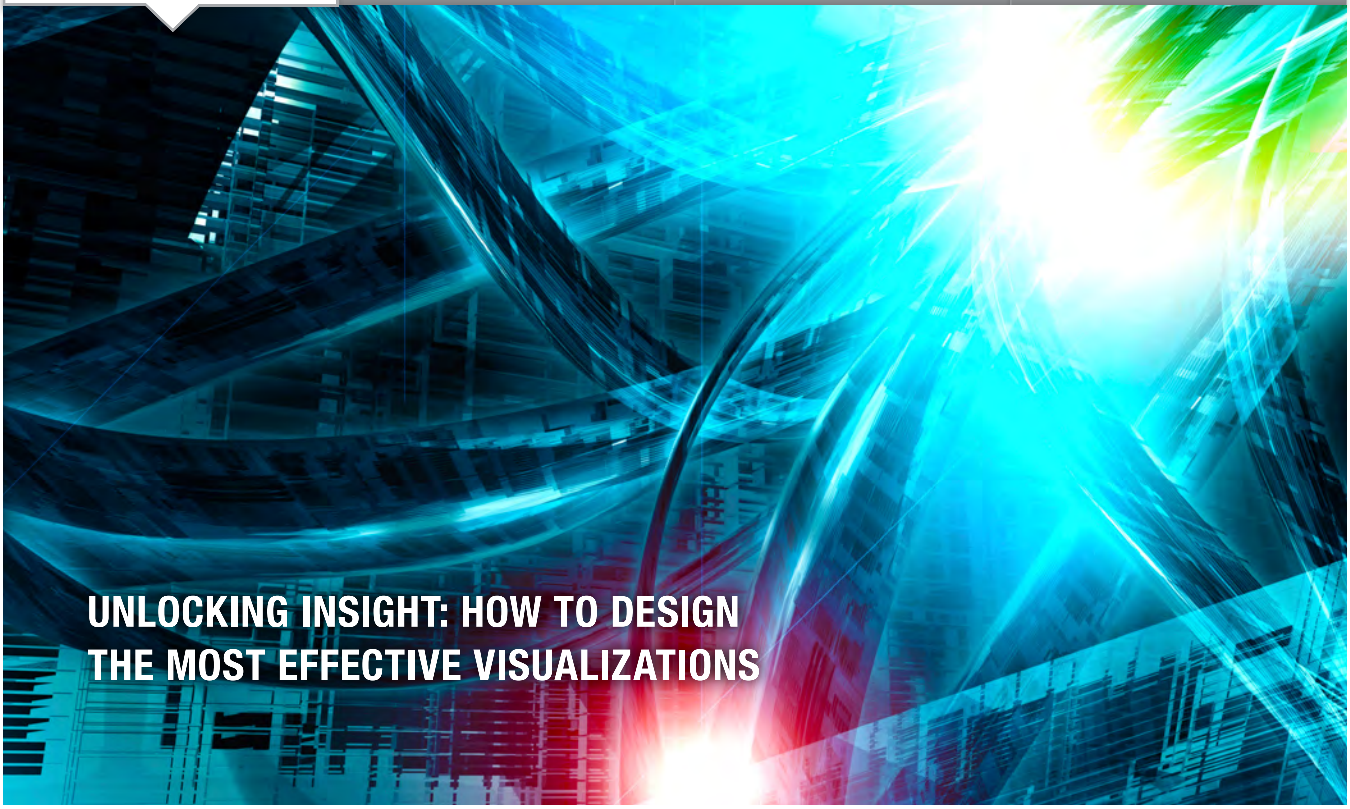
TDWI E-BOOK

DECEMBER 2013

TOP CONSIDERATIONS FOR EFFECTIVE VISUALIZATIONS

tdwi.org

tdwi
THE DATA WAREHOUSING INSTITUTE



UNLOCKING INSIGHT: HOW TO DESIGN THE MOST EFFECTIVE VISUALIZATIONS

From mobile interfaces to defining a purpose, two experts from IBM Business Analytics offer the best practices you need to design visualizations that maximize insight.

In a world of big data, the more you can make the results of big data analytics understandable, the more your users will prosper. Dr. T. Alan Keahey and Noah Iliinsky from IBM Business Analytics explain the key factors in creating visualizations that maximize your users' grasp of their data and the role of advanced visualizations in working with and understanding big data.

TDWI: What are the key factors to creating an effective visualization?

Iliinsky: The most successful visualizations have a few things in common. They have a clear purpose; they include only relevant, focused content; and they present the content in a manner that reveals and highlights the interesting relationships in the data.

To achieve greatness, there's a design process that must be followed. Much of the design effort happens long before the point of sketching possible layouts and picking fonts and colors. First, a purpose must be defined. Second, content that supports that purpose must be selected. Third, an appropriate layout or graph type must be chosen. After that, other choices regarding elements such as fonts, colors, and icons can be made.

Here are the steps in the process:

1. Purpose (why?)
2. Content (what?)
3. Structure (how?)
4. Everything else
5. Iterate (the first option is rarely the optimal solution)

The first thing to focus on is purpose. Ask yourself the following questions: What is the reason for creating this visualization? Who will be using it? What do they need to learn from it, and what actions will that cause them to take? How will it be consumed: print, big screen, mobile?

In order to be successful, you must make the information consumer happy, and in order to do that, you must understand who they are and what they need. You can't hit a target that doesn't exist. The more specific your purpose is and the better you can articulate it, the more likely it is that you'll be able to craft a solution that is valuable to the user.

For example, a not-so-useful purpose is "show the data." Which data? Why? To whom? On which device? A much better purpose is "show the slowest connections on the field techs' iPads, so they can upgrade the hardware." This latter, more focused purpose tells you what data to include, implicitly what to exclude, who cares about it, and how they'll be consuming it. Describing the purpose at that level of detail gives you huge amounts of guidance for the design process and improves your chances of being successful.

A good practice is not to start by asking "What data should we show?" but rather "What actions do we need to enable and what questions do we need to answer?" Once you have those more abstract questions answered, focusing on the true need and selecting the relevant data will be much easier.

The next major consideration to keep in mind is that when it's time to look for specific data points, any extra data is the same as noise. Resist the temptation to be comprehensive and show everything; instead focus on serving up concise, relevant answers. If there is a mandate to include everything, consider options to filter or hide the bulk of the data and let your customer focus on one specific facet at a time.

After a purpose has been established and relevant content selected, you can move on to selecting the appropriate graph type and encoding your data visually.

When trying to select the most appropriate visualization for the data set I am representing, what considerations should I keep in mind?

Iliinsky: There is a scientific basis for selecting the right graph types and visual encodings for your data. Our eyes and brains are extremely good at perceiving some differences and not very good at all at comparing others. For example, we're excellent at comparing relative position or length, but not as good at comparing brightness or angle, and we're terrible at comparing arc length. Some properties are interpreted by our brains to be ranked or quantified (such as position or length); others imply categorization (color or shape) or relationship (connecting lines or enclosing boxes).

Because our brains have these inherent interpretations of visual properties built in, we can be most successful in our designs when leveraging the meanings that our brains are going to impose on what we see. These built-in interpretations are why some graph types are strongly favored for displaying certain data types.

The classic visualization types—bar, line, and pie graphs as well as scatter plots—are classics because they reveal certain types of data and relationships very well:

- Compare values: Bar graph
- Show change over time: Line graph
- Show relationships between values: Scatter plot
- Show composition of a whole: Pie graph

Of course, often we may need to combine, nest, adjust, and mix these fundamental graph types, but they are a great place to start.

The most important two things to keep in mind when it comes to encoding are:

- Use position for your most important data and relationships
- Use color for category, not quantity

For much more detail on picking encodings to go with your data, download my white paper, "Choosing Visual Properties for Successful Visualizations."

What are the most common pitfalls that lead to ineffective visualizations?

Iliinsky: The most common pitfalls tend to be directly related to failures of the points above. Let's examine each case.

Lack of clear purpose, at best, yields a visualization that is a blunt tool, perhaps capable, but not well suited to any particular task. Symptoms of this include visualizations that require many steps to reveal answers that are interesting, and audiences who look at the visualization and say "So what?"

At worst, lack of purpose leads to visualizations that are incoherent messes. They tend to be designed by whim, with no awareness of task or design principles. These may be pretty but are often useless.

The second major failure mode is simply including too much data and making the interesting answers hard to find. The more work you do up front on behalf of your customer, the less effort they'll have to expend, and the happier they'll be.

The third failure mode is picking the wrong structure; good data displayed poorly is frustrating. Instead of going for radical and novel visualizations, start with the basics and only depart from tried and true formats when your solution is clearly worth the extra effort to you and to your customer.

Finally, don't use 3-D. At best it distorts the data, and at worst it distorts and hides data. If anyone requests graphics that are flashy, tell them you feel clarity of data is the strongest statement, and a sleek, modern, flat look is best for that.

What is the role of visualization in working with and understanding big data?

Keahey: Direct visualization of big data "in the raw" is almost never an effective strategy. There are two main reasons for this: first is the sheer scale of the big data and second is the complexity of it. In terms of sheer scale, how can we possibly fit a terabyte (10^{12}) or more of data into a screen having at most a few million (10^6) pixels? The general answer to that question is that we cannot do it very well. Basic data reduction and navigation techniques (such as sampling and filtering) can help manage the scale of the data, but they result in significant data loss and potential loss in information structure.

Data complexity is another factor that places limitations on our ability to directly visualize big data sets. There is a relatively small category of big data applications that involve relatively uniform data structures such as sensor networks and traffic monitoring. However, for many other categories of "big data," the information has a high degree of structure (such as Walmart's daily data) and that structure is a first-order property of the data that needs to be preserved during the analysis.

Where visualization can play the most effective role in big data analysis is in the use of specialized views that illuminate particular facets of the data. This can be done with moderately scaled visual representations that can show around 10^3 or 10^4 items on the screen simultaneously. One good example of this is the calendar heat map, which can be used to provide "at a glance" overviews of activity over time. Unlike the standard line chart view, the calendar heat map can bring to light daily, day of week, weekly, monthly, and even yearly patterns in a single compact representation.

More broadly speaking, visualization can play a key role in understanding big data sources. However, the specific visual metaphors should be carefully chosen to support the required analytic tasks. In addition, there are huge advantages to systems that integrate the visualization and analytics components of the solution so that the two are able to work together to provide a richer and deeper view of the data.

For more information on big data visualization, please refer to my IBM white paper, "Using Visualization to Understand Big Data."

Can you give a real-world business use case where richer advanced visualization has been used effectively?

Keahey: We recently created a novel visualization embedded in IBM's Global Process Services' (GPS) Compliance Analytics Toolkit (CAT). CAT, which is available as a service from GPS's supply chain management unit, leverages IBM SPSS, IBM ILOG, IBM Cognos, and Emptoris assets to allow the chief financial officer or chief procurement officer to understand the entire corporate supplier spending with actionable improvement recommendations. Our novel visualization allows the viewer to see in a single image the same information that formerly required navigation through a hierarchy of tabbed reports and mental accumulation of the main action areas within each page of the hierarchy.

With the new system, a CFO can quickly assess at a glance which areas have significant spending that is not in compliance with spending policies, and take immediate corrective measures. Furthermore, the visualization is interactive, so the CFO can make selections within it to get more detailed information in tooltips and sidebars without having to leave the overview page. This results in significant efficiency improvements due to reduced context switching and a better understanding of the global priority areas for action.

A screenshot of this visualization is shown here. The CAT tool enables a breakdown of company expenditures and segments the spending into several groupings. Each grouping is composed of many individual spending categories. Each category and grouping is scored by the analytics in terms of overall amount spent and the percentage that is in compliance with company policies. This visualization encodes funds spent by the size of the bubble and the percent that is noncompliant by color. The left bubble represents the spending for the entire company, the middle column of bubbles represents the analytical groupings, and the right column of bubbles shows the individual categories within each grouping.

How is the increasing use of mobile devices likely to affect the design and consumption of visualizations in the future?

Keahey: Obviously, the smaller display screens on tablets and smartphones will have a limiting effect on the amount of information that can be displayed. However, the relation between screen size and information content size is not a trivial linear one.

In terms of pixel count, much progress has been made in display hardware to increase pixel density. In the case of Apple's iPad, the initial displays had 1024x768 pixels, which works out to about 786K pixels at 132 pixels per inch (PPI). The retina display quadrupled that resolution to 2048x1536 pixels, which gives just over 3 million pixels at 264 PPI. However, 264 PPI is nearing the limits of perceptibility, particularly when taking into account individual variances in visual acuity. As a consequence, many of the additional pixels provided by the high-resolution displays serve a greater purpose to increase the fidelity of existing visual elements rather than to visually encode new ones.

The larger effect of this is that we are at or near saturation point for the amount of pixels that can be effectively distinguished in a tablet-size display, and visualization designers cannot rely on future hardware improvements to increase this limit.

Another significant factor in mobile involves the touch interfaces. Apple Human Interface guidelines call for a touch-point target size of around 44 points, which is equivalent to 0.61" or 15.5mm on each side. If we overlay a grid with those sized squares on a traditional 9.7" diagonal tablet, that gives us a touch grid of only 12x9 or 108 points. With some careful design, we can increase this number somewhat, perhaps doubling each dimension for a total of 400 touch points.

These two factors of asymptotic pixel density and limited interactivity resolution are going to place a premium on the effective use of screen real estate going forward. This, in turn, will lead toward a decluttering of the visual interface, eliminating many of the "chart junk" and nonessential elements. One key area where I see this happening is in the repurposing of the visualization metaphor as a tool not just for showing the information but also as the principal navigation mechanism for exploring that same information. This approach helps to eliminate the need for external controls (buttons, menus, etc.) and, if done correctly, allows for a fluid and intuitive means of information navigation.

How is IBM equipping report and dashboard authors and consumers with flexible visualization options to gain maximum insight from their ever-increasing amount and variety of data?

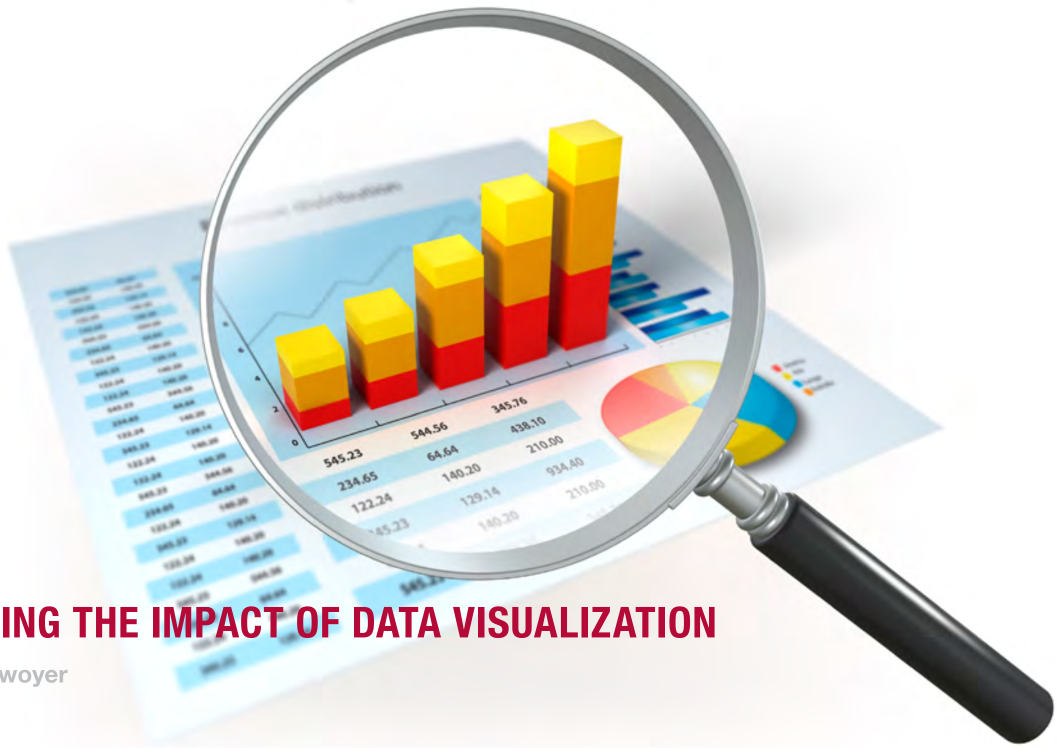
Keahey: IBM continues to provide rich and innovative visualization capabilities to business analyst and line-of-business users across our Cognos Business Intelligence portfolio. As we have discussed here, with the different types and amounts of data accelerating every day, the right visualization is key to unlocking insight. A visualization that was ideal for the data in your organization today may not yield the insight with the new data of tomorrow.

New, extensible visualization capabilities built into IBM Cognos Business Intelligence V10.2.1.1 solutions unleash report authors and business users from a static library of charts. No longer are users locked into only the in-product visualization options. Now you can easily augment your reports—both general and active reports—with a growing collection of visualizations to meet your data and insight requirements. With the new IBM Cognos Visualization Customizer, visualizations can be edited, including changing fonts, colors, or even icons to meet the most specific requirements.

Through the new Visualization Marketplace on IBM Analytics Zone, report authors can choose from over 30 visualization options, ranging from radar charts to heat maps and area charts. New visualization options will be regularly added to the Marketplace. You'll always find just the right visualization just a simple download away.

Dr. T. Alan Keahey has played a leading role in the research and development of highly innovative information visualization systems for close to 20 years. His experience spans a wide range of environments, including national labs research scientist, research director at a Lucent Bell Labs spin-off, and founder of his own visualization R&D company. He thrives on anchoring connections between the capabilities created in research environments and the real-world needs of business customers. Alan is currently a visualization science and systems expert at the IBM Business Analytics Office of the CTO, where he works to drive the innovative and effect use of visualization across IBM, including major partners and customers. He blogs about visualization and all things visual at HolisticSofa.com.

Noah Iliinsky strongly believes in the power of intentionally crafted communication. He has spent the last several years researching, writing, and speaking about best practices for designing visualizations. Noah works as a visualization expert at IBM's Center for Advanced Visualization. He is the co-author of *Designing Data Visualizations* and technical editor of, and a contributor to, *Beautiful Visualization*, published by O'Reilly Media. He has a master's degree in technical communication from the University of Washington and a bachelor's degree in physics from Reed College.



EVALUATING THE IMPACT OF DATA VISUALIZATION

By Stephen Swoyer

According to a Best Practices Report from TDWI Research, adoption of data visualization technologies is growing by leaps and bounds. Even so, most users continue to interact with data via static tables. That's changing, however, as data visualization is increasingly perceived—by stakeholders at all levels of the enterprise—as a high-value investment.

According to a Best Practices Report from TDWI Research, adoption of data visualization technologies is growing by leaps and bounds.

Even so, write industry veterans Wayne Eckerson and Mark Hammond, most users continue to interact with data via static tables: nearly two-thirds (65 percent) of users consume data primarily in the context of tables and text versus 35 percent who interact chiefly with visual charts or graphs, according to a companion survey conducted by TDWI Research.

"Although some data is better suited to non-visual rendering, it's clear that old habits die hard," write Eckerson and Hammond in *Visual Reporting and Analysis: Seeing Is Knowing*, part of TDWI Research's Best Practices Report series. What's surprising, the duo notes, is that most shops recognize that static text probably isn't the best way to present—to say nothing of consume—many kinds of information, yet they continue to do so anyway.

As Eckerson and Hammond write, it's a case of old habits dying hard. Very hard.

"Many executives and managers, for example, have run operations for years using text-based budgets and plans created in Excel," they write. "[B]usiness analysts, who are generally Excel jockeys, are often more comfortable with grids of data because that is how they've always interfaced with information."

Even so, one out of three—or 35 percent—is a good start.

Visual analysis isn't necessarily warranted in all cases, of course. Eckerson and Hammond recount the experience of Bruce Yen, director of business intelligence (BI) with GUESS Inc., a prominent designer and manufacturer of clothing. "We showed bar charts on how sales broke down by department, but some users said they could simply look at the numbers and get the same answer," Yen told the duo. "Maybe the data and analysis wasn't so complex that it needed a bar chart or some other type of visualization."

Insufficient complexity is just one barrier (or—as in Yen's example—reasonable objection) to the use of data visualization.

In many cases, visual interfaces suffer from poor design—for example, interfaces that frustrate or stymie users by forcing them to work harder to get at information in context—or what Eckerson and Hammond call "visual overload."

Over time, users tend to prefer to expose more data via their visual interfaces. *Initially*—that is, when they're first introduced to a tool—they're intimidated by too much data. It's best to start small and introduce additional features or functionality piecemeal—or, better still, to permit users to serve themselves. "Visualization tools that expose data and functionality on demand will have higher rates of adoption among users," the authors explain.

In practice, Eckerson and Hammond note, data visualization is perceived as a high-value technology investment. For example, nearly three-quarters (74 percent) of survey respondents assessed the impact of data visualization on business insight as "very high" or "high." Less than one quarter (23 percent) judged it "moderate."

Similarly, more than two-thirds (67 percent) of respondents said that data visualization has had a "very high" impact on user productivity. An additional 30 percent described its impact as "moderate"; just 3 percent put it at "low/very low."

"[D]ata visualization significantly improves business insights and user productivity (that is, accelerates time to insight) and increases user adoption of BI tools," according to the report. "These benefits help explain the growing popularity of data visualization in corporate environments."

Finally, visualization can help sell a solution, too. Although 70 percent of respondents said that data visualization had had a "very high" or "high" impact on past dashboard purchases, nearly four out of five (79 percent) expected it to have a similar impact on *future* purchases. "Once data visualization gains a beachhead in an organization and BI managers see the impact it has on user insights and productivity, its influence spreads wide and far," they write.

Stephen Swoyer is a contributing editor for TDWI.

IF DATA VISUALIZATION WERE CARS

By Dr. T. Alan Keahey

Data visualization is a science that has made a lot of progress over the centuries, and particularly over the last 100 years.

During that time, a few specific types of data visualizations have become very widely used. There have been many volumes written about how to choose the correct visualization for each task; however, I think it is also interesting to compare them to another well-known technology that has evolved significantly over the last 100 years: the automobile.

Just as with choosing the right vehicle for the given route and payload, it is important to choose the right visualization "vehicle" to carry your data to the viewer.

Bar Chart = Pickup Truck: The basic bar chart is capable of conveying a wide range of data types and can do it in efficiently sized chunks. This is a working man's visualization that can often get the job done all by itself. When in doubt, a bar chart is often a good place to start for your visualization needs.

Line Chart = Municipal Bus: Line charts are similar to bar charts, except that the horizontal axis always refers to time steps. This is similar to a bus or delivery van on a fixed schedule. They can haul a lot of the same data as a bar chart, but they do it at a fixed sequence of stopping points.

Scatter Chart = Larger Courier Truck: Scatter plots can show everything that bar and line charts show, but they can also show much higher volumes of data and different types of data. These are the serious haulers to use when you have a lot of data to view.

Bar/Line/Scatter Chart Matrices = Synchronized Fleet: Sometimes you just have more information than you can fit in a single chart. Combining multiple charts can let you introduce additional dimensions into your view. As with courier fleet control, you have to be careful in your planning to ensure that the charts are well synchronized to provide complete coverage without chaotic overlap.

Statistical Chart = Sporty Roadster: Specialized statistical distribution charts like the box-and-whiskers plot allow you to look at the basic properties of the data without having to view every single element. Like a nimble roadster, these can be a bit tricky to get the most out of, but they allow you to move through the data quickly and get a good feel for the curves in the road.

Choropleth Map = Trolley Car: A choropleth is a map that uses color to indicate the data values in specific regions. These are a good choice when you know that the geographical relation is a primary data property that you want to convey, just as city planners may choose a trolley car route to cover the most critical destinations.

Network Diagram = Race Car: Network diagrams can be sexy and fun to look at. They engage the viewer and draw him or her into wondering about the information within. They have a great deal of expressive power; however, they also have limitations. Network diagrams are highly dependent upon the choice of the layout algorithm being used. A layout for one application may be completely inappropriate for a different application. Similarly, race cars need to be set up uniquely for different track layouts. A Formula One street racer would fare poorly on an oval dirt track.

Pie Chart = Horse and Buggy: The venerable pie chart does almost nothing that cannot be done better with the more efficient bar chart. Some conventional wisdom has it that pie charts should be used to show “parts of a whole,” but in most such cases a bar chart will do a better job of that. The one possible exception is when showing that one or more parts add up to more than half of the total (however, even that has restrictions). These can be nice to look at, but are terribly inefficient.

3-D Bar/Line/Pie Chart = Horse-Drawn Automobile: In the early days of the automobile, some manufacturers created hybrid vehicles that could run on either gasoline or be drawn by horses. The result was an abomination that did poorly in either mode. There is no reason to add 3-D glitz to these charts, as it will make them harder to read values. If you would like to improve the aesthetics, you should instead explore colors, shading and textures to create a more pleasing visualization.

Dials and Gauges = The Edsel: Dials and gauges are often used in data dashboard design to convey a feeling of being in situational control. Usually these visual elements are “all show and no go,” similar to the huge fender flares and chrome elements of the failed Edsel car from the 1950s. While some limited use of dials and gauges can be justified in special situations (such as environmental monitoring), it is important to limit their impact so that they do not overwhelm with clutter and wasted space.

There are many other types of visualizations out there to choose from, and different types of vehicles to compare. While it is sometimes possible to transport a sheet of plywood strapped to the roof of a sedan, it’s rarely the best choice. Next time you are choosing a visualization method for a data set, consider carefully if you are matching the right chart for the data and viewing task.

Dr. T. Alan Keahey has played a leading role in the research and development of highly innovative information visualization systems for close to 20 years. His experience spans a wide range of environments, including national labs research scientist, research director at a Lucent Bell Labs spin-off, and founder of his own visualization R&D company. He thrives on anchoring connections between the capabilities created in research environments and the real-world needs of business customers. Alan is currently a visualization science and systems expert at the IBM Business Analytics Office of the CTO, where he works to drive the innovative and effect use of visualization across IBM, including major partners and customers. He blogs about visualization and all things visual at HolisticSofa.com.

www.ibm.com

tdwi.org

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 or analyticszone.com/visualization.

TDWI, a division of 1105 Media, Inc., is the premier provider of in-depth, high-quality education and research in the business intelligence and data warehousing industry. TDWI is dedicated to educating business and information technology professionals about the best practices, strategies, techniques, and tools required to successfully design, build, maintain, and enhance business intelligence and data warehousing solutions. TDWI also fosters the advancement of business intelligence and data warehousing research and contributes to knowledge transfer and the professional development of its members. TDWI offers a worldwide membership program, five major educational conferences, topical educational seminars, role-based training, on-site courses, certification, solution provider partnerships, an awards program for best practices, live Webinars, resourceful publications, an in-depth research program, and a comprehensive website, tdwi.org.