



Introduction

Infographics and Visualization

The remarkable mechanisms by which the senses understand the environment are all but identical with the operations described by the psychology of thinking.

—Rudolf Arnheim, from *Visual Thinking*

The partnership of presentation and exploration

If you asked me to choose one mystery that has fascinated me through the years, it would be our brain's ability to create and understand visual representations with different degrees of abstraction: graphics that encode data, concepts, connections, and geographical locations. How is it possible that the brain, a wrinkled chunk of meat and fat squeezed into a cavity too small for its size, can accomplish such challenging tasks? This book is my personal attempt to answer that question.

The human brain has dozens of regions related to visual perception: densely interconnected groups of neurons devoted to the processing and filtering of information that we collect through our eyes. Evolution designed us such that no other activity demands more mental resources than visual perception and cognition. This fact permeates everything we do, and impacts the way we express ourselves. Go ahead and explain a difficult concept to a friend. In the moment she gets what you mean, she will exclaim, with a sparkle of relief and happiness in her eyes:

“I see!”

Her expression makes complete sense, because deep inside our minds, *to see* and *to understand* are intertwined processes. We understand *because* we see. This causal relationship is also true the other way around. As I will explain later, we see because we have previous understanding of certain things. Seeing precedes understanding, and this understanding precedes a better, deeper seeing down the road.

We are a visual species. We are also a *symbolic* species, if we follow Terrence W. Deacon's famous definition.¹ Everything our senses gather is transformed, deep inside our minds, into simple, manageable representations, or symbols. Some of those symbols are verbal or textual, encoded with one of the thousands of languages and dialects humanity has devised. They can also be expressed through what Harvard psychologist Steven Pinker called *mentalese*, the inner language the mind uses to talk to itself.²

1 Terrence W. Deacon, *The Symbolic Species: The Co-Evolution of Language and the Brain* (New York: W.W. Norton & Company, 1998).

2 Steven Pinker, *The Language Instinct: How the Mind Creates Language* (New York: Harper Collins Publishers, Inc., 1994). Pinker builds on Noam Chomsky's hypothesis of an innate universal grammar.

But not all symbols are verbal. **The brain doesn't just process information that comes through the eyes. It also creates mental visual images that allow us to reason and plan actions that facilitate survival.** Imagine a bus. Picture it in your head. Now, examine it: Is it yellow, or blue, or red? Does it have a license plate? Is it a plate of the state you live in? That's it. That's a mental image. Understanding the mechanisms involved in these brain processes can help you become a better communicator, visual or otherwise.

This is the first theme of the book you have in your hands.



The second theme is the common nature of infographics and information visualization. Some professionals and academics have erected a sharp distinction between the two disciplines. According to them, infographics present information by means of statistical charts, maps, and diagrams, while information visualization offers visual tools that an audience can use to explore and analyze data sets. That is, where infographics tell stories designed by communicators, information visualization helps readers discover stories by themselves.

In the following pages, I take an unorthodox approach. **Infographics and visualization exist on a continuum.** Let me explain.

Imagine two straight, black parallel lines. On the top line, put the word “Infographics” on the left tip and “Visualization” on the right. On the line at the bottom, write “presentation” on the left, and “exploration” on the right. All graphics present data and allow a certain degree of exploration of those same data. Some graphics are almost all presentation, so they allow just a limited amount of exploration; hence we can say they are more infographics than visualization (**Figure 1**), whereas others are mostly about letting readers play with what is being shown (**Figure 2**), tilting more to the visualization side of our linear scale. But every infographic and every visualization has a *presentation* and an *exploration* component: they present, but they also facilitate the analysis of what they show, to different degrees.

An excellent example: The beauty of Stefanie Posavec’s *Literary Organism* is based on its organic appearance and careful selection of typefaces and colors, and also

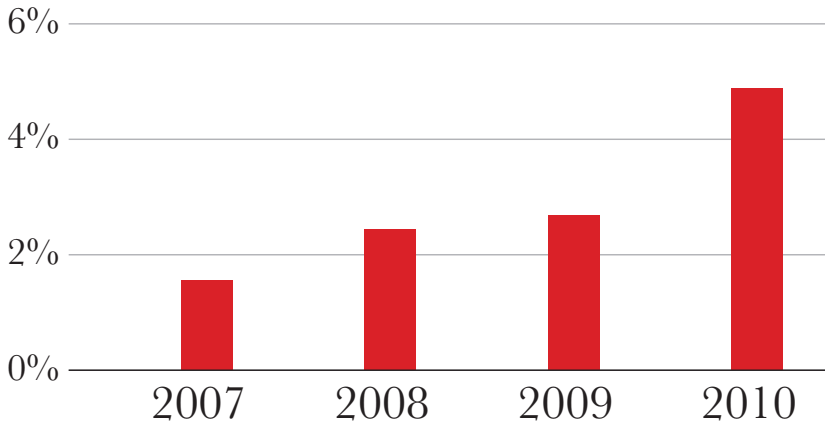


Figure 1 Unemployment rate in an imaginary country.

on the multiple readings you can extract from it. Each subdivision of this tree-like graphic represents a deeper, more granular level of organization in the book: chapters, paragraphs, sentences, and words. Colors correspond to the most common themes of the novel: travel, music, parties, sex, and so on.

After spending some time fathoming it, patterns emerge and convey a different message to each viewer. While the uninitiated in Kerouac's oeuvre will identify a big picture of the main topics the book discusses, the specialist or literary critic will be able to use this work of art as a tool to test hypotheses and intuitions. Is sex a prevalent theme in *On the Road*, for instance? What about the chapters that combine paragraphs about sex with paragraphs about work and survival?



A chart of mine in **Figure 3** also illustrates the complementarity between presenting and exploring. Each point of the color line represents a year between 1981 and 2010 in the history of Brazil. The position of a point on the horizontal axis is equivalent to the Gross Domestic Product, measured in billions of dollars. The position on the vertical axis is equivalent to the inequality level measured with the GINI index, developed by the UN. The farther to the right a point is, the bigger the GDP (adjusted for inflation and for purchase power parity); the higher the point is on the vertical scale, the higher the inequality in that particular year.

In other words, the graphic represents the covariation of economic development and social justice. The headline summarizes its central message: When the GDP

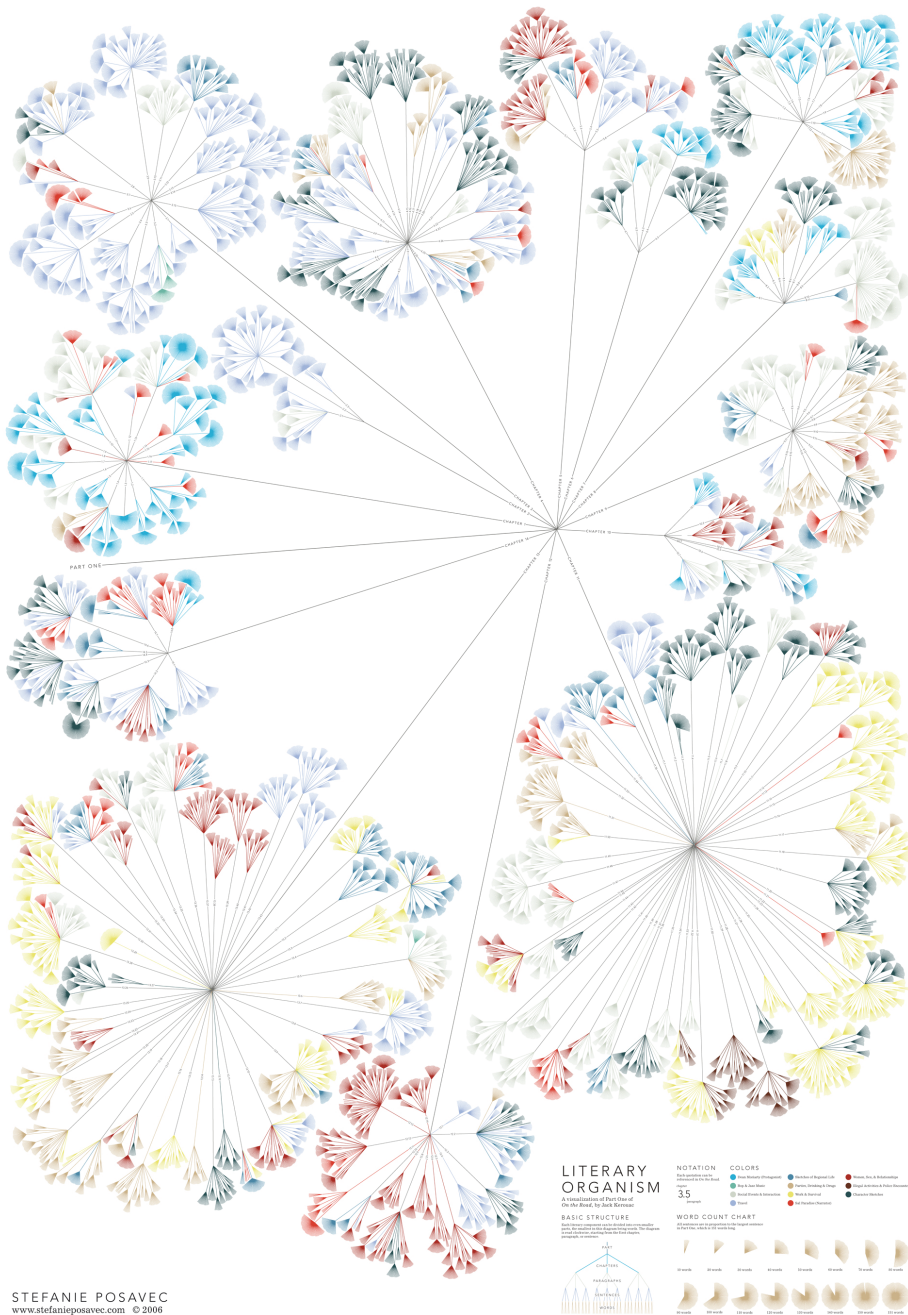


Figure 2 Stefanie Posavec *Literary Organism: a Visualization of Part I of "On the Road,"* by Jack Kerouac (www.itsbeenreal.co.uk). Reproduced with permission.

When the Brazilian Economy Improves, Inequality Doesn't Drop

The graphic below shows the correlation between Brazilian GDP (horizontal axis) and inequality (vertical axis) between 1981 and 2010. The position of the points, each representing a year, depends on how high GDP and inequality were. You can notice, for instance, that the economy grew between 1986 and 1989 because the line tends to move to the right, but inequality also grew, as the point representing 1989 is much higher than the ones before. You can also see that, during Lula da Silva's government, the economy expanded almost as much as during the terms of the other presidents who preceded him combined.

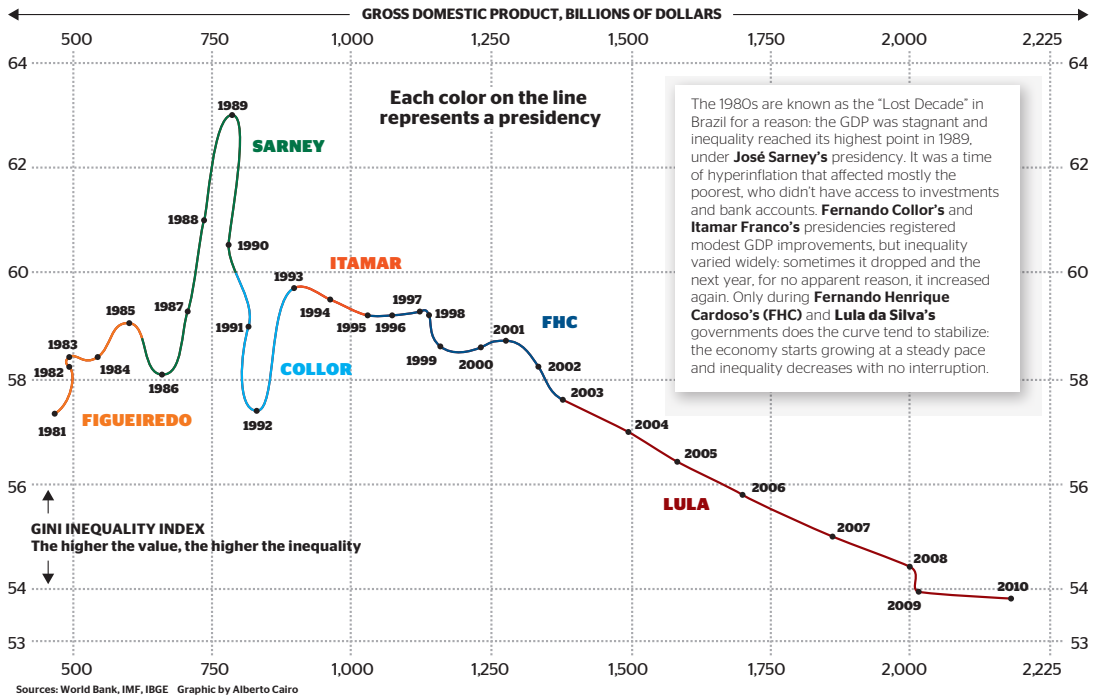


Figure 3 *Época* magazine. The co-variation of inequality and economic growth (Translated from Portuguese.) Reproduced with permission.

grows, inequality does not necessarily drop. It's one of the disgraces of recent Brazilian history that improvements in the economy don't always lead to a better living standard for everybody. In some years, particularly on the first half of the line, the opposite is true.

When I finished designing this project, I showed it to some middle-aged colleagues at *Época*, a weekly news magazine I worked for between 2010 and 2011. Their response was unanimous and encouraging. Although they were writers rather than designers or artists, all of them understood the graphic with a minimum of

effort. It confirmed for them facts that every Brazilian older than 40 remembers: the constant and stable economic growth the country went through during Lula's administration (2003–2011), when the government promoted several income distribution programs; the instability of José Sarney's and Fernando Collor's years, when the GDP barely improved but inequality varied erratically; the stabilization forced by Fernando Henrique Cardoso (FHC), who took Brazil out of the black hole of inflation; and so on.

Some of them even told me that the multicolored line was history-making, because it revealed the correlation between the two variables, which had never been shown before. The line looks wildly erratic between 1981 and 1992 (the chaos years); it smooths between 1993 and 2002 (the stabilization period); and it becomes perfectly straight after 2003, revealing an almost perfect relationship between better economic output and more equality in Brazil. To see self-proclaimed non-visual people in the process of unraveling such an uncommon graphic form, and getting satisfactory messages from it, was eye-opening for me.



The surprise reaction of my “text” colleagues as they read the graphic, rather than merely looked at it, has deep roots. This is the third theme of *The Functional Art: Graphics, charts, and maps aren't just tools to be seen, but to be read and scrutinized*. The first goal of an infographic is not to be beautiful just for the sake of eye appeal, but, above all, to be understandable first, and beautiful after that; or to be beautiful thanks to its exquisite functionality.

If you are among those journalists, designers, and artists who think that infographics and visualization consist of a bunch of data shaped into a spectacular and innovative form, keep reading. I hope I will be able to make you forget that simplistic idea. As Ben Shneiderman wrote once, “The purpose of visualization is insight, not pictures.”³ Images are the vocabulary of a language. They are means, not ends. You will never hear a writing journalist say that her goal is to strive for a good literary style by using elegant sentences and sophisticated structures. Her style is just a tool to facilitate comprehension and to wake up emotions in readers' minds so they'll absorb difficult ideas with ease. Aesthetics do matter, but aesthetics without a solid backbone made of good content is just artifice.

³ Stuart Card, Jock Mackinlay and Ben Shneiderman, *Readings in Information Visualization: Using Vision to Think* (London: Academic Press, 1999).



In this book, you will see that I write quite a bit about visual journalism. That's because I am a journalist, and I am convinced that many of the challenges news media face in using graphics are common to other professions that also use them on a regular basis, such as marketing, advertising, business intelligence, data analysis, and so on.

In newspapers and magazines, infographics have traditionally been created within art departments. In all of those I'm familiar with, the infographics director is subordinate to the art director, who is usually a graphic designer. This is not a mistake *per se*, but it can lead to damaging misunderstandings. In Brazil, the country where I lived while writing part of this book, journalists and designers call graphics "art." They would say, with that charming musicality of South American Portuguese, "Vamos fazer uma arte!" ("Let's make a piece of art!") Thinking of graphics as art leads many to put bells and whistles over substance and to confound infographics with mere illustrations.

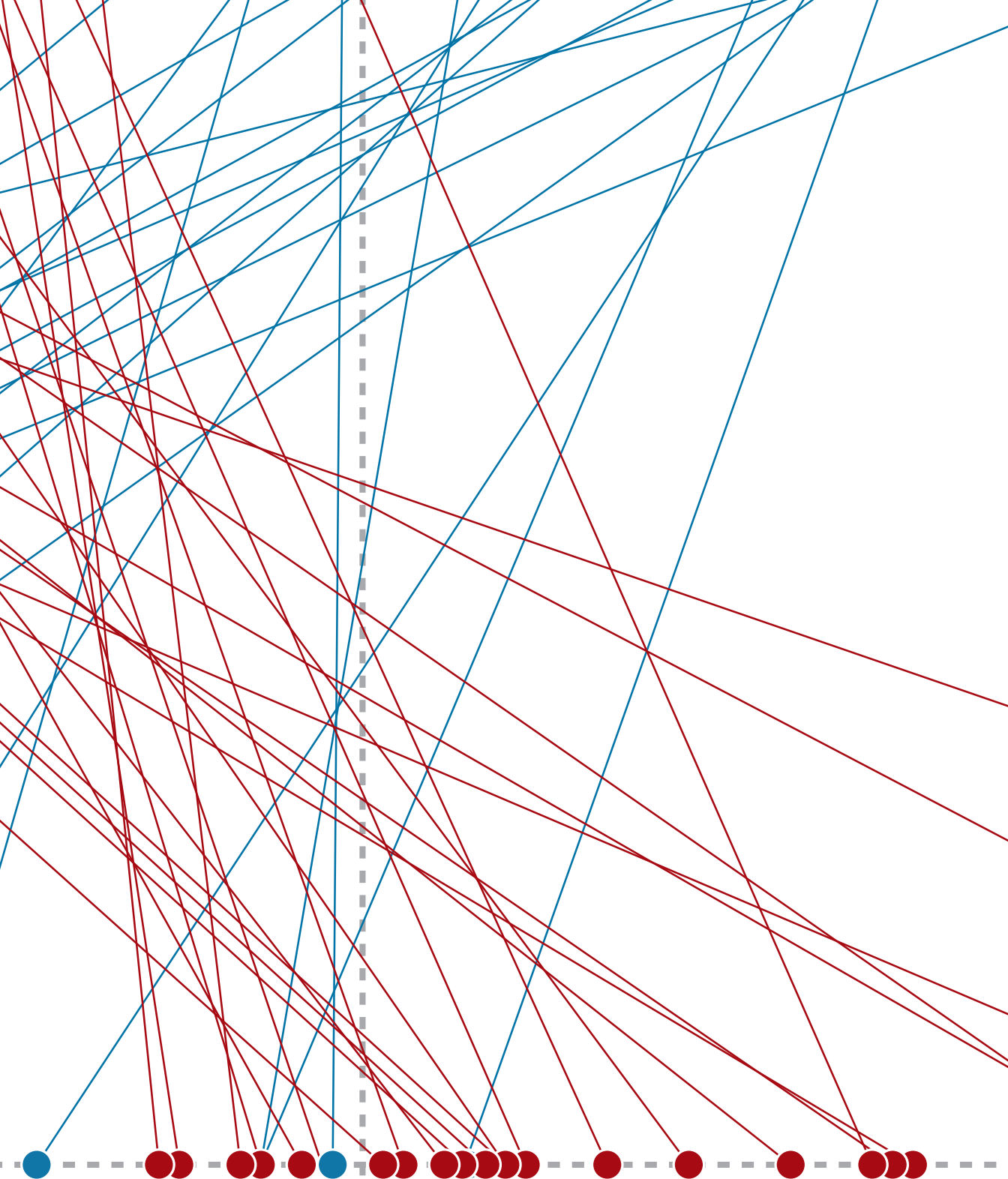
This error is at least in part the result of a centuries-long tradition in which visual communication has not been as intellectually elevated as writing. For too many traditional journalists, infographics are mere ornaments to make the page look lighter and more attractive for audiences who grow more impatient with long-form stories every day. Infographics are treated not as devices that expand the scope of our perception and cognition, but as decoration. As Rudolf Arnheim wrote, this tradition goes back to ancient Western philosophy, whose Greek thinkers such as Parmenides and Plato mistrusted the senses deeply.⁴ Unfortunately today, 40 years after Arnheim's masterful *Visual Thinking* was published, the philosophy is still in very good health.

The fourth theme of the book, therefore, is **the relationship between visualization and art, which is similar to the linkage of journalism and literature**. A journalist can borrow tools and techniques from literature, and be inspired by great fiction writing, but she will never allow her stories to *become* literature. That notion applies to visualization, which is, above all, a functional art.

Let's get started.

Miami, Florida. June 2012

⁴ Rudolf Arnheim, *Visual Thinking* (Berkeley: University of California Press, 1969).





PART I

foundations



Why Visualize: From Information to Wisdom

My expertise has always been my ignorance, my admission and acceptance of not knowing. . . . When you can admit that you don't know, you are more likely to ask the questions that will enable you to learn.

—Richard Saul Wurman, from *Information Anxiety 2*

Recently someone asked me what personality trait best characterizes those interested in a career in visualization and infographics. My answer: “An insatiable, childish curiosity.”

Curiosity, combined with a tendency to try to explain everything using reason, led me to a career in journalism and, later, to specialize in information graphics. It is not possible to be a good communicator if you have not developed a keen interest in almost everything as well as an urge to learn as much as you can about the strangest, most varied, unrelated topics. **The life of a visual communicator should be one of systematic and exciting intellectual chaos.** In my case, it consists of regular shifts between journalism, cognitive psychology, international

politics, and history. In your case, it might be sports, music, architecture, or just about anything else.

Let me give you an example of how far healthy curiosity can take you.

Rational Optimism

My original plan for this chapter was to open it with a few formal definitions for information visualization, information design, and infographics. But something changed my mind. While randomly navigating *The New York Times* website, I stumbled on a review of a book called *The Rational Optimist: How Prosperity Evolves*, by British science essayist Matt Ridley (2010).

The review¹ was mostly critical, but the hypotheses Ridley proposed sounded intriguing. The book honored its title, making a case for optimism about our future as a species. I had read other Ridley books in the past and had loved his *Genome: The Autobiography of a Species in 23 Chapters* (2006), so I was positively biased.

My curiosity ignited, I grabbed my Kindle and purchased the book. One minute later, it had finished downloading. For the rest of the day, I traded book writing for reading. (Note to my editor: I did it to make this chapter better, I promise!) The book is so well written that it was difficult to put down before the end.

One chapter about the fertility rate, or the average number of children born to women in each country, caught my eye. You may have heard or read the stories of Malthusian doomsayers who claim that rising fertility in poor regions is the reason the Earth has to support 7 billion people, with a forecast of 9 billion two decades from now, and even more in the far future.

Other doomsayers focus on the aging populations of developed countries where fertility rates are below 2.1 children per woman, a number that is known as the “replacement rate.” If the replacement rate in a country is significantly below 2.1, the population will shrink over time. If it’s much higher than 2.1, you’ll have a much younger population down the road, which can cause problems. Younger populations, for example, show greater rates of violence and crime.

Ridley contradicts both kinds of apocalyptic thinking by discussing two interesting trends. On average, fertility in rich countries is very low, but in the past few years it has trended slightly upward. On the other hand, poor countries show a *decrease* in average fertility. Contrary to conventional wisdom, in many countries

1 William Easterly, “A High-Five for the Invisible Hand,” *The New York Times*, June 11, 2010.

that verge on becoming first-world economies, such as Brazil, the drop is dramatic: the fertility rate has trended from more than six children per woman in 1950 to less than two in 2010.

Ridley suggests that, due to these two complementary trends, fertility rates everywhere will converge around 2.1 in a few decades, and the world population will stabilize at 9 billion people. It's counterintuitive, isn't it?

Ridley's case is compelling and supported by prospective data from reliable sources, such as the United Nations (UN) and The World Bank.² But something made me uncomfortable as I read his arguments. It took me a while to figure out what it was. Ridley writes about curves and lines and trends, but the chapter on fertility and population includes just one graphic, similar to the one in **Figure 1.1**.

Percentage increase in world population

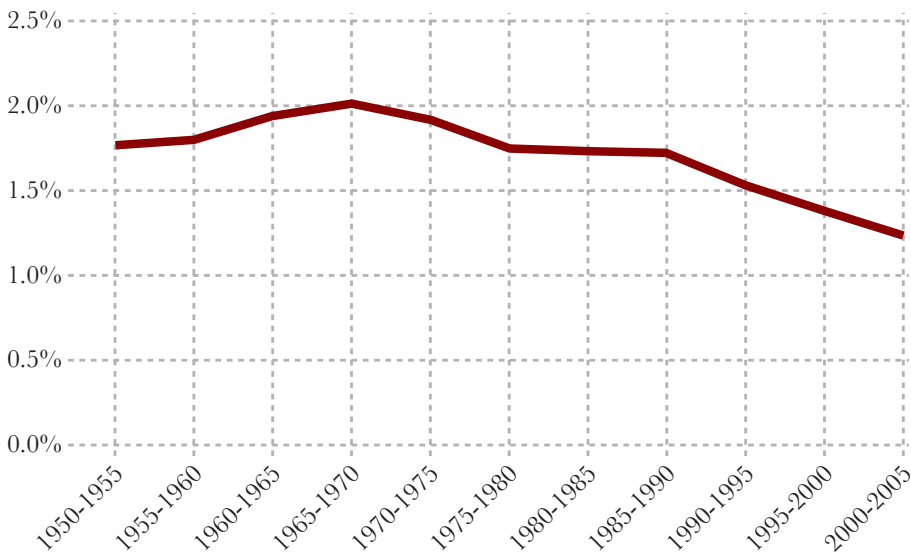


Figure 1.1 How much world population increases compared with the previous year.

The graphic is simple and clear, but also insufficient to support the claims Ridley makes. All it shows is that when you plot population change as a time-series chart, the trend is negative. The closer we get to the present, the lower the worldwide population growth. **The fact that the graphic is an aggregate of the data of all countries in the world impedes our ability to see the multiple patterns**

² Matt Ridley, *The Rational Optimist: How Prosperity Evolves* (New York: Harper Collins).

Ridley discusses. Where are those rich countries whose fertility rate is slowly recovering? Where is the evidence for the assertion that developing countries such as Brazil, China, and India are stabilizing their populations?

I told you before that I am curious. I didn't just take a day off to read *The Rational Optimist*. I also looked for the data Ridley used for the chapter on population. With that data, would I be able to prove his hypotheses?

Low-Tech Visualization to the Rescue

The first thing I did was go to the United Nations website. If you search for “UN data,” you will quickly find <http://data.un.org/>, which is a kind of paradise if you like to dig into huge databases on population, education, economics, and social development. Within the UN site, I searched for the fertility rate data. The UN's figures begin in 1950; the projections extend as far as 2100.

I decided to exclude the years after 2010, playing with actual data (at that time) rather than forecasts. Using the filters the site offers, I asked for a table that included the more than 150 countries on which the UN has complete research. **Figure 1.2** shows a screenshot of what I obtained.

The screenshot shows the UNdata website interface. The main content is a table titled "Total fertility rate (children per woman)". The table has columns for "Country or Area", "Variant", and years from 1950-1955 to 2005-2010. The table is filtered to show 229 records. The table is sorted by "Country or Area" in ascending order. The table shows data for various countries, including Afghanistan, Africa, Albania, Algeria, Angola, Argentina, Armenia, Aruba, Asia, Australia, Australia/New Zealand, Austria, Azerbaijan, Bahamas, Bahrain, Bangladesh, Barbados, and Belarus. The data points are numerical values representing the fertility rate for each country in each year.

Country or Area	Variant	1950-1955	1955-1960	1960-1965	1965-1970	1970-1975	1975-1980	1980-1985	1985-1990	1990-1995	1995-2000	2000-2005	2005-2010
Afghanistan	Estimate variant	8	8	8	8	8	8	8	8	8	8	7	7
Africa	Estimate variant	7	7	7	7	7	7	6	6	6	5	5	5
Albania	Estimate variant	6	6	6	5	5	4	3	3	3	2	2	2
Algeria	Estimate variant	7	7	7	7	7	7	6	5	4	3	3	2
Angola	Estimate variant	7	7	7	7	7	7	7	7	7	7	7	6
Argentina	Estimate variant	3	3	3	3	3	3	3	3	3	3	2	2
Armenia	Estimate variant	4	4	4	3	3	3	2	3	2	2	2	2
Aruba	Estimate variant	6	5	4	3	3	2	2	2	2	2	2	2
Asia	Estimate variant	6	5	6	5	5	4	4	3	3	3	2	2
Australia	Estimate variant	3	3	3	3	3	2	2	2	2	2	2	2
Australia/New Zealand	Estimate variant	3	3	3	3	3	2	2	2	2	2	2	2
Austria	Estimate variant	2	2	3	3	2	2	2	1	1	1	1	1
Azerbaijan	Estimate variant	5	5	6	5	4	4	3	3	3	2	2	2
Bahamas	Estimate variant	4	4	5	4	3	3	3	3	3	2	2	2
Bahrain	Estimate variant	7	7	7	7	6	5	5	4	3	3	3	2
Bangladesh	Estimate variant	7	7	7	7	7	7	6	5	4	3	3	2
Barbados	Estimate variant	5	5	4	3	3	2	2	2	2	2	2	2
Belarus	Estimate variant	3	3	3	2	2	2	2	2	2	1	1	1

Figure 1.2 UN data table.

I downloaded the table and decided to try a low-tech visualization exercise to show my students—mostly journalists with no technology background—that learning to create information graphics is not all that difficult.

Using OpenOffice (an open source software suite that includes a spreadsheet program), I reorganized the data and cleaned up the table a little. Some cells were missing, so the process involved a bit of manual tweaking—no big deal. **Figure 1.3** shows an excerpt of the result.

	A	B	C	D	E	F
1	Country of	1950-1955	1955-1960	1960-1965	1965-1970	1970-1975
126	Senegal	5.97	6.46	6.75	7.25	7.50
127	Serbia	3.22	2.75	2.57	2.43	2.36
128	Sierra Leo	5.52	5.60	5.70	5.77	5.84
129	Singapore	6.40	5.99	4.93	3.46	2.62
130	Slovakia	3.52	3.27	2.89	2.50	2.51
131	Slovenia	2.80	2.39	2.32	2.32	2.19
132	Somalia	7.25	7.25	7.25	7.25	7.10
133	Spain	2.57	2.75	2.89	2.92	2.86
134	Sri Lanka	5.80	5.80	5.20	4.70	4.00
135	Sudan	6.65	6.65	6.60	6.60	6.60
136	Suriname	6.56	6.56	6.56	5.95	5.29
137	Swaziland	6.70	6.70	6.75	6.85	6.87
138	Sweden	2.21	2.23	2.32	2.16	1.89
139	Switzerland	2.28	2.34	2.51	2.27	1.82
140	Syrian Arab	7.30	7.45	7.60	7.60	7.52
141	Tajikistan	6.00	6.20	6.30	6.72	6.83
142	Thailand	6.35	6.35	6.34	5.99	5.05
143	Timor-Les	6.44	6.35	6.37	6.16	6.15

Figure 1.3 My UN data table after tweaking.

Still with me? Now the fun begins. We have the table in the computer. Is it possible to make sense of it? Hardly. Extracting meaning from a table is tough. Can you see any interesting trends just by reading the figures? If you can, congratulations. You have an extraordinary memory. Most of us mortals have brains that didn't evolve to deal with large amounts of data. Let me prove it to you: Look at Figure 1.3 again and tell me in what years between 1950 and 1975 did the difference between the fertility rates of Spain and Sweden grow, and in what years did it drop? This apparently simple task forces you to do something extremely difficult: look up a number, memorize it, read another one, memorize it and compare it with the previous one, and so forth until you get to the end of the series. I wouldn't bother. But what if I designed a simple chart with the data in the spreadsheet? The result (**Figure 1.4**) is a visual tool that helps answer my question. The message in that graphic is clear: Spain started 1950 with an average number of children per woman higher than Sweden's. But then fertility in Spain fell drastically after 1970 and only recovered partially in the last five years of the series. On the other hand, Sweden's fertility rate has remained pretty stable over the last 60 years, although it is well below the replacement level of 2.1 children per woman.

By giving numbers a proper shape, by visually encoding them, the graphic has saved you time and energy that you would otherwise waste if you had to use a table that was not designed to aid your mind. **The first and main goal of any**

A Comparison of Spain's and Sweden's Fertility Rates

Average number of children per woman over her lifetime

Source: UN

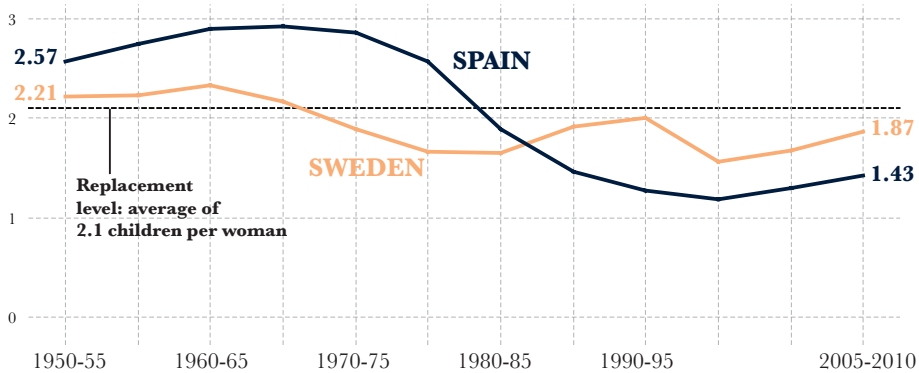


Figure 1.4 So much for Spanish pride.

graphic and visualization is to be a tool for your eyes and brain to perceive what lies beyond their natural reach.³

However, presenting data for two countries is far simpler than presenting data for one hundred of them, which is what we may need to do in order to put some of Matt Ridley's ideas to the test. Once we represent the lines for all countries in our data set, we get something similar to **Figure 1.5**. This colorful spaghetti dish may look interesting, but it's totally useless for our purposes. This is what you get when you let a software program do the hard work for you.

Remember, what we want to reveal is the projected confluence of the lines of rich countries (trending slightly up in recent years) and those of poor countries (trending down) around the 2.1 children per woman line. If you look at Figure 1.5 long enough, you may be able to tell the lines apart, but it's more likely you will just give up.

The way to solve this problem is to add some visual hierarchy to the mix. Obviously it makes no sense for all lines to be equally visible. **In information graphics, what you show can be as important as what you hide.** I put the chart generated in OpenOffice into Adobe Illustrator, where I highlighted a few rich countries and a few developing and poor countries.

I made other countries' lines light gray, so they remain on the scene but don't obscure the message. Why not get rid of them? Because they provide context to

³ This idea has inspired some of the best books out there, including those of Edward Tufte, William Cleveland, Stephen Few, and Stephen Kosslyn, among others. See the Bibliography for references.

Fertility Rate

Average number of children per woman over her lifetime
Showing all countries for which complete data is available

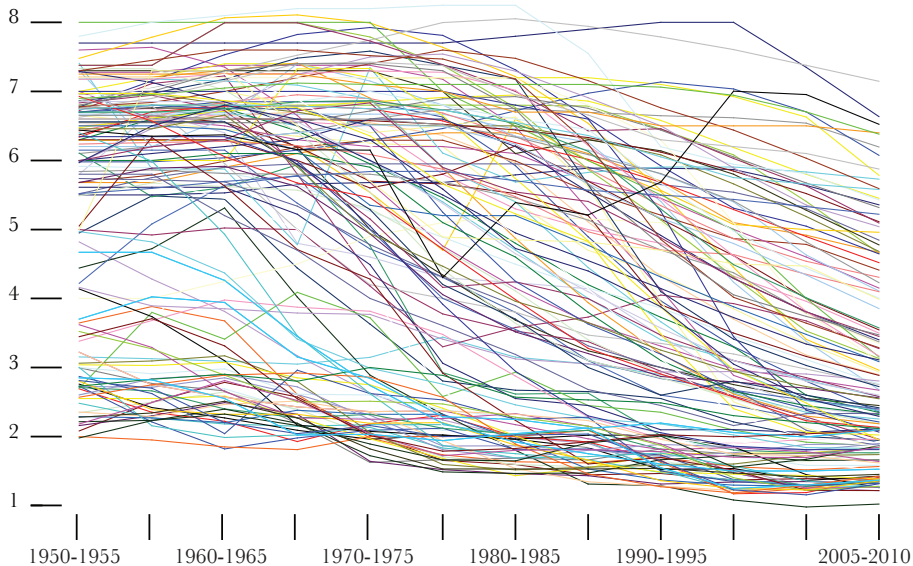


Figure 1.5 Too many lines obscure the message.

the cases that I highlight. While changing all the background lines to one color makes it impossible to see them as independent entities, collectively they show an overall downward trend in the data—you can see that many lines begin between the 6.0 and 8.0 children per woman point in 1950, but just a handful of them remain at that height in the vertical axis when they reach 2010. The final sketch in **Figure 1.6** looks much more user-friendly than the previous one.

Excited by what I was revealing, I explored other assertions made in *The Rational Optimist*. Ridley argues that a sudden drop in a country's fertility rate is usually precipitated by several factors: an increase in average per capita income, women getting better access to education, and the shrinking of infant mortality figures. The facts that more children survive their first years of life and that women are spending more time in school are positively correlated to better family planning.

On the economic side, Ridley explains that in rich countries, leisure options are everywhere, and they are cheap and accessible; the distractions of the modern world free us, albeit partially, of our primary impulse to reproduce with no control. We can explain this phenomenon in bogus academic jargon: the average number

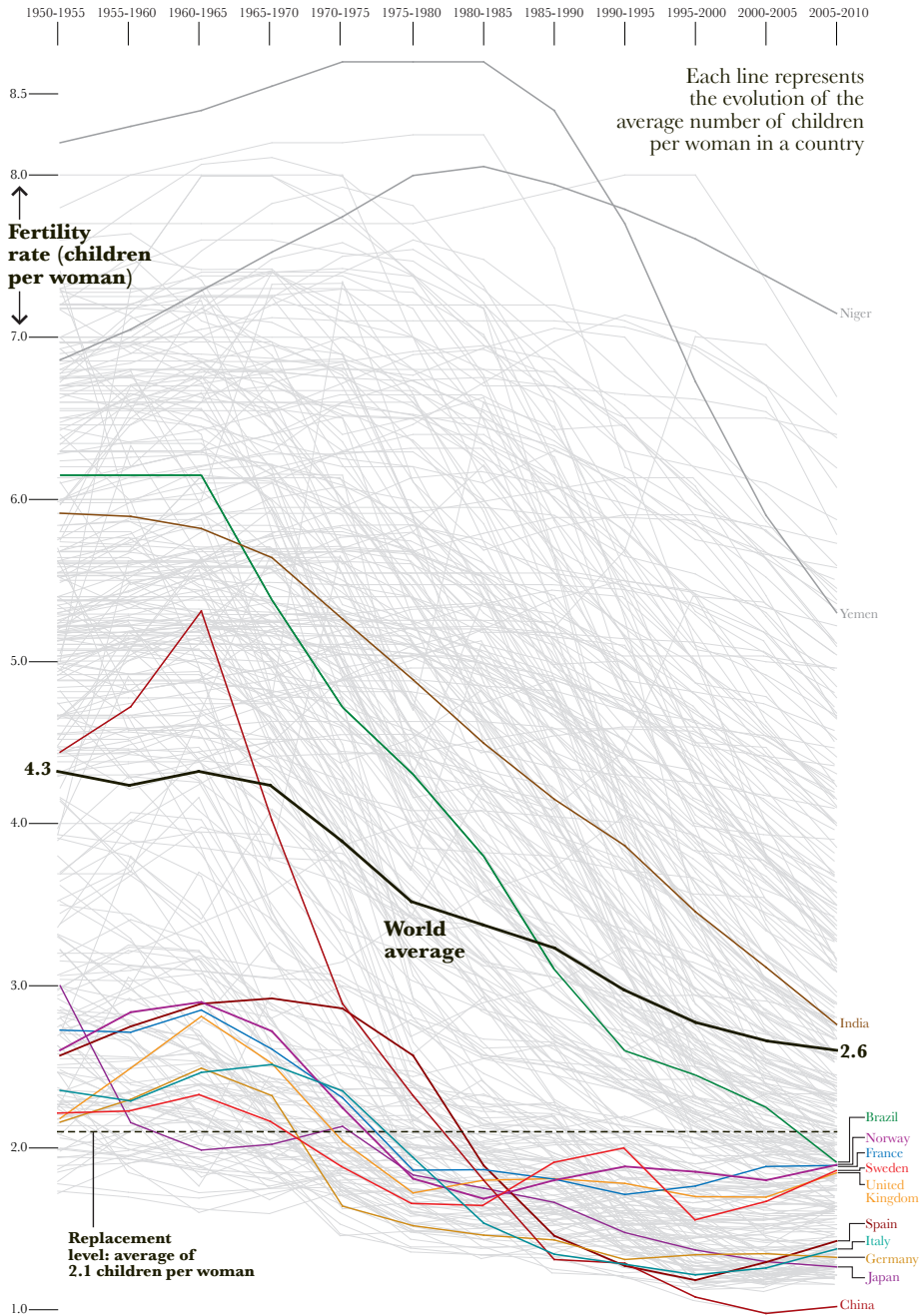


Figure 1.6 Highlighting the relevant, keeping the secondary in the background.

of children per couple is inversely proportional to the average amount of hours each member of that couple spends in front of any kind of screen. No kidding.

To prove the correlation between fertility, income, and women's schooling, I designed two small scatter-plots in the same spreadsheet software I used before. In **Figure 1.7**, each little circle represents a country. The position of each country on the horizontal axis is proportional to its fertility rate. The position on the vertical axis equals average income per person (first chart) and the percentage of students in middle school who are female (second chart).

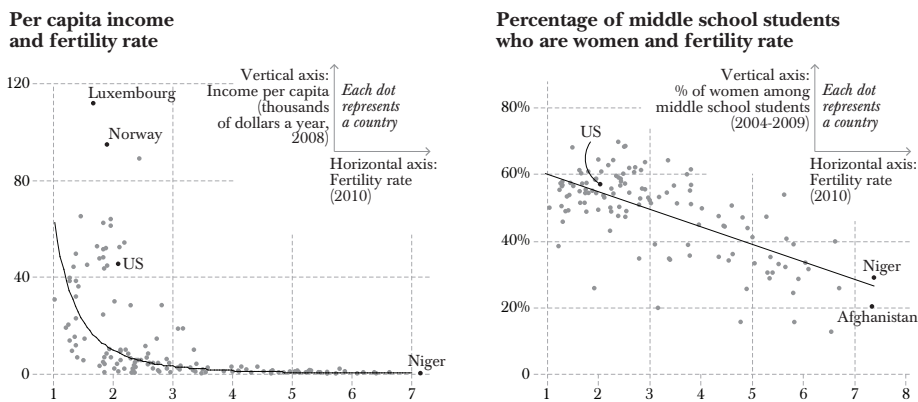


Figure 1.7 The more educated and rich you are, the fewer children you'll have.

The black line running between the dots is called a *trend line* or *regression line*: the closer the dots are to this line, the stronger the correlation between the two variables represented. You can see that the dots are pretty close to the line, so the variables are related. On average, the richer people are, the fewer children they tend to have; and the fewer girls who attend middle school, the more children on average they have in that particular country.

Here's the lesson I learned from this exercise: In just three or four hours of work, I completed a personal project that allowed me to see the evidence supporting Matt Ridley's discussion on the evolution of fertility. His hypotheses seem to have some basis after all. **But if you don't present your data to readers so they can see it, read it, explore it, and analyze it, why would they trust you?** This is a question many journalists, particularly those who write opinion columns, should ask themselves more often.

Drowning in Data? Only If You Don't Know How to Swim

The example I just gave you was not chosen randomly. It helps us delve into topics that are essential to understanding the present state of graphics as a form of communication. Isn't it amazing that we can read graphics at all, and use them to discover realities otherwise invisible to the bare eye? In the past two or three decades, psychology has unveiled many of the brain mechanisms involved in this kind of understanding. Another thought: Did you notice that the data I manipulated is available on the Internet *for free*? And isn't it wonderful that the tools I used to prove Matt Ridley's assertions are also broadly available to anyone and can be learned in a matter of hours?

Those three trends have converged to put visualization in the mainstream. The design of graphics is not just more democratic than it was a decade ago, when data was sparse and software expensive and difficult to use. We also have the potential to use graphics more intelligently because the principles informing their design have never been clearer.

Embracing graphics and visualization is no trivial endeavor. Citing research by International Data Corp (IDC), *The Economist* reported that the information generated in 2010 alone reached 1,200 exabytes,⁴ an amount equivalent to thousands of billions of issues of the venerable British magazine. The story added that the total amount of extant digital information totals several zettabytes. Here's a simple explanation to help you grasp the challenge we're dealing with:

1 bit, or binary digit, is the basic unit of information in computing. It represents either 0 or 1.

1 byte (the amount of information necessary to encode a letter or a number)
= 8 bits

1 kilobyte = 1,000 bytes

1 megabyte = 1,000 kilobytes or 1,000,000 bytes (10^6)

1 gigabyte = 1,000 megabytes or 1,000,000,000 bytes (10^9)

1 terabyte = 1,000 gigabytes or 1,000,000,000,000 bytes (10^{12})

1 petabyte = 1,000 terabytes or 1,000,000,000,000,000 bytes (10^{15})

1 exabyte = 1,000 petabytes or 1,000,000,000,000,000,000 bytes (10^{18})

1 zettabyte = 1,000 exabytes or 1,000,000,000,000,000,000,000 bytes (10^{21})

1 yottabyte = 1,000 zettabytes or 1,000,000,000,000,000,000,000,000 bytes (10^{24})

⁴ "All too much: monstrous amounts of data." *The Economist*, Feb. 25, 2010.

Confused? Don't worry. You're not alone. A yottabyte of information is such a huge number that it is impossible to imagine. In August 2010, Erich Schmidt, former CEO of Google, announced in a conference that between the beginning of time and 2003, humanity generated roughly five exabytes of data, whereas we now produce the same volume of bits *every two days*.

“The information explosion is so profoundly larger than anyone ever thought,” said Schmidt. Five exabytes is more than 200,000 years of DVD-quality video.⁵

To be fair, not all that “information” is what you would call information in a colloquial conversation. Most of it is the product of automated processes and communications between computers, mobile phones, and other devices—nothing that a human brain can understand. But still.

Let's catch our breath here and move on.

From Information to Wisdom

In the 1970s, years before access to the Internet was universal, Richard Saul Wurman, then a professor of architecture in North Carolina, predicted that the oncoming information explosion would require the intervention of a new breed of professionals trained in organizing data and making sense of it. According to Wurman, the biggest challenge our species was about to face was to learn how to navigate the upcoming tsunami of bits that was cresting the horizon.

Wurman called these people *information architects*. Their discipline, *information architecture*, has been defined by others as:

- The structural design of shared information environments;
- The combination of organization, labeling, search, and navigation systems within websites and intranets;
- The art and science of shaping information products and experiences to support usability and findability;
- An emerging discipline and community of practice focused on bringing principles of design and architecture to the digital landscape.

Wurman suggests that one of the main goals of information architecture is to help users avoid *information anxiety*, the “black hole between data and knowledge.”

People still have anxiety about how to assimilate a body of knowledge that is expanding by the nanosecond.... Information anxiety is produced by the

⁵ Google Atmosphere 2010 conference.

ever-widening gap between what we understand and what we think we should understand.⁶

The gap is better represented through the diagram in **Figure 1.8**, which shows the steps separating the two extremes of Wurman’s maxim. It is based on several models known as DIKW Hierarchies (Data, Information, Knowledge, Wisdom). Although the models have been criticized as simplistic and vague,⁷ they are useful for explaining what visualizations and graphics are about.

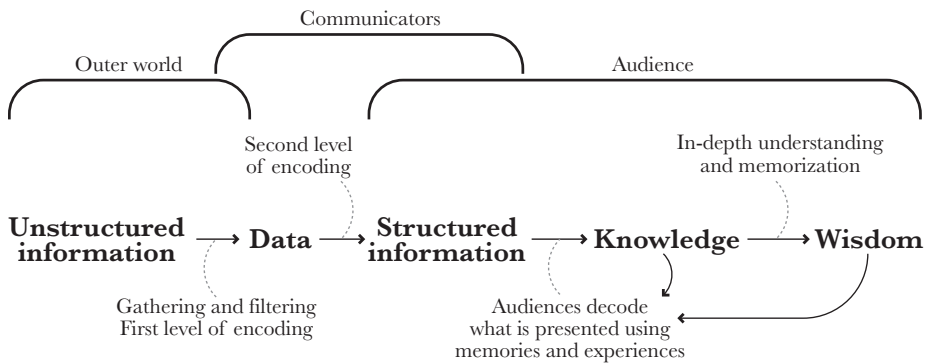


Figure 1.8 From reality to people’s brains.

In the diagram, **unstructured information** means **reality**, the world out there in all its glorious complexity. Every phenomenon that can be perceived or measured can be described as information.

Data are records of observations. Data can be encoded as symbols (numbers and words) that describe and represent reality. In between unstructured information and data, you can see a **first level of encoding**. Imagine a researcher studying the fertility rate. The data would be the records the researcher makes in a spreadsheet, for instance: 2, 5, 6, 2, 2, 2, 1, 1, 4, 3, 3 (and so on) children per woman.

The **second level of encoding** takes us from data to **structured information**. This happens when a communicator (a researcher, a journalist, or whomever) represents data in a meaningful way, using text, visuals, or other means. We can also say that this communicator has given shape to data, so that relevant patterns become visible.

⁶ Wurman, p. 14.

⁷ David Weinberg, “The Problem with the Data-Information-Knowledge-Wisdom Hierarchy,” *Harvard Business Review*, February 2010.

Information consumption can lead to higher **knowledge** on the part of the audience, if its members are able to perceive the patterns or meaning of data. It is not a passive process; our brains are not hard drives that store stuff uncritically. When people see, read, or listen, they assimilate content by relating it to their memories and experiences.

We reach **wisdom** when we achieve a deep understanding of acquired knowledge, when we not only “get it,” but when new information blends with prior experience so completely that it makes us better at knowing what to do in other situations, even if they are only loosely related to the information from which our original knowledge came. Just as not all the information we absorb leads to knowledge, not all of the knowledge we acquire leads to wisdom.

Every step in our diagram implies higher **order**. When we see the world, we unconsciously impose organization on the unstructured information that our eyes gather and transmit to the brain. We create hierarchies. We don't perceive everything in front of us at once, as we'll see later in this book. A moving entity, for instance, attracts our attention more than a static one, because movement may suggest an approaching threat. We therefore process the position and identity of the moving object before paying attention to anything else. Our brain gives **meaning** to the object, even if we are not aware of the reason why.

In the words of Kevin Kelly, a famous philosopher of technology, in his book *What Technology Wants* (2010):

Minds are highly evolved ways of structuring the bits of information that form reality. That is what we mean when we say a mind understands; it generates order.

So, without conscious effort, the brain always tries to close the distance between observed phenomena and knowledge or wisdom that can help us survive. This is what cognition means. **The role of an information architect is to anticipate this process and generate order before people's brains try to do it on their own.**

Making Reality Visible

Today, *information architect* refers broadly to professions with very different sets of tools and theoretical traditions. Outside academic circles, an information architect can be someone who writes technical handbooks, a software engineer, a web developer, a wayfinding designer (yes, that's a profession; who do you think creates public spaces that can be navigated with ease?), and that nerdy guy who makes charts on fertility just for the fun of it.

All of those professions share the goal of making the world easier for audiences and users, but that's too broad a goal to put them all in the same bag. For my purposes in this book, information graphics and visualization is a form of information architecture. But how can we be more precise in describing the relationship between the branch and the trunk?

Look at **Figure 1.9** and imagine information architecture as a big circle. Inside is the set of disciplines devoted to dealing with information. Among the most relevant disciplines is *information design*, defined by Stanford University's Robert E. Horn as "the art and science of preparing information so that it can be used by human beings with efficiency and effectiveness."⁸ The goal of the information designer is to prepare documents (both analog and digital) and spaces so they can be navigated effortlessly.

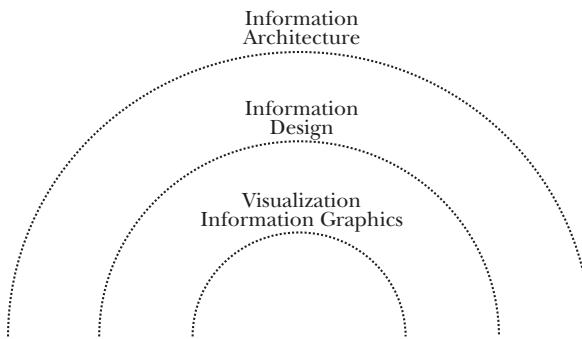


Figure 1.9 Information graphics is a form of information design. Information design branches from information architecture.

A significant part of information design is information graphics and visualization. Academic literature sometimes separates infographics from visualization and defines the latter as "the use of computer-supported, interactive, visual representations of data to amplify cognition,"⁹ but I prefer the more straightforward take of Joan Costa, a famous Spanish professor of design: To visualize is "to make certain phenomena and portions of reality visible and understandable; many of these phenomena are not naturally accessible to the bare eye, and many of them are not even of visual nature."¹⁰

⁸ Robert Jacobson, ed. *Information Design*, (Cambridge: MIT Press, 2000).

⁹ Card, Stuart, Jock, Mackinlay, and Ben Shneiderman. 1999. *Readings Information Visualization: Using Vision to Think*. San Francisco: Morgan Kaufmann.

¹⁰ Joan Costa, *La esquemática: visualizar la información* (Barcelona: Editorial Paidós, Colección Paidós Estética 26, 1998). One of the best books I've read on the visual representation of information. No English version available, unfortunately.

Why does Costa add that second part about not being necessarily of visual nature? Because graphical displays can be either figurative or non-figurative. To understand figurative displays, think of a map as a scaled portrait of a geographical area, or a manual that explains through illustrations how to use your new washing machine, or a news infographic on a catastrophic plane crash, like the one in **Figure 1.10** (pages 20 and 21), a superb project by *Público*, a medium-sized Spanish newspaper with a small but extremely talented graphics desk.

Other graphics that display abstract phenomena are **non-figurative**. In these, there is no mimetic correspondence between what is being represented and its representation. The relationship between those two entities is *conventional*, not natural (see **Figure 1.11**). The unemployment rate doesn't really resemble a grid of multicolored rectangles, does it?

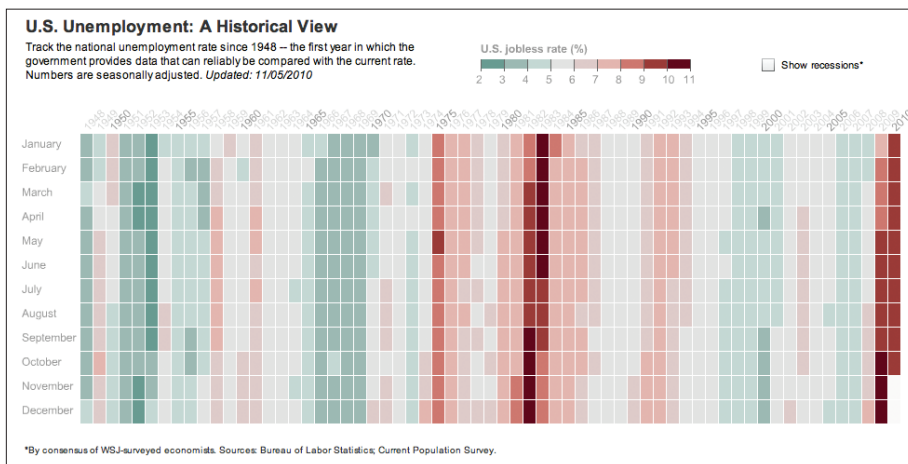


Figure 1.11 *The Wall Street Journal*. Web chart U.S. Unemployment: A Historical View. (Reprinted with permission of *The Wall Street Journal*, Copyright © 2010 Dow Jones & Company, Inc. All Rights Reserved Worldwide.)

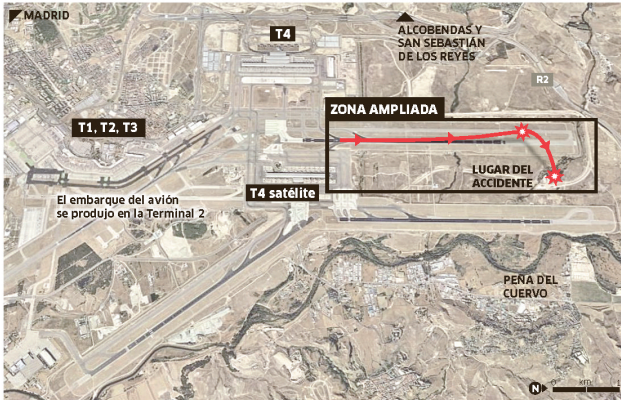
Visualization as a Technology

Let me introduce an idea crucial to the premise of this book: **Visualization should be seen as a technology**. That may sound odd. When someone mentions technology in a routine conversation, we usually think of machines: MP3 players, cars, refrigerators, electric toothbrushes, lawn mowers, computers. But what do all those devices have in common? I don't mean *physically*, but in their very essence:

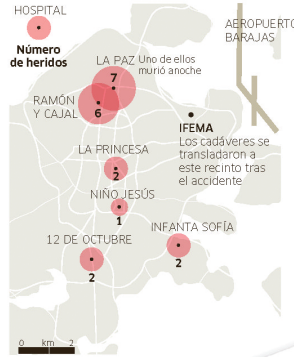
Radiografía de la catástrofe

El vuelo 5022JK de Spanair se salió de la pista de despegue a las 14:45 con 172 personas a bordo // El accidente activó todos los servicios de emergencias de Madrid // 153 ocupantes han perdido la vida

Localización



Dónde se trasladó a los heridos



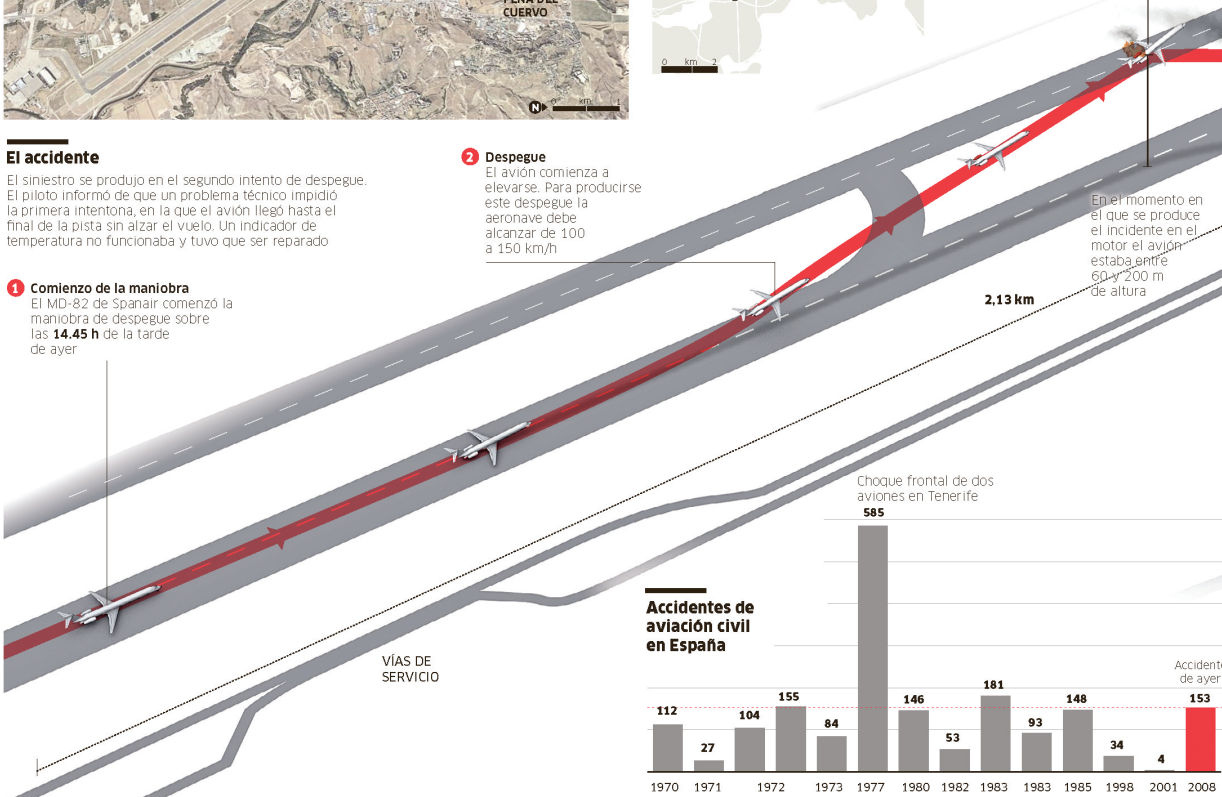
3 Incidente
 En los primeros momentos en el aire surge un problema en el motor izquierdo. Algunos testigos afirman que se produjo una explosión y que el avión comenzó a caer como una bola de fuego, si bien el problema no estaba confirmado al cierre de esta edición

El accidente

El siniestro se produjo en el segundo intento de despegue. El piloto informó de que un problema técnico impidió la primera intentona, en la que el avión llegó hasta el final de la pista sin alzar el vuelo. Un indicador de temperatura no funcionaba y tuvo que ser reparado

2 Despegue
 El avión comienza a elevarse. Para producirse este despegue la aeronave debe alcanzar de 100 a 150 km/h

1 Comienzo de la maniobra
 El MD-82 de Spanair comenzó la maniobra de despegue sobre las 14.45 h de la tarde de ayer



FUENTE: AGENCIAS, BO BING, CONSEJERÍA DE INTERIOR DE LA COMUNIDAD DE MADRID, SPANAIR, GOOGLE EARTH Y ELABORACIÓN PROPIA

Figure 1.10 Público (Spain). Plane crash in Barajas airport, Madrid, August 21, 2008. Infographic by Chiqui Esteban, Mónica Serrano, Álvaro Valiño.

El dispositivo de emergencia



Balance de víctimas

En el avión viajaban **172** personas, **163** eran pasajeros y los otros **nueve** miembros de la tripulación



4 Caída

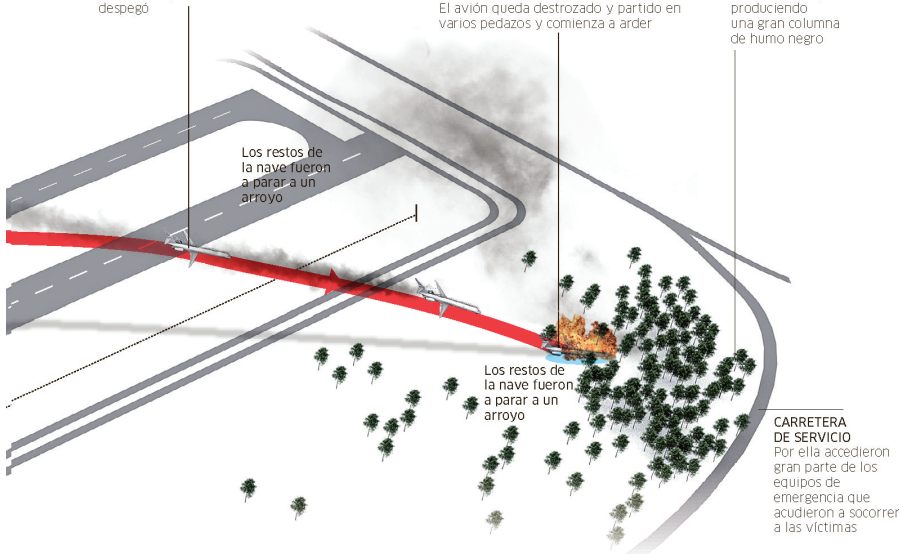
La nave se precipita hacia la derecha de la pista en la que despegó

5 Impacto

El avión termina en una zona arbolada entre las dos pistas de despegue y aterrizaje más cercanas a la Terminal 4. El avión queda destrozado y partido en varios pedazos y comienza a arder

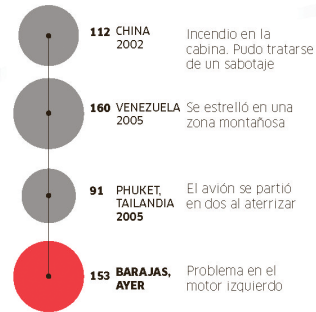
6 Incendio

El fuego se propaga por la zona arbolada, produciendo una gran columna de humo negro



El historial negro del MD-82

La nave, que ya se ha dejado de construir, cuenta con un historial de cerca de 500 muertos en estos últimos seis años



El avión

El avión iba bastante cargado de combustible para cubrir la distancia del itinerario previsto



MD-82 CONFIGURADO PARA 172 PASAJEROS

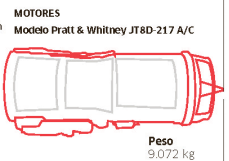
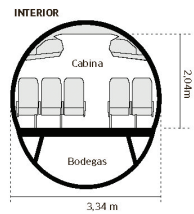
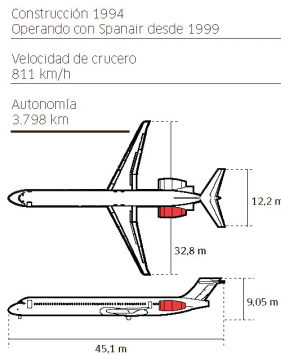


GRÁFICO: CHIQUI ESTEBAN, MÓNICA SERRANO, ALVARO VALIÑO

1. **They are extensions of ourselves.** Canadian media thinker Marshall McLuhan was the first to advance this notion half a century ago. A lawn mower helps us keep our garden neat without having to use our bare hands. An electric toothbrush reaches small spaces in between our teeth that would be impossible to reach with a toothpick—which is also a technology, by the way. An MP3 player is not just a player, but also a device that helps us remember the songs that define the best and worst moments of our lives.
2. **They are means to reach goals.** Sometimes it's just one goal—a freezer keeps food fresh—but it can be several. Think of a computer, whose functionality depends on other technologies such as software that we install. Technologies can harbor other technologies.

Visualization as technology has these same traits. Also, the word *technology* has various meanings and is potentially open to interpretation. I use it here with the sense given by W. Brian Arthur in his foundational *The Nature of Technology* (2009). According to Arthur, we can use the word *technology* in three different senses, illustrated in **Figure 1.12**: Technologies-singular, technologies-plural, and technology-general.

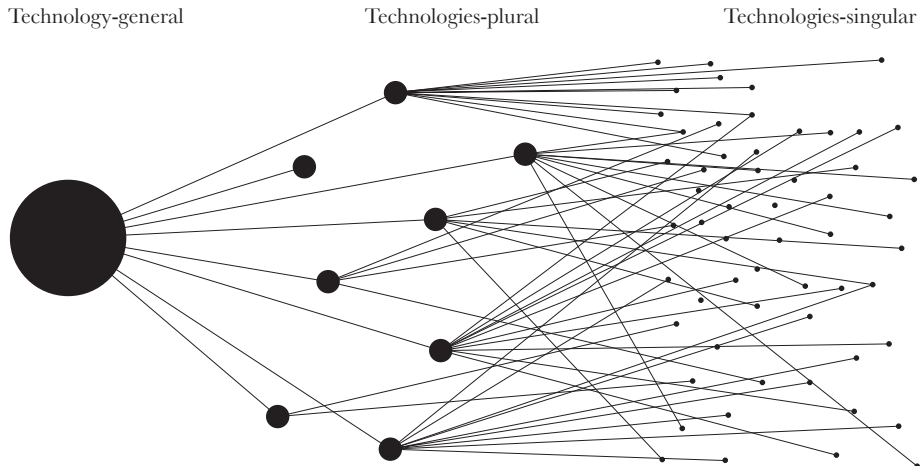


Figure 1.12 The three kinds of technology: general, plural, and singular.

What Arthur means is that technology can be, first, any object, process, or method devised to aid in a task, “a means to fulfill a human purpose.” This can be called *technology-singular*. The refrigerator and other devices described above are examples. The algorithms that run a software program as well as the letters, sentences, and

paragraphs that help me communicate with you are also technologies-singular. Even a humble nail is a technology.

Technologies-plural describe “assemblages of practices and components.” In other words, they are groups of technology-singular items that, when put together, make something new emerge. Electronics, biochemistry, and the varieties of engineering are all technologies-plural. Arthur also refers to technology-plural groups as “bodies of technology.”

The third meaning of technology is “the entire collection of devices and engineering practices available to a culture,” or *technology-general*.

How is this relevant for information graphics and visualization? First, as emerging disciplines, they are still a formative hodge-podge of concepts, methods, and procedures borrowed from many areas: the principles of map design (from cartography); guidelines on how to better display data on a chart (from statistics); rules on best practices for the use of type, layout, and color palettes (from graphic design); principles of writing style (from journalism); and more, including a wide array of software tools.

Second, and more important, **individual information graphics are also technologies, means to fulfill purposes, devices whose aim is to help an audience complete certain tasks**. This apparent no-brainer will have consequences later on. If you accept that a visualization is, above all, a *tool*, you are implicitly accepting that the discipline it belongs to is not just art, but *functional art*, something that achieves beauty not through the subjective, freely wandering self-expression of the painter or sculptor, but through the careful and restrained tinkering of the engineer.