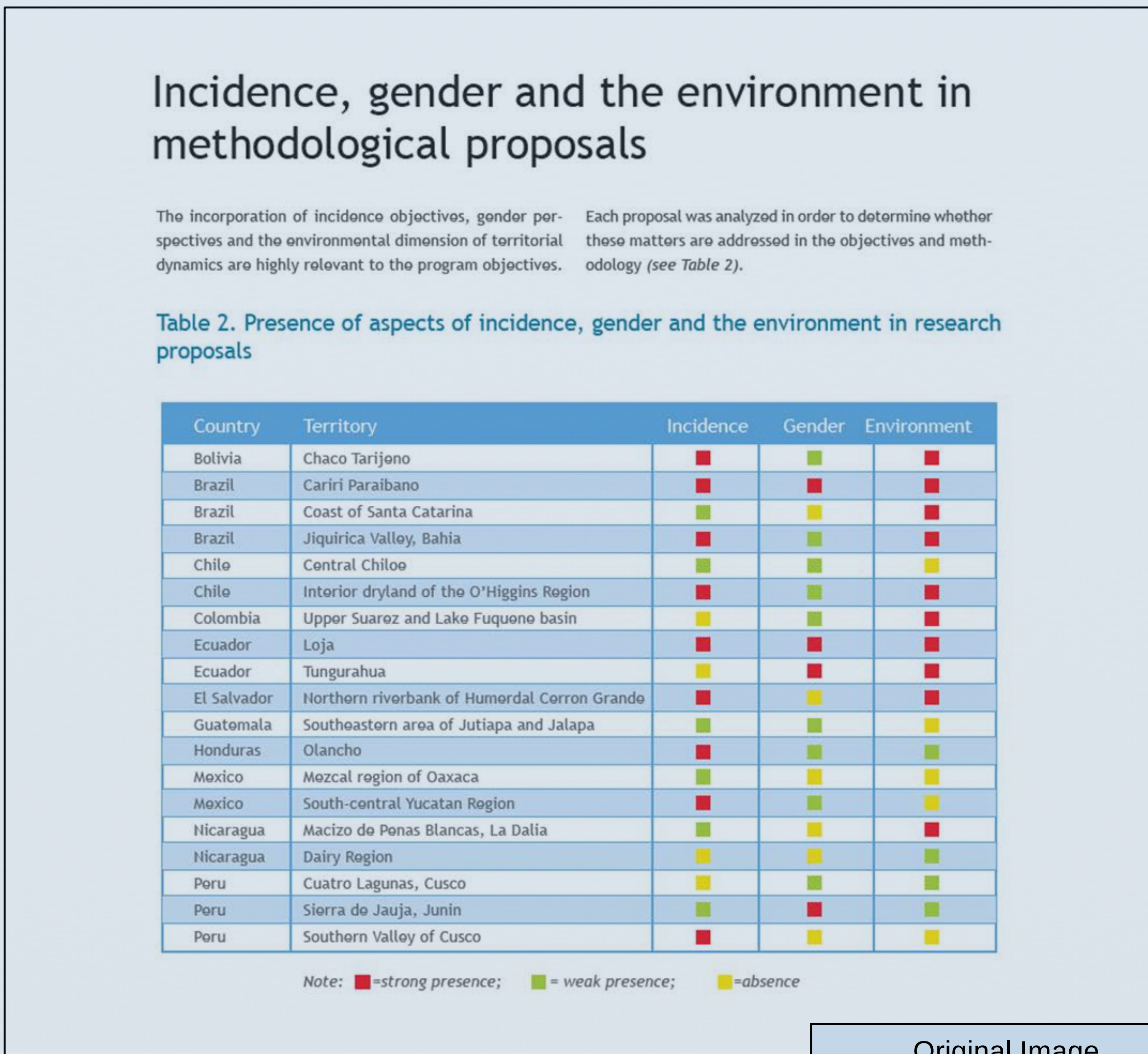
Models for Success

Many IDRC-staff and research partners have commented that it is useful to refer to good design models when deciding what types of visualizations to apply to their work. It can also be extremely useful to see how designers interpret and adapt designs based on their experience with data visualizations. Based on this rationale, this study took the opportunity to utilize Amanda Cox's extensive knowledge to redesign four of the reviewed examples, to demonstrate good practice. The final example in this series takes on a different form than the original example, and is meant to highlight the potential uses for applying more interactive designs to communicate research findings.

Rational for Redesigns

The following examples demonstrate ways to use color, sorting and charting forms to visualize data effectively. In each of the redesigns, I (Amanda Cox) tried to stay reasonably faithful to the original visualizations. For example, the overall sizes of static graphics were not changed. In three of the four cases, I used the typography and color palettes from the original designs. My assumption was that the visualizations would remain parts of larger reports or Web sites, so I did not worry about sourcing or deep explanations of methodology.

1. Incidence, gender and environment in methodological proposals



Incidence, gender and the environment in methodological proposals

The incorporation of incidence objectives, gender perspectives and the environmental dimension of territorial dynamics are highly relevant to the program objectives. Each proposal was analyzed in order to determine whether these matters are addressed in the objectives and methodology *(see Table 2)*.

Table 2. Presence of aspects of incidence, gender and the environment in research proposals

Country	Territory	Incidence	Gender	Environment
Bolivia	Chaco Tarijeno	red	green	red
Brazil	Cariri Paraibano	red	red	red
Brazil	Coast of Santa Catarina	green	yellow	red
Brazil	Jiquirica Valley, Bahia	red	green	red
Chile	Central Chiloe	green	green	yellow
Chile	Interior dryland of the O'Higgins Region	red	green	red
Colombia	Upper Suarez and Lake Fuquene basin	yellow	green	red
Ecuador	Loja	red	red	red
Ecuador	Tungurahua	yellow	red	red
El Salvador	Northern riverbank of Humerdal Cerron Grande	red	yellow	red
Guatemala	Southeastern area of Jutiapa and Jalapa	green	green	yellow
Honduras	Olancho	red	green	green
Mexico	Mezcal region of Oaxaca	green	yellow	yellow
Mexico	South-central Yucatan Region	red	green	yellow
Nicaragua	Macizo de Penas Blancas, La Dalia	green	yellow	red
Nicaragua	Dairy Region	yellow	yellow	green
Peru	Cuatro Lagunas, Cusco	yellow	green	green
Peru	Sierra de Jauja, Junin	green	red	green
Peru	Southern Valley of Cusco	red	yellow	yellow

Note: red =strong presence; green = weak presence; yellow =absence

Original Image

What's in the proposals?

Each proposal was analyzed in order to determine whether incidence, gender and environment perspectives are addressed in the objectives and methodology. Only four of the proposals address gender matters with specific questions or activities. As a result, the program's 2010 Plan of Action includes a specific line of work aimed at strengthening gender analysis in territorial dynamics.

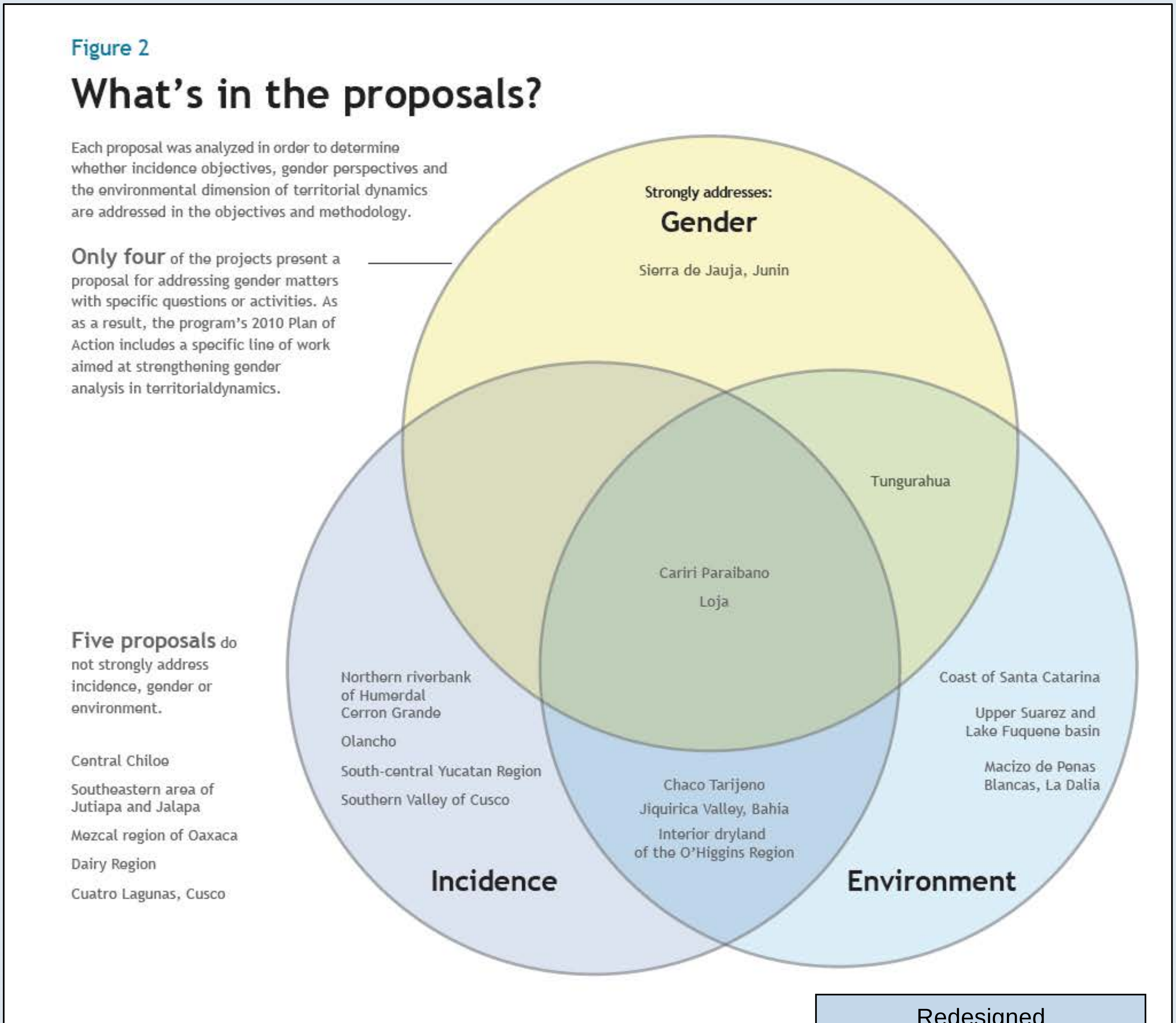
Table 2. Incidence, gender and the environment in research proposals

Country	Territory	Incidence	Gender	Environment
Bolivia	Chaco Tarijeno	Strong presence	Weak presence	Strong presence
Brazil	Cariri Paraibano	Strong presence	Strong presence	Strong presence
	Coast of Santa Catarina	Weak presence	Absence	Strong presence
	Jiquirica Valley, Bahia	Strong presence	Weak presence	Strong presence
Chile	Central Chiloe	Weak presence	Weak presence	Absence
	Interior dryland of the O'Higgins Region	Strong presence	Weak presence	Strong presence
Colombia	Upper Suarez and Lake Fuquene basin	Absence	Weak presence	Strong presence
Ecuador	Loja	Strong presence	Strong presence	Strong presence
	Tungurahua	Absence	Strong presence	Strong presence
El Salvador	Northern riverbank of Humerdal Cerron Grande	Strong presence	Absence	Strong presence
Guatemala	Southeastern area of Jutiapa and Jalapa	Weak presence	Weak presence	Absence
Honduras	Olancho	Strong presence	Weak presence	Weak presence
Mexico	Mezcal region of Oaxaca	Weak presence	Absence	Absence
	South-central Yucatan Region	Strong presence	Weak presence	Absence
Nicaragua	Macizo de Penas Blancas, La Dalia	Weak presence	Absence	Strong presence
	Dairy Region	Absence	Absence	Weak presence
Peru	Cuatro Lagunas, Cusco	Absence	Weak presence	Weak presence
	Sierra de Jauja, Junin	Weak presence	Strong presence	Weak presence
	Southern Valley of Cusco	Strong presence	Absence	Absence

KEY: Strong presence Weak presence Absence

Redesigned

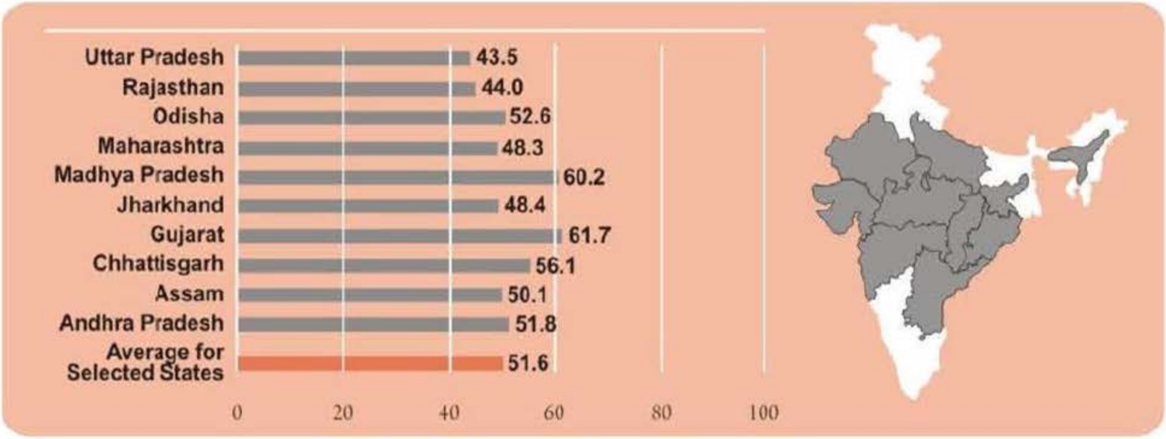
I have attached two versions of this table. The first (shown above) makes only simple, cosmetic changes. Most importantly, I updated the color scheme so it is more intuitive. I moved the key so it is closer to the corresponding information and easier to reference. I removed duplicate country labels and used the rules in the table to separate the countries. This makes patterns by country easier to see. For example, the proposals from Brazil are much more likely to strongly address gender, incidence or environmental perspectives than those from Peru. I adjusted the language in the precede to address the findings of the study, and the actions taken as a result of the data.



Redesigned

The second version is a venn-diagram. This version drops the weak presence indicator and the country names from the table. If either of these columns is critical, this is a bad idea. But removing some information can make patterns easier to see. For example, it is now immediately obvious that gender is addressed less frequently than environment. A text call-out emphasizes this point. Different types of overlap (e.g. which proposals address incidence and environment, but not gender?) are not immediately obvious.

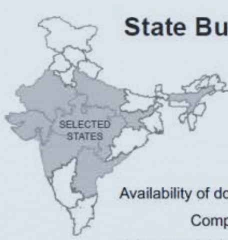
2. Summary scorecard on Budget Transparency



Summary Scorecard on Budget Transparency in Selected States

Transparency Parameters	Andhra Pradesh	Assam	Chhattisgarh	Gujarat	Jharkhand	Madhya Pradesh	Maharashtra	Odisha	Rajasthan	Uttar Pradesh	Average for Selected States
	Average Transparency Score										
Availability of Budget Documents	68	67	65	87	72	68	65	68	80	64	70
Completeness of the Information	75	74	81	85	74	81	77	75	56	69	75
Facilitating Understanding and Interpretation of the Information	51	50	39	65	64	35	70	47	71	42	53
Timeliness of the Information	59	51	77	77	53	84	53	69	25	33	58
Audit and Performance Assessment	39	29	55	39	23	67	35	31	35	35	39
Scope for Legislative Scrutiny	50	55	43	55	38	62	41	60	36	36	47
Practices relating to Budgeting for Disadvantaged Sections	49	44	71	63	37	70	29	43	30	40	48
Practices relating to Fiscal Decentralisation	24	31	19	24	27	14	17	29	19	29	23
Overall Budget Transparency Score (in %)	**51.8**	**50.1**	**56.1**	**61.7**	**48.4**	**60.2**	**48.3**	**52.6**	**44.0**	**43.5**	**51.6**

Original Image



State Budget Transparency Scorecard

KEY: 0% 33 67 100

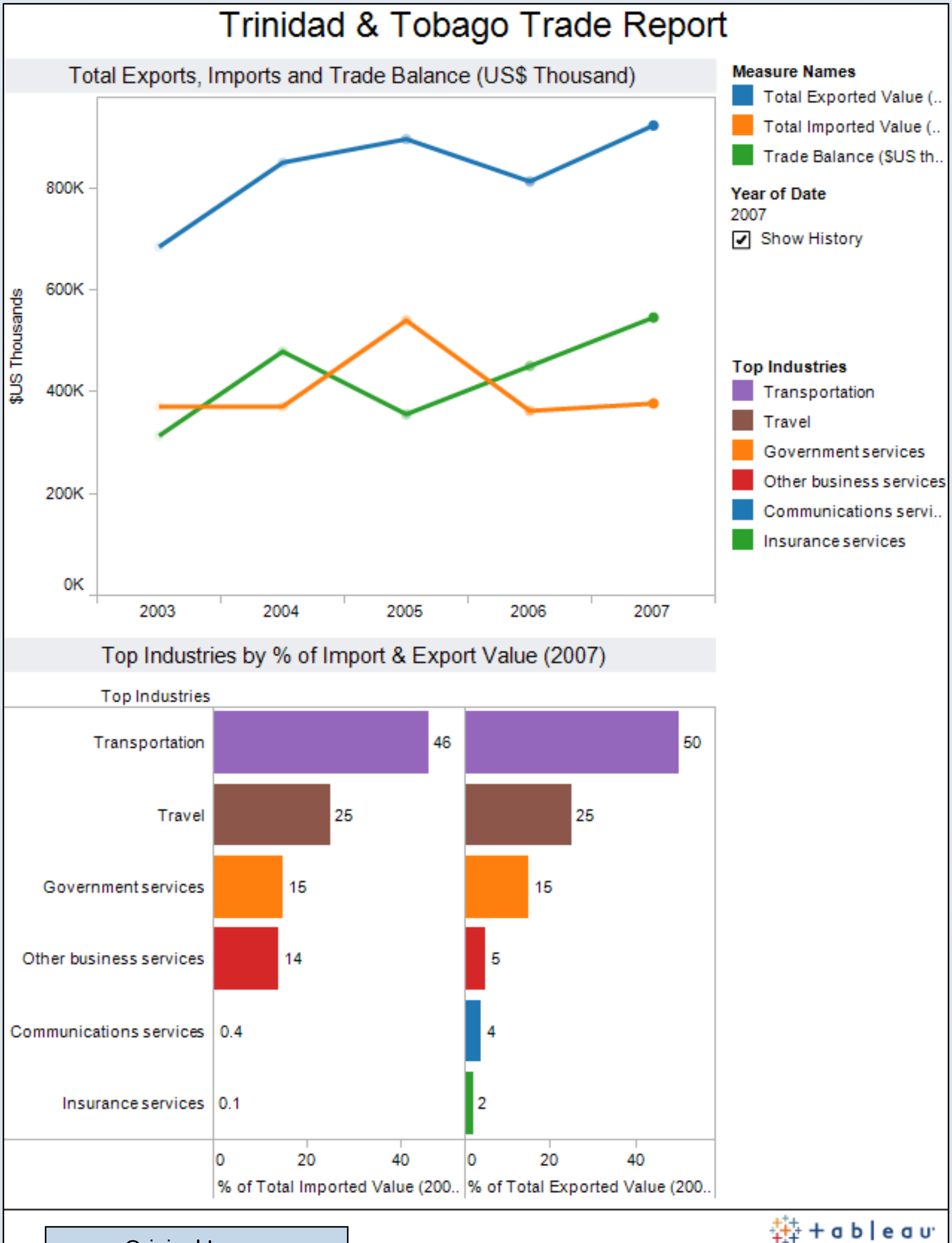
	Gujarat	Madhya Pradesh	Chhattis-garh	Odisha	Andhra Pradesh
Availability of documents	87	68	65	68	68
Completeness	85	81	81	75	75
Facilitating understanding	65	35	39	47	51
Timeliness	77	84	77	69	59
Audit and performance assessment	39	67	55	31	39
Scope for legislative scrutiny	55	62	43	60	50
Budgeting for disadvantaged sections	63	70	71	43	49
Fiscal decentralisation	24	14	19	29	24
Average score	**62**	**60**	**56**	**53**	**52**

Assam	Jharkhand	Mahara-shtra	Rajasthan	Uttar Pradesh	
67	72	65	80	64	Availability of documents
74	74	77	56	69	Completeness
50	64	70	71	42	Facilitating understanding
51	53	53	25	33	Timeliness
29	23	35	35	35	Audit and performance assessment
55	38	41	36	36	Scope for legislative scrutiny
44	37	29	30	40	Budgeting for disadvantaged sections
31	27	17	19	29	Fiscal decentralisation
50	**48**	**48**	**44**	**44**	**Average score**

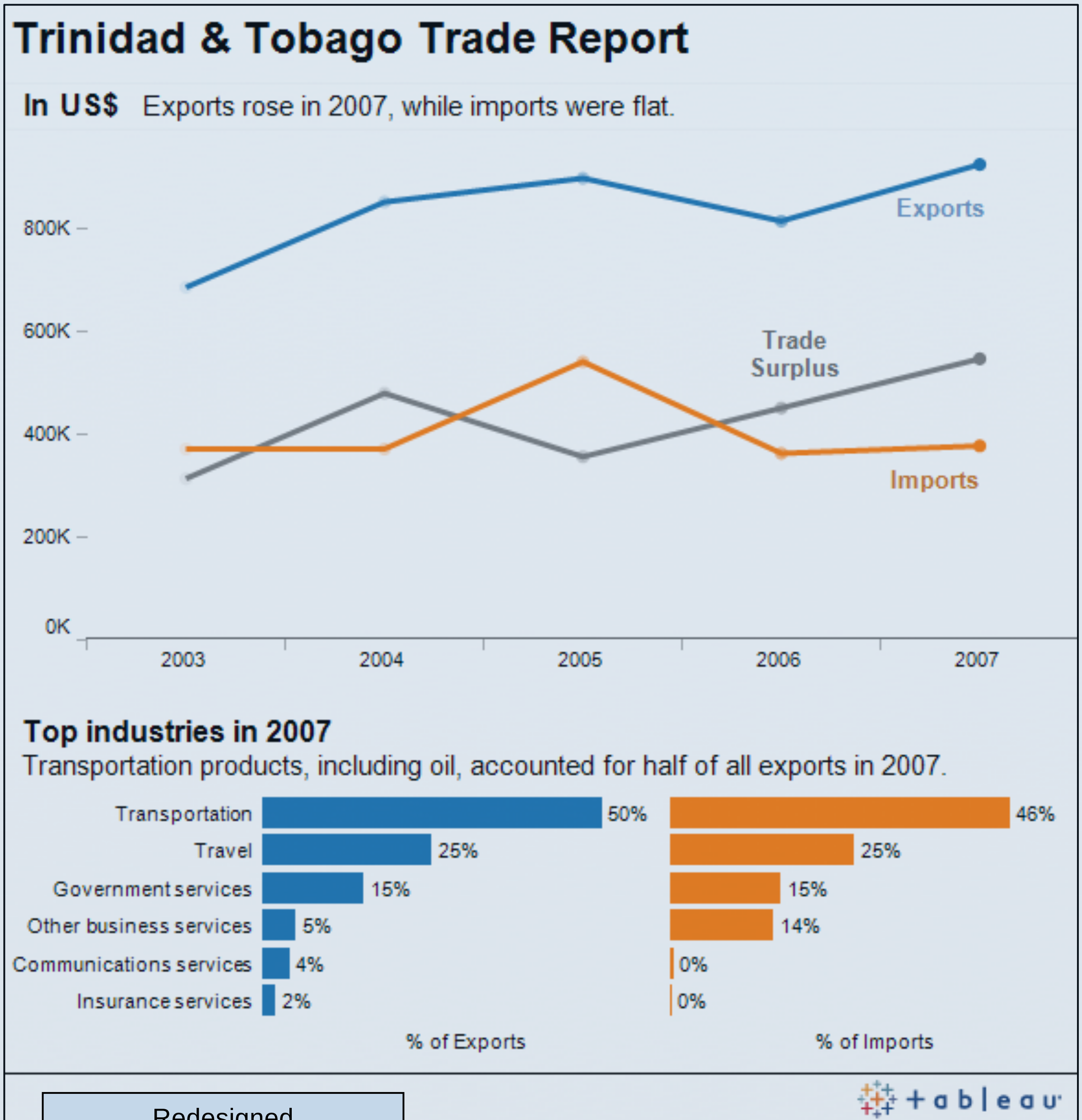
Redesigned

I sorted the states by average budget transparency score, not alphabetically. This means the order of the table corresponds to the value of interest; alphabetical sorting rarely reveals patterns. Vertical text is difficult to read, so I replaced it with horizontal text. I removed the decimal from the average score. Because of the way the data was collected, (using “A”, “B” or “C” grades and a corresponding value for each) this amount of precision is unlikely to be meaningful. I made the map of India smaller so its size is consistent with the amount of information it contains. I slightly shortened some of the labels, so they can be closer to the data. Most importantly, I added shading to the table, so patterns are easy to see. The breaks in the key correspond to grades in the original evaluation. Without studying the table, it is now obvious that document availability and completeness had high scores, while fiscal decentralization had low scores. The highest-scoring states generally had high scores for timeliness. Given its overall score, Odisha's audit and performance assessment score is very low, as is Rajasthan's score of timeliness. In the previous version of the chart, seeing these sorts of patterns and outliers requires close study.

3. Trinidad & Tobago Trade Report



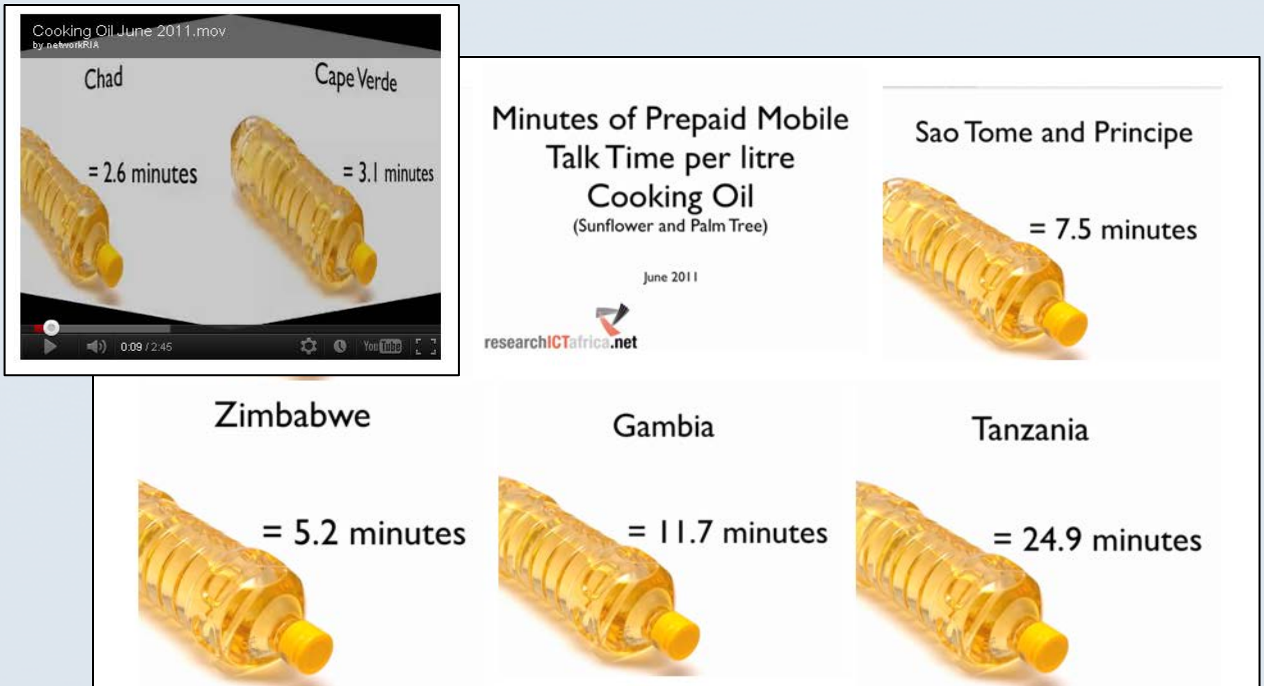
Original Image



Redesigned

For inexperienced developers, Tableau can be a very useful tool for adding basic interactivity to charts. Recognizing this, I wanted to leave this chart in Tableau, but redesign it, so that it is better suited for sharing with a broader audience. I removed unnecessary legends, and labeled the lines directly so understanding the chart is less of a decoding exercise. I changed the colors of the bar charts so they are linked to the top charts in a meaningful way. I renamed the variables so they are easier to read. For example, it is unnecessary to state that the data is shown in US dollars in eight unique places, including rollover. I made the initial view more informative by removing unnecessary interactivity. I added simple sentences describing what the data shows to take some of the burden off of the reader. With more data, I would have liked to have taken advantage of some more of Tableau's strengths. For example, including charts over time for each industry might help explain why the overall changes are happening. An option to show the trends in either U.S. dollars or local dollars would help explain how much of the changes are simply due to currency fluctuations.

4. Minutes of Prepaid Mobile Talk Time per Litre Cooking Oil



Original Interactive Display: researchictafrica.net

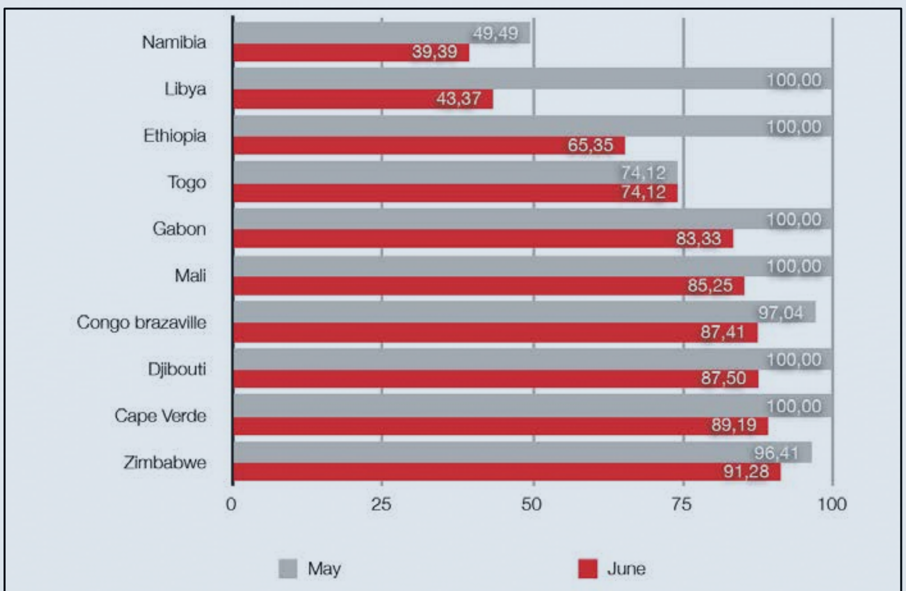
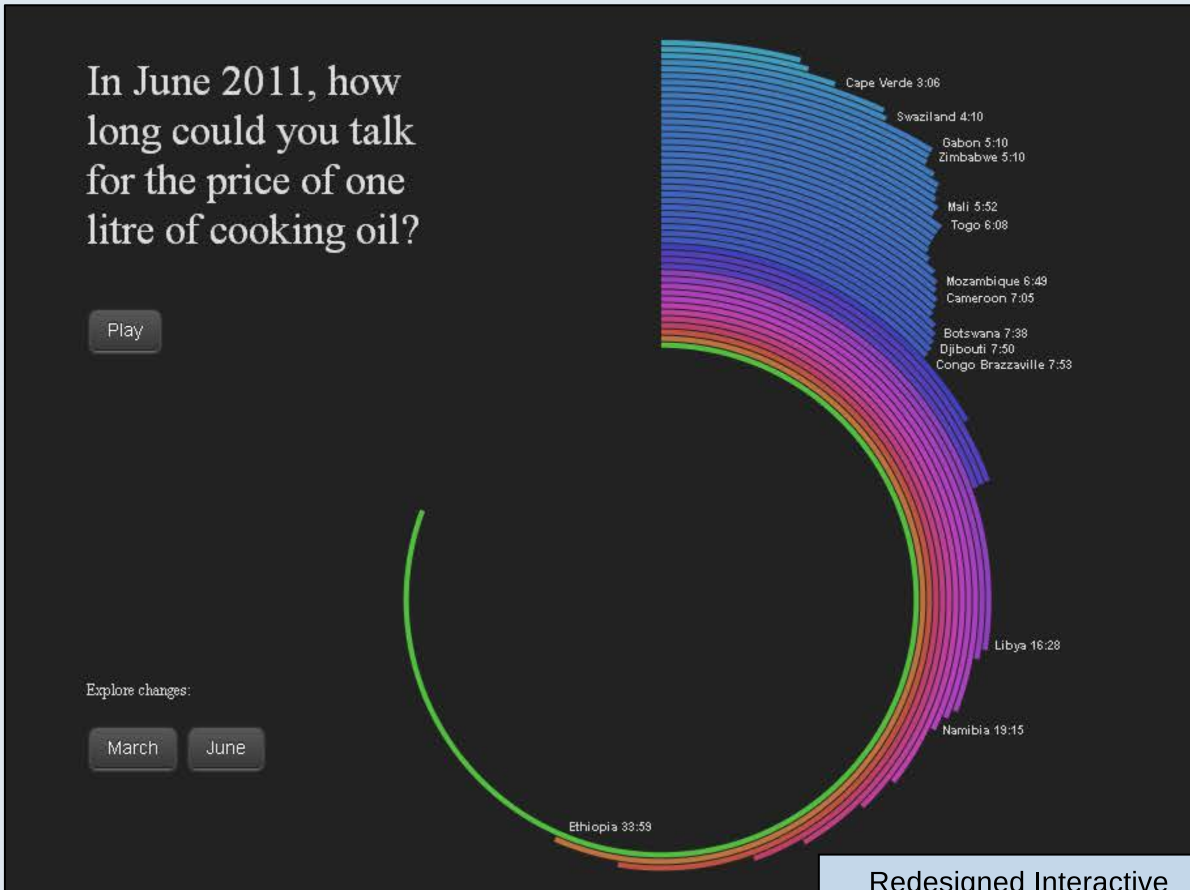


Figure 1: May/June 2011 Local price Index (base = March 2011)

In this case, I attempted to use strengths from two very different visualizations of the original data. One, a video of rotating numbers, is friendly, but it does not reveal any patterns within the data. After watching the videos for two different time periods, it is nearly impossible to know what has changed. The second, a bar chart, conveys relative magnitudes well, but it is perhaps less engaging than the video.



Redesigned Interactive

The redesign has two basic modes. The first, a play button, is similar to the video. But because the numbers change in a way that is linked to the data, it is easier to see that there are clusters of countries.

The shape of the chart, and how fast it moves, reflects the data. While the chart is playing, users can interact with the arcs, but this form could be turned into a straight video, which would make it compatible with older versions of Internet Explorer. A slider could also be provided for further control.

The second mode is better suited for exploratory use. It shows two time periods, literally highlighting countries that have changed. Outliers are easy to identify, and explanations for some countries are provided when a user interacts with an arc. For example, text is shown clarifying that, in Namibia, the dominant provider, MTC, cut prices, while in Ethiopia, prices are politically determined. The chart anticipates questions that readers are likely to have.

Redesign Conclusion

Unlike a bar chart, this chart exploits qualities that are unique about the data: most notably, the data is measured in units of time. Using what is unique about your data is a good way to make the visualization memorable.

Conclusion

Overall the findings from this 3 staged review provided important insights into how IDRC is currently implementing and using data visualizations. One of the largest limitations to section 2 is that it was unable to assess whether the visualizations were in fact effective at capturing the attention, or influencing the actions, of the intended audience. While it was not within the scope of this review to unpack this broader question of *influence*; this review does assess the degree to which IDRC-supported research has used data visualizations effectively to communicate. Assessing appropriate use and design is therefore an important step before evaluating data visualization influence. The hope is that further finding on the influence of data visualizations will start to emerge from the field, and perhaps from IDRC's strategic evaluation on communicating research for influence.

This review found that IDRC-supported research is engaging with data visualization use and using them into nearly half of all research outputs. The focus has largely been on standard line, bar and pie charts, and there appears to be a lack of understanding of how to tailor these visualizations so that they communicate a more focused and compelling message. Although data visualization use, still appears to be at a novice level, many of the Centre-staff and partners who were interviewed were very enthusiastic about the potential of data visualizations, but spoke to a need for greater knowledge and skills around how to strategically use data visualizations as effective communication tools.

It is therefore the intention of this report to draw attention to some of the basic principles of data visualization, and encourage further conversations amongst Centre staff and partners about what resources are required to more effectively communicate using data visualizations. This report is thus only the start of the conversation, and should provide some initial guidance as to where partners and IDRC staff can go to acquire further information and support.

References

Bederson, B.B., and B. Shneiderman (2003). *The Craft of Information Visualization: Readings and Reflections*. San Franciso: Morgan Kaufman.

Bresciani, Sabrina, and Martin J. Eppler, “The Risks of Visualization: A Classification of Disadvantages Associated with Graphic Representations of Information”, *ICA Working Paper # 1*, (2008).

Burch, Michael, and Julian Heinrich, Natalia Konevtsova, Markus Hoferlin, Daniel Weiskopf, “Evaluation of Traditional, Orthogonal, and Radial Tree Diagrams by an Eye Tracking Study” *IEEE Transactions on Visualization and computer graphic,* v.17, n. 12 (2011).

Checkland, Peter, and John Poulter, “Softsystems methodology” in M. Reynolds and S. Howell (eds.), systems approaches to Managing Change (London: Open University, Springer-Verland, 2010), 191-242.

Chen, Chaomei, *Information Visualization and Virtual Environments* (London: Springer, 1999).

Chen, Chaomei, *Information Visualization: Beyond the Horizon*, 2nd Ed. (London: Springer-Verlag, 2006).

Chen, Chaomei, “Top 10 Unsolved Information Visualization Problems”, *IEEE Computer Graphics and Applications* (July/August 2005), pp.12-16

Chen, Chaomei, “Overview: Information Visualization”, *WIREs Computational Statistics*, v.2 (July/August 2010), pp.387-403.

Chen, Chaomei, and M. Czerwinski, “Empirical evaluation of information visualizations: An introduction”, *International Journal of Human-Computer Studies*, v. 53 (2000), pp.631-635.

Chen, C. and Y. Yu, “Empirical studies of information visualization: A meta-analysis”, *International Journal of Human-Computer Studies*, v.53, n.5, (2000), pp.851-866.

Chen, Chun-houh, Wolfgang Hardle, and Antony Unwin (eds.), *Handbook of Data Visualization* (London: Springer-Verlag, 2008).

Friendly, Michael, “A Brief History of Data Visualization” , Chapter II.1 in Chun-houh Chen, Wolfgang Hardle, and Antony Unwin (eds.), *Handbook of Computational Statistics* (Berlin: Springer 2008), pp.15-56.

Few, Stephen, *Information Dashboard Design: The Effective Visual Communication of Dat*a. O’Reilly, 2006.

Few, Stephen, “What Ordinary People Need Most from Information Visualization Today”, *Perceptual Edge: Visual Business Intelligence Newsletter*, August 2008.

Few, Stephen, *Now You See It: Simple Visualization Techniques for Quantitative Analysis* (Oakland, CA: Analytics Press, 2009).

Iliiinsky, Noah, “On Beauty”, Ch. 1 in Steele and Iliinsky (eds.), *Beautiful Visualization* (Cambridge: O’Reilly, 2010), pp.1-13.

Kosara, Robert, and Silvia Miksch, Helwig Hauser, "Focus + Context Taken Literally", *IEEE Computer Graphics & Application (CG&A), Special Issue on Information Visualization,* v. 22 n.1 (2002): 22-29.

Lindquist, Evert. "Grappling with Complex Policy Challenges: Exploring the Potential of Visualization for Analysis, Advertising and Engagement", *Discussion Paper for HC Coombs Policy Forum Round tables, (2011).*

Lindquist, Evert. "Surveying the World of Visualization" *Background Paper prepared for the HC Coombs Policy Forum, (2011).*

McCandless, David, *Information is Beautiful* (London: Collins, 2009). Also published as *The Visual Miscellaneum: A Colorful Guide to the World's Most Consequential Trivi*a (2009).

Malik, Abish and Ross Maciejewski, Ben Maule, David S. Ebert, "A Visual Analytics Process for Maritime Resource Allocation and Risk Assessment", *IEEE Presentation (2011).*

Quora, "Data Visualization vs. Information Visualizations", www.quora.com, (2012).

Segel, Edward, and Jeffrey Heer, "Narrative Visualization: Telling Stories with Data", *IEEE Transactions on Visualization and Computer Graphics*, v.16. n.6 (Nov-Dec 2010), pp.1139-48.

Shneiderman, Ben, "Foreward" in Chaomei Chen, *Information Visualization: Beyond the Horizon* (London: Springer-Verlag, 2006).

Shneiderman, Ben, "The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations", paper prepared for the Proceedings of the Visual Languages 1996 Conference (September 1996).

Steele, Julie, and Noah Iliinsky (eds.), *Beautiful Visualization: Looking at Data Through the Eyes of Experts* (Cambridge: O'Reilly, 2010).

Stone, Maureen. "Choosing Colors for Data Visualization", Business Intelligence Network (2006).

Treisman, A, "Preattentive Processing in Vision", *Computer Vision, Graphics, and Image Processing*, v.31 n.2 (1985): 156-177.

Treisman, A, "Features and Objects in Visual Processing", *Scientific American,* v. 255 n.5 (1986): 114-125.

Tufte, Edward R., *Envisioning Information* (Chesire, Conn.: Graphics Press, 1990).

Ware, Christopher. *Information Visualization: Perception for Design*. San Francisco: Morgan Kaufmann, 2000; 2nd Edition, 2004.

Ware, Christopher, *Visual Thinking for Design*. San Francisco: Morgan Kaufmann, 2008.