

PART II

How Should You Design It?

Now that we know how to think about *why* and *what* we design, both in terms of our purposes and the relationships that define visualizations, let's talk about the *process* of designing visualizations.

Ultimately, the key to a successful visualization is making good design choices. Artists and designers of all kinds study their craft to learn how to do just that. So do programmers. While there is certainly room in these disciplines for creativity (yes, some programming languages are designed to have One Right Way to do things, but most allow for some flexibility and discernment on the part of the programmer), ultimately there is a *process* one follows to determine the optimal outcome for a given project.

Even the most whimsical creators—Jackson Pollock, for instance—use a process. Yes, there is room within that process for playfulness, beauty, and happenstance, but a series of decisions provides a structure that leads from point A to point B. Our aim in this section of the book is to lead you through the process of designing good data visualizations. This should get you 80% of the way to genius; that other 20% is the creative spark you'll have to provide on your own.

In data visualization, the number one rule of thumb to bear is mind is: *Function first, suave second*. Your visualization may look really slick, but if it's not communicating the information you're encoding (and doing so efficiently), then it's just so much visual noise—a pretty picture to put on the wall, maybe, but not a useful tool for imparting information.

Going for sexy, suave, and slick first is an easy trap to fall into. You may begin to notice (with surprise, or not) how very many visualizations, even some by professional designers, fall into this trap. Our goal is to help keep *you* out of it.

But we're using a visual medium, right? So beauty and aesthetics matter?

Absolutely, but there is nothing quite so beautiful as the methodical, streamlined transportation of particulars from one mind to another. This is elegance. This is efficiency. This is your brain on data visualization. And there's a process for that.

Which is, of course, to say that you *can have* suave. You just have to do your due diligence first, so that the visual (or literal!) bells and whistles that you creatively add later on *enhance* and *enable* the function of your design, rather than distract or detract from it.

This part of the book looks at how to do the process properly—from determining your goals, to selecting which dimensions of your data to present and highlight, to planning and implementing a successful encoding—so that you will be successful.

Function first, suave second. Let's go!

Determine Your Goals and Supporting Data

Knowledge Before Structure

OK, here we go, ready to build a visualization! And the first step is...step back and think. Successful products (and software) begin with well-thought-out designs, and spending time thinking before diving into implementation provides a huge return on the time invested up front.

The very first thing to think about is the goal of your visualization. To be useful, that goal must be defined before the implementation phase has started. The visualization (like any product or communication) *must* then be designed with that goal in mind.

The goal of your visualization is going to be informed by your own goals and motivations as well as the needs of your reader. A visualization's goal is usually to satisfy a need for specific information on the part of your reader (an informative visualization, as discussed in [Chapter 2](#)). But the goal may also be to change the reader's opinions or behavior in some way (a persuasive visualization, also covered in [Chapter 2](#)).

Examples of goals for visualizations include: to monitor systems, find bargains, compare company performances, select suitable solutions, track populations, tell stories, find specific data points, find outliers, show trends, support arguments, or simply give an overview of the data.



Goals should be stated in terms of the knowledge that can be acquired from the visualization, and should avoid any references to specific content or implementation—don't box yourself into a specific approach yet.

To begin defining the goal of your visualization, ask yourself, *What information need am I attempting to satisfy with this visualization?* Related questions include:

- What values or data dimensions are relevant in this context?
- Which of these dimensions matter; matter most; and matter least?
- What are the key relationships that need to be communicated?
- What properties or values would make some individual data points more interesting than the rest?
- What actions might be taken once this information need is satisfied, and what values will justify that action?

You eventually should be able to *concisely* state the goal of the visualization in *specific terms*, and without motioning any details of implementation. The goal is a pure statement about *function*. Consider which statement in each pair below is the better statement.

- Show the sales figures / Show which product lines are performing best and worst in each region, for each of the last five quarters
- Compare the demographics of Twitter and Facebook users / Compare the ages, education, and income levels of Twitter and Facebook users
- Show a timeline of the performance of every NFL team for the 2011 season / Allow users to compare individual performance metrics for any pair of teams or for the entire league for the 2011 season

While it's obvious that the second statement in each pair is more specific, stop and think for a moment about why they're also better goal statements. They indicate *specific values and relationships* that the visualization should make available to the reader. This level of specificity will directly guide the design process in a particular direction, indicating what data dimensions should be included, how they should be related to each other, *and why*.

In contrast, the first statement of each pair conveys a general intention, but no particular direction, and lends no specific guidance in determining what to include, how it should relate to anything else, or how to tell if and when you've achieved your goals.



If some statements seem to have conflicts in priority, you may have defined your goals too broadly.

Note also that the better statements have no mention of particular implementations. They don't specify using a timeline or bubble plot or pie graph. They talk about *what* knowledge is displayed, but *not how it is structured*.

As soon as your goal statement includes any detail of implementation, you've killed your own potential for creativity. You have preemptively selected one particular solution out of many that are possible, and constrained all of your future thinking to within the bounds of that box. If you want to be creative and innovative and come up with

superior solutions, delay the decisions about implementation as long as possible, until you've really got a good handle on the function you want to provide and the goal you need to achieve. Then you can select (or invent!) the best solution for the parameters of each given situation. The easiest way to think outside the box is to not put your self into a box in the first place.



Pro tip: this applies to programming (and relationships) as well. For best results, have a complete conversation about desired functionality before you're allowed to discuss implementation.

Having a clear understanding of the goals for your visualization will help you achieve your overall communication goals as well.

Avoiding TMI

Look at your inbox. There are probably too many messages there. A few of them are *really* important. Some are probably somewhat useful. A lot of them are likely noise, and make finding the relevant messages more difficult. The data in a visualization works the same way. The more information that's included, the harder it is to find the information that you (or your readers) care about.

Extraneous content will obscure the message and complicate the extraction of knowledge.

It is tempting to show as much data as possible. We have access to *a lot* of data, and it's easy to be seduced into including more data because it would be cool, or because it's more impressive that way, or simply because we can. But the choice of what to include must be based on the priorities we've discussed: your goals, the needs of your reader, and the shape of your data. It's more powerful to reveal knowledge and facilitate insight than it is to overwhelm with sheer volume.

When reading a visualization (or any other kind of communication), your reader has a limited amount of brainpower to dedicate to the problem (Figure 3-1). Some of this brainpower will be dedicated to *decoding* the visualization; any brainpower that is left may then be used to *understand* the message (if the reader hasn't yet given up in frustration).

In order to be successful, your job is to facilitate understanding by minimizing the amount of extraneous searching and decoding necessary to get the message.

One important way that you can limit the amount of extraneous searching and decoding your reader has to do is to choose encodings that most are easily decoded by the brain. This is the subject of [Chapter 4](#) and [Chapter 5](#), so we will cover it in much more detail shortly.

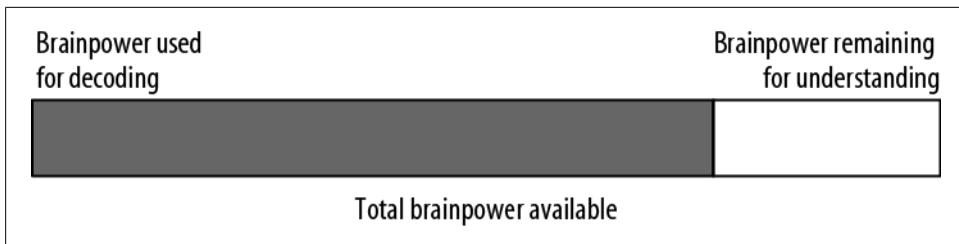


Figure 3-1. Readers have a finite amount of brainpower to apply to your visualization. Any resource used on decoding depletes what is available for understanding.

The other important way that you can limit the amount of extraneous searching and decoding your reader has to do is to be selective about the dimensions of data you choose to encode. Each one should contribute something directly to your message. If you’re adding labels or shading, or another full dimension of data, that doesn’t contribute to your main message, you’re just adding noise. We suggest, instead, embracing a clean, modern aesthetic, and making your important data easier to access.

Choose Appropriate Visual Encodings

Choosing Appropriate Visual Encodings

As we discussed in “[Data](#)” on page 16, once you know the “shape” of your data, you can encode its various dimensions with appropriate visual properties. Different visual properties vary—or may be modified—in different ways, which makes them good for encoding different types of data. Two key factors are whether a visual property is *naturally ordered*, and how many *distinct values* of this property the reader can easily differentiate. Natural ordering and number of distinct values will indicate whether a visual property is best suited to one of the main data types: *quantitative*, *ordinal*, *categorical*, or *relational* data. (*Spatial* data is another common data type, and is usually best represented with some kind of map.)

Natural Ordering

Whether a visual property has a natural ordering is determined by whether the mechanics of our visual system and the “software” in our brains automatically—unintentionally—assign an order, or *ranking*, to different values of that property. The “software” that makes these judgments is deeply embedded in our brains and evaluates relative order independent of language, culture, convention, or other learned factors; it’s not optional and you can’t* design around it.

For example, position has a natural ordering; shape doesn’t. Length has a natural ordering; texture doesn’t (but pattern density does). Line thickness or weight has a natural ordering; line style (solid, dotted, dashed) doesn’t. Depending on the specifics of the visual property, its natural ordering may be well suited to representing *quantitative* differences (27, 33, 41), or *ordinal* differences (small, medium, large, enormous).

Natural orderings are not to be confused with properties for which we have learned or social conventions about their ordering. Social conventions are powerful, and you

* Or shouldn’t try to: that way madness lies.

should be aware of them, but you *cannot* depend on them to be interpreted in the same way as naturally-ordered properties—which are not social and not learned, and the interpretation of which is not optional.

Color is not ordered

Here's a tricky one: Color (*hue*) is *not* naturally ordered in our brains. Brightness (*lightness* or *luminance*, sometimes called *tint*) and intensity (*saturation*) are, but color itself is not. We have strong social conventions about color, and there is an ordering by wavelength in the physical world, but *color does not have a non-negotiable natural ordering built into the brain*. You can't depend on everyone to agree that yellow follows purple in the way that you can depend on them to agree that four follows three.

The misuse of color to imply order is rampant; don't fall into this common trap. In contexts where you're tempted to use "ordered color" (elevation, heat maps, etc.), consider varying brightness along one, or perhaps two, axes. For example, elevation can be represented by increasing the darkness of browns, rather than cycling through the rainbow (see Figure 4-1[†] and Figure 4-2[‡]).

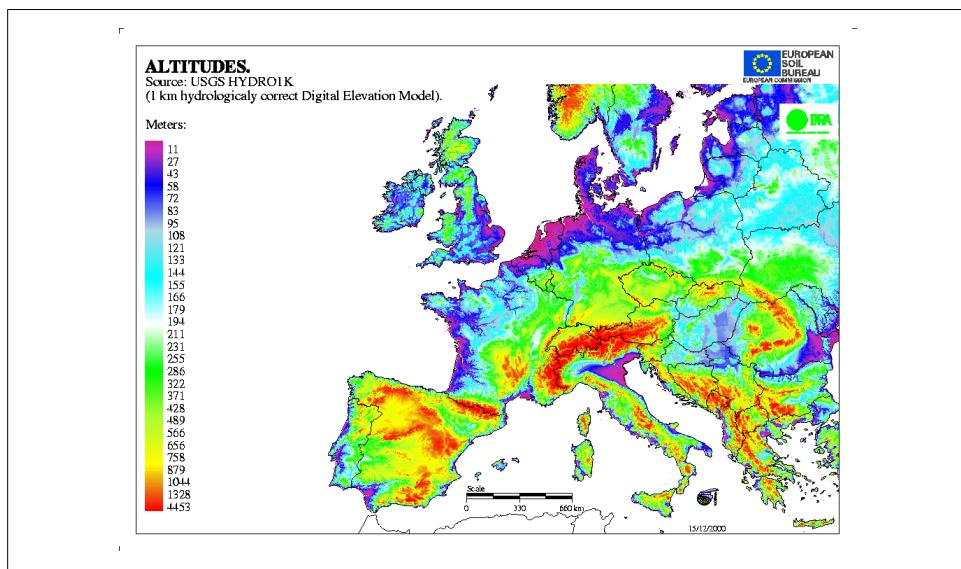


Figure 4-1. A rainbow encoding leads to a map that is very difficult to understand. Does red mean the Alps are hotter than the rest of Europe?

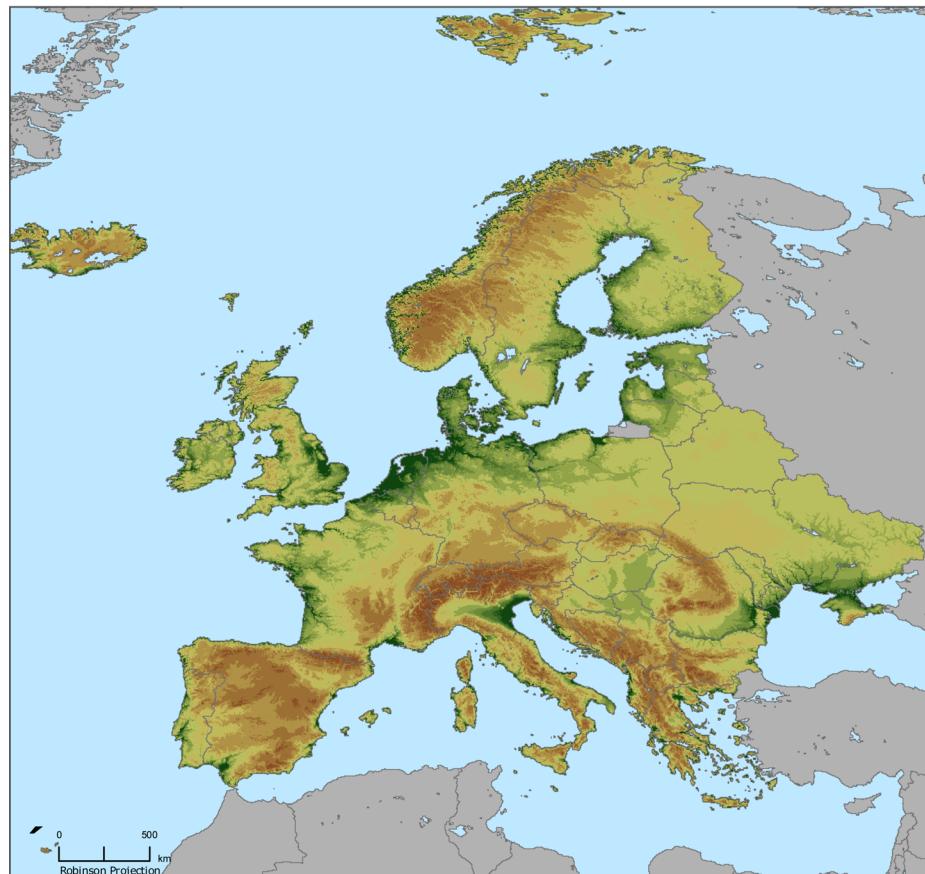
[†] European Soil Bureau. Copyright © 1995–2011, European Union. Used with stated authorization to reproduce, with acknowledgment. <http://eusoils.jrc.ec.europa.eu/>

[‡] Center for International Earth Science Information Network (CIESIN) (2007). Copyright © 2007, The Trustees of Columbia University in the City of New York. Columbia University. Population, Landscape, and Climate Estimates (PLACE). Used under the Creative Commons Attribution License. <http://sedac.ciesin.columbia.edu/place/>

Elevation Zones

Europe

PLACE II
Population, Landscape, and Climate Estimates



Digital elevation data (meters above mean sea level) were obtained as a 1 kilometer resolution elevation\bathymetry raster product from ISciences, L.L.C. Elevation zones were created by aggregating ranges of land elevation values into 12 thematic elevation classes. The 2004 ISciences data were resampled from their native 30 arc-second resolution to match GPW's population and land area 2.5 arc-minute spatial footprint. Source: ISciences, L.L.C. 300 N. Fifth Ave. Suite 120., Ann Arbor, MI 48104 <http://www.isciences.com/>.



Copyright 2007. The Trustees of Columbia University in the City of New York.
Source: Center for International Earth Science Information Network (CIESIN),
Columbia University, Population, Landscape, and Climate Estimates (PLACE).

Further information available at <http://sedac.ciesin.columbia.edu/place/>

Elevation

< 5 m	200 - 399.9 m
5 - 9.9 m	400 - 799.9 m
10 - 24.9 m	800 - 1499.9 m
25 - 49.9 m	1500 - 2999.9 m
50 - 99.9 m	3000 - 4999.9 m
100 - 199.9 m	>= 5000 m

Publish Date: 03/13/07



This document is licensed under a
Creative Commons 2.5 Attribution License
<http://creativecommons.org/licenses/by/2.5/>

Figure 4-2. In this example the colors diverge from one point, clearly indicating low, medium, and high elevations.



For help in choosing appropriate color palettes, a great tool is Color-Brewer2.0, at <http://colorbrewer2.org>.

Distinct Values

The second main factor to consider when choosing a visual property is how many *distinct values* it has that your reader will be able to perceive, differentiate, and possibly remember. For example, there are a lot of colors in the world, but we can't tell them apart if they're too similar. We can more easily differentiate a large number of shapes, a huge number of positions, and an infinite number of numbers. When choosing a visual property, select one that has a number of useful differentiable values and an ordering similar to that of your data (see [Figure 4-3](#)).

Example	Encoding	Ordered	Useful values	Quantitative	Ordinal	Categorical	Relational
	position, placement	yes	infinite	Good	Good	Good	Good
1, 2, 3; A, B, C	text labels	optional alpha or num	infinite	Good	Good	Good	Good
	length	yes	many	Good	Good		
	size, area	yes	many	Good	Good		
	angle	yes	medium	Good	Good		
	pattern density	yes	few	Good	Good		
	weight, boldness	yes	few		Good		
	saturation, brightness	yes	few		Good		
	color	no	few (<20)			Good	
	shape, icon	no	medium			Good	
	pattern texture	no	medium			Good	
	enclosure, connection	no	infinite			Good	Good
	line pattern	no	few				Good
	line endings	no	few				Good
	line weight	yes	few		Good		

Figure 4-3. Use this table of common visual properties to help you select an appropriate encoding for your data type.

Figure 4-4 shows another way to think about visual properties, depending on what kind of data you need to encode. As you can see, many visual properties may be used to encode multiple data types. Position and placement, as well as text, can be used to encode any type of data—which is why every visualization you design needs to begin with careful consideration of how you’ll use them (see Chapter 5).

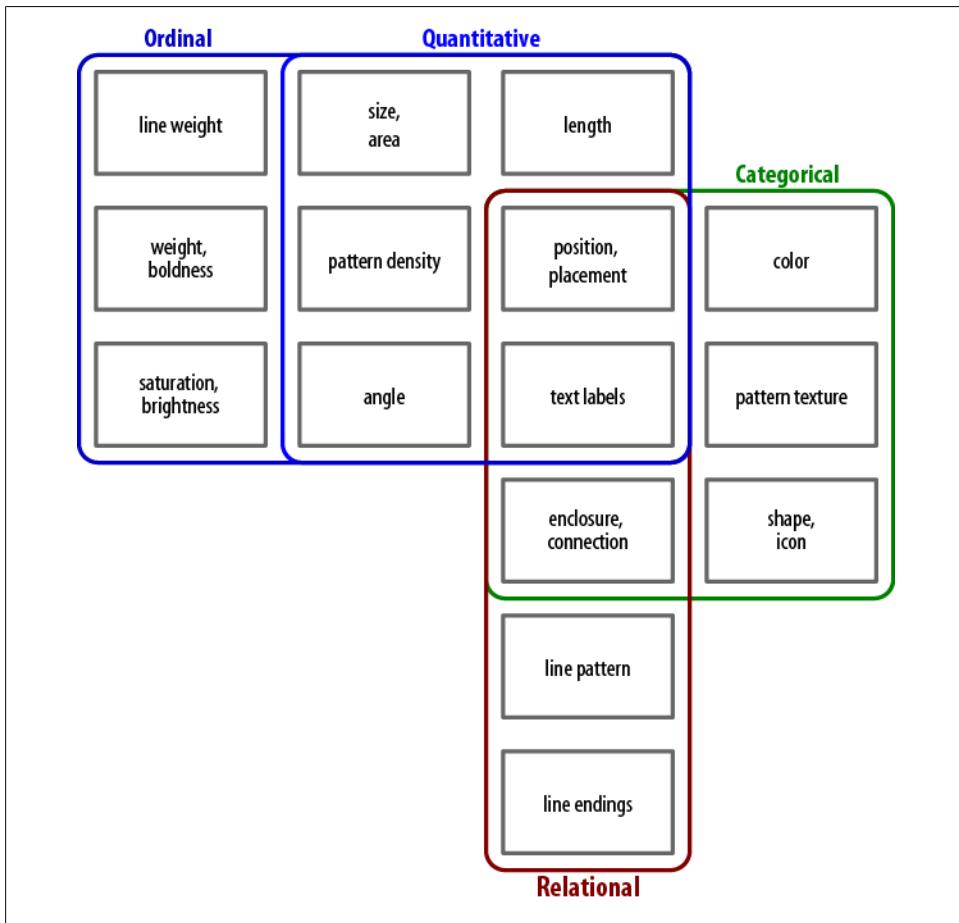


Figure 4-4. Visual properties grouped by the types of data they can be used to encode.

Redundant Encoding

If you have the luxury of leftover, unused visual properties after you’ve encoded the main dimensions of your data, consider using them to redundantly encode some existing, already-encoded data dimensions. The advantage of *redundant encoding* is that using more channels to get the same information into your brain can make acquisition of that information faster, easier, and more accurate.[§]

For example, if you've got lines differentiated by ending (arrows, dots, etc.), consider also changing the line style (dotted, dashed, etc.) or color. If you've got values encoded by placement, consider redundantly encoding the value with brightness, or grouping regions with color, as in [Figure 4-5](#)^{ll}.

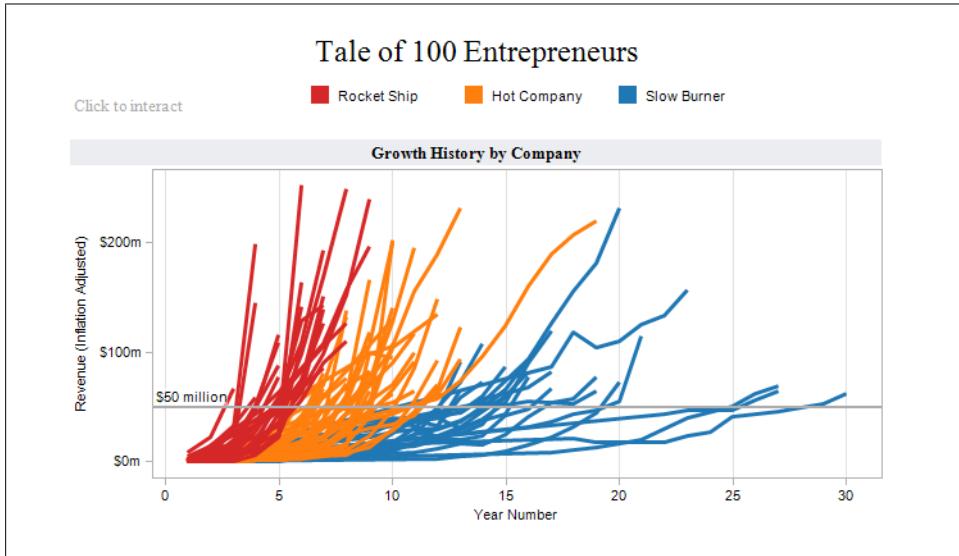


Figure 4-5. Color redundantly encodes the position of groups of companies in this graph.

To be totally accurate, in [Figure 4-5](#), adding color more strongly defined the groupings that weren't strongly defined before, but those groups are a subset of the information already provided by position. For that reason, in this case color adds slightly more informational value beyond mere redundancy.

Defaults versus Innovative Formats

It is worth noting that there are a lot of good *default encodings* and *encoding conventions* in the world, and with good reason. Designing new encoding formats can cost you a lot of time and effort, and may make your reader expend a lot of time and effort to learn. Knowing the expected defaults for your industry, data type, or reader can save you a lot of work when it comes to both figuring out how to best encode your data, and how to explain it to your readers. However, if we all used existing defaults all the time, not much progress would be made. So when should you use a default, and when should you innovate?

[§] Ware, *Information Visualization: Perception for Design* (Morgan Kaufmann), p. 179.

^{ll} Tableau Software Public Gallery. Copyright © 2003–2011 Tableau Software. <http://www.tableausoftware.com/learn/gallery/company-performance>

In writing, we often advise each other to stay away from clichés; don’t use a pat phrase, but try to find new ways to say things instead. The reason is that we want the reader to *think* about what we’re saying, and clichés tend to make readers turn their brains off. In visualization, however, that kind of brainlessness can be a help instead of a hindrance—since it makes comprehension more efficient—so conventions can be our friends.



Purposely turning visual convention on its head may cause the reader’s brain to “throw an exception,” if you will, and this technique can be used strategically; but please, use it sparingly.

The choice comes down to a basic cost-benefit analysis. What is the expense to you and your reader of creating and understanding a new encoding format, versus the value delivered by that format? If you’ve got a truly superior solution (as evaluated by your reader, and not just your ego), then by all means, use it. But if your job can be done (or done well enough) with a default format, save everyone the effort and use a standard solution.

Readers’ Context

In [Chapter 2](#), we discussed how important it is to recognize that you are creating a visualization for someone other than yourself—and that the reader may show up with a mindset or way of viewing the world different from yours.

First, it’s important to point out that your audience will likely be composed of *more than one* reader. And as these people are all individuals, they may be as different from each other as they are from you, and will likely have very different backgrounds and levels of interest in your work. It may be impossible to take the preconceptions of all these readers into consideration at once. So choose the most important group, think of them as your *core* group, and design with them in mind. Where it is possible to appeal to more of your potential audience without sacrificing precision or efficiency, do so. But, going forward, let us be clear that when we say *reader*, what we really mean is a representative reader from within your *core audience*.

Okay, now that we’ve cleared that up, let’s get specific about some facets of the reader’s mindset that you need to take into account.

Titles, tags, and labels

When selecting the actual terms you’ll use to label axes, tag visual elements, or title the piece (which creates the mental framework within which to view it), consider your reader’s vocabulary and familiarity with relevant *jargon*.

- Is the reader from within your industry or outside of it? What about other readers outside of the core audience group?

- Is it worth using an industry term for the sake of precision (knowing that the reader may have to look it up), or would a lay term work just as well?
- Will the reader be able to decipher any unknown terms from context, or will a vocabulary gap obscure the meaning of all or part of the information presented?

These are the kinds of questions you should ask yourself. Each and every single word in your visualization needs to serve a specific purpose. For each one, ask yourself: why use this word in this place? Determine whether there is another word that would serve the purpose any better (or whether you can get away without one at all), and if so, make the change.

Related to this, consider any *spelling preferences* a reader might have. Especially within the English language, there may be more than one way to spell a word depending on which country one is in. Don't make the reader's brain do extra work having to parse "superfluous" or "missing" letters.

Colors

Another reader context to take into account is color choice. There is quite a bit of science about how our brains perceive and process color that is somewhat universal, as we saw earlier in this chapter. But it's worth mentioning in the context of reader preconceptions the significant *cultural associations* that color can carry.

Depending on the culture in question, some colors may be lucky, some unlucky; some may carry positive or negative connotations; some may be associated with life events like weddings, funerals, or newborn children.

Some colors don't mean much on their own, but take on meaning when paired or grouped with other colors: in the United States, red and royal blue to Republicans and Democrats; pink and light blue often refer to boys and girls; red, yellow, and green to traffic signals. The colors red, white, and green may signal Christmas in Canada, but patriotism in Italy. The colors red, white, and blue are patriotic in multiple places: they will make both an American and a Frenchman think of home.

Colors may also take on special significance when paired with certain shapes. A red octagon means *stop* in many places (see [Figure 4-6[#]](#)), but not all.

Color blindness

Of course, we know that there are many variations in the way different people perceive color. This is commonly called *color blindness* but is more properly referred to as *color vision deficiency or dyschromatopsia*. A disorder of color vision may present in one of several specific ways.

[#]Christian Caron (2011). Copyright © 2011, Christian Caron.



Figure 4-6. This stop sign from Montreal is labeled in French, but no English speaker is likely to be confused about its meaning.

Although prevalence estimates vary among experts and for different ethnic and national groups, about 7% of American men experience some kind of color perception disorder (women are much more rarely affected: about 0.4 percent in America).^{*} Red-green deficiency is the most common by far, but yellow-blue deficiency also occurs. And there are lots of people who have trouble distinguishing between close colors like blue and purple.



A great resource for help in choosing color palettes friendly to those with color blindness is the Color Laboratory at <http://colorlab.wickline.org/colorblind/colorlab>. There you can select color swatches into a group (or enter custom RGB values) and simulate how they are perceived with eight types of dyschromatopsia. Note: the simulation assumes that you yourself have typical color vision.

Directional orientation

Is the reader from a culture that reads left-to-right, right-to-left, or top-to-bottom? A person's habitual reading patterns will determine their default eye movements over a

* Montgomery, Geoffrey, for Howard Hughes Medical Institute. *Seeing, Hearing, and Smelling the World*. Chevy Chase, MD: 1995.

page, and the order in which they will encounter the various visual elements in your design.

It will also affect what the reader perceives as “earlier” and “later” in a timeline, where the edge that is read from will be “earlier” and time will be assumed to progress in the same direction as your reader typically reads text.

This may also pertain to geographic maps: many of us are used to seeing the globe split somewhere along the Pacific, with north oriented upward. This suits North Americans just fine, since—scanning from left to right and starting from the top of the page—we encounter our own country almost immediately. The convention came about thanks to European cartographers, who designed maps over hundreds of years with their own continent as the center of the world.

Occasionally, other map makers have chosen to orient the world map differently, often for the same purpose of displaying their homeland with prominence (such as Stuart McArthur’s “South-Up Map,” which puts his native Australia toward the center-top) or simply for the purpose of correcting the distortion effect that causes Europe to look bigger than it really is (such as R. Buckminster Fuller’s “Dymaxion Map”).

Compatibility with Reality

As with so many suggestions in this chapter, a large factor in your success is making life easier for your reader, and that’s largely based on making encodings as easy to decode as possible. One way to make decoding easy is to make your encodings of things and relationships as well aligned with the reality (or your reader’s reality) of those things and relationships as possible; this alignment is called *compatibility*. This can have many different aspects, including taking cues from the physical world and from cultural conventions.

Things in the world are full of *inherent* properties. These are physical properties that are not (usually) subject to interpretation or culture, but exist as properties you can point to or measure. Some things are larger than others, have specific colors, well-known locations, and other identifying characteristics. If your encodings conflict with or don’t reflect these properties, if they are not compatible, you’re once again asking your reader to spend extra time decoding and wondering why things are “wrong;” why they don’t look like they’re expected to (for example, see the boats and airplanes in Figure 4-7).

Figure 4-8 shows an example from <http://html5readiness.com/>.

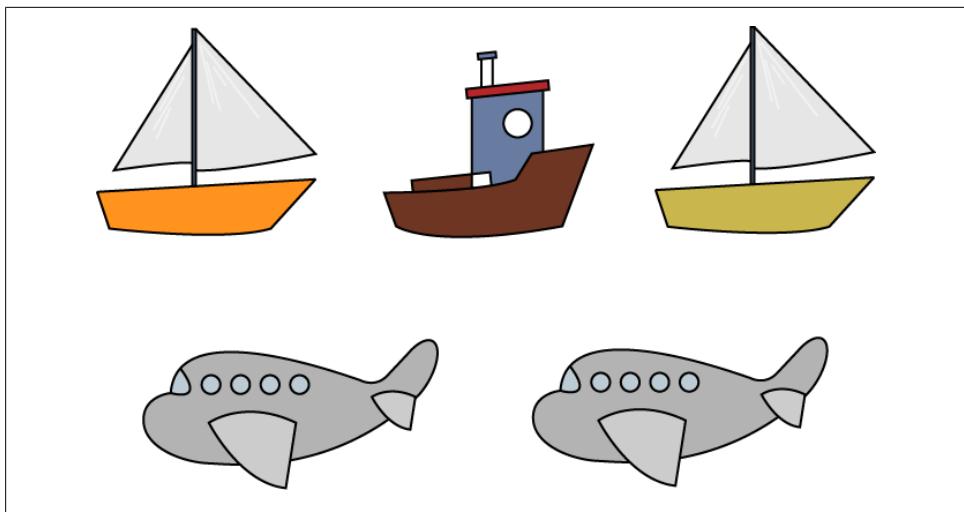


Figure 4-7. The visual placement of boats above airplanes is jarring, since they don't appear that way in the physical world.

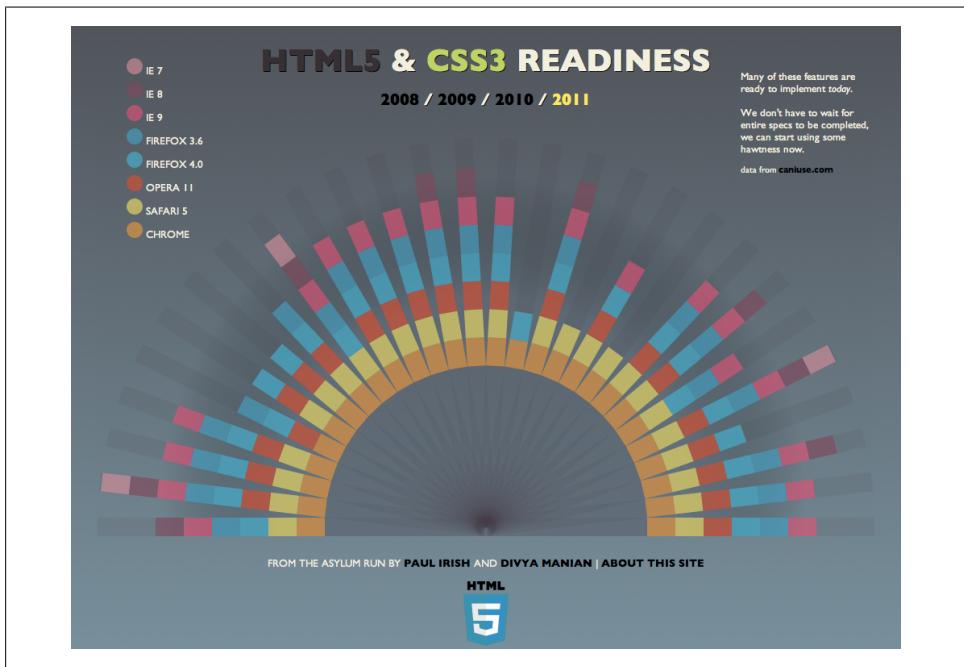


Figure 4-8. Representation of browser capabilities.

Notice how the colors they've chosen map to the browser icons, as shown in [Figure 4-9](#).

Browser	Existing color	Better color
		
		
		
		
		

Figure 4-9. The representative colors differ greatly from the colors in the browser icons. Other choices would better reflect the icons' colors.

The encodings they've chosen aren't very compatible with the reality of the browsers' icons and branding. IE, with a blue and yellow icon, is shown in shades of purple. Firefox, with a blue and orange icon, is shown in blue—which is fine, but curious, given the other browser icons that also contain blue and might be better contenders for the blue encoding. Safari, with a blue icon, is encoded with yellow. Chrome—which has red, blue, green, and yellow, but no orange in its icon—is orange. Opera, with its red icon and corresponding red label, has the only encoding that makes sense. An improved set of encodings that more closely match the reality of the browser icons shown in the last column of [Figure 4-9](#).

Beyond physical or *natural* conventions, there are *learned*, cultural conventions that must also be respected. These may not be as easy to point to, but are no less important.

Note that, as we advised in the section on natural ordering, you should not rely on social or cultural conventions to convey information. However, these conventions can be very powerful, and you should be aware that your reader brings them to the table. Making use of them, when possible, to *reinforce* your message will help you convey information efficiently. Avoid countering conventions where possible in order to avoid creating *cognitive dissonance*, a clash of habitual interpretation with the underlying message you are sending.

To use colors as an example of some of these learned conventions, red and green have strong connotations for bad and good, or stop and go. (See the Color section in [Chapter 6](#) for more on common color associations.) Beyond color, consider cultural conventions about spatial representations, such as what left and right mean politically, or the significance of above and below. Also consider cultural conventions about the meaning of square versus round, and bright versus dark.

All sorts of metaphorical interpretations are culturally ingrained. An astute designer will think about these possible interpretations and work with them, rather than against them.

Direction and reality

Direction is an interesting property to consider because it has both inherent and learned conventions. How many times have you looked at an emergency exit map in a hallway, and realized that the exit, displayed to the left on the map, was to your right in reality, because the map was upside down relative to the direction you were facing?[†] You may also run into maps that, for various reasons, don't put north at the top of the map. Even though the map may be fully accurate and not violating compatibility with physical reality, this violation of cultural convention can be enormously disorienting.

Patterns and Consistency

The human brain is *amazingly* good at identifying patterns in the world. We easily recognize similarity in shapes, position, sound, color, rhythm, language, behavior, and physical routine, just to name a few variables. This ability to recognize patterns is extremely powerful, as it enables us to *identify stimuli* that we've encountered before, and *predict behavior* based on what happened the last time we encountered a similar stimulus pattern. This is the foundation of language, communication, and all learning. The ability to recognize patterns and learn from them allows us to notice and respond when we hear the sound of our name, to run down a set of stairs without hurting ourselves, and to salivate when we smell food cooking.

Consequently, we also notice *violations of patterns*. When a picture is crooked, a friend sounds troubled, a car is parked too far out into the street, or the mayonnaise smells

[†] Your authors take particular interest in examining information design in the world, take every opportunity to do so, and hope that everyone else will start to do the same.

wrong, the patterns we expect are being violated and *we can't help but notice* these exceptions. Flashing lights and safety vests are intentionally designed to stand out from the background—we notice them because they are exceptions to the norm.

Practically speaking, this pattern and pattern-violation recognition has two major implications for design. The first is that *readers will notice patterns* and assume they are intentional, whether you planned for the patterns to exist or not. The second is that when they perceive patterns, *readers will also expect pattern violations to be meaningful*.

As designers, we must be extremely deliberate about the patterns and pattern violations we create. Don't arbitrarily assign positions or colors or connections or fonts with no rhyme or reason to your choices, because your reader will always assume that you meant something by it. If you change the order or membership of a list of items, either in text or in placement, it will be perceived as meaningful. If you change the encoding of items, by position, shape, color, or other methods, it will be perceived as meaningful.

So how should you avoid the potential trap of implying meaning where none is intended? It all comes down to three simple rules.

- Be consistent in membership, ordering, and other encodings.
- Things that are the same should look the same.
- Things that are different should look different.

These sound simple, and yet violations of these rules are everywhere. You can probably think of a few already, and will probably start to notice more examples in your daily life. Maintaining consistency and intention when encoding will greatly enhance the accessibility and efficiency of your visualization, and, as with any good habit, will make your life easier in the long run.

Selecting Structure

Just as we don't write PhD dissertations in sonnet form, or thank-you notes like legal briefs complete with footnote citations, it's important that the structure of your visualization be appropriate to your data.

The structure of a visualization should *reveal* something about the underlying data. Take, for example, one of the most classic data visualizations: the Periodic Table of the Elements ([Figure 4-10‡](#)). This is arguably one of the most elegant visualizations ever made. It takes a complex dataset and makes it simple, organized, and transparent. The elements are laid out in order by atomic number, and by wrapping the rows at strategic points, the table reveals that elements in various categories occur at regular intervals, or *periods*. The table makes it easier to understand the nature of each element—both individually, and in relation to the other elements we know of.

‡ Michael Dayah (1997). Copyright © 1997 Michael Dayah. <http://www.ptable.com>

Periodic Table of Elements

Figure 4-10. This rendition of the classic table makes good use of color and line.

Perhaps because it is so elegant and iconic, the Periodic Table is also one of the most frequently imitated visualizations out there. Designers and satirists are constantly repurposing its familiar rows and columns to showcase collections of everything from typefaces to video game controllers, and, ironically, visualization methods. This phenomenon is a particular peeve to your authors precisely because it violates the important principle of selecting an appropriate structure. With the possible (yet questionable) exception of Andrew Plotkin's Periodic Table of Desserts,⁸ copycat designers are using a periodic structure to display data that is *not periodic*. They are just so many derivative attempts at cleverness.



If you're using a particular structure just to be cute or clever, you're doing it wrong.

If you are tempted to use a periodic table format for your non-periodic data, consider instead a two-axis scatter plot or table, where the axes are well matched to the important

§ <http://ehlong.com/zarf/periodic/index.html>

aspects of your data. This will lead you to a more accurate, and less derivative, final product.^{||}



For another chemistry-oriented example of a specific structure with an entirely different purpose, check out the Table of Nuclides: http://en.wikipedia.org/wiki/Table_of_nuclides

Beyond that, we must refer you to other tomes (we suggest the books by Yau and Kosslyn listed in [Appendix A](#) to begin with, and Bertin for more dedicated readers) to help you select just the right structure for your particular circumstance; as you can see from [Figure 4-11](#), there are too many to address each one directly within the scope of this short book. But here are some general principles and common pitfalls to guide your selection process.C

Comparisons Need to Compare

If you intend to allow comparison of values, set the representations up in equivalent ways, and then *put them close together*. You wouldn't ask people to look at two versions of a photo in different rooms; you'd put them side-by-side. The same goes for visualizations, particularly with quantitative measures. If you want people to be able to meaningfully compare values, put them as near to each other as possible.

Another important comparison principle is that of *preservation*. Just as you would isolate variables in a clinical trial by comparing a test group to a control group—which is similar to the test group except for one variable—you need to isolate visual changes by preserving other conditions, so that the change may be easily and fairly interpreted.

A good example of this is in comparing two graphs. Beware of what scales you use on your axes so that the reader can fairly interpret the graph data. If one graph has a scale of 0 to 10 and the other has a scale of 0 to 5 ([Figure 4-11](#)), the slopes displayed on the graphs will be very different for the same data. Using unequal scales for data you are attempting to compare makes comparison much more difficult.

Some Structures Are Just Inherently Bad

Some formats are just *bad*, and should never be used under any circumstances. Many of the formats that fall into this category do so because they *distort proportion*. There are certain things that our brains are and aren't good at: for example, we are terrible at comparing lengths of curved lines and the surface areas of irregularly-shaped fields. For this reason, concentric circle graphs (see, for example, <http://michaelvandaniker.com/>

^{||} Astute readers will note that the periodic table is also a two-axis layout with carefully chosen axes that reflect, and facilitate access to, the relevant properties of the data.

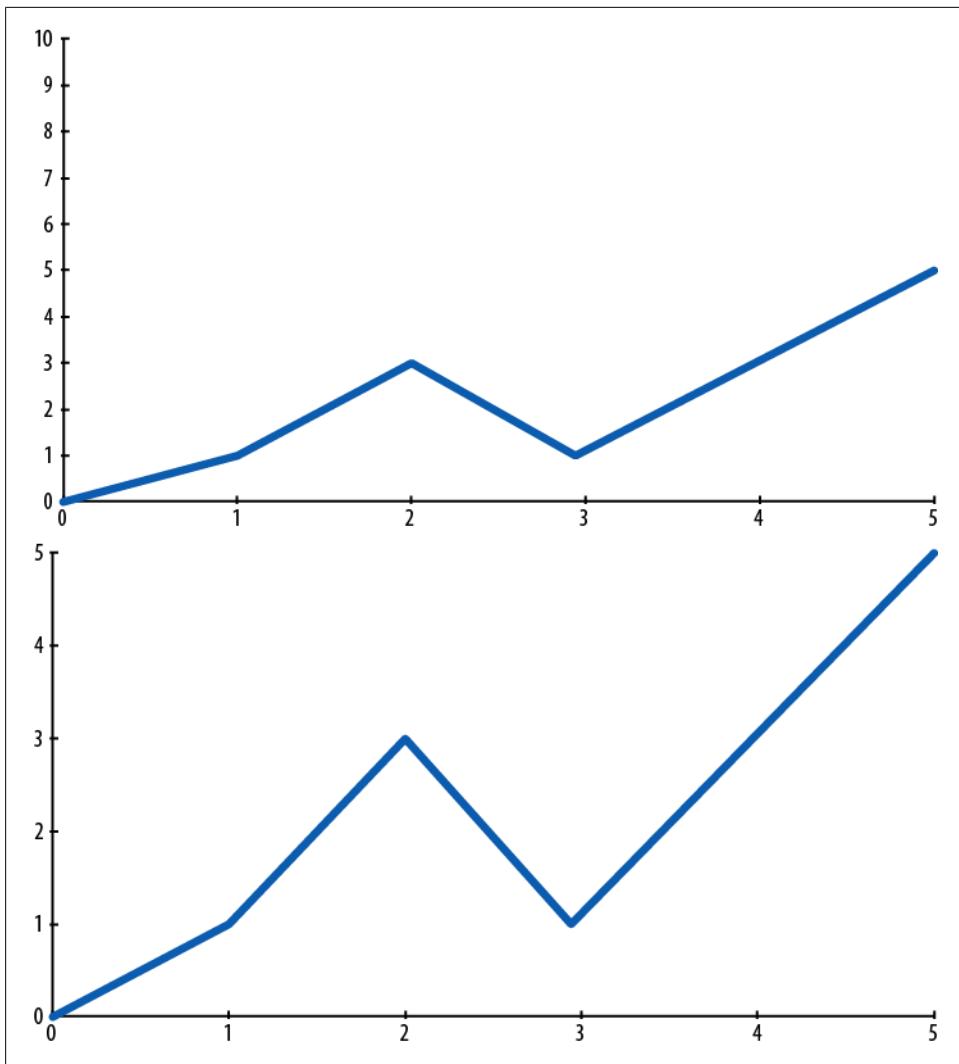


Figure 4-11. The same data appears flatter (top) or steeper (bottom) depending on the scales chosen. If we were attempting to compare these data sets, the unequal axes would introduce distortion that made comparison more difficult.

blog.2009/10/31/visualizing-historic-browser-statistics-with-axiis/) are one of the worst offenders in the world of data presentation structures.[#]

If I show you a section of the ring in the middle that represents a huge percentage, it still looks objectively shorter than a section of the outer ring that may represent a much

[#]We care so much about this issue that we dedicate a section in [Chapter 5](#) to good and bad uses of circular layouts.

smaller percentage. Also, having all of these lines wrapped in a circle makes it difficult to compare their lengths anyway. They only way you can really grasp the information represented in this graph is to read the percentage numbers in the labels. In this case, we may as well just have a table of numbers—it would be faster to read and easier to make comparisons with.

Similarly, the ringed pie graph format known as Nightingale's Rose (for its creator, Florence—see [Figure 4-12](#)), is almost completely useless. Comparing the areas of the sliced pie wedges is nearly impossible to do accurately. Line graphs or stacked bar graphs would have served much better.

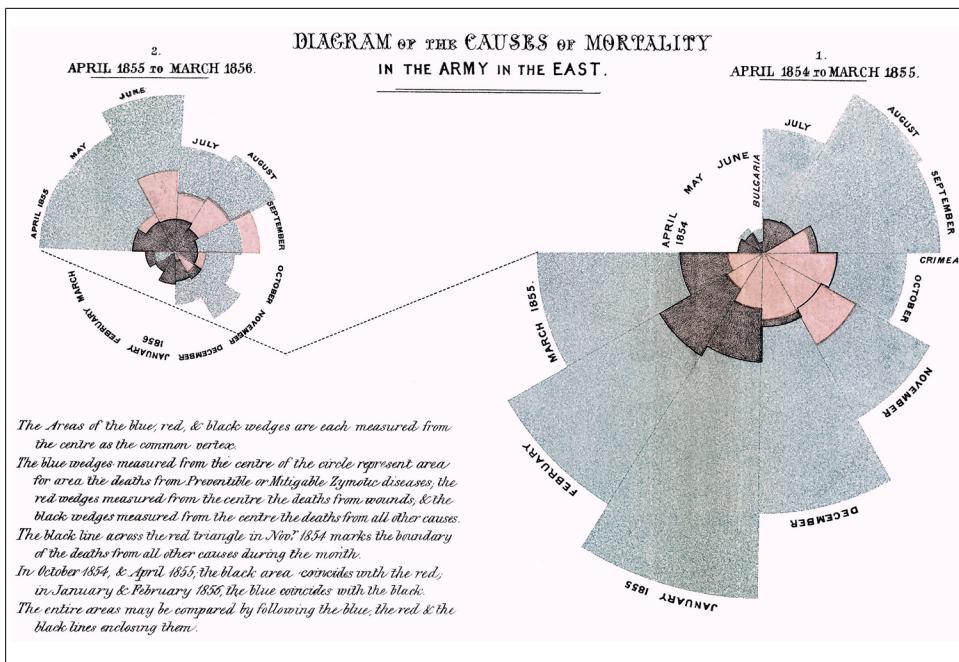


Figure 4-12. Nightingale's Rose.

Unfortunately, this format continues to be reinvented in all sorts of modern contexts. See [Figure 4-13](#) for an equally useless implementation using the same variously sized pie wedges.

Some Good Structures Are Often Abused

There are bad formats, and then there are good formats frequently misused. Like the Periodic Table, pie graphs are useful for a very specific purpose, but quickly devolve into unhelpful parody when drafted into extended service.

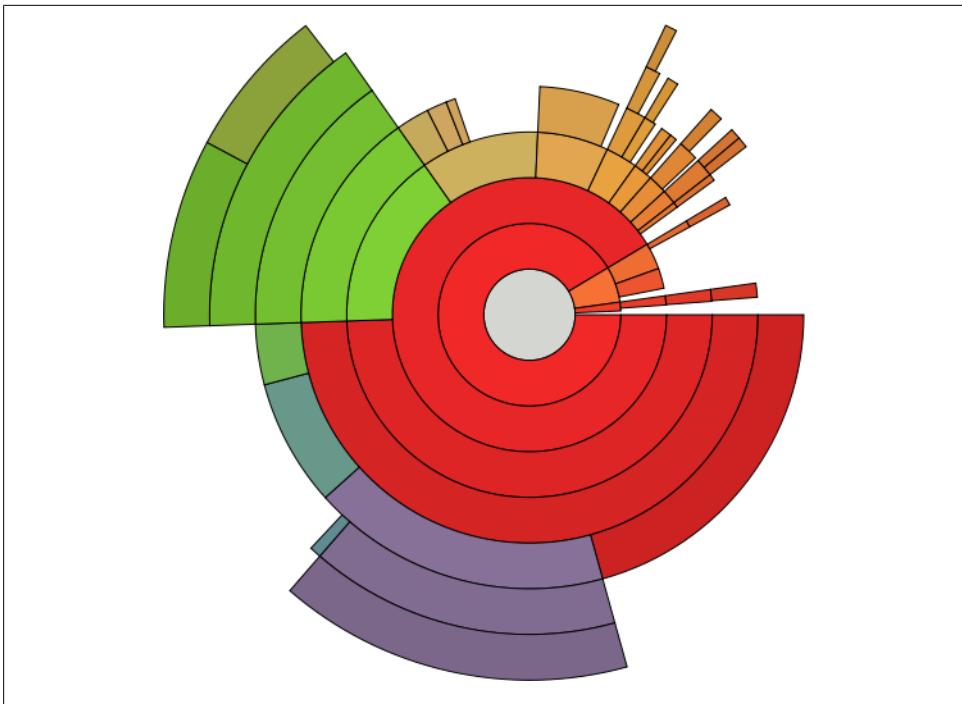


Figure 4-13. A radial layout distorts the data and renders this disk usage map totally ineffective for all but the coarsest comparisons.

The specialty of a pie graph is comparison—specifically, comparison of a few parts to a larger whole. We've already established above in our discussion of concentric circle graphs and Nightingale's Roses that the human brain is lousy at comparing the lengths and surface areas of curved or irregularly-shaped fields; pie graphs fall directly into this category.

Another common pitfall is the use of a geographic map for any and all data that includes a location dimension. Sometimes the use of a map will actually distort your message—such as when the surface area of each region fails to correspond to your population data (see the section on physical reality in [Chapter 5](#)). If your data is tied to population but your display is based on regional size, the proportionally larger surface areas of some regions may inflate the appearance of trends in those regions. Consider using a table or bar graph instead.



If you wish to show regional trends, remember that you don't have to position states or countries alphabetically; it's okay to group them by region or along some other appropriate axis.

Keep It Simple (or You Might Look Stupid)

We talked about careful selection of visual content in [Chapter 3](#), and will talk about selecting and applying encodings well in [Chapter 6](#). But editing (in the sense of minimizing noise to maximize signal) is also a key concept to bear in mind for selecting a useful structure (and *keeping it useful*).

Consider [Figure 4-14](#), which shows an organization chart developed in 2010 by the Joint Economic Committee minority, Republicans. The chart, titled “Your New Health Care System,” depicts the Democratic party’s proposed health care system, and displays a bewildering array of new government agencies, regulations, and mandates, represented by a tangled web of shapes and lines.

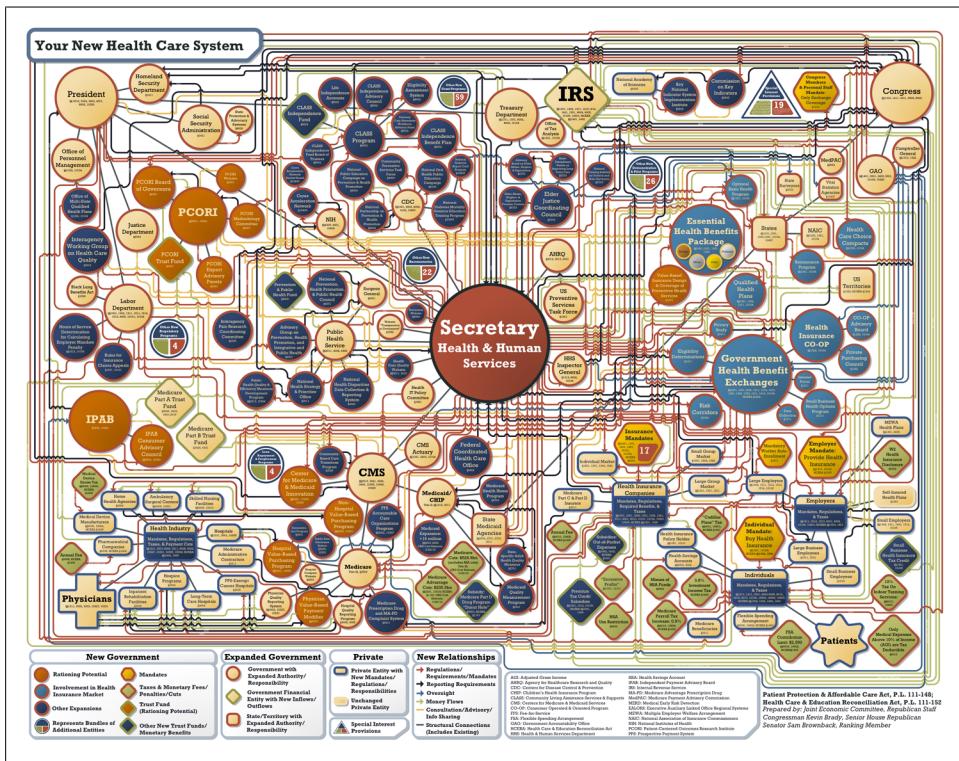


Figure 4-14. This rendition of the healthcare plan clearly revels in and aims to exaggerate the system’s complexity.

It’s fairly obvious that political motivations dominated the design choices for this visualization; it clearly falls into the category of persuasive visualization (rather than informative). The chart itself doesn’t leave the reader with any actual information other than, “Wow, this system is complicated.” When we consider the title of the press release

in which this was unveiled—"America's New Health Care System Revealed"—we know those responsible to be disingenuous.

A citizen designer, Robert Palmer, took it upon himself to make a different, cleaner visual representation of the same proposed health care plan ([Figure 4-15*](#)). His chart is strikingly different from the one created by the Joint Economic Committee minority.

Palmer explained his motivation in an open letter to Rep. John Boehner (R-OH) on Flickr (<http://www.flickr.com/photos/robertpalmer/3743826461/>):

By releasing your chart, instead of meaningfully educating the public, you willfully obfuscated an already complicated proposal. There is no simple proposal to solve this problem. You instead chose to shout "12! 16! 37! 9! 24!" while we were trying to count something.[†]

There is no doubt that national healthcare is a complex matter, and this is evident in both designs. But Palmer's rendition clearly aims to pare down that complexity to its essential nature, for the purpose of making things *easier to understand*, rather than purposefully clouding what is happening under the abstracted layer. This is the hallmark of effective editing.

Sometimes a designer will make the visualization more complicated than it need to be, not because he is trying to make the data look bad, but for precisely the opposite reason: he wants the data to look as good as possible. This is an equally bad mistake.

Your data is important and meaningful all on its own; you don't have to make it special by trying to get fancy. Every dot, line and word should serve a communicative purpose: if it is extraneous or outside the scope of the visualization's goals, it must go. *Edit ruthlessly*. Don't decorate your data.

* Robert Palmer (2010). Copyright © 2010, Robert Palmer. <http://rp-network.com/>

† <http://www.flickr.com/photos/robertpalmer/3743826461/>

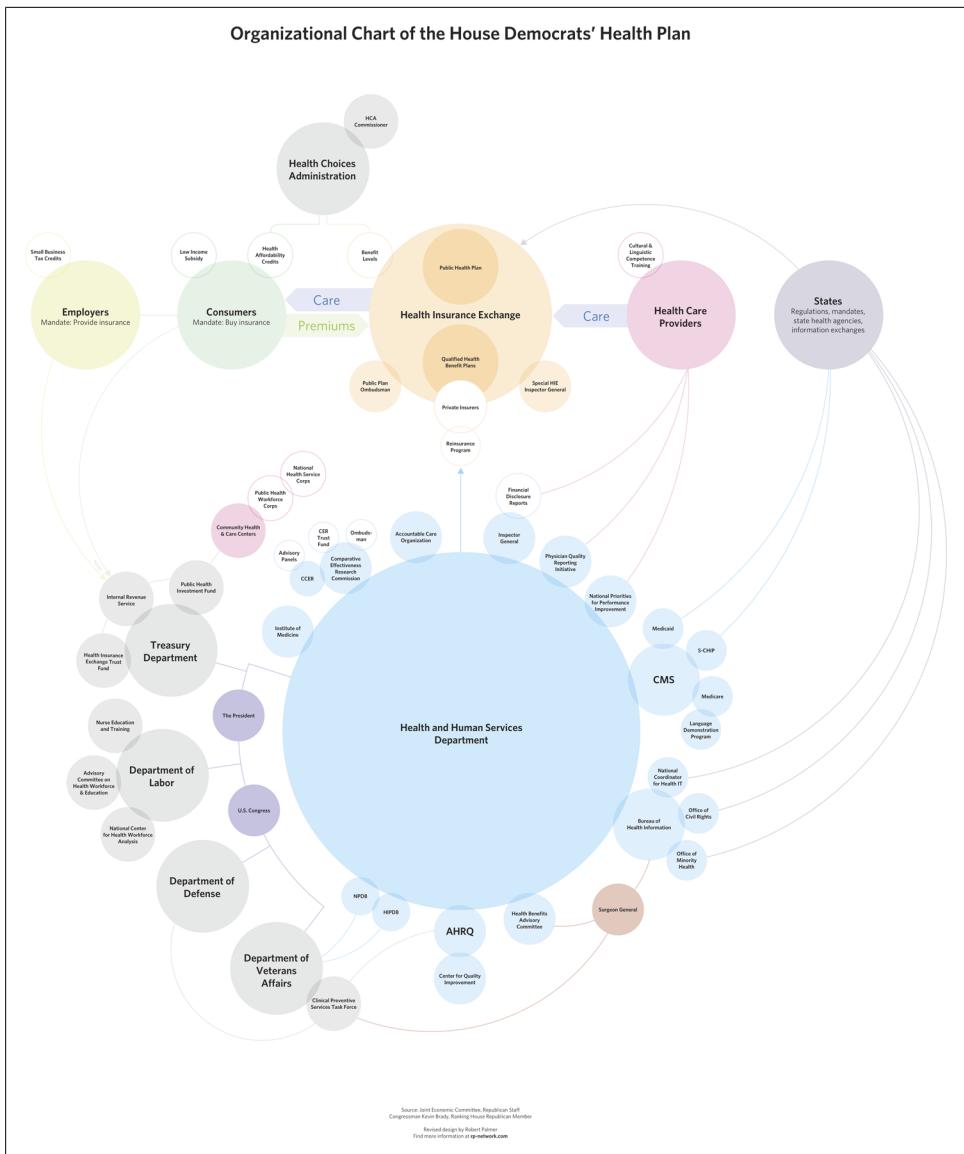


Figure 4-15. Palmer's representation of the same healthcare plan doesn't oversimplify, but is much easier to parse.