

Choosing a successful structure for your visualization

By Noah Iliinsky, IBM Visualization Expert



Contents

- 2 Introduction: Structure
- 2 A step back
- 3 Why is structure so important?
- 4 Actual selection of a structure
- 18 Extensible visualization
- 19 Conclusion
- 19 About the author

Introduction: Structure

One of the most important design decisions for visualization, and certainly one of the most visible, is the layout that you use to frame your data. Should you use a graph, a map, a diagram, a chart or something else? When you answer this question, you have determined the **structure** of your visualization. The structure defines the landscape for presenting your data and consequently defines what sort of information will be most readily available from your visualization. Therefore, selecting the right structure is of utmost importance to communicating your data effectively. Also important is being able to choose the structure you want and not being limited to a few prepackaged visualizations that are included in the release of your business analytics software and applications.

A step back

Before you select a structure, you must make two other major design decisions. Each of these affects all subsequent decisions, so these two decisions are imperative. The common and tempting action is to start the design process by saying “We’ll create a timeline” or “We’ll use a pie graph” without fully considering the options and implications; however, you must defy convention and temptation and make a careful, considered choice.

First, you must know your purpose, or why you are creating this visualization. That includes aspects such as who it is for, what they need to learn from it, what actions or decisions you're trying to enable, and how it will be viewed. Designing for big screen, mobile and print impose very different constraints. The more specific you can be about your purpose, the more likely you are to be successful.

Second, after you have a clear definition of your purpose, you can select the data you need to visualize. How much should you use? The most likely answer is less than you think. When someone is looking for a specific answer, any extraneous data presented to that person is the same thing as noise. Unless you have a definite mandate to show everything, and that usually indicates a too-broad purpose, you probably should select and show only a useful and relevant subset of all the possible data. This subset of data that you've selected for your visualization is referred to as your content.

For more information about appropriate purpose and content, visit <http://complexdiagrams.com/4pillars>.

Why is structure so important?

Given all of the possible visual encodings of data, position, size, color, shape and many others, why this focus on structure? In a word: science. Back in 1984, in "Graphical Perception: Theory Experimentation and Application to the Development of Graphical Methods," William Cleveland and Robert McGill measured the human ability to accurately judge differences of quantities encoded by various visual properties. These were the rankings for the properties they measured:

1. Aligned and then non-aligned position (non-aligned usually implies orthogonal axes)
2. Non-aligned length (bars that do not start at the same base line)
3. Angle and slope (such as pie graphs)
4. Area
5. Volume and saturation (roughly equivalent)
6. Hue¹

They found that humans can estimate and compare the magnitude of position most accurately. Interestingly, the brain does not interpret some visual properties, such as hue, as quantitative at all. Therefore, it is best to use position, the most powerful encoding, to convey the most important data and relationships. A graphical structure that reveals the most important data positionally is going to be most effective. After you have selected a structure to reveal your key data and relationships, you can then add other data dimensions with other visual properties, such as size, shape, color and more.

To learn more about applying other visual properties to represent data, see my white paper, “[Choosing visual properties for successful visualizations](#),” and related materials at complexdiagrams.com/properties.

Actual selection of a structure

Knowing that position is important is a good start, but this knowledge offers little concrete guidance. Luckily, if you have a well-defined purpose, various tools and guides can help you select the right structure. One of these is the [list of visualization options in IBM's Many Eyes tool](#) (Table 1).

















Need	Option
See relationships between data points	 Scatterplot  Matrix chart  Network diagram
Compare a set of values	 Bar chart  Block histogram  Bubble chart
Track rises and falls over time	 Line graph  Stack graph  Stack graph for categories
See the parts of a whole	 Pie chart  Tree map  Tree map for comparisons
Analyze text	 Word tree  Tag cloud  Phrase net  Word cloud generator

Table 1: Many Eyes visualization options.

The visualization types are grouped under headings, and those headings are common, archetypal business purposes. The visualization structures are under each specific heading for good reason. Each listed structure supports its parent purpose by virtue of how the human brain perceives, interprets and applies meaning to those structures. Although in some situations, some flexibility in graph selection is present, the listed groups represent best practices for conveying information effectively. Do you want to see values changing over time? Use a line graph. Are you interested in what expenditures make up your budget? Use a tree map or pie graph. If you want to compare sales figures for products, regions and quarters, use bar graphs.

In some situations, your information need might be more complex and the purposes might be more involved. In these cases, the visualization types can be combined to support these special needs. However, before investigating exotic options, consider clarifying your purpose or dividing the functionality into multiple, simpler visualizations. This is often more effective than trying to achieve everything with a single complex visualization.

In praise of bar graphs

In many situations, a bar graph works. Strictly speaking, vertical bars make “column graphs” and horizontal bars make “bar graphs”; however, in this paper, “bar graph” means both. Bar graphs are a simple, powerful and standard graph type. Therefore, they are easy to create and easy for your customers or other business users to understand. They are broadly applicable to very common data with the following forms:

- A main value and one category (such as sales, product)
- A main value and several categories (such as sales, product, quarter)

Bar graphs make it possible to represent and compare these basic values. Bar graphs are also flexible in how they facilitate understanding of assorted data perspectives. Note that position has meaning in the ordering of the bars. The example in Figure 1 could be sales per product (color), grouped by week. You can use this bar graph to emphasize and compare the per-week performance of each product.

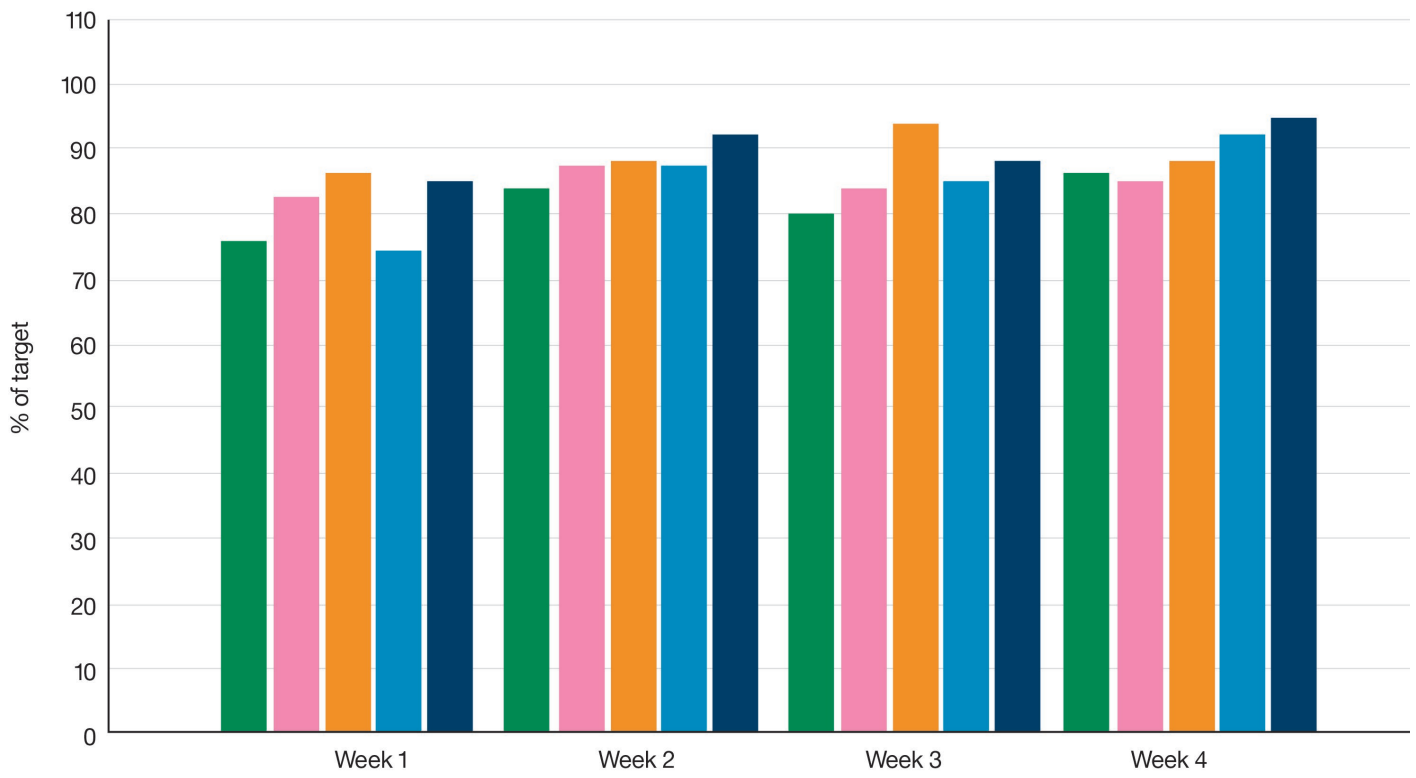


Figure 1: Bar graph that shows a grouping by week

Or, you could group the weekly sales by product (Figure 2); this would be better for comparing the performance per product over several weeks.

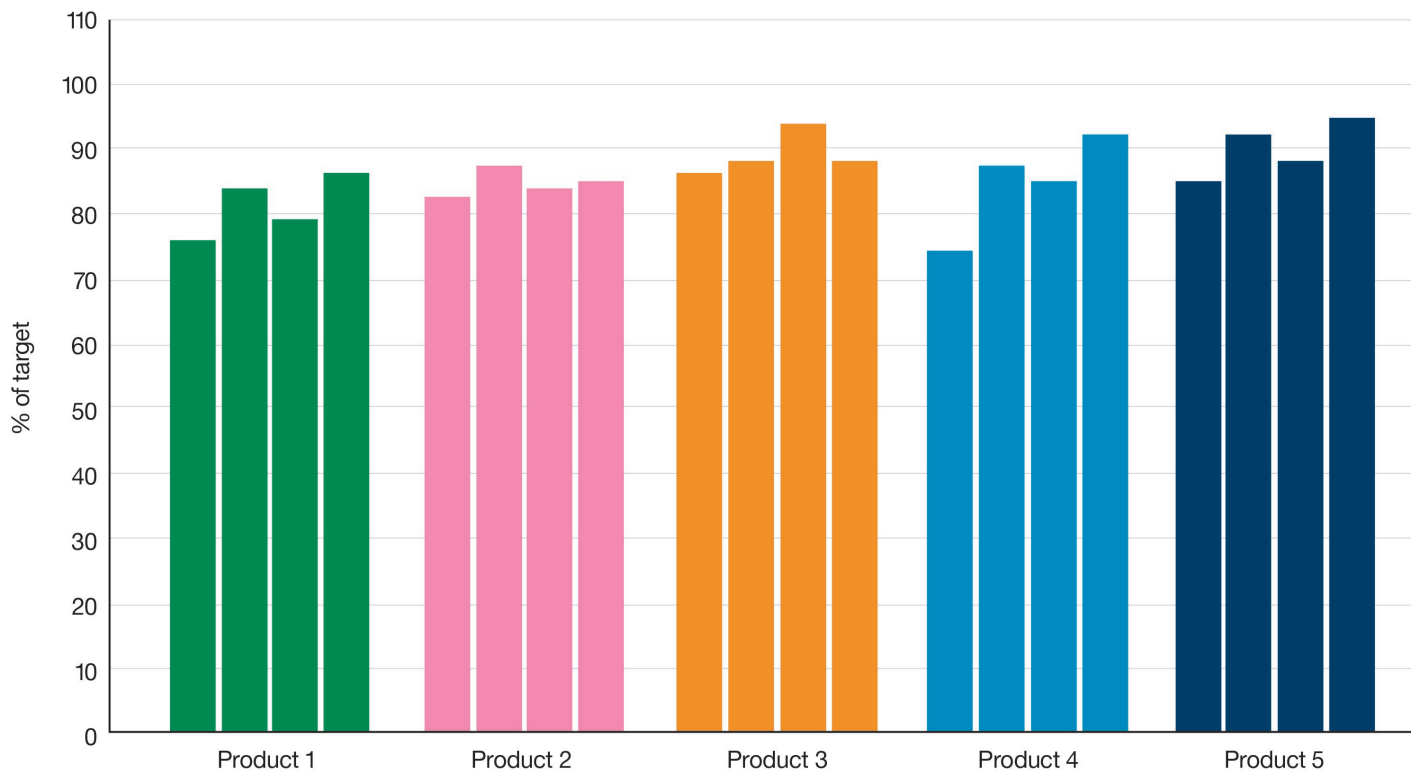


Figure 2: Bar graph that shows product grouping

If more subtle meaning is required, bar graphs can be annotated with points or lines to show targets or thresholds (Figure 3).

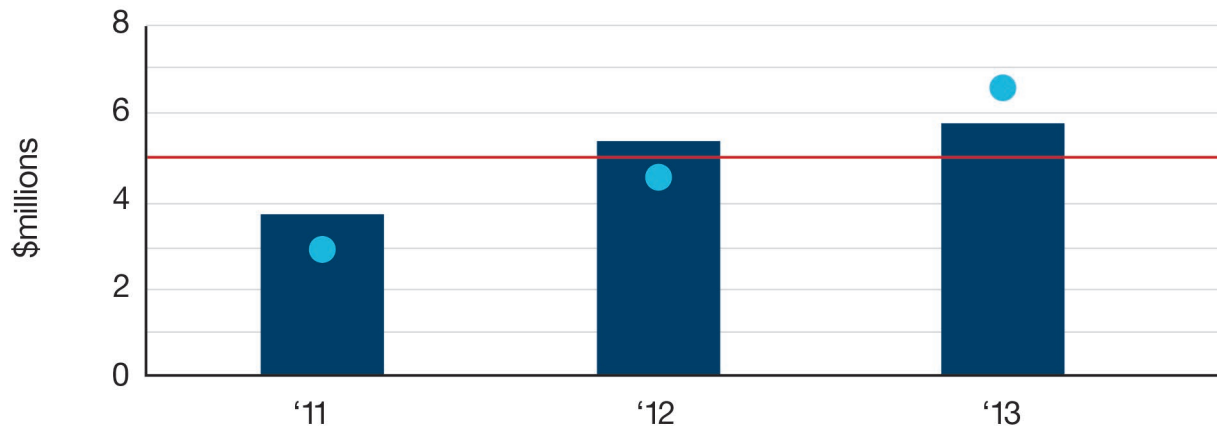


Figure 3: Annotated bar graph

Bullet graphs (Figure 4) are bars with shaded background regions to represent qualitative results.

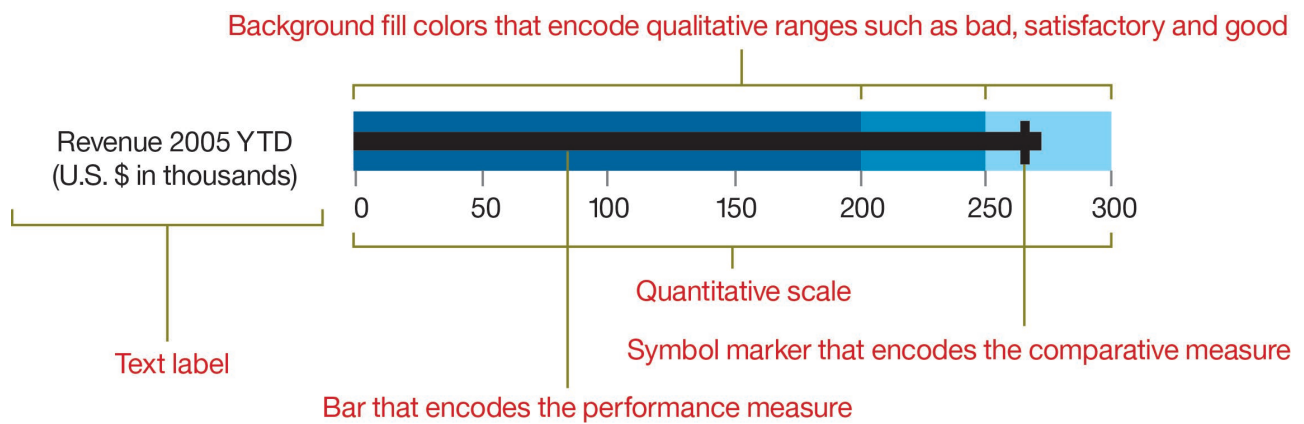


Figure 4: Bullet graph

Source: https://en.wikipedia.org/wiki/Bullet_graph

Line graphs, dot plots, scatter plots and matrices

Here are some other simple, useful, recommended graph styles.

For data that has a continuous relationship, usually one that changes over time, line graphs (Figure 5) are the standard structure. This is the type of graph you usually see when you are viewing data that varies continuously, such as temperature, employment numbers and stock prices. (Data with a continuous relationship should not be confused with an axis scale that allows any continuous value but does

not imply continuity in the data.) You can show continuous data with non-continuous graphs, such as separate bars to represent per-week sales. However, you really should not connect non-continuous data, such as separate products, with a continuous line graph.

Line graphs can get messy when a large number of lines are being displayed. Consider minimizing the saturation of any background colors or grids and using medium gray for most data lines while highlighting the lines of interest.

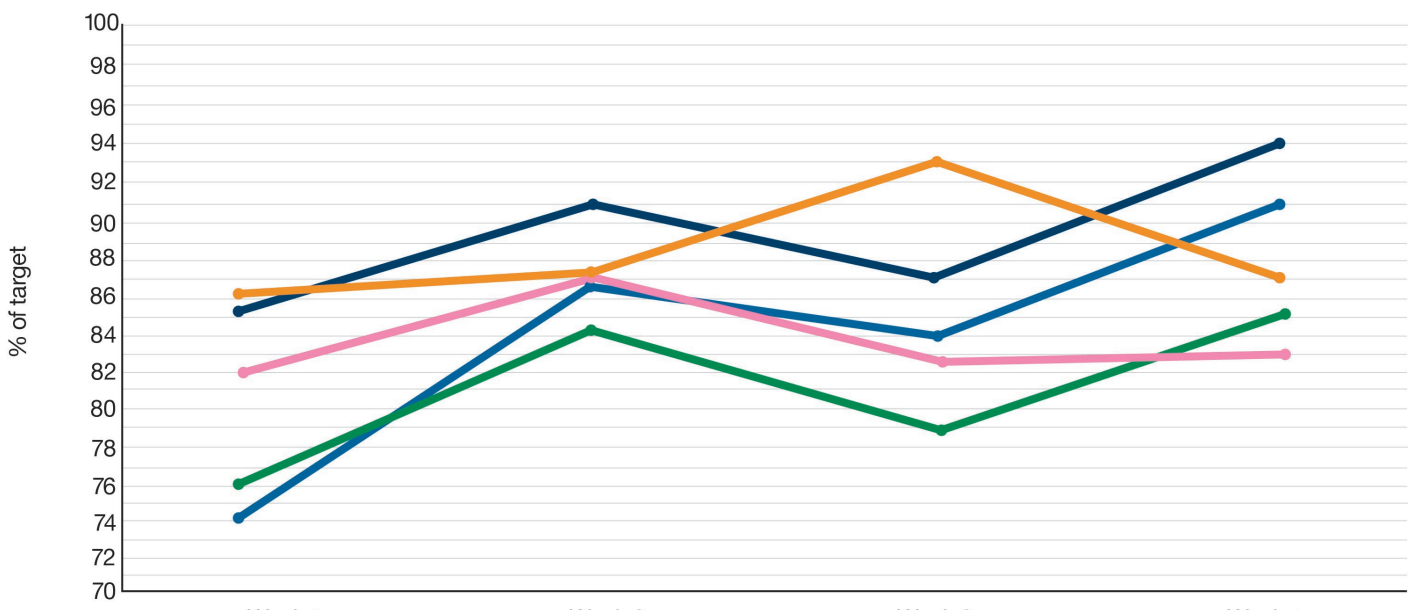


Figure 5: A line graph

Dot plots are like bar graphs without the bars. These plots rely solely on dot position to convey value, instead of length. They work very well for high-density display of data (Figure 6).

Note that the dot plot in Figure 6 displays the same information as the previous graphs in a much smaller space.

However, isolating single data points can be challenging if they overlap. Individual dots can easily use size to encode magnitude or color to show category. Dot plots typically have one continuous, quantitative axis (the y axis in this case) and an ordinal or categorical axis (the x axis in this case). Finally, dot plots can be used to show an overview of a distribution of values.

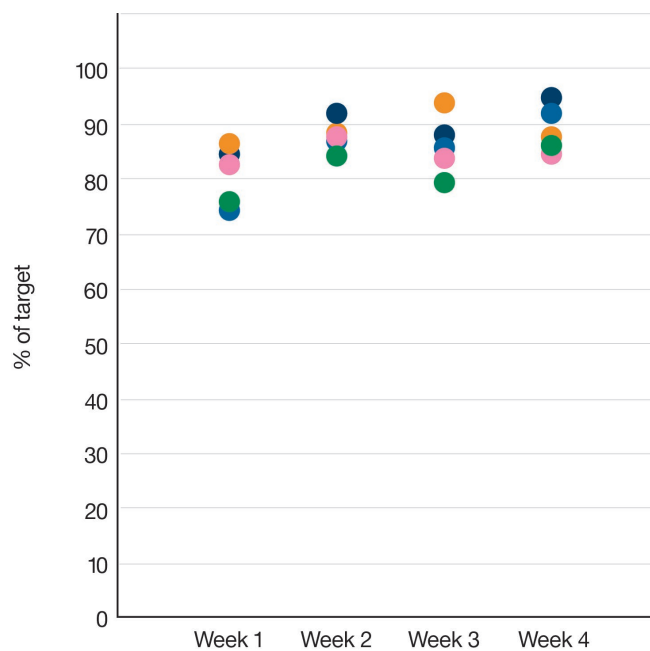


Figure 6: Dot plot that is being used for a high-density display

If you make a dot plot with two continuous axes, you have created a scatterplot. Scatterplots (Figure 7) are superior at revealing trends, relationships or correlations between data dimensions.

Figure 7 displays time-series data, although time is not on either of the major axes. Instead, each point is one value in the time series, and proximity ties the points together.

If time series data is shown and the data points are far apart, a line or ribbon may be drawn to connect the points in chronological order.

After the points are plotted into the two-dimensional space of the scatter plot, other standard encodings can be added; these commonly include size, color, shape and highlight. These encodings enable you to simultaneously display 4-6 dimensions of data. However, the plot can look cluttered if the encodings are applied without careful consideration.

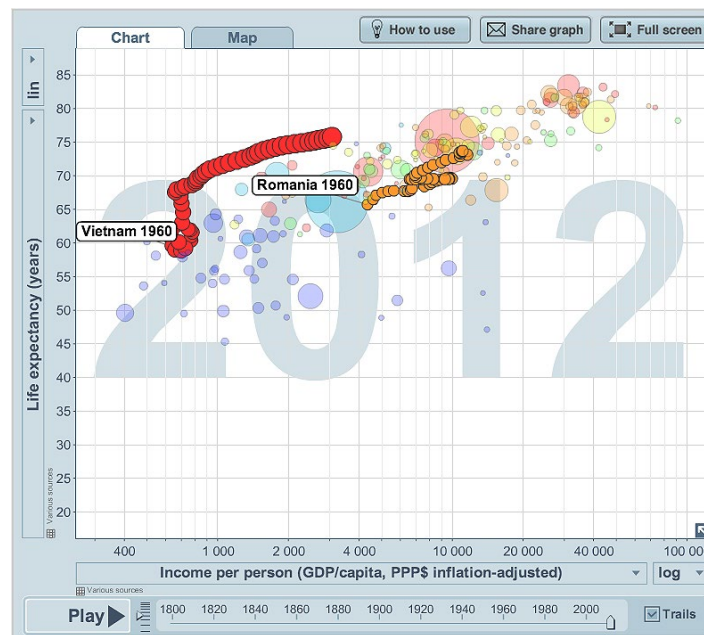


Figure 7: A scatter plot with time series data

Source: Free material from www.gapminder.org

If you have two categorical axes, you can create a simple matrix graph (Figure 8) or grid, which can accommodate values of any sort at the intersection: smaller graphs, specific quantities or binary values (battleship exists: yes or no).

Matrix graphs can be excellent for comparing arbitrary combinations of input factors, such as region and category. In this matrix graph, you can see values changing by time of day and day of week.

Twitter sample: ● drunk, drinking ● hungover, hangover

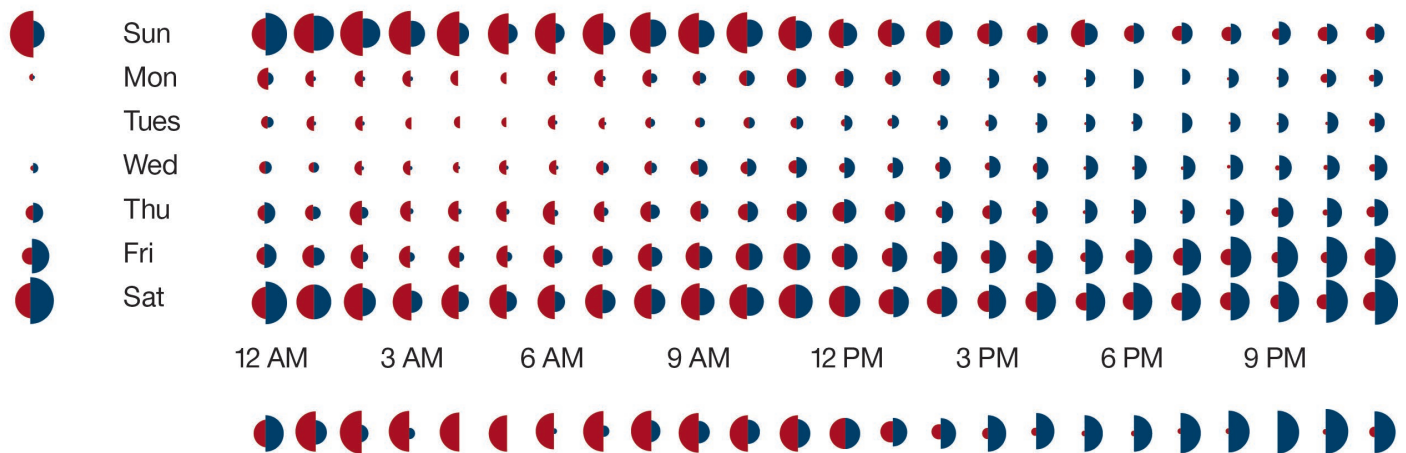


Figure 8: Matrix graph with values that change by time of day and day of the week.

Source: © 2013 LuckySort inc. Used with permission. <http://infomati.co/post/70121615444/graph-of-periodic-drinking-activity-and>

Composition: Pie graphs and tree maps

Graphs that show composition, by definition, demonstrate how the whole of a collection is made up of its component parts. The parts must be variations of the same sort of component, must be mutually exclusive, and must add up to 100 percent. For example, you can't make a valid composition graph out of sales and expenditures because they are different types of components (income and outlay) and don't add up to 100 percent.

The most common visualization of composition is the pie graph. Pie graphs can be valid and successful when a few conditions are met (Figure 9). Successful pies:

- Present a limited number of slices (10 is a rough maximum; the fewer, the better).
- Include only a few slices that are relevant for comparison, even if the pie actually contains more than those few slices.
- Apply to situations where precision is not required.
- Order slices from largest to smallest.

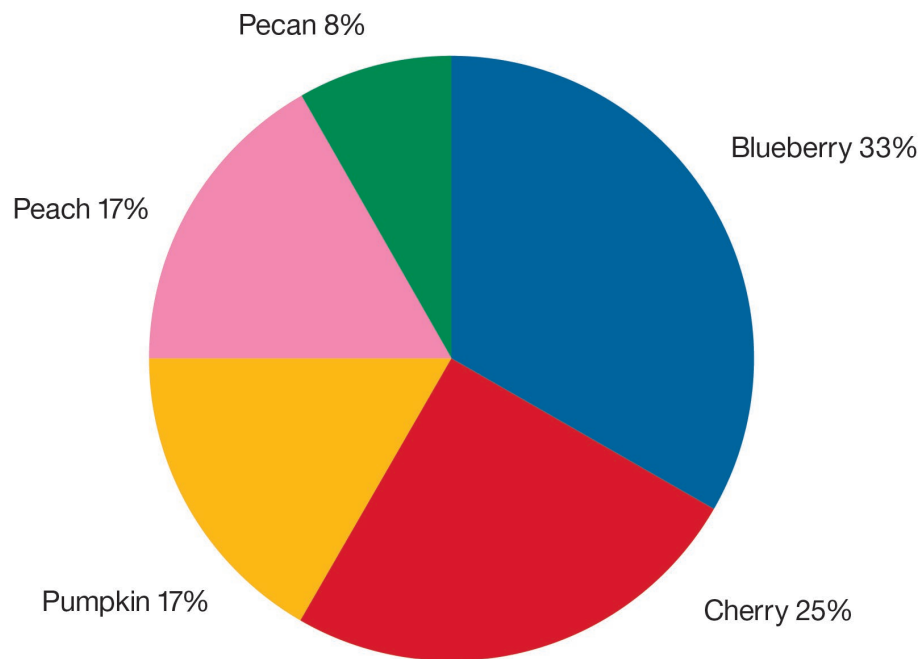


Figure 9: Pie chart

If precision is required, numerous slices are relevant or the number of slices is large, meaningful comparison becomes very difficult (Figure 10). In any of these situations, bar graphs are a better choice.

The reason that bars are better in these sorts of situations is because people are not as good at comparing angles as they are at comparing position and length. That said, the central angles on pie graphs are still somewhat useful; “donut graphs,” with the central angles missing, are even more challenging than pies when it comes to accurate perception.

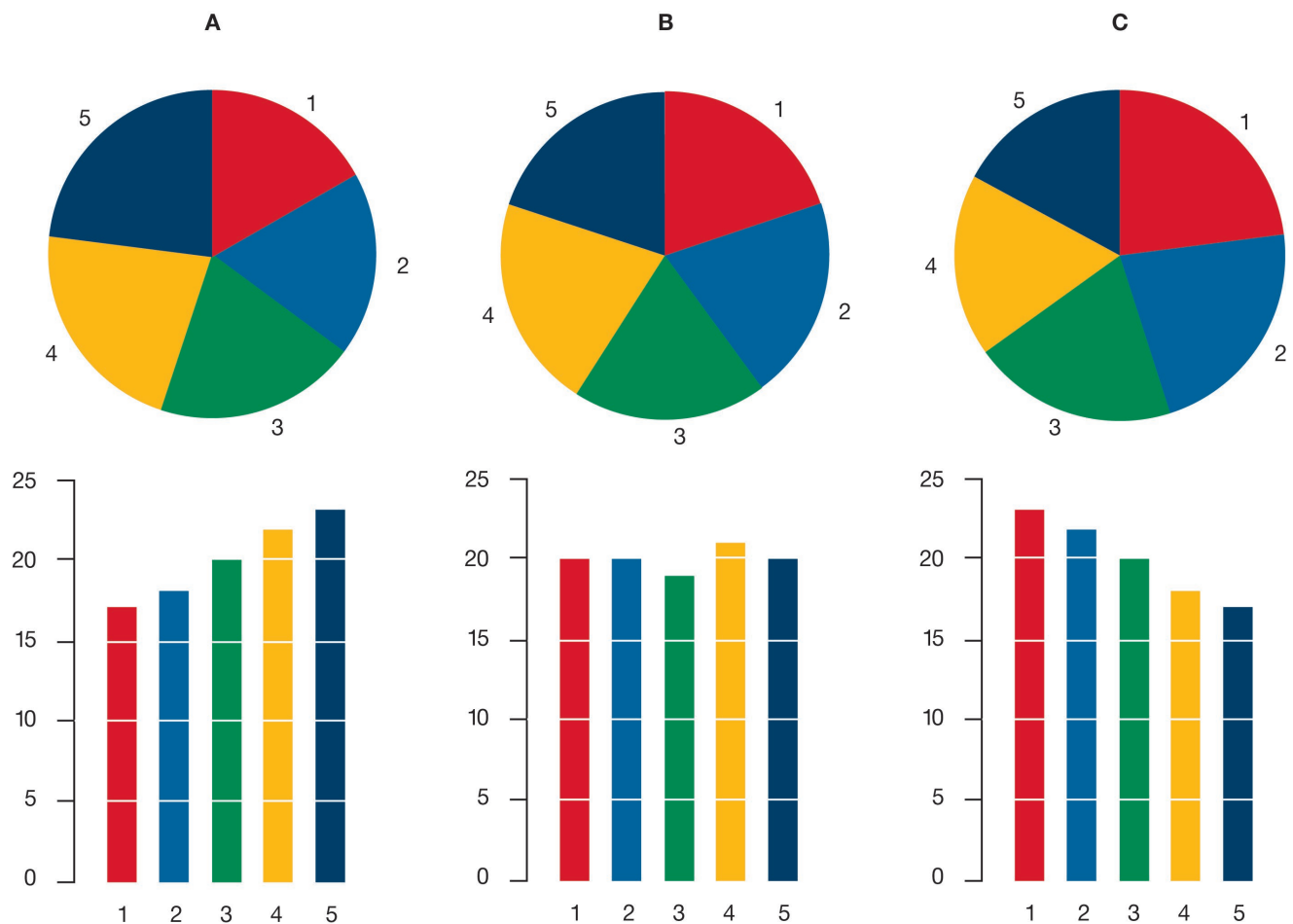


Figure 10: Complicated pie charts

Source: <http://en.wikipedia.org/wiki/File:Piecharts.svg>

Tree maps (Figure 11) enable representation of composition to be combined with hierarchy, so the composition of various levels of the hierarchy can be compared. Typically, tree maps have a few levels of hierarchy, use color to show categories and sometimes encode magnitude with intensity.

Although tree maps can show huge amounts of data, they also make precision comparisons difficult. Remember that Cleveland and McGill showed that humans are worse at comparing area than they are at comparing angles, length or position.



Figure 11: Tree map example

Source: <http://en.wikipedia.org/wiki/File:GradientGroupedTreemap.jpg>

Visualization structures to avoid

What visualizations should you not use? You should avoid structures in the following list because of the various ways they distort data or make conveying real information much more difficult:

- Radar charts (falsely imply meaning, unless you actually have cyclical data, like time of day)
- Circular bar graphs (potentially difficult to interpret intent, very challenging to compare curved lengths)

- Circular area graphs (extremely challenging to compare sections)
- 3D anything (massive distortion possible)

Figure 12 shows examples of some visualizations that would probably work better as a bar graph

In the situations where you might be considering these types of structures, you most likely should use a bar graph.

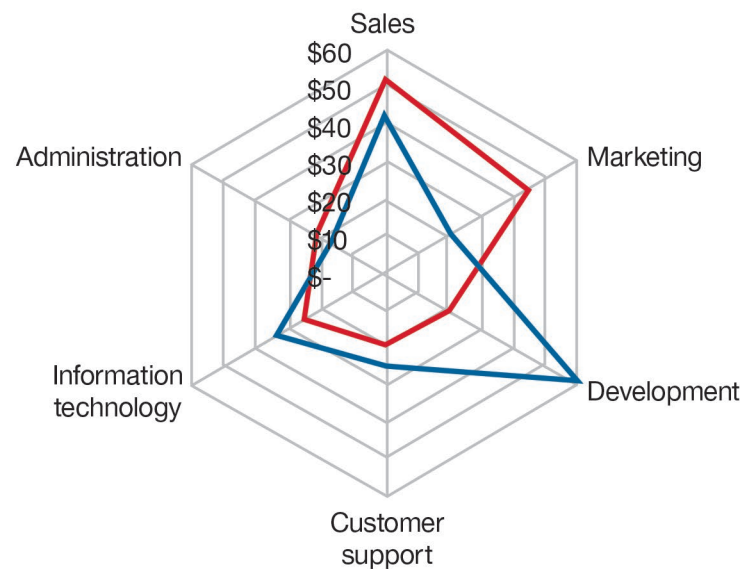
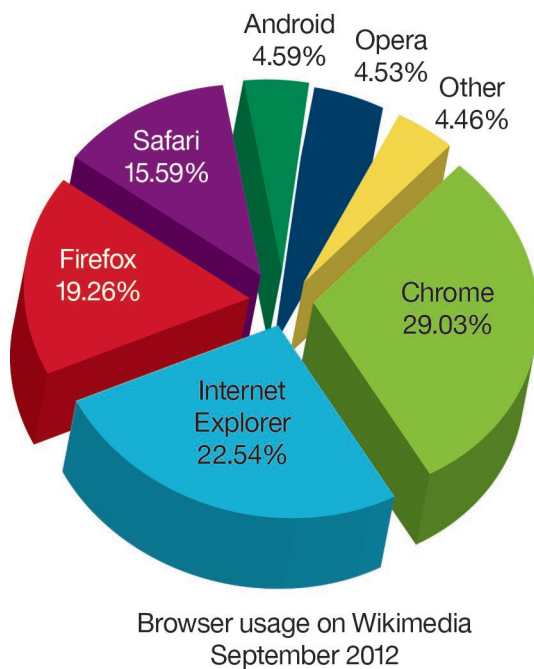


Figure 12: Pie chart and spider chart

Extensible visualization

You might have a handle on your data and know what type of structure you need, but the prepackaged visualizations that are part of your software or application might be holding you back. Perhaps they cannot provide you with the flexibility to create the specific type of bar graph, pie chart, matrix or other recommended structure that you need. Or, they might not include the more innovative and newest visualizations created to deal with the unique challenges presented by big data.

IBM has developed a solution that addresses big data, structure limits and prepackaged options. A visualization rendering engine, RAVE, is now available in the IBM® Cognos® family of business intelligence software, IBM business analytics tools and other IBM products. As a descriptive language, RAVE enables rapid development of new visualization types and flexibility to customize visualizations. RAVE underpins extensible visualization, a new capability in Cognos Business Intelligence. With extensible visualization, you can download and edit a wide variety of graph and chart templates whenever you need them. No longer do you have to wait for a new product or version release to provide you with new graph types.

Extensible visualization enables “pluggable” interactive visualizations that can be added between product releases. You can augment your reports—both general and active reports—with a growing collection of visualizations to meet your data and insight requirements. In addition, with the new IBM Visualization Customizer, you can edit and customize your visualizations to change properties, such as color or icons.

Extensible visualization and big data

The new and varied visualizations available through extensible visualization can be used to interpret and understand big data. The visualizations that are available as part of the IBM extensible visualization library can be used to encode networks of relationships between items such as products, individuals or groups over a very long period of time. For more information about how visualization can help you make sense of big data, see T. Alan Keahey’s white paper, “[Using visualization to understand big data.](#)”

Extensible visualizations and the IBM Visualization Customizer are available, at no charge, on [AnalyticsZone.com](#), including versions of each graph structure recommended in this paper.

Conclusion

The layout you use to frame your data in visualization is known as the structure; structure is important because of the way the human mind interprets placement. Because of this, selecting the right structure is of utmost importance to communicating your data effectively. Position is the most accurate property to use when you are representing quantities.

When you have made a considered decision about the structure of your visualization based on relevance and the information you need to convey, being able to use the structure you want is paramount. You should not be limited to a few prepackaged visualizations that are included in the release of your software and applications. Extensible visualizations from IBM make it possible for you to select just the right structure for your data — and that includes big data — and edit it so that it tells your audience what you want them to know in the best way possible.

About the author

Noah Iliinsky is currently a Visualization Expert at the IBM Center for Advanced Visualization. Noah strongly believes in the power of intentionally crafted communication. He has spent the last several years thinking, writing and speaking about best practices for designing visualizations, informed by his graduate work in user experience and interaction design. He is a frequent speaker in industry and academic contexts. He is the co-author of *Designing Data Visualizations* and technical editor of and a contributor to *Beautiful Visualization*.

About IBM Business Analytics

IBM Business Analytics software delivers data-driven insights that help organizations work smarter and outperform their peers. This comprehensive portfolio includes solutions for business intelligence, predictive analytics and decision management, performance management and risk management.

Business Analytics solutions enable companies to identify and visualize trends and patterns in such areas as customer analytics that can have a profound effect on business performance. They can compare scenarios; anticipate potential threats and opportunities; better plan, budget and forecast resources; balance risks against expected returns and work to meet regulatory requirements. By making analytics widely available, organizations can align tactical and strategic decision making to achieve business goals. For more information, see ibm.com/business-analytics.

Request a call

To request a call or to ask a question, go to ibm.com/business-analytics/contactus. An IBM representative will respond to your inquiry within two business days.



© Copyright IBM Corporation 2014

IBM Corporation
Software Group
Route 100
Somers, NY 10589

Produced in the United States of America
January 2014

IBM, the IBM logo, ibm.com and Cognos are trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the Web at "Copyright and trademark information" at www.ibm.com/legal/copytrade.shtml.

This document is current as of the initial date of publication and may be changed by IBM at any time. Not all offerings are available in every country in which IBM operates.

THE INFORMATION IN THIS DOCUMENT IS PROVIDED "AS IS" WITHOUT ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT. IBM products are warranted according to the terms and conditions of the agreements under which they are provided.

1 W. S. Cleveland and R. McGill, "Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods." *Journal of the American Statistical Association*, 1984, 79:531-554.



Please Recycle