IDRC Data Visualization Review

Written by Amanda Cox

Data visualization is not a new concept for IDRC, or for its partners. IDRC-supported research has dabbled in visualization 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 remain the same regardless of whether your data visualization is static or dynamic.

As IDRC-supported research has used data visualizations, the Evaluation Unit commissioned a series of visualizations to be review by Amanda Cox, Graphics Editor at the New York Times. The set are data visualizations nominated as being good quality by IDRC staff and from a review of recent documents filed in IDRC's Digital Library.

The examples highlighted represent a range of forms and levels of difficulty, but are not meant to represent an exhaustive assessment of data visualization use. Instead, this review is organized to assist learning, and thus uses each example to illustrate larger design principles which could be applied to data visualizations across the Centre's work.

A framework for success

The framework for this review 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.

For the purposes of this review, clarity is defined by four main questions: 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 relates to the following questions: Does the language used in the visualization support at least one specific idea? Do design choices such as colours, 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?

Finally, 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?

Summary

A review of 21 data visualizations produced by IDRC'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 ap-

propriate 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. Colour, typography, or highlighted areas were rarely used to draw attention to points of interest. Using colour 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. These 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

Example 1

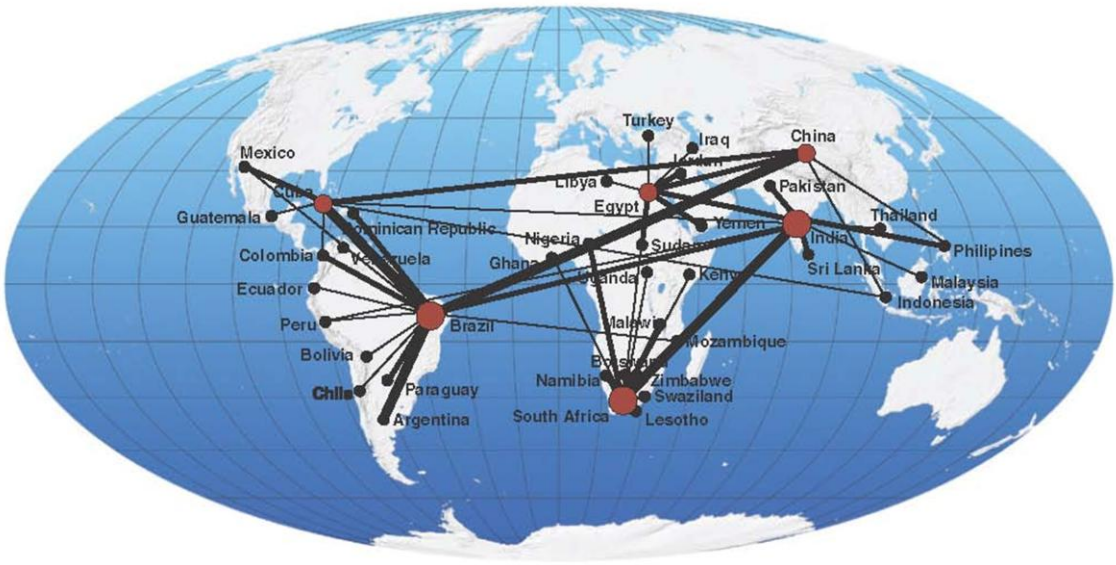


Figure 3 Collaboration network of health biotech firms in South-South collaborations. The size of each node represents the total number of South-South collaborations for the country, while the width of each line represents the number of collaborations between the two linked countries. For clarity, only linkages of two or more collaborations were included on this map.

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 to 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: colour, 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.

Colour

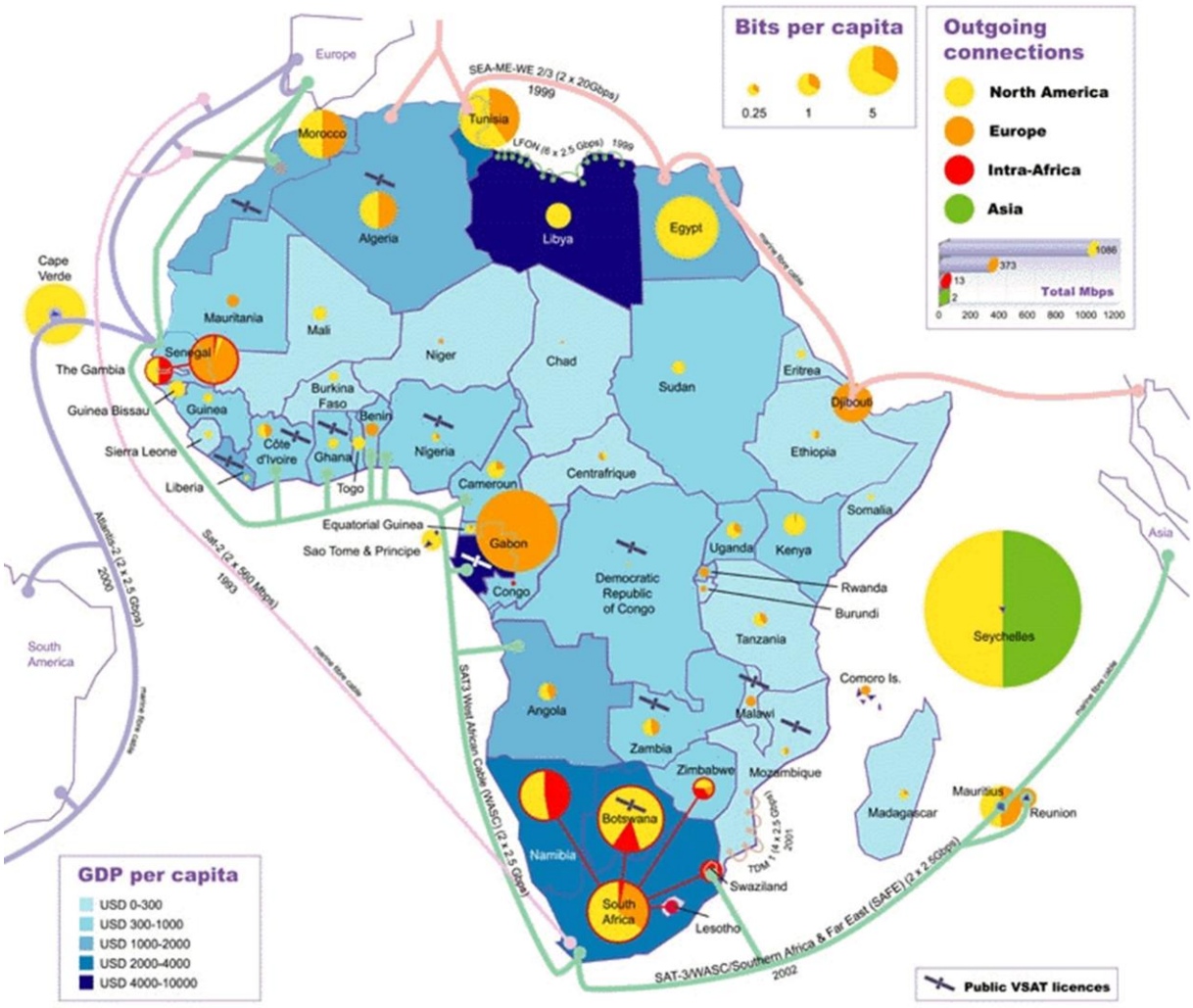
Example 1: Designing for Emphasis

Design choices should help a reader determine what is important. In the example above, 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 coloured 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 may further reduce clutter.

Example 2



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 the map above 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 colours chosen to represent each country's wealth make the fact that Libya is wealthy jump out.

Example 3

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 Humerdal 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

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 colour, is very useful in immediately conveying relative importance.

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

Example 3: Matching Expectations

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

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

But, above, 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 colour-blind readers, who would be able to distinguish between some kind of presence (red and

Example 4

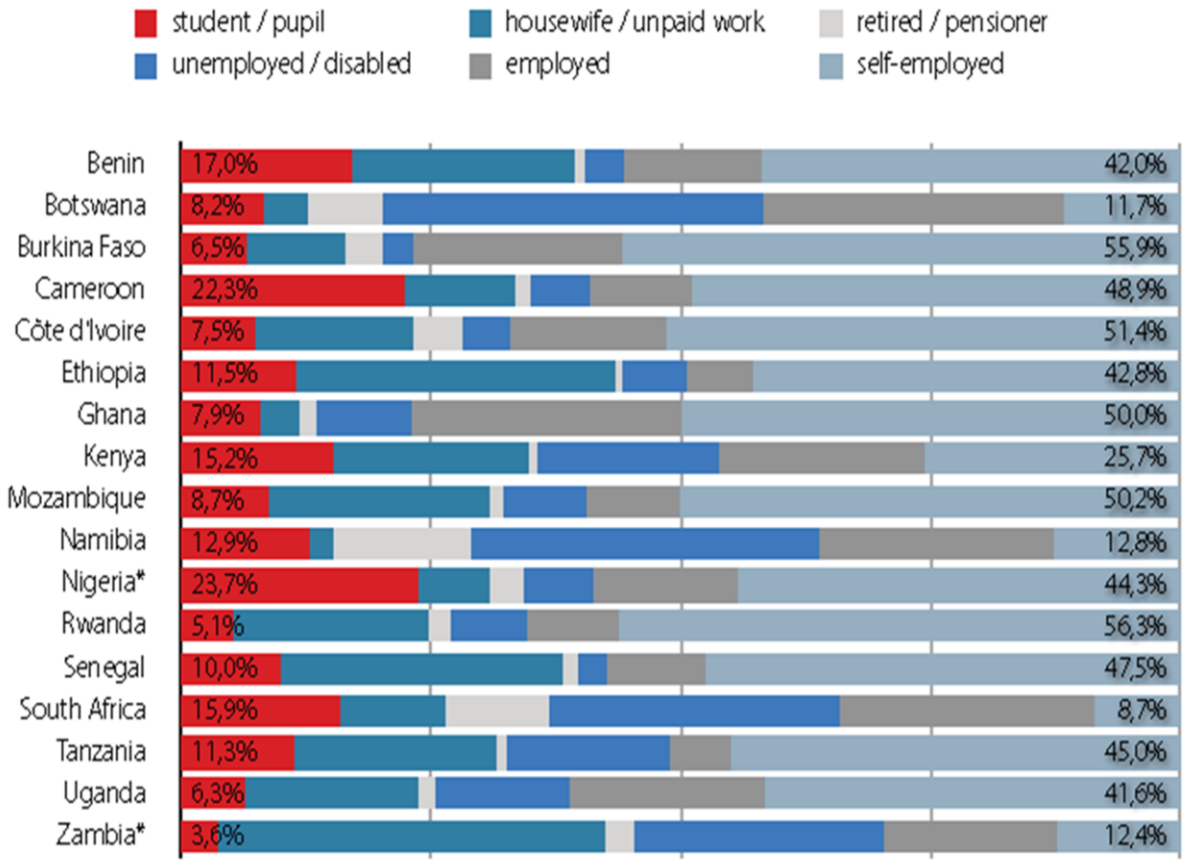


Figure 1: Main activity during past six months

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 colour

Alphabetical sorting is usually not the best choice in a chart with numerical data. In the example above, colour 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, colour should be used thoughtfully. Here, certain categories are more similar to others. For

Example 5

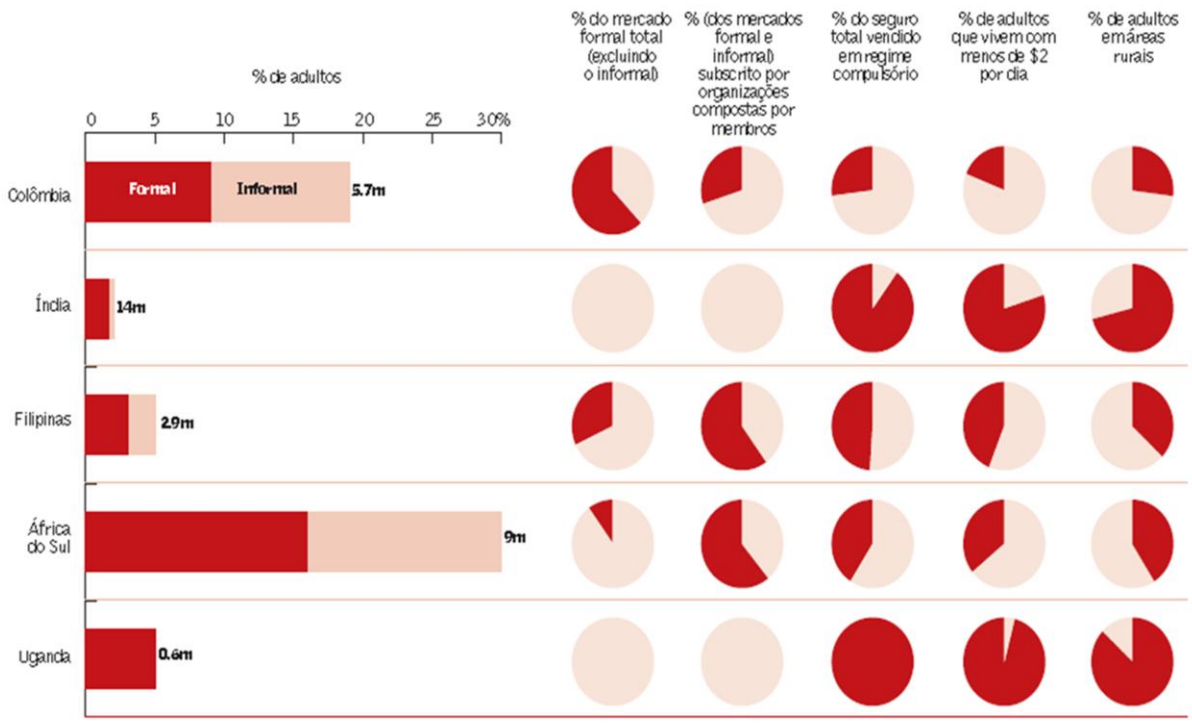


Figura 1: Características do mercado de microsseguro nos cinco países[1]

example, the unemployed, employed, and self-employed are all part of the labor force. Thoughtful colour 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.

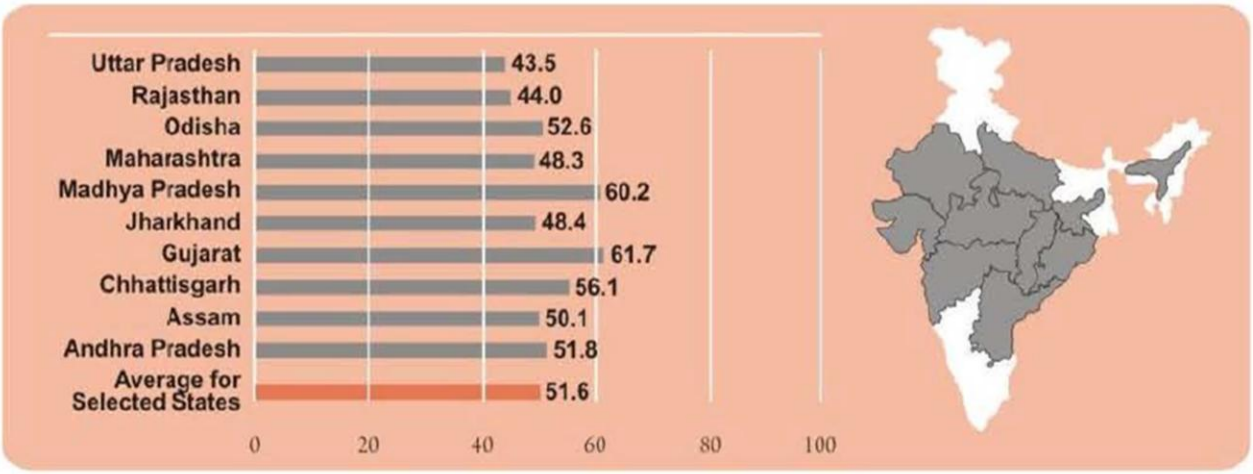
Example 5: Consistency

This example above 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 counter clockwise. 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



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**

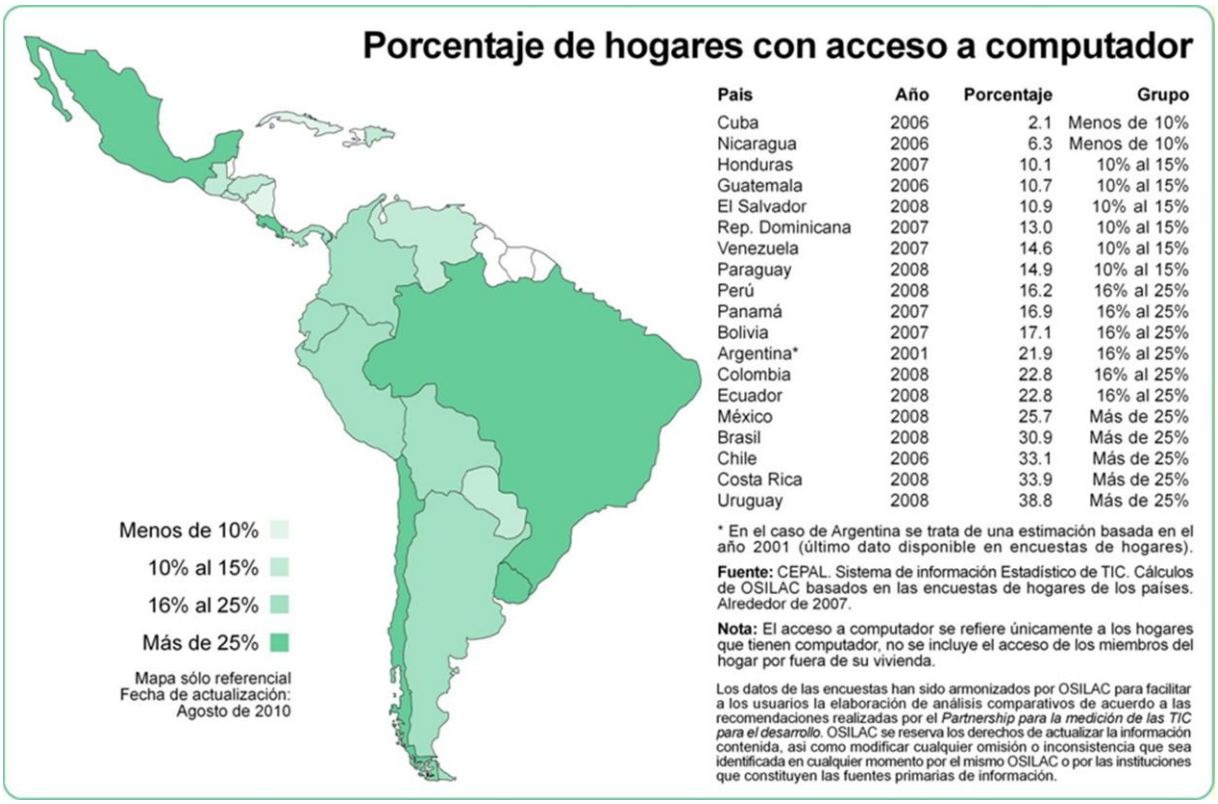
Example 6: Using the same information twice

Labels should not be separated from data. In this example above, 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 with in 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, la-

Example 7



Pais	Año	Porcentaje	Grupo
Cuba	2006	2.1	Menos de 10%
Nicaragua	2006	6.3	Menos de 10%
Honduras	2007	10.1	10% al 15%
Guatemala	2006	10.7	10% al 15%
El Salvador	2008	10.9	10% al 15%
Rep. Dominicana	2007	13.0	10% al 15%
Venezuela	2007	14.6	10% al 15%
Paraguay	2008	14.9	10% al 15%
Perú	2008	16.2	16% al 25%
Panamá	2007	16.9	16% al 25%
Bolivia	2007	17.1	16% al 25%
Argentina*	2001	21.9	16% al 25%
Colombia	2008	22.8	16% al 25%
Ecuador	2008	22.8	16% al 25%
México	2008	25.7	Más de 25%
Brasil	2008	30.9	Más de 25%
Chile	2006	33.1	Más de 25%
Costa Rica	2008	33.9	Más de 25%
Uruguay	2008	38.8	Más de 25%

bels 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.

Example 7: Repetitive Information

Good charts anticipate questions that readers are likely to have. Above, 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 colour 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 colours on the map.

Example 8: Precision

Including actual footprints of the research projects on this map (see page 10) would make the

Example 8

RTD program research projects

In early 2009, the program completed the analysis of dynamics of social and economic change in over 10,000 municipalities (or their equivalents) in 11 countries. Using the *Elbers et al.*[1] method, geographic areas characterized by different development outcomes were identified. Based on these maps, 19 territories in 11 countries were selected as sites for the program based on their dynamics of economic growth and social inclusion, in order to concentrate research and capacity building activities in them. Table 1 lists the territories selected and Figure 2 shows their geographic locations.

The project coordination teams presented proposals that were evaluated by two anonymous reviewers and then adjusted, based on the comments received. Each research project fits into the general methodological framework of the applied research component, which has been modified progressively according to the partial results obtained and the needs of the research activities.

Table 1. Selected territories

	Country	Territory	Surface in Km²	Population
1.	Bolivia	Chaco Tarijeno	13.072	225.366
2.	Brazil	Cariri Paraibano	7.075	119.430
3.	Brazil	Costa de Santa Catarina	15.000	1.500.000
4.	Brazil	Jiquirica Valley, Bahia	12.414	309.192
5.	Chile	Central Chiloe	3.412	89.000
6.	Chile	Interior dryland of the O'Higgins Region	2.153	20.093
7.	Colombia	Upper Suarez and Lake Fuquene basin	483	35.337
8.	Ecuador	Loja	10.793	404.835
9.	Ecuador	Tungurahua	3.369	441.034
10.	El Salvador	Northern riverbank of Humedal Cerron Grande	570	70.048
11.	Guatemala	Southeastern area of Jutiapa and Jalapa	570	70.000
12.	Honduras	Olancho	1.009	36.375
13.	Mexico	Mezcal region of Oaxaca	18.220	490.745
14.	Mexico	South-central Yucatan Region	628	29.900
15.	Nicaragua	Macizo de Penas Blancas, La Dalia	462	126.209
16.	Nicaragua	Dairy region	546	16.404
17.	Peru	Cuatro Lagunas, Cusco	954	35.000
18.	Peru	Sierra de Jauja, Junin	2.100	60.000
19.	Peru	Southern Valley of Cusco	3.749	88.926

[1] *Elbers, C., Lanjouw, J. O., Lanjouw, P. 2003. "Micro-level Estimation of Poverty and Inequality." Econometrica 71(1): 355-364.*

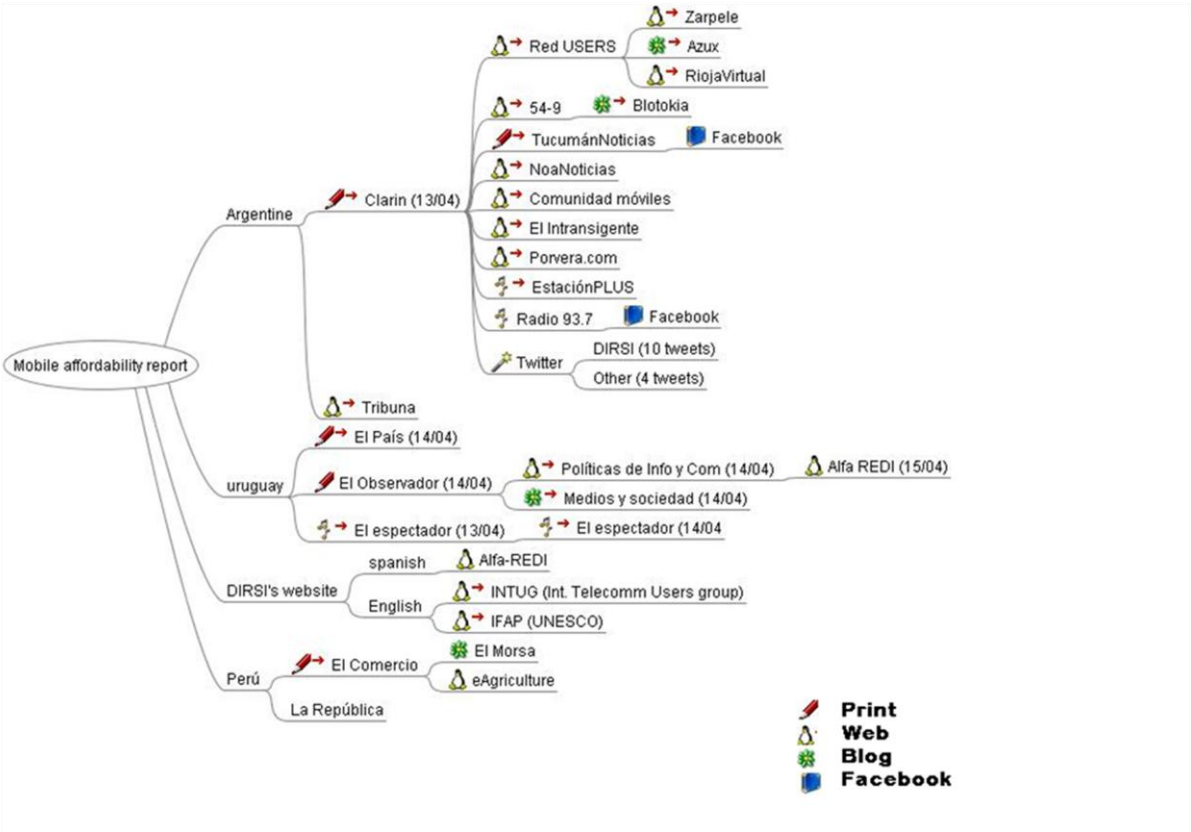
visualization more sophisticated, especially since one of the purposes of this table seems to be to show the size of the projects. This might not be meaningful for the smaller territories, but it would certainly be possible for the larger ones.

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 it easier to recognize any patterns within the data. 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 to sort by one of the columns in the table.

Example 9

Tariffs and the affordability gap in mobile telephone services y Latin America and the Caribbean



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 above, 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 "tariffs" nor an "affordability gap" seem to appear in the diagram, though it is difficult to tell.

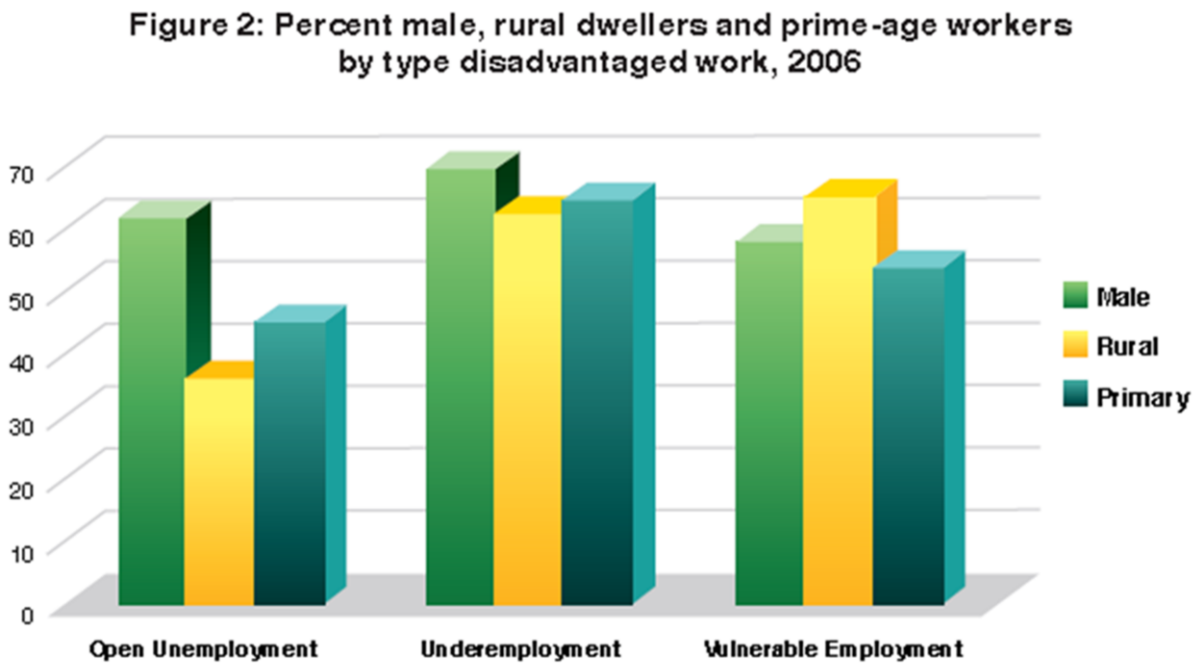
Drawing coherent icon sets is quite difficult. But when well-known icons are available – for example, in the case of Twitter and Facebook – it often makes sense to use them.

Example 10: Avoiding 3D

Almost all visualization experts recommend avoiding 3D. The reason is simple: it makes charts more difficult to read accurately. For the example on page 12, 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 10



Example 11: Avoiding scavenger hunts

Flash is a poor way to display large amounts of text. For this example, iGuide - New Mechanisms for Linking Research and Policy, if you are interested in the entire guide 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 co-worker 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

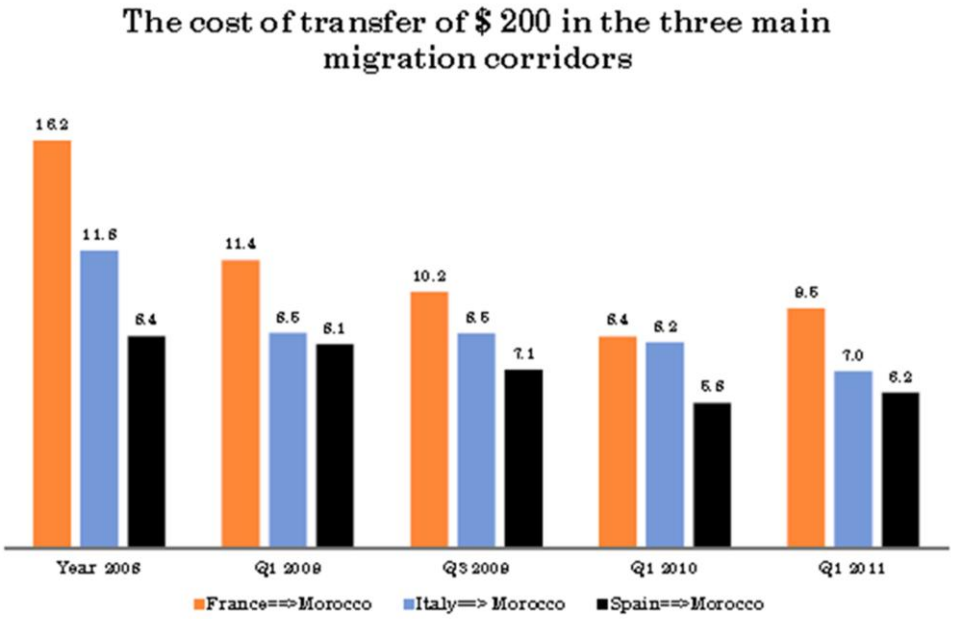
Research on how people interpret charts suggests that line charts are best at conveying movement across time, particularly when the quantity being measured does not start over at zero with each new time period. A line chart would make the patterns shown in this chart (see page 13) more immediately obvious. 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 Fair Mobile Index, 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.

Example 12



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 (see page 14), 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 (see page 15) 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:

Example 14

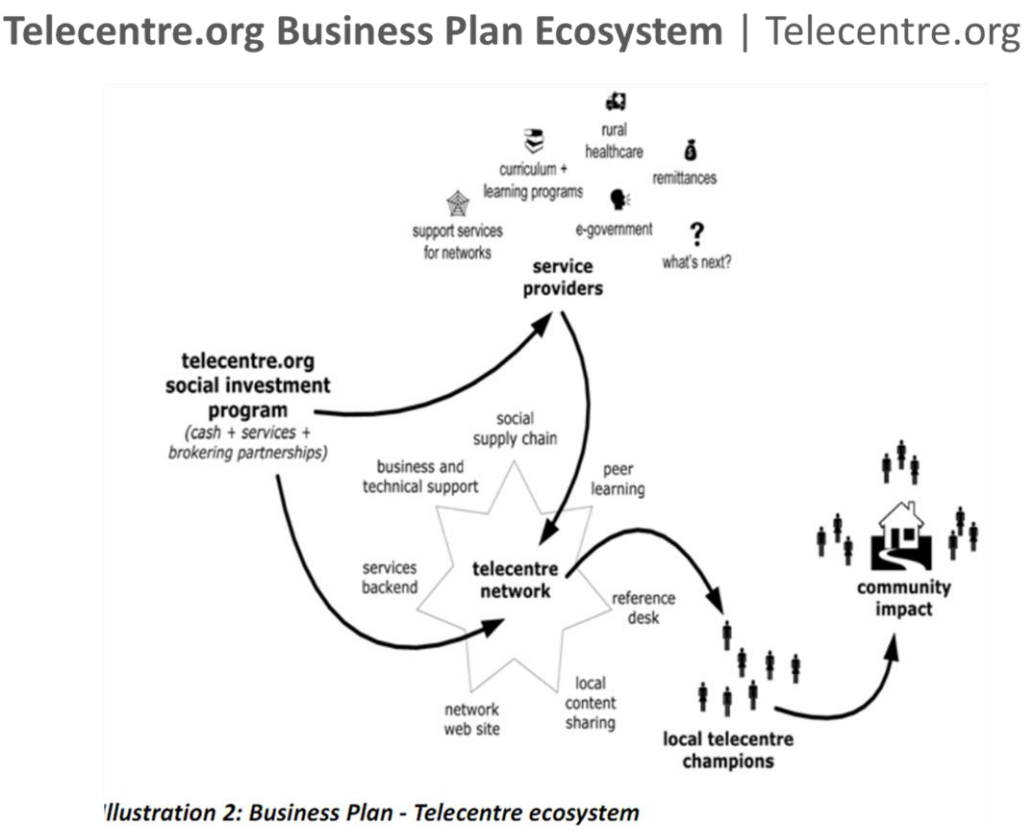


Illustration 2: Business Plan - Telecentre ecosystem

"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.

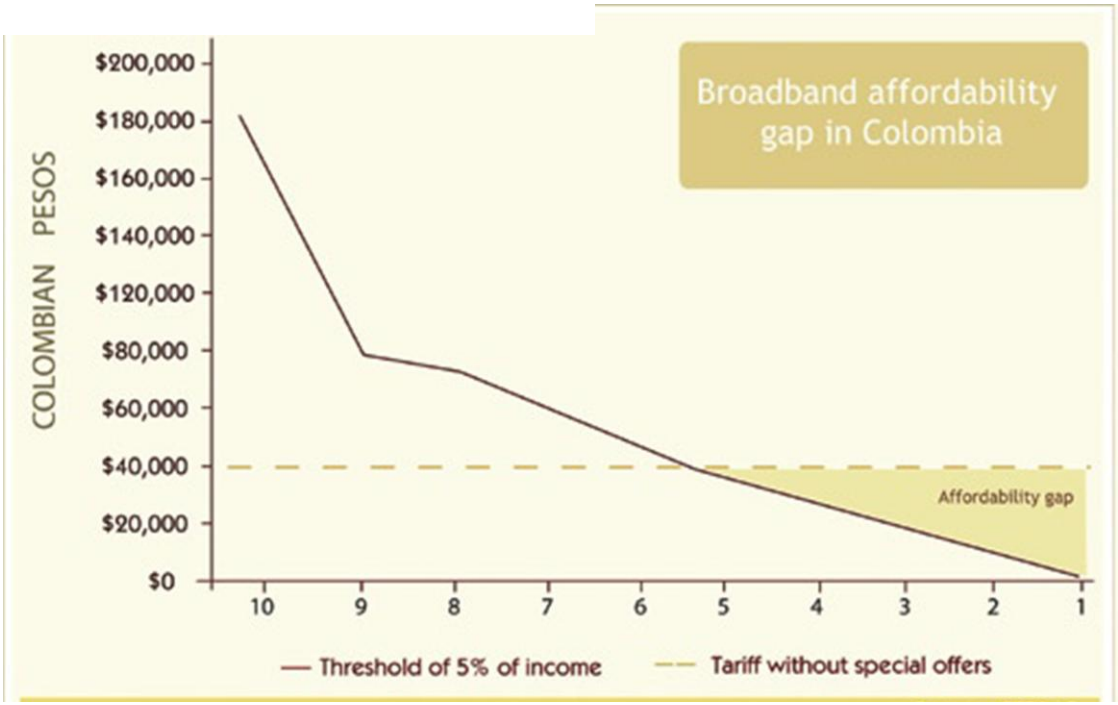
Example 16: But what does it mean?

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 these visualizations on mapping Wikipedia's languages 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 Swahili 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 15



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 (see page 16), 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.

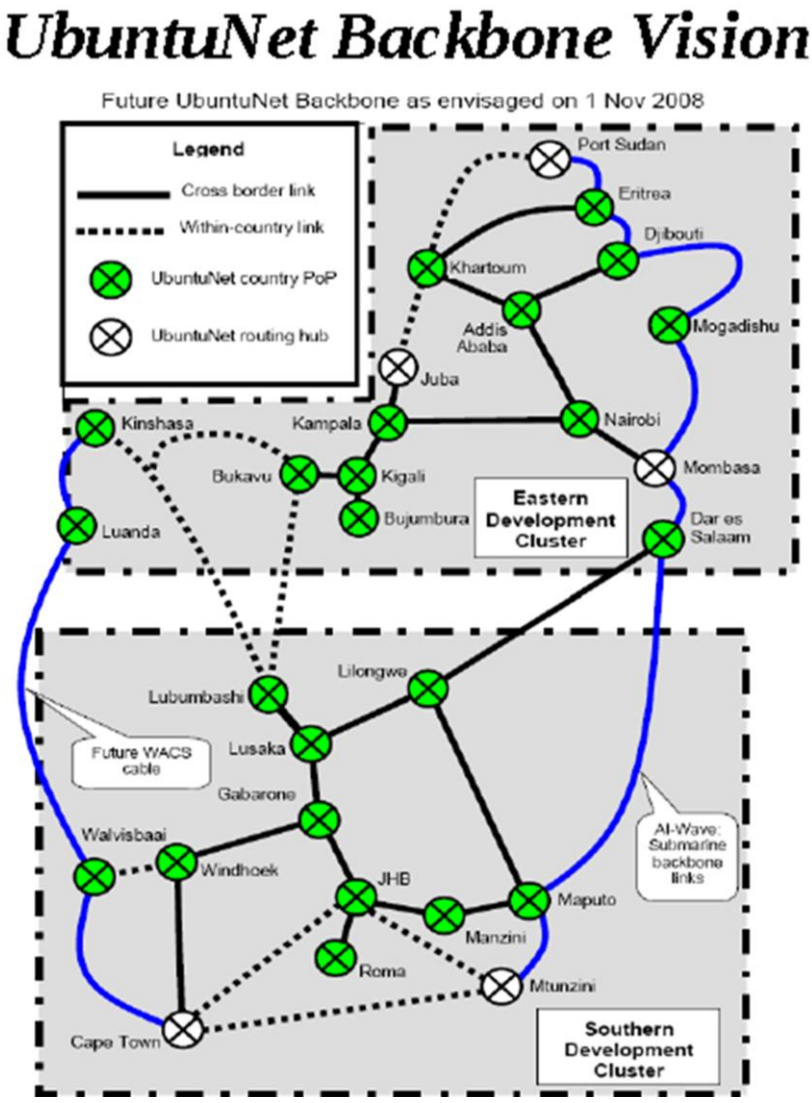
Interaction

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 from Global Trade Alert 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.

Example 17



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.

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, Trinidad and Tobago Trade Report. Keys get cut off. The legend for the bar chart (see page 17) 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 called Global Impact Study ICT Venue Inventory Database, 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, HarrassMap, 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 effectiveness of IDRC-supported 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 colour 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. 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.

About Amanda Cox

Amanda is a graphics editor at the New York Times. She joined the Times in 2005, after completing a masters in statistics from the University of Washington. Amanda has a strong background in data visualization theory and design and has a fondness for conceptual pieces. She has warmly been referred to as the "Queen of InfoVis", and has served as a keynote speaker, university lecturer, graphic consultant, and a judge for data visualization competitions. Amanda, and her colleagues have also received several awards for their dynamic work, including top honors at Malofiej, the largest international infographics competition.

This review was commissioned by IDRC's Evaluation Unit with assistance from Jacqueline Strecker and Tricia Wind.

Annex 1

	Example	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21		Yes	No	%Yes
Clear																										
Charting form appropriate?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes		16	5	**76**
Units familiar to audience?	No	Yes	Yes	Yes	Yes	No	Yes	No	No	Yes		Yes	Yes		No	Yes		Yes	Yes	Yes	Yes		13	5	**72**
Anticipate questions?	No	No		No	No	No	Yes	No	No	No	Yes	Yes	No		Yes	Yes		No	No	Yes	Yes		7	11	**39**
Titles appropriate?	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes		19	2	**90**
Focused																										
Language support specific ideas?	No	No	No	No	No	No	No	No	No	No		No	Yes		Yes	Yes	Yes	No	No	No			4	14	**22**
Design support specific ideas?	No	No	No	No	No	No	Yes	No	No	No	Yes	No	Yes	No	Yes	Yes	Yes	No	No	No	Yes		7	14	**33**
Hierarchy established?	No	No		Yes				Yes	No		No					Yes	No			Yes	Yes		5	5	**50**
Compelling																										
Audience will want to talk about?	Yes	Yes						No			Yes	Yes	Yes		Yes	Yes		Yes		Yes	Yes		10	1	**91**
Sufficient variation?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes		Yes	Yes	Yes		Yes	Yes	Yes	Yes	No	Yes	Yes		17	2	**89**
Use annotations or photography?	No	No	No	No	No	No	No	No	No	No	Yes	No	No		Yes	Yes	No	No	No	No	Yes		4	16	**20**